UCLA
UCLA Electronic Theses and Dissertations

Title
Design and Verification of a Closed-loop-ready High-channel-count Neuromodulation Unit

Permalink
https://escholarship.org/uc/item/2hb7h78b

Author
Hokhikyan, Vahagn

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Design and Verification of a Closed-loop-ready High-channel-count Neuromodulation Unit

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biomedical Engineering

by

Vahagn Hokhikyan

2017

© Copyright by

Vahagn Hokhikyan

2017

ABSTRACT OF THE DISSERTATION

Design and Verification of a Closed-loop-ready

High-channel-count Neuromodulation Unit

by

Vahagn Hokhikyan

Doctor of Philosophy in Biomedical Engineering

University of California, Los Angeles, 2017

Professor Dejan Marković, Chair

According to a report by the World Health Organization (WHO), neuropsychiatric disorders affect about one billion people worldwide[1] and are the leading cause of disability in the U.S.[2]. To address this global epidemic, neuroscientific initiatives are being developed globally[3].

Deep brain stimulation (DBS) has been successful alternate treatment modality for some neuropsychiatric disorders (Parkinson's disease, essential tremor, dystonia, and obsessive-compulsive disorder) when traditional treatment options failed (resection, medication, and psychotherapy). According to one long term study targeting Parkinson's disease, patients' motor function and daily activity-scores improve about 50% while off medication and receiving only DBS therapy[4]. This therapy is delivered through large electrodes in the form of fixed-frequency rectangular stimulation pulses in an always-on, open-loop fashion - ignoring the disease state, medication status, or side effects.

In this work, we present a more advanced neural implant which, while being backward compatible with traditional DBS therapy, 1) is capable of sensing neural signals while delivering stimulation, 2) has high-channel-count, and 3) can generate non-rectangular stimulation waveforms. These key features are enabled by our custom-designed neural sensing and stimulation integrated circuits (ICs), which along with a few passive components, a miniature printed circuit board, and our user-friendly graphical interface comprise a capable neuromodulation system.

Our sensing IC's ability to capture neural signals and stimulation artifacts without saturation at implant-level power is unique. Sampled local field potential (LFP) signals can be used to close the knowledge gap about the disease biomarkers and be fed into closed-loop algorithms for automatic tuning of stimulation parameters based on the disease state. The resulting autonomy will reduce or eliminate the need for periodic clinical visits for re-adjusting stimulation parameters to ameliorate neural network's habituation effects.

Our high-channel-count stimulation IC, when paired with high-density probes, will substantially increase the spatial resolution of DBS, which can improve the therapeutic index and potentially result in lower power consumption for achieving the same therapeutic benefits[5]. Also, the ability of our stimulation IC to generate custom, non-rectangular waveforms can lead to increased implant battery life, as some non-rectangular waveforms seem to be more energy efficient[6],[7].

We believe that the proposed system has the potential to improve the quality of patient care and to further our understanding of neuropsychiatric disorders, and we hope that it will soon find an increased use in various clinical and research environments.

The dissertation of Vahagn Hokhikyan is approved.

James Bisley

Nader Pouratian

Nanthia Suthana

Dejan Marković, Committee Chair

University of California, Los Angeles

2017

To the Architect of the Universe,

Who put me in a wonderful world and gave me the curiosity to explore it.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

Many names come to mind that were directly or indirectly involved in supporting the work leading to this dissertation over the last few years. I am grateful to Divine Providence for placing me at UCLA, in Parallel Data Architectures Group led by Prof. Dejan Marković right at the time when the group had a growing interest in advancing the neuroengineering field. This also happened to be the time when the U.S. government deemed important to unlock "*the mystery of the three pounds of matter that sits between our ears*"[8] and announced the BRAIN initiative. The timing and the environment could not have been better for realizing the vision outlined in my university entrance application - designing autonomous implantable medical devices.

I am grateful to Prof. Dejan Marković for taking me under his wings during all these years, for providing me with guidance and financial support, and for giving me the freedom and the unique opportunity to participate in eight mixed signal IC designs. I am indebted to him for the digital circuit design and optimization knowledge that he imparted to me through his two classes – EE216A (Design of VLSI Circuits and Systems) and EE216B (VLSI Signal Processing). If it was not for his encouragement, on occasion or two, I could have quitted the Ph.D. program, and looking back I am thankful that he held me back.

I would like to thank Dr. Nader Pouratian, Dr. James Bisley, and Dr. Nanthia Suthana for their time and effort in serving on my Ph.D. committee and evaluating this work.

At Parallel Architectures Group, I was surrounded by much smarter people, who besides offering their fellowship, helped me in various aspects of this work. Dr. Vaibhav Karkare familiarized me with the digital ASIC flow and allowed me to design the packetization engine and

the wireless transceiver controller for the first version of the neural sensing IC. This neural sensing front end underwent three more iterations. Later, other system level blocks were added for system usability. Wenlong Jiang improved the VCO ADC core design, Sina Basir Kazeruni designed the adaptive stimulation artifact rejection (ASAR) DSP block, Hariprasad Chandrakumar designed the power-on-reset (POR) and crystal oscillator circuits. He also co-designed the low dropout voltage regulators with Weiyu Leng. Our visiting scholars Pinwen Su and Yinghua Hu helped me with the design documentation and testing of the nonlinearity correction (NLC) block. I would like to thank my colleague Dejan Rozgic for entrusting me with the design of digital infrastructure for the neural stimulator IC. For the last revision of this IC, we are grateful to our visiting scholar Dr. Ippei Akita who helped Dejan Rozgic with the analog circuit design. I would also like to thank Sam Barclay for designing NMU test PCB and Wenhao Yu for designing the second version of the NMU PCB.

Our collaborators from Lawrence Livermore National Laboratories helped us with the design and fabrication of the neural probes and implant encapsulation packages. I would like to specifically mention Sat Pannu, Kedar Shah, Vanessa Tolosa, Angela Tooker, Timir Baran Datta-Chaudhuri, and Razi-Ul Muhammad Haque for their contributions to this effort.

I would like to thank our group's administrative assistant – Kyle Jung, for being the person behind the scenes for many of us. To support us, the students, he went an extra mile beyond his responsibilities and many times we learned about those instances after the fact and from others.

Although my family was not directly involved in this work, they did their share of heavy lifting. My parents Albert and Svetik were my main encouragers and supporters for pursuing a

Ph.D. degree before I got married. My engineering bent is probably mostly due to my father, who programmed the first-generation computers that occupied an entire room and took me to his workplace on several occasions. Although an engineer herself, my mother wanted me to become a medical doctor. I hope that a biomedical engineering degree is a good compromise for both of them.

I have two feelings when thinking about my wife and three children during my Ph.D. student years – guilt and gratitude. My wife Anait had to put-off her graduate school plans, spent many lonely evenings when I was in the lab, double her parenting efforts, and among all these circumstances encourage me (thank you, sweetheart! I couldn't have done this without you). Our little wonders – Arpi, Emma, and Albert, never knew how many times they pulled me out of despair just by running up to me and hugging me when I came home at night (love you, my little ones!).

This work was funded by two of the DARPA BRAIN initiative programs – restoring active memory (RAM) and Systems-Based Neurotechnology for Emerging Therapies (SUBNETS)[9]. The views, opinions, and findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Vita

1997-2001 Management, Armenian State University of Economics, Yerevan, Armenia

2004-2008 B.S., Computer Engineering, California State University, Northridge, U.S.A

2008-2010 M.S., Electrical Engineering, California State University, Northridge, U.S.A

2011-2017 Ph.D. Candidate, Biomedical Engineering, University of California, Los Angeles, U.S.A

Publications

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 Proc. IEEE Int. Solid-State Circuits Conference (ISSCC'16), Feb. 2016, pp. 484-485.

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," in IEEE Journal of Solid-State Circuits (JSSCC'17), vol. 52, no. 1, pp. 173-184, Jan. 2017.

D. Rozgic, V. Hokhikyan, W. Jiang, S. Basir-Kazeruni, H. Chandrakumar, W. Leng, D. Marković, (in press) "A True Full-Duplex 32-Channel 0.135cm^3 Neural Interface," in 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS'17).

Certifications

Dec. 2011 Professional Engineer (PE) in Electrical Engineering, License E 19873, by State of California Board for Professional Engineers, Land Surveyors, and Geologists

1 Introduction

Many neuropsychiatric disorders do not have cures but are rather chronically managed by different treatment modalities during the patients' lifespan. Neuromodulation – a process of regulating neuronal populations through electrical stimuli is one of these treatment modalities and is chronically applied through deep brain stimulation (DBS) implants. These implants have come a long way since their introduction in 1960's. They are currently FDA approved to treat Parkinson's disease, essential tremor, and dystonia and are being studied for treating a host of other neurological disorders (Tourette syndrome, chronic pain, depression, obsessive-compulsive disorder, epilepsy, Alzheimer's disease, etc.[10]). However, the hardware delivering DBS has not changed substantially during these years. Standard DBS probes (Medtronic DBS 3387/3389) have four (4) large ring-shaped stimulation sites which result in pea-size or larger activation volumes. In most cases, only one of these sites is being utilized delivering fixed-frequency rectangular constant-voltage or constant-current stimulation pulses. This stimulation paradigm is applied in always-on, open-loop fashion, ignoring the disease state, medication status, or side effects. Constant application of same stimulation pulses to a neural target activates the habituation of that network and, as a result, the stimulation settings become less effective over time. This habituation effect is usually addressed by occasional clinical visits to re-adjust the stimulation parameters. This re-adjusting process is a little more complex for some neurological disorders due to the long response times. In the case of dystonia, it often takes weeks to observe the effectiveness of stimulation parameters and is thus a tedious process for both - the patient and the clinician. On the other hand, while the open-loop stimulation paradigm improves patients' quality of life in

most cases, it may cause side effects. Hallucinations, manic responses, low mood, anger, and even suicidal thoughts are some of the side effects of DBS. Thus, sending continuous electrical impulses without any mechanism to detect the effectiveness of therapy may be detrimental to patient's mental and physical well-being. The two main reasons why stimulation is not applied automatically based on the disease state are 1) the lack of sensing capability in implantable pulse generators (IPG), and 2) the knowledge gap about the disease biomarkers, where the latter is mostly due to the former. On the stimulation side, recent studies conducted to test the effectiveness of higher density probes (eight sites vs. four) show that the increase in spatial resolution can improve the therapeutic index and potentially result in lower power consumption for achieving the same therapeutic benefits[5]. The status quo of the traditional rectangular stimulation waveforms is also being challenged - some non-rectangular waveforms seem to be more energy efficient[6],[7]. Taking all these recent developments into account, in this work we present a more advanced neural implant which, while being backward compatible with traditional DBS therapy, 1) is capable of sensing neural signals while delivering stimulation for providing feedback to closed-loop algorithms, 2) has high-channel-count for increased spatial resolution, and 3) can deliver custom, non-rectangular stimulation waveforms for improved energy efficiency.

1.1 Organization of This Dissertation

Chapter two presents the system-level design of our closed-loop ready high-channel-count neuromodulation unit. In Section 2.1 the non-traditional system architecture of the implant is discussed, and its advantages over traditional neural implant architectures are presented. Section 2.2 describes the design of our VCO-based neural sensing IC and our solution for increasing its

dynamic range for capturing both neural signals and stimulation artifacts. Section 2.3 presents our stimulation IC paying attention to special considerations pertaining to closed-loop neurostimulators such as patient safety and waveform parameter updates. Non-rectangular waveform generation mechanism is also discussed. Chapter three focuses on the verification of the designed closed-loop ready high-channel-count neuromodulation unit. Section 3.1 presents the NMU performance tests while Section 3.2 describes the hardware, firmware, and software designed for the NMU verification.

2 System Design

As discussed in Chapter 1, currently available neuromodulation systems improve the quality of patient care. However, there are still areas where the introduction of a closed-loop, high-channel-count system could make a significant impact. This chapter presents the design of a neuromodulation system, which we believe addresses the current shortcomings of DBS. Although the main focus of this work is the electronics, system architecture and implantation schemes are briefly reviewed to justify the need for miniaturization of the electronics.

2.1 The Neuromodulation Unit

The proposed neuromodulation unit (NMU) houses our custom designed high-dynamic-range sensing front-end and stimulation ICs. The design of these components will be discussed in more detail in Sections 2.2 and 2.3. This section focuses on system-level design features, and how they compare to those of the existing closed-loop DBS systems.

2.1.1 Nontraditional System Architecture

Traditional DBS implants consist of an IPG attached to a neural probe. After a small hole is drilled into the skull (burr hole), the neural probe is inserted through this hole to the neural target using stereotactic surgical techniques. The other end of the neural probe is routed under the skin to the upper chest area, where it connects to the implanted IPG (Figure 2.1). If more than one neural targets need to be stimulated, another pair of an IPG and a neural probe is implanted. This implantation scheme has been refined during the last several decades, works for several indications, and results in minimum amount of skull bone removal (burr hole).

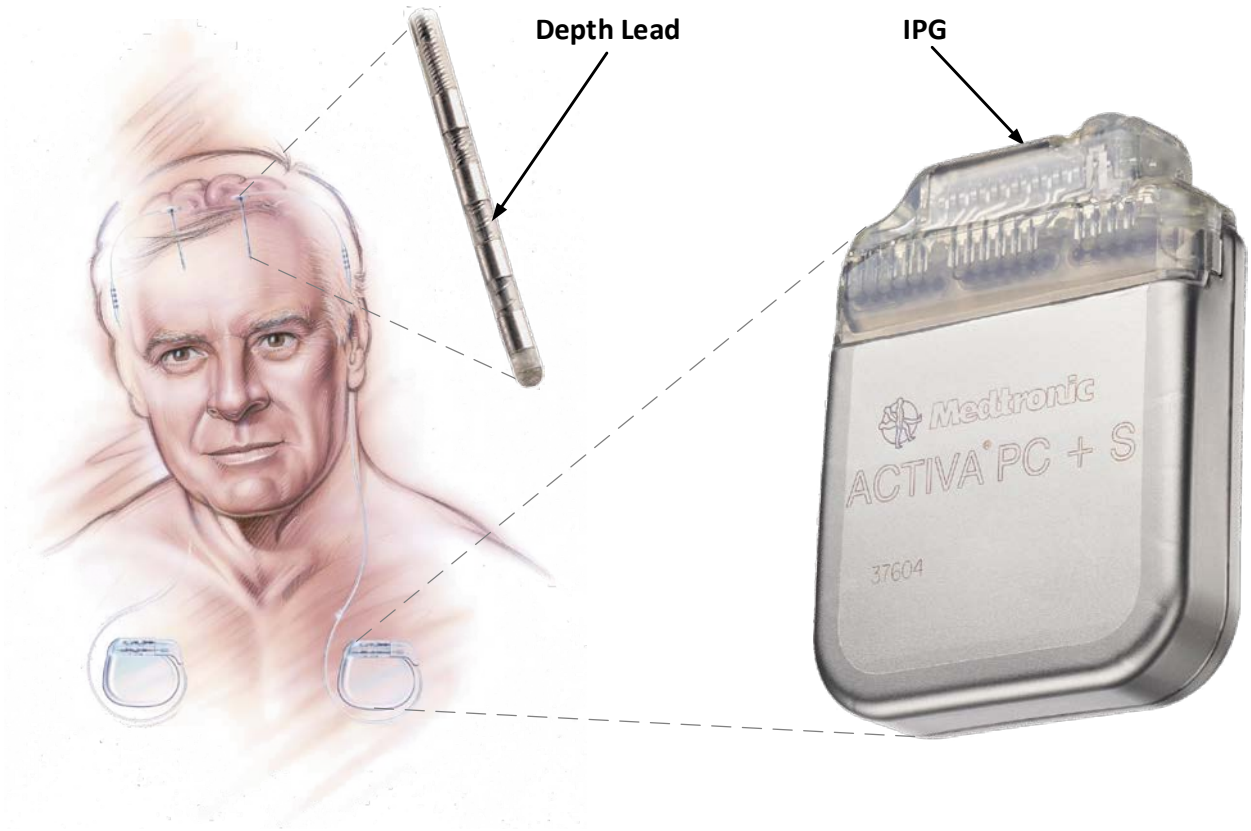

Figure 2.1 Traditional DBS IPG and probe implantation scheme (Medtronic Activa PC+S shown).

