UNIVERSITY OF CALIFORNIA

Los Angeles

Energy-efficient DSP Solutions for Simultaneous Neural Recording and Stimulation

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Electrical Engineering

by

Sina Basir-Kazeruni

2017

ABSTRACT OF THE DISSERTATION

Energy-efficient DSP Solutions for

Simultaneous Neural Recording and Stimulation

by

Sina Basir-Kazeruni
Doctor of Philosophy in Electrical Engineering
University of California, Los Angeles, 2017
Professor Dejan Marković, Chair

An increased interest in the investigation of the inner workings of the brain, together with recent technological advancements have been great catalysts for the development of neural stimulation and signal recording systems. These neural interfaces have enabled a better understanding of underlying neurological diseases, and provide promising therapeutic interventions for various neurological disorders.

As discoveries and technological advancements continue, new challenges and opportunities emerge. One of the major challenges is the development of small, portable, and power-efficient closed-loop neuromodulation systems. The ability to simultaneously stimulate and record is a key capability required in enabling such systems.

A closed-loop neuromodulation system is comprised of mainly four elements: (a) Stimulator: an energy-efficient and flexible stimulation engine, (b) Sensing: Low-power, high dynamic-range analog front-ends, (c) Digital Signal Processing (DSP): energy-/area-efficient digital signal processing units, and (d) Wireless transfer: an energy-efficient wireless power and data transfer

unit. In summary, efficient and concurrent stimulation, sensing, processing, and transfer of neural signals are required. Design efforts are in full effect to realize leading edge stimulation, sensing, and wireless transfer technologies; however, one common difficulty in realizing concurrent stimulation and recording of neural signals is the presence of stimulation artifacts observed at the sensing end. Existing solutions (e.g., blanking the recording channel during stimulation or self-cancelling stimulation electrodes) have not answered all the challenges and lack the ability of continuous signal recording during the stimulation phase, thus rendering a critical portion of the data unusable.

In this work we propose an energy-efficient, implantable, real-time Adaptive Stimulation Artifact Rejection (ASAR) engine, capable of adaptively removing stimulation artifact for varying stimulation characteristics at multiple sites. Additionally, a blind artifact template detection technique is introduced, which in combination with the proposed ASAR algorithm, eliminates the need for any prior knowledge of the temporal and structural characteristics of the stimulation pulse; this technique also enables us to effectively battle the non-linear mapping of brain tissue, and non-idealities of electrode interfaces, with linear filtering.

Two engines, implemented in 40nm CMOS, achieve convergence of <42μs for Spike ASAR and <167μs for LFP ASAR, and can attenuate artifacts up to $100mV_{p\text{-}p}$ by 49.2dB, without any prior knowledge of the stimulation pulse. The LFP and Spike ASAR designs occupy an area of $0.197mm^2$ and $0.209mm^2$, and consume 1.73μW and 3.02μW, respectively at 0.644V.

The LFP ASAR is integrated in a 64-channel sensing chip used in a state-of-the-art implantable, closed-loop neuromodulation unit (NM).

The dissertation of Sina Basir-Kazeruni is approved.

Danijela Cabrić

Gregory J. Pottie

Nanthia A. Suthana

Dejan Marković, Committee Chair

University of California, Los Angeles

2017

“Between stimulus and response there is a space.

In that space is our power to choose our response.

In our response lies our growth and our freedom.”

– Victor E. Frankl

To my sister

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation would not be possible without the support, guidance, and help of numerous individuals, to whom I will always be grateful.

I would like to take this opportunity to thank my advisor, and PhD committee chairman, Professor Dejan Marković. His continuous support and guidance has been a great catalyst, in research and teaching, during my time at UCLA; I cannot thank him enough.

I would like to thank my committee members. Professors Danijela Cabrić, Gregory Pottie, and Nanthia Suthana, have been invaluable sources of knowledge, and I will be forever grateful for their help and support.

I have significantly benefited throughout various collaborations during my PhD studies. I am grateful for the opportunity to collaborate with Professor Ali H. Sayed, and Adaptive Systems Laboratory during the development of the ASAR algorithm. Dr. Itzhak Fried and his lab provided many of the human patient data that were used in the design and verification of this work; these recordings proved to be vital. Additionally, I have had the privilege of connecting and gaining insights from numerous faculty at UCLA; I thank each and every one of them.

The staff of the Electrical Engineering department at UCLA provide the day-to-day support necessarily for students. I am deeply appreciative of all the assistance, support and friendship that I have received from Deeona Columbia, Kyle Jung, Jamie Khang, Il Hee Choi, Mandy Smith, Lori Miller, Ryo Arreola, and Julio Romero.

I consider myself lucky to have been able to be part of a very diverse group of exemplary researchers; I have heavily benefited, not only from their scientific and technological knowledge, but more importantly their friendships. Thank you Dejan Rozgić, Richard Dorrance, Neha Sinha, Hariprasad Chandrakumar, Wenlong Jiang, Vahagn Hokhikyan, Alireza Yousefi, Ippei Akita,

Zahra Aghajan, Michelle Tran, Nathan Pham, Yuta Toriyama, Fengbo Ren, Vaibhav Karkare, Uneeb Rathore, Sumeet Nagi, Uroš Topalović, Aryan Navabi, Farid Shirinfar, Hani Esmaeelzadeh, Arash Mirhaj, Abishek Manian, Ali Titiz, Emily Mankin, Stefan Vlaski, Hawraa Salami, Sarah Gibson, Zoltan Romocsa, Amal Katrib, Fang-Li Yuan, Tsung-Han Yu, Cheng C. Wang, Ahmed Alzuhair, Trevor Black and many others that made this journey more rewarding.

I am forever indebted to my friends. Thea Weiss, Sina Kalbassi, Behzad Amiri, Mahdi Ashktorab, Sourav Basu, Esmaeel Paryavi, Leon Tchikindas, Amir Mansouri, Kasra Vakilinia, Mania Aslan, Abhinav Sarna, Sina Eizadshenass, Rana Tehrani-Yekta, Kurt Groesser, Stephanie Delgado, Brenden Gomez, Hamed Afshari, Amin Rashidian, Kia Karbasi, and Hannah Ahders; without your friendship I could not have completed this chapter.

Last but not least, I want to take this opportunity to thank my family. Dad, who has always been supportive of me; Mom, who has been my closest friend in life; and my sister; I love you with all my heart.

VITA

2007	Electric Power Student, Honeywell Aerospace, Mississauga, Ontario, Canada
2008	IC Design Intern, Synopsys Inc., Mississauga, Ontario, Canada
2009	Hardware Intern, NVIDIA Corporation, Santa Clara, CA
2010	BASc, Electrical Engineering University of Waterloo
2011-2012	Teaching Assistant, Electrical Engineering, University of California, Los Angeles
2011-2012	UCLA Graduate Division Fellowship, University of California, Los Angeles
2012	Samueli Fellowship, Electrical Engineering University of California, Los Angeles
2012	2011-2012 Henry Samueli Excellence in Teaching Award
2012	MS, Electrical Engineering University of California, Los Angeles
2012-2013	Teaching Associate, Electrical Engineering, University of California, Los Angeles
2011-2012	UCLA Graduate Division Fellowship, University of California, Los Angeles
2013-2017	Teaching Fellow, Electrical Engineering, University of California, Los Angeles
2016	UCLA Dissertation Year Fellowship (DYF)
2016-2017	Teaching Assistant Consultant (Head TA), Electrical Engineering University of California, Los Angeles
2011-2017	Graduate Student Researcher, Electrical Engineering, University of California, Los Angeles
2017	2016-2017 Henry Samueli Excellence in Teaching Award

PUBLICATIONS

D. Rozgic, V. Hokhikyan, W. Jiang, **S. Basir-Kazeruni**, H. Chandrakumar, W. Leng, D. Marković, “A True Full-Duplex 32-Channel 0.135cm^3 Neural Interface,” *Proc. IEEE Biomedical Circuits and Systems Conference (BioCAS’17)*, Turin, October 2017.

Sina Basir-Kazeruni, Stefan Vlaski, Hawraa Salami, Ali H. Sayed, Dejan Marković, “A Blind Adaptive Stimulation Artifact Rejection (ASAR) Engine for Closed-Loop Implantable Neuromodulation Systems,” *Proc. 8th Intl. IEEE EMBS Conf. on Neural Engineering, (NER’17)*, pp. 186-189, Shanghai, China, May 2017.

R. Harjani, D. Cabric, D. Marković, B. Sadler, R. Palani, A. Saha, H. Shin, E. Rebeiz, **S. Basir-Kazeruni**, F.-L. Yuan, “Wideband Blind Signal Classification on a Battery Budget,” *Communications Magazine, IEEE*, vol.53, no.10, pp. 173-181, October 2015.

F.-L. Yuan, R.K. Palani, **S. Basir-Kazeruni**, H. Shih, A. Saha, R. Harjani, D. Marković, "A Throughput-Agnostic 11.9-13.6GOPS/mW Multi-Signal Classification SoC for Cognitive Radios in 40nm CMOS," *Proc. IEEE Symp. VLSI Circuits (VLSI’15)*, pp. C150-C151, Kyoto-Japan, June 2015.

Sina Basir-Kazeruni, Hao Yu, Fang Gong Yu Hu, Chunchen Liu, Lei He, "SPECO: Stochastic Perturbation Based Clock Tree Optimization Considering Temperature Uncertainty," *Integration, the VLSI Journal, (Elsevier)*, vol. 46, no. 1, pp. 22 - 32, January 2013.

Fang Gong, **Sina Basir-Kazeruni**, Hao Yu, Lei He, "Stochastic Behavioral Modeling Analysis of Analog/Mixed-Signal Circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD’12)*, vol. 32, no. 1, pp. 24-33, January 2013.

Fang Gong, **Sina Basir-Kazeruni**, Lara Dolecek, Lei He, "A Fast Estimation of SRAM Failure Rate Using Probability Collectives", *Proceedings of the 2012 ACM International Symposium on Physical Design (ISPD ’12)*, Napa, CA, March 2012.

CHAPTER 1

Introduction

1.1. Neuropsychiatric Disorders

Neuropsychiatric disorders are the leading category of diseases in the United States [1] (Figure 1.1). In fact, millions of patients worldwide battle various mental, behavioral, and neurological disorders, which include depression, Alzheimer's, Parkinson's, Obsessive Compulsive Disorder (OCD), epilepsy, paralysis, and many more.

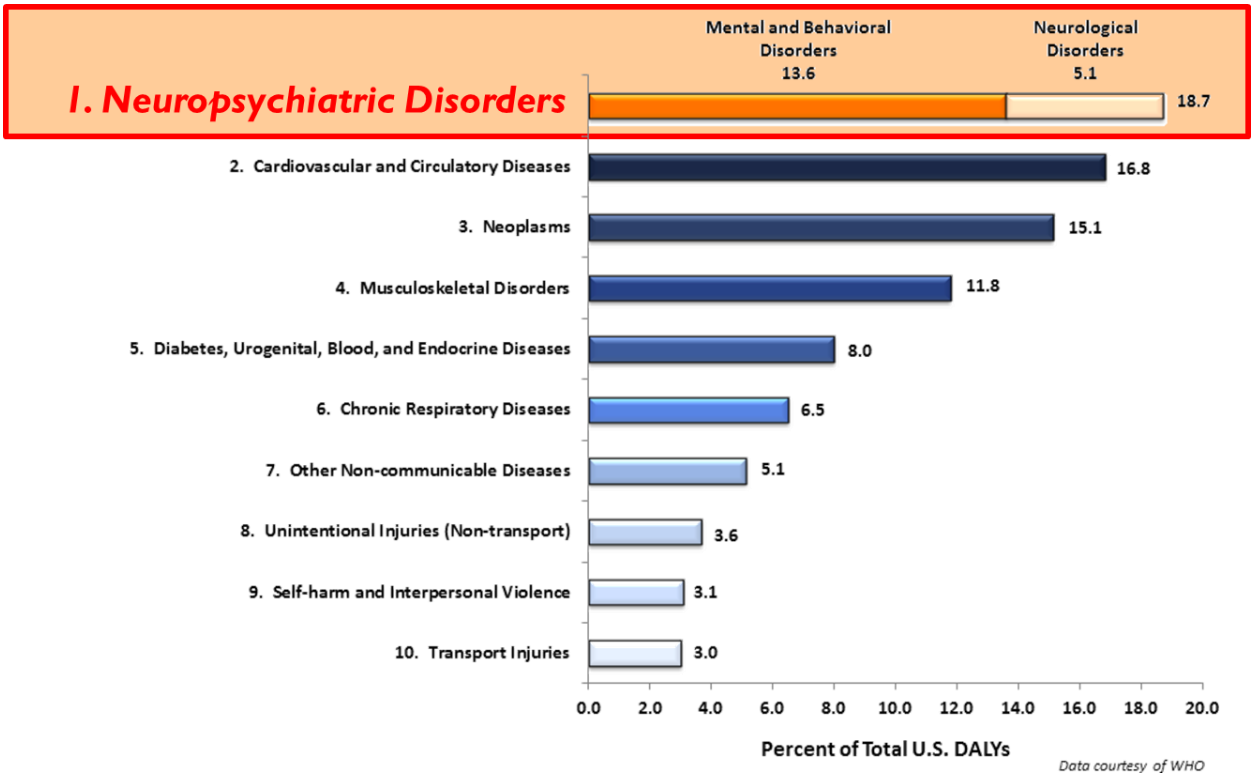

Figure 1.1: U.S. Leading Categories of Diseases/Disorders [1]

Traditionally therapeutic options include pharmacological and surgical approaches. While effective in many cases, these approaches are very limited, as they often only alleviate the worst effects of illness, are imprecise, and are not universally effective. In addition, patients can develop resistance to medications, and surgical solutions over time.

The limitations of traditional therapeutic options for neurological disorders, along with technological and scientific advancements, have been a catalyst for scientists and engineers alike to look for better alternatives. Various neural interface solutions have been developed as a result.

1.2. Neural Interfaces

Development of neural interfaces that help us study the brain, has been in progress in recent years. An ever-existing curiosity in the inner workings of the brain, together with technological advancements, have further intensified development of neural stimulation and signal recording systems. These systems, in return have helped greatly – mainly by looking at the electrical activity of neurons – in decoding of the human brain functions [2], [3]. As a result, a better understanding of underlying neurological diseases exists that can provide promising therapeutic interventions for various neurological disorders, such as epileptic seizures, and creation of brain-machine interfaces and for neuro-prosthetic technologies to aid paralyzed patients [4], [5].

1.2.1. Deep Brain Stimulation (DBS)

Deep Brian Stimulation, commonly referred to as DBS, was introduces deceases ago; Approximately 100,000 people around the globe live with a DBS implant [6]. In DBS systems, a neural stimulator is used to stimulate a particular region of the brain with the aim of providing therapy and relief for various neurological disorders.

Since their introduction in the late 1980s, systems similar to the Medtronic Activa DBS system (shown in Figure 1.2), have proven immensely helpful in providing therapy for patients with some specific neurological disorders. For disorders like Parkinson’s, and Dystonia, DBS systems have proven to be an alternative treatment method, working where traditional approaches failed.

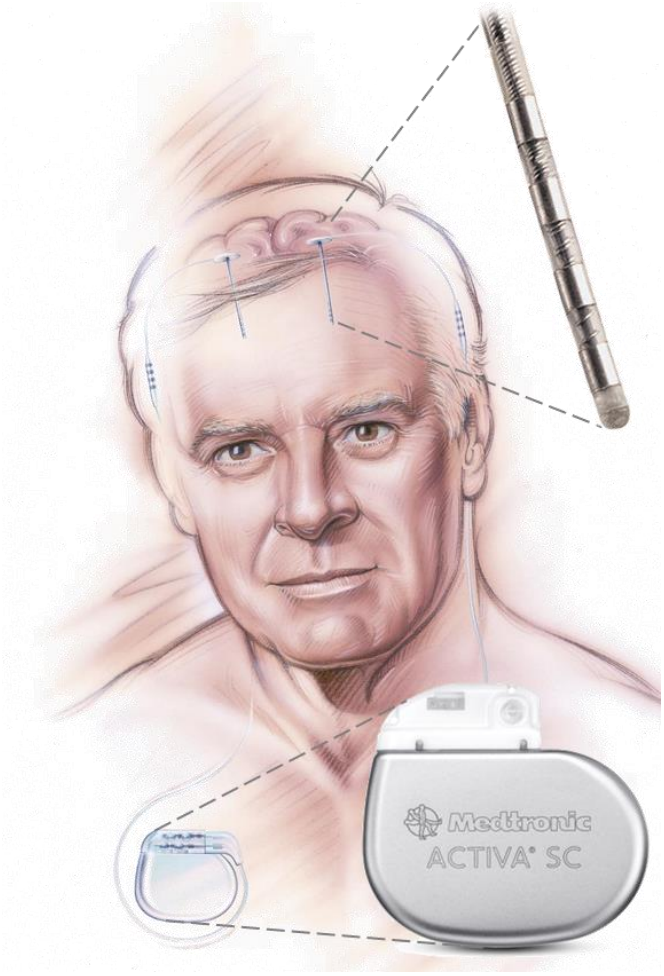

Figure 1.2: Deep Brain Stimulation (DBS) System. (Medtronic Activa SC System)

Although successful, DBS is not without shortcomings:

- The probes used on these systems only provide a few (~4-8), large stimulation sites. This results in low resolution, and low specificity.
- These systems provide continuous stimulus to patients, and over time the physiological response can be diminished, resulting in habituation.
- It may take weeks to observe the effectiveness of stimulation parameters, for some conditions such as dystonia.

- Continuous stimulation over long periods of time can lead to various undesirable side-effects (e.g., depression, behavioral disorders, etc.).

Additionally, DBS' effectiveness has only been shown in Parkinson's and other movement disorders where its open-loop stimulation can target anatomically focal regions. It is, however, insufficient for all other neurological disorders; to provide therapy for most neurological disorders, precise localization, over a distributed neural network is required.

Therefore, new solutions are required that can provide closed-loop neural stimulation and recording.

1.2.2. Neuromodulation Devices

More recently, neural recording and stimulation systems have been able to provide treatments for epilepsy and spinal cord injuries through closed-loop control of neural stimulation [7]. It is safe to say that, engineers and scientists, are trying to "close the loop" with the brain, and create much more capable systems, that can help with study and treatment of many disorders.

Human memory is an excellent example: The medial temporal structures, including the hippocampus and the entorhinal cortex, are critical for the ability to transform daily experience into lasting memories; and it has been shown that stimulation of the entorhinal region, enhances memory of spatial information when applied during learning [8].

There is abundance of therapeutic opportunities, current neuromodulation devices (e.g. NeuroPace RNS-300 system shown in Figure 1.3), however, are lagging behind. These systems do not provide the required resolution, and are very bulky requiring large components to be surgically implanted in the wires. As a result, these devices are hard to implant, not comfortable

for the patients, and do not provide high resolution recording and stimulation over various regions of the brain.

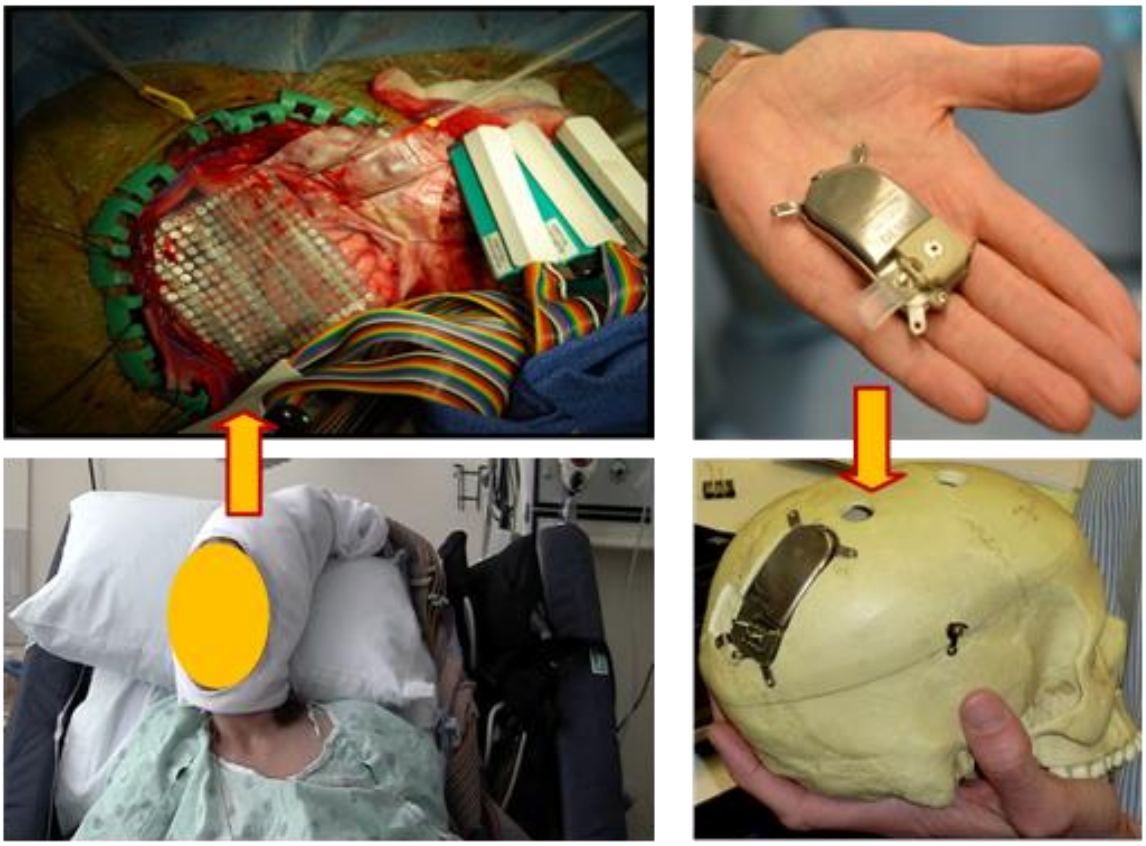

Figure 1.3: Current Neuromodulation (NM) Devices. (NeuroPace RNS-300 System) [9]

Most importantly, all neuromodulation devices released to date, lack a critical component to enable a true closed-loop solution: ability to record neural activity in the presence of stimulation artifacts [10], or simultaneous recording and stimulation.

For a system with full promise of therapy for multiple neurological disorders, that patients can take advantage of comfortably in their daily lives, we need a wireless, implantable, closed-loop neuromodulation system that enables simultaneous stimulation and recording.

1.2.3. Closed-loop Neural Recoding and Stimulation

A closed-loop neuromodulation system is comprised of mainly four elements: (a) Stimulator: an energy-efficient and flexible stimulation engine, (b) Sensing: Low-power, high dynamic-range analog front-ends, (c) Digital Signal Processing (DSP): energy-/area-efficient digital signal processing and feature extraction units, and (d) Wireless transfer: an energy-efficient wireless power and data transfer unit. In summary, efficient and concurrent stimulation, sensing, processing, and transfer of neural signals are required.

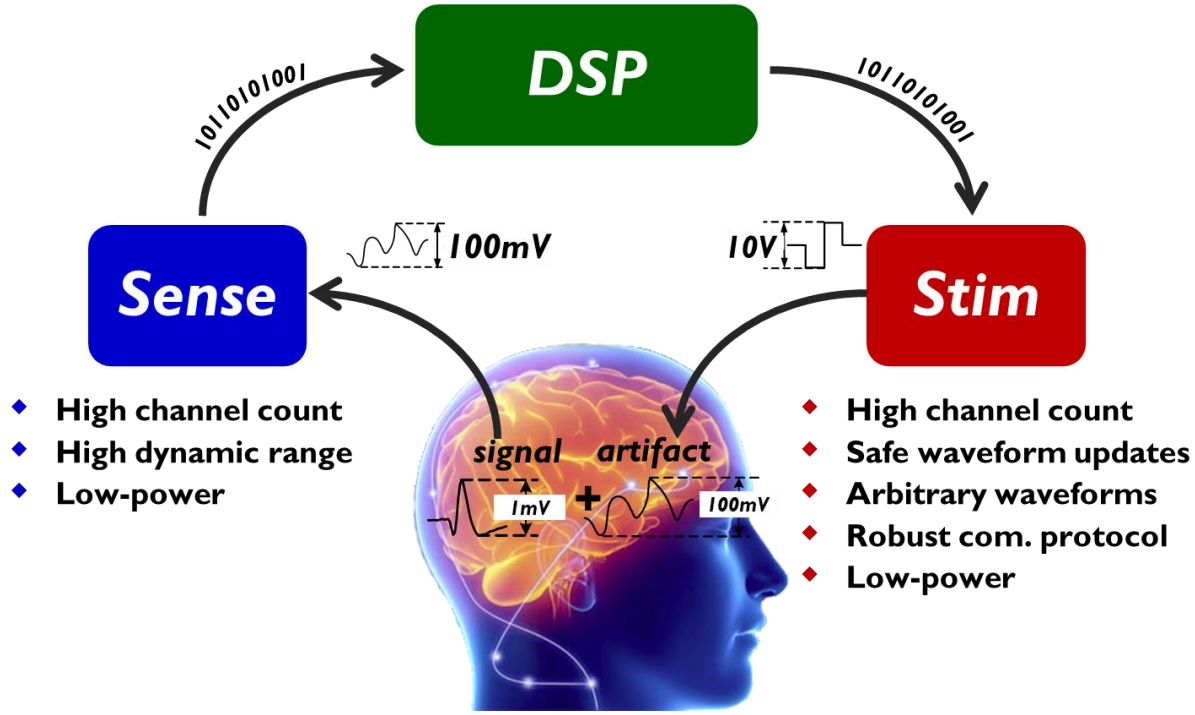

Figure 1.4: Closed-loop Neuromodulation System

One common difficulty in realizing concurrent stimulation and recording of neural signals is the presence of stimulation artifacts (<100mV) observed at the sensing end alongside neural signals of interest (<1mV) (see Figure 1.4). Existing solutions have not answered all the challenges

and lack the ability of continuous signal recording during the stimulation phase, thus rendering a critical portion of the data unusable.

Therefore, a DSP solution that can adaptively remove the stimulation artifacts in real-time is an essential component of any modern neuromodulation system with aspirations for simultaneous neural recording and stimulation.

1.3. Dissertation Outline

The remainder of this dissertation is organized in the following manner:

- **Chapter 2: Background**, provides background information on signal of interest in the neuromodulation system and design of stimulation artifact solution. The importance of stimulation artifact rejection is summarized, and key challenges and requirements for a successful implantable, real-time stimulation artifact rejection solution is reviewed. At the end, a review of prior art is provided.
- **Chapter 3: Proposed Adaptive Stimulation Artifact Rejection (ASAR) Algorithm**, provides the motivation and mathematical foundation for the proposed Adaptive Stimulation Artifact Rejection (ASR) algorithm. In this chapter, algorithmic decisions rising from the need for better performance are discussed. Template detection method is introduces, and the algorithm is formulated and organized in two phases. Simulation results verifying the operation and validity of ASAR are presented.
- **Chapter 4: ASAR Hardware Implementation**, explains the hardware implementation considerations and design choices in different design blocks. Hardware implementation for both phases of ASAR operation, along with their corresponding hardware timing diagram illustrations are provided.

- **Chapter 5: IC Implementation and Measurement Results**, presents the implementation of two ASAR design (LFP ASAR and Spike ASAR) in 40nm CMOS technology. Furthermore, detailed chip measurement results are included in this chapter. A new experiment devised to measure the true noise (artifact) suppression of the ASAR filter is introduced and its results are shared. Lastly, ASAR is compared with all other stimulation artifact methods, and a qualitative comparison with state-of-the-art adaptive filtering method in this field is included.
- **Chapter 6: ASAR Integration: Sensing IC and Neuromodulation Unit (NM)**, delivers an overview of ASAR integration within our 64-channel neural sensing chip, its other building blocks, and real-time measurements. Miniaturized neuromodulation unit (NM), combining this sensing solution with a stimulator IC for a fully implantable, closed-loop neuromodulation solution, is shown.
- **Chapter 7: Conclusion**, concludes this thesis. Research contribution and future work are discussed.

