WIRELESS IMPLANTABLE ICs FOR ENERGY-EFFICIENT LONG-TERM AMBULATORY EEG MONITORING

AL FREEMAN

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE

GRADUATE PROGRAMME IN ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE

YORK UNIVERSITY

TORONTO, ONTARIO

JUNE 2020

ABSTRACT

This thesis presents the design, development, and experimental characterization of wireless subcutaneous implantable integrated circuits and systems for long-term ambulatory EEG monitoring. Application-, system- and circuit-level requirements for such a device are discussed and a critical review of the state-of-the-art academic and currently available commercial solutions are provided. Two prototypes are presented:

The first prototype presented in Chapter 2 is an 8-channel wireless implantable device with a 2.5×1.5 mm^2 custom-designed integrated circuit implemented using CMOS 180nm technology at its core. The microchip is fabricated and the measurement results showing its efficacy in EEG signal recording in terms of input-referred noise, voltage gain, signal-to-noise ratio, and power consumption are presented. The chip is implemented together with a BLE 5.0 module on the same platform. Our vision and discussions on biocompatible encapsulation of this system, as well as its integration with a microelectrode array as also provided.

The second prototype, also implemented in CMOS 180nm technology and presented in Chapter 3, employs a novel EEG recording channel architecture that enables long-term implantation of EEG monitoring devices through significant improvement of their energy efficiency. The channel leverages the inherent sparsity of the EEG signals and conducts recording in an activity-dependent adaptive manner. Thanks to the proposed fully dynamic spectral-compressing architecture, the recording channels power consumption is drastically reduced. More importantly, the proposed architecture reduces the required wireless transmission throughput by more than an order of magnitude. Our test results on 10 different patients' pre-recorded human EEG data shows an average of 12.6× improvement in the device's energy efficiency.

DEDICATION

Dedicated to the loving memory of my father Dr. Ali Erfani who inspired me to be a better person. You will always be missed.

ACKNOWLEDGEMENT

I wish to express my deepest gratitude to my supervisor Professor Hossein Kassiri for his patience, recommendations and guidance. His intellectual and financial support has been tremendous help during this project and I feel very privileged having worked with him during the course of this program.

I would also like to thank my supervisor committee and defense committee members, Professor Ebrahim Ghafar-Zadeh & Professor Eleftherios Sachlos for reviewing this thesis and providing me with their valuable feedback.

I would like to thank the technical and administrative staff and faculty members of Lassonde School of engineering for their assistance and support.

I wish to acknowledge and thank my colleagues and friends Tayebeh Yousefi, Alireza Dabbaghian, Mansour Taghadosi, Tania Moeinfard and Fatemeh Eshaghi for their continuous help.

My most profound gratitude goes to my beautiful wife Miriam Freeman and my loving son Brayden Freeman who patiently supported me throughout the years. You have been the inspiration for my every effort.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

Introduction and Motivation

1.1 Motivation and Objective

The ever-increasing demand for understanding human brain functions and dysfunctions requires methods and tools for monitoring brain activity, analyzing the acquired data, and in some cases, providing the proper feedback to stop/prevent an undesired neurological event. Unlike signal processing and neuro-modulation that are only required in diagnostic/treatment applications, brain neuro-monitoring is the essential first step toward any research on the brain. Depending on the required spatial and temporal resolution as well as the level of tolerated invasiveness, various brain imaging/monitoring technologies are available. Table 1 1, summarizes and compares different brain imaging/monitoring methods in terms of their spatial and temporal resolution, spatial coverage and invasiveness.

Technology	Spatial Resolution (mm)	Spatial Coverage	Temporal resolution (s)
MRI	0.5-100	0.5 cm^2<	500<
PET	5-100	50 cm^2<	10<
fMRI	2-100	20 cm^2<	0.5<
MEG	5-100	50 cm^2<	0.05<
EEG	10-100	100 cm^2<	0.05<
ECoG	1-50	1-500 cm^2	0.005<
Depth EEG	0.1-0.5	0.001-0.2 cm^2	0.0005<
Patch Clamp	0.01-0.03	1-20 neurons	0.0001<

Table 1 1: Various brain neural activity monitoring methods compared in terms of their spatio-temporal resolution and spatial coverage.

Electroencephalography (EEG) is a standard test to acquire information about the metabolic and electric status of the brain. It is widely used for both the diagnostics and the monitoring of the state of human cognitive system, and provides brain activity recordings with high temporal and moderate-to-high spatial resolution. It is typically done in a hospital environment where the patient is tethered to a benchtop recording system and their neuro-electrical activity is sensed using several scalp electrodes (Figure 1 1), all placed based on a standard convention [1]. The test requires a trained technician to prepare the patient and the recording system, conduct the test, and perform post-test data collection and translation. As such, it is only practical for short periods (e.g., <1-2 hours).

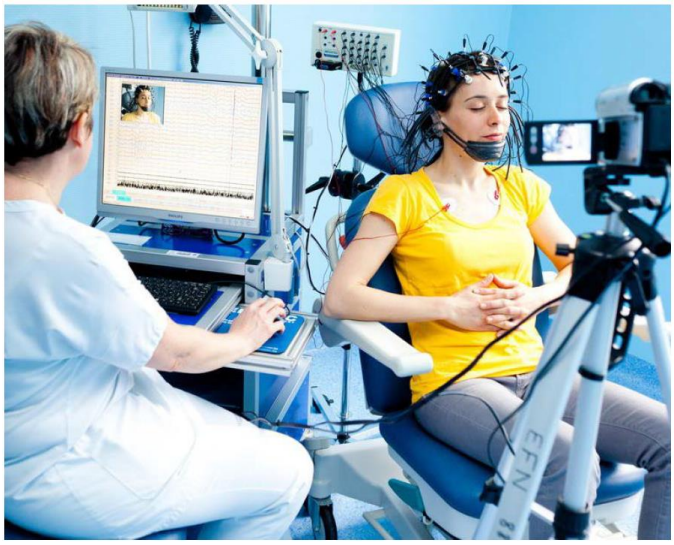

Figure 1 1 : A typical EEG recording session in a medical facility

However, many applications demand long-term continuous monitoring of the brain activity to obtain a deeper insight into the root cause of particular neurological disorders. For example, prior to a brain resection surgery in epileptic patients, a long-term monitoring needs to be done in order to precisely localize the source of very-infrequently-occurring seizures. Long-term EEG

monitoring is also used for monitoring vigilance level [2], sleep stages [3], and indices of impending severe hypoglycaemia [4] [5]. This became the motivation for development of a wireless medical device solution that can perform all the required tests while the patients go about their normal routines and daily activities and to eliminate the need for long term hospitalizations.

For most cases, intra-cranial EEG implants (i.e., deeply implanted in the brain), are too invasive to be acceptable. Therefore, over the past decade, several surface EEG recording systems in the forms of headsets, headbands, etc. have been introduced, and some of them are commercialized. The most important common issues with all of these solutions are (1) none of them are subtle enough to be used by the patient constantly during their daily routine without being obtrusive, and especially not comfortable enough to be worn during sleep at night; (2) all of them suffer from constant instability of recording electrode due to many reasons (e.g., talking, chewing, frowning, blinking, etc.), which cause recording artifacts that are orders of magnitude larger that the EEG signals. Therefore, the recording quality in terms of signal-to-noise ratio is typically not sufficient to be a reliable source for clinical applications.

Motivated by this, in this project, we studied the feasibility and challenges of the design and implementation of a minimally-invasive subcutaneous wireless implantable device, which (a) is not as invasive as deep-brain implants, (b) provides an unobtrusive continuous long-term EEG monitoring, and (c) its recording quality is not affected by motion-induced artifacts.

Figure 1 2 depict the high-level block diagram and envisaged implantation of the proposed solution. As shown, the device is comprised of an electronic backend connected to a flexible electrode array. The backend module will host a microchip that is responsible for multi-channel low-noise signal amplification and analog-to-digital conversion. Additionally, discrete components will be used to enable wireless data communication and embedded signal processing.

An inductive coil will be wound at the outer edges of the backend module for wireless power reception.

Two prototypes will be reported; first one presenting the design, development, and experimental characterization of an 8-channel custom-designed implementation of a neural recording microchip as well as a miniaturized platform for wireless communication and programmable neural signal processing. Next we will present a novel architecture for loss-less compressive activity-dependent neural data acquisition designed to improve the energy efficiency of the EEG recording device by more than an order of magnitude.

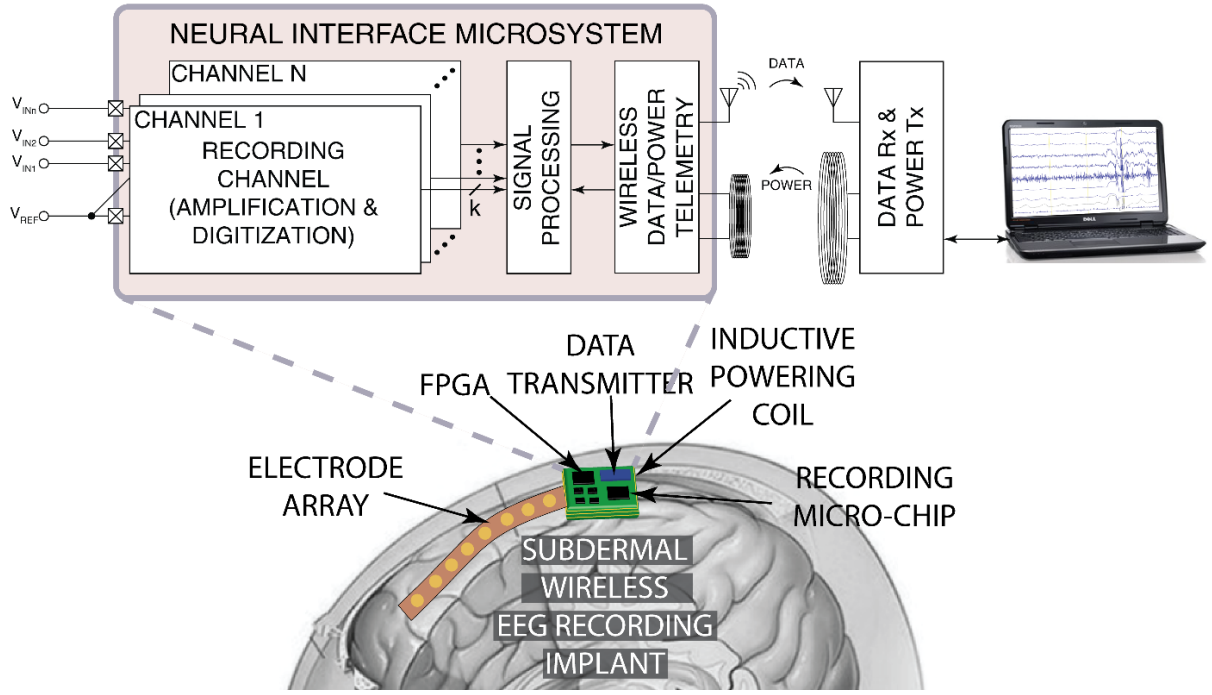

Figure 1 2 : Top-level block diagram and envisage implantation of the proposed subdermal implant.

1.2 Design consideration

1.2.1 Application-level

From the application-level point of view, the most important criterion is to minimize the solution's invasiveness. This involves both minimizing the device's weight and form factor, as well as its biocompatibility. Ideally, the device implantation surgery should be carried out by making a small incision on the scalp skin. The weight and size of the implant should be kept at a level that it is unobtrusive or minimally-obtrusive for the patient. Electrode array's material must be selected to be bio-compatible to prevent any damage to the neighboring tissue, and the electronic backend must be encapsulated using a bio-compatible material such as parylene-c or similar. Additionally, the device, particularly the electrode array, must be made using flexible material to accommodate various patients scalp size and curvature. Since this subcutaneous-EEG micro-system is considered as an outpatient care solution, the device must be designed to operate automatically with no need for patient's involvement. For the same reason, it is ideal for the device to be remotely controllable/configurable by the physician/technician.

1.2.2 System-level

From the system-level point of view, a sufficient number of recording channels must be integrated onto the device to capture the brain neural activity with a reasonable spatial resolution. Multi-channel recording is also required for multi-variate algorithms that use signals from different locations on the brain to perform seizure detection. In addition, to add diagnostic capability to the solution, the device should be equipped with a signal-processing unit that could be programmed with cross-patient or patient-specific diagnostic algorithms, depending on the application. Wireless communication is also an unavoidable feature that needs to be included in the system. Wireless

link specifications such as the link throughput and being uni- or bi-directional are decided based on (a) the application and the size of data that needs to be transmitted/received in real time, and (b) whether or not certain aspects of the system needs to be reconfigured wirelessly.

Perhaps, the most important system-level aspect of such an implantable device is the energy delivery. Envisaged to be fully unobtrusive, there can't be any external module to be mounted outside of the scalp. Therefore, continuous wireless powering of the device, which requires an external energy transmitter is not an option, and a battery is unavoidable. However, a battery that can last for several months is orders of magnitude larger (in terms of volume) to be implanted under the skin. Therefore, a hybrid approach where a small rechargeable battery that lasts ~1 day along with overnight wireless charging is ideal.

As it is discussed in details in [6], embedding signal processing on the device results in a significant reduction in decision-making latency and removes the need for a power-hungry high-throughput data transmitter that otherwise would be required to communicate raw EEG signals. However, it has the disadvantage of limited available computational power, which could lead to sacrificing the computation accuracy. Generally, a significant portion of signal processing (e.g., data compression, feature extraction and seizure detection) is conducted on an embedded processor, and only the signal processing results are communicated wirelessly. All of the system-level modules (i.e., recording channels, signal processing, and wireless transceiver) must be connected to each other using the least number of wires possible to realize a minimal form factor for the device. From a system-level point of view, this means using serializers/deserializers at the interface point of these modules to time-multiplex multi-bit data that is being communicated. Furthermore, to ensure that the solution is fully self-contained, all the necessary DC bias voltages and clock waveforms required for the operation of the integrated analog and digital circuits must

be generated using on-device components such as digital-to-analog converters (DACs), voltage regulators, and crystal oscillators.

1.2.3 Circuit-level

From the circuit-level perspective, each recording channel should be able to amplify and digitize surface EEG signals with a typical amplitude range of 10µV to a few mV and frequency content of up to 500Hz. The sensing front-end circuit must have (a) a differential architecture, (b) an input-referred noise that does not add substantially (i.e., more than 10%) to the noise that already exists at the recording electrodes (10µVrms) [7], (c) input impedance that is orders of magnitude larger than the electrode-tissue interface impedance, (d) a reasonably high voltage gain, and (e) a mechanism for rejecting large (up to 100s of mVs) DC offset between the recording and reference electrodes. The input signal dynamic range demands a minimum ADC resolution of 8 bits.

In addition to the above design specifications that are roughly the same for implantable EEG recording front ends, a subcutaneous EEG-recording device has an important requirement that is mainly attributed to their large size where signals could travel a few centimeters between different modules. This makes the signals prone to various types of noise and interference from the environment and from neighbor components on the device. As such, the recording circuit should be able to strongly attenuate any noise or interference that is common between the recording and reference electrodes.

1.3 State-of-the-art

To date, there hasn't been any report of a subdermal implant for long-term EEG monitoring. However, since many of the application, system, and circuit-level design

considerations are similar to wearable wireless ambulatory EEG devices, reviewing these works could provide useful intuition in design of the proposed subdermal implant. Table 1.2 summarizes some of the recently reported academic and commercial wireless monitoring headsets that aim to provide medical-grade EEG signals. Looking at these devices, it seems that the designs are either performance-oriented at the cost of sacrificing some of the application-level concerns (e.g., ease of use, comfort, lightweight, quick setup time, etc.), or user-oriented at the cost of reducing the system-level capabilities (Figure 1 3).

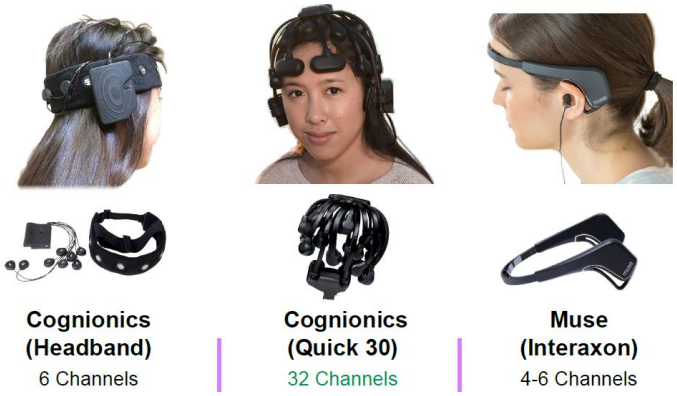

Figure 1 3 : Commercial examples of wearable EEG recording devices.

For example, the device presented in [8] meets the channel count, noise, bandwidth, and wireless connectivity requirements. The device is also designed to be adjustable to different scalp shapes and sizes. However, the weight and bulkiness of the device limits its usage to in-lab tests and inappropriate for ambulatory EEG recording. Also the device does not feature a mechanism for removing artifacts due to the physical motions of the patient, which could significantly degrade the recording quality. An example of the other end of spectrum is the Muse headband [9]. While this solution features a small form factor and lightweight, it only has 4 recording channels and uses passive electrodes for recording and integrates all the active circuitry on one side of the device, causing significant susceptibility to different types of noise and interference sources. Overall, none

of these devices is ideal to be used as a quickly-mountable medically-relevant ambulatory EEG monitoring and analysis headset for the applications described in the previous section.

Specification	*[14]*	*[18]*	*[17]*	*[17]*	*[15]*
# of channels	8	5	6	30	4
Signal Processing	Yes	Yes	Yes	Yes	No
BW (Hz)	0.5-100	0.5-100	<262	<262	N/R
Wireless Tech	BT 2.1	BT 2.1	BT 4.0	BT 4.0	BT 2.1
Motion Artifact removal	Digital	Digital	No	No	No
Weight (grams)	200	N/R	110	610	61

Table 1 2: Recently-reported wireless monitoring headsets for medical-grade EEG recording.

A number of commercially-available closed-loop implantable neural interface microsystems have been reported to enable some of the aforementioned functions. They are mostly designed for a very specific application and their primary use is for deep brain stimulation and recording, hence, lack the spatial coverage that is required for many applications [10].

1.4 Thesis organization

Chapter 2 presents the design, implementation, and experimental validation of an 8-channel custom-designed EEG microchip implemented in TSMC 180nm CMOS technology. It also reports a details analysis on the system-level design challenges of the subdermal implantable device, with a particular focus on its energy breakdown and bottlenecks. The chapter also presents design, and experimental validation of a miniaturized prototype that hosts a bidirectional wireless data communication module as well as a programmable computing module for neural signal processing.

Chapter 3 presents the design and simulation results of a novel channel architecture that targets significant improvement of system-level energy efficiency of the implantable device. The proposed design benefits from a fully-dynamic architecture and performs loss-less compression of the recorded EEG signals as it is performing the amplification and digitization. The compression of the signal is done in a fashion that EEG signal epochs with higher level of activity are recorded

with higher quantization resolution, and the idle periods (which are >90% of the time during an EEG recording) are recording with a low resolution. We have shown that this method leads to orders of magnitude energy efficiency improvement, particularly for cases that EEG recording is performed for a specific application, such as a neurological event (e.g. an epilepsy seizure) detection.

Chapter 4 concludes the thesis and discusses possible future directions for this research work.

CHAPTER 2:

A Wireless Subcutaneous Implantable Microsystem for Ambulatory EEG Monitoring

As discussed in the introduction, a subdermal implantable device seems to be an optimal solution for long-term ambulatory EEG monitoring. On one hand, it is not as invasive as a cortical or sub-cortical implanted device, and on the other hand is not prone to motion artifacts and interference like surface EEG devices.

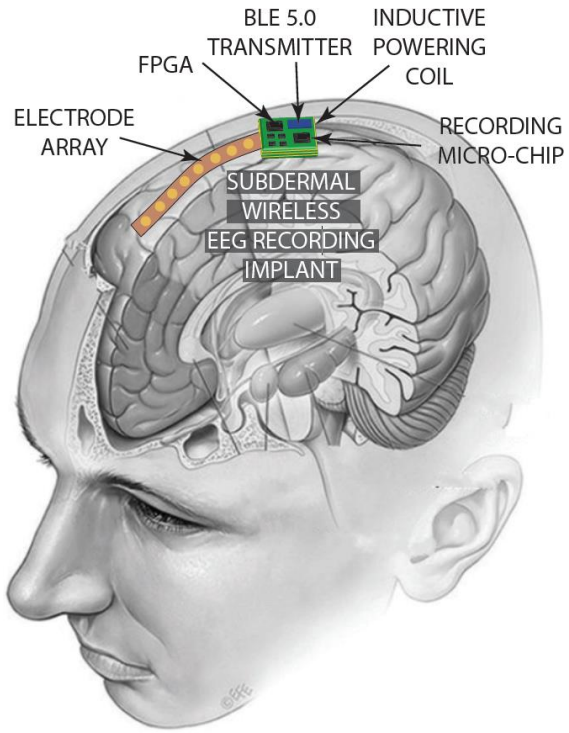

Figure 2 1 : Placement of the proposed EEG recording system.

Figure 2 1 shows the envisaged implantation of the subdermal device. As shown, eight recording electrodes are implemented on an organic substrate in the form of a narrow strip, which is connected at the tail to a mini-PCB that hosts the active electronics. The organic electrode array is custom-designed and supplied by our industry partner in this project, Panaxium Inc. As will be described in this chapter, the mini-PCB is envisioned to host blocks for neuro-physiological recording, wireless data transmission, and wireless energy reception, management, and storage.

The implanted device will operate fully automatic and records, digitizes, and wirelessly transmits the EEG activities to an external computer (hand-held or desktop) in real time, with no need for external manual control. Since this solution is planned to be used for long-term applications, it is desired for the patient's comfort and psychological health that there will be no external modules attached to the body. Meanwhile, a battery that can last several months will be too large to be implanted. Therefore, we decided to use a rechargeable battery that lasts ~24-48 hours, giving the patient the option to only need to recharge at night. The direct correlation between the battery size (which is by far the largest component in this device) and its storage capacity further emphasizes on the importance of energy-efficient design for the recording and wireless communication blocks.