Since the introduction of Neuropace RNS neuromodulation system (November 2013), there is another implantation scheme in use – IPG is directly implanted in the cranium (under the scalp). This system has cortical and depth leads which are implanted in the area of the epileptic seizure focus. The depth lead is implanted through a burr hole using standard neurosurgical techniques. The cortical lead is implanted through craniectomy (removing part of the skull). This implantation scheme is targeted for epilepsy only and requires a significant amount of skull bone removal.

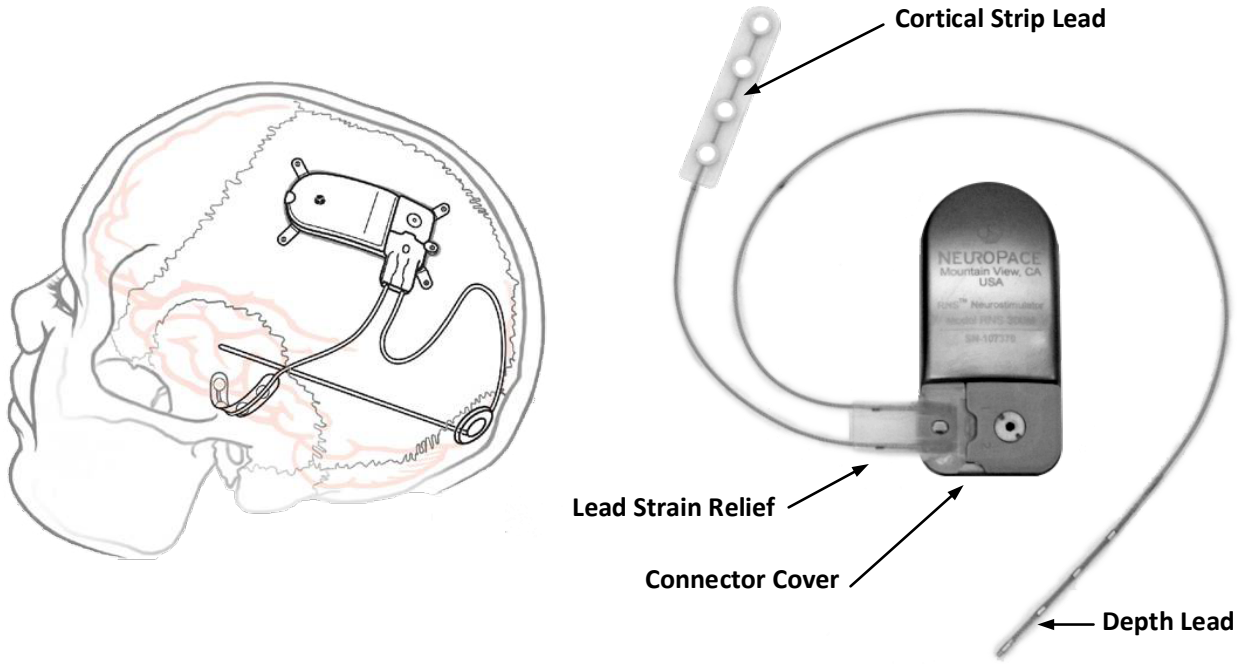

Figure 2.2 Neuropace RNS responsive neuromodulation system and its implantation scheme[11].

The proposed system is modular and has a slightly different implantation scheme. A neuromodulation unit (NMU) is connected to a neural probe and implanted epicranially next to the burr hole. Due to its small footprint, the NMU is affixed to the cranium by making a thin carving rather than removing it completely. The power and communication cable of the NMU is then routed to an aggregator module (AM) - also implanted epicranially using a small amount of cranial carving. If more than one neural targets need to be stimulated or recorded from, another neural probe and NMU pair (smart lead) can be implanted and terminated at the same AM. The cable running from AM to control module (CM) implanted in the chest, contains the power lines as well as the combined data stream from all NMU modules (Figure 2.3).

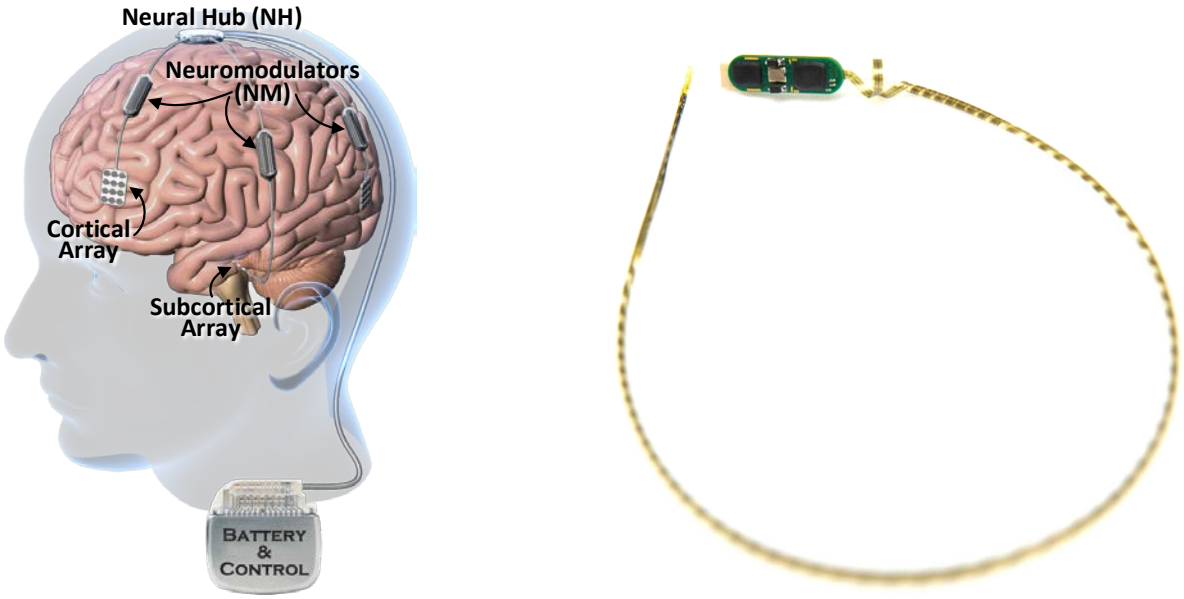

Figure 2.3 The proposed DBS system and its implantation scheme (left). Smart lead assembly (right).

2.1.2 Advantages of the Proposed System Architecture

When comparing our system to traditional DBS and Neuropace RNS implantation schemes in terms of surgical invasiveness, modularity, and proximity of the sensors and drivers to their neural targets, the following observations can be made (summarized in Table 2.1). The proposed implantation scheme becomes surgically less invasive than existing alternatives when three or more brain areas need to be targeted. This modular, distributed architecture allows increasing/decreasing the number of NMU satellites depending on a clinical application without modifying the system design. From the perspective of potential therapeutic effects, the main advantage of our system architecture compared to the DBS implantation scheme is the proximity of sensors and drivers to their neural targets. This advantage is less pronounced when comparing our system with Neuropace RNS since the latter is implanted in the cranium rather than in the chest. For sensing, proximity to the neural target means higher signal-to-noise ratio (SNR) and

for stimulation - lower power losses in cables. Reduced power leads to dissipating less heat inside the surrounding tissue, which according to American Association of Medical Instrumentation (AAMI) should not exceed 1-2°C for implanted medical devices.

Table 2.1 The proposed system's implantation scheme compared to traditional DBS and Neuropace RNS implantation schemes. Cells highlighted in blue denote the qualitative advantage of a system for a given category.

Implant	Traditional DBS Implants (Medtronic Activa PC+S or similar)	Neropace RNS	This work
Surgical Invasiveness for One to Two Neural Targets	**Minimal**	Maximal	Moderate
Surgical Invasiveness for Three or More Neural Targets	Maximal	Maximal	**Moderate**
Modular?	No, for more neural targets an additional system is required.	No, for more neural targets an additional system is required.	**Yes, for more neural targets only an additional smart lead is required.**
Proximity of Sensors and Drivers to Neural Targets	Minimal	Moderate	**Maximal**

2.1.3 Alternate System Topologies Utilizing Our NMU

The modular architecture shown in Figure 2.3 can be used for treating many neuropsychiatric disorders. In cases where the stimulation and sensing targets are highly localized, one NMU will suffice for therapy, and alternate system architectures are possible. The system shown in Figure 2.4, for example, resembles cochlear implant architecture. It is conceptualized for short-term memory enhancement application, targets hippocampal region of the brain, and consists of a fully implanted miniature IPG housing NMU assembly along with a near-field wireless data transceiver. Since the power management unit (PMU) residing in the stimulation IC has a mode for receiving wireless inductive power, the addition of a coil is all that is required to break free of an implanted battery requirement and its periodic replacement surgeries. The power and control, in this case, are provided through the external earpiece hidden behind the ear.

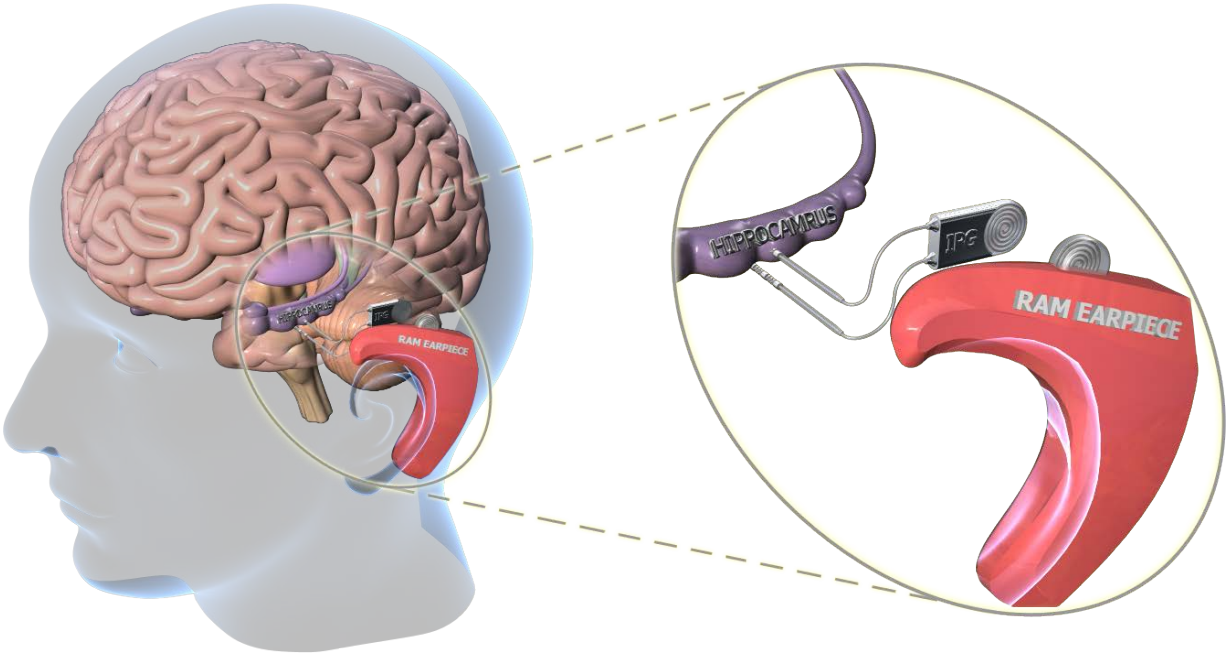

Figure 2.4 Alternate system topology utilizing the proposed NMU.

2.1.4 Advantages of the Proposed NMU Electronics

When comparing our proposed NMU electronics with that of currently available commercial and academic closed-loop neurostimulators, several features stand out – high-channel-count, ability to generate custom, non-rectangular stimulation waveforms, and ability to sense during stimulation (Table 2.2 and Table 2.3). These features are discussed in Sections 0 – 2.1.4.3 in more detail.

*Table 2.2 Proposed NMU electronics compared to that of currently available commercial closed-loop neurostimulators. Cells highlighted in blue denote the qualitative or quantitative advantage of a system for a given category. *The proposed system has eight (8) independent stimulation engines, which can be mapped to 64 electrodes by modifying stimulation IC configuration register.*

Implant	Medtronic Activa PC+S	Neropace RNS	This work
No. of Stim. Channels	4	8	**8 x 64***
Output Current (mA)	**0-25.5**	6-11.5	0-5.1mA
Compliance Voltage (V)	10.5	**12**	10
Frequency (Hz)	2-250	1-333	**2.4-16,667**
Pulse Width (µs)	60-450	40-1000	**10-1280**
Custom, Non-rectangular Stim. Waveforms?	No	No	**Yes**
No. of Sensing Channels	4	4	**32**
Simultaneous Sensing During Stimulation?	Only for certain stimulation site and parameter combinations	No	**Yes**
Battery Type	**Rechargeable**	Non-rechargeable	**Rechargeable**
Recorded Data Monitoring	**Wireless**	**Wireless**	**Wireless**

Table 2.3 Proposed NMU electronics compared to academic closed-loop neurostimulators. Cells highlighted in blue denote the qualitative or quantitative advantage of a system for a given category.

Reference		[12] **ISSCC'13**	[13] **TBioCAS'15**	[14] **ISSCC'16**	**This work**
Application		Cortical	Cortical	Spinal Cord	**Cortical / Sub-cortical**
Implant Size		>3cm^3	N/A	0.5cm^3	**0.135cm^3**
Stimulation	**No. Channels / Engines**	8 / 1	8 / 1	**160 / 40**	64 / 8
	Stim. Mode	**Current**	Voltage	**Current**	**Current**
	Custom, Non-rectangular Stim. Waveforms?	No	No	No	**Yes**
	Max. Current / Resolution	30µA (fixed)	0.23mA / 5-bits	0.5mA / 7-bits	**5.1mA / 8-bits**
	Max. Freq / Pulse-width / Resolution	N/A	220Hz / 440µs / 40 µs	**20kHz / 8ms / 10 µs**	20kHz / 1.26ms / 10µs
Sensing	**No. of Channels**	8	8	16	**32**
	Simultaneous Sensing During Stimulation?	No	No	No	**Yes**
	Linear Input Range (mV_{pp})	10	1	36	**100**
	SFDR	N/A	N/A	N/A	**81dB**
	ENOB (bit)	9.57	6.5	8.5	**12.8**
	Integrated Noise (μV_{RMS})	5.23	**1.97**	7.68	4
	Area/Ch. (mm^2)	0.38	0.35	N/A	**0.12**
	Signal BW (Hz)	1-7k	500-3k	1-7k	1-250

2.1.4.1 High-Channel-Count

Standard DBS probes (Medtronic DBS 3387/3389) have four (4) large ring-shaped stimulation sites which result in pea-size or larger activation volumes. Recent studies conducted to test the effectiveness of higher density leads (eight electrodes vs. four) show that the increase in spatial resolution can improve the therapeutic index and potentially result in less power consumption to achieve the same therapeutic benefits[5]. The Medtronic-Sapiens lead is currently the highest density probe in the industry and is being investigated for its therapeutic benefits. It is currently unclear, however, how the Medtronic-Sapiens electrodes are going to be paired and programmed. Neither Medtronic nor its competitors offer an IPG with such high channel count to be used with this probe. Our single NMU electronics can interface with 64-electrode leads, simultaneously sense neural data from 32 channels, and deliver eight (8) independent stimulation waveforms multiplexed to 64 neural sites. As a comparison, the two neurostimulators in the market with sensing capability - Neuropace RNS and Medtronic PC+S, can sense from four (4) channels only[15], and the highest number of stimulation sites in commercial DBS IPGs is 16.