CHAPTER 2

Background

2.1. Signals of Interest

Before reviewing design requirements, challenges, and limitations of prior-art, we need to introduce the signals of interest in the intended closed-loop neuromodulation system. Most invasive neuromodulation systems and the physicians that use these devices require and work with one of the following signals in the human brain:

- Local Field Potentials (LFP)
- Active Potentials (AP, or commonly referred to as Spikes)

To provide a complete solution, we consider both of these signals. Figure 2.1 shows various types of biological signals, along with their bandwidth and amplitudes [11], [12], [13].

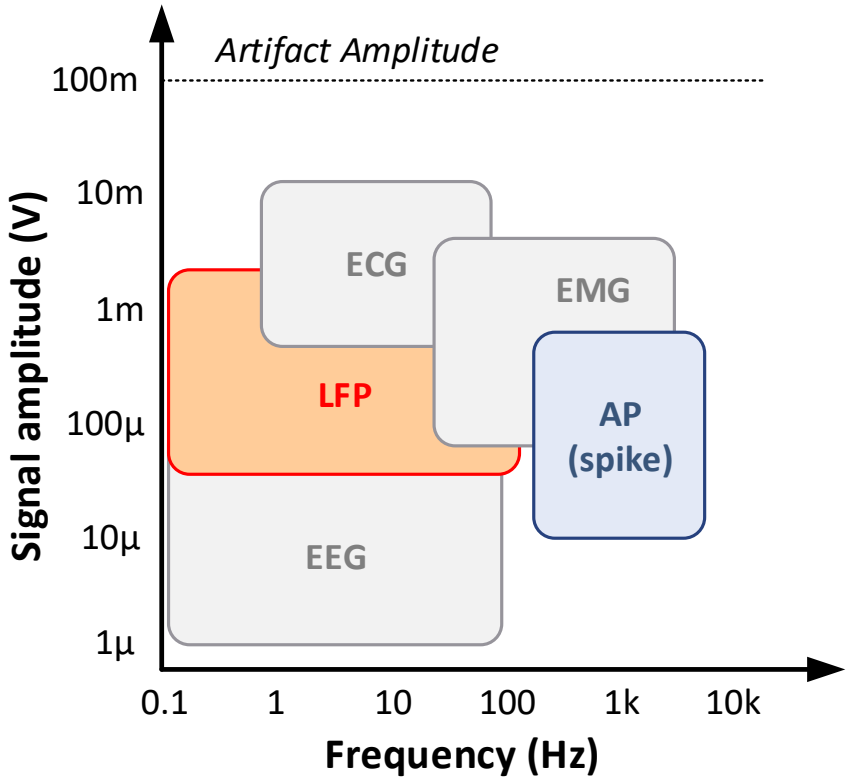

Figure 2.1: Various Biological Signals [13]

An immediate observation from Figure 2.1, is the overlap of the frequencies of the signals of interest and the stimulation artifacts; this has traditionally been a major source of difficulty in removal of stimulation artifacts. We can also observe that the stimulation artifacts possess a much higher amplitude when compared to the neural signals of interests; this is another obstacle that will be discussed in more detail.

Table 2.1 shows specific bandwidth and amplitudes for the Local Field Potential (LFP), and Action Potential (AP). These parameters will be considered as specifications for our designs.

Table 2.1: Neural Signals of Interest

Neural Signal	**Bandwidth**	**Peak Amplitude**
Local Field Potential (LFP)	1Hz – 200Hz	~1 mV
Action Potential (AP)	200 Hz – 5 KHz	~100 µV

2.2. Importance of Stimulation Artifact Rejection

Closed-loop stimulation is desired to build an effective neuromodulation system. This operation requires concurrent stimulation and recording of neural signals (<1mV) in the presence of stimulation artifacts (10s of mV).

Traditionally the neural recording front-ends could not provide a high enough dynamic range to avoid saturation in the presence of stimulation artifacts. As a result, conventional neural recording systems only view stimulation artifact rejection primarily as a method to mitigate saturation in neural recording front-ends. Conventional recording systems, therefore, do not allow for recording of signals of interest during stimulation; making the recovery of signal in these regions impossible.

The problem exists today: various neural interfaces introduced in the last few years (e.g., [14], [15], and [16]) still suffer from this problem and cannot offer neural recording under stimulation artifact. The system shown in [17], also has limited input signal range (~2-3mV) with no means to avoid saturation

Recently, front-ends capable of recording neural signals with much higher dynamic range have been designed. For example the VCO front end in [18] and [19] are capable of recording Local Field Potentials (LFPs) with up to ±50mV linear-input-range, while the neural recording front-ends in [20], [21], [22], and [23] promises recording of neural signals (both Local Field Potentials and Action Potentials) with ±20mV up to ±40mV linear-input-range. Most recently, [24] introduces an implantable neural recording front-end solution with ±100mV linear-input-range. These designs overcome the saturation problem in conventional systems, however, large stimulation artifacts remain in the digitized signal.

In order to recover the neural signal, an effective stimulation artifact rejection method is needed to enable concurrent stimulation and recording while achieving the following:

- Maximizing usable neural data
- Minimizing time delays for closed-loop response
- Helping reduce the design constraint of various other blocks in the system
    - After the removal of stimulation artifacts, the sampling rate and number of bit can be reduced, without any loss in data, resulting in more relaxed design constraints in following signal processing blocks and wireless transfer.

We propose that these new front-end designs are used followed by an Adaptive Stimulation Artifact Rejection (ASAR) for concurrent stimulation and recording, and more importantly, enabling investigation of the instantaneous neural response to stimulation.

Figure 2.2 illustrates observation of both stimulation artifact (a) and desired neural signal (s) as the recorded signal (d) at the input of the recording front-ends, as well as the difference between the conventional neural recording systems and one that was proposed.

ASAR, as a result, becomes a very important building block of the proposed recording solution, and subsequently a complete closed-loop neuromodulation system.

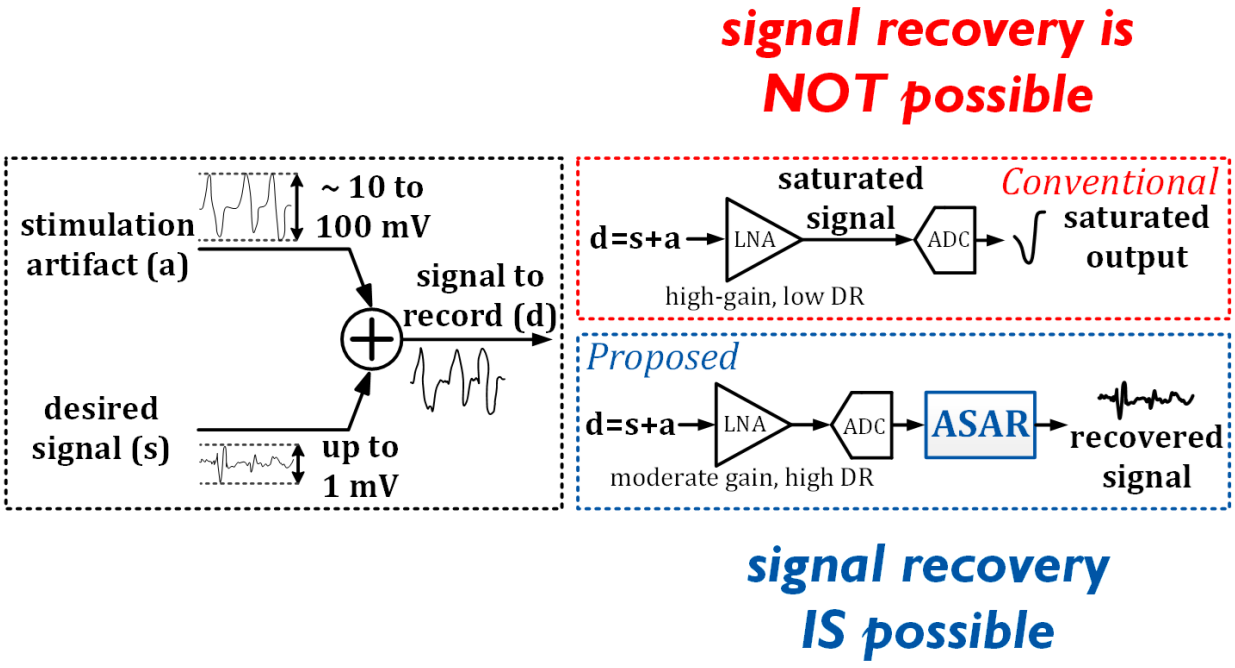

Figure 2.2: Neural Recording Systems - Conventional vs. Proposed

2.3. Key Design Challenges and Requirements

In this section key design challenges for adaptive stimulation artifact rejection is briefly discussed and these requirements are used to assess prior arts, as well as our solution and how effective they are for a closed-loop, implantable, neuromodulation system.

2.3.1. Concurrent recording and stimulation

Any sufficient stimulation artifact rejection solution, would require the ability to enable recording during stimulation. This may seem obvious, but as discussed earlier, some traditional solutions will not meet this requirement, as they only view artifact rejection as a method to mitigate the saturation problems in the recording front-end.

2.3.2. Robustness

It is imperative that a successful solution for stimulation artifact rejection is able to function under varying conditions.

Varying stimulation pulse characteristics – both structural and temporal – present the greatest need for robustness; these variations could occur multiple times within a short period of time by design (e.g., algorithm choice, physician, etc.), or as a result of uncontrolled variables (e.g., non-linear brain tissue mapping, electrode movement, etc.).

2.3.2.1. Timing Assumption

Assuming a fixed, or fixed-range, of timing delay between stimulation and when the stimulation artifact is observed on the recording side, can lead to failure.

Various factors that can contribute to changes in timing parameters are:

- Varying stimulation pulse characteristics – structural and temporal.
- Unknown brain tissue mapping, that changes from person-to-person, and over time.
- Non-ideal electrode interfaces.
- Varying recording and stimulation sites.

A solution that is agnostic to these timing variations and makes minimal timing assumptions would be desirable.

2.3.3. Performance

The difficulties with measuring the true attenuation of stimulation artifacts will be discussed in more detail in the following chapters, however, some basic performance metrics need to be met for any acceptable solution.

2.3.3.1. Size

Any solution will require to be an implant scale; while this is not strictly defined, it does exclude all complex algorithmic solutions implemented on PC, FPGA or other development board.

2.3.3.2. Convergence

A closed-loop neuromodulation system would require a real-time rejection of stimulation artifacts, as the decision to provide therapeutic stimulation is made in real-time, based on the available neural recording across the various regions of the brain. This requires real-time convergence of the solution and excludes solution that need repeated offline tuning/training.

2.3.3.3. Power

Any solution for an implantable closed-loop neuromodulation device, requires to be extremely low power, as the maximum power density of implantable devices are limited before they start to damage neurons [11].

Wireless transmission dominates the power of traditional neural implants, however, by introduction of various DSP blocks, this power can be significantly reduced by transmission of signal features at much lower rates than the raw signal (DSP helps in removing stimulation artifact, compressing raw data, and feature extraction, resulting in significant data-rate compression [25]). In implantable neural devices that employ batteries, low power operation, is even more essential in order to eliminate the need for frequent battery replacements. This has led into developments of alternative technologies, such as inductive power transfer or thermal energy harvesting [26] [27].

Figure 2.3 shows and compares estimated system powers for various options for action potentials. This figure also shows, that in our approach, all the DSP components need to consume less than ~6μW per channel to enable realization of a fully implantable and wirelessly powered neuromodulation system.

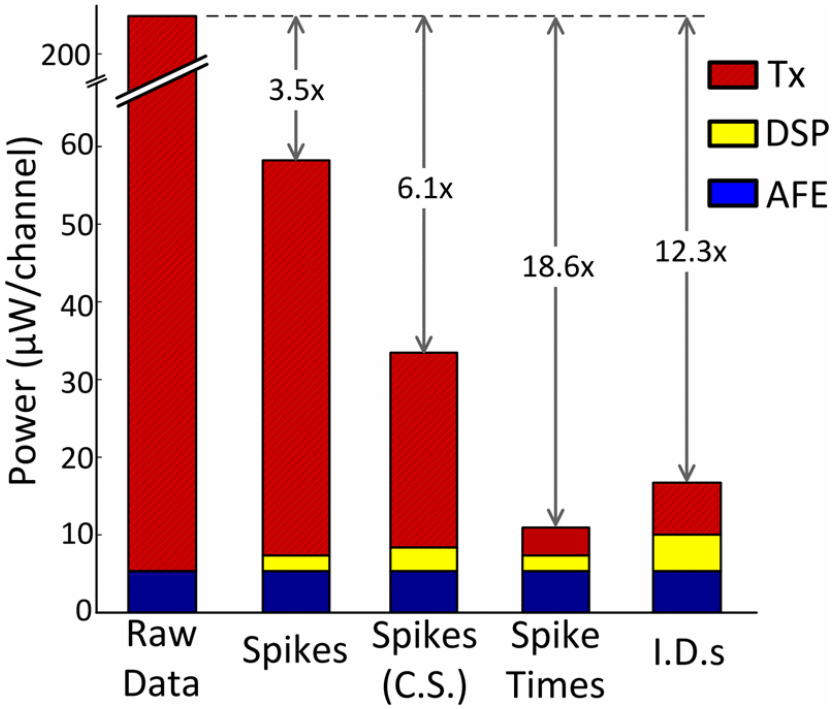

Figure 2.3: Estimated System Power for Various Output Options for Spike Recordings [28]

The focus of this work is the development of an energy-efficient ASAR algorithm to adaptively remove stimulation artifact, however, other necessary DSP blocks, such as digital filters to separate LFP and neural Spikes have also been included (See Appendix A). In addition, for various neurological disorders, and their therapeutic needs, other signal processing blocks may need to be considered in the future. Therefore, in order to meet our power requirements and provide some margin for additional feature extraction and processing blocks, we have constrained ourselves to ~3µW per channel for stimulation artifact rejection. Any solution requiring more than 6µW per channel will definitely not be suitable for implantable neuromodulation systems.

2.4. Review of Prior Art

Several previous works have focused on the problems associated with the presence of Stimulation Artifacts in neural recording systems. Some of the key ideas introduced by these papers are briefly summarized below. In reviewing the prior art, we will try to evaluate them based on the design requirements and challenges that were introduced in Section 2.3.

2.4.1. Blanking the recording channel during stimulation

Some previous works have tried to mitigate the stimulation artifact issue by "blanking" the recording channel during or immediately after stimulation. These methods, do this, in order to reduce the burden on the analog front-ends that cannot support very high dynamic ranges necessary to capture the neural signal alongside stimulation artifact. For example, in [29], an overload recovery technique is employed (Figure 2.4).

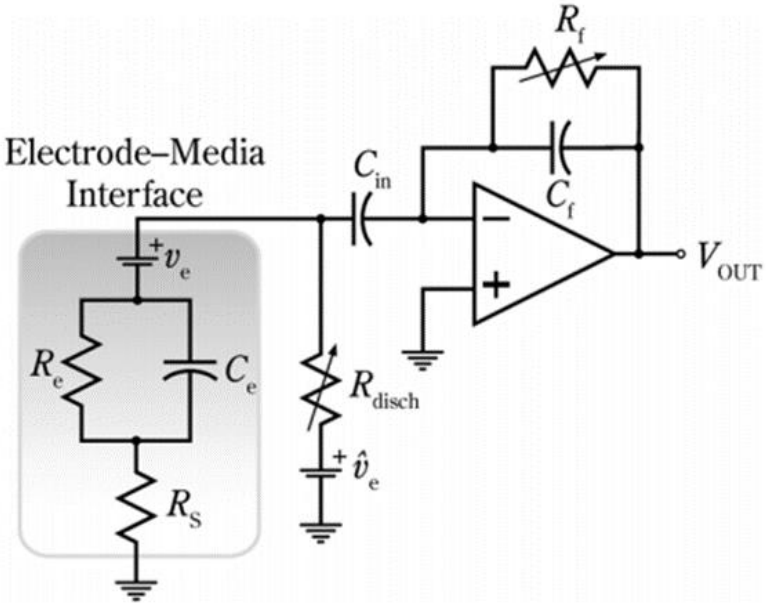

Figure 2.4: Overload Recovery [29]

Here a shunt resistance is used to de-polarize the electrode, immediately after stimulation ends. In this example and most other cases a quick recovery from saturation is achieved, however, it is clear that capability to record during stimulation onset is not provided and a critical portion of neural recording is rendered unusable.

Blanking the recording channel during stimulation is a very common method used in various sensing solution in neuromodulation systems. Some more recent examples are were blanking is utilized are [30], and [31].

2.4.2. Polynomial Curve Fit

A local curve fitting approach is shown in [32], in order to try and suppress the stimulus artifact. This algorithm, called SALPA, also blanks the output of the recording during saturated regions during stimulation (see Figure 2.5), and require PC hardware. Although its performance is significantly better than low-pass Butterworth (BW-L), high-pass Butterworth (BW-H), and linear phase filters, it is still not satisfactory. Furthermore, the limited degree of freedom to model the artifact can hinder its applicability.

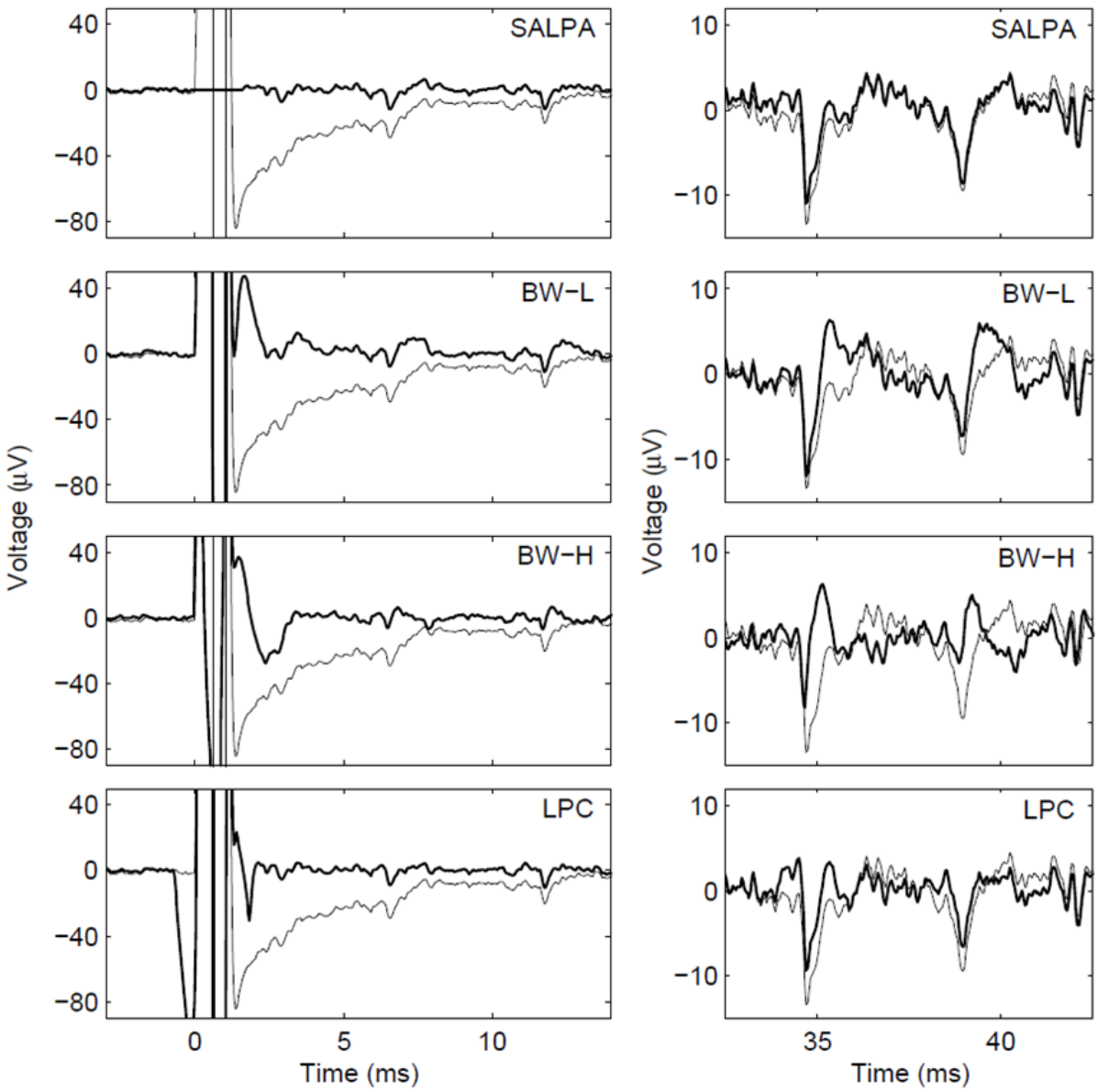

Figure 2.5: Comparison of Various Curve Fitting Methods [32]

Most importantly, this solution is an offline algorithm. Meaning that a bulk of data needs to be acquired before any portion of it can be cleaned. This constraint renders such systems inapplicable to a majority of applications where (a) memory is limited, (b) computational power is limited or (c) the cleaned data needs to be available in real-time.

2.4.3. Artifact Cancellation

Another interesting method was introduced in [33] were artifact templates, are created and updated regularly and subsequently canceled in the recording using a feedback DAC. Figure 2.6 shows a top-level summary of this method.

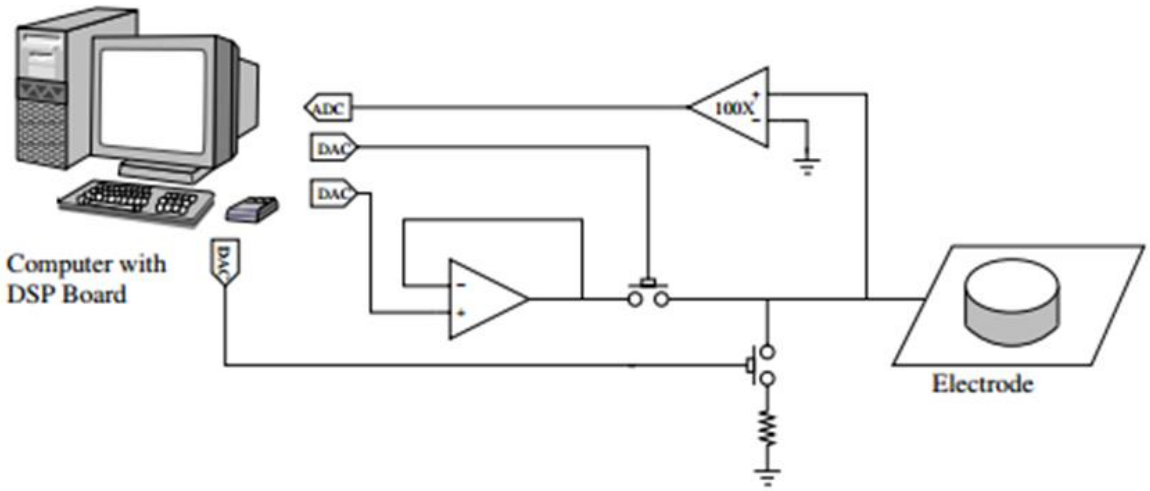

Figure 2.6: Artifact Cancellation [33]

Although intriguing, few issues remain: (a) need for an offline training/update of the stimulation artifact, and (b) cancelling using a feedback DAC on the sensing side of the recording system significantly increases noise at the input of analog front-end and cause major system level complexities. Lastly, in this work only biphasic stimulations were considered.

2.4.4. Self-Cancelling Stimulation Electrode Configuration

Self-cancelling stimulation electrode configurations have been previously proposed. In [34], recording electrodes are placed differentially around stimulation electrodes (Figure 2.7), and common-mode passive filters are utilized to attenuate the stimulation artifact.

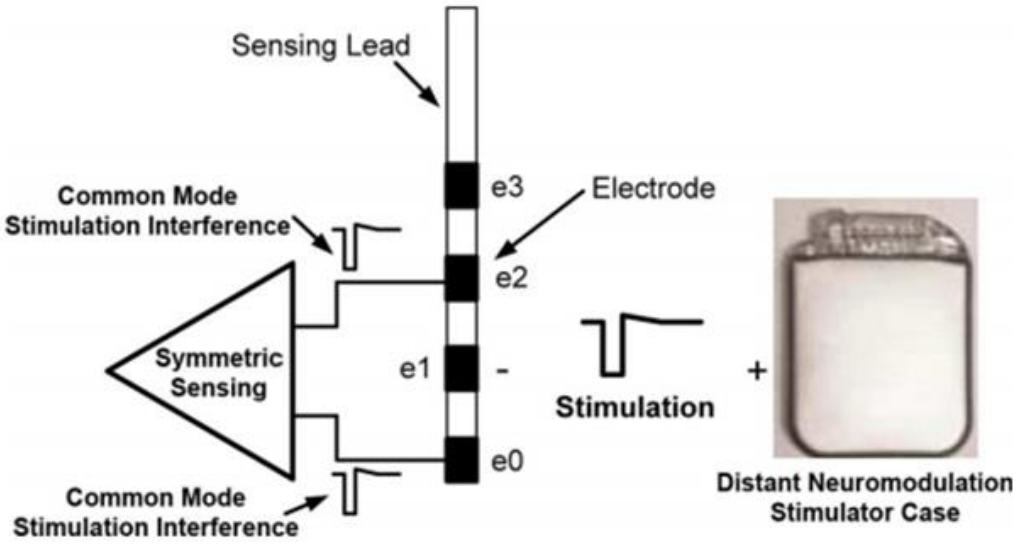

Figure 2.7: Self-cancelling Stimulation Electrode Configuration [34]

Clearly, this solution is very application-specific, and unless it is produced with a set configuration in mind, cannot be effectively evaluated. Furthermore, electrode configuration requirements can limit the flexibility that is offered to the user, as far as stimulation and recording sites are concerned.

Another noteworthy observation in this work is the required off-chip components which are undesirable. Additionally, in this work the artifacts reside in frequency ranges that did not overlap the signals of interest, which is not the case in many practical situations.

2.4.5. Echo-cancellation

Echo-cancellation based methods to remove stimulation artifact in neural systems have been suggested in [35] [36], and have shown great promise in concept, however their performance is not nearly sufficient to be implemented in an implantable, real-time, closed-loop neuromodulation system. Most recently, the method in [36] (shown in Figure 2.8), as well as the method in [10],

claimed attenuation of 8-sample long artifacts by up to 42dB. This claim is very misleading, as it only relies on the insertion of an 8-bit DAC at the input of the front-end, and is not a real measure of filter attenuation. The best attenuation in [10] [36], seems to be 24dB from their results and measurements; that measurement also relies on max amplitudes of the neural signals and stimulation artifact amplitudes, which we believe is a not a great way to measure this filter performance. More detailed discussion about this is included in Section 5.3.

Performing the cancellation of estimated artifact at the input of the amplifier, using a DAC as shown in Figure 2.8, introduces the same input noise issues that were previously discussed in Section 2.4.3.