In the remaining of this chapter, we will describe the first prototype designed and developed in this project. First, the system-level design, considerations, and analysis are presented. Next the design of a custom-made 8-channel neuro-recording device is presented, followed by the characterization measurement results for the chip. The chapter ends with the measurement results for the wireless communication module, validating its performance.

2.1 System level design and consideration

During the course of this research, we worked closely with industry experts and our partner Panaxium to properly define system level criteria and metrics to capture the EEG signal successfully from the organic electrode array. Additionally, the required voltage amplification gain, input referred noise, input impedance, CMRR, PSRR, and the effective resolution of the digitized data have been defined to meet the requirements for effective high-quality neural recording. This has to be accomplished while the power consumption and physical size of the device are minimized.

A typical implanted neural recorder consists of an array of analog recording frontend channels that are responsible for low noise amplification and preconditioning (e.g., filtering) of biomedical signals, data converters to digitize the acquired data, and wireless data telemetry unit. Depending on the application, a digital signal processing block might also be included in the system. This block allows for pre-processing of the raw-recorded data before transmission. This way, only the processing results, which could be the detection of a certain neurological event will need to be transmitted instead of the raw recordings. The signal processing module could also relax the transmission throughput requirement by applying data compression algorithms on the raw recorded signals.

Figure 2 2, shows the top-level block diagram of the presented device. As shown, the microsystem includes 8 recording channels, each equipped with a low-power low-noise amplifier, a 10-bit analog to digital converter, and a parallel-to-serial converter that facilitates the data transition to the wireless transceiver. All channels are integrated into a single recording microchip and will be mounted on the front side of a miniaturized PCB board. BLE transceiver and the FPGA controller along with the battery will also be placed on the PCB board, while the electrode array

is connected to the PCB from one side. For biocompatibility, all electronics will be encapsulated with approved materials. Figure 2 3 shows the envisaged final prototype.

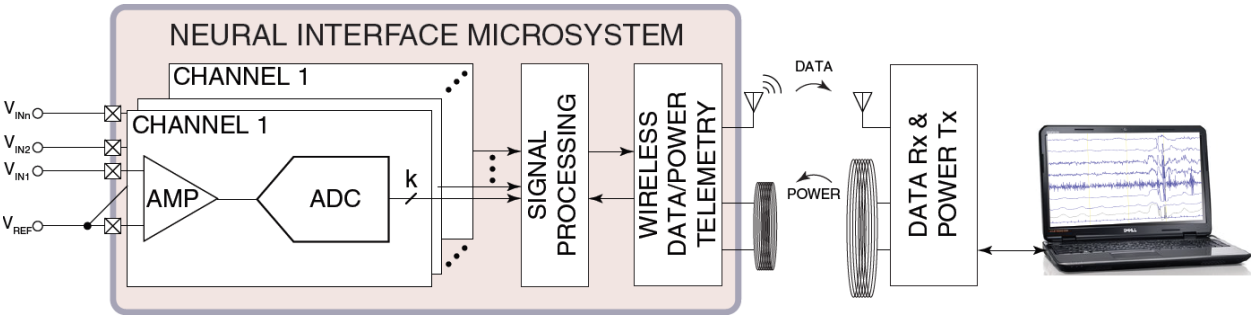

Figure 2 2 : Top-level block diagram of the presented device.

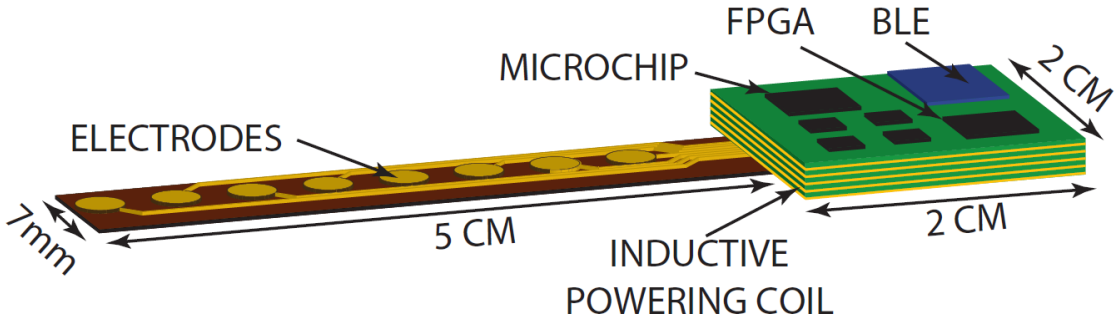

Figure 2 3 : Illustration of the envisioned final prototype.

Parameter	Target
Voltage gain	> 200 V/V
Bandwidth	1 Hz to 5 kHz
DC offset removal	$\pm 100mV$
Integ. Input referred noise	<10 μVrms
CMRR	> 60dB
Power	< 10 μA × VDD

Table 2 1 : Early design specifications

2.1.1 Wireless Data Transmitter

As shown in Figure 2 1, the transmission of the data is enabled by a standalone and off-the-shelf Bluetooth Low Energy (BLE) chip which is controlled by the onboard FPGA controller. In addition to competitively high transmission data-rate, BLE offers highly standardized communication protocol that widens the application of this microsystem and can be paired with widely available electronics such as cell phones or any other sophisticated biomedical monitoring systems. [11]

Perhaps, the most important reason for adopting the Bluetooth technology in this project is their energy efficiency. The importance of energy efficiency of all active electronic blocks in the solution's overall size and feasibility was described earlier. However, it must be emphasized that the wireless communication module is by far the most important block in terms of power consumption. To help putting this in perspective, it should be noted that a typical commercially available neural recorder has a power consumption in the order of 100µW [12], a typical wireless data transmitter that is specifically designed for low-power biomedical applications has an average of 10 mW of power dissipation in continues transmission mode. It can be deduced that overall battery life of an implantable EEG recorder is overwhelmingly limited to the efficiency of the data telemetry.

Motivated by this, an extensive research was conducted to identify the most suitable transceiver solution for this application (please see table 2 1). Based on our survey summarized below, we decided to use the Nordic Semi microchip, mainly due to its optimal performance in terms of power consumption and maximum data-rate.

Manufacturer Model No.	data throughput	Nominal power for 0dB output power	Package size	Additional consideration
Nordic semi nRF8002 series	2 Mbps	9.7 mA @ 1.8V	5.00 x 5.00 mm 32 pins	Built in MCU BLE enabled
TI CC2640R2F	2 Mbps	6.1 mA @ 3V	7.00 x 7.00 mm 48 pins	Built in MCU BLE enabled
TI CC1101	500 kbps	16.8 mA @ 1.8V	4.15 x 4.15 mm 20 pins	SPI interface 2FSK
Microsemi ZL70103	256 kbps	5.3 mA @ 2 V	7.00 x 7.00mm 48 pins	SPI interface 2FSK

Table 2 2 : List of competitive contenders for wireless data telemetry

Figure 2 4 depicts a simplified block diagram of the prototype board that has been designed to validate the system-level performance of the presented solution, with a particular interest in the wireless transmission module and the way it is interfaced with the wireless receiving stationary module. As shown, the digitized EEG signals are first fed to the signal processing module for initial decoding and processing. The output of this block could be used to further optimize the data acquisitions (e.g., channel omission, sampling rate variation, etc.) as well as serialized and fed to the on-board wireless transceiver microcontroller unit using a serial peripheral interface (SPI) protocol. The microcontroller receives and organizes the raw data from the FPGA and prepares it for transmission.

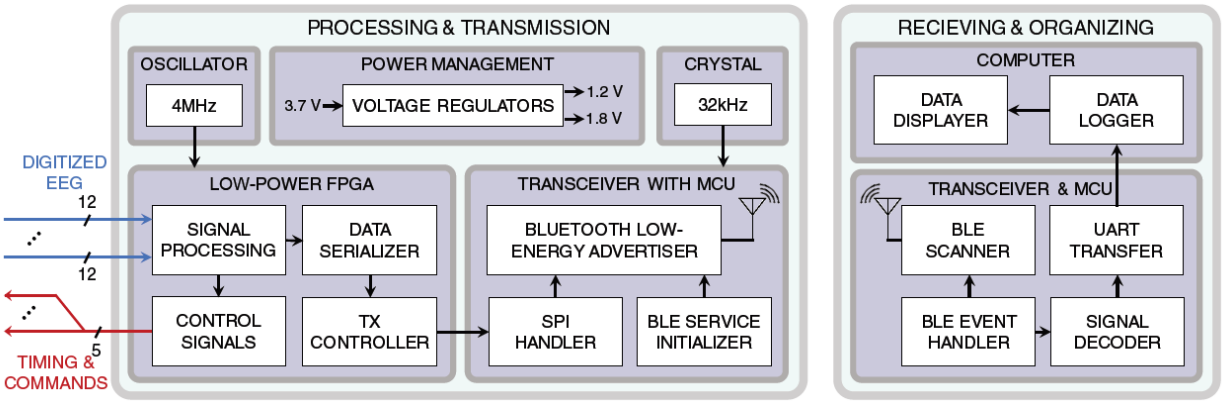

Figure 2 4 : Block diagram of the prototype board

In this work, we have used the Rigado BMD-350 module that uses the Nordic microchip introduced above (Table 2 1). The module integrates both the transmitter Bluetooth chip and its planar antenna. Table 2 2 compares this transceiver module with the most popular wireless module for medical applications used in the literature. On the receiver side, a development board from Rigado is used to receive and decoded the data and send it to the USB port of a bench-top computer using a universal asynchronous receiver-transmitter interface.

Comparison of the off-the-shelf available Bluetooth microchips/modules for medical sensory applications.

Transceiver	NINA-B112	BL652	ZL70103	CC2640R2F	BMD-350
	[32]	[33]	[34]	[35]	[36]
Energy Efficiency (nJ/b)	9-19	9-19	15-19	11-23	12-25
I_{DC} at 0dBm (mA)	5.3	5.3	5.3	6.1	7.1
VDD (V)	1.7-3.6	1.7-3.6	2.8-3.5	1.8-3.6	1.7-3.6
Duplex	N/R	Full	Half	Full	Full
Physical Size (mm^3)	14×10×4	14×10×2	6×5×2	7×7×2	8.7×6.4×1.5
Antenna	Internal	Internal	External	External	Internal
Max DR (Mbps)	1	1	0.8	1	2

Table 2 3 : Comparison of the popular off-the-shelf wireless modules

Figure 2 5, shows the top and bottom view of a 13mm×17mm PCB implementation of the presented backend module. The board is populated with the low-power Microsemi FPGA as the main processor, making the device capable of hosting small-to-medium-size (in terms of computational expense) neural signal processing algorithms, a Bluetooth low energy (BLE) 5.0 transceiver module with a built-in antenna for wireless communication (Rigado BMD350 built based on the nRF52832 SoC from Nordic Semiconductor).

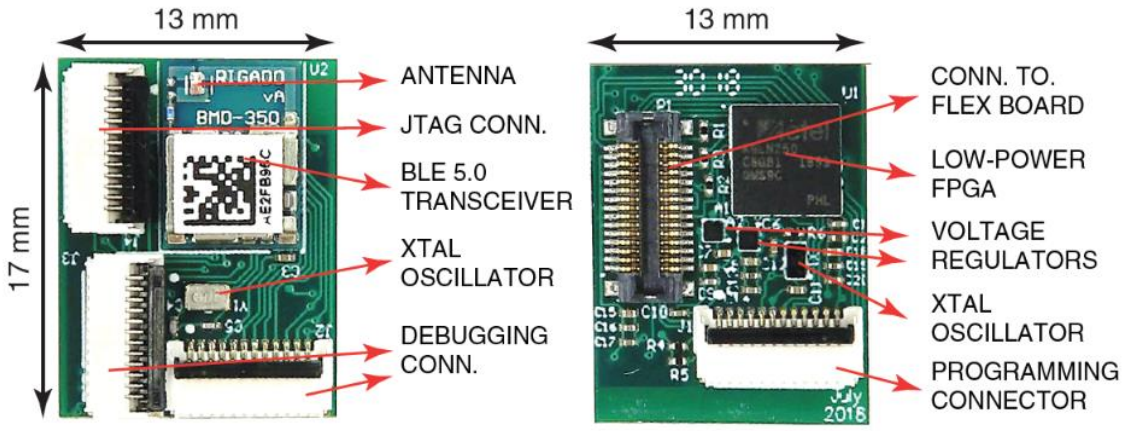

Figure 2 5 : The PCB prototype board

Figure 2 5, also shows other components on the backend PCB including voltage regulators, connectors, crystal oscillators and peripheral passive components to ensure that the board can operate as a self-contained module and only requires a 3.7V supply voltage. The full list of components used in this implementation is provided in Table 2 3.

Component	Company	Model
Regulator-3.3	Analog device	ADP121-ACBZ33R7
Regulator-1.2	Analog device	ADP121-ACBZ12R7
Adj Regulator	Linear Technology	LT3020EMS8xPBF
BLE Tx	Rigado Inc	BMD-350-A-R
FPGA	Microsemi	AGLN250V2-CSG81
XTAL 32kHZ	ECS Inc	ECS-.327-6-12-C-TR
Oscillator 4MHz	SiTime	8021AI-J4-XXS-4.0E
Vertical Connector	Panasonic	AXT430124
Vertical Connector	Panasonic	AXT430324
FFC Connector	Molex	5034801200

Table 2 4 : List of components used

2.1.2 Power Management

Similar to the wireless module, we carried out research to identify optimal battery that can power this system. Ideal batteries should be compact and capable of max discharge rate of greater than 11 mA (~10mA for BLE chip in addition to the estimated 1mA for the EEG recoding chip). More importantly, we needed to ensure that the battery has the sufficient capacity to last at least 12 hours before recharging. Based on the above, DTP301120 polymer Li-ion rechargeable battery, which can support nearly 4 hours of continuous data transmission at max data rate is the most attractive choice for this application. We decided to sacrifice the battery lifetime to avoid increasing the solution size beyond the budget set for us by our collaborators. Later in Chapter 3, we will show that how we leveraged the sparsity of EEG signals to achieve an order of magnitude longer battery cycles with the same battery.

Model #	Capacity	Max Discharge rate	size
DTP301120	40 mAh (3.7V)	40mA	11.5 x 22 x 3.2 mm
031015	16 mAh (3.7V)	48 mA	10 x 15 x 3 mm
Bottoncell:MS920SE	11 mAh (3 V)	0.8 mA	9.5 mm x 2.1 mm
Bottoncell:MS621FE	5.5 mAh (3 V)	0.25 mA	6.8 mm x 2.1 mm
Eagle Picher - D-00020	50 mAh (3.65 V)	25 mA	18x12x4.5 mm
Wyon - W102	18 mAh (3.7 V)	36 mA	10x6.5x4 mm

Table 2 5 : Comparison of various batteries for capacity, discharge rate and size

2.2 Analog front-end and signal preconditioning

Accuracy of EEG monitoring system heavily relies on the performance of its Analog front-end and mainly low noise amplifier. While the number of recording channels are specified by the spatial resolution required by the application, the electrical performance requirement for each

channel is relatively similar for all neuro-physiological recording circuits. The circuit should have sufficient frequency bandwidth, that is 500Hz if designed for local field potential (LFP) recording, and 5 kHz if designed for recording action potentials (APs). The voltage gain must be large enough to amplify the smallest neural activity (~10uV) sufficiently so that a low-power ADC is able to digitize it with the required resolution. On the other hand, the gain should not be so large that a strong neural signal (e.g., 1mV amplitude) saturate the amplifier.

The recording circuit architecture must be differential to enable the use of a reference electrode for removing common-mode signals and interference. To ensure maximum common-mode rejection and minimal loss of signal at the electrode interface, the front-end circuit must have the highest possible input impedance. Additionally, the input-referred noise must be minimized to obtain the highest possible signal to noise ratio. This includes both thermal and flicker noise that are generated by the circuit and referred to the input. Another challenge in the design of neural amplifiers is removing a DC offset voltage at the input that is generated due to chemical reactions between brain cells and microelectrodes, and can saturate the amplifier. As it has been mentioned before, all of the above must be done while maintaining the area and power consumption within the budget specified by the application or the desired battery life-cycle.

To make the chip capable of both LFP and AP recording (the AP recording is for future applications), we decided to design the amplifier capable of capturing EEG signal within the frequency spectrum of 1 – 5 KHz and amplitude ranging from 10μV to 5mV while eliminating any possible DC offset that may be caused by electrochemical activities at the recording site. [13]

To maximize CMRR, fully differential folded cascaded OTA, shown in Figure 2 6, topology was chosen for the first amplification stage with large PMOS devices at the input to minimize the flicker noise. [14] It's worth mentioning that choosing large PMOS devices also

facilitates lowering the overdrive voltage for these devices when biasing them in subthreshold region of operation. It's well known that transconductance "gm" of MOS transistors is inversely proportional to their overdrive voltage and putting input devices in subthreshold region for the same drain current (i.e., same power consumption) will help maximizing their transconductance.

$$gm = \frac{2I_D}{V_{ov}}$$

To remove the DC offset, input transistors are AC coupled to the organic electrodes using two on - chip metal-insulator-metal (MIM) capacitors. The voltage gain of this stage is set by the ratio of C1/C2. C2 was set to 300 fF, which is the minimum possible MIM cap that could be implemented in this technology. Accordingly, to C1 was set to 15pF to realize a voltage gain of 50 V/V. C2 also influences the amplifier's bandwidth. The low-frequency high-pass pole of the amplifier is set by $1/(R_2C_2)$. Given the small value of C2, to achieve a high-pass pole of 1Hz, R_2 needs to be greater than 1GΩ. Implementing such a large resistance on the chip as a passive component is not feasible due to its area that will be orders of magnitude larger than the entire active circuit. Instead, we decided to use pseudo-resistors that are implemented using diode-connected long thick-oxide NMOS transistors, as shown in Figure 2 6. A common-mode feedback (CMFB) circuit, depicted in Figure 2 7, is used to set the DC output of the amplifier and ensures a DC level of V_{CM}=0.9V at all times.

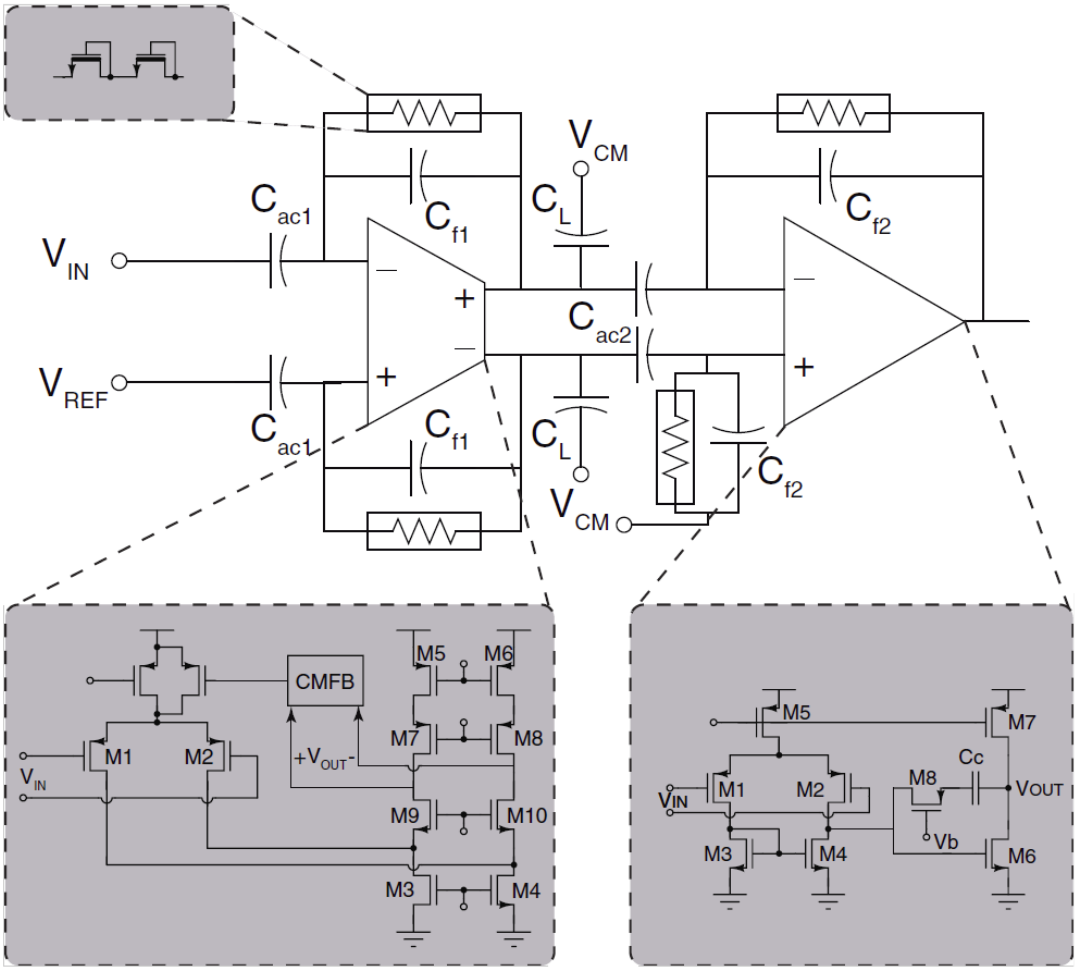

Figure 2 6 : Circuit view of the analog front-end including fully differential folded cascaded OTA

Device	Width (µm)	Length (µm)	Multiplier
First stage : M1 , M2	2	1.5	25
First stage : M3 , M4	2	15	1
First stage : M5 , M6	4	10	1
First stage : M7 , M8	1	2	1
First stage : M9 , M10	1	1	1
2nd stage : M1 , M2	2	5	1
2nd stage : M3 , M4	1	10	1
2nd stage : M5	1.6	3	1
2nd stage : M6	1	10	1
2nd stage : M7	2	3.5	1
2nd stage : M8	0.9	3.5	1

Table 2 6 : Summary of transistor sizing for AFE

As shown in Figure 2 6, a second stage of amplification is added to the recording channel to increase the overall voltage gain. Similar to the first stage, a capacitively-coupled negative feedback amplifier topology is used. Since the input referred noise of the second stage is significantly attenuated by the first-stage gain, its noise performance requirement is much more relaxed. As depicted, for this stage, we used a two-stage OPAMP.