2.1.4.2 Non-Rectangular Waveforms

Currently available DBS IPGs deliver therapy in the form of rectangular shaped pulses (Figure 2.5). There are two approaches to controlling these pulse amplitudes – 1) constant-voltage, and 2) constant-current. Constant-voltage IPGs output fixed voltage during the stimulation, regardless of electrode-tissue impedance changes. If electrode-tissue impedance changes over-time (e.g., scar tissue formation), a constant-voltage pulse will deliver varying levels of current (hence charge) to the neural target. This change will result in a different therapy than initially programmed and

will need to undergo a readjustment by a clinician. Constant-current IPGs will vary the pulse voltage to attain the programmed amount of current up to a maximum value (compliance voltage). The electronics for these IPGs is a little more complex; however, constant charge injection into the neural target is ensured.

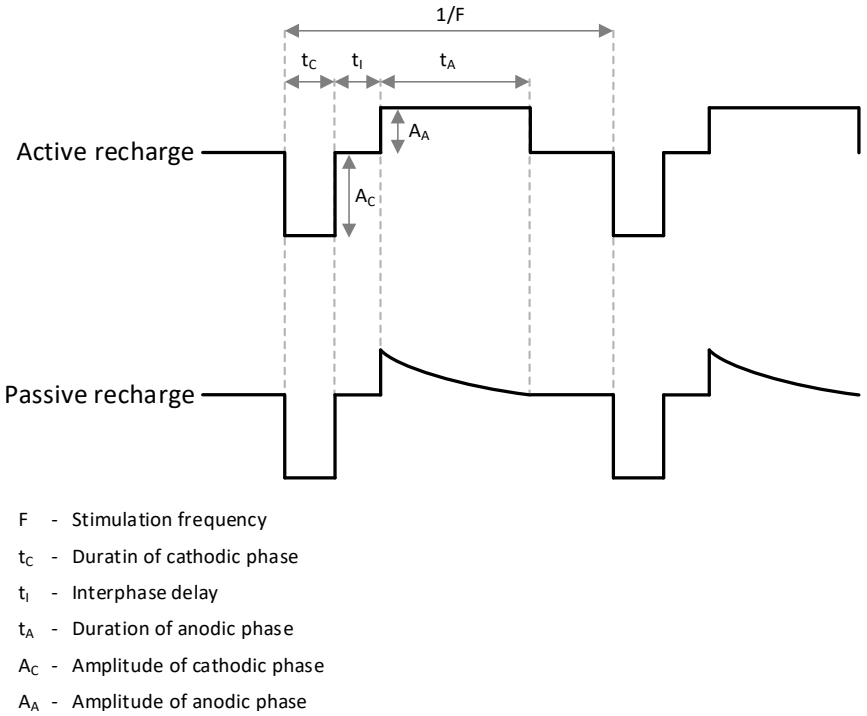

Figure 2.5 Constant-voltage and constant-current rectangular stimulation pulses used in current DBS systems.

Rectangular shaped pulses have become the de-facto standard for neural stimulation since the inception of the neuromodulation field. The reason behind this decision is not the effectiveness of delivered therapy but perhaps has to do with the ease of generating such pulses from the engineering perspective. As a result, the strength-duration curves and charge injection capacities of neural electrodes are defined for rectangular pulses. Recent studies challenged the status quo of the traditional rectangular stimulation waveforms and found that some non-rectangular waveforms seem to be more energy efficient[6],[7]. It is hard to assess additional clinical benefits

of non-rectangular stimulation pulses due to the lack of hardware capable of generating such waveforms. To provide maximum flexibility to the pioneers of the neuromodulation field, our NMU electronics features non-rectangular constant-current biphasic waveform generation with selectable coarseness settings (Figure 2.39). Traditional rectangular stimulation pulses are also supported.

2.1.4.3 Simultaneous Sensing and Stimulation

Among currently available DSB implants, only two have neural sensing capability – Neuropace RNS and Medtronic PC+S. Neuropace RNS is an FDA-approved device that targets epilepsy and can record from up to four (4) electrodes simultaneously. Although the electrical specifications of the recording analog front end are not revealed, the user manual talks about the saturation events while sensing and their elimination from processing algorithms by using saturation event detectors[11]. Signal saturation is due to an insufficient dynamic range of the recording analog front end, and can be prevented by disconnecting sensing amplifiers from the signal source - "blanking" (Figure 2.6). "Blanking" of course results in data loss and is probably employed by Neuropace RNS during stimulation based on device functionality illustration[16].

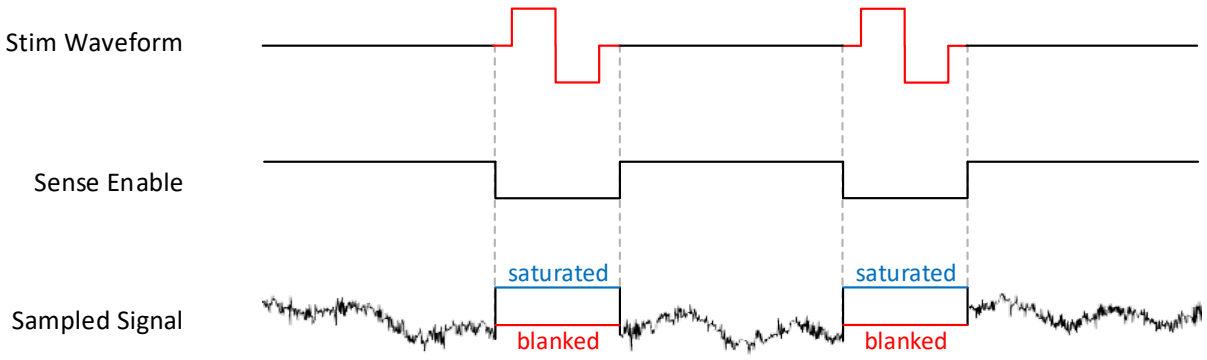

Figure 2.6 Illustration of "blanked" neural sensing.

Medtronic PC+S is an investigational device not yet approved by FDA and is meant to be used for research purposes[17]. This device can simultaneously sense LFP signal from four(4) channels while stimulating[18]. However the caveat is that a specific electrode configuration and stimulation parameters should be used[15].

Our NMU allows simultaneous sensing of LFP signals from 32 channels while stimulating without "blanking" or restrictions on recording sites and stimulation parameters. Since our sensing front end captures neural signals along with stimulation artifacts, it can help to understand probed neural network's dynamics during stimulation. More neuronal populations can be investigated for biomarker content around the implanted probe due to the absence of recording site selection restrictions. Lastly, patient therapy will not need to be compromised due to a limited set of stimulation parameters.

2.1.5 NMU Physical Dimensions

Two versions of NMU printed circuit board assemblies (PCBA) were designed (Figure 2.7). The NMU PCBA v1 was assembled with earlier versions of the stimulation and sensing ICs which were designed for 32-electrode neural probes. The NMU PCBA v1 is smaller than a penny (18.8mm x 5.8mm x 1.8mm) and occupies about 135mm^3 of volume. The stimulation and sensing ICs along with only six (6) passive components are placed on the top side of the printed circuit board (PCB), while the bottom side is reserved for power, communication, and neural electrode connections. The 180° phase shifted AC power is delivered through two (2) conductors located near the communication lines. In this version of the NMU PCBA, communication with the sensing and stimulation ICs is achieved through three (3) and four (4) line serial peripheral interfaces (SPI)

respectively. The neural interface consists of 34 pads – 32 for neural targets and two for "common" and "reference" electrodes. The NMU PCBA v2 is smaller than a quarter (22.5mm x 4.5mm x 2mm), and was designed by Wenhao Yu for 64-electrode neural probes. This version shares the same topology as v1 - components on top and implant feedthrough contacts at the bottom. The communication section of the feedthrough differs from that of v1 – stimulation and sensing ICs both share a 3-wire SPI interface instead of using two separate SPI interfaces. This design change is covered in more detail in Section 2.1.7.

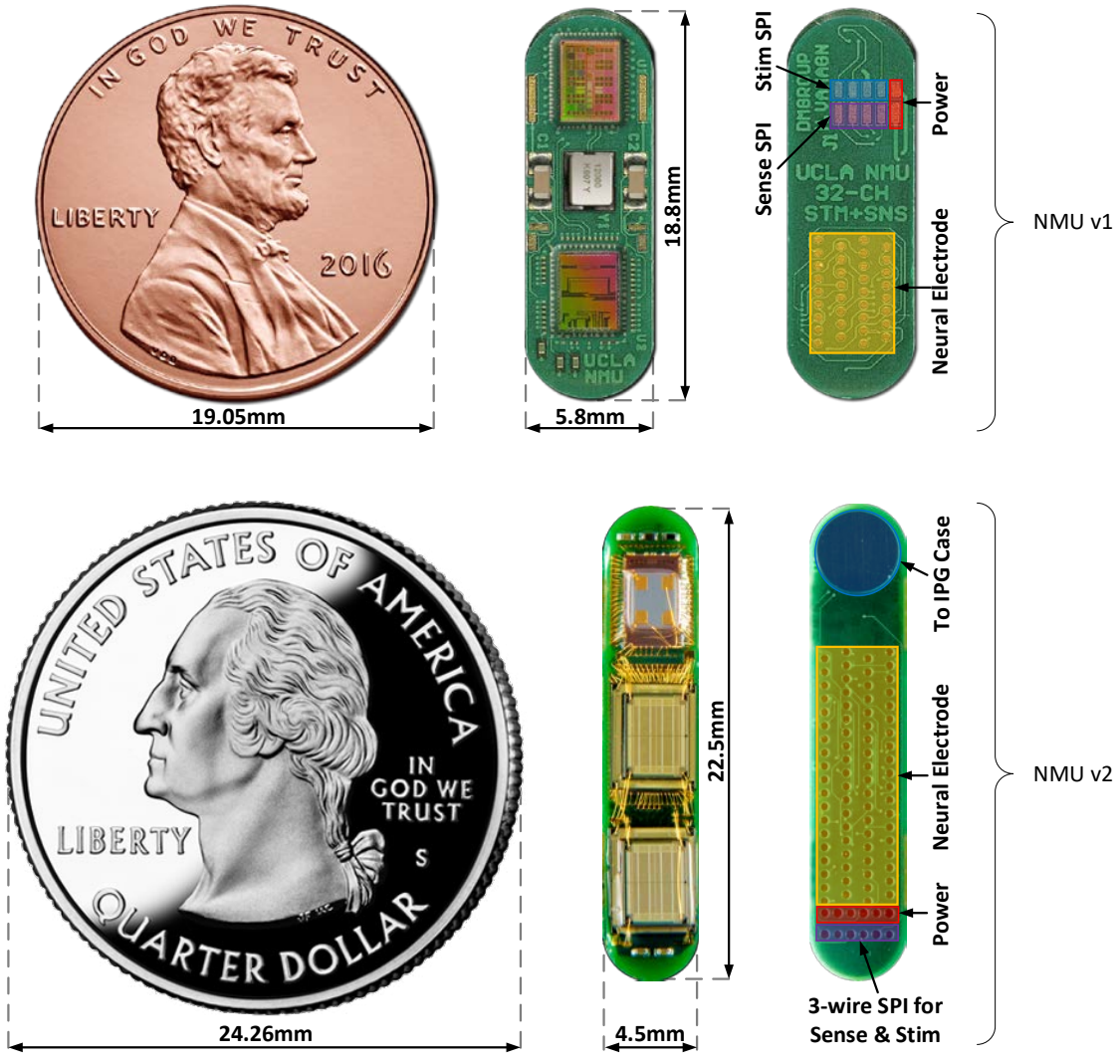

Figure 2.7 The top and bottom sides of v1 and v2 of the proposed NMU PCB assembly.

2.1.6 NMU PCB Design

The NMU PCB required advanced features for fabrication largely due to very small form-factor and the high number of interconnections between the stimulation IC, sensing IC, and the neural lead feedthrough interface. The eight-layer PCB is 31mil with 3mil minimum spacing, 3mil minimum track width, and uses blind-and-buried vias for routing. The board was surface finished with electroless-nickel-electroless-palladium-immersion-gold (ENEPIG) coating both for ease of sensing and stimulation IC wire-bonding as well as for neural probe attachment process compatibility.

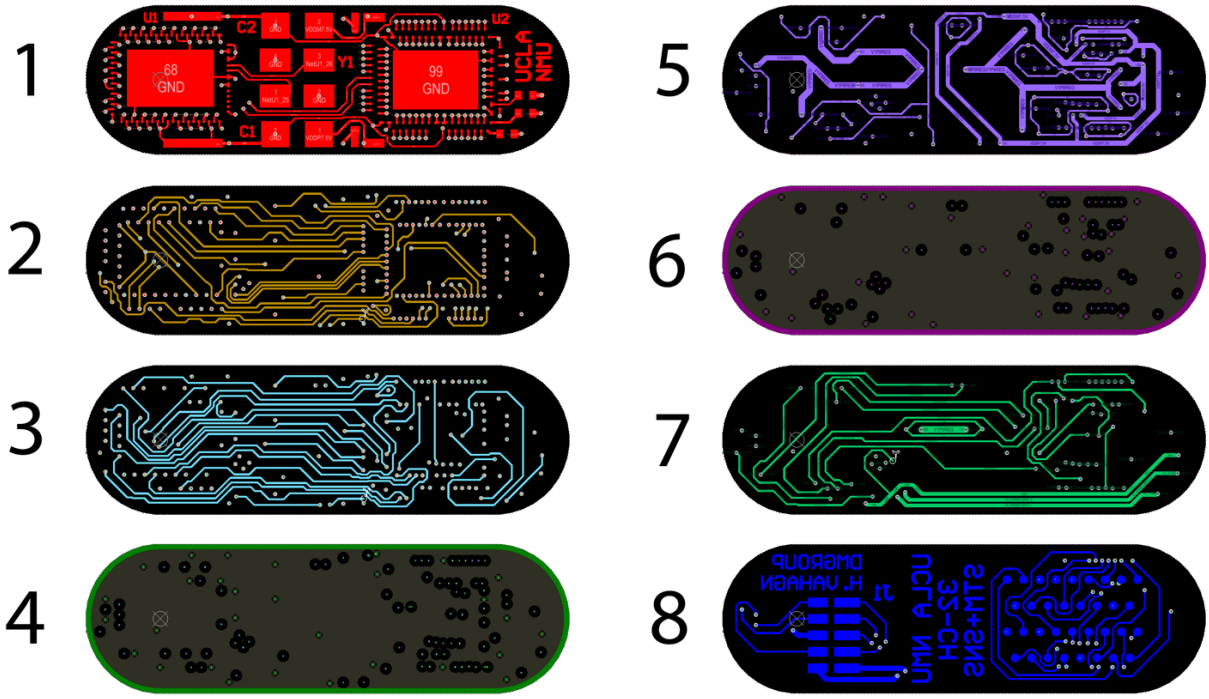

Figure 2.8 The NMU v1 PCB layers. Displayed from the top (#1) to bottom (#8).