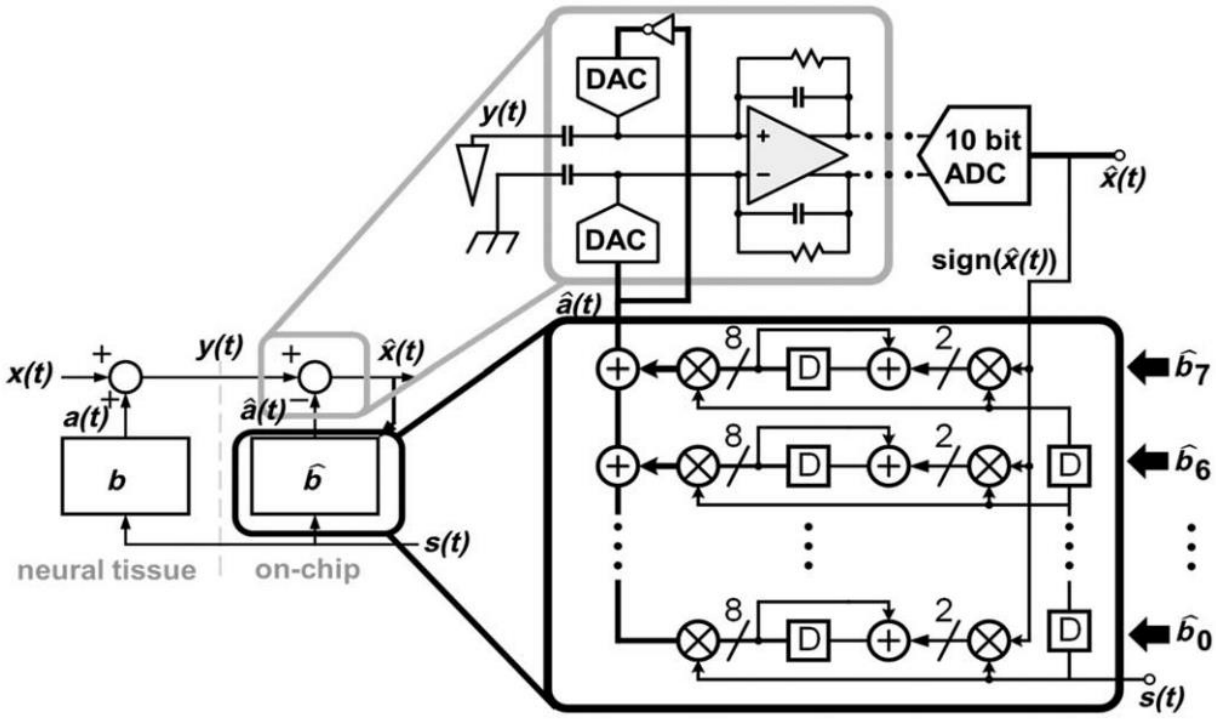

Figure 2.8: Echo-cancellation Based Stimulation Artifact Cancellation [10]

Additionally, this method, demonstrated long convergence times (~3 seconds in [36] and up to ~12 seconds in some of the measurements in [10]), which is not practical in real life situations, where stimulation/recording sites along with stimulation pulse's structural and temporal characteristics can change at any moment.

The feedback nature of artifact subtraction dictates a maximum delay between stimulation onset and when it is seen at the sensing end. If the delay is any longer, the method could fail, unless stimulation is periodic. Lastly, in all the results shown in [10] and [36], since u(t) is a single bit parameter, only monophasic stimulation patterns are supported. This is contrary to their claim of supporting monophasic stimulation patterns.

While the echo-cancellation is a promising method, major modifications are required to improve its attenuation, convergence, and overall performance before it can be used in a closed-loop neuromodulation device.

2.4.6. Other Methods

Other methods that are not discussed in as much detail include the following:

- Authors in [37], present a phase-space model approach, in which they model a periodic artifact as a noisy oscillator. They then approximate the periodic components by fitting the phase space model of the oscillator. Although this is a thought-provoking approach, it significantly suffers from its limited degree of freedom expected to model artifacts only.
- The system presented in [38], employs a (rapid) overload recovery in their front-end. As we previously discussed in Section 2.4.1, this is an insufficient solution and is usually done in order to alleviate the front-end saturation. In the same work, in their closed-loop experiment, which in addition to their implantable IC, includes an FPGA and a PC, a multi-

channel PCA denoising algorithm is implemented to remove artifacts. For PCA's use in this application to be valid, an assumption needs to be made that the stimulation artifact is purely noise and the neural signal recorded during stimulation behaves similar to the segment of the data before and after stimulation. In other words, this solution ignores any neural response that can be caused by stimulation. Validity of this assumption is not clear; however, even if valid the multi-channel PCA algorithm is not implemented on chip and therefore is not implantable.

- A similar artifact removal strategy is adopted in [39] with a slightly different algorithm (this implementation is on an ARM Cortex-M3 processor, which is part of their SoC FPGA, and is not in the implantable section of their system). Averaged artifacts are used to generate artifact flags. Upon arrival of data to the SoC FPGA, artifacts are detected based on the artifact flags, and a linear interpolation is used to clean those artifacts. Use of linear interpolation in this case, effectively is a very intricate "blanking" solution, where the recording during stimulation is not zero, but interpolated using the before and after data points. Results look visually acceptable in time-domain recordings; however, there is no certainty that those recording are valid responses (and hence recording) of neurons during stimulation.

CHAPTER 3

Proposed Adaptive Stimulation Artifact Rejection (ASAR) Algorithm

Our proposed algorithm embeds the findings from echo-cancellation and other methods mentioned in section 2.4.5, into a comprehensive adaptive filtering framework and aims to provide a complete Adaptive Stimulation Artifact Rejection solution for modern low-power, closed-loop neuromodulation systems. The implementation of the energy-efficient ASAR algorithm aims to clean neural recordings in the presence of stimulation artifact, by utilizing adaptive filtering techniques, as motivated in Figure 3.1.

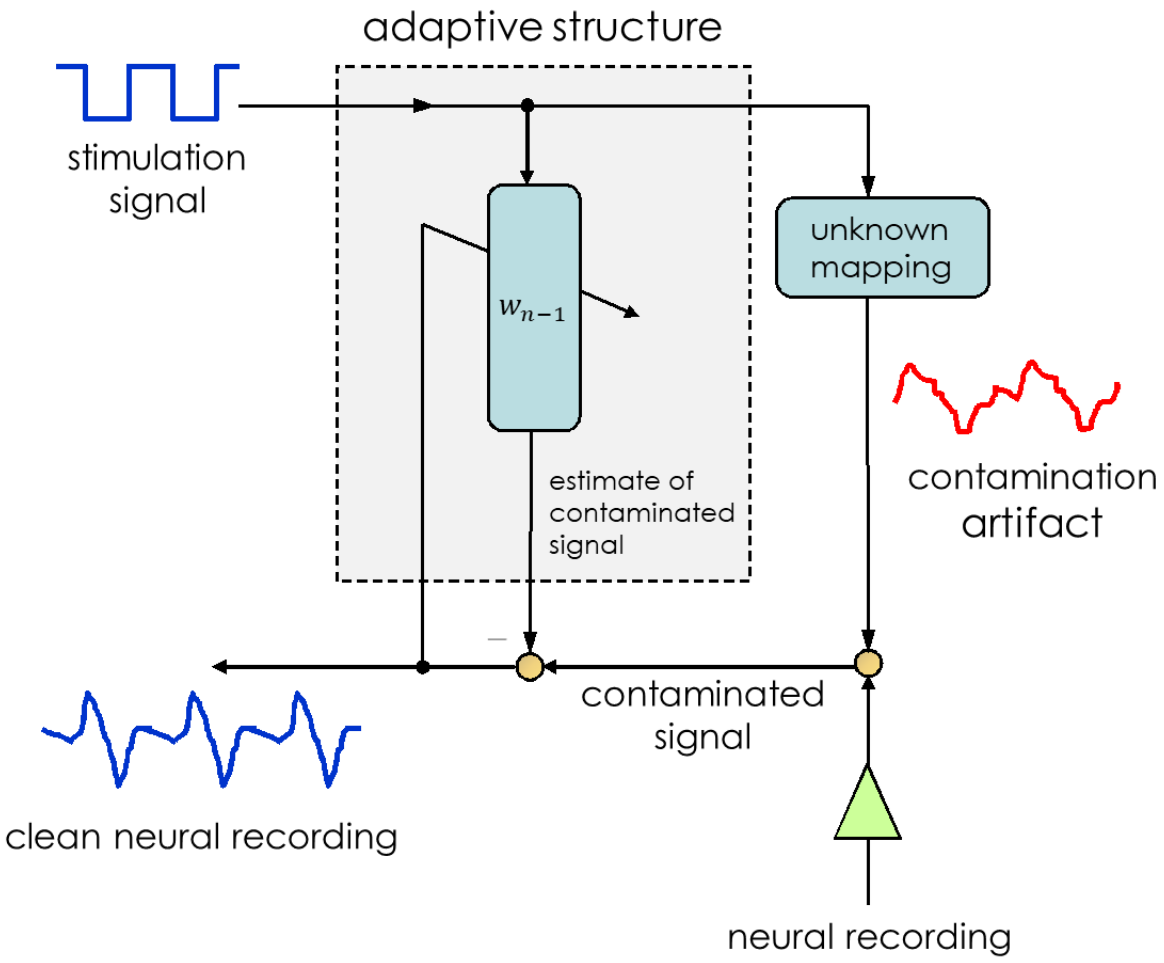

Figure 3.1: ASAR Motivation

3.1. ASAR Algorithm

We are considering a scenario where K measurement sites are distributed spatially. At each particular electrode k and time instance i, the artifact is modeled as

$$a_k = u_{k,i} w_k \tag{1}$$

where $u_{k,i}$ and w_k are vectors of size M. Note that this representation assumes that the artifact is generated as a linear transformation from $u_{k,l}$ through w_k. The most straightforward approach would be to populate $u_{k,l}$ from the stimulation pattern, which is commonly done in methods similar to [36]. This has several major drawbacks, namely that (a) the stimulation needs to be known, (b) the algorithm can only cancel linear transformations of the stimulation pattern, which is not sufficient in practice and (c) the method is highly susceptible to errors stemming from non-idealities, asynchrony and misalignment.

One of the main contributions of our proposed method is that we will show how to obtain $u_{k,l}$ directly from measured data and use this information to cancel artifacts in real-time without any prior knowledge about the nature of the stimulation. Furthermore, the proposed algorithm absorbs the non-linearity correction implicitly into the computationally inexpensive generation of $u_{k,l}$, effectively allowing us to perform non-linear artifact rejection at approximately the cost of a linear adaptive filter (more detail on this is provided in Section 3.1.1.3). Before we proceed to elaborate on the estimation of $u_{k,l}$, we will finish the derivation of the adaptive core of the algorithm.

In our application, neural recording, some electrodes are clustered in close spatial proximity. It would then be desirable to allow for the incorporation of this information into the calculation of the weight vector.

The algorithm allows for the incorporation of multiple measurements by approximately solving for the weight vector w_k as:

$$w_{N_k}^{o} = \text{argmin}_w \sum_{l \epsilon N_k} \mathrm{E}\left\| \boldsymbol{d_l}(i) - \boldsymbol{u_{l,i}} w \right\|^2 \quad (2)$$

where we denote by N_k the set of electrodes in close proximity to electrode k. The assumption here is that $w_k \approx w_l$ for k in N_k. Note that the algorithm allows, but does not require multiple measurements to be utilized. In practice, electrode geometry and computational restrictions can determine whether one or multiple measurements should be utilized. Although very powerful, hardware cost considerations call for the lowest number of recording channels to be used at any given time, as any additional recording channel not only adds hardware cost in this algorithm's implementation, but also increases the overhead on the overall system (i.e., recording front-end, data transfer, control circuitry).

The case where only measurements from electrode k are used to clean said electrode is admissible as a special case. In this case, the above sum collapses to a single element. This special case is the one that we used for the hardware implementation of our proposed solution. The performance lost due to using a single measurement is heavily outweighed by the introduction of our template detection (Section 3.2.2) method, which is a novel method of obtaining $u_{k,l}$ directly from measured data (one additional electrode measurement k').

3.1.1. Adaptive Filtering

The evaluation of the expression in equation (2) is infeasible in hardware; instead the estimated weight vector $w_{N_k}^{o}$ is iteratively calculated, which can be implemented in hardware using adaptive filtering. This estimated weight is:

$$\boldsymbol{w}_{k,i} = \boldsymbol{w}_{k,i-1} + \frac{\mu}{\|\boldsymbol{u}_i\|^2 + \epsilon}\boldsymbol{u}_i^T\left[\underbrace{\boldsymbol{d}_k(i) - \boldsymbol{u}_i\boldsymbol{w}_{k,i-1}}_{e}\right] \quad (3)$$

and the cleaned neural signal is then estimated by the following:

$$\widehat{\boldsymbol{s}_k}(i) = \boldsymbol{d}_k(i) - \boldsymbol{u}_i w_{k,i} \quad (4)$$

Where, i is the time index, k is the electrode index (channel), w represents filter coefficients, u is a signal correlated with the artifact, d is the measured signal, and $\hat{s}$ is the cleaned neural signal.

A simple representation of an adaptive filter [40] is shown in Figure 3.2, which is used as the basis for our implementation.

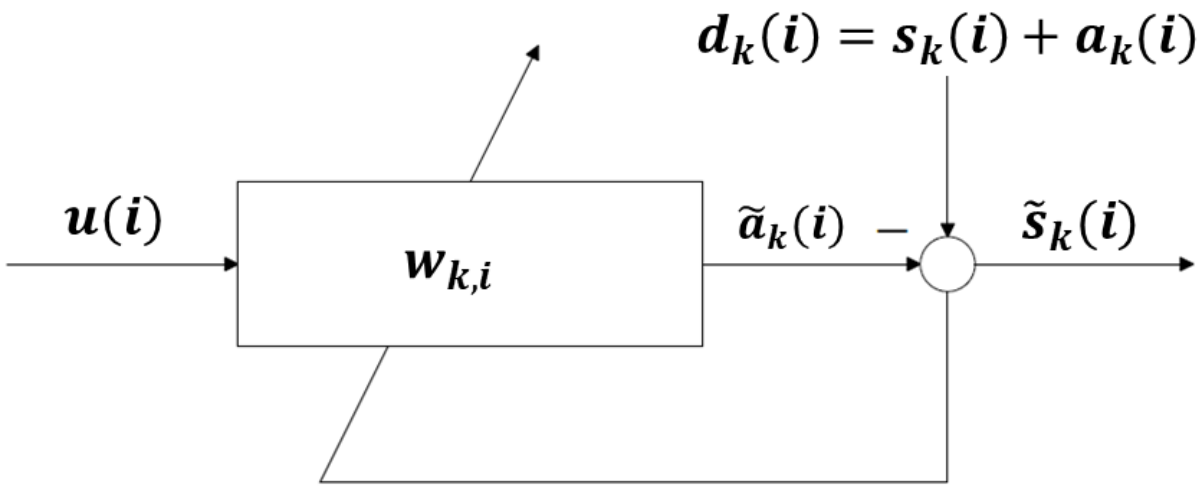

Figure 3.2: Adaptive Filter

The design and implementation of this algorithm has been done with full consideration of hardware and application requirements. It is therefore important to note some the significant components and modifications that were implemented to enable this solution to achieve the fastest

convergence reported for an adaptive stimulation artifact rejection solution while enabling it to work with arbitrary stimulation pulses without a need for prior knowledge of stimulation pulse characteristics or stimulation/recording timing information.

3.1.1.1. Normalized Least Mean Square (NLMS)

In prior implementations of LMS adaptive filtering solutions for stimulation artifact rejection in neuromodulation applications, a fixed step-size (μ') is used for calculation of error signals and the filter coefficients (w). This, we found, to be one of the reasons that the method in [36], and others, suffer from very long convergence times or low accuracy. First major difference is that we calculate an appropriate step-size each time a new sample is received (using norm calculations), and avoid dealing with accuracy vs convergence time trade-off. This results in faster convergence times while maintaining the accuracy of the results.

Additionally, the ASAR is implemented, as will be shown in the following sections and chapter, in a fully digital feed-forward manner, which avoids injecting noise at the input of the front-end and does not limit the filter's attenuation as no feedback DAC is required.

3.1.1.2. Posteriori Error

In an adaptive filter implementation of (3), the priori error (denoted by e), can be and is commonly used as the estimate of the neural signal. While for really small step sizes, the priori error and posteriori error are roughly equal in steady-state, the choice between them becomes an important one for us since our step size is not fixed and is chosen aggressively to improve the convergence rate in comparison to the state-of-the-art. It has, indeed, been shown that using

posteriori error as oppose to the priori error can lead into better performance of the algorithm, as it is less sensitive to varying step size values [41].

The feed-forward nature of ASAR, is another reason we chose to use the posteriori error. As a result we obtain the estimate of the neural signal, $\widehat{s_k}(i)$, using the more recent coefficients, $w_{k,i}$, as shown in (4).

The choices to go with the normalized full least mean square and use the posteriori errors, each would require an additional M adder and multipliers (where M is the order of the implemented filter); however, they result in faster convergence times and better performance.

3.1.1.3. The choice of template u_i

The choice of the template u_i is critical. The stimulation pulse itself, as employed in [10], [36], is not suitable for this purpose, because: (a) the mapping from stimulator through stimulation electrode, brain tissue and sensing electrode is highly non-linear, resulting in the need for complex filters and long convergence times, and (b) prior knowledge about the structural and temporal shape of the stimulation pulse is required. To remedy both drawbacks, the blind template detection method was developed, which operates without information on the stimulation waveform. By obtaining a template from an adjacent electrode and learning only the mapping between adjacent recordings, a linear NLMS filter with 16 taps turns out to be sufficient. The difference between the conventional choices for template versus our choice for the template has been visualized in Figure 3.3.

This method enables us to assume no characteristics for stimulation and sets no limits for delays between stimulation pulses that are observed at various recording sites, close or far. Most importantly, this enables our implementation to work with any arbitrary stimulation

pulse. We believe this is the main reason why our linear LMS adaptive filter, can so effectively, estimate and resolve a non-linear mapping (of the brain tissue); hence enabling us to offer an innovative solution at a much lower computational complexity/cost.

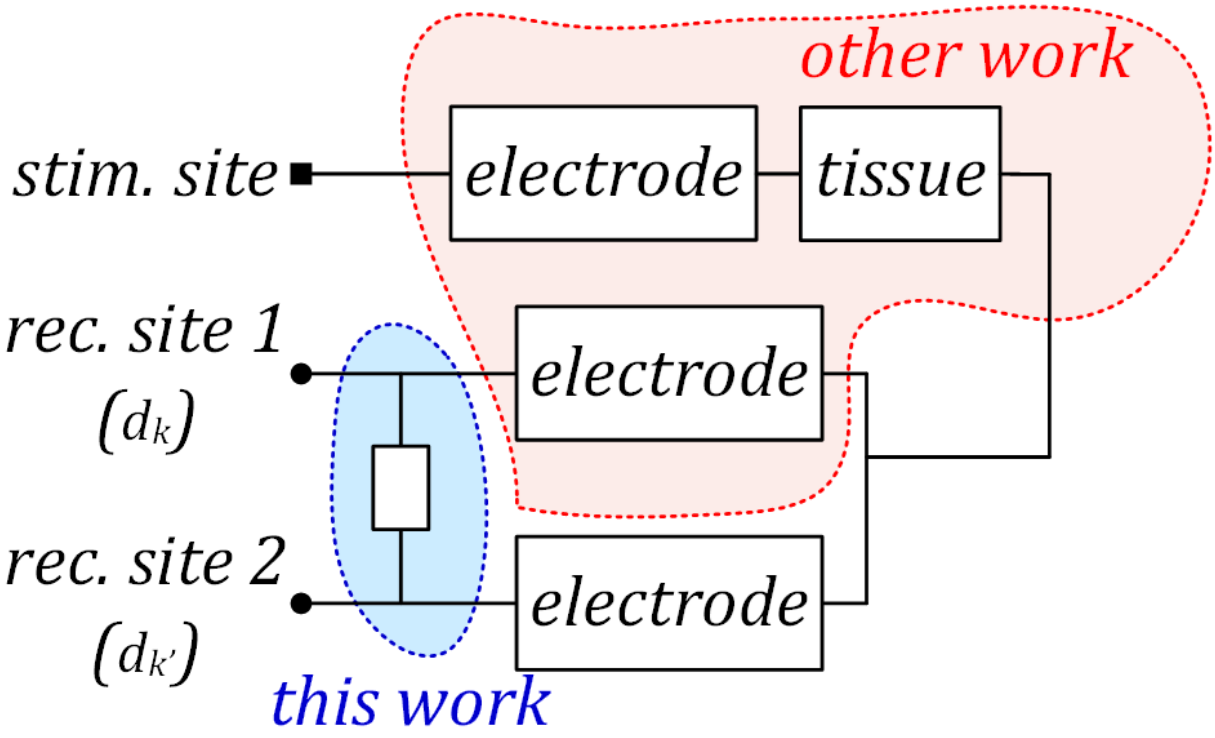

Figure 3.3: The Choice of Template u_i

3.2. Proposed Blind Adaptive Stimulation Artifact Rejection Solution

Figure 3.4 shows the block diagram for our proposed blind Adaptive Stimulation Artifact Rejection (ASAR) solution. ASAR operates in two phases: (I) statistics calculation (training), and (II) template detection and adaptive filtering.

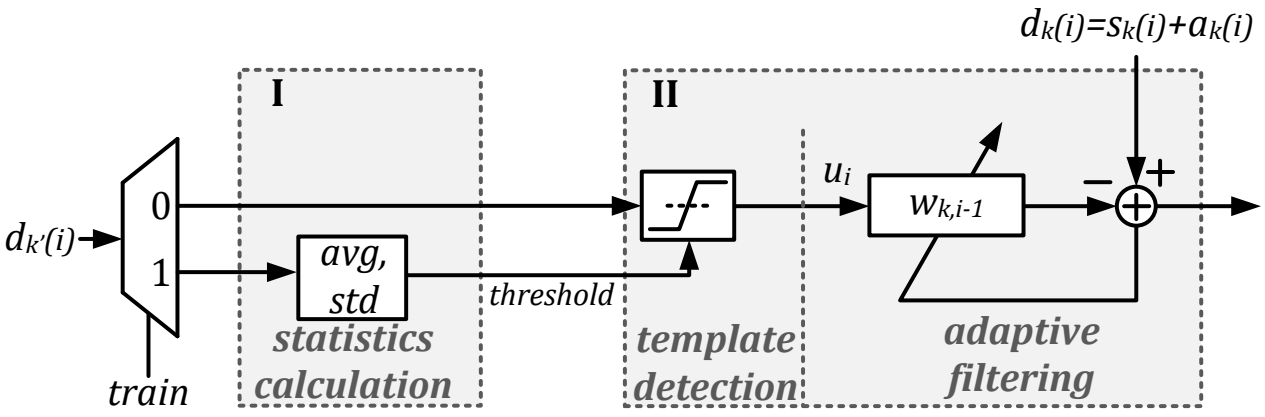

Figure 3.4: Blind Adaptive Stimulation Artifact Rejection (ASAR) Block Diagram

3.2.1. Statistics Calculation

Statistics of the neural signal, from an adjacent recording channel, are calculated in the absence of artifacts during the first N samples, and an appropriate threshold value is set. This is done by recursively updating the values:

$$\boldsymbol{S}(i) = \boldsymbol{S}(i-1) + x(i) \tag{5}$$

and,

$$\boldsymbol{T}(i) = \boldsymbol{T}(i-1) + x^2(i) \tag{6}$$

where $x(i)$ is the input sample at time i. Mean (avg) and standard deviation (std) at time $i = N$ are then calculated as (for N large enough):

$$avg = \frac{\boldsymbol{S}(N)}{N} \tag{7}$$

$$std \approx \sqrt{\frac{1}{N-1}(\boldsymbol{T}(N) - N.avg^2)} \quad (8)$$

The number of samples N is chosen as $N = 2^n$ for some n in order to reduce multiply/divide into shift operations, resulting in a more efficient hardware implementation.

3.2.2. Template Detection and Adaptive Filtering

To clean the measurement $\boldsymbol{d}_k(i)$ at electrode k, we choose a nearby electrode k' and determine a template $\boldsymbol{u}_i \in \mathbb{R}^{1\times 16}$. Based on the threshold obtained in the previous phase, $\boldsymbol{u}_i(l)$, the l-th element of $\boldsymbol{u}_i$ is estimated from $\boldsymbol{d}_{k'}(i)$ through blanking within $\alpha \cdot std$ of the mean:

$$\boldsymbol{u}_i(l) = \begin{cases} \boldsymbol{d}_{k'}(i-l), & if\ |\boldsymbol{d}_{k'}(i-l) - avg| \geq \alpha \cdot std \\ 0, & otherwise \end{cases} \quad (9)$$

An example of template detection is shown in Figure 3.5, where $\boldsymbol{d}_k(i)$, is the input to the template detection block. On the left, after the statistics calculation phase in completed, template detection is enabled, and based on the appropriate threshold set based on those statistics (shown in red), a template, $\boldsymbol{u}_i$, is extracted on the right.

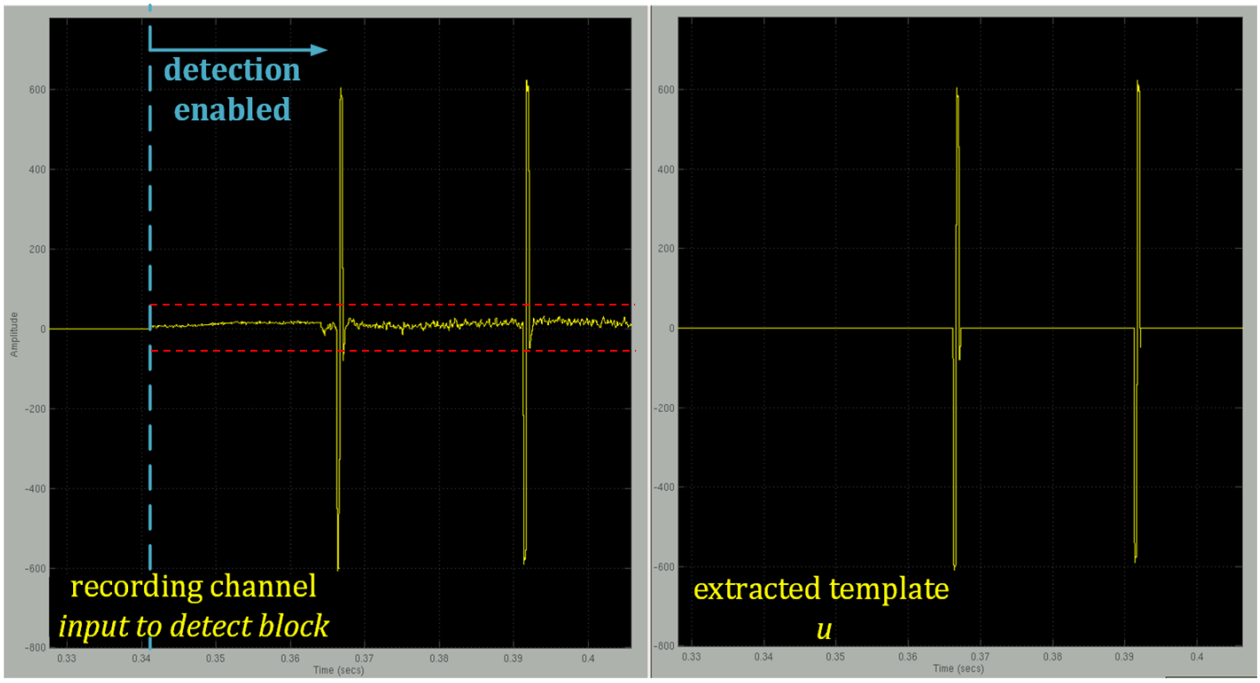

Figure 3.5: Template Detection Example

After an appropriate template, $\boldsymbol{u}_i$, is extracted, it is them applied to a Normalized Least Mean Square (NLMS) 16-tap adaptive filter. The clean neural signal $\widehat{\boldsymbol{s}_k}(i)$ is then obtained by subtracting the estimated artifact $\boldsymbol{u}_i \boldsymbol{w}_{k,i}$ from $\boldsymbol{d}_k(i)$ as shown in equations (3) and (4).

In the remainder of this chapter, we review some simulations that were performed to evaluate the effectiveness of this proposed algorithm. The hardware implementation details, as well as the fabricated integrated circuit along with further measurement results are presented in the next chapter.