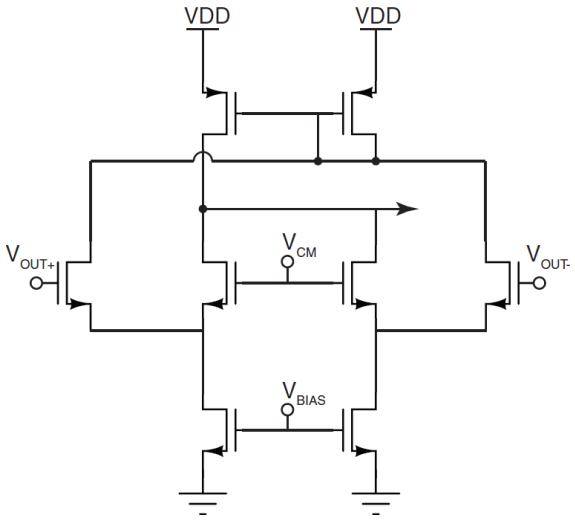

Figure 2 7 : Circuit view of the common mode feedback

Figures 2 8 and 2 9, show the gain-bandwidth simulation results for the first and second stage amplifiers. The first stage, which uses a fully differential folded cascade OTA, has a closed-loop gain of 33.5dB, for the 1Hz to 24 kHz bandwidth. The circuit's stability is also verified and it shows a phase margin of 60 degrees.

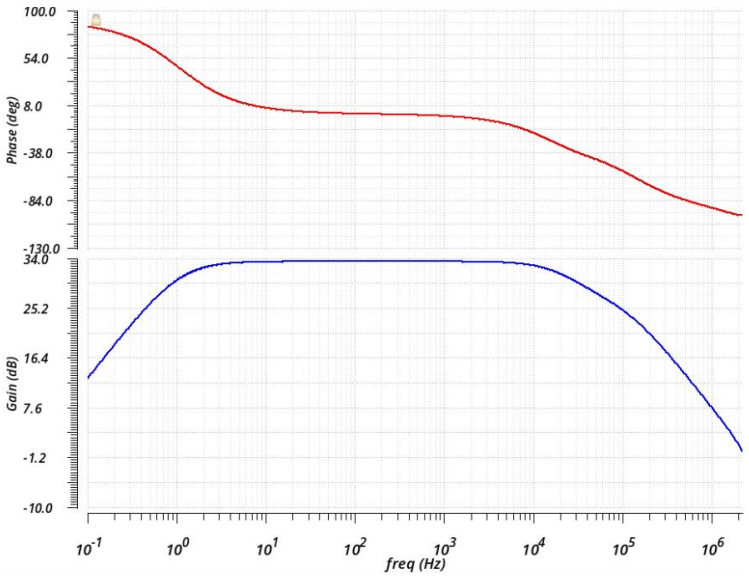

Figure 2 8 : Gain and phase plot for first stage folded cascaded OTA

The second gain stage, employing a low power and wide swing OpAmp , has a closed loop gain of 15.5 dB , and was also stable with a phase margin of 50 degrees as shown in figure 2 9, The 3dB bandwidth for this stage was 0.4 Hz to 45 KHz.

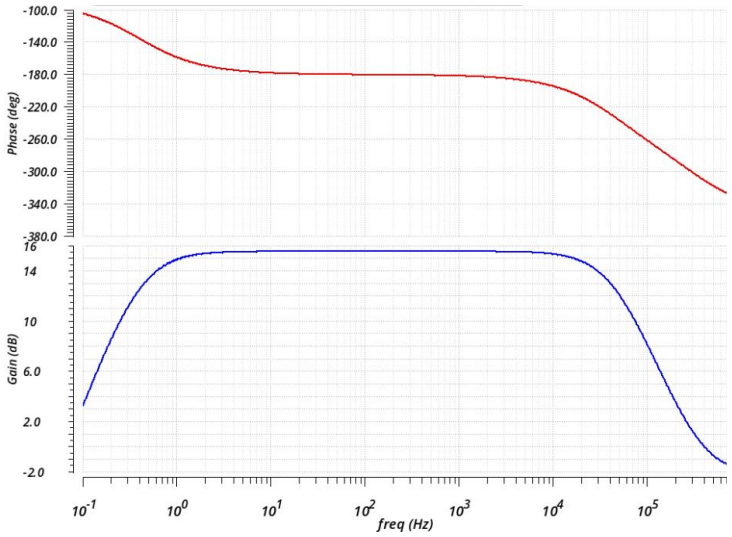

Figure 2 9 : Gain and phase plot for second stage OpAmp

Overall, both amplification stages have a combined gain of 49 dB with overall bandwidth of 1 Hz to 20 KHz. Figure 2 10, shows the simulation results for the voltage gain and bandwidth of the two-stage amplifier, which both meet/exceed our initial targets.

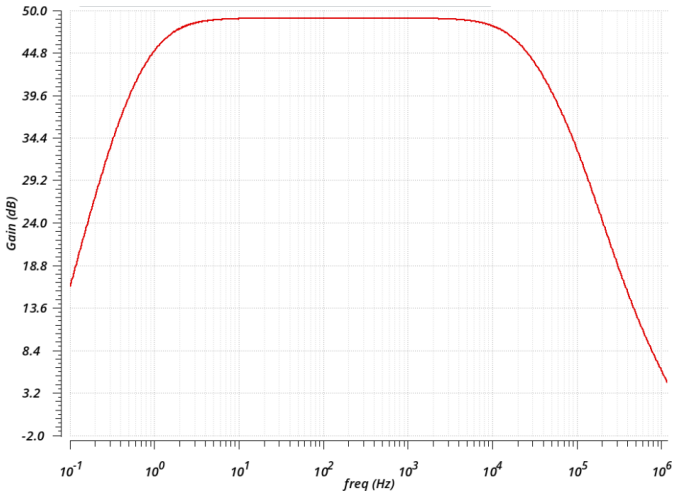

Figure 2 10 : Gain bandwidth plot for overall gain stages

The input - referred noise of the analog front end is simulated and plotted in figure 2 11. The high frequency noise floor is at 46nV/sqrt(Hz) and the integrated input referred noise from 1Hz to 5 KHz was calculated at approximately 6 μV RMS. It's worth noting that as shown in figure below flicker noise 1/f becomes the dominant source of noise for frequencies below 100Hz as expected.

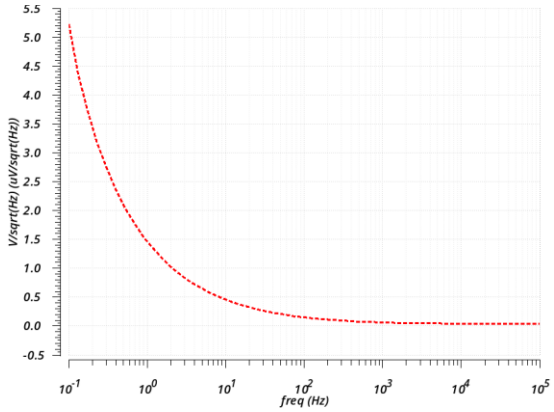

Figure 2 11 : Input referred noise for the analog front end

We should also mention that the input impedance is maximized as the amplifier's inputs are connected to the gates of PMOS devices with fA leakage current, leading to a super-GΩ input impedance. Such a high Z_{in} along with the fully differential architecture of the amplifier guarantees

a high CMRR. Simulation results were conclusive that the designed analog front-end met the system level requirements and criteria and are ready to be passed on to the data converter.

2.3 Data conversion using SAR ADC

As mentioned earlier, the EEG data needs to be digitized and serialized before wireless transmission. Considering the EEG's low-frequency signal content and to minimize the power consumption, the successive approximation register (SAR) topology was selected for the ADC. The ADC should be designed with a dynamic range that covers the entire magnitude range of the EEG signals, when amplified by the two-stage amplifier. Given the voltage gain of 240 V/V, and the desired ranged of 10uV - 5mV at the amplifier's input, the ADC's input dynamic range should be equal to 2.4mV - 1.2V. Accordingly, the required ADC resolution is calculated as:

$$required\ quantization\ levels = \frac{1.2V}{2.4mV} = 600$$

$$requied\ ADC\ resolution = log_2^{600} = 9.23$$

Considering minimum required resolution of 9.23 bits and accounting for non-idealities in practical ADCs such as quantization errors and mismatches caused by parasitic, a resolution of 10 bits was chosen for this design.

SAR ADCs convert sampled values from a continuous analog waveform into digital quantized values through a binary search process where each sampled value is compared against a sequentially-approximated voltage level. The digital output bits, from MSB to LSB, are set based on the result of the consecutive comparisons. The reference voltage level for the next comparison is also set based on the result of the previous comparison. After n repetitions are done, where n is the ADC's number of bits, the conversion is completed. Number of bits don’t affect the data

conversion flow however higher number of bits will lengthen the binary search, hence, make the ADC slower.

The first stage, that is, the track and hold (see Figure 2 12), was implemented using a transmission gate with analog signal at input and a capacitor as the storage that will be charged to the sample level and will retain this voltage as SAR comparison commences. Schematic view of the sample and hold circuit is displayed below. (For simplicity, connection to other terminals of the comparator are not marked.)

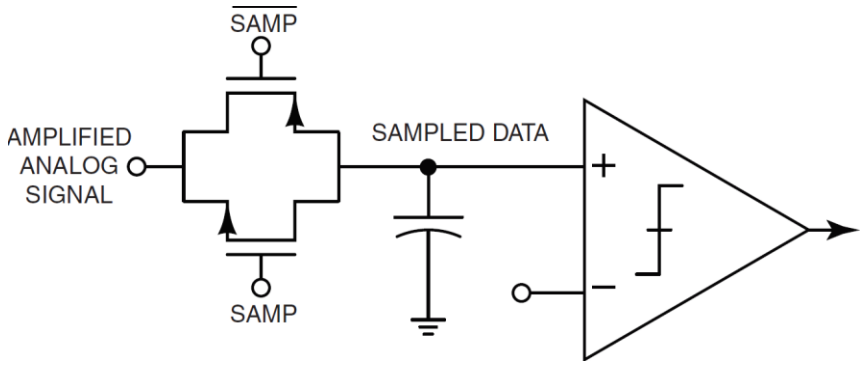

Figure 2 12 : Sample and hold circuit

Sampled level needs to be compared to all possible quantization levels which will be generated by a capacitive DAC. A 10-bit binary-weighted capacitive DAC is employed in this design to generate the required 1024 voltage levels. All capacitors have a common terminal connected to the comparator which is also the output of the DAC and their other terminal is individually controlled (charged / discharged) by SAR logic, and thus any possible voltage level within the target resolution can be synthesized. A simplified view of the implemented capacitive DAC is illustrated below.

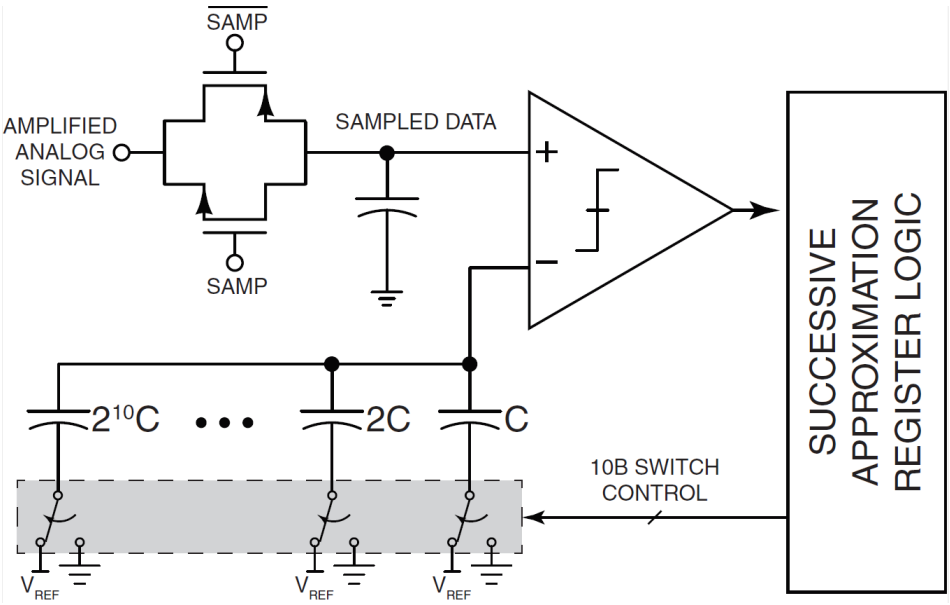

Figure 2 13 : Capacitive DAC with binary weighted capacitors

The successive approximation starts with setting the most significant bit (MSB) to 1. Then this value is sent to the DAC which generates the corresponding analog value. If comparator determines that sampled value is greater than DAC output, the bit is left as 1 and we move on to the next bit.

The comparator used in this project was designed based on the “strong arm” architecture. The circuit, shown in Figure 2 14, has a fully-dynamic power consumption, benefits from a pre-amplifier stage to minimize the kick-back noise, and uses all-node-reset method to achieve a minimal hysteresis. Once the “clk” has fallen to 0, the output node of the comparator is charged to 1. In this mode, all the other nodes in the circuit are also reset to VDD/GND to ensure no residual charge is left in the parasitic capacitors before the comparison starts. This will ensure minimal offset and hysteresis for the comparator. Once the “clk” is raised to 1, the input differential pair, as well as the cross-coupled positive feedback loop are enabled and the two outputs slew away from each other to VDD and GND, indicating the comparison result as a 1 or a 0. [15]

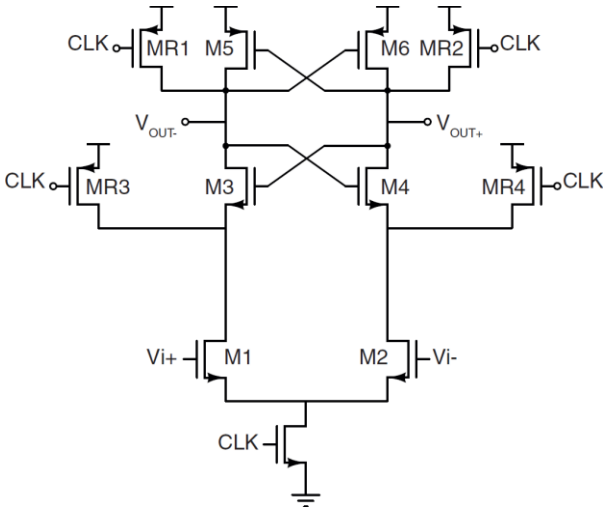

Figure 2 14 : Strong Arm Comparator

Device	Width (nm)	Length (nm)	Multiplier
MR1, MR2, MR3, NR4	420	250	4
M1, M2	420	250	1
M3, M4, M5, M6	420	250	1
Tail NMOS CLK	420	250	2

Table 2 7 : Summary of transistor sizing for ADC's comparator

In this design, the comparator output could be fed to a shift register to generate a parallel output after 10 clock cycles, or could be directly used as the serial output. If used as a serial output, it should be noted that the data is valid for certain clock cycles, and invalid for the others. Therefore, timing information of the SAR ADC must be sent along its serial output bits to the block that will use them (e.g. a signal processing block, or a wireless transmitter). We should mention that since the BLE transceiver requires single-ended data, the other terminal of the comparator was connected to a dummy load so not to introduce mismatches in the design.

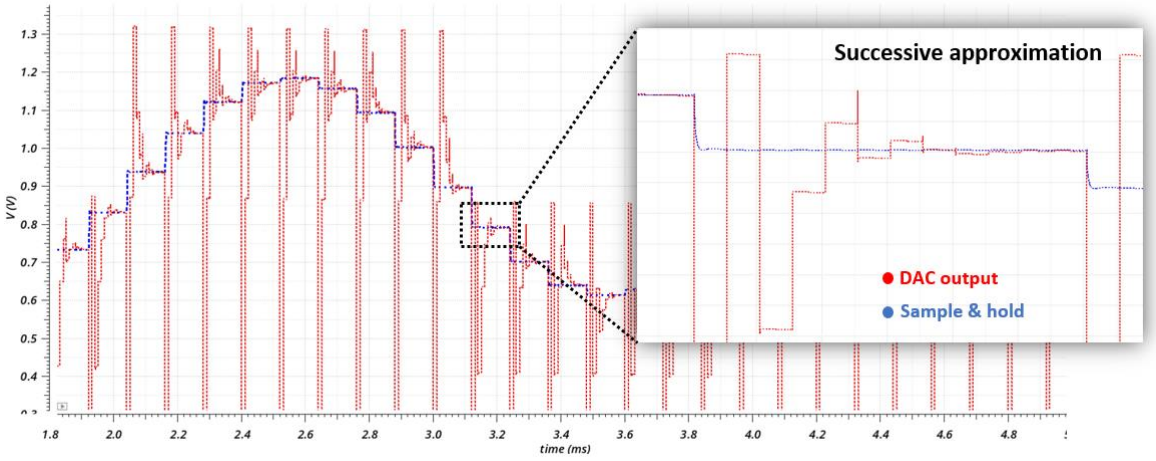

Figure 2 15 : Transient analysis of the SAR ADC

Figure 2 15 shows an example transient simulation results of the SAR ADC's output in response to a sinusoidal input. The blue trace represents the sampled and hold value of the 500-Hz sinusoidal wave and the red trace is the output of the SAR DAC which tracks the sampled value successfully. Based on the post layout simulation results, the effective number of bits (ENOB) was calculated as,

$$ENOB = \frac{SNR - 1.76}{6.02} = 7.85 \text{ Bit}$$

Effective number of bits is directly related to the parasitic mismatch in the SAR ADC's capacitive DAC, choosing larger capacitors will significantly improve ENOB at the cost of higher power consumption and larger area.

2.4 Integration and fabrication

Once front-end and Data converter blocks were successfully verified to meet the system level criteria, they were integrated to form a single recording channel depicted in Figure 2 2. Eight recoding channels were implemented and their serial outputs were fed to a multiplexer that interfaces with the BLE chip for wireless data telemetry.

The parallel to serial converter was comprised of an 8-to-1 multiplexer which is controlled by a “clk” with 8 times the frequency of the generated data. Once transmitted, the bit stream will be deserialized inside a microcontroller unit to reconstruct each channel’s data.

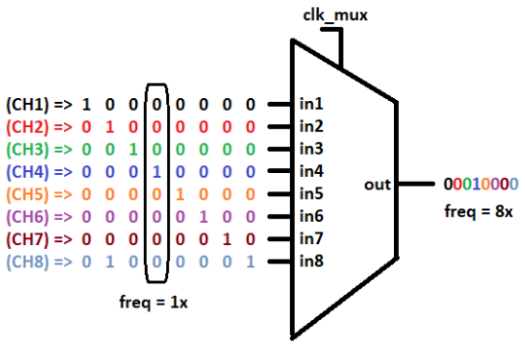

Figure 2 16 - 8:1 Mux used for data serialization

Figure 2 17, shows the layout of the described recording channel with major blocks annotated. As shown, each channel includes a 10 bit SAR ADC, a fully differential OTA and an OpAMP for additional gain which all together take an area of less than 0.14 mm^2.

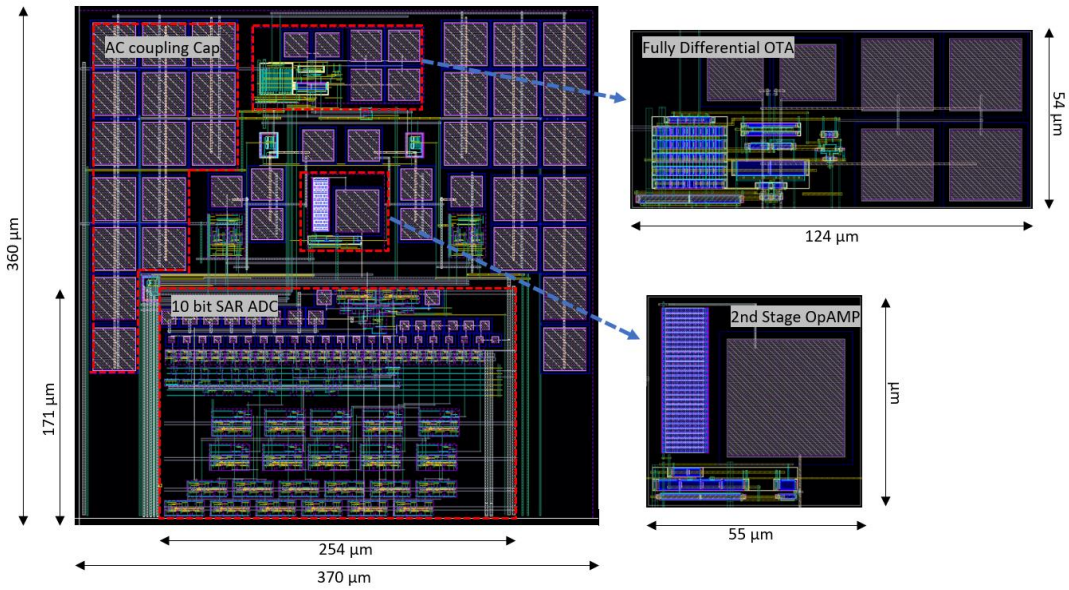

Figure 2 17 : Layout of the EEG recording channel (0.36 x 0.38 mm)

Figure 2 18, shows the full layout of the chip with all 8 recording channels, the parallel-to-serial data converter, as well as additional blocks for testability, signal buffering, and ESD protection.