When routing two traces close to each other, the current flowing through one trace (aggressor) induces a current in the other trace (victim). This "crosstalk" can cause undesirable effects in the circuitry at the victim node. For this reason, potential aggressor traces were shielded from

surrounding victim traces located on the same layer. Since the AC power signals were high on the list of aggressor traces (5Vpp, 1MHz), they were grouped in one power layer, and besides being isolated from other power traces on that layer, they were also shielded from the layers above and below by two ground planes (Figure 2.8).

2.1.7 Using Three Wires for Communicating with Two SPI Slaves

Both sensing and stimulation ICs are implemented as SPI slaves. When only one SPI slave is connected to an SPI master, the minimum number of required physical lines is three (3) – serial clock (SCK), master-out-slave-in (MOSI), master-in-slave-out (MISO). When connecting additional SPI slaves to the system one slave-select (SS) signal should be added for each slave. For a two-slave configuration, this brings the number of needed physical conductors to five (5) (Figure 2.9).

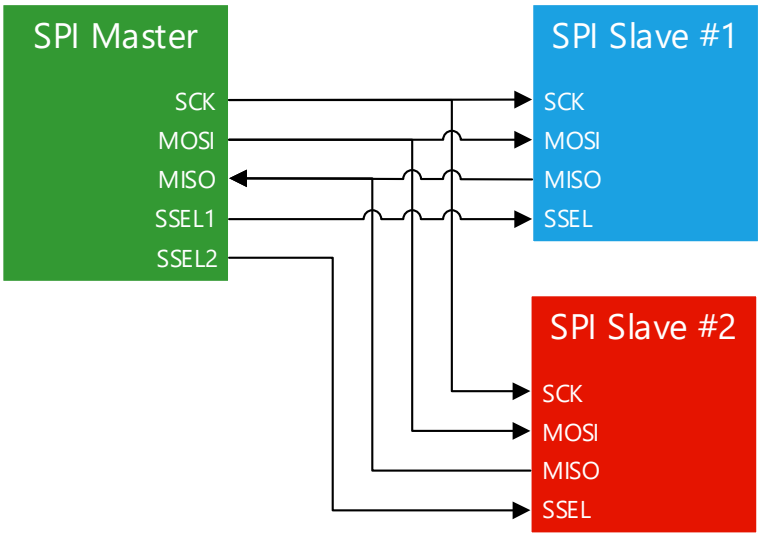

Figure 2.9 SPI master connected to two SPI slaves. For avoiding contention, SSEL1 and SSEL2 lines decide which slave drives the MISO line.

Assuming four (4) NMU assemblies per system, the number of physical conductors running from NMU to the AM module just for digital communication will be 20. This may not seem many, but the feedthrough-interconnect takes significant space on the AM package. To reduce this space and make the AM smaller, while conforming to the regular SPI bus specifications we came up with a new multiplexed bus architecture based on SPI which allows two (2) slave communication only through three (3) wires instead of five. This modification reduced the number of physical conductors running from NMU to AM from 20 down to 12 (Figure 2.10).

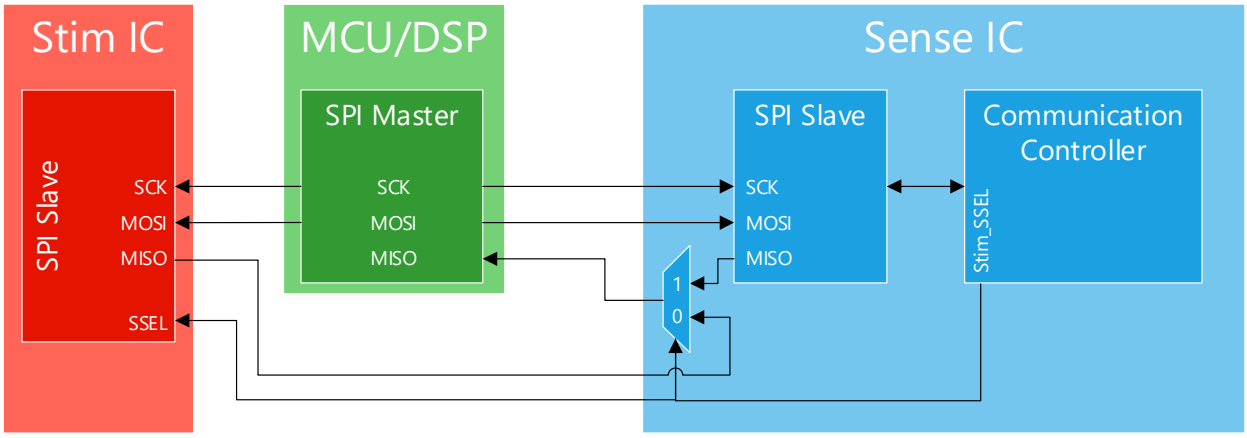

Figure 2.10 Our scheme of controlling two (2) SPI slaves (sensing IC & stimulation IC) using only three (3) wires.

This low-pin-count SPI bus was achieved by the introduction of a small overhead in the communication protocol. In the case of the conventional SPI communication, the asserted SS line tells the slave that it was selected for data exchange. The data from the master is then transferred through the MOSI line, and the slave sends a simultaneous reply on the MISO line. In our case, we assigned the sensing IC to be the communication gateway. The SPI communication controller (SCC) hosted in this IC assumes that as a slave it is always selected and passes the received SPI bytes through its decoding logic. When an SPI packet starts with a STIM_SELECT token (0x22),

then the SCC asserts the SS line of the stimulation IC and selects the reply of the stimulation IC (MISO) for routing to the SPI master (MCU located inside the AM) (Figure 2.11). The SCC will also set itself in the pass-through mode and not process the upcoming packet bytes. When the packet does not start with a STIM_SELECT token, then the SCC assumes the communication is addressed to the sensing IC and processes the following packet bytes.

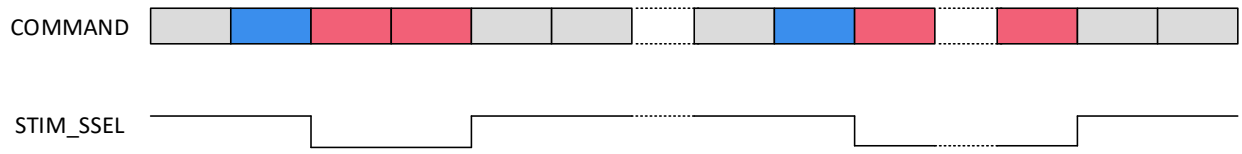

Figure 2.11 SPI communication controller processing the STIM_SELECT tokens (blue).

2.1.7.1 Absence of SPI SSEL Signal and Recovery from Partial Packets

The above-described custom SPI communication protocol works for well-formed packets. This cannot be assumed to be the case all the time, and the corresponding risks need to be mitigated at the SCC level. Let's assume the SPI master starts sending a packet to one of the NMUs, but due to some system malfunction only part of the packet is transmitted, or due to noisy environment extra unintended data bits are shifted into the NMU's receive buffer. Regardless of how the current packet will be processed, the next packet will be out of synchronization between AM and NMU. With a conventional SPI protocol, the SS line takes care of these scenarios by indicating to the internal SPI slave finite state machine (FSM) that it is about to receive the first byte in the packet. This is the reason why many single slave systems will still employ the four-line SPI (adding the SS line), even though they can get away with only three lines. To recover from partial packets, the proposed SCC employs a timeout counter for receiving packet contents. A 100ms timeout activates after each received packet byte. If all packet bytes are received, the

timeout counter deactivates. However, even if one byte is not received, the timeout event will trigger 100ms after the last byte, and the slave FSM will return to initial state. Before returning to the initial state, the slave FSM can also reset the IC if the RESET_ON_ERROR option is set in the configuration register.

2.2 The High-Dynamic-Range Sensing IC

Our high-dynamic-range sensing front-end enables the proposed system to concurrently capture both – neural signals of interest and stimulation artifacts. This section outlines why that is challenging for low-power neural sensors required for implants, and presents our solution.

2.2.1 Challenge of Capturing Neural Signals and Stimulation Artifacts

The two types of neural signals that can be accessed through implanted neural probes are local field potentials (LFP) and action potentials (AP) (Figure 2.12). APs are also known as single-units and as the latter name implies contain information about the activity of single (or several) neurons. APs are thus highly localized, and in order to convey information about a specific neural network state, very-high-channel recording is needed followed by a digital signal processing engine for integration. Another factor that makes APs impractical for implanted system usage is the over-time scar tissue formation around the ultra-fine electrodes. The latter dramatically decreases the signal quality of the recordings rendering them useless. LFPs, on the other hand, contain information about the dynamics of large groups of neurons, are recorded from much larger electrodes, and are less prone to the over-time signal degradation. These factors make LFPs a more preferable choice for implanted systems.

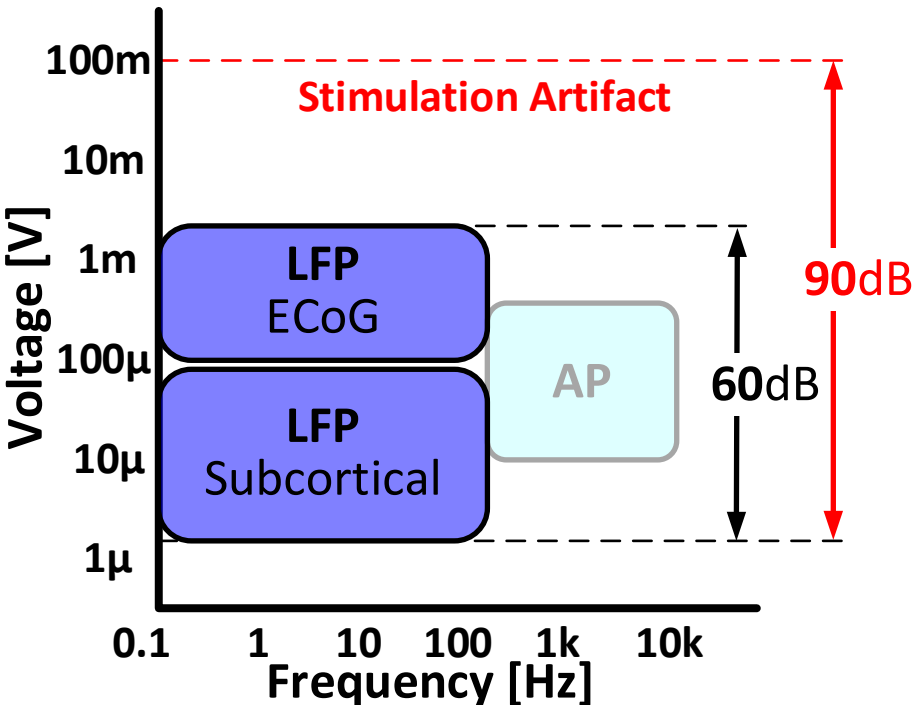

Figure 2.12 Frequency and amplitude of the neural signals of interest.

LFPs span a bandwidth of 0.1-100Hz, and an amplitude of 1µV-1mV. Thus, an analog-front-end (AFE) designed to capture LFPs should have a dynamic range of about 60dB. The LFPs are, however, not the only electrical signals present in a DBS implanted brain. The therapeutically injected stimulation pulses can induce up to 100mV signals into the same neural tissue from which neural sensing is performed. This would cause the signal chain to saturate. Until the stimulation artifacts are removed and the AFE recovers from saturation, all the neural signal information will be lost. To avoid this, one could of course design an AFE with sufficient dynamic range to capture both the neural signals and the stimulation artifacts. The challenge is that a typical implant battery can not sustain the amount of power required to operate this circuit. Technological innovation is required to achieve the required dynamic-range while consuming implant-level power (Figure 2.13) - which is exactly what our sensing front-end does.

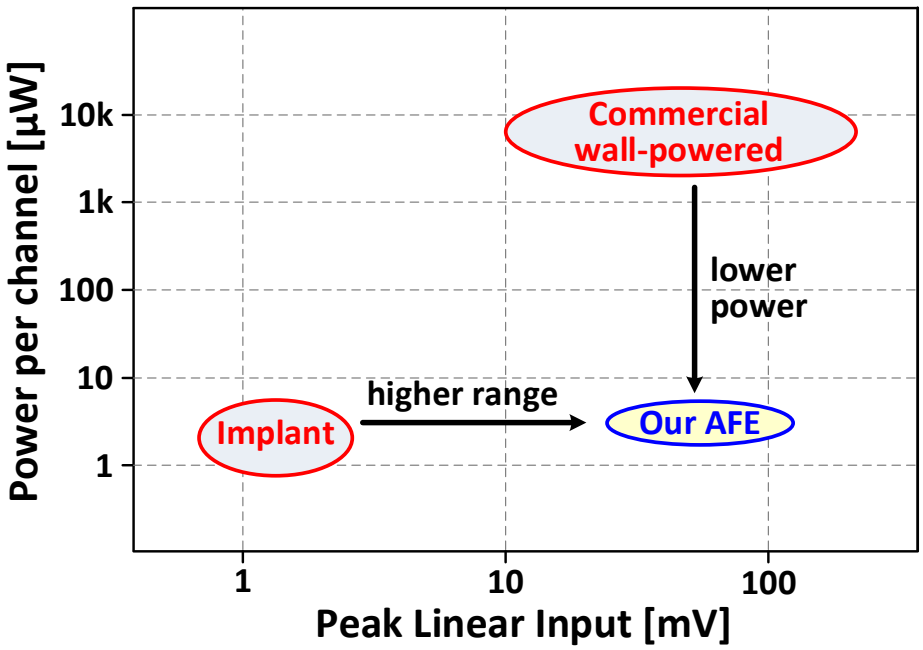

Figure 2.13 Power and dynamic range tradeoff of a sensing front-end.

2.2.2 Our Sensing Front-End Architecture

Conventional neural sensing front-ends are optimized for recording neural signals only[19],[20],[21]. Prior to conversion, these AFEs are typically proceeded with a large gain instrumentation amplifier (IA) stage. The IA stage provides a high input impedance and also high input gain, which relaxes the noise requirement - hence power consumption, for the next stages. Inputs that exceed the designed input range will result in signal saturation which can take tens of milliseconds to recover given the large time constants involved in neural recording. This is the main reason why some systems do not even attempt recording while stimulation is on ("blanking"). A solution proposed by Vaibhav Karkare [22] and later improved by Wenlong Jiang [23] is as follows. VCO-based ADCs rely on phase-domain increments of the oscillator output. Since there is no physical bound on the maximum phase of a waveform, the output of a VCO-ADC is not prone to saturation (Figure 2.14).