3.3. Simulation Results

We planned to implement two ASAR solutions: (a) for LFP signals based on the analog front-end presented in [18] (called LFP front-end in this thesis), and (b) for the analog front-end presented in [20] that also covers Spikes (called Spike front-end for the remainder of this thesis). This is

because, with the availability of the mentioned front-ends, we will have the opportunity to integrate our method to achieve the proposed sensing solution presented in Figure 2.2, and a complete neuromodulation system. Therefore our simulations and testing should match those conditions.

The LFP and Spike front-ends operate at 6kHz, and 24~30kHz, respectively. Simulations were ran for both cases to make sure the algorithm can handle artifact rejection in all conditions.

For testing of the algorithm, our colleagues and collaborates in UCLA's neurosurgery department, provided real human data, with various stimulation pulse characteristics, which we ran through the algorithm for verification. The range for these tests have been summarized in Table 3.1.

Table 3.1: Testing Coverage of Stimulation Characteristic

Stimulation type	**Current amplitude**	**Pulse width**	**Pulse frequency**	**Stimulation pattern**	**Sampling frequency**
Micro & Macro	150µA - 2mA	200µs - 300µs	50Hz - 300Hz	Theta burst, pulse train	6 kHz & 30 kHz

In the next few figures, simulations for some of the test cases covered in the data shown in Table 3.1 are shared.

Continuous stimulation pulse train are a common modality used in various stimulation paradigms. Figure 3.6 shows a human patient recording with a continuous stimulation pulse of 1.2mA. To mimic the conditions for the LFP and Spike front-ends, as explained earlier, this recording is cleaned using ASAR at 6kS/s and 30kS/s, respectively. These cleaned signals are shown in Figure 3.6 (b) and (c).

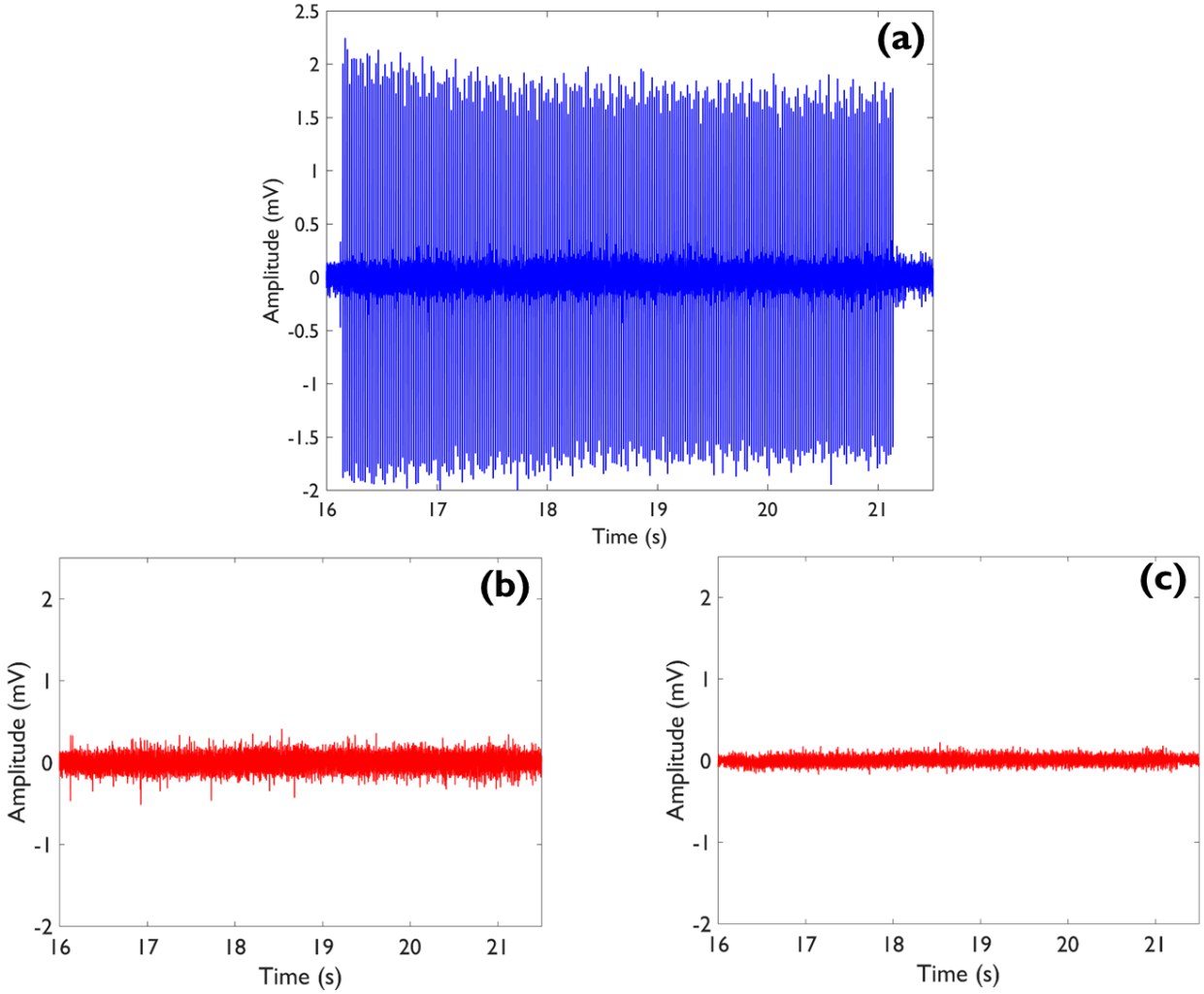

Figure 3.6: ASAR Simulation: Continues Stimulation Pulse (1.2mA), **(a)** Recorded Signal, **(b)** Cleaned Signal at 30kHz, **(c)** Cleaned Signal at 6kHz

Another stimulation modality is the use of burst stimulation pulses, theta-burst in our case. Figure 3.7 shows a human patient recording with a theta-burst stimulation pulse of 150µA. To mimic the conditions for the LFP and Spike front-ends, this recording is cleaned using ASAR at 6kS/s and 30kS/s, respectively. These cleaned signals are shown in Figure 3.7 (b) and (c).

A theta-burst stimulation protocol has been shown to be optimal for inducing Long-Term Potentiation (LTP) [42], and its use as a stimulation protocol in the human entorhinal area has been shown to improve memory specificity [43].

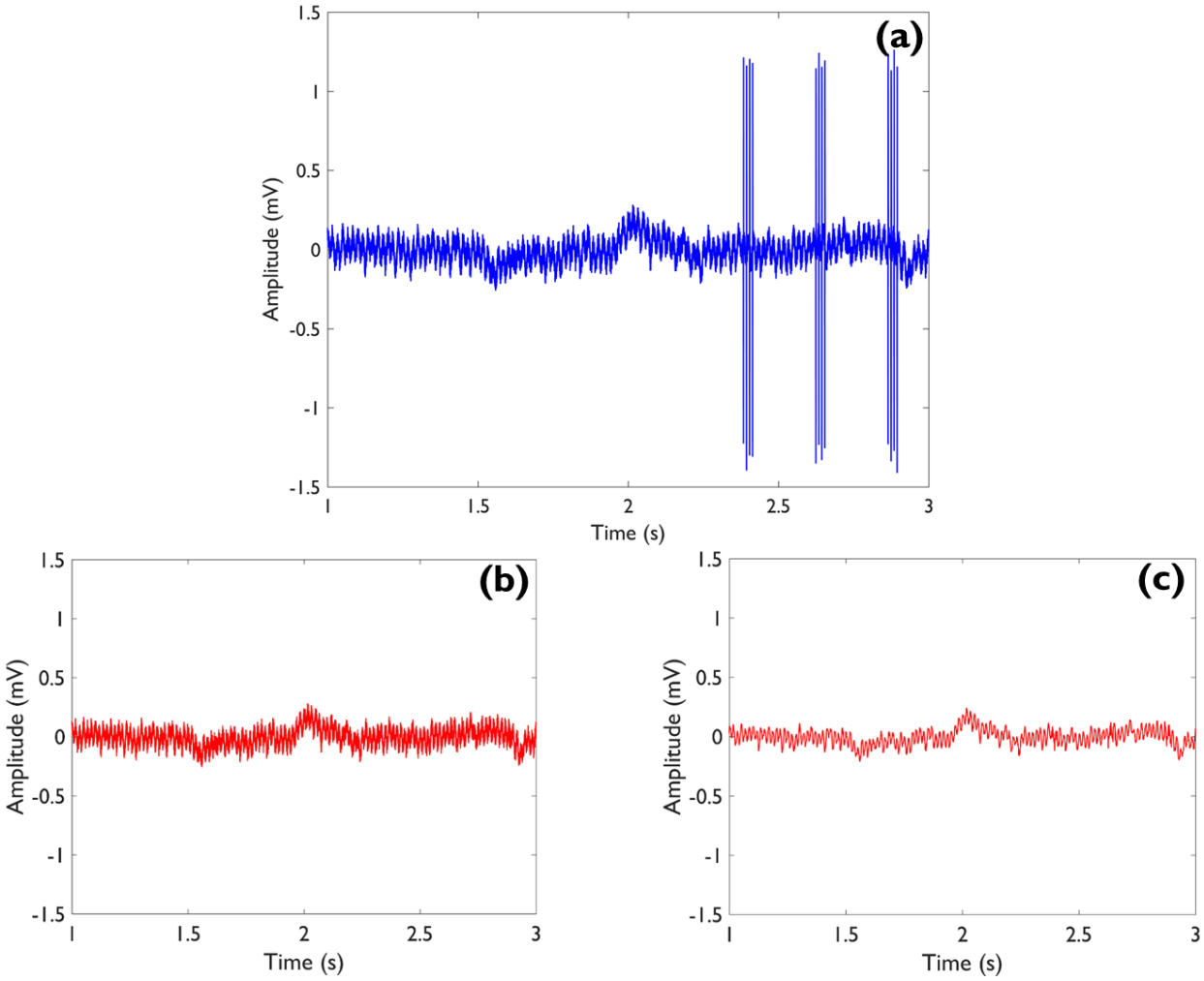

Figure 3.7: ASAR Simulation: Theta-burst Stimulation (150μA), **(a)** Recorded Signal, **(b)** Cleaned Signal at 30kHz, **(c)** Cleaned Signal at 6kHz

A different human patient recording with a theta-burst stimulation pulse, with a much higher stimulation current of 2mA, is provided in Figure 3.8. ASAR's cleaned output corresponding for the Spike and LFP front-ends, are shown in Figure 3.8 (b) and (c), respectively.

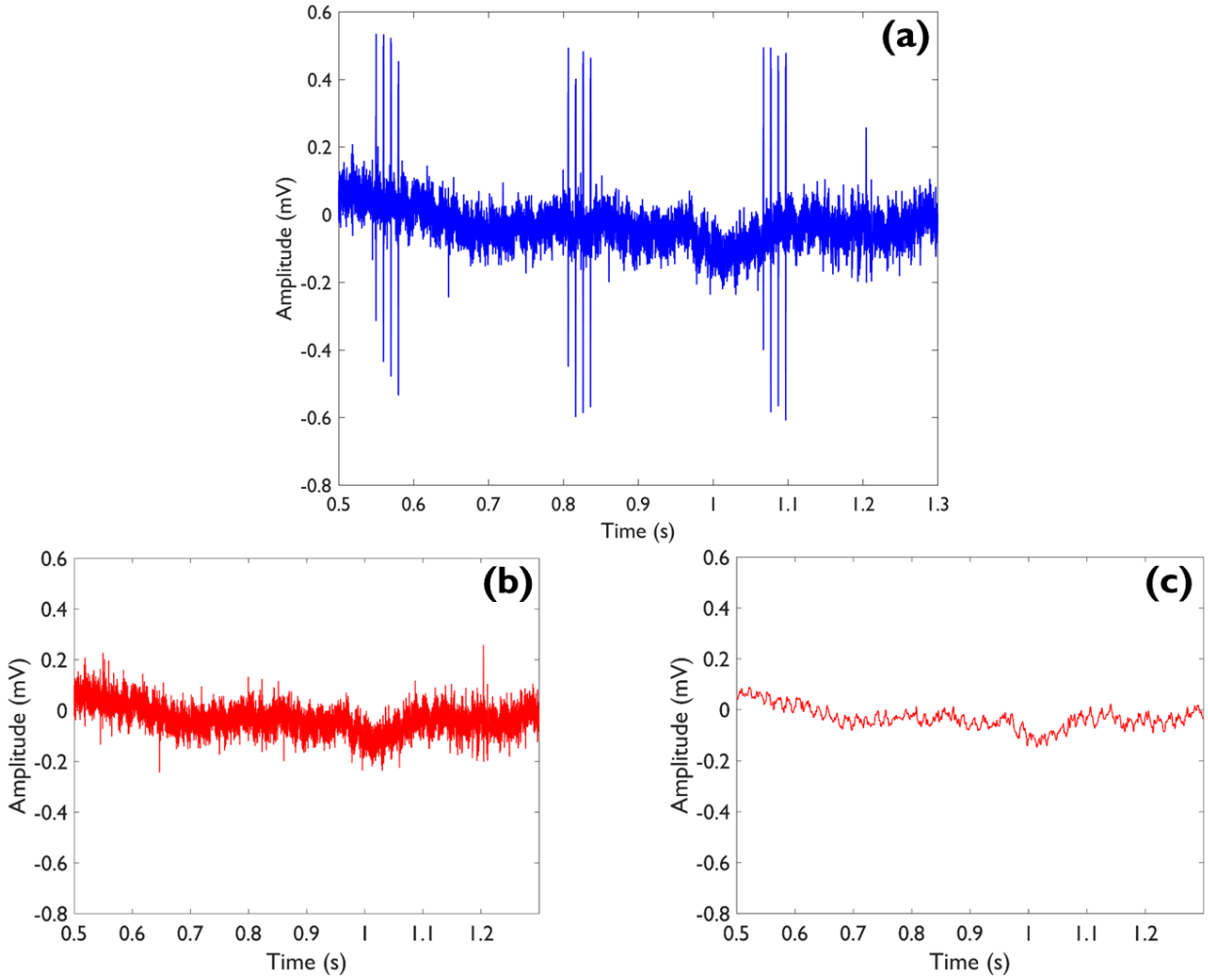

Figure 3.8: ASAR Simulation: Theta-burst Stimulation (2mA), **(a)** Recorded Signal, **(b)** Cleaned Signal at 30kHz, **(c)** Cleaned Signal at 6kHz

Although not commonly done, we obtained a data set with varying (increasing) stimulation pulse amplitude to observe the results in the simulation. Figure 3.9 shows this simulation result, where the blue signal is the recorded signal containing neural signal and artifact, and the red signal is the ASAR output, cleaned signal.

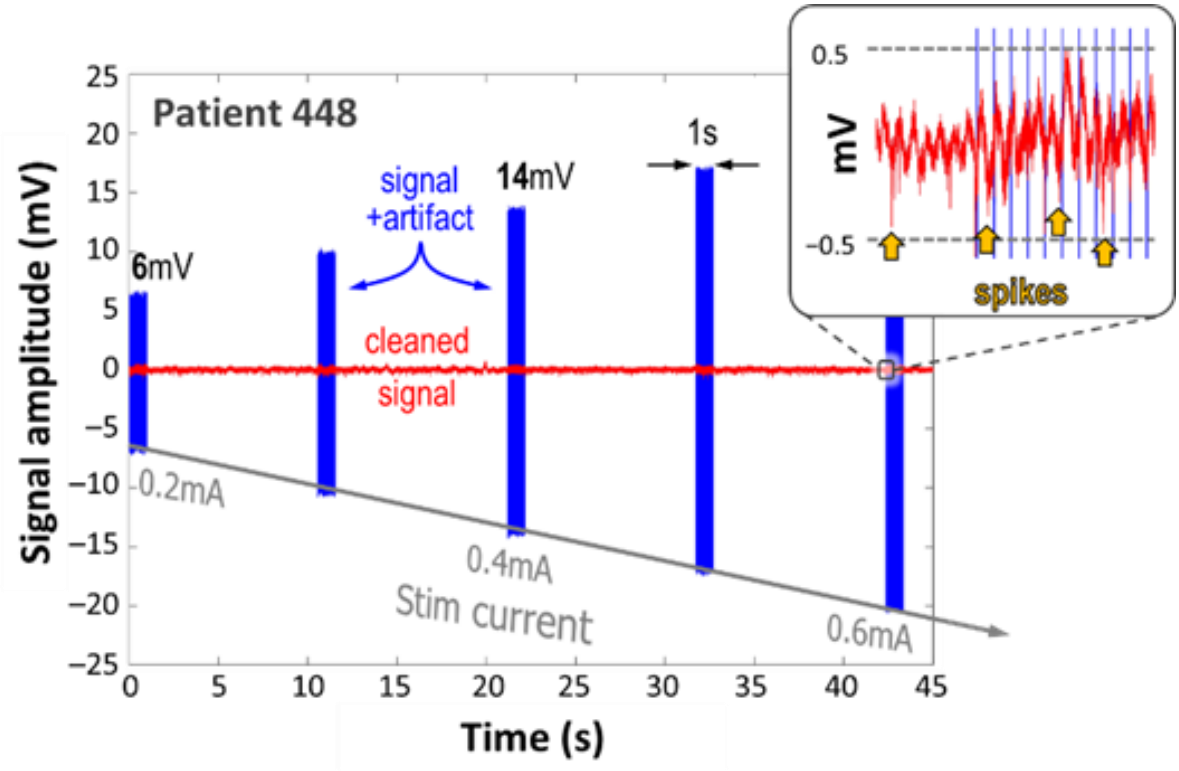

Figure 3.9: ASAR Performance with Varying Stimulation Levels

3.3.1. Difficulties in measuring true attenuation of the filter

One cannot simply measure the attenuation of this adaptive filter for stimulation artifact rejection, and in fact no one has sufficiently addressed this in the literature. This is mainly due to the fact, that neural signals are not a “known signal” and we cannot model them as easily as some signals in the communication field for example.

Some previous work observe the max amplitudes of the artifact and compare that to the amplitude of the cleaned neural signal to report attenuation numbers. This is not a true measure of the filter attenuation, because there is no distinction between the desired neural signal and the residual artifact present in the cleaned signal that may have not been removed. We try to address this in our chip measurements, and believe that the method presented in Section 5.3 comes very close to providing the best attenuation measure for these types of filters.

Another method, would be to look at the spectrum of the cleaned signal and compare it to the signal of interest. Here the main difficulty is the overlap of the artifact frequency components with the signals of interest. Therefore, this cannot be done for neural signals with clinical recordings. However, as my colleagues have been designing stimulator and sensing front-end solutions, which our ASAR solution will be integrated with, we had access to custom in-vitro measurements.

An in-vitro, concurrent stimulation and sensing, experiment was done where a continuous stimulation pulse of 3mA, with 2ms inter-pulse duration was recorded in conjunction with a ±3.5mV, 7Hz signal tone (mimicking a neural signal); this recording is shown in Figure 3.10(a). To test capability of our proposed ASAR algorithm, we cleaned this measurement signal and produced a spectrogram; an attenuation of up to 114dB in the dominant tone was achieved, as shown in Figure 3.10(b) (more details are available in [44]). Given the ideal nature of this test, we will never assume to get such great results in our hardware implementation and with human neural recordings, but this helps verify our algorithm.

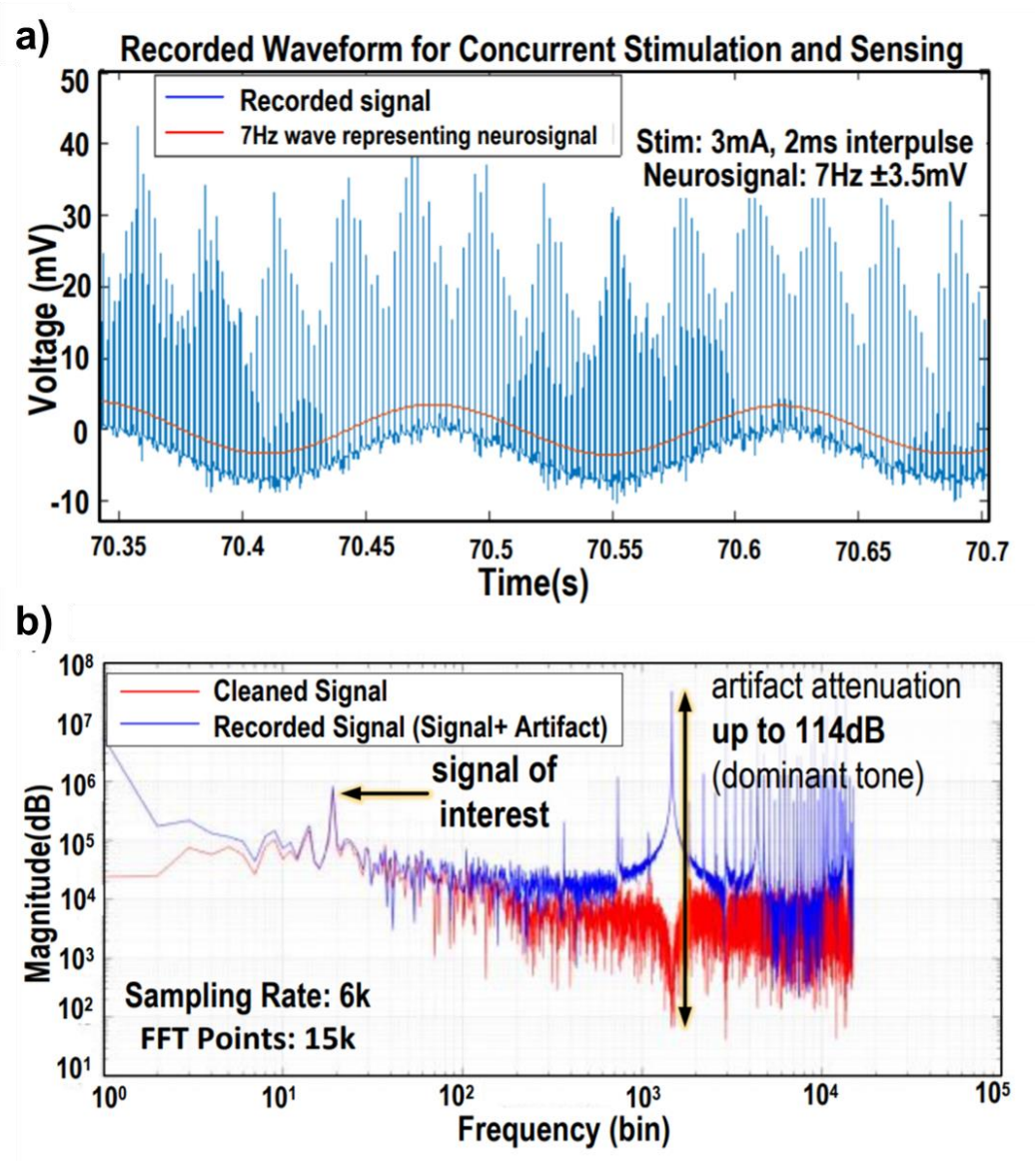

Figure 3.10: (a) Time-domain Waveform for Concurrent Stimulation and Sensing, **(b)** In-band Artifact Suppression Using ASAR Algorithm

CHAPTER 4

ASAR Hardware Implementation

In this chapter, the hardware implementation details and considerations are presented. Two designs were implemented, named the Spike ASAR and LFP ASAR for simplicity. Before going over implementation details, it is worthwhile to summarize the design specifications for the two ASAR designs (Table 4.1). These specifications are in line with all the requirements that have been previously mentioned and in line with the output specifications of the analog front-ends in [18], and [20], as mentioned previously. This would allow for integration of these ASAR designs with the mentioned analog front-ends to provide a complete sensing solution.

Table 4.1: Design Specification for Two ASAR Implementations

Design Specifications	**LFP ASAR**	**Spike ASAR**
Signal(s) of interest	LFP	Spike + LFP
Bandwidth (Hz)	1-200	1-5K
Sampling Rate (KS/s)	6	24
Input Resolution	16bits	12bits
Artifact amplitude (mV_{p-p})	<100	<40

As discussed in Section 3.2, and shown in Figure 3.4, ASAR operates in two phases: (I) statistics calculation (training), and (II) template detection and adaptive filtering.

4.1. Phase I: Statistics Calculation

The training phase during which the statistics of the neural signal is calculated using equations (7) and (8) in the absence of stimulation is straightforward. They were some hardware implementation considerations, however, that are summarized below.

4.1.1. The choice of N

N is the number of samples during which statistics calculation is performed at the initiation of ASAR operation, in the absence of stimulation. There are multiple division and multiplication operations involving N in the statistics calculation. As a result, number of samples N is chosen as $N = 2^n$ for some n in order to reduce multiply/divide into shift operations, resulting in a more efficient hardware implementation.

The larger the N, the better understanding of the statistical values we will have; however, a larger N value, also implies a longer training phase at the initiation of ASAR, during which no stimulation can be applied.

During extended testing of all the available data sets, we found $n = 13$ ($N = 2^n = 2^{13}$) to be a great compromise, applicable to both our designs. With this choice, we were able to reduce multiply/divide operations in statistics calculation into shift operations.

4.1.2. Implementation of Square Root Operator

Neural signals are very small in value, especially when compared to the stimulation artifacts and the full dynamic range available in the system to accommodate large artifacts. Additionally, the standard deviation calculated for neural signals is used to set a threshold value in the template detection method. It is therefore, imperative, that precise standard deviation values are obtained

for very small neural signals to avoid false positive template detection. As it can be seen in (8), these directly translates into the ability for improved precision of small input values for the square root operation.

This led to implementation of the square root operation as a look-up table. Here is a brief description of this implementation:

- Look –up table of 798 entries: $\left[\sqrt{256}, \dots, \sqrt{1026}\right]$.
- Incoming inputs to the square root operation are normalized by even shifts (2^N) left or right: $x = 2^{2N} \times x_n$.
    - Here, $256 \leq x_n \leq 1024$
- Square root can then be calculated: $\sqrt{x} = 2^N \times$ table entry $[int(x_n) - 256]$.

With this implementation, improved precision for smaller numbers are achieved while maintaining an upper bound on percentage error.

4.2. Phase II: Template Detection and Adaptive Filtering

4.2.1. Template detection

The template detection method, along with its benefits, have been discussed in Section 3.2.2. The implementation of this block is shown in Figure 4.1. Although, simple in implementation, the addition of this novel idea, allows us to employ a linear NLMS filter with only 16 taps, to resolve a very non-linear "channel" of electrode interfaces and brain tissue, by obtaining a template from an adjacent electrode and learning only the mapping between adjacent recordings. Template detection, in addition, enables our implementation to work with any arbitrary stimulation pulse.