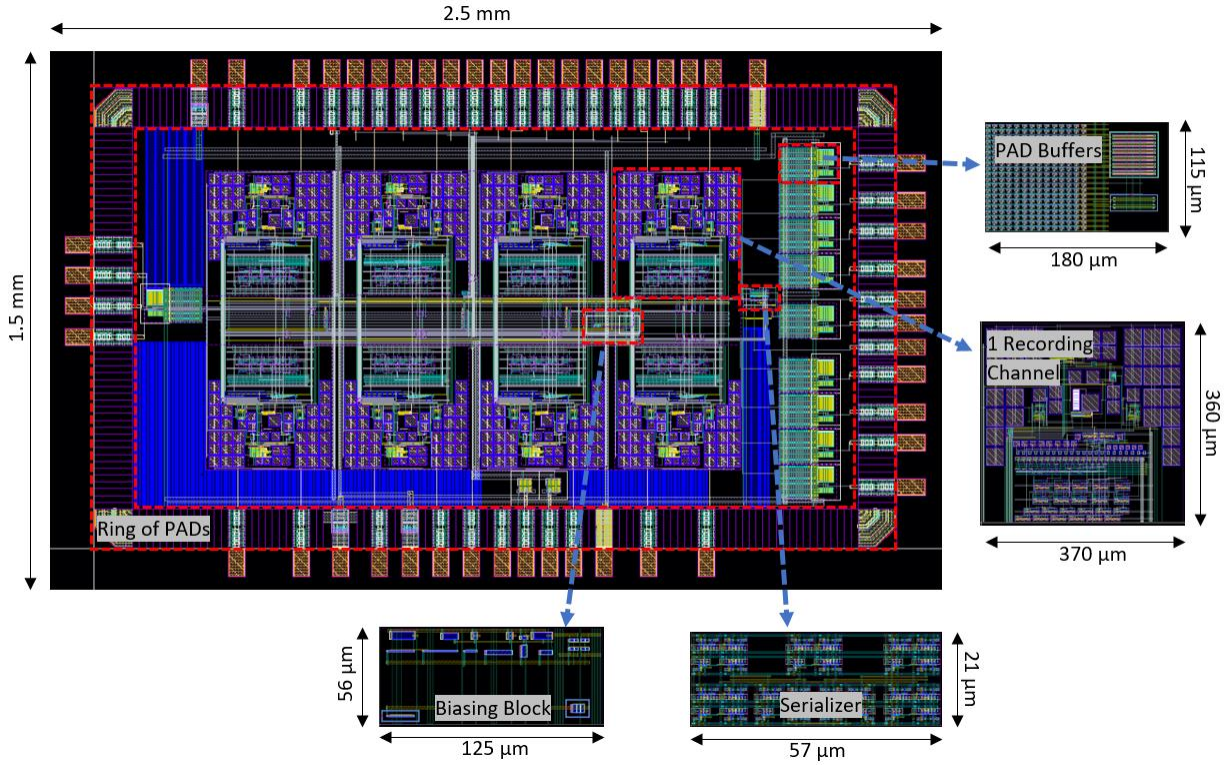

Figure 2 18 : Full chip layout of 8 channel EEG recorder

2.5 Experimental Measurement Results

The microchip was fabricated in TSMC 0.18 µm technology through CMC microsystems services. Following fabrication, the chip was packaged (PGA-68) and was tested using a development test-bench PCB, which was custom-designed and fabricated for this project as shown in figure 2 19.

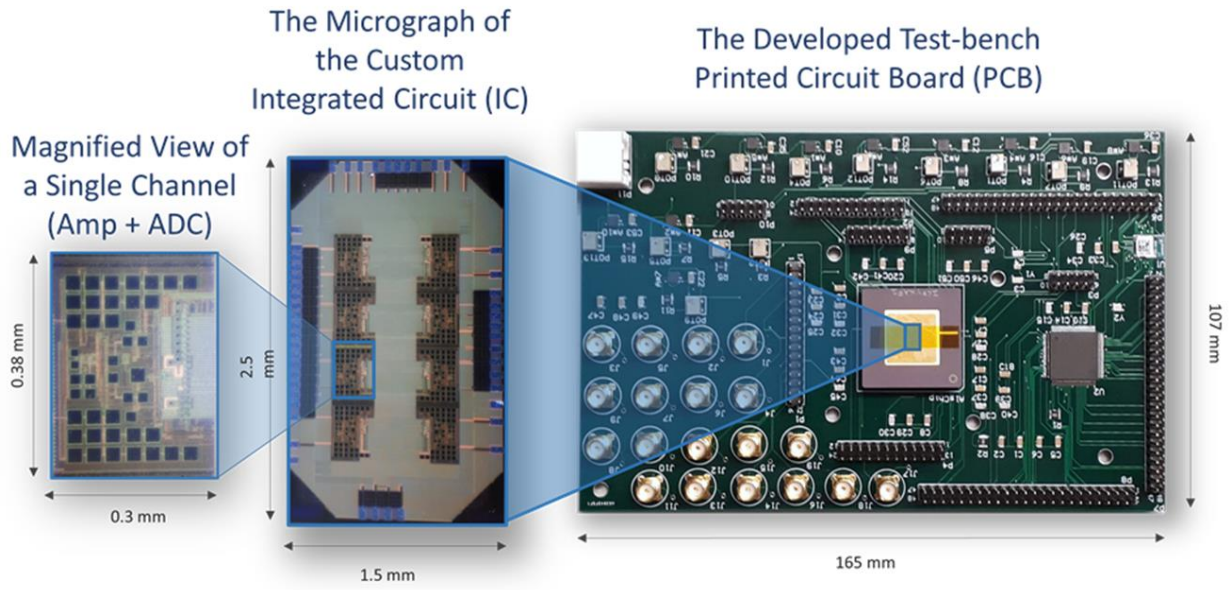

Figure 2 19 : Magnified view of the fabricated chip along with test-bench PCB

To measure the gain and bandwidth of the analog front end, sinusoidal function generator with capability of generating differential source was connected to the input of subthreshold OTA (using BNC connector for robustness) with differential peak to peak amplitude of 2mV, then frequency of the input wave was swept from 0.1 Hz to 40 KHz and the output amplitude after the single ended OpAmp was probed and measured. Result of this measurement is plotted in Figure 2 20, Gain of 48.8 dB was recorded which was very close to the post layout simulation results of 49dB. However, 3dB bandwidth deviated from the expectation by 5 KHz and was recorded at approximately 15 KHz. It was observed during the design stage that bandwidth is sensitive to the variation of reference current and also post fabrication parasitic and PVT variations could also contribute to the mismatch between bandwidth expectation and measurements. Nevertheless, our system level requirement bandwidth of 5 KHz is clearly achieved and the chip performance won't be affected by the aforesaid variation.

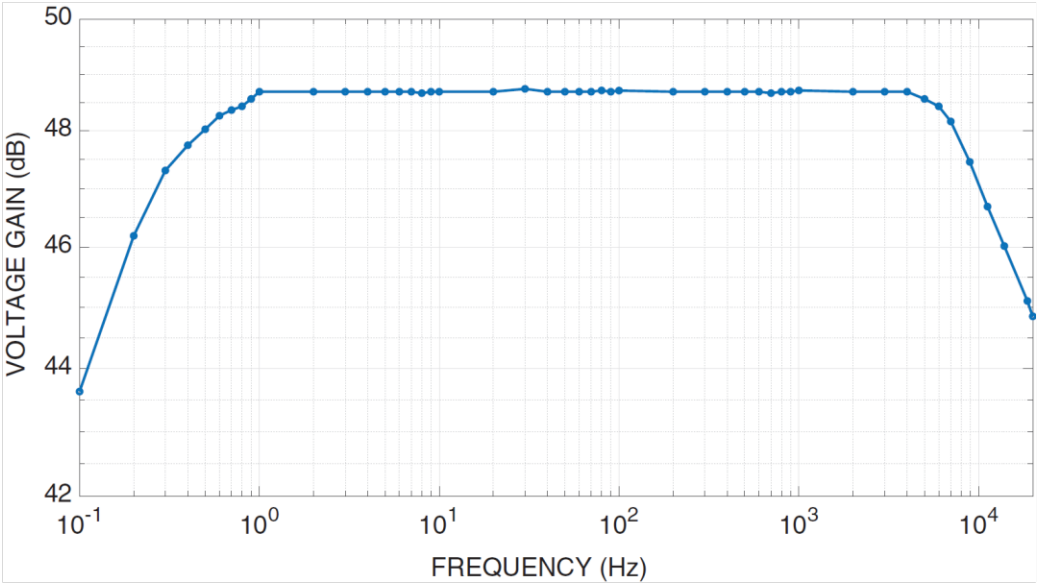

Figure 2 20 : Measured gain – bandwidth graph post fabrication

Input referred noise was measured by shorting inputs of the input amplifier to the ground and measuring the noise magnitude at the output, the measured noise is then divided by the overall front end's gain to evaluate the input refereed noise. The results of this measurement are plotted in figure 2 20. At lower frequencies, the measured noise was higher than the simulation results i.e. at 8Hz, measurement: 1μV/$\sqrt{Hz}$ vs. simulation: 0.5 μV/√Hz, however the gap between measured and simulated results reduces at higher frequencies i.e. at 110Hz, measurement: 128 nV/√Hz vs. simulation: 136 nV/√Hz. Noise measurement can be significantly affected by environmental elements such as Electromagnetic interference (EMI), and also process corner variation will contribute to any mismatches between simulated and measurement results.

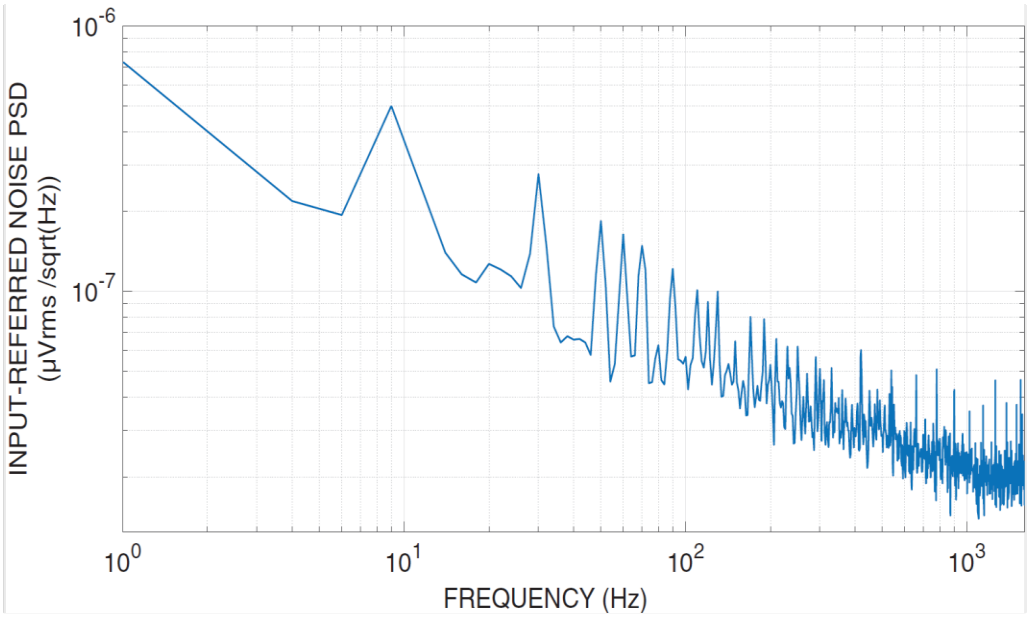

Figure 2 21 : Measured input referred noise

Table 2.7 summarizes the overall specification of fabricated chip at both channel-level and the chip-level. Measurement results confirm that the chip meets the expectations for high-quality neural recording. Nonetheless as discussed in section 2.3, this was an engineering design decision to priorities lower power consumption and smaller chip area over higher ENOB for SAR ADC.

Overall specification summary	
ASIC area	3.75 mm^2
Technology	TSMC 0.18 µm
Analog front end power consumption (per channel)	3.7μA RMS
ADC power consumption (per channel)	14.4μA RMS
Gain	48.8 dB
Bandwidth	15 KHz
Int. input referred noise	6 µV RMS (1 Hz – 5 KHz)
ADC resolution	10 bit
ADC ENOB	7.85 bit
# of channels	8 Channels

Table 2 8 : Overall specification summary

CHAPTER 3:

An Activity-Adaptive Loss-Less-Compressive EEG Recording Channel Architecture

3.1 Introduction and motivation

As discussed in the previous chapter, wireless data transmission is responsible for the majority (i.e., >80%) of the implantable device overall power consumption. Despite adopting one of the most energy efficient wireless technologies (i.e., BLE 5.0), our calculations show that the implantable device presented in chapter 2 can only last ~4-5 hours of continuous operation without the need for a battery recharge. It was also discussed that for practical reasons, a minimum of 12 hours of operation (ideally, >24 hours) is required for ambulatory long-term EEG recording devices. It should also be considered that the system presented in the previous chapter has only 8 recording channels. Similar to surface EEG, carrying out a clinical-grade recording with sufficient spatial resolution and coverage, more electrodes needs to be added to the system, hence, more number of recording channels. This means more EEG data is acquired at any moment, which demands higher communication data-rate, further reducing the battery life.

The above further emphasizes on the vital role of data transmission energy efficiency in overall success of the brain-implantable device. Reviewing the literature, we noticed that due to the wide availability of far-field radio transmitters and a myriad of different infrastructures (e.g., Bluetooth Low Energy, Wi-Fi, etc.), these transmitters can be employed conveniently, and are expected to exhibit robust operation [16]. However, even the best low-power radios consume >1 nJ/b [17] [18], which is an order of magnitude larger than what we require for a sub-dermal EEG.

As an alternative to far-field radios, impulse radio ultra-wideband (IR-UWB) transmission has been proposed and employed in a few recent reported devices. UWB radios do not require to generate a carrier signal and generally use a very wide band of frequencies (3.1-10.6GHz) and yield high data-rates with a few tens of pJ/b energy efficiency [19] [20]. However their high peak transmission power requires a large high-capacity battery for operation. They also need the receiver to be at a cm-scale distance (e.g., on skin), making them less practical for our application of interest. The same issue is valid about the near-field backscattering transmitters that are highly energy efficient, but require the external module (i.e., the interrogator) to be placed on the skin [21] [22] [23].

Considering the above discussion, it seems that while the more energy-efficient technologies such as UWB and near-field backscattering might be an attractive solution for implantable devices with in-lab or short-term applications, far-field radios are the only practical option for a long-term ambulatory applications, where the patient's comfort and normal day-to-day life is of greater priority. This motivates for devising strategies that allow for using the far-field transmitters, without sacrificing the battery life.

In the rest of this chapter, we will first introduce our proposed idea for lossless EEG data compression, and later, describe the circuit implementation, characterization, and integration.

3.2 LFP signals Sparsity

In defining the characteristics of EEG signal for long term monitoring applications, it's essential to review and distinguish between two different classes of bio potentials. "Action potential" is the result of individual cell's depolarization which in turn ripples through adjacent cells (Figure 3 1). Action potentials are the fundamental means to cell-to-cell communication. [24] By nature, action potentials are very unique to individual cells and recording them requires invasive direct access to the cell with use of fine, needle-shaped probes that are inserted into the brain.

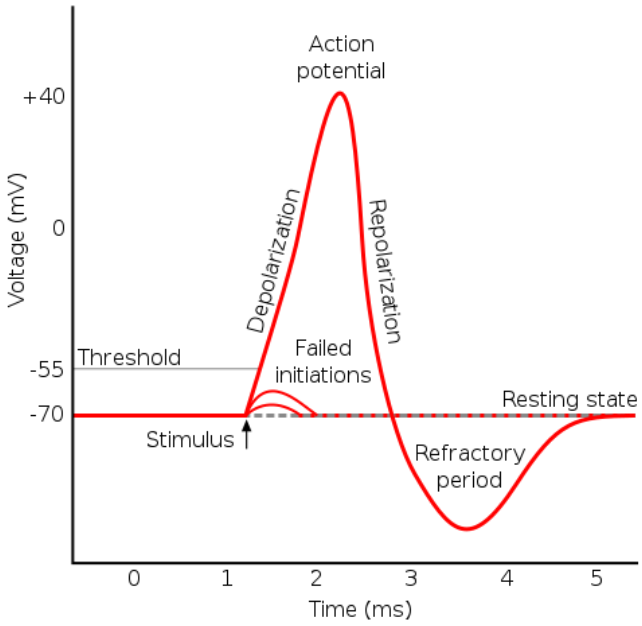

Figure 3 1 : Diagram showing phases of an action potential over time [25]

On the other hand, local field potentials (LFPs) are extracellular signals that could be defined as the spatially-averaged sum of action potentials of a cluster of cells in a region of interest.

LFPs are monitored by placing electrodes in the vicinity of a group of neurons and as the name implies, the field created by the cells in that spatial radius will be visible to the neurological recorder. Since direct access to individual cells are not required, this method is significantly less invasive in compare to recording action potentials, making it more appropriate for long-term ambulatory EEG recording [26].

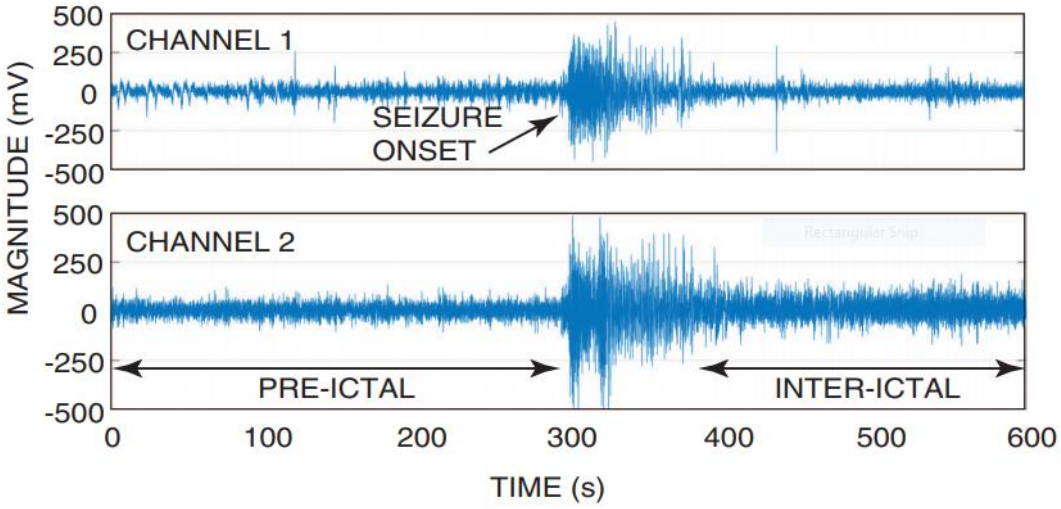

Figure 3 2 : LFP EEG recording from CHB-MIT scalp EEG database [27]

As illustrated in Figure 3 2, LFP EEG activities of the brain are represented with large transient spikes in amplitude of the local potential field which are sparse in time domain. [28] As mentioned earlier sampling sparse signals at a continuously high frequency are not efficient and will lead into significant overhead for the transmitter. By properly defining the sampling windows and adjusting the sampling rate accordingly it's possible to drastically reduce the data generation and extend the implantable device's battery life. It's also worth noting that since Nyquist condition is met for each sampling window, this technique is theoretically loss less and the sampled data can be completely reconstructed once transmitted.

3.3 EEG Recording with Activity-Adaptive Quantization Resolution

Generally, recording of sensory signals, including biomedical signals, starts with low-noise signal amplification, followed by signal conditioning stages such as spectral filtering and noise reduction, and ends with the digitization of the data before being ready for processing on- or off-chip [29]. Considering the unpredictable and non-periodic nature of the LFP EEG signals ,the recording process is required to be done continuously to ensure high temporal resolution. However, depending on the level of signal's activity (i.e., transient dynamic range) the required sampling rate to fully capture the signal's dynamics might vary. Of course, adopting a variable sampling rate should be done without violating the Nyquist rate. A crucial principle in sampling any continuous-time signal is the Shannon-Nyquist sampling theorem which states, “A band-limited signal with bandwidth X can be perfectly reconstructed if it’s sampled with a frequency 2X” [30].

One way to deal with this is to leverage the fact that the Nyquist rate varies for different intervals in an EEG recording. Therefore, it'd be possible to acquire signals with a variable sampling rate (i.e. non-uniform sampling), which results in an adaptive temporal resolution that it is correlated with the signal "level of activity". Adopting a non-uniform sampling approach means that the maximum sampling rate (hence, data rate) is only used for a fraction of time, which for the typically-sparse EEG signals would be a very small fraction. Consequently, it results in a significant reduction in the required data transmission throughput.

While the advantages of this idea are clear, it has a fundamental issue that constrains its efficacy in practice. Due to the unpredictability of the EEG signals, the sampling rate variations must be done based on the signal's average level of activity for the past x number of samples. Therefore, the quantized data must be fed to an activity monitoring block that decided to increase/decrease the sampling rate. The problem happens when a sudden activity such as a seizure

starts right after a low-activity period. In this situation, due to the recent past low activity, the circuit is sampling at a low rate. Since the activity monitoring block is using an averaged metric, it is very likely for the circuit to take a long time to understand that it must increase the sampling rate. Hence, a time-sensitive sudden short neurological event might be completely missed or the detected very late.

This motivates for a recording technique that adjusts the quantization resolution rather than the sampling frequency. Using this method, the signal is always sampled at the Nyquist rate, but the resolution of the digital output is adaptive to the signals level of activity. While the idea is simple and elegant, it's effective and efficient implementation could be quite challenging with two fundamental problems.

The first problem is that generally, nyquist-rate ADCs are designed for a certain resolution that is set by the application. For example, for a SAR ADC, the capacitor bank sizing, the sampling stage precision, the comparator dynamic settling error, and many more parameters are set based on the required quantization resolution. This means that a SAR ADC with a variable resolution will still have to meet the area and power requirements that are determined by the maximum resolution. This means that the benefits of an adaptive resolution ADC will only be limited to the wireless transmission data-rate and the recording channels power consumption will still be more or less the same. The second issue is related to the fact that the sequential operation of these ADCs are generally designed for a set resolution, and a variable resolution could cause serious complexity in terms of post-ADC data handling. Such irregular data handling requires complex digital timing controllers with substantial dynamic power consumption, which is against the initial goal of energy efficiency.