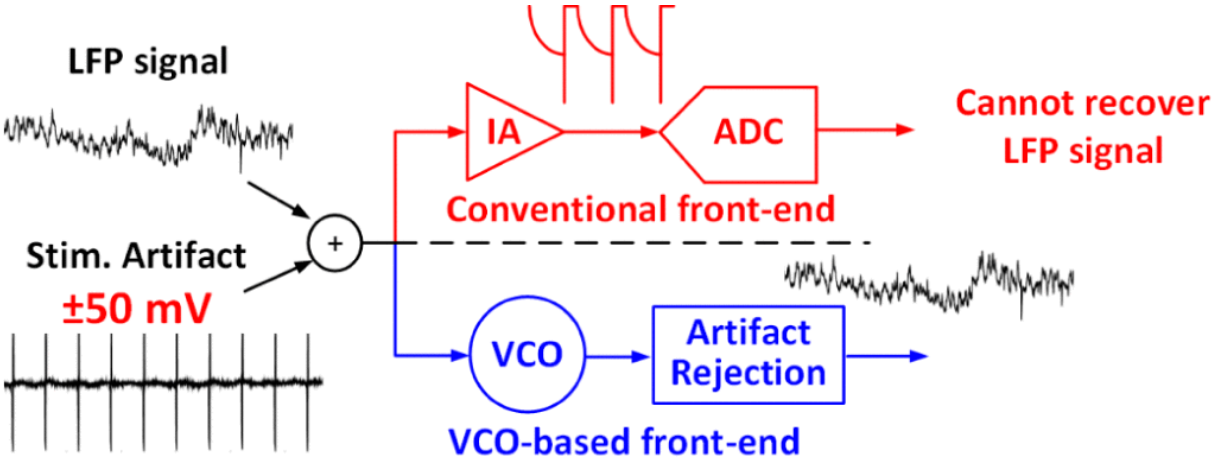

Figure 2.14 Conventional sensing AFE vs. VCO-based AFE.

Figure 2.15 shows the high-level schematic of the mentioned VCO-based ADC. The design of this ADC is not going to be discussed in this work, for more details refer to V. Karkare's and W. Jiang's publications[22],[23].

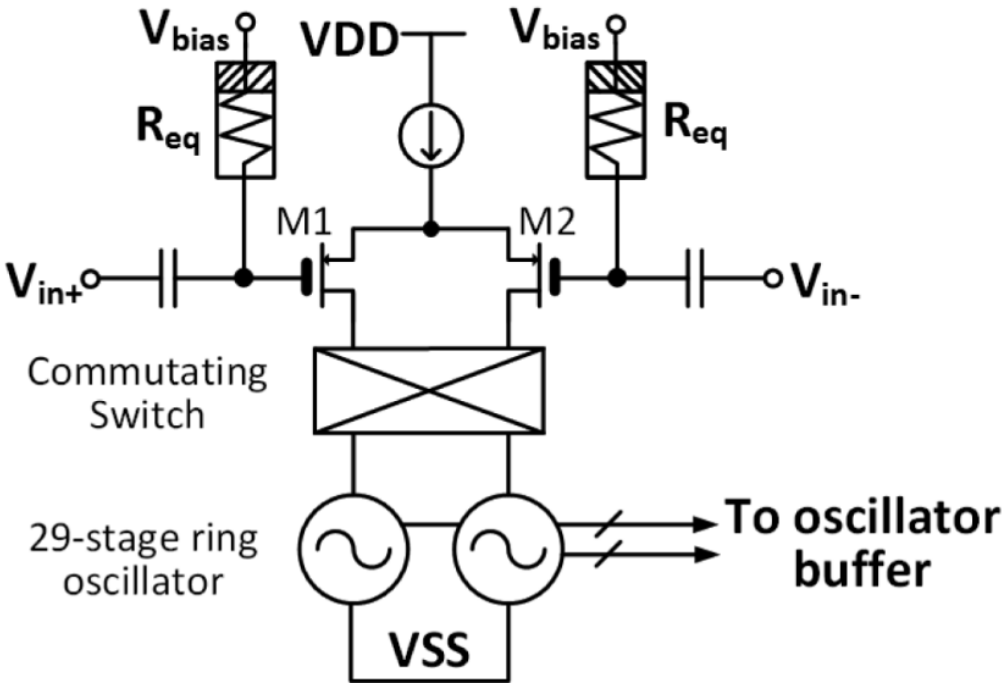

Figure 2.15 Schematic of a VCO-based ADC used in our system.

2.2.3 Challenges of Using VCO ADCs for High-Resolution Applications

While the selected VCO-ADCs architecture is not prone to saturation, it does not have a linear voltage-to-frequency tuning curve (Figure 2.16). This limits the effective resolution of the VCO ADC to approximately seven (7) bits and makes it unsuitable for simultaneous recording of neural signals and stimulation artifacts. The good news is, however, that this nonlinearity can be corrected post-conversion. In general, two calibration methods can be applied – 1) foreground, and 2) background. For foreground calibration, a mapping function is applied in a feedforward fashion on the conversion results. Background calibration, on the other hand, has a feedback loop which updates the mapping function periodically to ensure mapping function didn't drift over time. Accuracy-wise this method is preferable to foreground calibration however it requires more power and silicon area. Foreground calibration, on the other hand, can be realized with simpler building blocks if mapping drift can be accounted for. The contributors of VCO nonlinearity are 1) the nonlinear voltage-to-frequency relationship, 2) VCO supply voltage variations, 3) temperature and 4) the physical aging of the circuit. Since the voltage-to-frequency relationship is time-invariant, the supply voltage can be easily fixed using a low-dropout (LDO) voltage regulator, and the effects of aging contribute very little to the overall nonlinearity and can be manifested after many years of constant operation, the main remaining contributor of the mapping drift is the operating temperature variation. As an implanted neural implant, the entire electronics is going to be housed inside human body which has a temperature stability of about $\pm 2^{\circ}C$. Given a typical VCO temperature sensitivity of 50ppm/C[24], the introduced readout drift of 100ppm will be buried under the noise floor of a 13 effective number of bits (ENOB) ADC. For all the

mentioned reasons, the foreground nonlinearity calibration is acceptable for an implantable neural recording system.

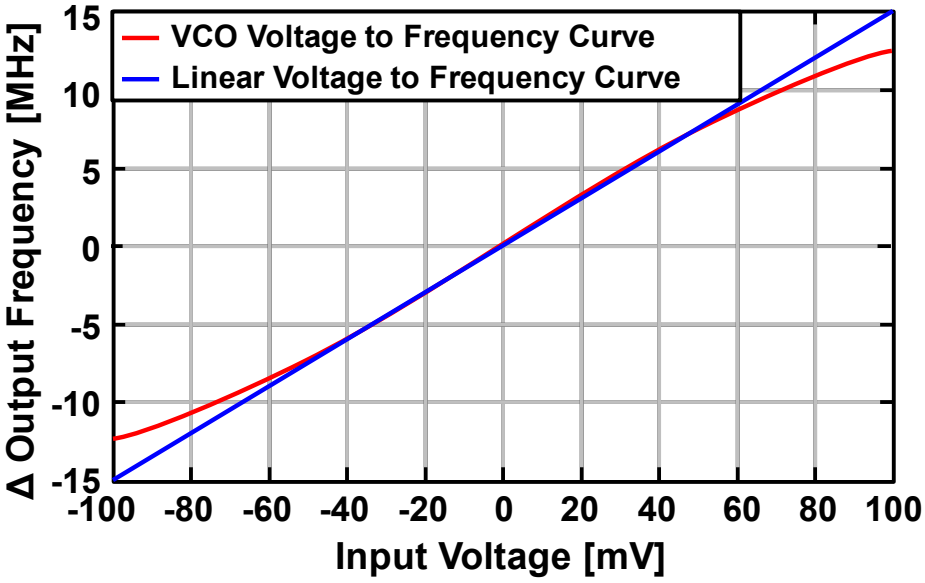

Figure 2.16 Simulated voltage-to-frequency tuning curve of a VCO-ADC.

2.2.4 Previously Used Methods for VCO ADC Nonlinearity Correction (NLC)

Over time different methods have been proposed to improve the linearity of VCO-based quantizers. In [25] the VCO-based quantizer linearity is improved by placing it inside a continuous time sigma-delta feedback loop. Another approach is using a replica VCO to perform correlation based background calibration [26]. In [27] and [28] a look-up table is used to invert the nonlinearity of the ADC. In a high-channel-count system, however, replica VCOs and look-up tables (LUT) will take a significant area. For example, our 32-channel ADC with a raw resolution of 21-bits and expected output resolution of 15-bits will require a memory with 32 x 2^{21} locations and 15bits of word length. Assuming 0.124μm^2 [29] for 6T SRAM cell area, 124.8mm^2 silicon area will be

needed, which is prohibitively large. For comparison, each VCO ADC core takes only 0.12mm^2 silicon area.

2.2.5 The Proposed NLC Approach – Foreground Calibration Using a Polynomial Computation Engine

Provided that the VCO-ADC maintains monotonicity in its transfer function, this nonlinearity can be corrected by learning and inverting the VCO tuning curve. A nonlinear function such as a polynomial can be mapped to this inverted tuning curve (Figure 2.17). Using this foreground calibration approach, we will need to store only the polynomial coefficients; however, we need to calculate polynomial function value for each ADC sample.

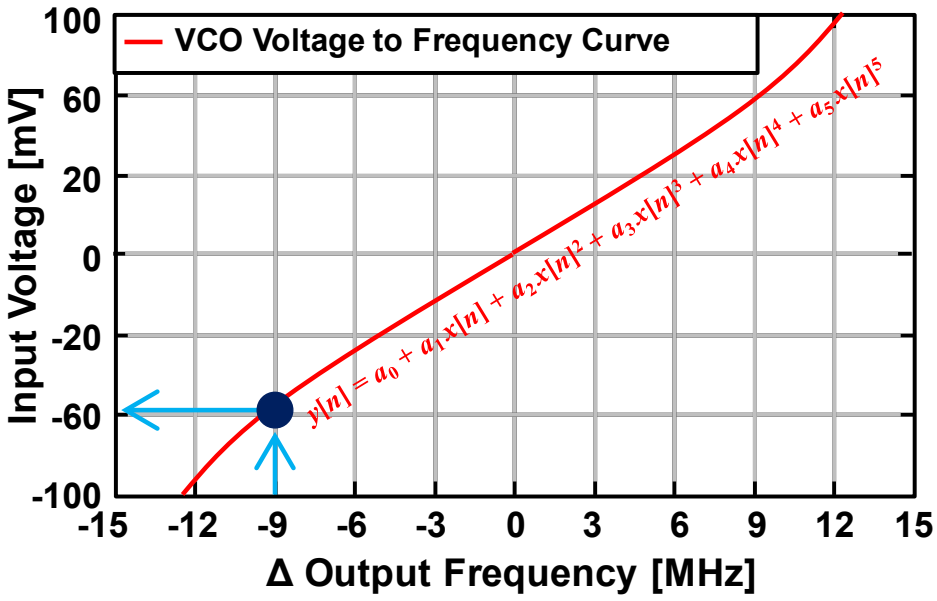

Figure 2.17 Finding reverse frequency-voltage mapping of VCO ADC.

2.2.5.1 Engine Order and Hardware Implementation

The hardware complexity of this calibration technique depends on the polynomial order. Since the VCO distortions are dominated by third and fifth order nonlinearities: a fifth-order polynomial

should be sufficient to correct for them. The general form of our fifth-order polynomial is the following:

$$y[n] = a_0 + a_1x[n] + a_2x[n]^2 + a_3x[n]^3 + a_4x[n]^4 + a_5x[n]^5$$

where $x[n]$ is the NLC input sample, $a_0 - a_5$ are the polynomial coefficients, and $y[n]$ is the NLC output. For better accuracy, we use single-precision floating-point operations to compute the polynomial. A straight-forward implementation would require too many floating-point multipliers that take up significant area even in an advanced technology node. Noticing that the sampling rate of the quantizer is very low, a tradeoff of execution time and area can be made in our favor by adopting Horner's method, which allows iterative computation of a polynomial using a single multiplier-accumulator (MAC) kernel. With Horner's method, our polynomial can be rewritten as

$$y[n] = a_0 + x[n](a_1 + x[n](a_2 + x[n](a_3 + x[n](a_4 + x[n](a_5 + x[n] \cdot 0)))))$$

Thus, $y[n]$ can be computed starting with the innermost MAC operation $a_5 + x[n] \cdot 0$, then expanded outward until the whole polynomial is computed. The schematic representation of the implementation is shown in Figure 2.18. When the quantizer output is ready, the raw sample $x[n]$ is sent to the NLC engine and the input valid signal is asserted for one clock cycle. The engine then starts the first iteration, where it multiplies the input sample $x[n]$ with the accumulator register content (which is initialized to 0), adds the fifth-order coefficient a_5 to it, and stores the result back into the accumulator register. In the second iteration, the input $x[n]$ is multiplied with accumulator register content (which is now nonzero); the fourth-order coefficient a_4 is added, and the result is stored into the accumulator register again. This process continues until the whole

polynomial $y[n]$ is computed. After the last (sixth) iteration is complete, the output valid signal is asserted to indicate that the nonlinearity corrected sample is now ready. This iterative polynomial solution with only one MAC kernel reduces the NLC area to be much smaller than the classical LUT approach.

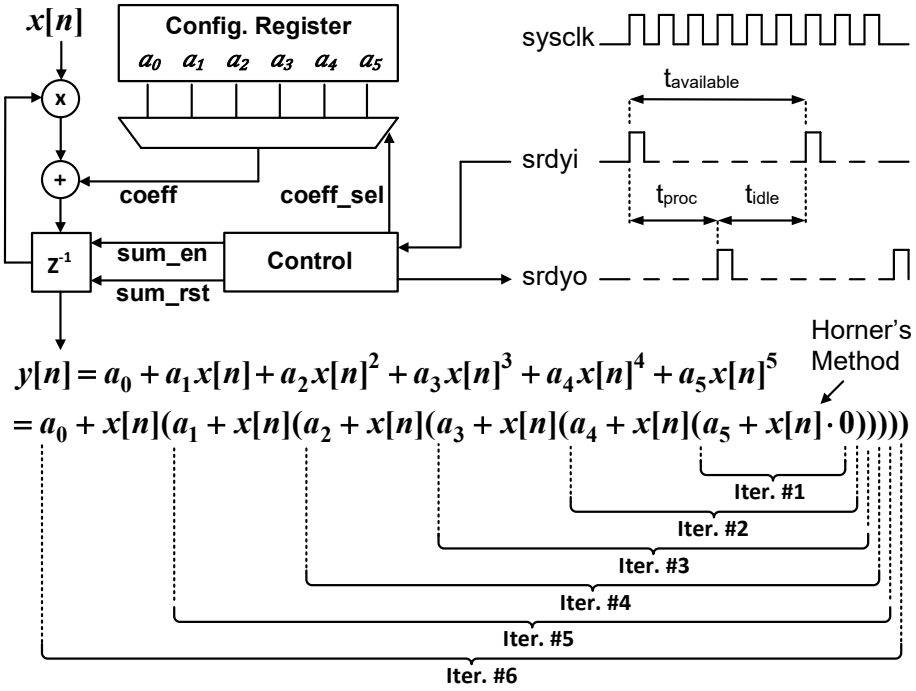

Figure 2.18 NLC engine architecture. Source-ready-in (srdyi) and source-ready-out (srdyo) signals specify whether the input and the output of the NLC block are valid.

2.2.5.2 Engine Order Revisited

The NLC block for the first version of the sensing IC was designed as a 6^{th} order polynomial engine. This was due to the 3^{rd} and 5^{th} order nonlinearities being dominant terms for VCO distortions. For the next IC iteration among other improvements, we also increased the polynomial order to make sure it was not the limiting factor for the ADC linearity. This was because the IC design CAD tool simulations would show much higher nonlinear distortions at the output of the VCO ADC than the expected 5^{th} order terms. The redesigned polynomial engine split the input

range of the VCO ADC into four (4) sections, each one having a 10^{th} order polynomial engine (Figure 2.19).