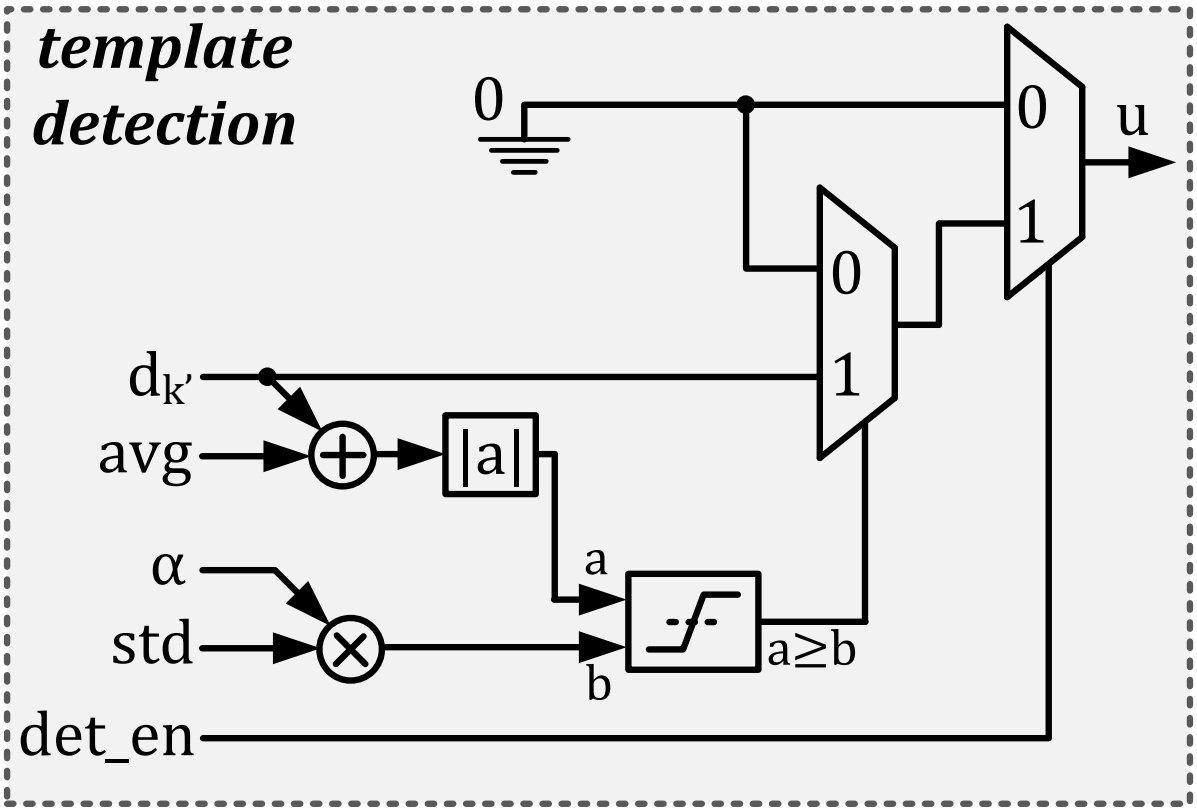

Figure 4.1: Template Detection

4.2.2. Adaptive Filtering

The hardware implementation of the adaptive filter encompasses all the components that were discussed in Section 3.1.1. As a result, a 16-tap NLMS filter was implemented, which uses the posteriori error to obtain the estimated (clean) neural signal. This filter adaptively adjusts the filter step size based on the norm calculations to obtain the best convergence speed and accuracy balance.

Figure 4.2 shows the implementation of the adaptive filter, governed by equations (3), and (4). The weight update block is highlighted in red, and the filtering region in blue. Paths for calculation of both the error signal ($e(i)$) and the cleaned neural signal ($\hat{s}(i)$) are show in this

figure. To avoid congestion, full details on calculation of u is not shown in the figure, but should easily follow from (3).

The number of bits used to implement this 16-tap filter is also indicated in the figure. Places where multiple numbers are shown (e.g., 12, 16) indicate the number of bits that are used for the implementation of Spike ASAR, and LFP ASAR, respectively.

As shown in the implementation the posteriori error is used to calculate the estimated neural signal ($\hat{s}(i)$), while the priori error is used the weight update, as it is done in conventional implementation of adaptive filters.

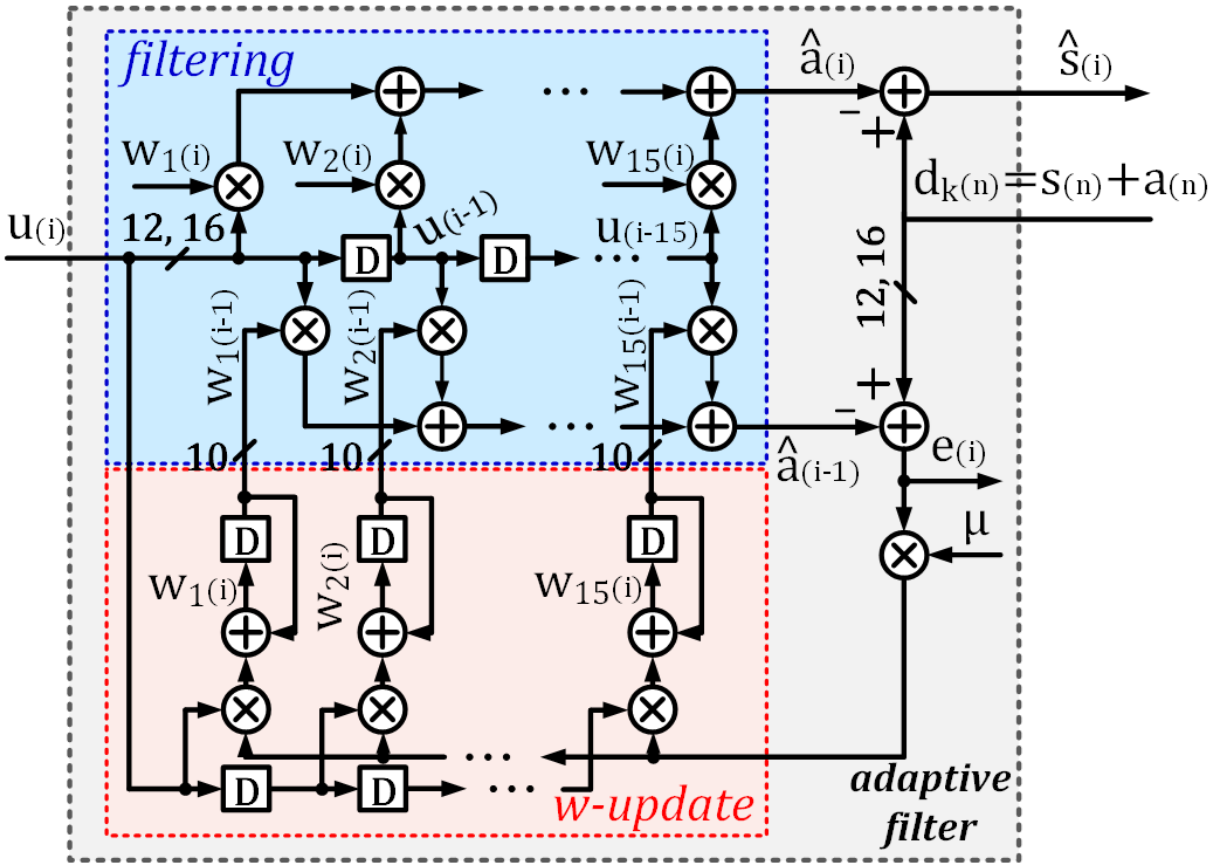

Figure 4.2: Adaptive Filter

4.3. ASAR Timing Diagram

The ASAR algorithm and the hardware implementation for various blocks have been explained. Next, the operation and execution of this algorithm in hardware is reviewed by utilizing timing diagrams for different phases of operation. For simplicity, these timing diagrams only show some of the critical signals to express the operation of the system.

At the initiation of ASAR, through chip start-up or reset, phase I is activated to calculating statistics of neural signal (training). The timing diagram for this phase is shown in Figure 4.3.

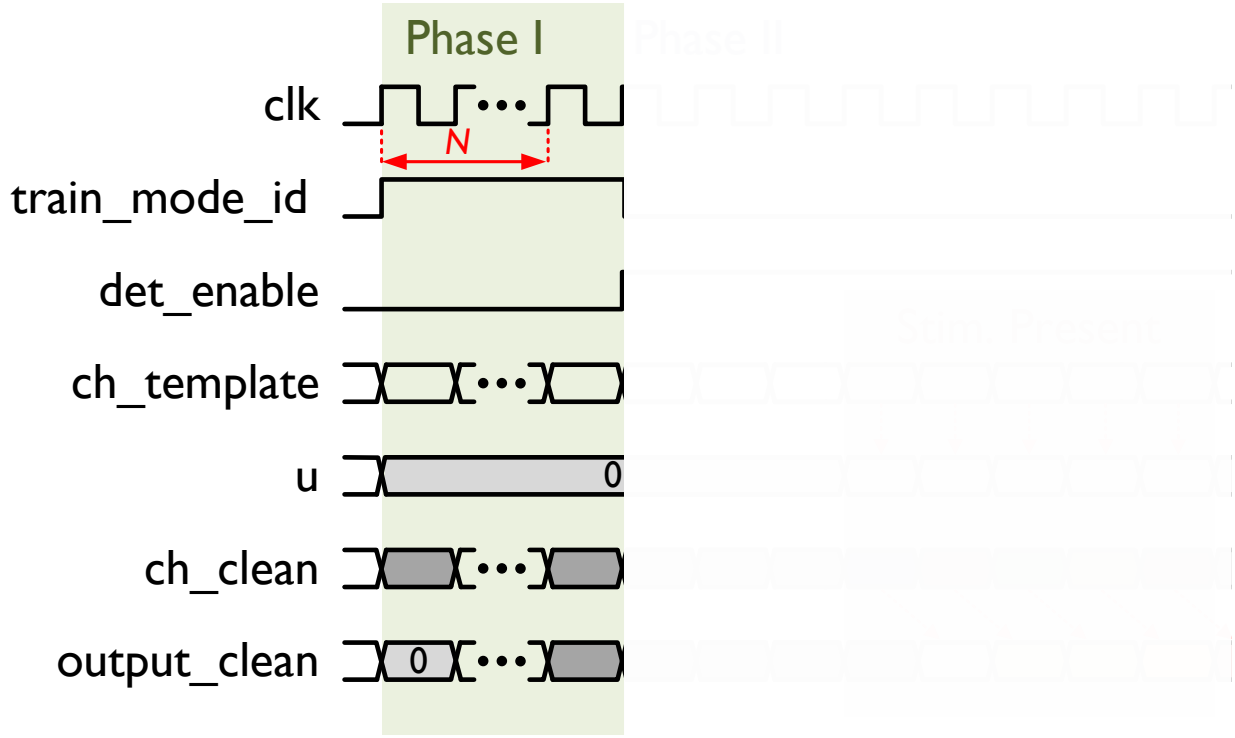

Figure 4.3: ASAR Timing Diagram: Phase I

During this phase which lasts $N + 1$ clock cycles, the statistics of neural signals are calculated during the first N samples, and stored (in the following clock cycle). The `train_mode_id` is set to 1, and `det_enable` is set to 0 during phase I. Here, the template, `u`, is zero regardless of

the value of the adjacent recording channel (`ch_template`). Lastly, the output signal (`output_clean`) is just the delayed version of the input recording channel (`ch_clean`).

At the end of phase I, the desired threshold is obtained and system enters phase II of the operation (template detection and adaptive filtering). During phase II, stimulation may or may not exist. We first look at the timing diagram for phase II, without stimulation present in Figure 4.4.

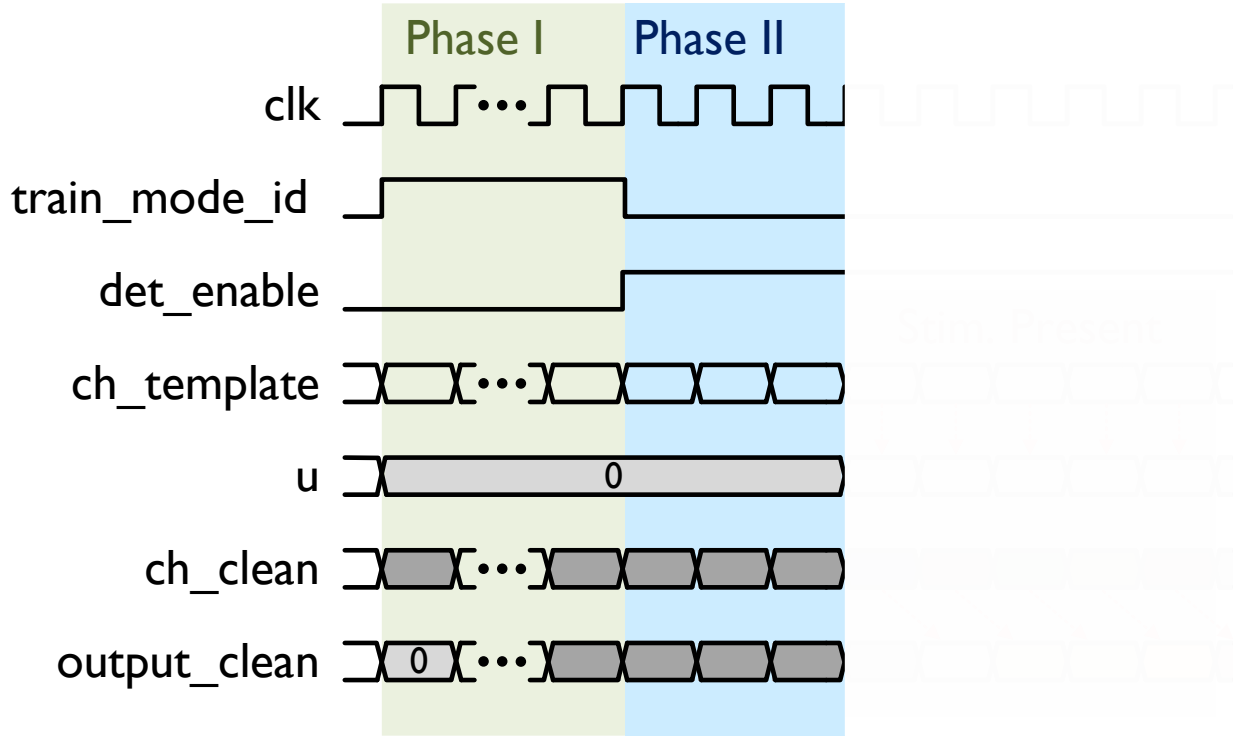

Figure 4.4: ASAR Timing Diagram: Phase II, No Stimulation

In phase II, training has ended (`train_mode_id = 0`) and template detection is enabled (`det_enable = 1`). In the absence of stimulation, however, no template is detected and therefore template, `u`, stays at zero, and the output signal (`output_clean`) is just the delayed version of the input recording channel (`ch_clean`).

Figure 4.5 shows the timing diagram for when ASAR encounters a stimulation in phase II operation and a template is detected on the adjacent recording channel (`ch_template`). This adjacent recording is then used as the template, `u`, for adaptive filtering. Note that due to the fully combinational implementation of the template detection method as shown in Figure 4.1, there are no clock cycle delays added during this operation. Here, the adaptive filter block uses the detected template and estimates the cleaned neural signal (`output_clean`) from the input recording channel (`ch_clean`).

ASAR has 4 clock cycle latency at the output, however, only one shown in the timing diagrams.

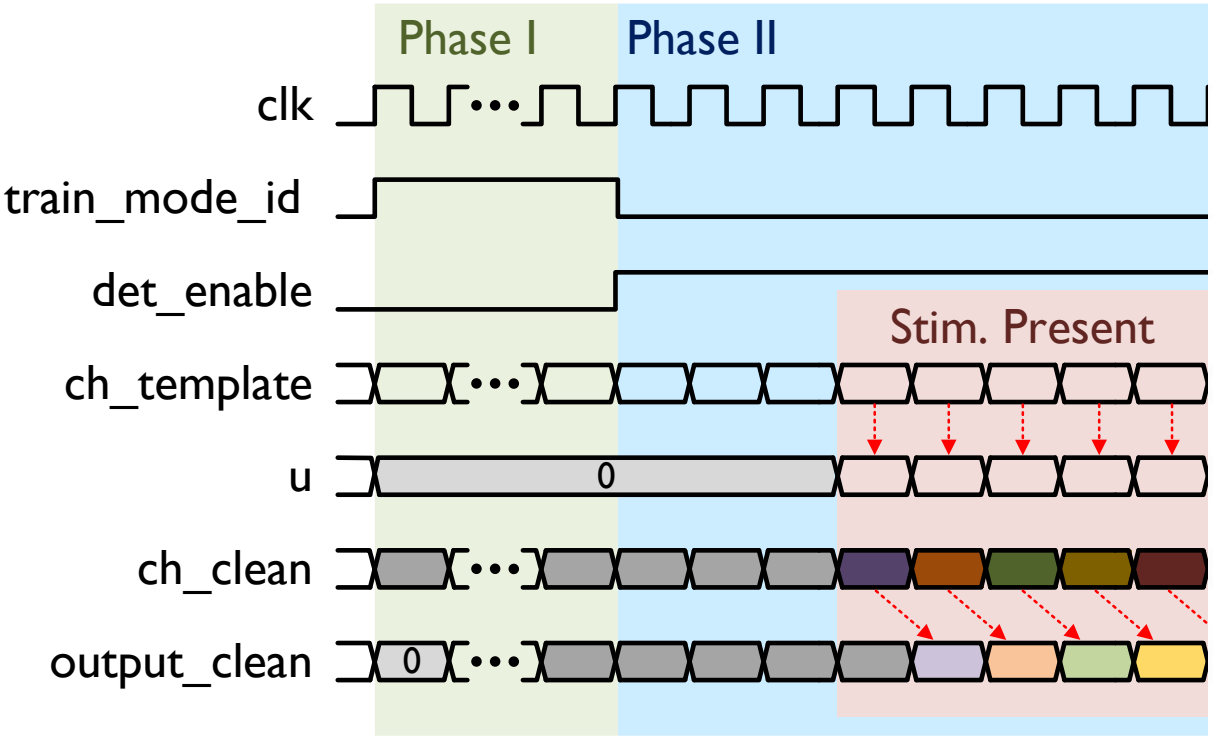

Figure 4.5: ASAR Timing Diagram: Phase II, Stimulation Present

Once the stimulation pulse subsides, the template becomes zero again. This has been shown in Figure 4.6.

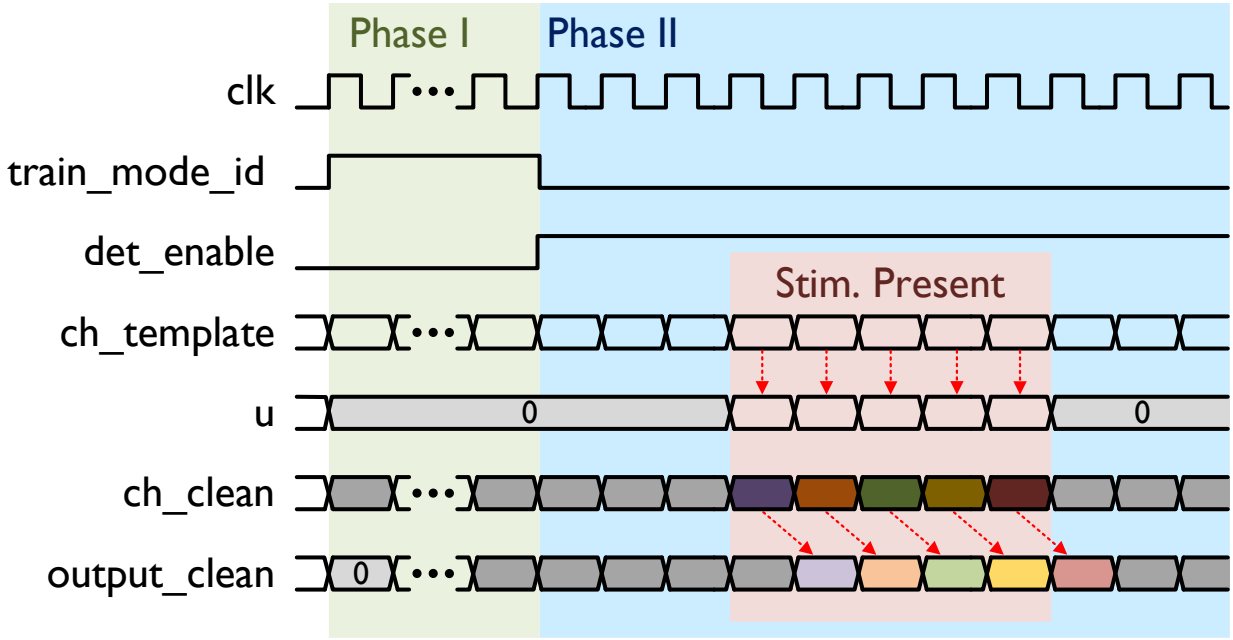

Figure 4.6: ASAR Timing Diagram: Phase I and II

Additional provisions are included for the practical use of the ASAR hardware, which were not illustrated in the timing diagrams. One example is the addition of a `calc_rst` signal that can reset statistics calculation, or in other words re-train the algorithm. Although a single training is sufficient, patient movement, passage of time, or other factors can introduce various changes and offsets in the system; it would be beneficial to have the option to re-train the system at any given point. `calc_rst` can be utilized without the need for a global chip reset.

Next chapter includes the IC implementation of ASAR hardware for both the LFP and Spike ASAR designs in 40nm CMOS technology. Chip measurements for this implementation is also included in the same chapter.

CHAPTER 5

IC Implementation and Measurement Results

5.1. IC Implementation

Two ASAR designs, governed by the design specifications summarized in Table 4.1, were designed and fabricated in 40nm CMOS technology. Table 5.1 summarizes the input/output signals for these designs.

Table 5.1: ASAR IC Implementation Input and Outputs (IO)

Signal Type	Signal		Comment
	LFP ASAR	**Spike ASAR**	
Input	`clk`		clock signal
	`ch_clean <15:0>`	`ch_clean <11:0>`	recording channel to be cleaned. $\boldsymbol{d}_k$ in (4).
	`ch_template <15:0>`	`ch_template <11:0>`	adjacent recording channel to be cleaned. $\boldsymbol{d}_{k'}$ in (4).
	`calc_rst`		calculation reset signal (sync), used for re-training.
	`thresh_scale <1:0>`		threshold scaling factor. α in (9).
	`global_rst_n`		asynchronous global reset
Output	`train_mode_id`		training mode indicator
	`output_clean <15:0>`	`output_clean <11:0>`	output, cleaned neural signal

It is important to note, that although, the designs allocated 16-bit and 12-bit, for recording inputs and output of LFP and Spike ASAR, respectively, due to limited number of PADs available during the tape-out, all multi-bit input and outputs were serialized and a single bit was used to implement them. Subsequently, this resulted in the addition of Serial-to-Parallel blocks at the input interface, and Parallel-to-Serial block at the output interface of the design. Additionally the implementations were ran at 16x and 12x higher clock frequencies that initially required to accommodate these changes.

The implementations used a single supply (VDD) for their operation and were designed with supply scaling in mind; however, the increased clock frequency limited the supply scaling to 644mV.

Figure 5.1 is the chip micrograph of both ASAR designs. The LFP ASAR (shown in blue) is 749µm by 263µm in dimension, occupying a total area 0.197mm^2, while the Spike ASAR (shown in red) has dimension of 794µm by 263µm, resulting in total area of 0.209mm^2. These area numbers include the Serial-to-Parallel and Parallel-to-Serial circuitries, that will not be required once the solution is integrated in a sensing IC solution, as all the connections will be done internal to the chip.

Table 5.2 summarizes all the implementation details for both designs.

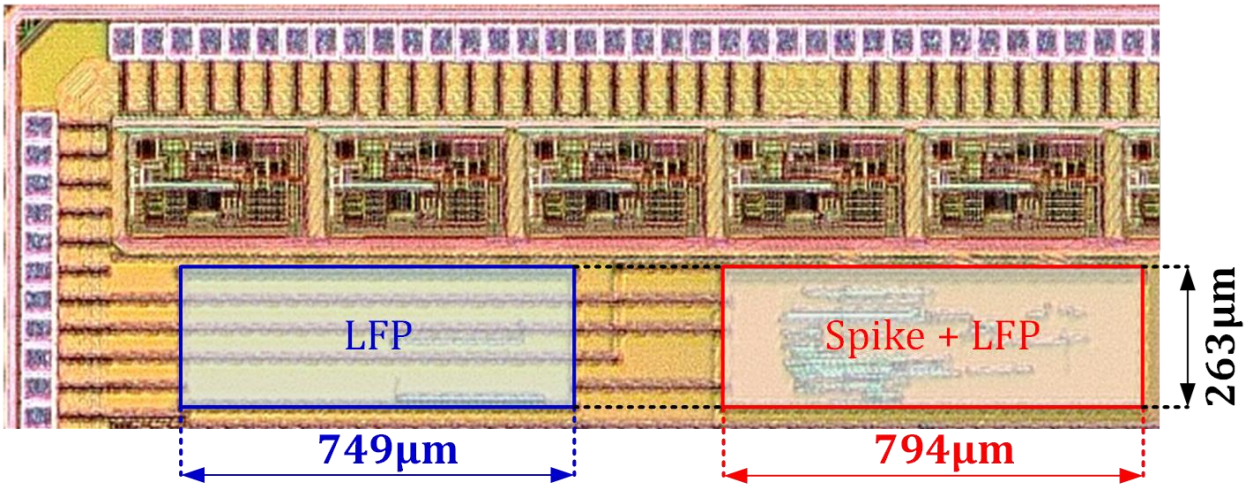

Figure 5.1: ASAR Chip Micrograph

Table 5.2: ASAR IC Implementation Details

	LFP ASAR	**Spike ASAR**
Technology	40nm LP HVT CMOS	
Supply voltage (V)	0.644	
Area	$0.197mm^2$	$0.209mm^2$
Operating frequency (KHz)	$96 = 16 \times 6$	$288 = 12 \times 24$

5.2. Testing Setup

To test the ASAR chip, a ROACH (Reconfigurable Open Architecture Computing Hardware) FPGA board [45] was used. An ASAR test Printed Circuit Board (PCB) was designed and manufactured. Figure 5.2 shows the ASAR chip test setup.

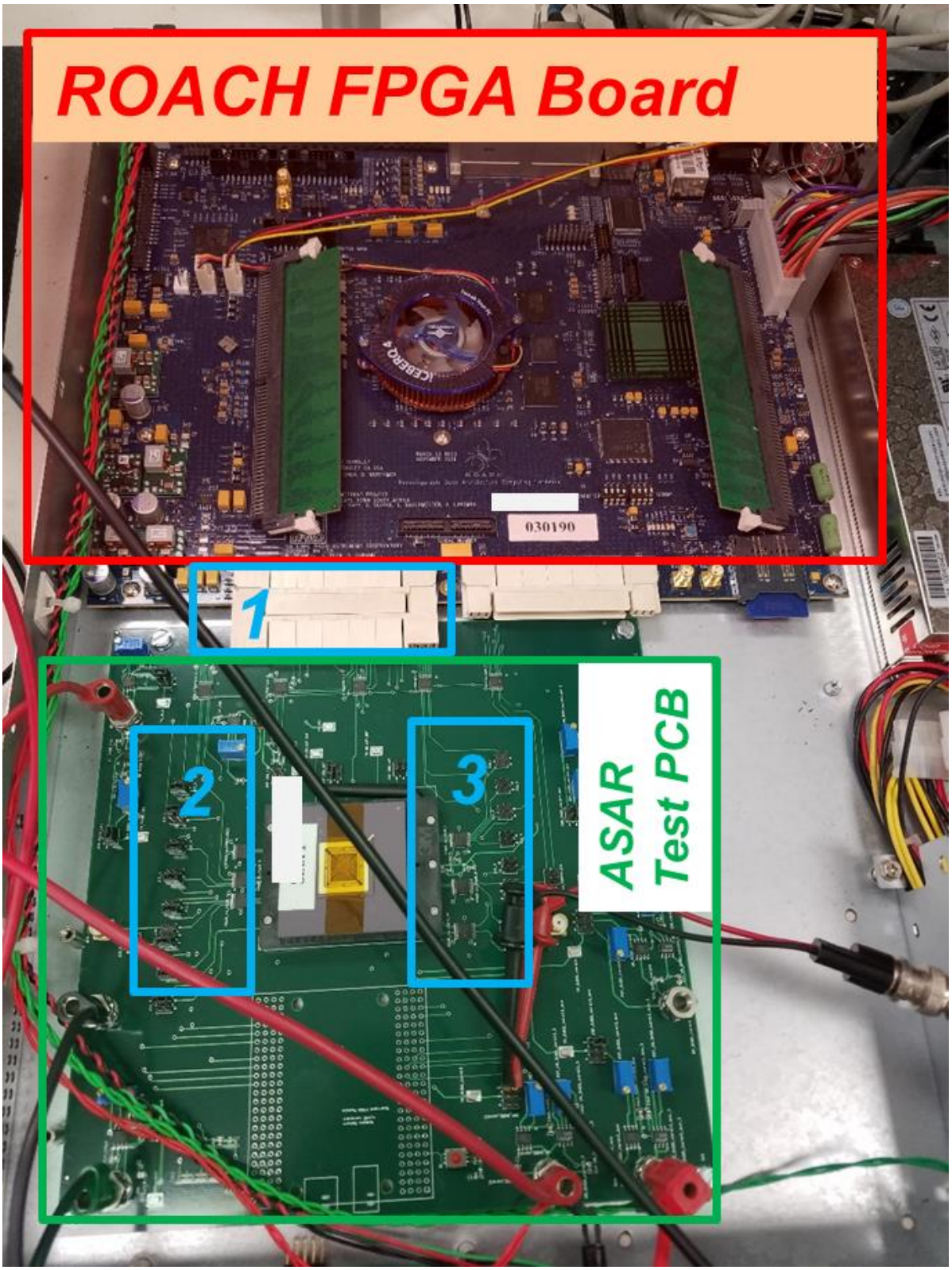

Figure 5.2: ASAR Chip Test Setup

The ROACH board hosts a Xilinx Virtex 5 FPGA, and is connected to the ASAR test PCB, using a Z-DOK connector interface (marked in region 1 of Figure 5.2). The packaged ASAR chip for testing is located in the center of the test PCB. Regions marked 2 and 3 in Figure 5.2, show the components and interfaces for testing the Spike and LFP ASAR designs respectively.