Driven by the above discussion, in this work, we propose to use an oversampling spectral-shaping amplifying ADC that (a) has a variable adaptive resolution that can be adjusted with a simple clock frequency variation, and does not require post-ADC data re-synchronization, (b) integrated a dynamic amplification stage inside itself, making the amplifier's power consumption adaptive as well, (c) leverages single-bit activity monitoring computation, which significantly reduces the dynamic power consumption of this module.

3.4 System-Level Design

It is expected that by employing the described activity-adaptive variable data rate recording strategy, the wireless transmission throughput, hence, the power consumption reduces significantly. This will bring the power consumption of this block to the same level as the array of recording channels. Therefore, it is crucial to ensure that the recording channels' power consumption is also scaling down proportional to the activity slow down. This demands for a recording channel architecture with a power consumption that is dominantly dynamic. Therefore, an increase/decrease in the clock frequency, not only changes the oversampling ratio of the quantizer (hence, its resolution), but also directly varies the power consumption of the recording front-end.

Conventionally, the EEG recording front-ends are designed based on an amplifier+ADC architecture, such as the circuit described in Chapter 2. However, in recent year, a new generation of fully-dynamic neural ADCs have been reported where amplification and digitization are both done in a single stage [31] [32]. Adopting this type of front-ends makes the recording channel's power consumption proportionate to the required sampling rate, hence, adaptive to the signal's activity. The top-level block level diagram of the proposed system is shown in Figure 3 3.

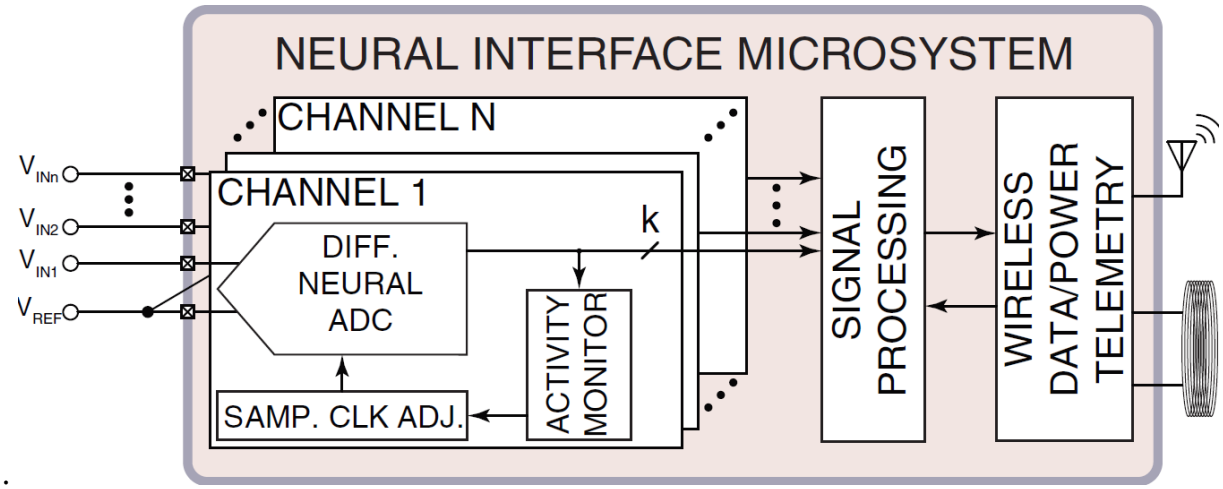

Figure 3 3 : Block level diagram

As shown, the EEG signals are fed to an array of differential fully-dynamic neural ADCs, which are responsible for low-noise amplification and quantization. The digital output of the recording channels are fed to an activity monitoring block where the required quantization resolution is calculated and fed back to the front-end neural ADC to make its performance adaptive to the EEG level of activity. The output of the recording channels together with the resolution information are fed to the wireless transmission block.

Since determining the level of EEG signal's activity is a major factor in the performance and efficacy of the proposed system, the activity-level evaluation criteria must be clearly defined. Based on the various EEG-based neurological event detection systems reported in the literature (e.g., [33] [34] [35]), it is commonly accepted that sudden bursts in LFP signals beyond the noise floor can be counted as events (in this text LFP “event” and LFP “activity” are used interchangeably). Accordingly, any period with no event is defined as idle period.

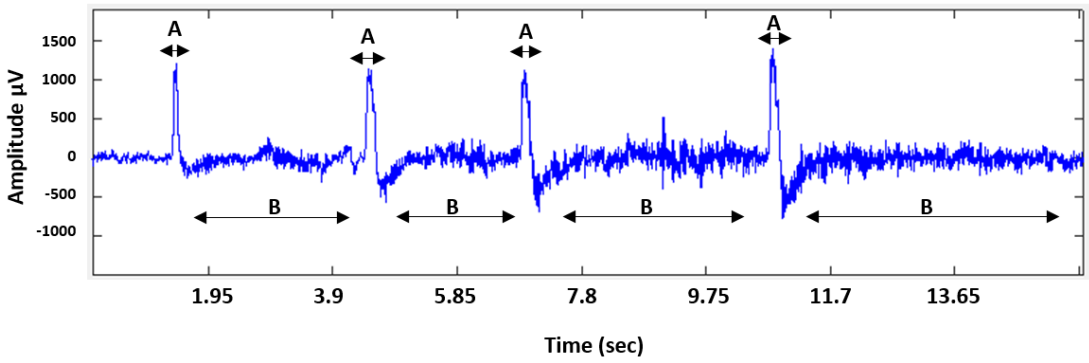

Figure 3 4 : LFP EEG recording with an event (A) and idle period (B) marked

Therefore, activity-level evaluation requires continuous comparison of the digitized signal's magnitude with different threshold levels. These threshold levels that are placed at different offsets above and below the signal's DC level are calculated based on a down-sampled window-averaged version of the signal itself, hence, are continuously updated. Crossing these thresholds raises the flag for the activity monitor that there might be an upcoming event. Of course, it is also considered that a single artifact-like spike might also trigger these thresholds and we have devised strategies to avoid such false positives as much as possible. Nevertheless, since the main task of this block is lossless compression of the signal, not event detection, mistaking an artifact for a neurological event has a near-zero importance for our purpose as the detection is done at a later stage. Therefore, for this stage we try to maximize the sensitivity of the event detection and allocate less priority to maximizing specificity.

3.5 Design and Implementation of the Recording Front-End

As mentioned earlier, the compressive sensing strategy requires a recording front-end that its quantization resolution varies with the signal's level of activity. In other words, the signals o high activity (e.g., neurological events) should be recorded with the maximum resolution, and the idle periods should be recorded with the lowest possible resolution. We also discussed that non-

uniform sampling, such as the work reported in [36] are not a good candidate for this application as they could result in late detection or completely missing an event.

Adopting an adaptive resolution strategy is expected to reduce the wireless transmitter's throughput requirement by at least an order of magnitude (more details in the next section). Such a drastic reduction makes the transmitter's power consumption in the same order of magnitude as the array of recording channels. Therefore, making the power consumption of each recording channel similarly adaptive to the signal's activity level could significantly reduce the system's overall power consumption.

In this work, we present a fully dynamic neural front-end oversampling Delta-modulated ADC that will be used to both amplify the input LFP signals and quantize them with an activity-dependent variable resolution. The signal to noise ratio (SNR) of the oversampling ADCs are calculated as

$$SNR = 6.02N + 1.76\,dB + 10\,\log(OSR)$$

Where N is the number of bits of the quantizer (e.g., N=1 for a simple comparator), and the oversampling ratio (OSR) is defined as $f_s/(2.BW)$, where f_s is the sampling frequency and BW is the input signal bandwidth. It is clear from the above equation that the SNR, and accordingly, the ADC's resolution, has a direct relationship with the OSR. For a fixed signal bandwidth (e.g., 500Hz for LFP signals), changing the sampling frequency, f_s, will result in changing the quantization resolution. Therefore, by simply changing the f_s, and keeping the decimation period intact, the ADC provides a digital output with different resolutions that are proportional to the f_s.

Additionally, we replaced the conventional ac-coupled two-stage OTA & OpAmp architecture with a single stage rail-to-rail differential difference pre-amplifier (DDA). The

differential difference architecture makes the design needless of the large AC-coupling capacitors (discussed in chapter two) and removes the DC offset by the assistance of the digital output of the delta modulator. The co-dependence of the pre-amplifier and the oversampling quantizer makes them a single differential-difference oversampling neural ADC stage.

3.5.1 System-level Idea

As shown in Figure 3 5, the neural activity throughout the Electrocorticography (ECoG) and local field potentials (LFPs) spectra range (< 500 Hz) has a non-flat spectral density, in which the signal magnitude increases with a factor of 1/f as the frequency approaches zero Hertz [37] [38]. Therefore, the required dynamic range (DR) for a neural-ADC easily exceeds 50-dB disregarding stimulation artifacts, which could easily increase the required dynamic range up to 70-dB. Such a high dynamic range will significantly increase the ADC power consumption. To address this challenge, we propose a neural ADC architecture, which Δ- equalizes the input neural signals spectra to an almost flat spectrum, reducing the dynamic range required for neural signal.

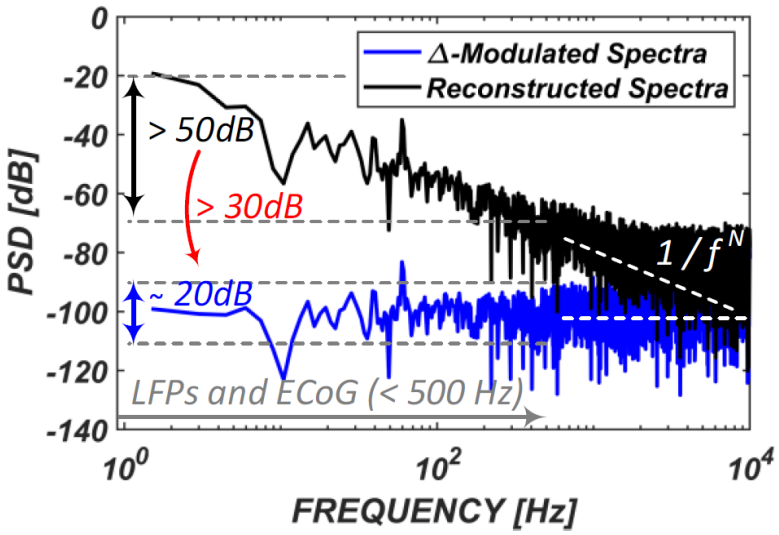

Figure 3 5: Spectra comparison of Δ-modulated and Σ-demodulated neural signal.

Delta modulation is based on quantizing the change in the signal from sample to sample rather than the absolute value of the signal at each sample. Since the output of the integrator in the feedback loop of Figure 3 6, tries to predict the input x(t), the integrator works as a predictor. The prediction error term, x(t) – $\bar{x}$(t), in the current prediction is quantized and used to make the next prediction.

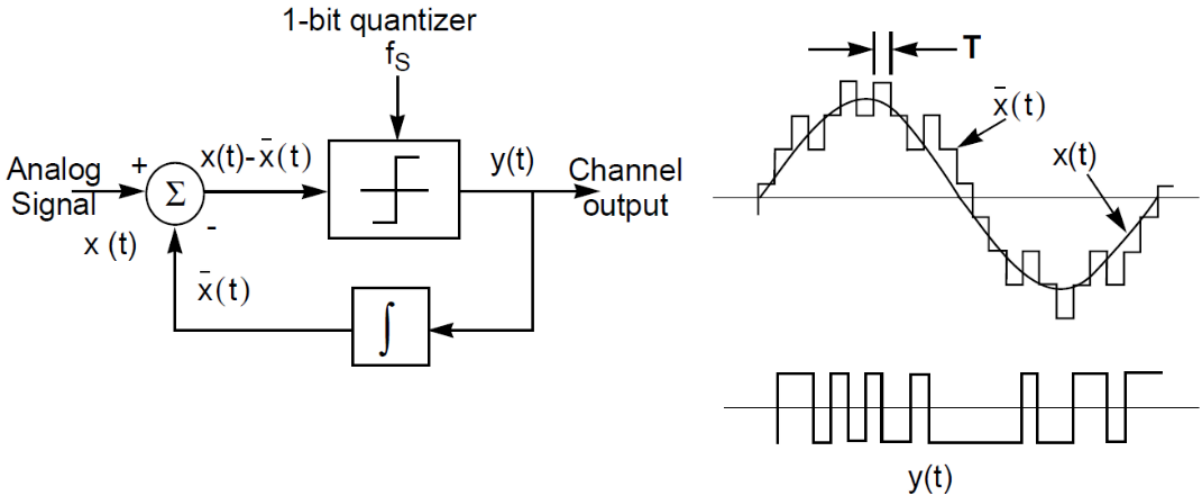

Figure 3 6 : Delta modulation principle [39].

Figure 3 7, shows the block diagram of the proposed differential-difference Δ-modulated ADC. From the system-level point of view, the same prediction as in a normal delta modulator is performed here, with the main difference that the feedback loop tries to predict the difference between the two differential inputs. Therefore, one input of the comparator is an amplified version of the difference between V_{in+} and V_{in-}, and the other input is the predicted version of the same signal based on the previous sample. Given the steady nature of DC offset, it is expected that the offset is perfectly predicted. Therefore, the difference between the two inputs of the comparator is basically the small change in the differential signal from one sample to the next. Therefore, an output "1" of the comparator means that the differential signal is increased, and an output "0"

means a decrease in the differential signal. Consequently, the output bitstream, $D_{MOD}[n]$, represents the variation of the signal (i.e., its derivative), rather than its magnitude, hence, the name Δ modulator. We will describe in the next section that a simple integration can reconstruct the signal magnitude from this bitstream.

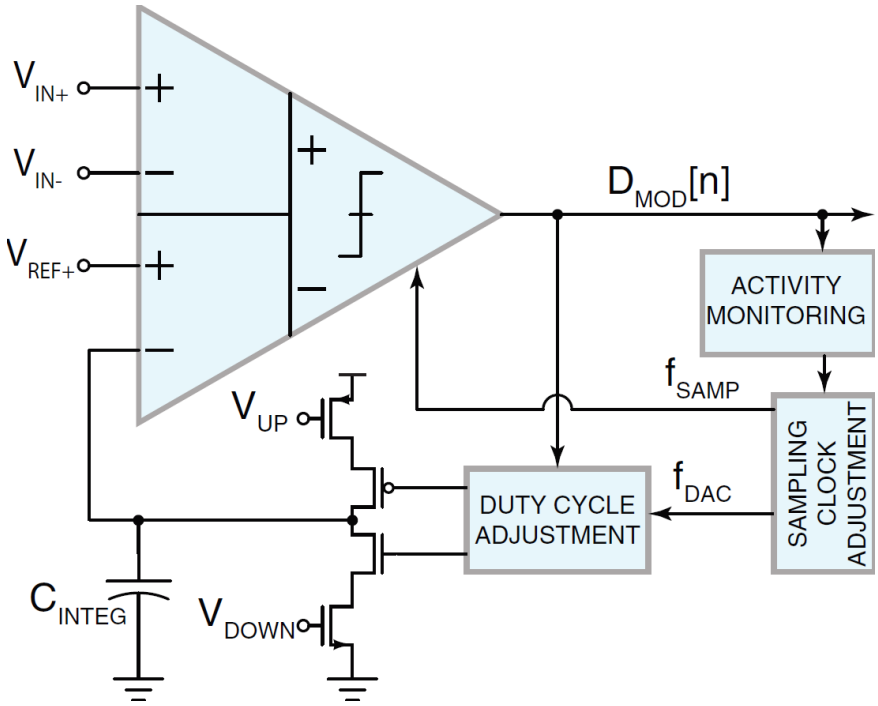

Figure 3 7 : Block diagram of the proposed differential-difference Δ-modulated ADC

Since the LFP signals frequency bandwidth are limited to 500Hz, utilizing a high OSR would significantly help reducing quantization noise. Therefore, a binary comparator as the quantizer would be sufficient to achieve the required SQNR. To realize spectrum equalization, input signal is Δ-modulated before quantization, and then demodulated in the digital domain. The Δ-modulation is implemented using a feedback loop as shown in Figure 3 7. This approach reconstructs the delayed replica of the input signal by the integrator loop filter and subtracts it from the input signal by a binary comparator.

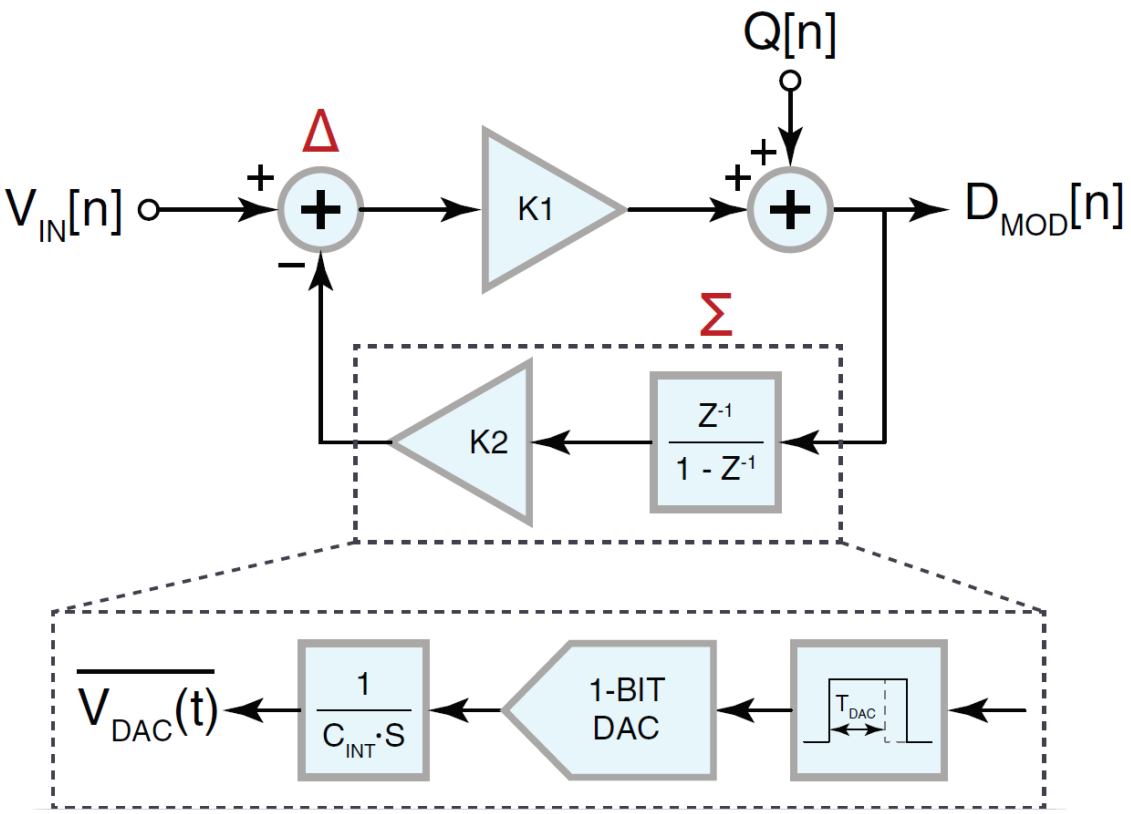

Figure 3 8 : Feedback architecture of the proposed design

From a discrete-time signal analysis point of view, the feedback architecture of the proposed design is shown in Figure 3 8. $D_{MOD}[n]$ is the modulator's output binary sequence. It is fed back to loop filter to generate a delayed replica of input signal, *i.e.* $\overline{V_{DAC}(t)}$. As shown, in the feedback path, the binary output of the 1-bit quantizer is fed to a 1-bit digital-to-analog converter (DAC), and then is integrated. Prior to the 1-bit DAC, a duty-cycling stage is placed that can determine the integration period. As shown in Figure 3 7, the 1-bit DAC is implemented using a push-pull charge pump, and the integrator is implemented using a capacitor that is connected between the output of the charge pump and ground.

The above denotes that: 1) gain of the DAC, 2) integrator capacitor charging time by either a negative or positive polarity, 3) size of the integrator capacitor, and 4) sampling frequency are

the design knobs in the proposed neural-ADC. Therefore, assuming constant loop filter gain at low frequency, the size of the integrator capacitor can be reduced significantly at a high OSR as its value is inversely proportional to sampling frequencies. This would enable replacing the large (tens of μF) feedback capacitor of the conventional differential-difference neural amplifier (DDNA) by a compact capacitor in the pF range.

3.5.2 Circuit Implementation

Figure 3 9, presents the circuit diagram of the proposed OpAmp-less neural recording channel. It includes a DDNC, which consists of a preamplifier and a single-bit comparator (depicted in Figure 2 14) and a low-power feedback integrator charge pump as the modulator's DAC. The role of the clocked-preamplifier is to provide: 1) a low-noise bandlimited neural signal amplification gain before quantization, 2) a high input impedance isolation between the electrodes and ADC, and 3) a differential stage, which scales well with technology and voltage supply and operates rail-to-rail removal of common-mode interferers. The preamplifier's transconductance gain is programmable by changing the DC currents flowing through each branch (I_N and I_P). Symmetric design and layout are other important aspects of the differential-difference preamplifier to achieve high common-mode rejection ratios (CMRR). Therefore, multi-fingered common-centroid layout techniques were considered to make the layout compact and minimize gate resistance and local gradients.