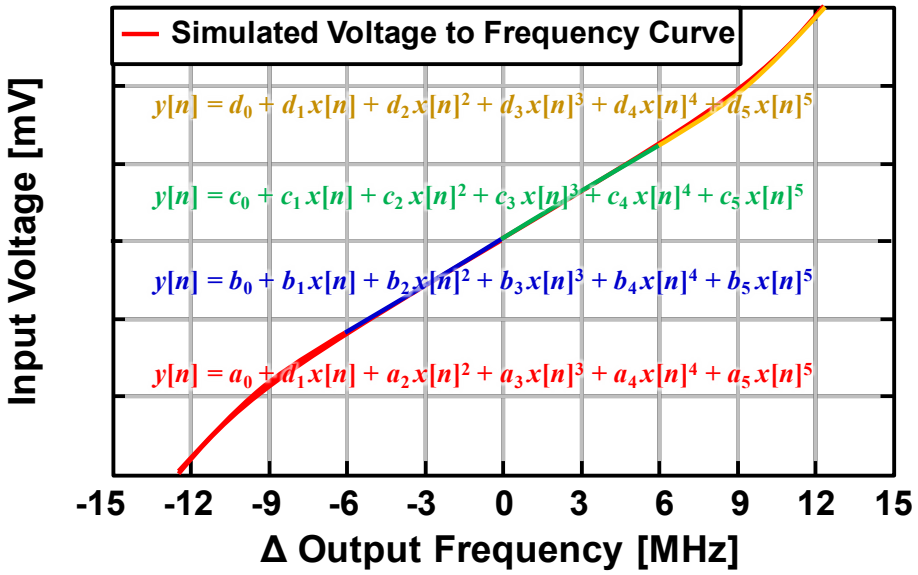

Figure 2.19 Dividing the NLC curve into four sections.

Two different metrics were used to evaluate the fitting performance of the new engine – 1) mean square error (MSE), and 2) ENOB extracted from the fast Fourier transform plots. For the first test, we set out to find out whether dividing the nonlinearity curve into a different number of sections (up to four) results into fitting improvements. Since this engine takes four different polynomial coefficients, we can effectively make one, two, or three section engines by setting some section coefficients to be equal to each other. The measurement results revealed that one-section calibration had the best ENOB and MSE performance (Figure 2.20). This result is a little surprising. If the one-section engine is sufficient for the fitting, the two, three, and four sections should perform as well and not worse than one-section since they provide more degrees of freedom for the fitting. Practically, however, the degraded results for the two, three, and four section tests were due to the curvature mismatch at the section division points.

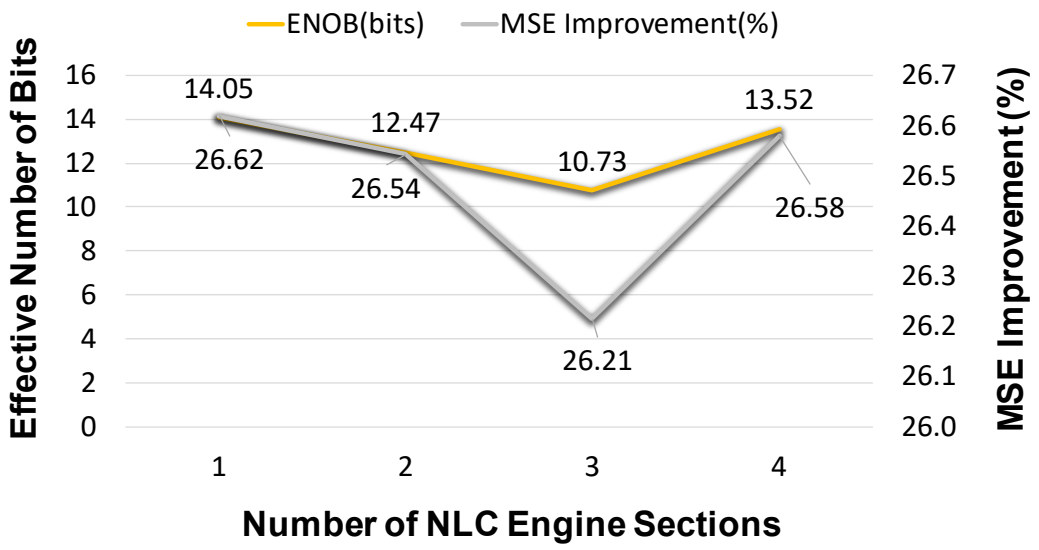

Figure 2.20 Polynomial fitting performance of four-section NLC engine when utilizing different number of sections.

By fixing the section number to one, the next performed test was to identify the required polynomial order for the engine. The same ADC data was applied to a one-section engine, however, in each case, the polynomial fitting function was constrained by a different degree of freedom. The measurement results showed that the fitting performance plateaus sharply after 5^{th} order (Figure 2.21). This is consistent with our original analysis.

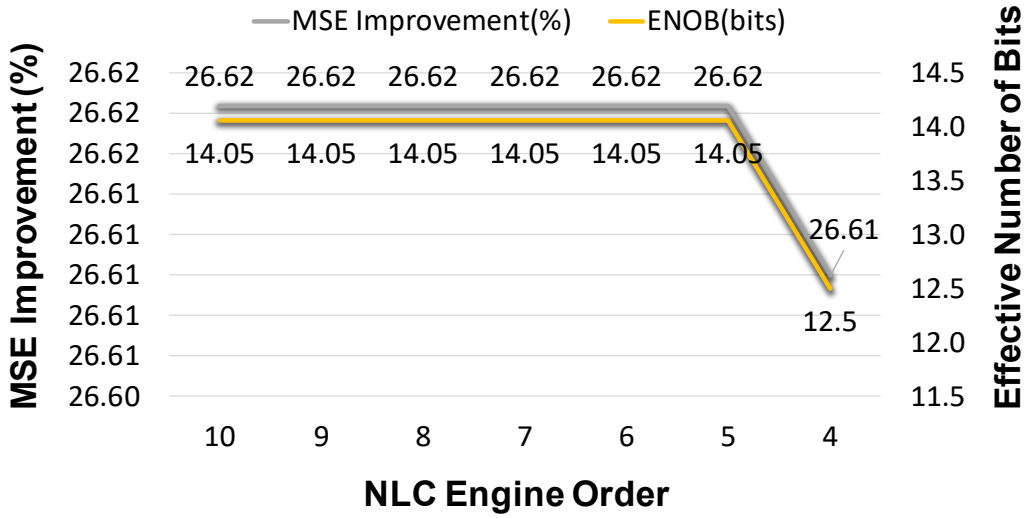

Figure 2.21 Polynomial fitting performance of NLC engine when varying the polynomial order.

2.2.5.3 Coefficient Sharing Between Channels

The NLC engine was implemented with a separate coefficient set for each channel. As previously mentioned in Section 2.2.3, the VCO ADC voltage-to-frequency nonlinearity sources are the silicon process, the supply voltage, and the operating temperature. Since the supply voltage and operating temperature are well regulated and the same across all channels, the only source of cross-channel variability remains the silicon manufacturing process. If the on-chip process variations were insignificant, we could use a single coefficient set across all ADC channels. For a 32-channel system, this would mean 32x reduction in coefficient memory size. Since this variability can not be simulated or analytically derived, we empirically measured it on fabricated ICs by applying the coefficients derived for one channel across all the other channels. We can see from Figure 2.22 that coefficient sharing among channels can result in up to 45dB SNR losses. For this reason, we kept the original arrangement of allocating separate coefficients for each channel.

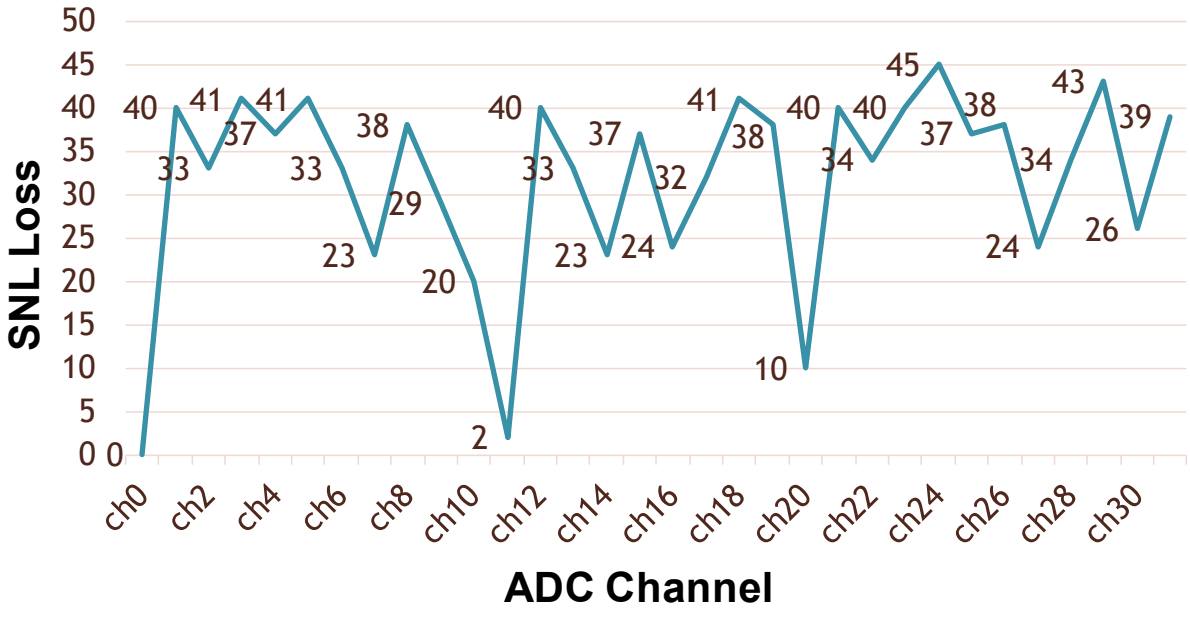

Figure 2.22 Resulting SNR losses when ch0 NLC coefficients are applied across all other channels.

2.2.6 Low-Power Optimizations

Before starting the low power optimizations, power consumption sources of the initial design need to be understood. Table 2.4 shows the power consumption of various digital blocks for the sensing IC including the four (4) NLC engines. About 86% of the power consumption is due to the NLC cores, and 60% of this is due to the leakage component. Thus, reduction of the leakage power of the NLC engines will be the main target for the power optimizations. Main contributors of the leakage power are the area of the design, transistor threshold voltage, and the supply voltage. Since threshold voltage is fixed for the digital standard cells provided by the foundry, proper library selection is key in leakage reduction. Area and the supply voltage have a linear relationship to leakage power, however, lowering the supply voltage also requires additional voltage-translation circuitry (hence power) for interfacing the voltage-scaled block to the rest of the system.

Table 2.4 Power consumption of sensing IC v1 digital blocks operating at 12.288MHz and fabricated with the TSMC45GP process. Cells highlighted in blue and red denote the main power consumption source and the main power component respectively.

Design Block Name	Switching Power (µW)	Internal Power (µW)	Leakage Power (µW)	Total Power (µW)	%
VCO timer	0.428	1.33	8.06	9.83	1.3
PISO	0.236	0.936	2.62	3.79	0.5
SIPO	0.346	1.7	3.55	5.59	0.7
Config. reg.	14.8	44.1	28	86.9	11.5
4 NLC cores	20.67	215.9	412	648	**85.8**
Entire design	36.7	264	455	755	100.0
%	4.9	35.0	**60.3**	100.0	

2.2.6.1 Technology Selection

The initial version of the sensing IC (v1) was fabricated with TSMC40GP fabrication process – where TSMC is the name of the foundry (Taiwan Semiconductor Manufacturing Company), 40 denotes the technology node (40nm) and GP means general-purpose transistors were used. GP device processes are meant for high-performance applications such as CPUs, GPUs, or FPGAs hence their power/delay curve is skewed towards lower delays at the expense of added power. This process was used for the v1 of sensing IC due to availability rather than suitability. This process inherently has higher leakage compared to low-power (LP) process of the same 40nm node due to thinner gate oxide layer. Since high throughput (lower-delay) is not needed in our design, just by porting the v1 design to this process would result in 97.6% less leakage power and reduce the total power consumption by 42.5% (Table 2.5).

Table 2.5 Power consumption of sensing IC v1 digital blocks synthesized at 12.288MHz for the TSMC40LP process. Note the total power of the entire design is 434µW, which is substantially less (42.5%) than in the design fabricated with the TSMC45GP process.

Design Block Name	Switching Power (µW)	Internal Power (µW)	Leakage Power (µW)	Total Power (µW)	%
VCO timer	0.922	1.85	0.208	2.98	0.7
PISO	0.346	1.3	0.061	1.71	0.4
SIPO	0.513	2.36	0.084	2.96	0.7
Config. reg.	22	62.2	0.654	84.8	19.5
4 NLC cores	30.3	301.4	10.08	341.7	78.7
Entire design	54.1	369	11.1	434	100.0
%	12.5	85.0	2.6	100.0	

2.2.6.2 Interleaving

Initial sensing IC version has four (4) parallel NLC engines each of which operates at system clock speed of 12.288MHz. This clock speed selection was made for ease of integration with other digital blocks operating at this frequency. The actual needed clock speed for each engine is the sampling rate of the VCO ADC (6.25kHz) multiplied with (N + 2) cycles, where N is the order of the NLC engine. Since N is six (6) for v1 of the sensing IC, the resulting required operating frequency of each NLC engine will be 50kHz. This operating frequency is about four (4) orders of magnitude slower from the maximum frequency of the TSMC40LP fabrication process resulting in a leakage dominated design where most of the energy is spent on keeping the IC powered rather than doing useful operations. For mitigating this inefficiency, an area-saving architectural transformation such as data stream interleaving can be applied to the design. This architectural transformation combines the data-streams of N channels sequentially and processes them in one processing element (PE). As a result, the logic area of the combined processing elements will decrease at the expense of the increased processing frequency. This relationship is described in Figure 2.23.

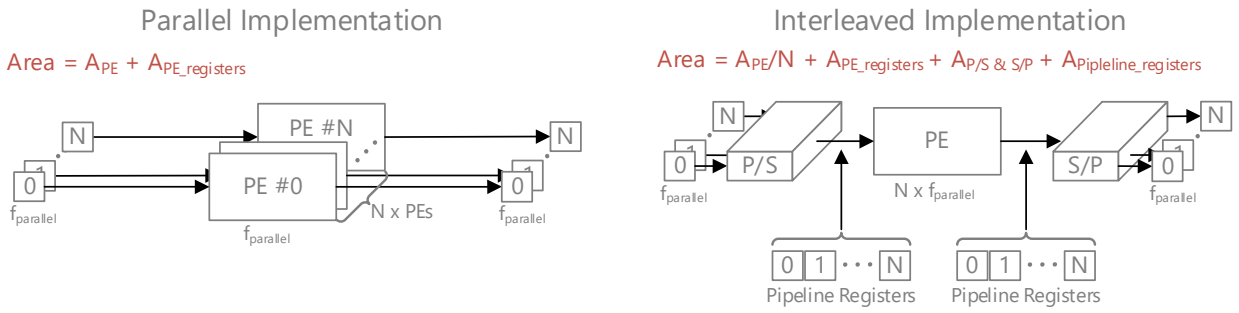

Figure 2.23 Parallel implementation vs. data-stream interleaving architecture.