The ROACH FPGA testing models were designed in MATLAB/Simulink environment using Xilinx Blockset library. The LFP ASAR testing model is shown in Figure 5.3.

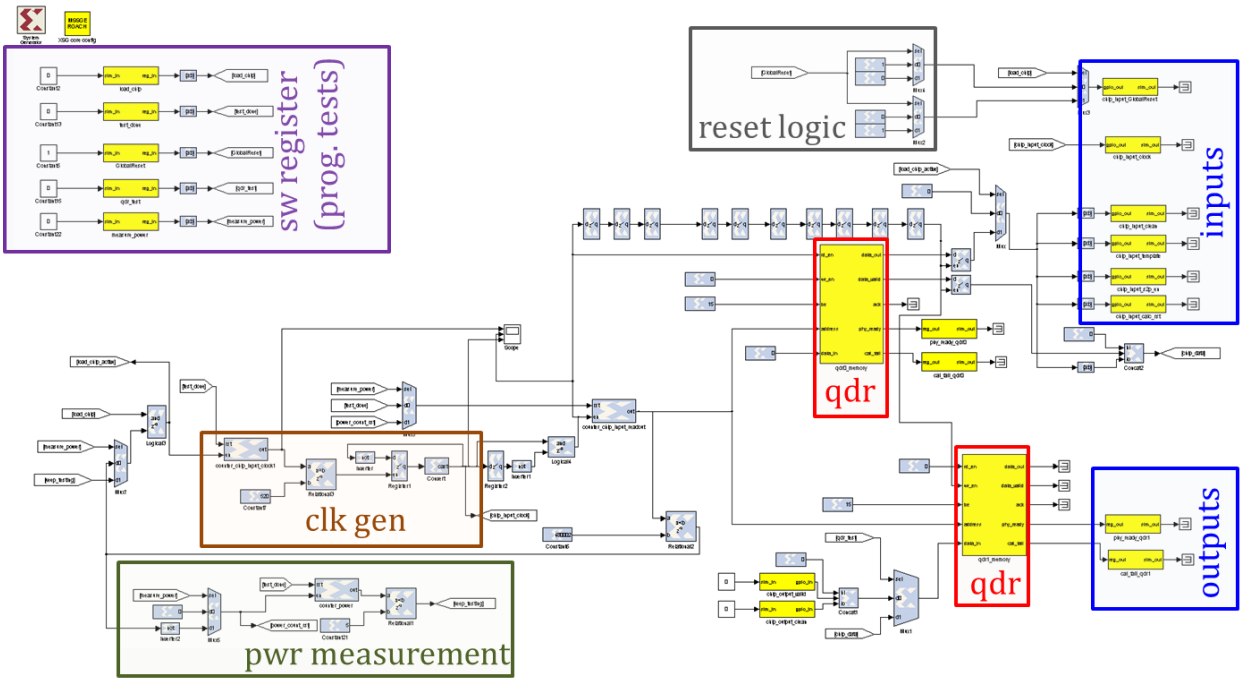

Figure 5.3: ROACH FPGA testing model (LFP ASAR)

5.3. Measurement Results

Clinical human patient data were used to test both ASAR designs. Chip measurement with sample data for LFP signals, with input resolution of 16 bits, and sampling frequency of 6kHz is shown in Figure 5.4. Phase I and phase II operations of ASAR are clearly indicated in this figure; at the end of stimulation pulses, a reset test was also performed, which is shown.

To perform the chip measurements for LFP ASAR across various data sets, we were able to scale the supply voltage down to 644mV, and while operating at 96kHz (Table 5.2), an average power consumption of 1.73µW was achieved.

Chip measurement with sample data for Spike ASAR, with recordings containing Spike and LFP signals, sampled at 24 kHz, with input resolution of 12 bits is shown in Figure 5.5. Again, phase I and phase II operations of ASAR are clearly indicated in this figure.

The supply voltage for the Spike ASAR was scaled down to 644mV, as well. An average power consumption of 3.02µW was achieved, while operating at 288kHz (Table 5.2).

At the time of measurement, recordings with artifacts up to $29mV_{p-p}$ (LFP) and $36mV_{p-p}$ (Spike) were available, resulting in artifact attenuation of up to 37dB and 40dB, respectively; however, the designs are capable of handling larger amplitudes (Table 4.1). These extremities were tested using synthetic data.

For average power calculations of both ASAR designs, stimulations are assumed to be ON for 5% of the total duration of the test, to provide a conservative measure.

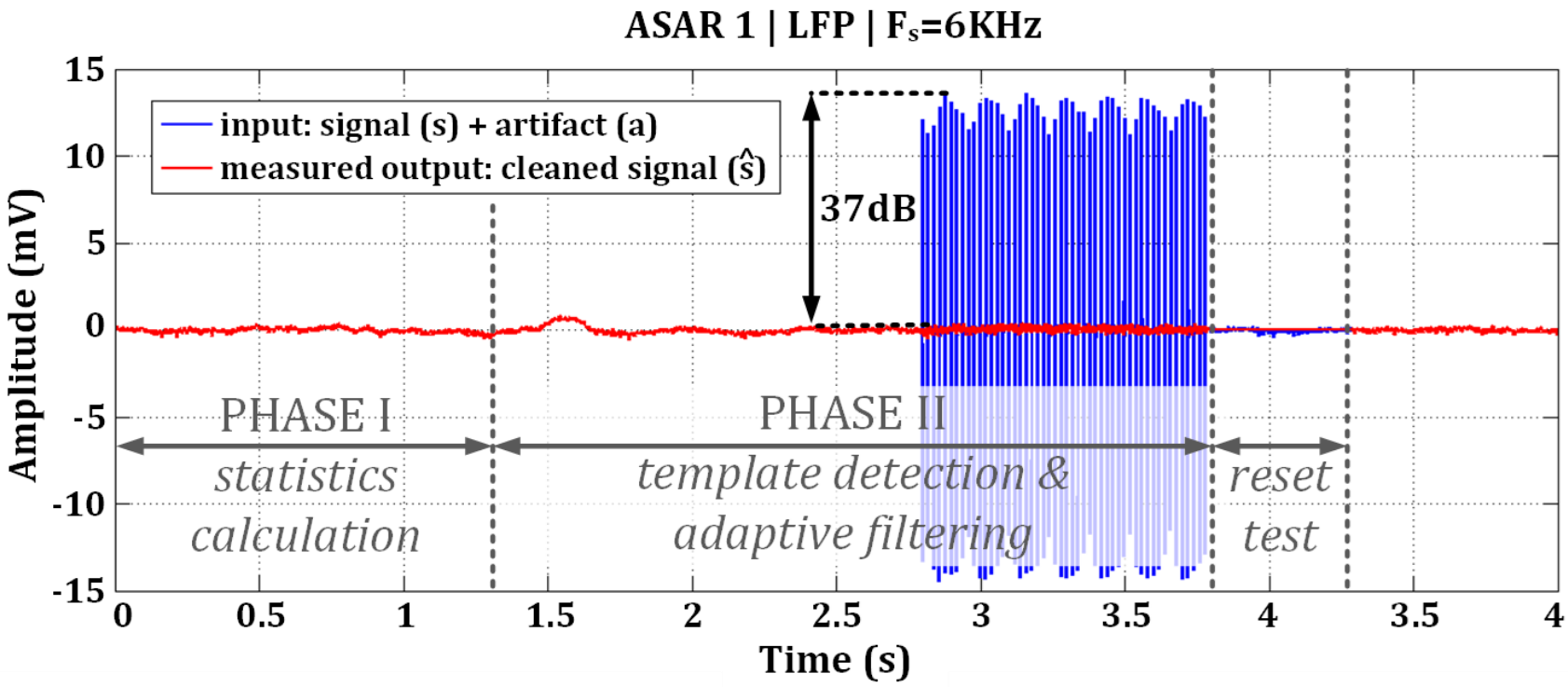

Figure 5.4: LFP ASAR Chip Measurement

(LFP recording, F_s=6kHz, input resolution=16 bits, stimulation artifact amplitude = 29mV_{p-p})

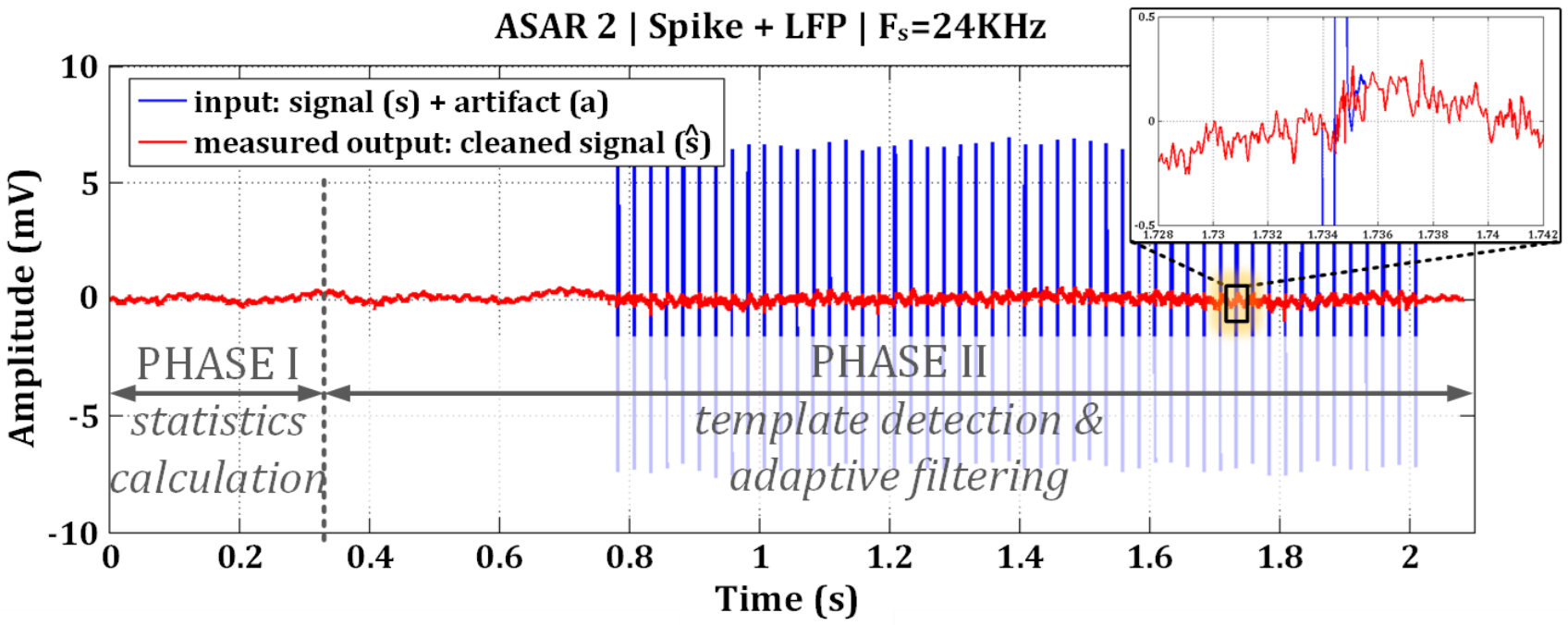

Figure 5.5: Spike ASAR Chip Measurement

(Spike + LFP recording, F_s=24kHz, input resolution=12 bits, stimulation artifact amplitude = 14mV_{p-p})

Figure 5.6 shows the time-domain recordings and spectrogram of neural signal (Spike and LFP) with $36mV_{p-p}$ stimulation artifacts with and without the ASAR activated.

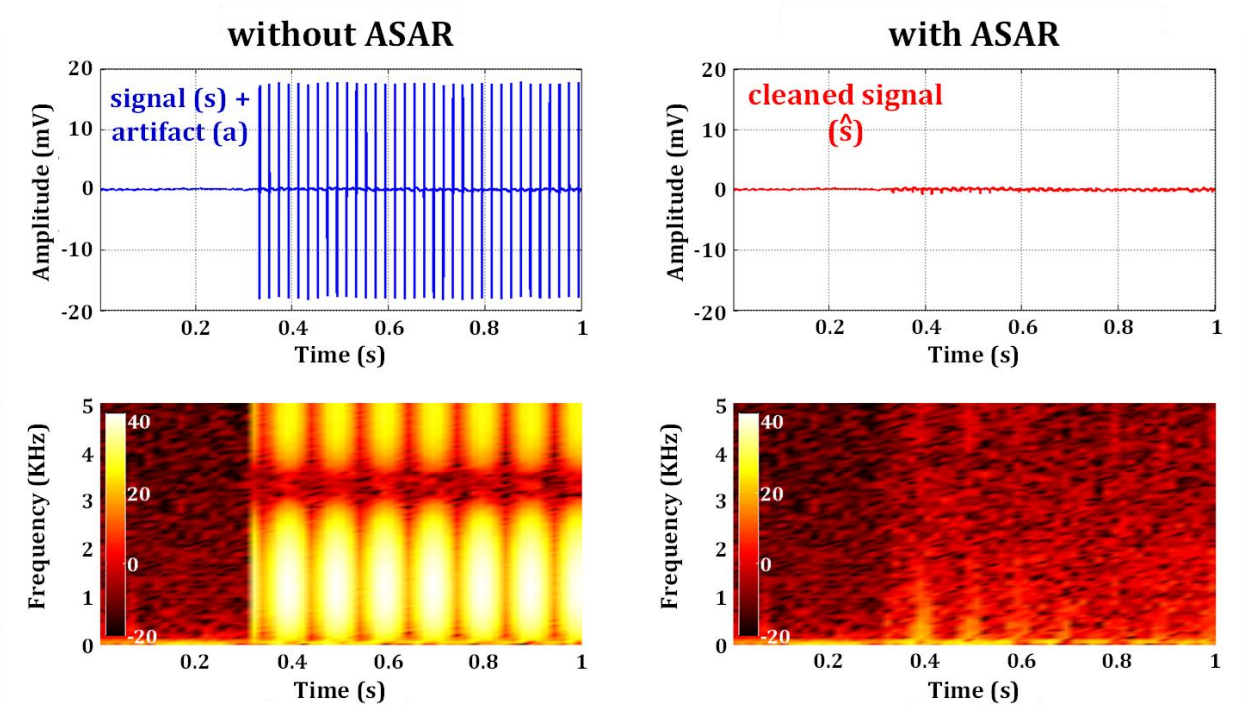

Figure 5.6: (left) Time-Domain Recordings and Spectrogram of Neural Signal (Spike + LFP) with $36mV_{p-p}$ Stimulation Artifact. (right) Measured ASAR Output Waveform and Spectrogram.

There is no unified way of reporting filter attenuation performance in stimulation artifact rejection applications. [36] reports attenuations of up to 42dB, primarily based on their use of 8-bit DAC for error subtraction; clearly this is a speculative attenuation figure and not an accurate one. In the same work and other methods, filter attenuation performance is sometimes reported by comparing the amplitude of the artifact and neural signals against each other. Although we do not find this satisfactory, these figures were provided for both our ASAR designs earlier in this Section

As it was previously mentioned in Section 3.3.1, a true measurement of the filter performance is difficult as the cleaned signal $\hat{\boldsymbol{s}}_k(i)$ includes the signal of interest combined with a small residual artifact, which are inseparable, making it impossible to obtain a ground truth.

In order to provide a better measure for the filter attenuation in this application, we devised a method (experiment) that calculates noise (artifact) power suppression based on SNR comparison of added known synthetic artifacts to clinical patient data (input) and cleaned signal by ASAR (output). This method along with some other measurement results in this chapter were introduced in [46].

To achieve this, two different set of human patient recording were obtained, from two separate patients. From one patient's data, neural signal (s) was extracted, in the absence of stimulation. Second patient's data was used to extract an artifact, called synthetic artifact (a) here. By addition of the neural signal (s) and synthetic artifact (a), we obtain an input signal similar to neural recordings in the presence of stimulation; one major difference is that in this method, neural signal and artifact are each are known quantities and can be completely separated with our knowledge.

Next step would be to use our input signal to test the ASAR. ASAR's output, the estimated cleaned neural signal ($\hat{s}$), can now be directly compared to the actual neural signal (s).

Signal-to-Noise Ratio (SNR) values were calculated at the input and output points, and compared to provide the noise (artifact) suppression that can be achieved by ASAR. This was done for varying stimulation artifact amplitudes; the results are shown in Figure 5.7. It can be observed that the ASAR can achieve noise power reduction of up to 49.2dB.

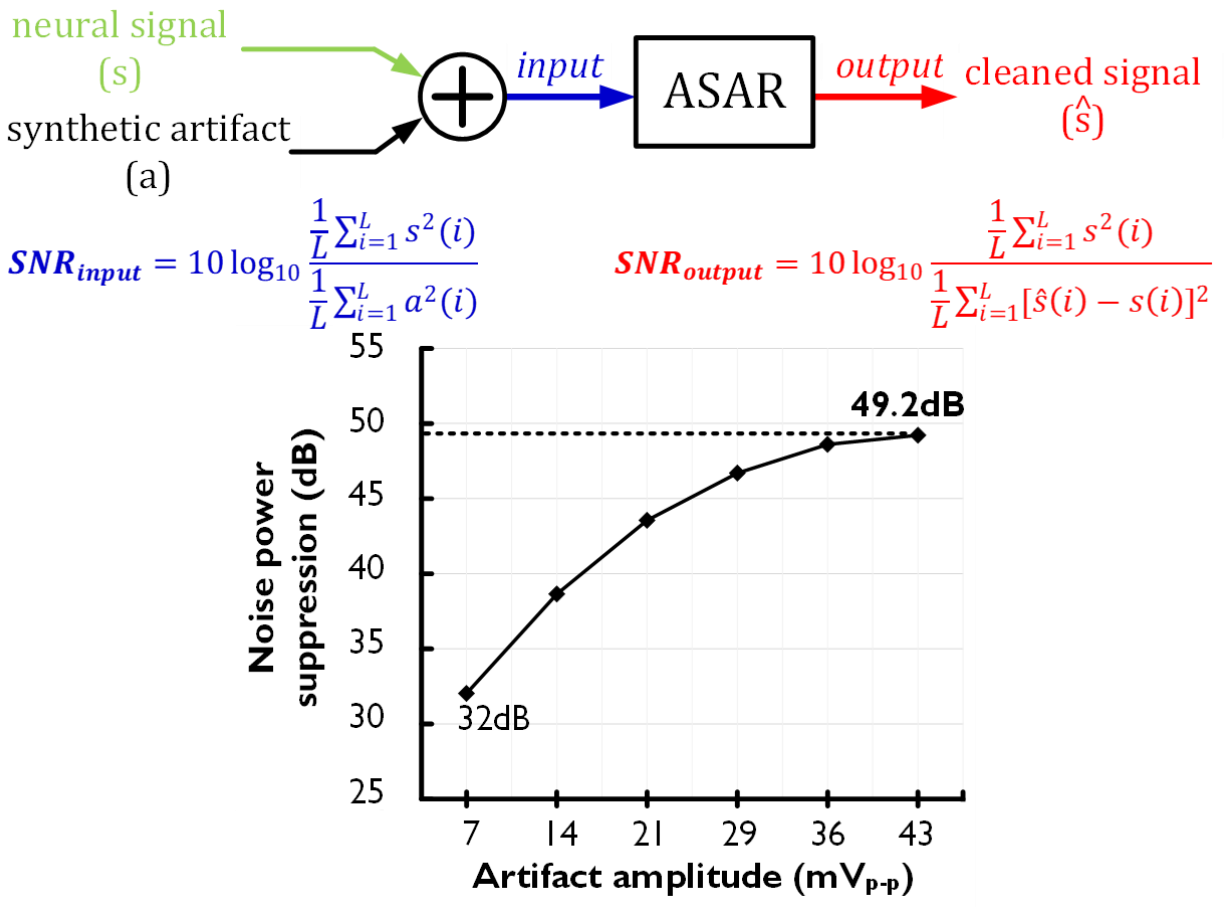

Figure 5.7: Noise (Artifact) Power Suppression

All the chip measurements were done in real-time and both ASAR designs achieve real-time convergence of <42μs for Spike ASAR and <167μs for LFP ASAR, making them suitable for closed-loop neuromodulation systems.

In conclusion of this chapter, two comparisons are provided: (a) a qualitative comparison of various stimulation artifact rejection methods (only implantable methods were considered) in Table 5.3 (green is desirable), and (b) comparison of ASAR with state-of-the-art in Table 5.4.

Table 5.3: Comparison of Various Implantable Stimulation Artifact Rejection Methods

Reference	**[29]**	**[34]**	**[10] , [36]**	**This work**
Method	**Blanking**	**Self-canceling**	**Adaptive Filtering**	**Adaptive Filtering**
Enable recording during stimulation	No	Yes	Yes	Yes
Handle in-band artifacts	Yes	No	Yes	Yes
Application specific	No	Yes	No	No
Agnostic to recording delay	Yes	Yes	No	Yes
Tolerate large artifact amplitudes	Yes	Yes	No	Yes
Need prior knowledge of stimulation pulse	No	No	Yes	No

Blanking methods are the most commonly used methods for stimulation artifact rejection, however, they do not deliver neural recording during stimulation, and depending on the implementation, could have a long recovery time after the stimulation has ended. The self-cancelling method presented in [34] can only handle out-of-band artifacts, while being very application specific and requiring large off-chip components. The adaptive filtering methods presented in [10] **,** [36] enable recording during stimulation and handle in-band artifact; however, they do not tolerate large artifacts, are not agnostic to recording delays, and need knowledge of the stimulation pulse's characteristics. ASAR overcomes all the shortcomings of the prior methods, by enabling recording during stimulation and handling in-band artifacts of much larger amplitude, and without needing any prior knowledge of the stimulation pulse and/or recording timing information.

Table 5.4 compares ASAR to the state-of-the-art. Both ASAR designs are penalized due to limited IO pads available, forcing them to run at much higher clock frequencies, and include Serial-to-Parallel and Parallel-to-Serial circuits; this results in a small area and a large dynamic power overhead and limits our ability to scale the supply voltage below 644mV. At the same time ASAR achieves much higher performance in attenuation, and noise (artifact) suppression, while achieving a significantly faster convergence speeds (>17000x faster). ASAR tolerates much higher amplitudes of artifact, and with the introduction and implantation of template detection method, it can operate "blindly", requiring no prior knowledge of stimulation pulse's structural or temporal characteristics.

Table 5.4: Comparison of ASAR with State-of-the-art

		JSSC'16 [10]	This work	
			LFP ASAR	Spike ASAR
Technology (nm)		180	40	
Area (mm^2)		0.17	0.197	0.209
Power (µW)		0.33	1.73 [a,b]	3.02 [a,b]
BW (Hz)		1-2K	1-200	1-5K
Signals of interest		ECoG	**LFPs**	**Spikes + LFPs**
Stimulation Artifact Rejection	**Architecture**	Mixed-signal Feedback	**Digital Feedforward**	
	Adaptive filter	Signed LMS, 8-tap, fixed step size	**Full LMS, 16-tap, adaptive step size**	
	Sampling rate (KS/s)	4	**6**	**24**
	Attenuation (dB)	24 [c]	**up to 37**	**up to 40**
	Noise power suppression (dB)	-	**up to 49** [d]	
	Operating Frequency (KHz)	4	96 (=16·6) [a]	288 (=12·24) [a]
	Tolerable amplitude (mV_{p-p})	<10	**<100**	**<40**
	Convergence time (µs)	>3,000,000	**<167**	**<42**
	Prior knowledge of stimulation	Yes	**No**	

[a] due to limited IO pads, Serial-to-Parallel and Parallel-to-Serial circuits were added and ASAR input and outputs were serialized. This resulted in operating frequencies 12x and 16x higher than the respective sampling frequency of each design, and higher power.

[b] power calculations assume stimulations to be ON for 5% of the total duration of the test.

[c] artifact attenuation of up to 42dB is based on the resolution of the 8-bit DAC and not a true measure of the filter attenuation (reported measured attenuation of 24dB was chosen for comparison).

[d] emulated by adding varying synthetic artifacts to clinical human patient neural signal.

CHAPTER 6

ASAR Integration:
Sensing IC and Neuromodulation Unit (NM)

The next natural step towards an implantable, closed-loop, neuromodulation system, is the integration of the ASAR engine in a sensing IC, and eventually an implantable neuromodulation unit (NM). As part of the DARPA SUBNETS programs, this opportunity was provided.

In this chapter, we showcase the sensing IC integrating LFP ASAR with the state-of-the-art LFP front-end demonstrated in [18], [19]. This sensing IC is then integrated in our miniaturized neuromodulation unit (NM) along with stimulation IC, and power management units illustrated in [44]. Finally, we briefly review the complete systems containing multiple NM modules, a Neural Hub (NH) and a Control Module (CM), which add more best-in-class technology, such as wireless data transfer in [47], to achieve a complete, 256-channel, implantable, and closed-loop neuromodulation system.

6.1. Sensing Chip

A 64-channel sensing solution was created by integrating the VCO front-end technology (LFP front-end) in [18], [19], with ASAR. This work was done in collaboration with Wenlong Jiang, Vahagn Hokhikyan, and Hariprasad Chandrakumar.

The schematic for the sensing chip is shown in Figure 6.1.

An on-chip clock generation schemed was designed to generate the main system clock of 12MHz using an external crystal; on-chip clock generation allows for major savings in power and area when compared to external crystal oscillators.

Various analog and digital supply voltages were required to power up different blocks in this chip. These voltages were generated and regulated on chip from a source of 1.8V supply, suing multiple LDOs.