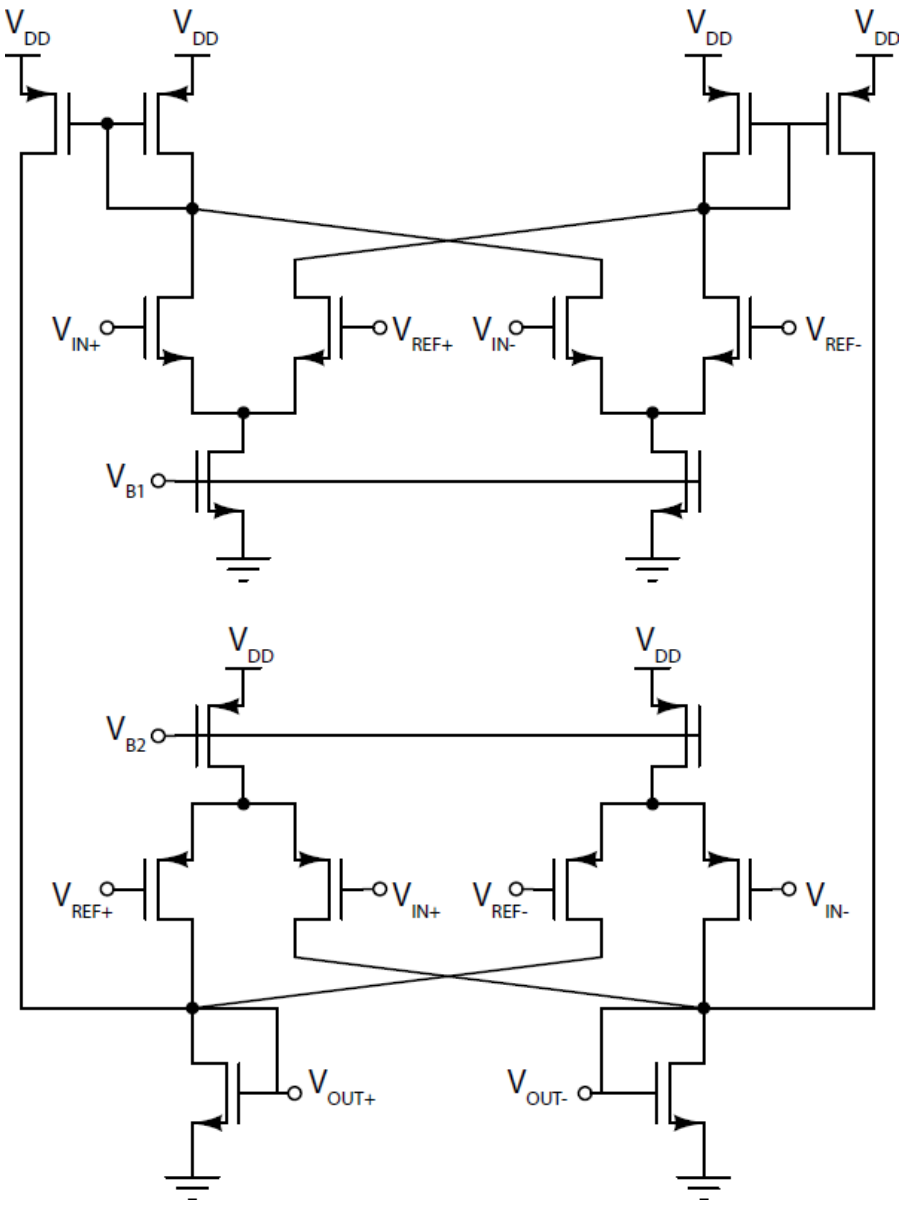

Figure 3 9 : Circuit diagram of differential difference amplifier

Given the fact that the preamplifier is integrated into the body of ADC, it does not amplify the large transients caused by stimulation or motion artifacts before the ADC, and hence is more tolerant of artifacts. The preamplifier, however, needs to provide a moderate low-noise amplification to relax the single-bit quantizer sensitivity and not adding much to the input referred noise to achieve a decent overall noise efficiency factor (NEF) including the noise sources associated with the ADC. Thus, the overall input referred noise (IRN) of ADC is comprised of

quantization noise, electronic noise of preamplifier, and feedback integrator. Low IRN is achievable by selecting a large (> 1 pF) feedback capacitor voltage integrator (C_{INT}) and input transistor devices while the gate-leakage current does not exceed the safety threshold corresponding to super-GΩ DC input impedance. Moreover, since the electrodes are directly connected to the gate of the transistors, DC offset voltage of input devices gets removed by Δ-modulation. Consequently, mixed-signal feedback makes the neural interface DC-coupled without large AC-coupling capacitors making it friendly with chopper stabilization (as the input impedance is proportional to C_S^{-1}). The preamplifier also completely isolates the electrodes from following stages noise such as kick-back, and more importantly, from the feedback loop eliminating the risk of charge injection or any short-circuit DC current flowing into the electrodes. Both PMOS and NMOS devices are utilized to enable rail-to-rail transconductance amplification as well as DC offset removal. More importantly, since the quantizer makes a binary decision, the linearity of preamplifier does not affect the overall linearity of the neural-ADC, and the preamplifier can operate without any common-mode feedback (CMFB) circuits as well.

The differential difference amplifier has open loop gain of 18.8dB. Based on the simulation results, strong-arm comparator is capable of reliably detecting voltage differences as small as $50\mu V$. Considering that the smallest EEG feature of $10\mu V$ as defined earlier in the specifications, the pre-amplifier's gain is large enough to make, the signal's amplitude sufficiently large for the comparator to yield the correct result. Frequency response analysis for this amplifier indicated 3db bandwidth of 27 KHz, which exceeds system level requirements.

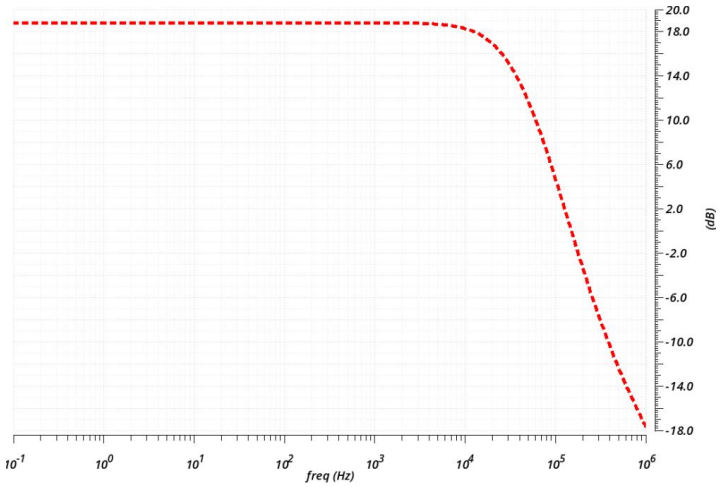

Figure 3 10 : Frequency response of the differential difference amplifier

Input referred noise performance of this block was verified using spectre noise analysis. High frequency noise floor was recorded to be 169 $nV/\sqrt{hz}$ and the integrated noise from 1Hz to 5 KHz is 20.6 μV. As shown in Figure 3 11, the input-referred noise is dominated by the flicker noise, which means that the addition of a chopper stabilization stage to the pre-amplifiers could reduce the noise level significantly. The proposed architecture has the advantage of not having the ac-couple input capacitors, therefore, adding chopper switches won't result in fold-back noise issues, as it does with the conventional architectures.

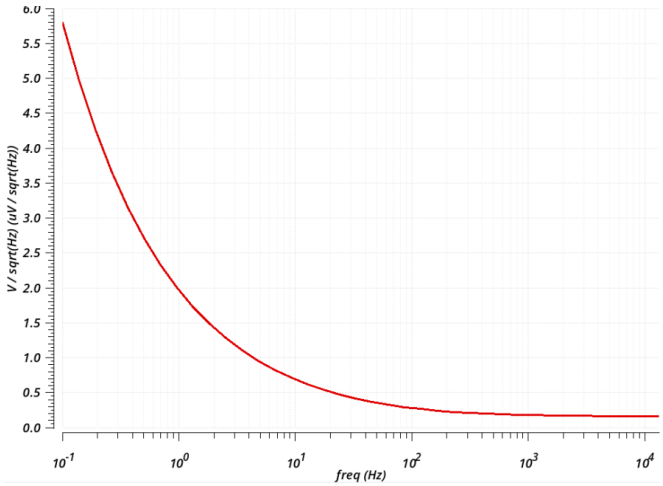

Figure 3 11 : Input referred noise of differential difference amplifier

3.6 Activity Monitoring Design and Implementation

As described in the previous section, the front-end neural ADC's output bit stream represents an oversampled derivative of the input signal. Therefore, in addition to decimation, which is required for any oversampling ADC, the bit stream must be integrated as well to obtain the signal's magnitude itself.

Figure 3 12, shows the top-level block diagram of the digital activity monitoring block. As shown, following the integration, a 10-bit nyquist-rate version of the signal magnitude is obtained and fed to a thresholding block that continuously monitors the relative change of the signal's magnitude with respect to a dynamically-calculated DC level. The DC level itself is also computed using the signal magnitude as well, as it will be described later.

The thresholding block generates a 4-bit number that represents the deviation of the signal magnitude from the DC level. Based on this 4-bit number, the activity level of the signal is evaluated and proper commands to adjust the quantization resolution are transmitted back to the recording front-end.

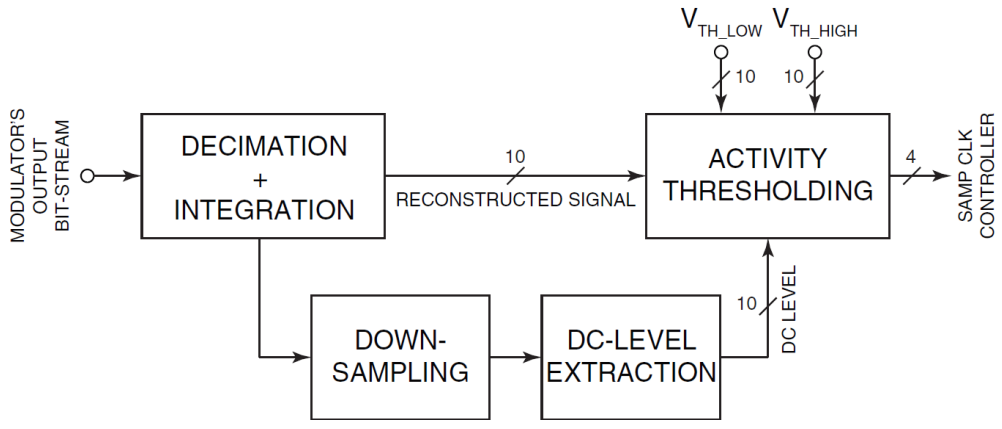

Figure 3 12: Digital back-end block diagram

As the first step, the proposed algorithm was implemented in MATLAB and was verified against real EEG database. This database was provided through the courtesy of [27], which was recorded in the Children's Hospital Boston. All signals were sampled at 256 samples per second and 16-bit resolution. Complete MATLAB code is included in the Appendix.

Using resampling function in MATLAB the data was up-sampled by a factor of 100, this will up-sample the recorded data from 256 samples/sec; to 25 kSample/sec. Resampling function simulates the effect of sampling clock adjustment. Initially, two threshold levels were defined and used to control the sampling rate. We should mention that the number of threshold levels, i.e., the number of different quantization resolutions is an arbitrary number and can be more than two, if desired. If the current LFP level is below threshold level one, sampling rate is reduced to 256 samples/sec; crossing the first threshold will result in increasing the sample rate to 2.5 kSample/sec; and ultimately, once the second threshold is passed, sampling rate is maximized at 25 kSample/sec.

DC offset is calculated and is used to update both threshold levels. The calculation is carried out by a weighted averaging process where a new sample of the magnitude is taken every 40 clock cycles, and the average of the current DC value and a new sample are calculated with weights of 7 and 1, respectively. The weighting and the down-sampling ratio are set to maximize the correlation between the normal and compressed signals, and both are adjustable. The results are plotted in figure 3 13, showing the original EEG signal at the top (blue) and the output of the algorithm at the bottom (red). Region "A" denotes a significant EEG event and therefore is sampled at 25 kSample/sec, region "B" represents a minor EEG event and is sampled at 2.5 kSample/sec however, region C is the idle region were sampling rate is further reduced to 256 samples/sec. Each region defines a time window with uniform sampling rate.

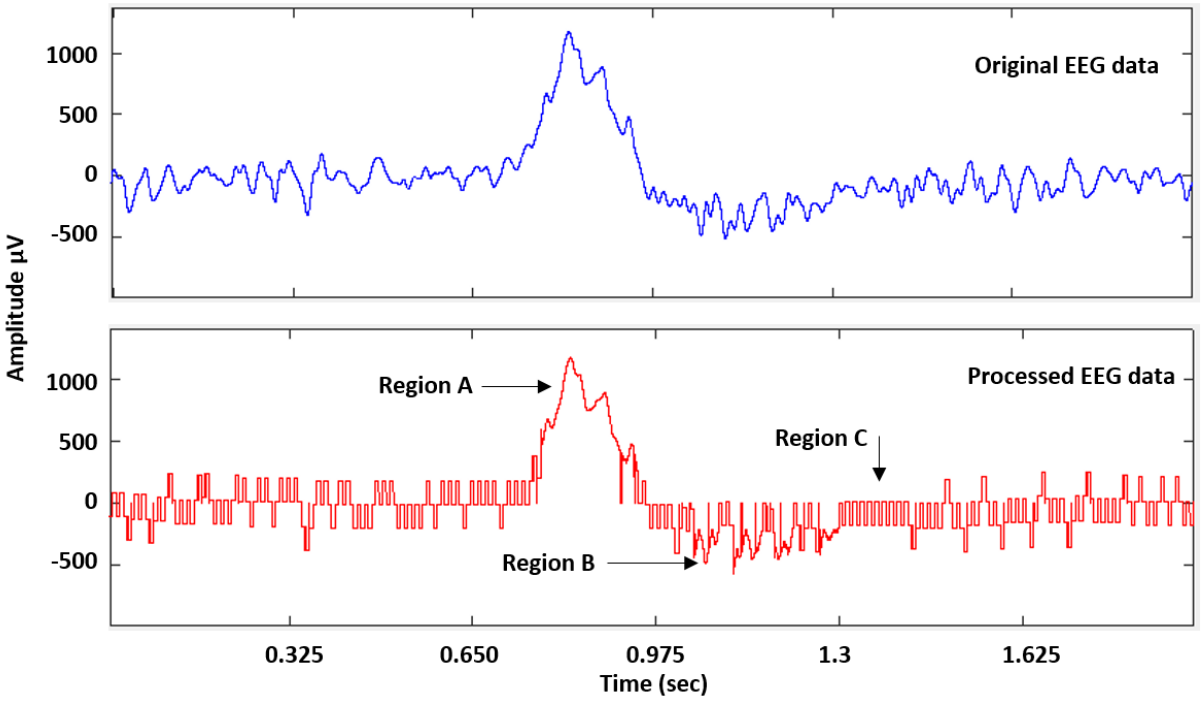

Figure 3 13 : Plot of sampling rate adjustment algorithm

As evident from the above example, the maximum sampling rate is only used for a short period of time. Since the quantization resolution is proportional to the sampling rate, this means that the recording data rate for the majority of the time is reduced by two orders of magnitude. Given that the wireless data transmission power dissipation is directly proportional to its throughput, such a dramatic reduction in the recording data-rate translates into an orders of magnitude improvement in their energy efficiency. Moreover, given the fully dynamic architecture of the recording front-end, the power consumption of the recording array is proportionally reduced with the sampling rate reduction.

The algorithm was verified by testing its efficacy on the data from multiple patients in the EEG databased and comparing the correlation between the normal and compressed signals for each case. Upon successful verification of the algorithm, the digital VLSI implementation was performed in TSMC 0.18um CMOS technology. Amplitude reconstruction block uses two 11-bit

binary up-counters, one that counts the number of 1s and the other counts the number of 0s received form the comparator. Number of 0s is subtracted from the number of ones to calculate the amplitude. It's worth noting that the sign of reconstructed EEG data is assigned according to "Two's complement" operation.

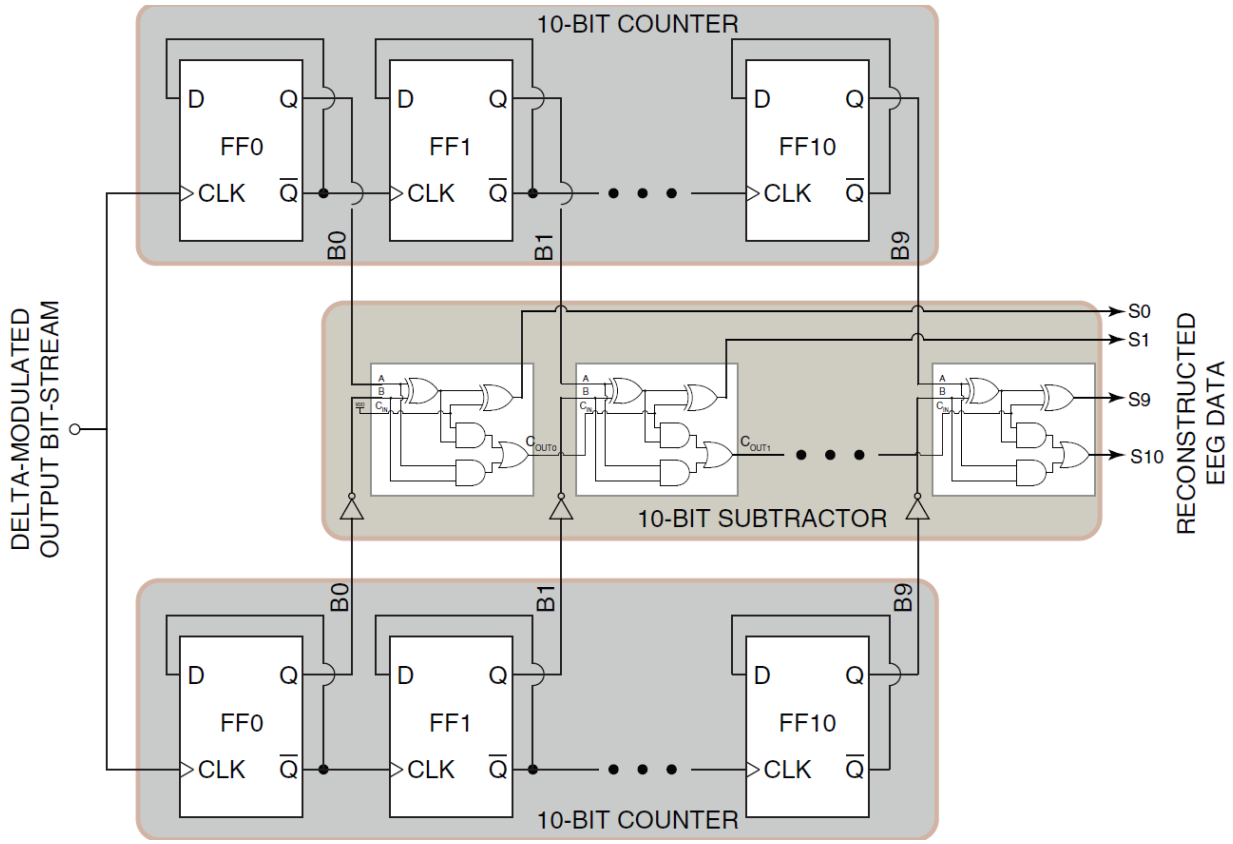

Figure 3 14 : Circuit diagram of amplitude reconstruction block

Results of transient analysis for the amplitude reconstruction block is shown in Figure 3 15. Only the first 4 bits of the output are plotted for better visualization. The reset for this block is active low and once reset is lifted, it continuously detects 1s and 0s and updates the amplitude accordingly.

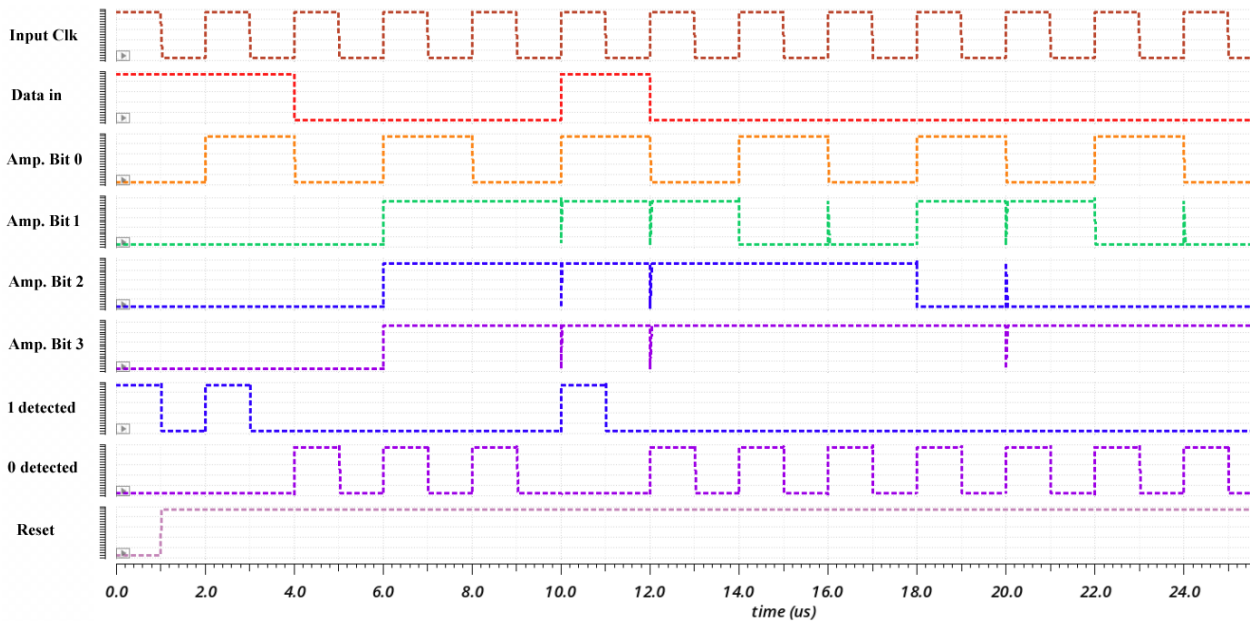

Figure 3 15 : Transient analysis for amplitude reconstruction block

As briefly mentioned earlier, to extract DC level form EEG signal's amplitude, a combination of down sampling and low-pass filtering was utilized to attenuate and remove any temporary fluctuations of the signal. An effective method to implement a low pass filter in digital domain is the use of averaging [40]. In its simplest form, a sampled data D_n is summed with the previously sampled value $D_{(n-1)}$ and also the subsequent sample $D_{(n+1)}$, dividing the sum by 3 provides a level representing the LPF of 3 sampled values.

$$D_{LPF} = \frac{D_{(n-1)}+D_n+D_{(n+1)}}{3} \text{ [40]}$$

In this simple form, the results are sensitive to large short-term variations in each sampled data. In other words a large deviation from the mean can shift the average value. This is equivalent of a high cut-off frequency in low-pass filtering. Increasing the number of samples will improve the results at the cost of higher power consumption and memory. Therefore, we decided to use a two-pronged strategy where both down-sampling and weighted averaging are used. In doing so,

$D_n$was added to $7 \times D_{(LPF-1)}$ and divided by 8 and new D_n is sampled every 40 clock cycles. It should be noted that initial value for $D_{(LPF)}$is programmable and can be set for different applications.

$$D_{LPF} = \frac{D_n + (D_{(LPF-1)} + 2D_{(LPF-1)} + 4D_{(LPF-1)})}{8}$$

To simplify circuit implementation of binary multiplication, $7 \times D_{(LPF-1)}$ was synthesized in the form of $D_{(LPF-1)} + 2D_{(LPF-1)} + 4D_{(LPF-1)}$. Therefore, the expensive 10-bit multiplication is replaced with two inexpensive bit-shifting operations, hence, improving both areas and power efficiency of implementation. Circuit level diagram for DC level extraction is shown in figure 3 16.