Since interleaved implementation requires an overhead area for merging and separating the data-streams, the benefits of interleaving will start diminishing after some $N_{optimal}$ level. Figure 2.24

shows the area and power of the parallel, 16x, and 32x interleaved NLC engine implementations. The lower limiting bound of the interleaving level is the pipelining level of the used floating point multiply-accumulate-unit (MAC) which is 15. Any interleaving number below 15 will still take the same area as 15, but the effective per-channel area will be higher because more interleaved engines will be required. Thus, the only possible interleaved NLC engine combinations for a 32-channel system are 1x32 and 2x16. As we can see from Figure 2.24 the 16-level and the 32-level interleaved implementations have comparable power consumptions with the 32-level design having smaller per-channel area. Although the 2 x 16-level interleaved implementation takes more silicon area, from system perspective it is more desirable since one of the 16-level interleaved engines can be disabled when not more than the half of the 32 VCO ADC channels are enabled for sampling.

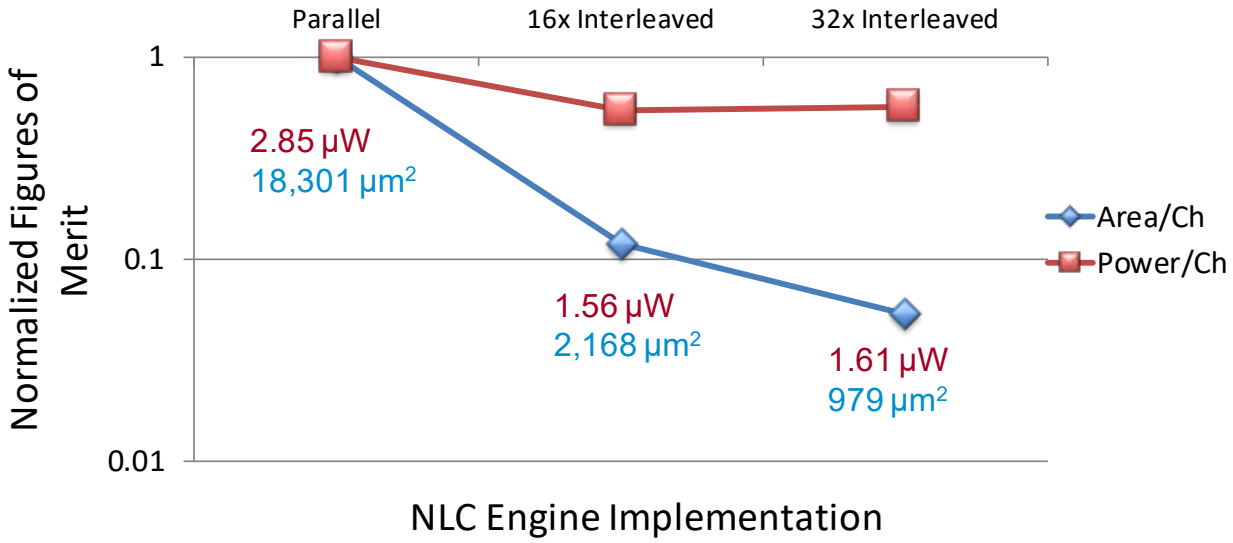

Figure 2.24 Area and power of parallel, 16x interleaved, and 32x interleaved NLC engine designs.

2.2.6.3 Clock Domain Partitioning and Clock Gating

Using a single clock domain across all design blocks results in a fully synchronous design which is easy to implement. The drawback of this approach is the excess dynamic power burnt in modules

which could operate at lower clock speeds. In optimized design, each block runs at the minimum clock speed it requires for operation. The challenge, in this case, is the synchronization between clock domains to avoid metastable states. When the data from one clock domain changes outside the safe setup and hold windows of the sampling flip-flop, the latter may be unable to settle into a stable '0' or '1' logic level before the next clock transition. For one-bit signals crossing clock domains, a simple structure called double flip-flop synchronizer can be used.

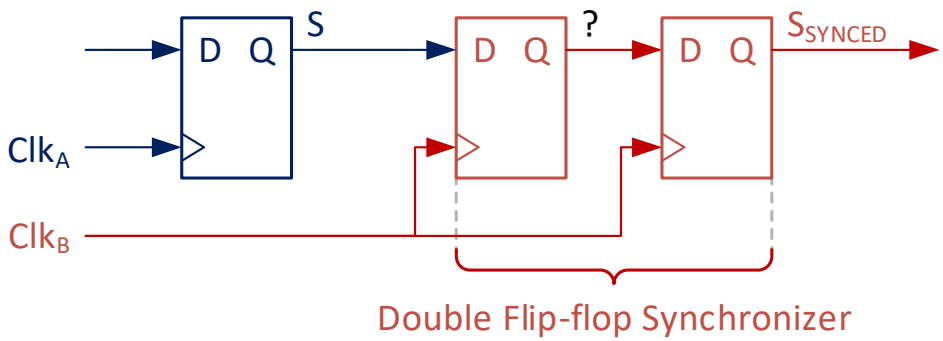

Figure 2.25 Double flip-flop synchronizer.

Using double flip-flop synchronizers for multi-bit signal synchronization can result in invalid values passed across clock domains as each of the bits in the signal can resolve to a different logic state initially. Asynchronous first-in-first-out memories (FIFO), and handshake mechanisms can be used for these cases. MUX recirculation synchronizer is a relatively simpler multi-bit synchronization structure which requires lower overhead. It uses a double flip-flop synchronizer for the "enable" signal when passing the multi-bit data. This "enable" signal will be asserted one cycle after the data transition to ensure data is stable for sampling when the "enable" signal is asserted. When the "enable" signal changes from '0' to '1' however it is resolved to '0' for the initial clock cycle, the MUX in front of the other clock domain will recirculate the stable output

of the flip-flop back into its input. When the "enable" signal is resolved to '1' the data from the cross-clock domain will already be stable for sampling, and no metastable state will arise when sampling it.

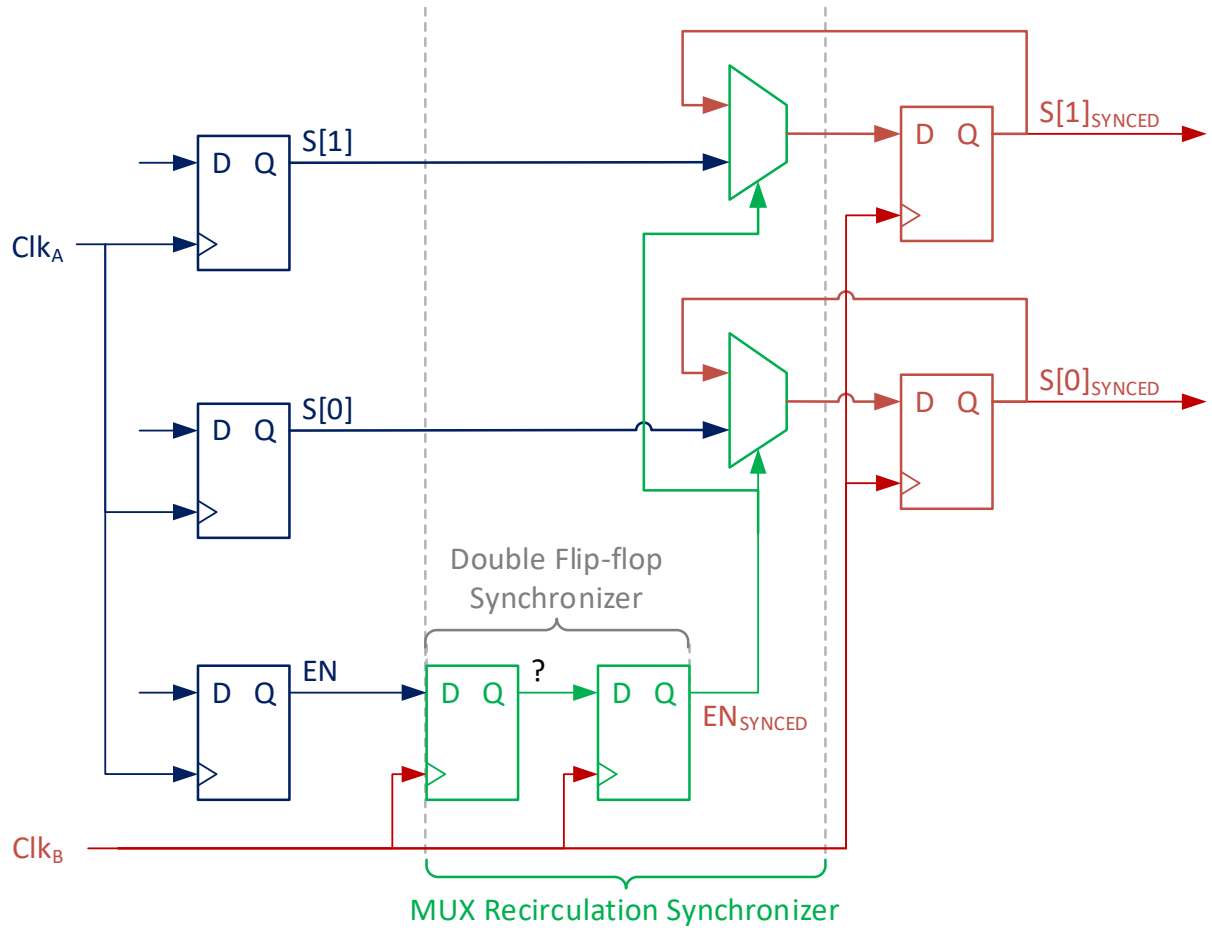

Figure 2.26 Architecture of the MUX recirculation synchronizer.

To further reduce the power consumption of the digital subsystems, they were partitioned into six (6) separate clock domains each running at their minimum required clock speeds (Figure 2.27). Double flip-flop and MUX recirculation synchronization techniques were used for passing single-bit and multi-bit data across the clock domains.

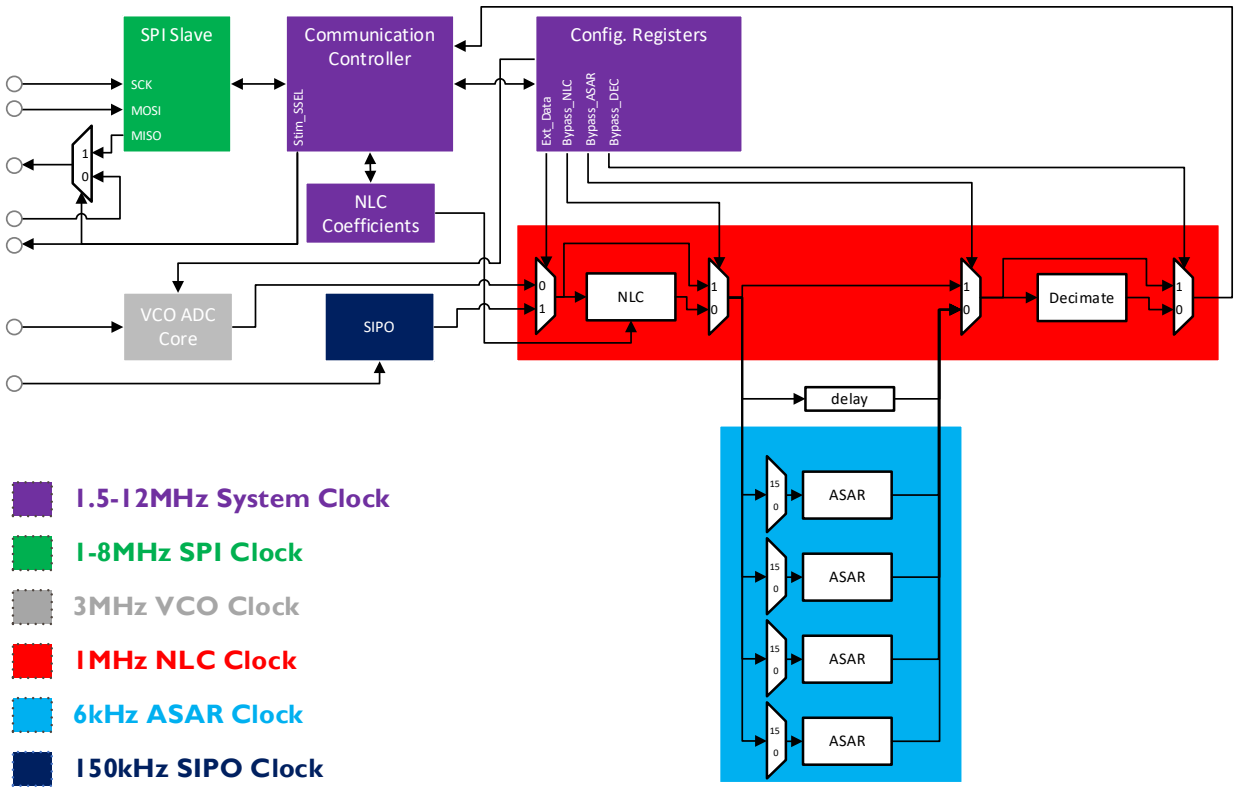

Figure 2.27 Clock domain partitioning of sensing IC's digital subsystem.

When a digital subcircuit is idling, the clock tree buffers consume significant power even if the data fed into the flip-flops do not change. Clock tree power consumption can comprise up to 50% of the total dynamic power in a modern process. Some parts of the digital subsystem such as configuration registers need to be programmed once or very few times. If the clock is fed to configuration registers all the time, however, the capacitances of that subsection of the clock tree will be charging and discharging even though the data fed to the flip-flop remains unchanged. If we can enable the clock to the subtree based on the data or "enable" signal we can save significant energy for low activity subcircuits. Modern electronic design automation (EDA) tools allow for automatic clock gating based on explicitly specified enable signals or based on data input. This technique was also used for the sensing IC's digital subcircuit optimization and resulted in about

60% power savings compared to non-clock gated design. Figure 2.28 shows the power consumption of the interleaved NLC engines and other digital subsystems after clock domain partitioning and clock gating.

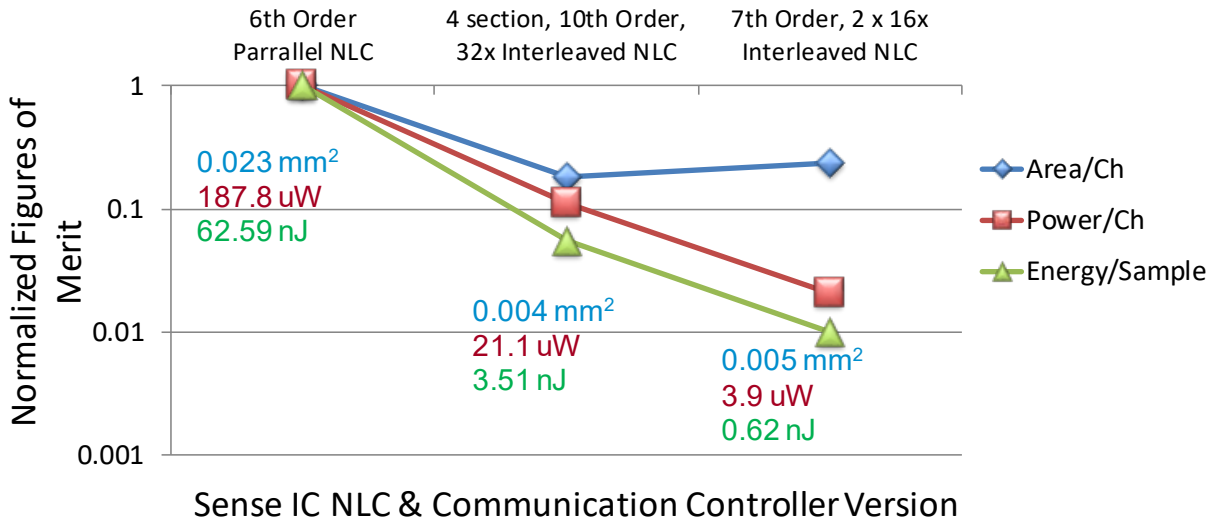

Figure 2.28 Sensing IC power consumption after clock domain partitioning and clock gating.