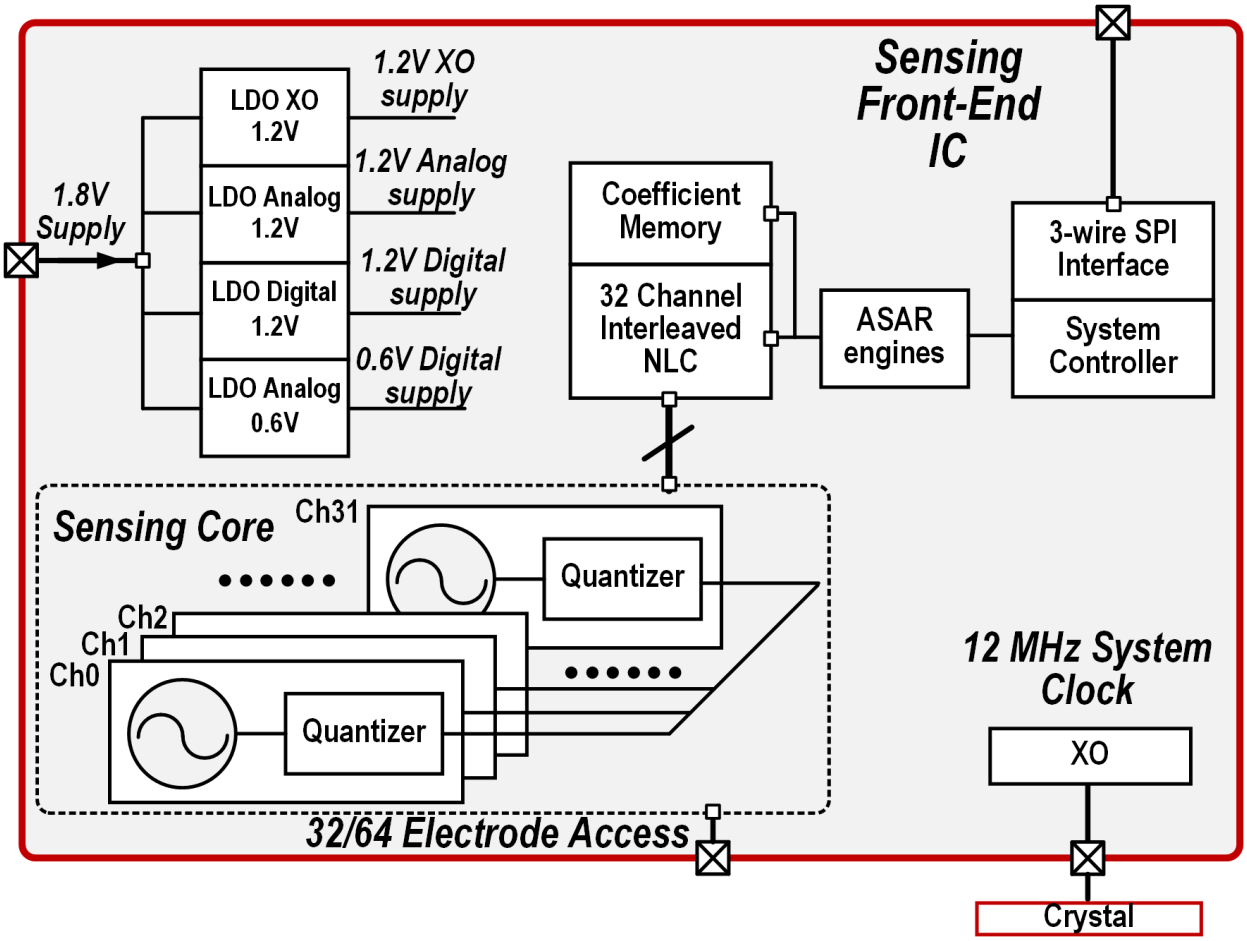

Figure 6.1: Sensing Chip Schematic

32 VCO front-ends were implemented in the sensing core. These cores provide the capability for 32-channel single-ended recording or 64-channel differential recording. Due to the non-linear nature of the VCO front-end outputs, Non-Linearity Correction (NLC) is an integral component in this system. 32 (interleaved) NLC channel are included in the system at the output of the front-end.

At the output of the NLC, reside 4 ASAR engines that can be configured to reject artifacts on any of the 32-channels.

A custom 3-wire SPI communication scheme is also designed which can packetize recording and other information in one of many different forms and communicate with the Stim IC as well as the aggregator module (i.e., Neural Hub (NH)). More details on the implementation of this SPI interface can be found in [48].

Lastly, a system controller block is designed which controls the communication and interactions between various blocks.

It is important to note that the output of ASAR blocks poses a much lower dynamic range than the front-end recordings, because of the removal of the stimulation artifacts, and do not require sampling rates higher than the Nyquist rate of the signals of interest. This is to say, that the output of the ASAR can easily loose a few bits and be down sampled to significantly reduce the data rate. This is done by addition of decimation filters in the digital control section of this chip (not shown in Figure 6.1). The reduced data rate can ease the constraint in wireless transfer blocks and any other signal processing that may be added down the stream in the future.

The sensing IC was designed with consideration for testability of individual blocks, multiple blocks together, as well as the whole system. The sensing IC block diagram shown in Figure 6.2 illustrates this; bypass multiplexers and signals are provided at the input of NLC, ASAR, and decimation filter blocks. This provides the ability to skip one or multiple of this blocks during tests. Additionally, by bypassing earlier blocks, external data can be directly inputted into NLC, ASAR, and decimation blocks to test them individually.

A huge power penalty would be incurred if all the blocks in the sensing chip used the main 12MHz system clock. To avoid this, sub-clocks appropriate for individual blocks were generated internally as shown in Figure 6.2. The ASAR cores operate at 6kHz.

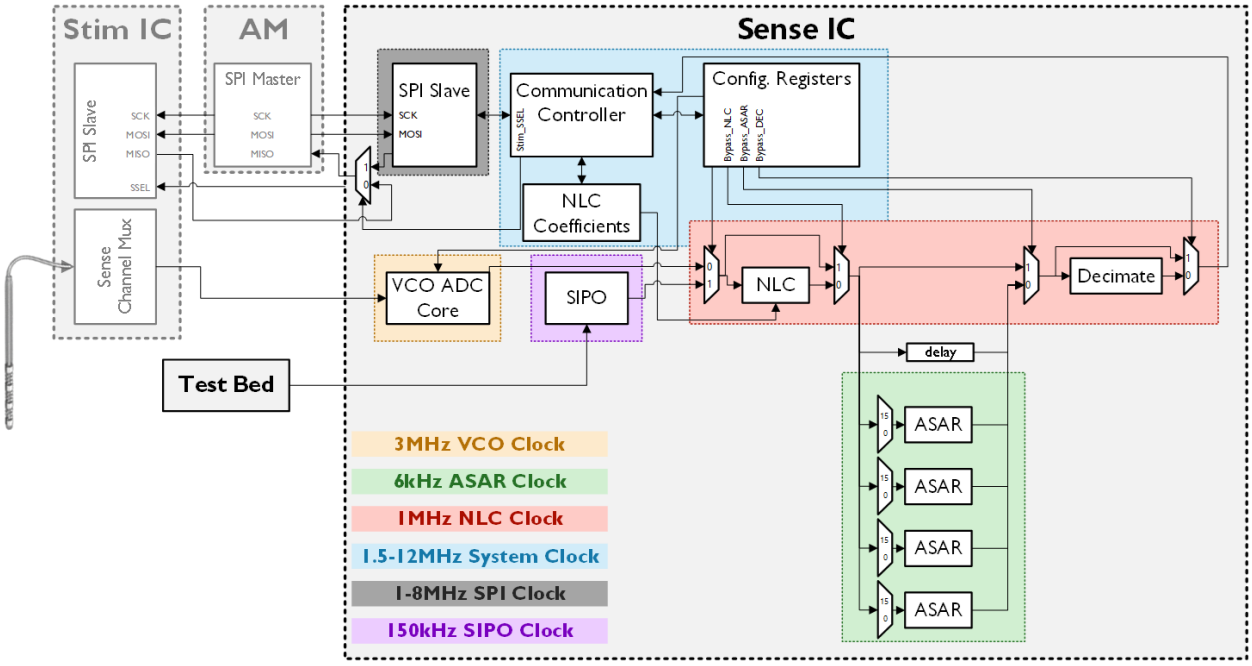

Figure 6.2: Sensing Chip Block Diagram

The sensing chip was fabricated in 40nm CMOS technology. Sensing chip's dimensions are 2637µm by 4459.5µm, occupying a total area of 11.76mm^2.

Figure 6.3 shows the sensing chip micrograph. In this figure the two clusters of 16 VCO front-end cores are shown in blue, resulting in a total of 32 VCO cores. All the digital circuitry including the NLC, ASAR engines, decimator filters and the SPI interface are shown in red at the center and the right side of the image. Other circuitry, including the LDOs, Oscillator, and Power-On-Reset (PoR) are on the top right side of the image, shown in purple.

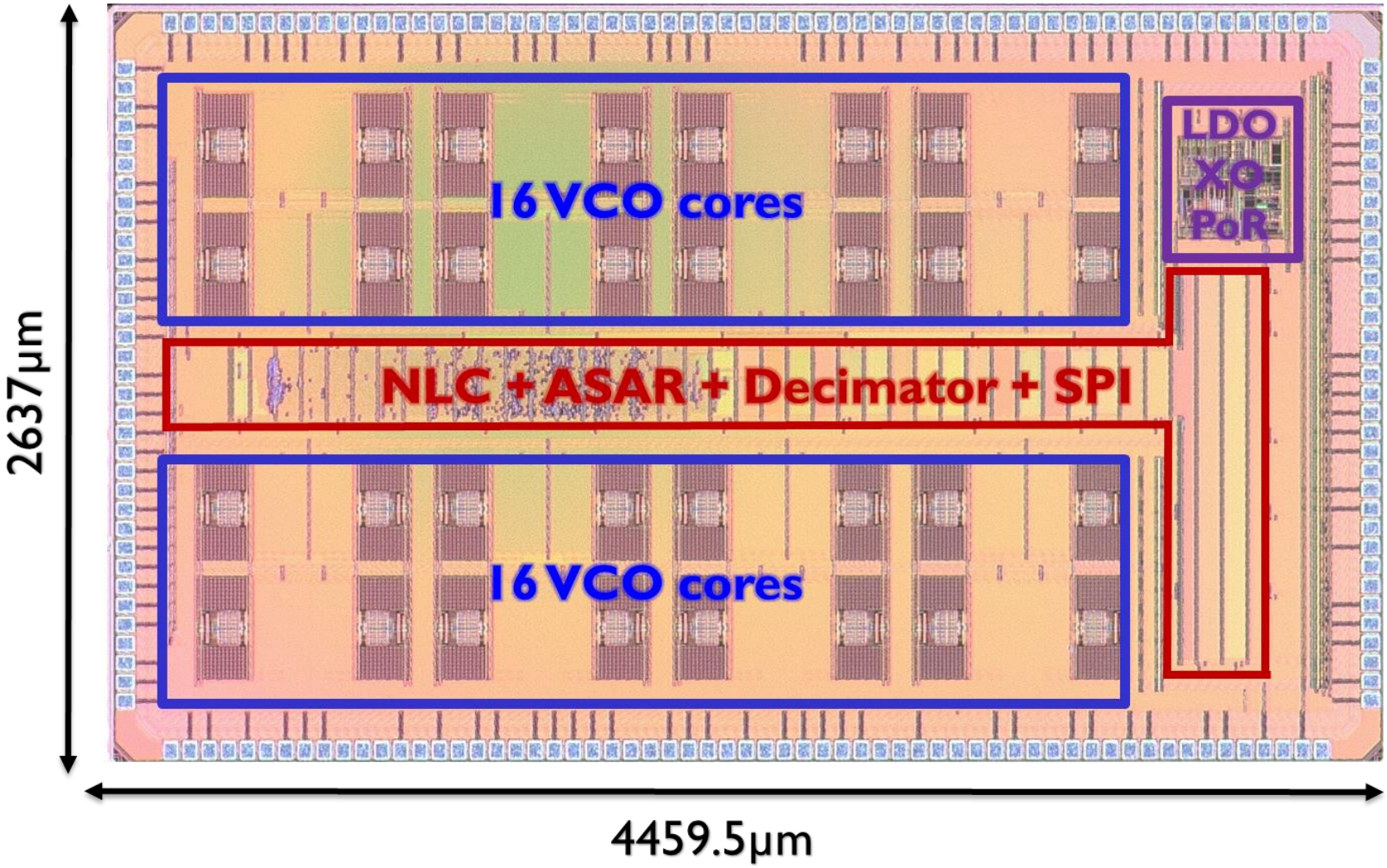

Figure 6.3: Sensing Chip Micrograph

6.1.1. Test Setup and Measurements

In order to test the sensing chip, a field programmable gate array (FPGA) board from Numato Lab called Saturn [49] was used. This FPGA board, hosting our custom test firmware on its Spartan 6 FPGA, is the link between the PCB test board and the test PC terminal. Figure 6.4 shows the complete test setup. Temperature chamber is used for testing of the VCO front-ends.

National Instrument's PXI platform [50] and dynamic signal generator card is a high precision instrument that is used to feed the input data for testing in this setup.

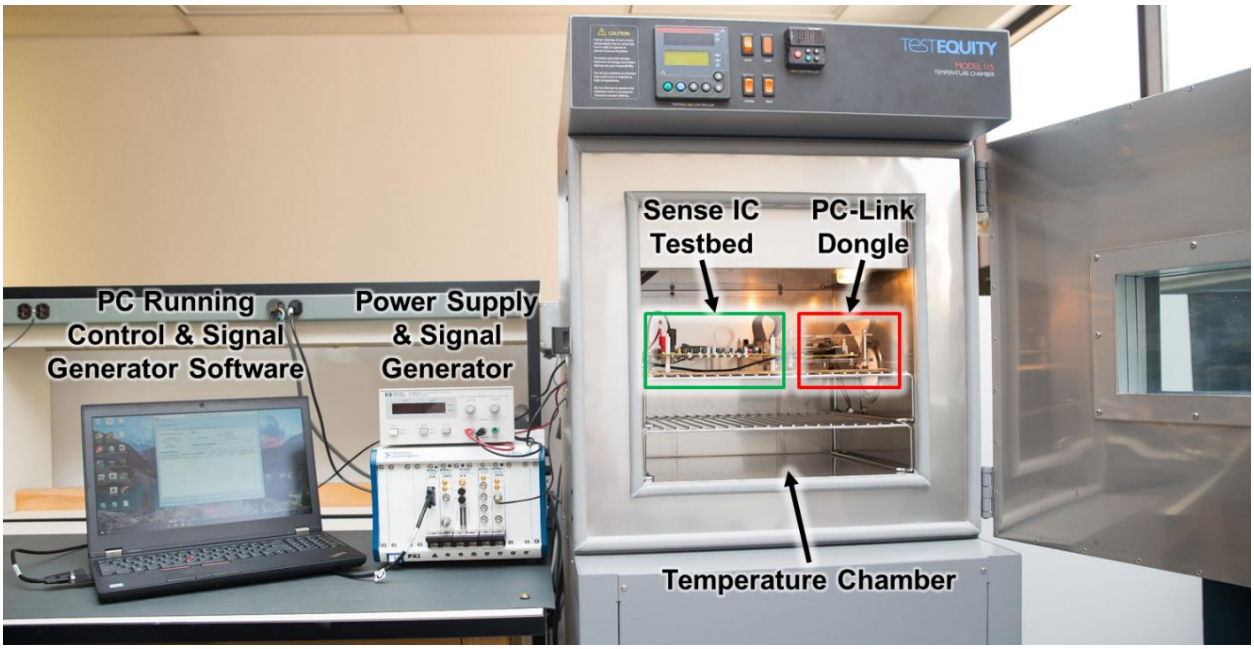

Figure 6.4: Sensing Chip Test Setup

The PC hosts a custom GUI, designed by Vahagn Hokhikyan, which enables us to program different parameters for various tests, easily. A sample screenshot of this GUI is provided in Figure 6.5. In this screen various configuration settings for sensing chip testing are set, and the channel configuration can be selected.

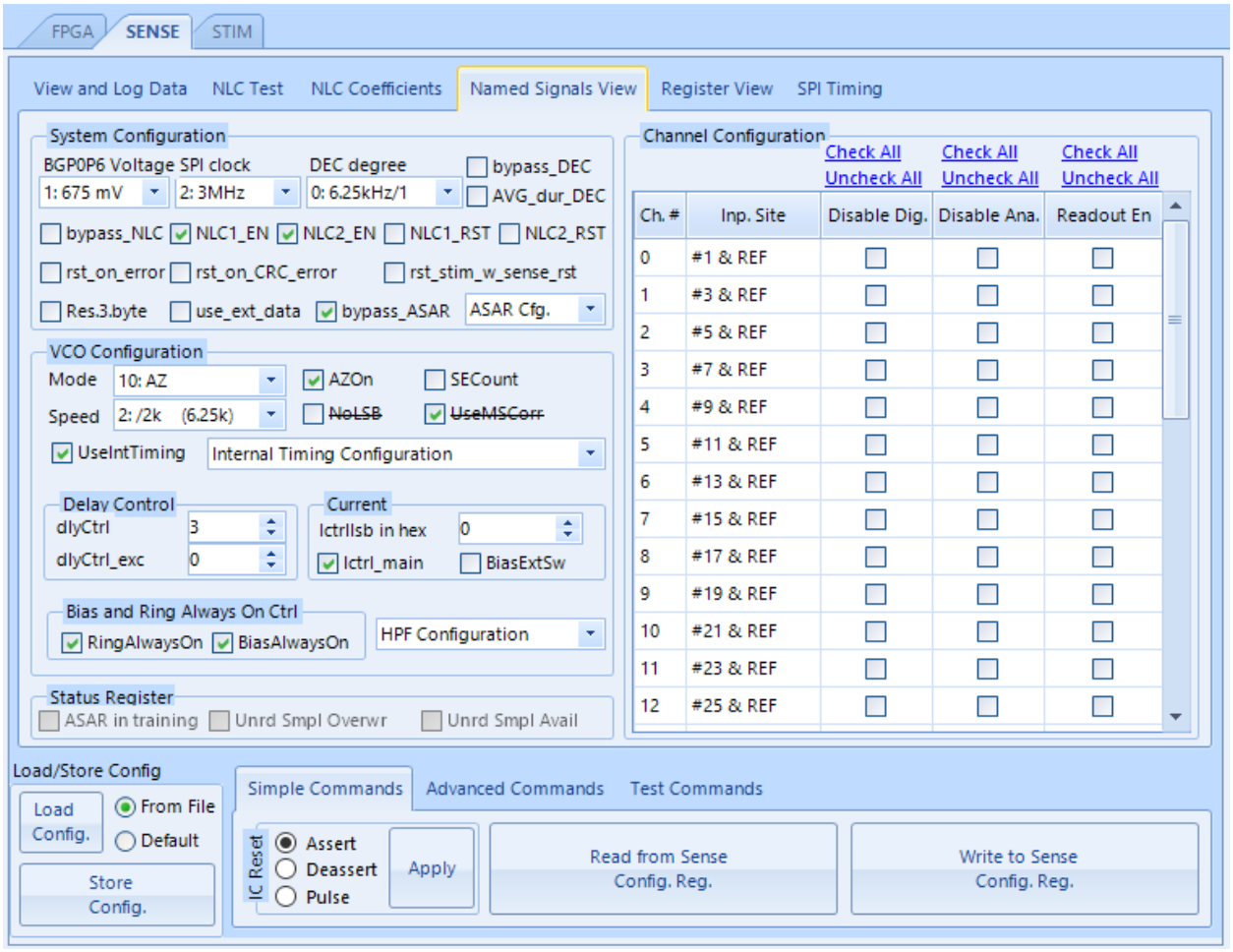

Figure 6.5: Sensing Chip Test GUI – Sense Configuration

ASAR along with other blocks, were individually tested using this platform, and their functionality was confirmed. For example, an ASAR chip measurement matching the data set used in Section 5.3 (LFP ASAR) is shown in Figure 6.6.

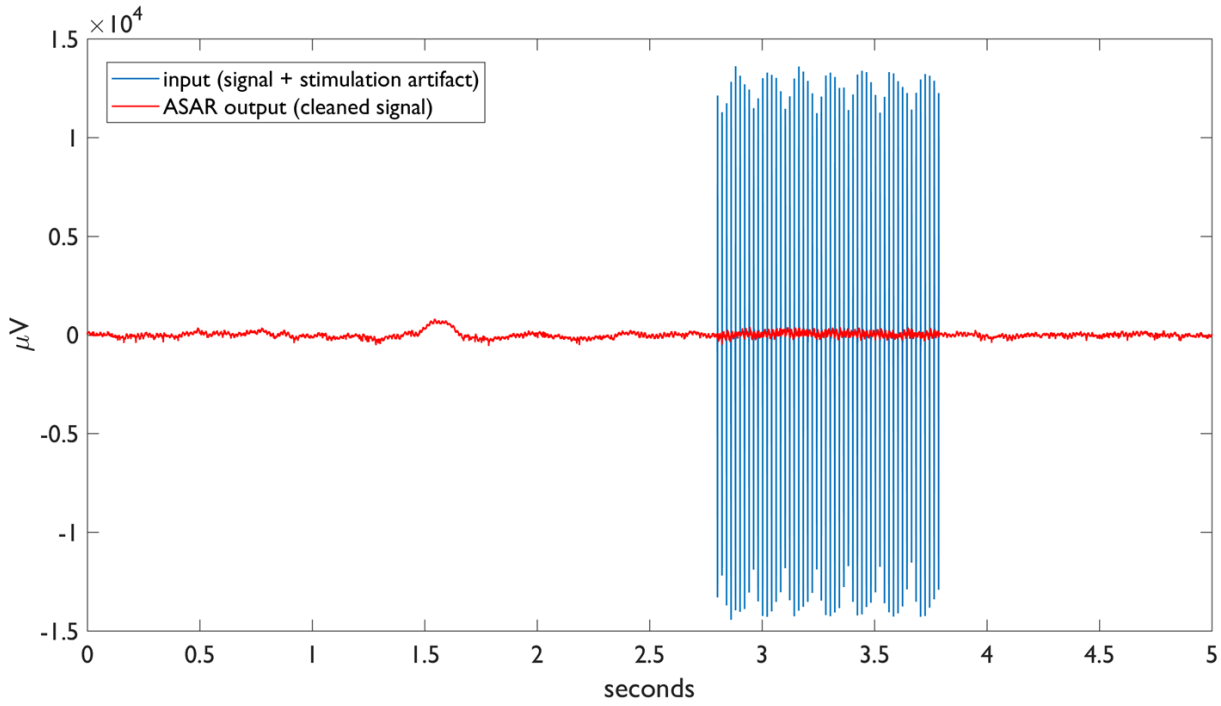

Figure 6.6: Sensing Chip Measurement – ASAR Only

Many other detailed measurements of the whole sensing system and front-end are included in [19], and [51]; however, it is important that ASAR's functionality is verified alongside all other components in the sensing chip. A final verification was done to ensure sensing chip is functioning as intended. The measurement shown in Figure 6.7 is an example of sensing chip operation with all main components (VCO front-end + NLC + ASAR) activated. This is a real-time, ~85 seconds measurement, with varying stimulation pulses. The stimulation current ranges between 0.3mA and 0.9mA, corresponding into recorded artifacts with amplitudes ranging from $20mV_{p\text{-}p}$ ~ $60mV_{p\text{-}p}$.

This sensing chip provides complete 32/64-channel sensing solutions for LFP signals. Next section, briefly overviews the integration of the sensing chip within the miniaturized neuromodulation unit (NM), and neuromodulation system.

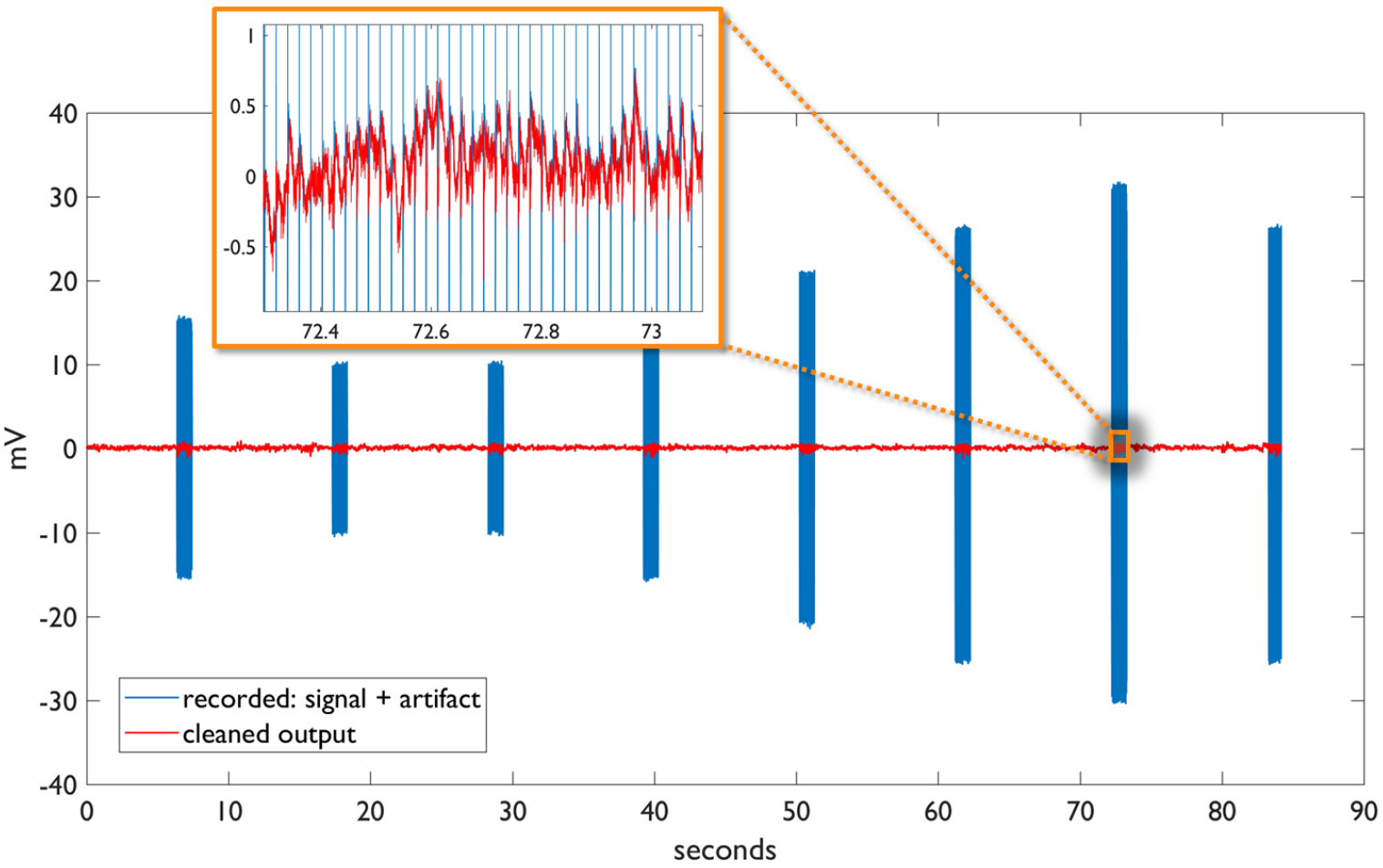

Figure 6.7: Sensing Chip Measurement – VCO front-end + NLC + ASAR

6.2. Neuromodulation (NM) Unit

Figure 6.8 illustrated the SUBNETS neuromodulation system. The Systems-Based Neurotechnology for Emerging Therapies (SUBNETS) program was created and funded by DARPA, and required interdisciplinary research and collaboration of experts from psychiatry, neurosurgery, neural engineering, microelectronics, neuroscience, statistics and computational modeling [52]. Our research team was responsible for the development of the neuromodulation units (NM) in this system, which is presented in this section. This has been a collaborative effort with Dejan Rozgic, Wenlong Jiang, Vahagn Hokhikyan, Hariprasad Chandrakumar and Wenhao Yu.

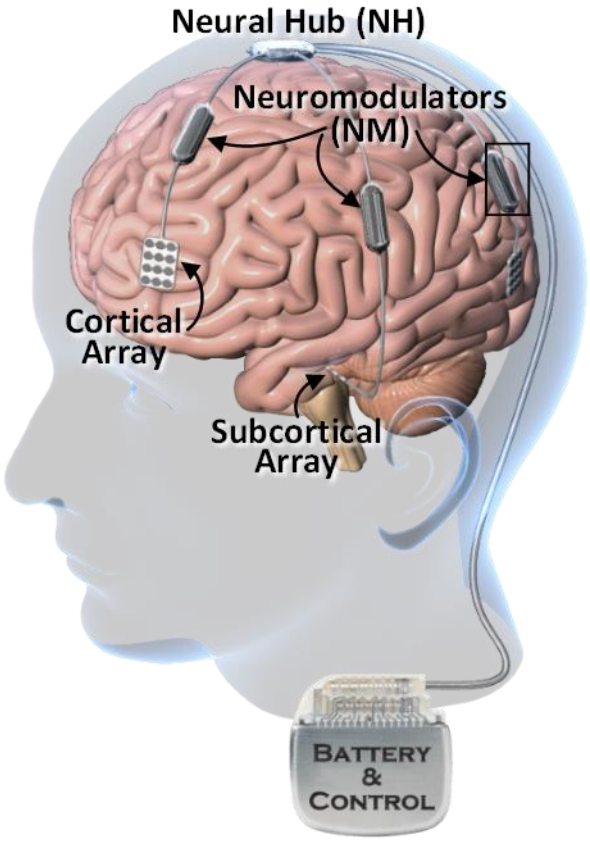

Figure 6.8: SUBNETS Neuromodulation System

In addition to the NM, this system includes support for high-precision, high-channel-count cortical and subcortical electrode arrays; provides a Neural Hub (NH) for aggregation of data from up to 4 NM modules, providing up to 256-channel recording capability, as well as battery and control module that hosts the wireless data transfer technology [47], amongst other components.