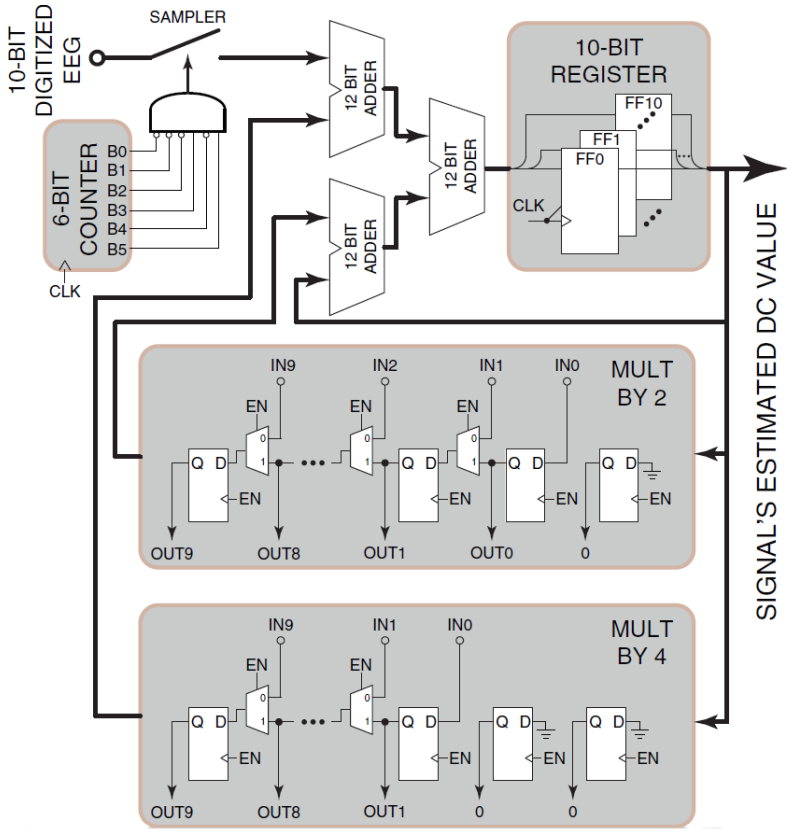

Figure 3 16 : Simplified circuit diagram of DC extraction block

Figure 3 17, shows an example of transient simulation results for this block. Input and output of DC extraction block are 11 bit binary buses and the values of these terminals are shown in decimal system using analog to digital conversion of Virtuoso's ADE. As shown below at $t = 7\mu sec$, value of the current sampled signal (i.e., D_n) is (10000000100 bin = 1028 dec) and the value of the output (i.e., the stored DC level before updating) is (01110010000 bin = 912 dec). The circuit updates the DC value and makes it available for the next 40 clock cycles to (01110011110 bin = 926) as expected:

$$Updated\ DC\ level = \frac{Sampled\ Data + (7\ \times Previous\ DC\ level)}{8} = \frac{1028 + (7 \times 912)}{8} = 926.5$$

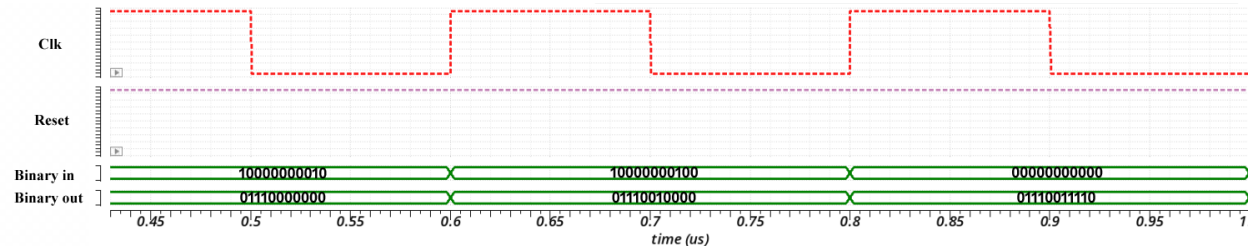

Figure 3 17 : Transient analysis of DC extraction block

Threshold values are defined with respect to EEG signal's DC level, and the sampling rate should be adjusted irrespective of the polarity at which threshold was crossed. Two separate threshold values are both added (to providing positive threshold) and then subtracted from (providing negative threshold) the signal's DC level that was extracted earlier. Subsequently, to compare the signal with these threshold values, the LFP's amplitude is subtracted from the positive threshold, and the negative threshold is subtracted from the amplitude for every clock cycle. Using subtractor as a comparator allows for the time-sharing of the subtractor that we have already implemented, hence saves silicon area. Based on the sign bit of the subtraction result, it can be determined if the threshold was crossed. A simplified diagram showing circuit view of threshold detection circuit is depicted in Figure 3 18.

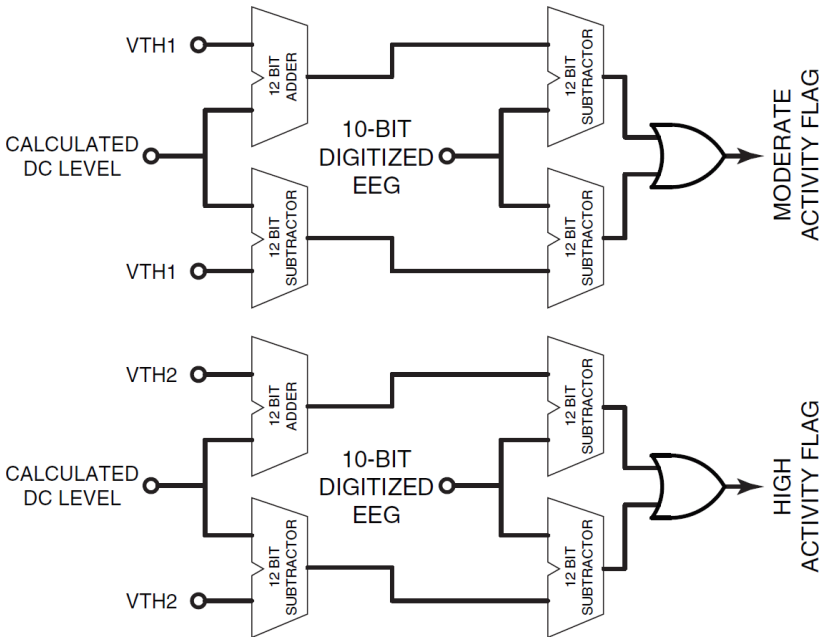

Figure 3 18: Threshold level detection circuit

Sampling clocks were generated form a reference clock with help of a D-flip-flop "DFF" based frequency divider. The divider consists of 7 DFFs in counter configuration and a 3 input mux is used to switch between undivided reference clock (for maximum sampling rate of $256 \times 2^7 \sim 32.7\ KHz$), output of 3rd DFF (for minor event rate of $256 \times 2^4 \sim 4\ KHz$) and output of the 7th DFF (for idle region with sampling rate of 256 Hz). The mux is controlled by the output of "threshold detect" block.

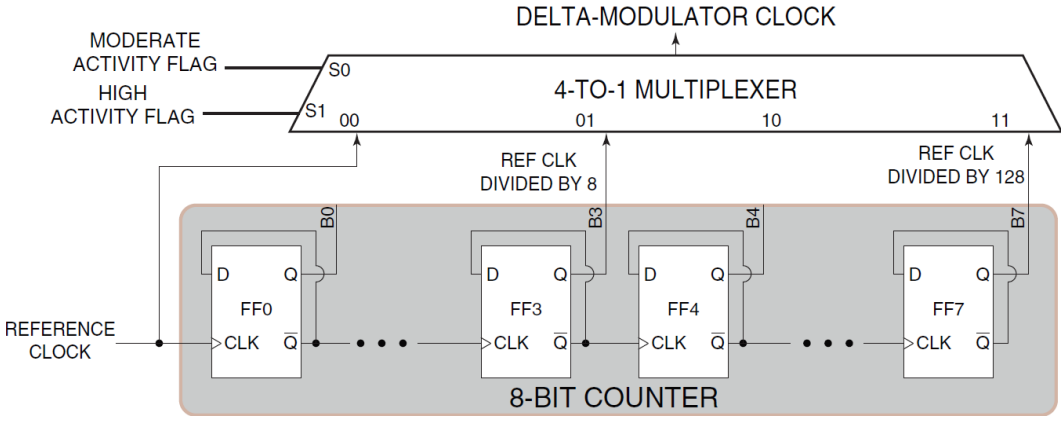

Figure 3 19: Circuit diagram of sample clock generator

Functionality of the sample clock generator block was verified using transient simulation. Here the results for Divide by 8 mode are plotted. It's worth noting that divided output is available both with 0 degrees and 90 degrees phase shift. The 0 degree clock is used by comparator to slice the data and the 90 degrees clock is used to sample the data.

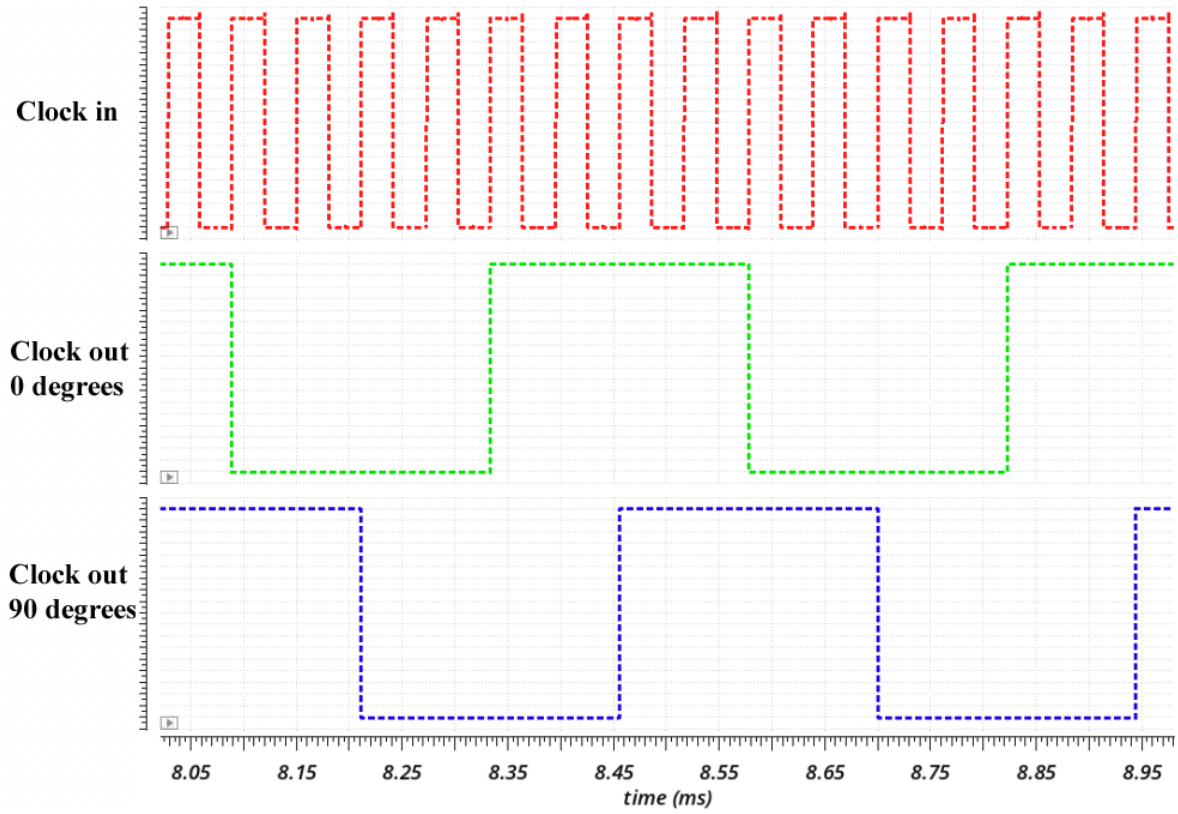

Figure 3 20 : Transient analysis for clock generator

3.7 Integration and Performance summary

The final step in the design of enhanced EEG ambulatory recorder with activity dependent sampling rate adjustment was the integration of all components together. All previously verified blocks were assembled together as illustrated in Figure 3 21.

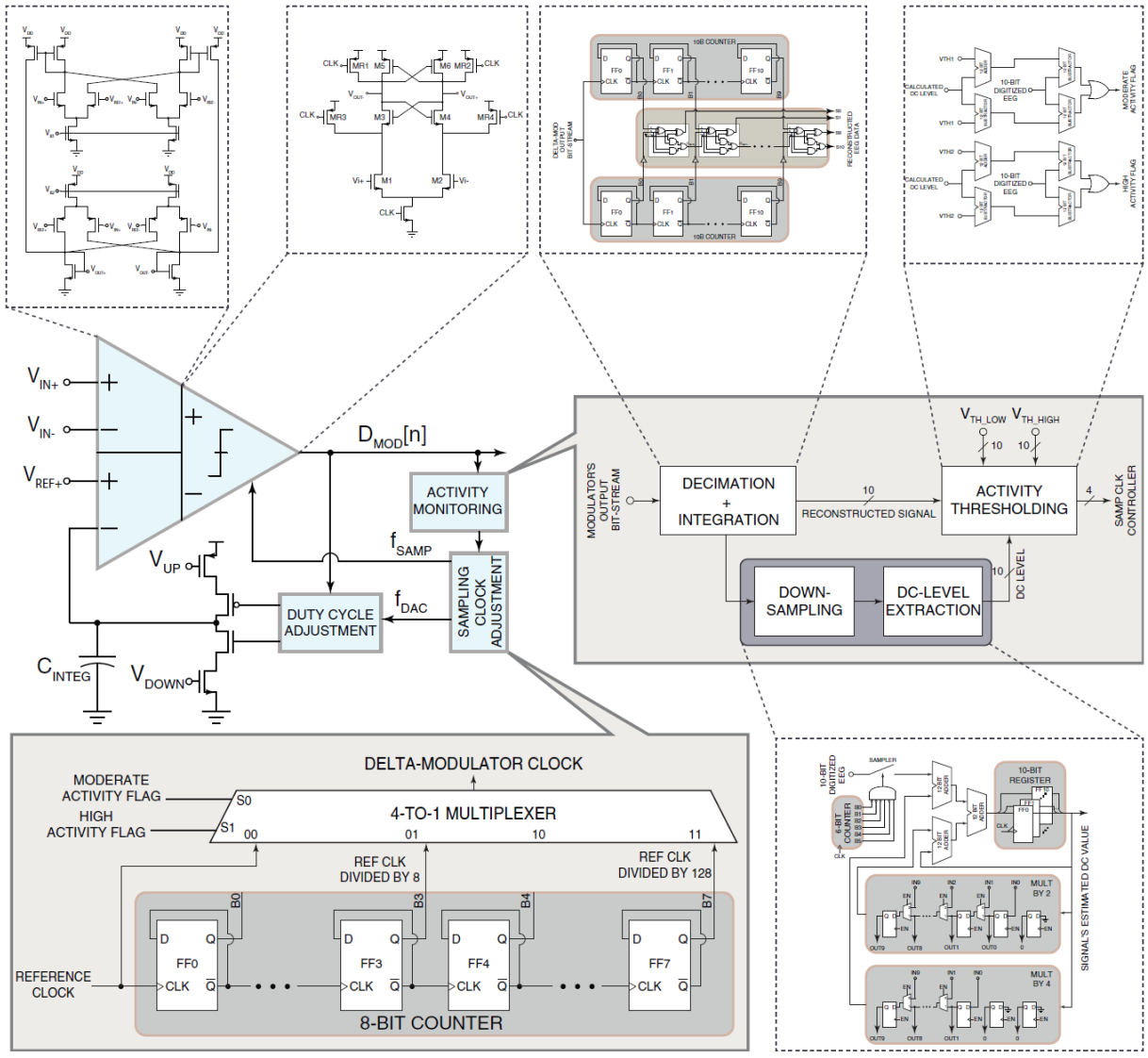

Figure 3 21: Integrated blocks diagram

The mixed-signal recording front-end consumes 2.7 μA RMS in max operation mode and as low as 400nA RMS in idle mode. The comparator and charge pump have a combined current consumption of 730nA RMS when fully operational. The remaining digital logic for activity monitoring and sampling rate adjustment draw 83 μA RMS current.

To evaluate the power savings resulted from the presented technique, we first characterized sets of offline pre-recorded EEG recordings for 10 pediatric subjects which were collected at

Boston children hospital [27]. Using this data and the threshold values from section 3.6, we first calculated the number of samples in each activity region (as defined in figure 3 13) for every patient and then converted the number of samples into duration of time. To make the results a better statistical representation of bigger population, the activity periods were averaged between all patients. according to the MATLAB results, on average over 56 minutes out of 1 hour of EEG recording was in idle region, and duration of all activities sum to just slightly less than 4 minutes. Based on these findings for nearly 93% of the time, sampling rate will be reduced to just 256 sample per second (a reduction factor of 100x). Since power dissipation of transmitter is directly proportional to sampling clock's frequency, this means that the power consumption is reduced to 7%+0.01*93%=7.93% of the conventional architecture (a reduction factor of 12.6x). It should be noted that this power reduction is valid for both the data transmission block (due to the same throughput reduction factor) and for the array of recording channels (due to its dully dynamic design).

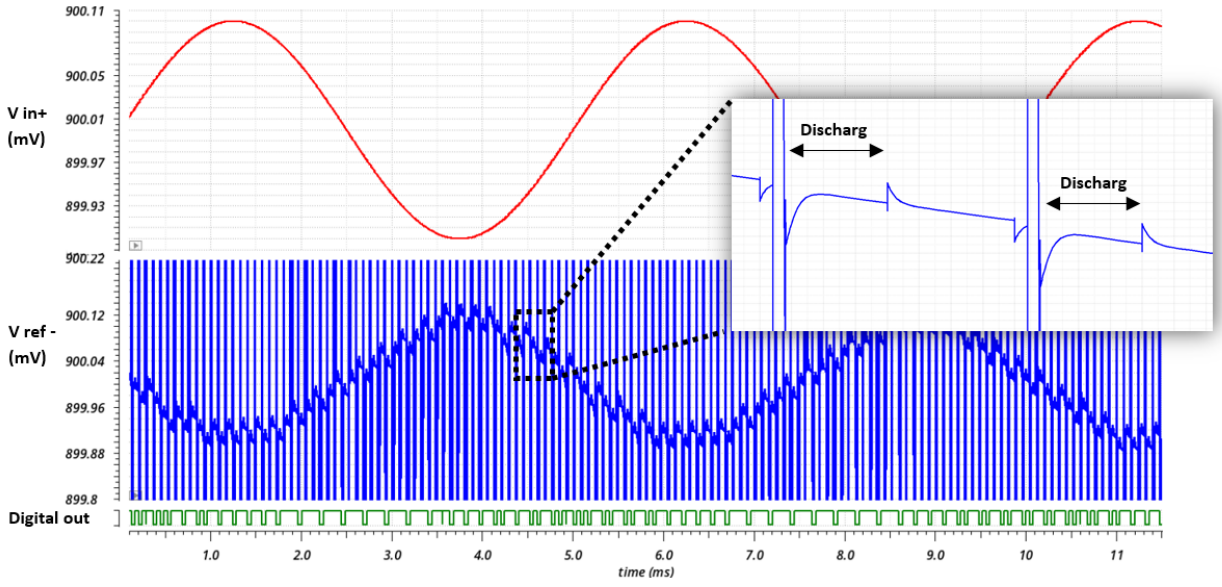

Figure 3 22 : Synthetization of V ref (–) in response to input sin wave at V in (+)

Once all sub blocks were individually verified for functionality, it's essential to validate the performance of newly integrated recording channel. Transient analysis was performed on the loop as illustrated in figure 3 22. In this test, a 200Hz sine wave was applied at the V_{in+} terminal of mixed-signal front-end. The V_{in-} & V_{ref+} terminals were set to the mid rail (0.9V) common mode potential. In this configuration, the expectation is for V_{ref-} to inversely follow the V_{in+}. As shown in figure 3 22, the charge pump charges and discharges the C_{INTEG} (see figure 3 21) to synthesize the V_{ref-} (which is plotted in blue) as close as possible to the inverse of the input signal. Additionally, looking at the digital output bitstream in figure 3 22 (Green curve), we observe increased rate of 0s as slope of V_{in+} increases and increased rate of 1s as slope of V_{in-} increases. This confirms our expected behaviour from the delta channel to have an output bitstream that reflects the derivative (i.e., the slope) of the input signal. As discussed earlier (Figure 3 14), these bits are later used to reconstruct the signal in the digital domain.

A high-level comparison between the projects that were discussed in chapter 2 & 3 are presented in table 3 1, the results are indicative of significant power savings in the latter design for AFE and ADC. As mentioned earlier, digital signal processing "DSP" in the dynamic sampling design adds 83 μA RMS current however this additional current is negligible in compared to potential savings of upto 9 mA (based on 10x power reduction) in the transmitter current. Additionally, it must be mentioned that since this was a proof-of-concept prototype, the DSP block was designed in a very performance-conservative fashion. We believe that the power consumption of this block (i.e., 83uA x VDD) could be significantly reduced in the next iterations of this design.

	Conventional Recording	Activity-Adaptive Resolution Recording
ASIC area per channel	0.14 mm^2	0.19 mm^2
Technology	TSMC 0.18 µm	TSMC 0.18 µm
Analog front end power consumption	3.7μA RMS	2.7 μA RMS
ADC power consumption	14.4μA RMS	730 nA RMS
DSP power consumptions	N/A	83uA
Data Tx estimated current consumption	10 mA (Max)	<1 mA (Average)
AFE Gain	48.8 dB	18.8 dB
Bandwidth	15 KHz	28 KHz
Int. input referred noise	6 µV RMS (1 Hz – 5 KHz)	20.6 µV RMS(1 Hz-5KHz)
ADC resolution	10 bit	5.67 bit Max
ADC type	SAR	Delta

Table 3 1: System-level performance comparison between the designs presented in Chapters 2 and 3.

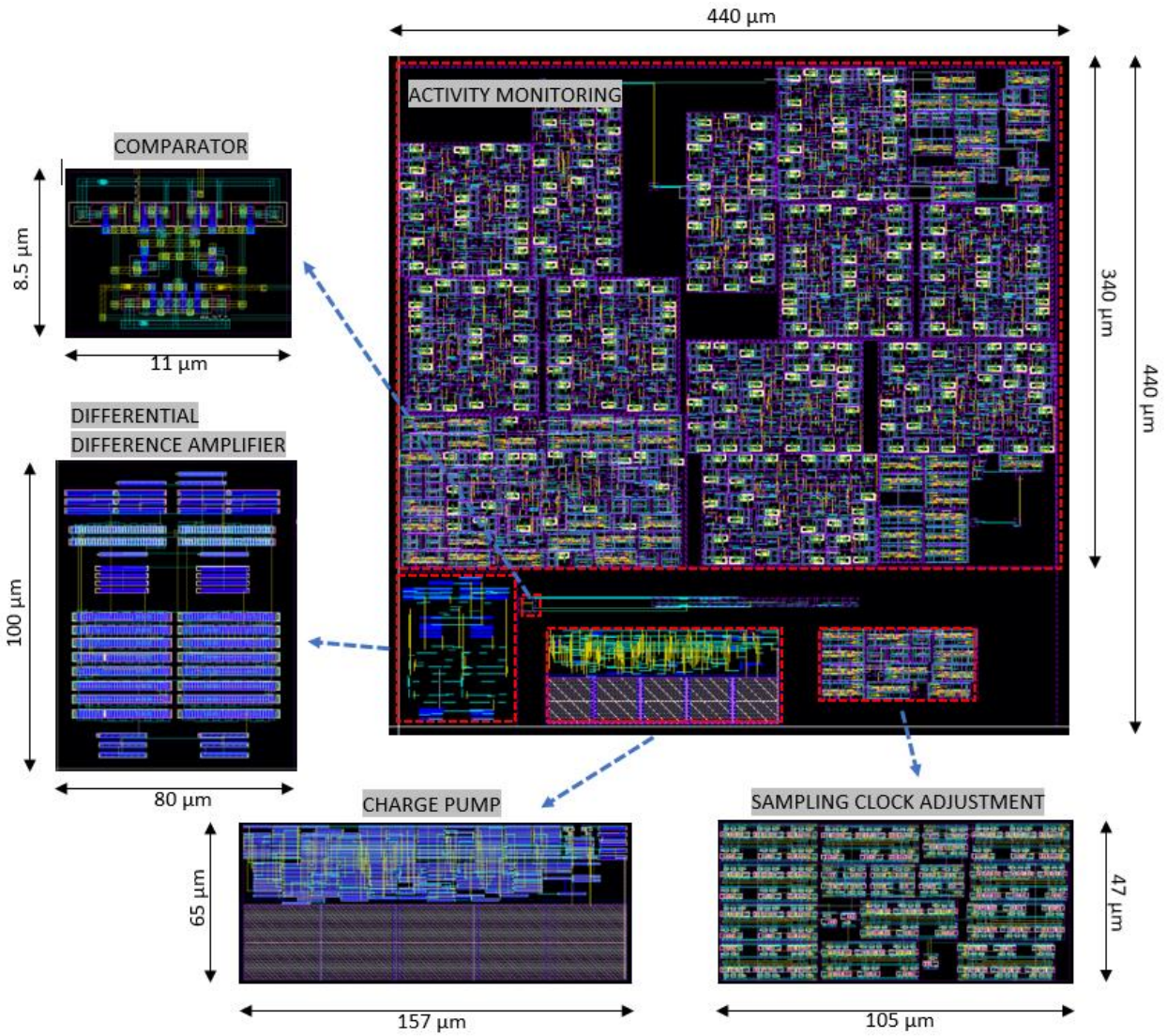

Figure 3 23 : Layout view of Activity-Adaptive EEG Recording channel

CHAPTER 4:

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Conclusion

This dissertation presented the design, implementation, and characterization of two implantable microsystems for long-term ambulatory EEG monitoring. The first design introduced an innovative 8 channel recording system with a Bluetooth LE module for wireless data telemetry and organic electrodes. The latter design was built on the learnings of the earlier work and was focused on extending the battery life by increasing the energy efficiency of the device with a particular focus on the wireless data link.

The work covered in chapter 2 discussed the system and circuit level design of a fully implantable EEG monitoring system for ambulatory application with 8 differential recording channels. This microsystem was fabricated in TSMC 180nm CMOS technology. Post fabrication, it was tested in laboratory with the help of custom made PCB board that was specifically designed for this application. The ASIC footprint occupied an area of 3.75 mm^2 and consumes 3.7μA RMS current for the analog front-end and 14.4μA RMS for data converter and other applicable digital logic. It was also shown that the BLE wireless module is by far the most power-hungry block in

this design, drawing up to 10 mA of current during transmission. This leads to sub-optimal battery charging cycles, hence, making the device less practical as an ambulatory solution.

Motivated by this, a holistic approach was taken to design and implement an energy-efficient implantable device, with a particular focus on the power consumption required for wireless data transmission. The innovative technique which was introduced in this chapter was focused on activity-dependent adaptive sampling and quantization. It was demonstrated that EEG signal's LFP activity footprints are very sparse in time domain. An algorithm was developed and tested in MATLAB that showed EEG data can be divided into different regions based on level of activity. It was also proven that by dynamically reducing the sampling rate in idle regions, the power consumption of the wireless transmitter can be reduced by more than an order of magnitude.

To maximize the efficacy of the proposed idea, the recording front-end circuit was also redesigned from scratch to have a fully-dynamic power consumption, hence, experience a similar power reduction benefit due to the adaptive sampling rate. The new mixed-signal front end, digital signal processing blocks and reference clock generation were implemented and validated both at the circuit and system levels. Subsequently, the new activity dependent recording channel with dynamic sampling rate adjustment was integrated and successfully validated against all design requirements.

4.2 Future work

Although both projects were executed as expected and measurement results and simulation data do validate these SoCs functionality and system level requirements, there are certain areas where improvements can be made for the next iteration of the prototype.

4.2.1 Integral transceiver

Design of the radio transceiver was beyond the scope of this project due to time constraints and extensive span of the design, therefore a standalone Bluetooth low energy module was used for data telemetry.

For next generation of this design, an integrated RF transceiver can be incorporated into the microchip for further miniaturization as well as potential power reduction. A custom made wireless data telemetry block can be optimized to take full advantage of activity based sampling rate adjustment that generates data with variable bandwidth. Current Bluetooth LE module operates with a constant clock for data transmission which is not optimal for this application.

4.2.2 Improvements to reference voltage generation

As discussed in section 3.5, reference voltage used for differential difference amplifier is generated with the help of a CMOS charge pump. Although every effort was made to ensure accurate matching between P & N current sources, due to PVT (process) variations it's not possible to realize this in practice. This over time may introduce systematic bit errors into the system. Even though, these systematic errors can be eliminated with further signal processing on chip or in the receiver, this issue may be resolved with improving the charge pump accuracy.

4.2.3 More optimized digital logic

In activity dependent sampling rate adjustment design, all digital signal processing blocks including "amplitude reconstruction", "DC extraction" and "threshold detection" were custom made and simulated in spectre environment. An alternative approach would be to synthesize the circuit using RTL "register transfer level" flow, in this method, logic is described in the form of Verilog code and is directly synthesized to layout. Although both solutions are identical in terms

of functionality, RTL generated layouts are more optimized in terms of size and power consumption.

4.2.4 Further in vitro and in vivo experiments

Due to time constraints and unexpected events such as Covid 19 pandemic, activity dependent sampling rate adjustment design was not sent for fabrication and the results presented in chapter 3 are based on simulation data. Further in vitro test results may be used to render this design more robust and reliable. Any biomedical implanted device can benefit from in vivo validation. The effectiveness of these neural microsystem can be practically proven with cooperation of medical professionals. In vivo experiment which involves freely moving animal will also be beneficial for ambulatory validation.

4.2.5 Encapsulation

For bio compatibility reasons, prior to any in vivo tests, proper encapsulation and packing of circuitry is essential. The barrier layer should withstand erosion from naturally occurring bio chemicals and also to protect the body from toxins and electrical current.

It’s worth reemphasizing that the electrode array used in this design is completely organic and biocompatible and the remainder circuitry consist of standard materials which are used in existing clinical products, and therefore compatible with existing encapsulation techniques such as titanium casing and Parylene-C encapsulation.

Bibliography

[1] H. O. L. H. H. J. a. C. E. G. H. Klem, ""The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology," *Electroencephalography and Clinical neurophysiology,* vol. 52, pp. 3-6, 1999.

[2] C.-H. C. C.-S. H. Y.-H. C. a. L.-W. K. C.-T. Lin, "Real-time assessment of vigilance level using an innovative Mindo4 wireless EEG system," *IEEE International Symposium on Circuits and Systems (ISCAS '13),* pp. 1528-1531, 2013.

[3] A. B. M. B. e. a. G. Pillar, "An automatic ambulatory device for detection of AASM defined arousals from sleep," *Sleep Medicine,* vol. 4, no. 3, pp. 207-212, 2003.

[4] K. H. R. E. e. a. C. B. Juhl, "Automated detection of hypoglycemia-induced EEG changes recorded by subcutaneous electrodes in subjects with type 1 diabetes—the brain as a biosensor," *Diabetes Research and Clinical Practice,* vol. 88, no. 1, pp. 22-28, 2010.

[5] L. F. R. E. e. a. L. S. Snogdal, "Detection of hypoglycemia associated EEG changes during sleep in type 1".

[6] H. K. N. S. H. H. J. L. P. V. a. R. G. M. T. Salam, "Tradeoffs between wireless communication and computation in closed-loop implantable devices," *IEEE International Symposium on Circuits and Systems*, pp. 1838-1841, 2016.

[7] H. K. e. al, "All-wireless 64-channel 0.013mm2/ch closed-loop neurostimulator with rail-to-rail DC offset removal," *IEEE International Solid-State Circuits Conference*, vol. 60, pp. 452-453, 2017.

[8] I. Cognionics, "Quick-30 Dry EEG Headset," Cognionics, Inc., [Online]. Available: http://cognionics.com/index.php/54- roducts/headsets. [Accessed 01 05 2020].

[9] M. inc., "introducing muse 2," [Online]. Available: https://choosemuse.com/muse-2/. . [Accessed 12 07 2019].

[10] S. T. M. T. S. Hossein Kassiri, "Closed-Loop Neurostimulators: A Survey and A Seizure-Predicting Design Example for Intractable Epilepsy Treatment," *IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS,*, vol. 11, no. 5, pp. 1026 - 1038, 2017.

[11] "Link LABS," [Online]. Available: https://www.link-labs.com/blog/bluetooth-vs-bluetooth-low-energy.

[12] i. technologies, "Products overview," intan technologies, [Online]. Available: http://intantech.com/. [Accessed 05 04 2020].

[13] H. Aurlien, I. Gjerde, J. Aarseth and B. Karlsen, "EEG background activity described by a large computerized database," *Clinical Neurophysiology,* vol. 115, p. 665–73, 2004.

[14] K. Abdelhalim, "WIRELESS NEURAL RECORDING AND STIMULATION SOCS FOR MONITORING AND TREATMENT OF INTRACTABLE EPILEPSY," *Thesis,* 2013.

[15] B. Razavi, "A circuit for all seasons - The Strong ARM Latch," *IEEE SOLID-STATE CIRCUITS MAGAZINE,* pp. 12-17, 2015.

[16] C. S. Mestais et al., "WIMAGINE: Wireless 64-channel ECoG recording implant for long term clinical applications," *IEEE Trans. Neural Syst Rehabil Eng.*, vol. 23, no. 1, pp. 10 - 21, 2015.

[17] F. Goodarzy, E. Skafidas and S. Gambini, "Feasibility of energy-autonomous wireless micro-sensors for biomedical applications: Powering and communication," *IEEE Rev. Biomed. Eng,* vol. 8, p. 17029, 2015.

[18] A. P. Chandrakasan and P. P. Mercier, Eds., Ultra-Low-Power Short-Range Radios, New York, NY, USA: Springer-Verlag, 2015.

[19] P. P. Mercier, D. C. Daly and A. P. Chandrakasan, "An energy-efficient all-digital UWB transmitter employing dual capacitively-coupled pulse-shaping drivers," *IEEE J. Solid-State Circuits,* vol. 44, no. No. 6, pp. 1679-1688, 2009.

[20] H. Ando et al., "Wireless multichannel neural recording with a 128 Mbps UWB transmitter for implantable brain-machine interfaces," *IEEE Trans. Biomed. Circuits Syst,* 2015.

[21] S. Mandal and R. Sarpeshkar, "Power-efficient impedance-modulation wireless data links for biomedical implants," *IEEE Trans. Biomed. Circuits Syst.,* vol. 2, no. 4, pp. 301-315, 2008.

[22] H. -M. Lee and M. Ghovanloo, "An integrated power-efficient active rectifier with offset-controlled high speed comparators for inductively powered applications," *IEEE Trans. Circuits Syst. I Reg. Papers,* vol. 58, no. 8, pp. 1749-1760, 2011.

[23] S. Ha, C. Kim, J. Park, S. Joshi and G. Cauwenberghs, "Energy-recycling integrated 6.78-Mbps data 6.3-mW power telemetry over a single 13.56-MHz inductive link," *IEEE Symp. VLSI Circuits Dig. Tech. Papers,* pp. 66-67, 2014.

[24] L. P. Barnett MW,, "The action potential," *Practical Neurology,* vol. 7, pp. 192-197, 2007.

[25] A. Khanna, "Action Potential," Teach me physiology, 05 08 2018. [Online]. Available: https://teachmephysiology.com/nervous-system/synapses/action-potential/. [Accessed 28 04 2020].

[26] K. Michmizos, D. Sakas and K. Nikita, "Prediction of the timing and the rhythm of the parkinsonian subthalamic nucleus neural spikes using the local field

potentials," *IEEE Transactions on Information Technology in Biomedicine.,* vol. 16, pp. 190-197, 2012.

[27] A. Goldberger , L. Amaral , L. Glass , J. Hausdorff and P. Ivanov , " Components of a New Research Resource for Complex Physiologic Signals," PhysioBank, PhysioToolkit, and PhysioNe, 09 06 2010. [Online]. Available: https://physionet.org/content/chbmit/1.0.0/. [Accessed 28 04 2020].

[28] P. Rigas, D. A. Adamos and C. Sigalas, "Spontaneous Up states in vitro: a single-metric index of the functional maturation and regional differentiation of the cerebral cortex," *development of intrinsic brain activity,* 2015.

[29] H. Kim, S. Kim, N. Van Helleputte, A. Artes, M. Konijnenburg and J. Huisken, "A CONFIGURABLE AND LOW-POWER MIXED SIGNAL SoC," *IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS,* vol. 8, no. 2, pp. 247-267, 2014.

[30] L. Hans Dieter, "The Origins of the Sampling Theorem," *IEEE Communications Magazine,* vol. 37, no. 4, pp. 106-108, 1999.

[31] H. Kassiri, M. T. Salam, M. R. Pazhouhandeh, N. Soltani, J. L. Perez Velazquez, P. L. Carlen and R. Genov, "Rail-to-Rail-Input Dual-Radio 64-channel Closed-Loop Neurostimulator," *IEEE Journal of Solid-State Circuits,* vol. 52, no. 11, pp. 2793-2810, Oct. 2017.

[32] C. Kim, S. Joshi, H. Courellis, J. Wang, C. Miller and G. Cauwenberghs, "Sub-μVrms-noise sub-μW/channel ADC-direct neural recording with 200-mV/ms transient recovery through predictive digital autoranging," *IEEE J. Solid-State Circuits,* vol. 53, no. 11, p. 3101–3110, Nov. 2018.

[33] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag and A. Chandrakasan, "A micor-power EEG acquisition SOC with integrated feature extraction processor for a chronic seizure detection system," *IEEE J. Solid-State Circuits ,* vol. 45, no. 4, pp. 804-816, Apr. 2010.

[34] J. Yoo, L. Yan, D. E. Damak, M. A. B. Altaf, A. H. Shoeb and A. P. Chandrakasan, "An 8-channel scalable EEG acquisition SoC with patient-specific seizure classification and recording processor," *IEEE J. Solid-State Circuits,* vol. 48, no. 1, pp. 214-228, Jan 2013.

[35] T. Zhan, S. Guraya and H. Kassiri, "A resource-optimized VLSI architecture for patient-specific seizure detection using frontal-lobe EEG," *Proc. IEEE Int. Symp. Circuits Syst.,* pp. 1-4, 2019.

[36] N. Van Helleputte, "VLSIx 2016: Adaptive Sampling," imec,, 14 05 2019. [Online]. Available: https://www.youtube.com/watch?v=YPYIUHXlb5U. [Accessed 25 04 2020].

[37] C. Bedard, H. Kroger and A. Destexhe, "Does the 1/f frequency scaling of brain signals reflect self-organized critical states?," *Phys. Rev. Lett.,* vol. 97, no. 11, p. 118102, Sep. 2006.

[38] K. Miller et al., "Power-Law Scaling in the Brain Surface Electric Potential," *PLoS Comput. Biol.*, vol. 5, no. 12, Dec 2009.

[39] S. Park, Principles of sigma-delta modulation for analog-to-digital converters, 1999.

[40] Valdez, M. E., "Digital Filters," GRM Networks, [Online]. Available: http://www.grm.net/~mikevald/Digfilt.html. [Accessed 01 05 2020].

Appendix

5.1 MATLAB code for activity dependent sampling rate adjustment design:

```
clc
format long
% EEG comparator

over_sampling_factor = 100;



v_in = val(1:10000);
% v_in = val(1,:);
v_in_fft = fft (v_in);
v_in = v_in.' ;



% resampling the data
v_in = resampler (v_in , over_sampling_factor);



%[P,Q] = rat(over_sampling_factor);
%v_in = resample(v_in,P,Q);


v_ref = zeros (size(v_in));

comp_accu = zeros (size(v_in));


% Vector to store activity location
activity_detect = zeros (size(v_in));

activity_threshold_1 = 210 ;
activity_threshold_2 = 400 ;


% Cap size = 20 PicoF
```

```
cap_size = 1e-12 ;
%cap_size = 20e-12 ;

% Clock freq
clk_freq = 256 * over_sampling_factor ;
pulse_width = 1 / clk_freq ;

% charge pump's current
cap_current = 50e-9;
%cap_current = 1e-6;

size_of_data_vec = size(v_in);
size_of_data = size_of_data_vec (1,1);

%comparator's output
comp_out = 0;

sample_counter = 0 ;
DC_offset = 0 ;

index = 1 ;

over_sampling_status = over_sampling_factor ;

step_size = 100 ;

while index < size_of_data


   if (index > (step_size))
      v_ref (index) = v_ref (index-step_size) + find_cap_v (cap_size ,( cap_current*comp_out) , pulse_width) ;

   end


   if (v_in (index) > v_ref (index))
      comp_out = 1 ;
   else
      comp_out = -1 ;
   end
   if index > 1
      comp_accu (index) = comp_accu (index-1) + comp_out ;
   end

   % DC level estimation

   if (sample_counter >= (40 * over_sampling_factor))
      DC_offset = (DC_offset + v_ref (index))/2;
      sample_counter = 0 ;
   end
   sample_counter = sample_counter + 1;

   % activity detection
```

```
    for i = 1:1:step_size
        v_ref (index+i) =  v_ref (index) ;
    end

    % Clock freq
    clk_freq = 256 * over_sampling_factor / step_size ;
    pulse_width = 1 / clk_freq ;


    if ((abs(v_ref (index)) >= activity_threshold_1 )&&((abs(v_ref (index)) < activity_threshold_2)&&(
step_size ~= 10)))
        step_size = 10 ;
    end

    if ( (abs(v_ref (index)) >= activity_threshold_2 )&&( step_size ~= 1))
        step_size = 1 ;
    end



    if ((abs(v_ref (index)) < activity_threshold_1 )&&( step_size ~= 100))
        step_size = 100 ;
    end


    index = index + step_size ;


end

v_ref = v_ref (1:size_of_data);


disp ('DC offset is : ')
disp (DC_offset)

figure(1)
plot (v_in,'b')

figure(2)
plot (v_ref,'r')


figure(4)
plot (comp_accu,'r')

function v_temp = resampler ( vin , over_sampling_factor   )

    % resampling the data
    [P,Q] = rat(over_sampling_factor);
    v_temp =  resample(vin,P,Q);
```

```
end

function V_cap = find_cap_v (cap_size , cap_current , pulse_width)
   total_charge = cap_current * pulse_width ;
   %total_charge = integral(cap_current,0,pulse_width);
   V_cap = total_charge / cap_size;
End
```