The NM module, in particular, hosts our 64-channel sensing chip presented in Section 6.1 with the ASAR engines integrated. In addition, state-of-the-art stimulation and power management technologies [9] have been included, providing support for high-channel-count stimulation, various stimulation modalities, and different power delivery options.

The NM PCB was designed to house the sense and stim ICs, as well as few passive components to support these circuits. Figure 6.9 shows the neuromodulation unit (NM) PCB design. This PCB's dimensions are 22.5mm by 4.5mm and when sealed in the implantable NM capsule occupies 552mm^3 of volume. The inner volume of this capsule, where active electronics are placed is 338mm^3. This unit includes stacked integrated capacitors to further downsize the overall NM module. SPI interface connections as well as connections for 66-contact neural electrodes are provided through the bottom side of this PCB.

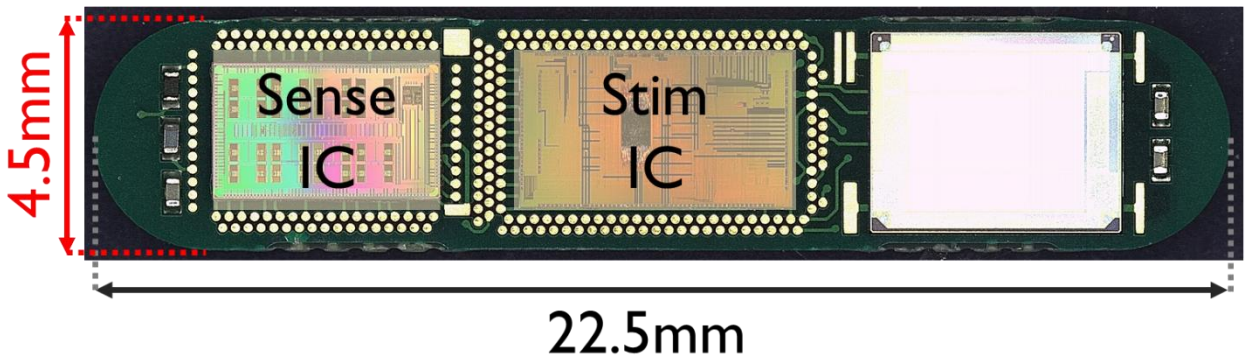

Figure 6.9: Neuromodulation Unit (NM) PCB

This NM unit was used to obtain the concurrent sinusoid (mimicking neural signal) and stimulation recording that was used earlier in Section 3.3.1, to verify ASAR algorithm's functionality. There are more tests ongoing on this unit, and we hope to be able to demonstrate the ASAR functionality in in-vivo measurements soon.

In summary, when compared to state-of-the-art neural interfaces in [14], [15], [16], and [53], our miniaturized neuromodulation unit (NM), is a superior solution that provides the following capabilities:

- Support for both cortical/sub-cortical applications
- Fully implantable 64-channel recoding and stimulation solution, operating in real-time
- High-dynamic range front-end with linear input range of $100mV_{p\text{-}p}$

Most importantly, with integration of ASAR within the sensing chip, this is the only implantable, high-channel-count neuromodulation unit that is capable of adaptively rejecting stimulation artifacts, in real-time.

CHAPTER 7

Conclusion

7.1. Research Contribution

This work demonstrated an energy-efficient, implantable, real-time Adaptive Stimulation Artifact Rejection (ASAR) engine, capable of adaptively removing stimulation artifacts for varying stimulation characteristics at multiple sites, while meeting the challenges of modern implantable, closed-loop, neuromodulation systems.

A blind artifact template detection technique was introduced, which in combination with the proposed ASAR algorithm, eliminated the need for any prior knowledge of the temporal and structural characteristics of the stimulation pulse. To the best of our knowledge, this is the first *blind* stimulation artifact rejection solution.

Two ASAR designs, LFP ASAR and Spike ASAR, have been implemented in 40nm CMOS technology. LFP ASAR achieves convergence of <167μs, occupies an area of 0.197mm^2, and consumes 1.73μW at 0.644V supply. Spike ASAR, achieves convergence of <42μs while occupying an area of 0.209mm^2 and consuming 3.02μW at 0.644V.

A novel method for measurement of noise (artifact) power suppression was presented for adaptive stimulation artifact rejection applications; using this measurement, it was verified that ASAR can attenuate artifacts up to 100mV$_{p\text{-}p}$ by 49.2dB, without any prior knowledge of the stimulation pulse.

Furthermore, integration of ASAR within a 64-channel sensing chip, and an implantable neuromodulation unit (NM) was presented.

Realization of ASAR is a significant contribution towards enabling concurrent neural stimulation and recording for state-of-the-art closed-loop neuromodulation systems, and can aid in maximizing usable neural data, and minimizing time delays for closed-loop response decision making. Most importantly, this can assist physicians and scientists in the investigation of

instantaneous neural response to stimulation, which could lead to new discoveries or therapeutic remedies.

7.2. Future Work

Areas of future work can be divided in two categories:

- Further integration and tests using the existing technology:
    - In the future, the spike ASAR design, can be integrated into a complete sensing chip solution, capable of recording both LFP and neural spikes.
    - In-vivo measurements with the miniaturized 64-channel neuromodulation unit (NM), would be the next logical step to fully showcase the capability of the LFP ASAR.
- Development of additional energy-efficient DSP blocks to enable closed-loop algorithms for therapy of various neurological disorders: with a complete neuromodulation platform in hand, focus can be shifted in DSP blocks required for extraction of biomarkers, and energy-efficient implementation of algorithms responsible for therapeutic decision makings. This is a particularly vast field and a challenging task; the opportunities, however, are endless.

APPENDIX A

Digital Filters for LFP and Neural Spikes

In Chapter 6, integration of the LFP ASAR design within a 64-channel sensing chip and consequently a miniaturized neuromodulation unit was discussed. However, and additional ASAR design, Spike ASAR, was presented in Chapter 5. A valuable future work would be to integrate the Spike ASAR with a state-of-the-art front-end, capable of recording both LFP and Spike signals with high-dynamic range. A suitable front-end is presented in [24].

However, to realize such sensing IC, additional digital filters are required to separate the LFP and Spike signals after the rejection of stimulation artifacts. Indeed, after consulting with and receiving feedback form our neurosurgery and neuroscience colleagues, requirements for these filters have been specified (summarized in Table A.1).

Table A.1: Digital Filter Requirements

Requirement	**LPF Filter**	**Spike Filter**
Passband corner frequency	130Hz	200Hz
Stop band frequency	400Hz	130Hz
Stop band attenuation	40dB	>80dB
Decimation factor	4	1
Input sampling factor	24KHz	24KHz
Filter phase response	**Linear phase is required**	**Linear phase is NOT required**

The analysis, and design of filters presented in this appendix was done in collaboration with Zoltan Romocsa, where he completed the verilog code as part of his MS project [54].

A.1. Digital Filters for LFP Band

There are post-processing algorithm in neuroscience (e.g., Phase-Amplitude Coupling (PAC) [55]), that require phase of the LFP signals to remain linear. This necessitates an FIR implementation for LFP band. Computational cost of various FIR filters for LFPs were compared and are shown in Figure A.1.

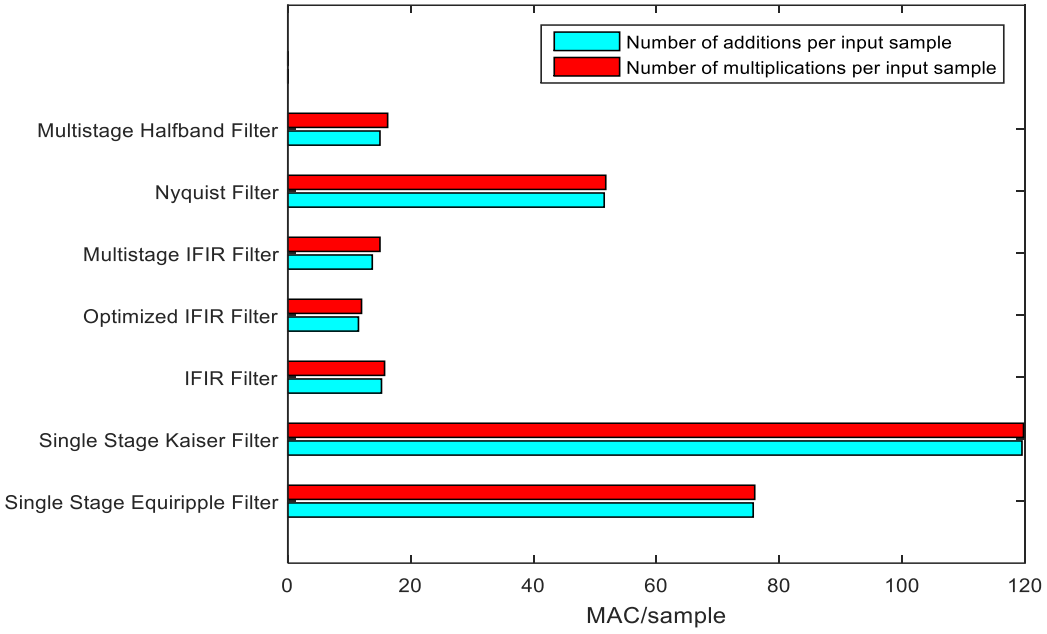

Figure A.1: Computational complexity of evaluated FIR filters

The optimized IFIR filter was chosen for implementation. Interpolating FIR filters, not only filter the input signal, but also reduce its sampling rate, which is ideal in our application after the removal of stimulations artifact. Although implementation details are included in [54], a block diagram of this optimized interpolating FIR filter is shown in Figure A.2.

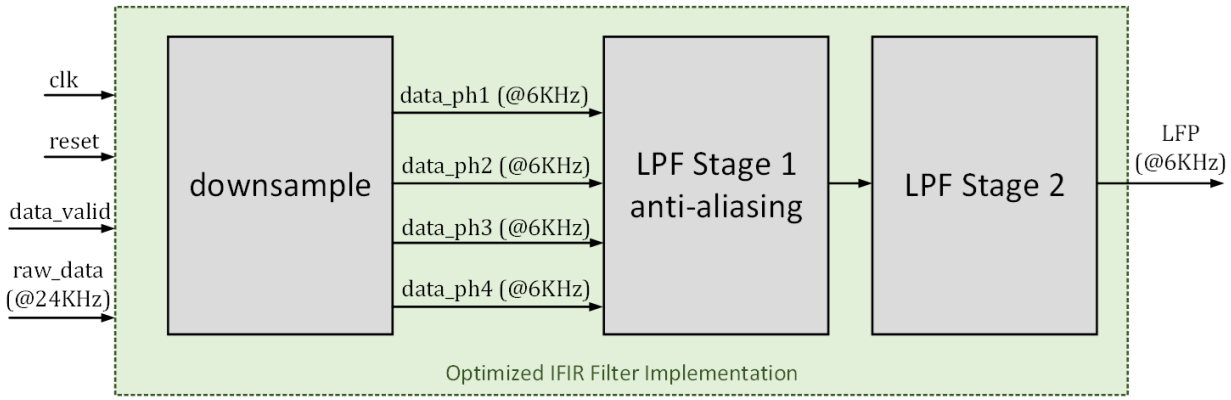

Figure A.2: Optimized IFIR Filter Implementation Block Diagram

A.2. Digital Filters for Neural Spike Band

For the design of these filter, we explored IIR filters, as there were no requirements on signal phase, and these filters are less expensive in hardware. Computational cost of candidate IIR filters for Spikes were compared and are shown in Figure A.3.

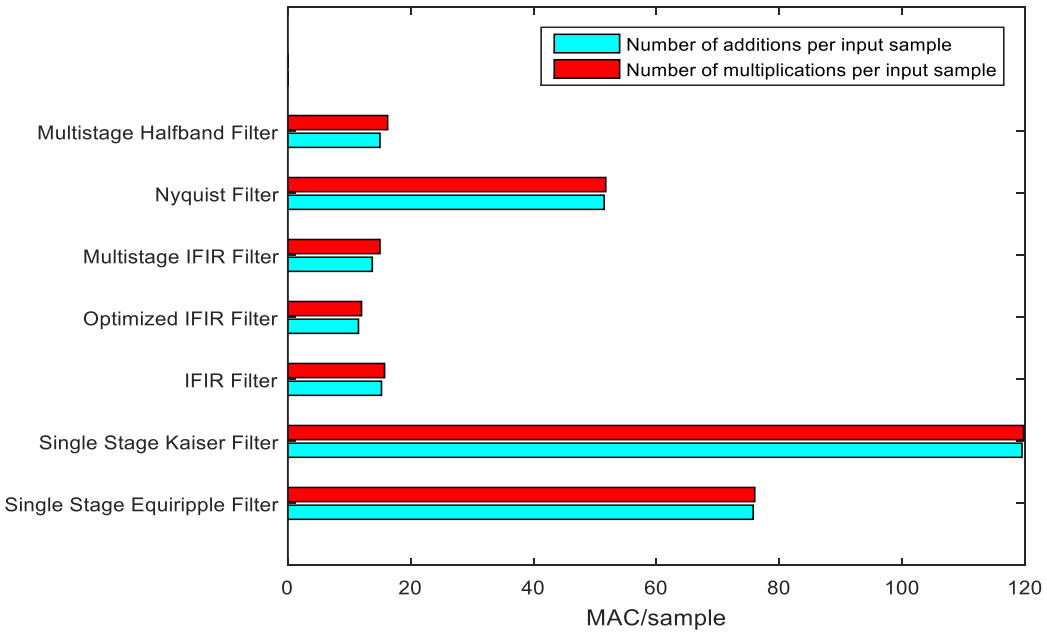

Figure A.3: Computational complexity of evaluated IIR filters

Here, the Chebyshev Type2 IIR filter was chosen. In summary, as future work objective, the Spike ASAR design presented in Chapter 5 can be integrated with a suitable Spike recording front-end, to create an state-of-art neural recording solution, capable of recording both LFP signals and neural spike activity, concurrently with stimulation and the ability to reject any stimulation artifact in real-time.

REFERENCES

[1] World Health Organization (WHO) , "National Institute of Mental Health," NIH, [Online]. Available: https://www.nimh.nih.gov/health/statistics/disability/us-leading-categories-of-diseases-disorders.shtml. [Accessed 15 08 2017].

[2] J. Jacobs, M. Kahana, A. Ekstrom and I. Fried, "Brain Oscillations Control Timing of Single-Neuron Activity in Humans," *The Journal of Neuroscience,* vol. 27, no. 14, pp. 3839-3844, 4 April 2007.

[3] U. Rutishauser, I. Ross, A. Mamelak and E. Schuman, "Human memory strength is predicted by theta-frequency phase-locking of single neurons," *Nature,* no. 464, pp. 903-907, 2010.

[4] A. Schwartz, X. Cui, D. Weber and D. Moran, "Brain controlled interfaces: Movement restoration with neural prosthetics," *Neuron,* vol. 52, pp. 205-220, 2006.

[5] J. Mavoori, A. Jackson, C. Diorio and E. Fetz, "An autonomous implantable computer for neural recording and simulation in unrestrained primates," *J. Neurosci. Meth.,* vol. 148, pp. 71-77, 2005.

[6] Medtronic, Inc., "[Press Release] New Research Shows Significant Benefits of Medtronic Deep Brain Stimulation Earlier in Progression of Parkinson's Disease.," Medtronic, Inc., 2013. [Online]. Available: http://newsroom.medtronic.com/phoenix.zhtml?c=251324&p=irol-newsArticle&ID=1802686&highlight=. [Accessed 16 09 2017].

[7] W. Chen and et. al., "A fully integrated 8-channel closed-loop neural-prosthetic SoC for real-time epileptic seizure control," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, San Francisco, 2013.

[8] N. Suthana and et. al., "Memory Enhancement and Deep-Brain Stimulation of the Entorhinal Area," *The New England Journal of Medicine*, pp. 502-510, Feb. 2012.

[9] D. Rozgić, "Flexible and Energy-Efficient Circuits for Implantable Biomedical Systems," PhD Thesis, Electrical Engineering, UCLA, Los Angeles, 2017.

[10] A. Mendrela and et. al., "A Bidirectional Neural Interface Circuit with Active Stimulation Artifact Cancellation and Cross-Channel Common-Mode Noise Suppression," *IEEE Journal of Solid-State Circuits (JSSC)*, pp. 955-965, Apr. 2016.

[11] R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 38, no. 5, pp. 958-965, Jun 2003.

[12] W. Wattanapanitch, M. Fee and R. Sarpeshkar, "An Energy-Efficient Micropower Neural Recording Amplifier," *IEEE Transactions on Biomedical Circuits and Systems (TBioCAS)*, vol. 1, no. 2, pp. 136-147, Jun 2007.

[13] R. R. Harrison, "A Versatile Integrated Circuit for the Acquisition of Biopotentials," in *Custom Integrated Circuits Conference (CICC)*, 2007.

[14] Y.-K. Lo and et al., "A 176-Channel 0.5cm^3 0.7g Wireless Implant for Motor Function Recovery after Spinal Cord Injury," in *IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, 2016.

[15] Y. P. Lin and et al., "A Battery-Less, Implantable Neuro-Electronic Interface for Studying the Mechanisms of Deep Brain Stimulation in Rat Models," *IEEE Transactions on Biomedical Circuits and Systems (TBioCAS),* vol. 10, no. 1, pp. 98-112, Feb 2016.

[16] P. Cong and et al., "A 32-channel modular bi-directional neural interface system with embedded DSP for closed-loop operation," in *40th European Solid State Circuits Conference (ESSCIRC),* Venice Lido, 2014.

[17] R. Muller and et al., "A miniaturized 64-channel 225μW wireless electrocorticographic neural sensor," in *IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, 2014.

[18] W. Jiang, V. Hokhikyan, H. Chandrakumar, V. Karkare and D. Marković, "A ±50mVp Linear Input-range, VCO-based Neural Recording Front-end with Digital Nonlinearity Correction," in *IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, 2016.

[19] W. Jiang, V. Hokhikyan, H. Chandrakumar, V. Karkare and D. Marković, "A ± 50-mV Linear-Input-Range VCO-Based Neural-Recording Front-End With Digital Nonlinearity Correction," *IEEE Journal of Solid-State Circuits (JSSC),* vol. 52, no. 1, pp. 173-184, Jan 2017.

[20] H. Chandrakumar and D. Marković, "A 2μW, 40mVpp Linear Input-range, Chopper-stabilized Bio-signal Amplifier with Boosted Input Impedance of 300MΩ and Electrode--offset Filtering," in *IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, 2016.

[21] H. Chandrakumar and D. Marković, "A High Dynamic-Range Neural Recording Chopper Amplifier for Simultaneous Neural Recording and Stimulation," *IEEE Journal of Solid-State Circuits (JSSC),* vol. 52, no. 3, pp. 645-656, Mar 2017.

[22] H. Chandrakumar and D. Marković, "A 2.8µW, 80mVpp Linear-Input-Range, 1.6GΩ Input Impedance, Bio-signal Chopper Amplifier Tolerant to Common-Mode Interferers up to 650mVpp," in *IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, 2017.

[23] H. Chandrakumar and D. Marković, "An 80mVpp Linear-Input-Range, 1.6GΩ Input Impedance, Low-Power Chopper Amplifier for Closed-Loop Neural Recording that is Tolerant to 650mVpp Common-Mode Interference," *IEEE Journal of Solid-State Circuits (JSSC),* vol. 52, no. 11, pp. 2811-2828, Nov 2017.

[24] H. Chandrakumar and D. Marković, "A 15.2b-ENOB Continuous-Time ΔΣ ADC for a 7.3µW 200mVpp-Linear-Input-Range Neural Recording Front-End," in *IEEE International Solid-State Circuits Conference (ISSCC),* San Francisco, 2018.

[25] K. Najafi, "Solid-state microsensors for cortical nerve recordings," *IEEE Engineering in Medicine and Biology Magazine,* vol. 13, no. 3, pp. 375-387, 1994.

[26] T. Denison, K. Consoer, W. Santa, A. Avestruz, J. Cooley and A. Kelly, "A 2 µW 100nV/rtHz Chopper-Stabilized Instrumentation Amplifier for Chronic Measurement of Neural Field Potentials," *IEEE Journal of Solid-State Circuits (JSSC),* vol. 42, no. 12, pp. 2934-2945, Dec 2007.

[27] D. Rozgić and D. Marković, "A 0.78mW/cm^2 Autonomous Thermoelectric Energy-Harvester for Biomedical Applications," in *IEEE Symp. VLSI Circuits (VLSI)*, Kyoto, Japan, 2015.

[28] V. Karkare, H. Chandrakumar, D. Rozgić and D. Marković, "Robust, reconfigurable, and power-efficient biosignal recording systems," in *Custom Integrated Circuits Conference (CICC)*, San Jose, 2014.

[29] E. Brown, J. Ross, R. Blum, N. Yoonkey, B. Wheeler and S. DeWeerth, "Stimulus-Artifact Elimination in a Multi-Electrode System," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 2, no. 1, pp. 10-21, Mar 2008.

[30] H. Gao and et al., "HermesE: A 96-Channel Full Data Rate Direct Neural Interface in 0.13 μm CMOS," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 47, no. 4, pp. 1043-1055, Apr 2012.

[31] W.-M. Chen and et al., "A Fully Integrated 8-Channel Closed-Loop Neural-Prosthetic CMOS SoC for Real Time Epileptic Seizure Control," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 49, no. 1, pp. 232-247, Jan 2014.

[32] D. Wagenaar and S. Potter, "Real-Time Multi-channel Stimulus Artifact Suppression by Local Curve Fitting," *Journal of Neuroscience Methods*, vol. 120, no. 2, pp. 113-120, Oct 2002.

[33] R. Blum, J. Ross, S. Das, E. Brown and S. DeWeerth, "Models of stimulation artifacts applied to integrated circuit design," in *IEEE 26th Annual International Conference of the Engineering in Medicine and Biology Society (IEMBS)*, 2004.

[34] S. Stanslaski and et. al., "Design and Validation of a Fully Implantable, Chronic, Closed-Loop Neuromodulation Device With Concurrent Sensing and Stimulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 4, pp. 410-421, Jul 2012.

[35] J. Gnadt, S. Echols, A. Yildirim, H. Zhang and K. Paul, "Spectral cancellation of microstimulation artifact for simultaneous neural recording in situ," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 10, pp. 1129-1135, Oct 2003.

[36] A. Mendrela and et. al., "Enabling Closed-Loop Neural Interface: A Bi-Directional Interface Circuit With Stimulation Artifact Cancellation and Cross-Channel CM Noise Suppression," in *IEEE Symposium on VLSI Circuits*, Kyoto, Japan, 2015.

[37] D. V. Nowicki, B. Piallat, A. Benabid and T. I. Aksenova, "Artifact Processor for Neuronal Activity Analysis during Deep Brain Stimulation," in *Lecture Notes in Computer Science: Artificial Neural Networks*, Berlin, Heidelberg, Springer, 2008, pp. 508-516.

[38] B. C. Johnson and et al., "An implantable 700μW 64-channel neuromodulation IC for simultaneous recording and stimulation with rapid artifact recovery," in *Symposium on VLSI Circuits (VLSI)*, Kyoto, 2017.

[39] A. Zhou and et al., "WAND: A 128-channel, closed-loop, wireless artifact-free neuromodulation device", arXiv:1708.00556 [q-bio.NC].

[40] A. H. Sayed, Adaptive Filters, NJ: John Wiley & Sons, 2008.

[41] M. Hur, J. Y. Choi, J.-S. Baek and J. S. Seo, "Generalized Normalized Gradient Descent Algorithm Based on Estimated a Posteriori Error," in *10th International Conference on Advanced Communication Technology (ICACT)*, Gangwon-Do, 2008.

[42] J. Larson, D. Wong and G. Lynch, "Patterned stimulation at the theta frequency is optimal for the induction of hippocampal long-term potentiation," *Brain Research,* vol. 368, no. 2, pp. 347-350, 1986.

[43] A. S. Titiz, M. R. H. Hill, E. A. Mankin, Z. M. Aghajan, D. Eliashiv, N. Tchemodanov, U. Maoz, J. Stern, M. E. Tran, P. Schuette, E. Behnke, N. A. Suthana and I. Fried, "Theta-burst microstimulation in the human entorhinal area improves memory specificity," *Human Biology and Medicine, Neuroscience,* 24 10 2014.

[44] D. Rozgić, V. Hokhikyan, W. Jiang, S. Basir-Kazeruni, H. Chandrakumar, W. Leng and D. Marković, "A True Full-Duplex 32-Channel 0.135cm^3 Neural Interface," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Turin, 2017.

[45] CASPER (Collaboration for Astronomy Signal Processing and Electronic Reseach), "ROACH," UC Berkeley, 16 Apr 2013. [Online]. Available: https://casper.berkeley.edu/wiki/ROACH. [Accessed 16 07 2016].

[46] S. Basir-Kazeruni, S. Vlaski, H. Salami, A. Sayed and D. Marković, "A Blind Adaptive Stimulation Artifact Rejection (ASAR) Engine for Closed-Loop Implantable Neuromodulation Systems," in *8th Intl. IEEE EMBS Conference on Neural Engineering (NER),* Shanghai, 2017.

[47] A. Yousefi, D. Yang, A. A. Abidi and D. Marković, "A distance-immune low-power 4-Mbps inductively-coupled bidirectional data link," in *Symposium on VLSI Circuits (VLSI)*, Kyoto, 2017.

[48] V. Hokhikyan, "Design and Verification of a Cloosed-loop-ready High-channel-count Neuromodulation Unit," PhD Thesis, Biomedical Engineering, UCLA, Los Angeles,2017.

[49] "Saturn – Spartan 6 FPGA Development Board With DDR SDRAM," Numato, [Online]. Available: http://numato.com/product/saturn-spartan-6-fpga-development-board-with-ddr-sdram. [Accessed 26 11 2017].

[50] "PXI Platform," National Instrument, 2017. [Online]. Available: http://www.ni.com/pxi/. [Accessed 29 11 2017].

[51] W. Jiang, "Multi-Channel High-Dynamic-Range Implantable VCO-Based Neural-Sensing System," PhD Thesis, Electrical Engineering, UCLA, Los Angeles, 2017.

[52] DARPA, "SUBNETS Aims for Systems-Based Neurotechnology and Understanding for the Treatment of Neuropsychological Illnesses," Defense Advanced Research Projects Agency (DARPA), 25 10 2013. [Online]. Available: https://www.darpa.mil/news-events/2013-10-25. [Accessed 14 08 2017].

[53] R. Muller and et al., "A Minimally Invasive 64-Channel Wireless μECoG Implant," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 50, no. 1, pp. 344-359, Jan 2015.

[54] Z. Romocsa, "VLSI implementation of digital filters for neural spike sorting," ENGR 299 Capstone Project Report, Electrical Engineering, UCLA, Los Angeles, 2015.

[55] D. Dellavale and et al., "Optimization of a Phase-to-Amplitude Coupling Algorithm for Real-Time Processing of Brain Electrical Signals," in *Information & Communication Technology Conferece*, Eindhoven, 2013.