2.2.6.4 Comparison with State-of-the-art

When we compare our VCO based sensing front-end to state-of-the-art biosignal recording front-ends in terms of power consumption, input range, and linearity, for comparable power budget our design achieves 10x more input range and 21dB more linearity. This linear input range allows recording of LFP signals in the presence of large interferers and is largely enabled by the digital nonlinearity calibration. The price we pay for this usability is 4µW of additional power per channel and a slight increase in ADC core area (0.007 mm^2 out of 0.12mm^2 total area). It is worth to note that our NLC engine is running off of a 1V power supply for the practicality of system integration. Due to a large design delay slack, we could have easily voltage scaled and powered the NLC engine from a much lower supply. For 0.5V operation, for example, the standalone NLC power

consumption would have gone down to 1µW (4x). While from individual block perspective this is advantageous, from a system perspective, the caveat is that the low-drop-out (LDO) voltage regulator needed to generate the 0.5V and the level shifters needed for logic level translation would add power overhead. The use of LDO voltage regulator would also limit the maximum power savings to 2x (vs. 4x). Thus, the benefits of voltage scaling optimization would be too low for the implementation costs from the system stand point.

Table 2.6 Comparison of our VCO-based sensing front-end with state-of-the-art biosignal front-ends. Cells highlighted in blue denote the linear input range allowing recording of LFP signals in the presence of large interferers and the price we pay in terms of power to achieve it.

Reference	[21] **JSSC'07**	[19] **ISSCC'14**	[30] **JSSC'16**	**This work**
Topology	CCIA (no ADC)	Chop. Amp + ADC	Chop. Amp + MADC	**Direct digitization VCO with NLC**
Signals	LFP	LFP	Spike/LFP	**LFP**
Technology	0.8µm	65nm	0.13µm	**40nm**
Supply	1.8V	0.5V	1.2V	**1.2V** (Analog) **1V** (Digital)
Area/Ch.	1.7 mm^2	0.025 mm^2	0.018 mm^2	**0.12 mm^2**
Power/Ch.	2µW	2.3µW	9.1µW	**3µW** (Analog) **3.9µW** (Digital)
In-ref. Noise	1µV	1.3µV	4.2µV	**4µV**
Peak Input	5mVp	±0.5mV	1mVpp	**±50mV (10x↑)**
SFDR	60dB	52dB	50dB	**81dB (21dB↑)**
Z_{in}(DC)	8MΩ	28MΩ	∞	∞ **(No C_{ext})**

2.2.7 Support for "Blanking" & Identification of "Blanked" Samples

While simultaneous sensing of stimulation artifacts sets our sensing front-end apart from other biosignal recording systems, it introduces a need for digital signal processing (DSP) algorithms to remove this large interferer from the useful neural signal before the neural signal content can be analyzed. In case neural activation pattern during stimulation is not of interest for an application, our sensing front can be used in traditional "blanking" mode where neural signals are sensed only when stimulation is off. To identify the "blanked" samples, sensing IC has an input for a trigger signal. When this trigger signal is logic high, the MSB of the ADC signals will be marked with "1". This feature can be used for two purposes – 1) marking the samples with stimulation artifacts for "removal" for artifact-free processing, and 2) synchronizing samples with some external event (e.g., behavioral experiment). For the "blanked" sample identification, the stimulation IC will need to have a trigger signal denoting the stimulation "on-time."

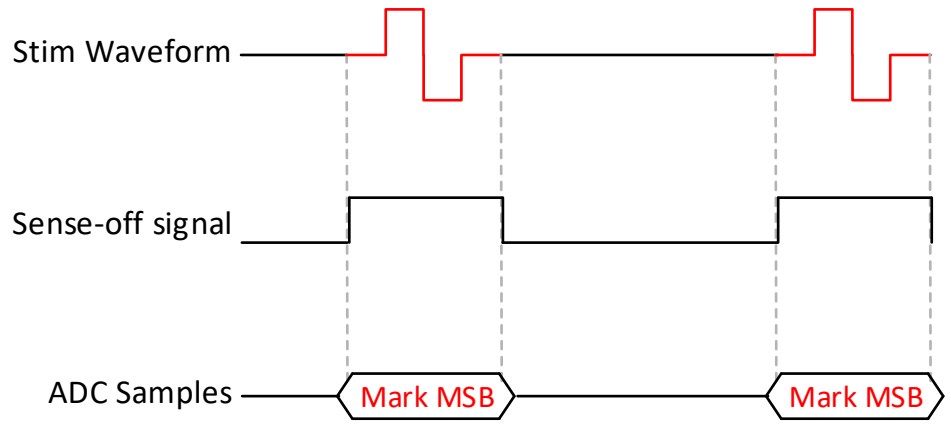

Figure 2.29 Marking ADC samples based on a trigger signal (sense-off).

2.2.8 Decimation Block for Data Reduction

The LFP band spans from low-frequency delta oscillations (1-4Hz) up to high-gamma oscillations (60-200Hz). If a sharp anti-aliasing filter is used, a 400Hz sampling rate should be sufficient for

capturing the entire LFP band. Stimulation artifacts have higher frequency signal components (around 1kHz) due to the sharp rectangular edges of stimulation waveforms. When factoring-in these stimulation artifacts, the required Nyquist sampling rate will be around 2kHz. Our VCO-based sensing front-end nominal sampling rate is 6.25kHz for capturing neural signals and stimulation artifacts without aliasing. When stimulation is off, or when stimulation artifact components are removed by the built-in adaptive stimulation artifact rejection (ASAR) DSP block, the sampling rate can be reduced using the decimation block (Figure 2.30). Supported decimation degrees and effective sampling rates are listed in Table 2.7. The ASAR block has been designed by Sina Basir Kazeruni and will not be discussed in this work.

Table 2.7 Effective sampling rate after decimation.

Decimation Degree	Effective Sampling Rate (Hz)
2	3,125
4	1,5625
8	781.25
16	390.625
32	195.3125
64	97.65625
128	48.828125

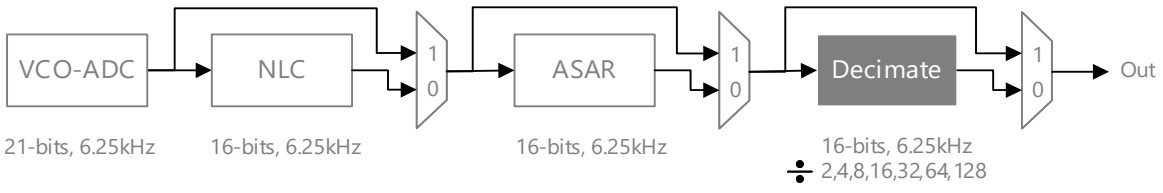

Figure 2.30 Decimation block in the sensing IC signal chain.

2.2.9 IC Micrographs

The v1, v2, and v3 of the sensing IC were fabricated in the GP, LP, and LP flavors of TSMC 1P10M 40-nm CMOS technology respectively. The chip micrographs are shown in Figure 2.31.

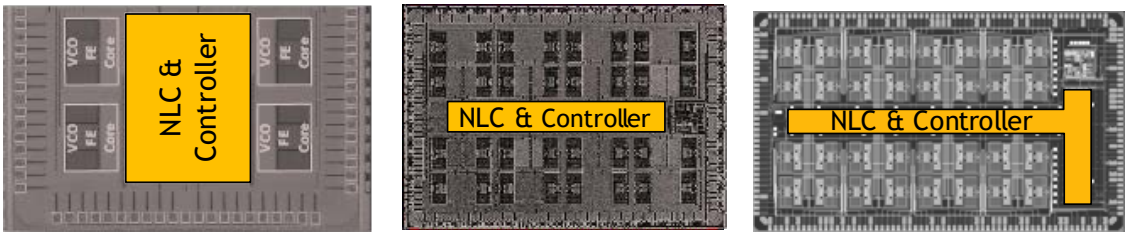

Figure 2.31 IC micrographs of sensing IC v1, v2, and v3 (left to right). NLC and other digital blocks highlighted in orange.

In the sensing IC v1, four front-ends have been implemented with parallel 6^{th} order NLC implementation. This version of the IC uses four-wire SPI master interface to transmit the read samples. The v2 of the sensing IC uses an SPI slave interface, which is more appropriate for sensor ICs. The NLC has four (4) sections of 10^{th} order engines and is 32-level interleaved in this version. Due to this area-efficient implementation, the effective NLC area ($0.035mm^2$) is ten times smaller than the area of the LUT approach with $2^{17.4}$ words and 15 bits of word length (assuming 0.124 μm^2 for a 6T SRAM cell[29]). Other system level blocks such as a power-on-reset circuit (designed by Hariprasad Chandrakumar), a crystal oscillator (designed by Hariprasad Chandrakumar), low dropout voltage regulators (designed by Hariprasad Chandrakumar and Weiyu Leng) were also integrated into this version for reducing the number of external off-the-shelf components required for the IC operation. In v3 of the sensing IC the NLC order was finalized to seven (7) and instead of single 32-level interleaved implementation two 16-level interleaved engines were used. This version supports a more robust communication protocol (acknowledgments and CRC checks were

added) and underwent clock domain optimization (from one to six domains). On the signal processing chain side, v3 also includes an adaptive stimulation artifact rejection engine (designed by Sina Basir Kazeruni) and a decimation block for sampling rate reduction.

2.3 Stimulation IC

This section presents the design of our stimulation IC paying attention to special considerations pertaining to closed-loop neurostimulators such as patient safety and waveform parameter updates. Our non-rectangular waveform generation mechanism is also discussed.

2.3.1 Stimulation Parameter Update Considerations for Closed-Loop Implants

Open loop stimulation is typically in an always on state. Some implants allow the patient to make changes within an allowable parameter range; however, these changes are relatively infrequent - a handful of times during the day. In closed-loop implants, the need to change stimulation parameters may be much more frequent depending on the disease state. Waveform update requests may be received every several seconds, and when they fall in between anodic and cathodic phases, they may result in charge-imbalanced waveforms (Figure 2.32). If these charge-imbalanced waveforms occur frequently, permanent tissue damage may occur over time due to oxidation-reduction reactions at the electrode-tissue interface.

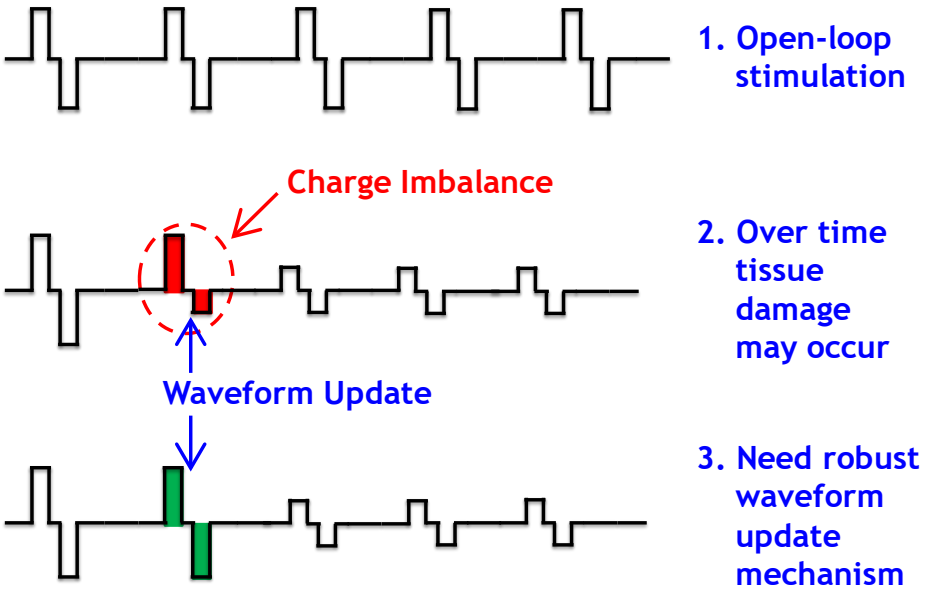

Figure 2.32 Possibility of creating charge-imbalanced stimulation waveforms during frequent stimulation parameter updates.

Even when the waveform update request is received before the first stimulation phase, there is still a possibility for creating a charge-imbalanced waveform. Since stimulation waveforms are described by multi-bit parameters, parallel parameter updates are practically not feasible. Hence, new stimulation parameters are shifted-in and applied serially (Figure 2.33).

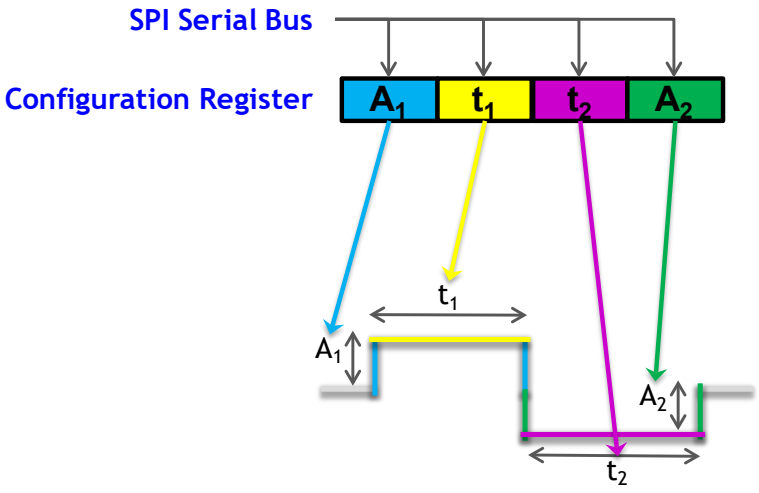

Figure 2.33 Serial update of stimulation parameters.

If the configuration registers storing the pulse timing and amplitude information are directly connected to the pulse generation engine, partial waveform updates may occur even when the update request occurs right before the first phase of the waveform. This can occur if the time required for writing the parameter set is greater than the generated duration of the waveform.

To address issues due to these potentially frequent waveform updates, we need to ensure a safe mechanism of updating waveform parameters. This means that all parameters related to an update request should be written quick enough and on the correct phase of the waveform.

2.3.2 Safe Waveform Parameter Updates in Commercial Off-the-Shelf Stimulation ICs

There is very little public information available about the stimulator ICs used in commercial implantable pulse generators (IPG). Design details are typically guarded from the competitors. For this reason, in this section, we will explore what it takes to safely update the waveform parameters in two commercially available stimulator ICs – 1) Intan technologies RHS2116 and 2) Cactus semiconductor CSI021.

2.3.2.1 Intan Technologies RHS2116 Stimulator/Amplifier

RHS2116 is a recent addition to Intan's portfolio and compared to other Intan amplifier ICs it includes a neural stimulator function. It is mainly intended for multichannel headstages used in electrophysiology experiments. In order to generate stimulation waveforms, an SPI master needs to program the "phase 1" and "phase 2" amplitudes, then constantly send commands over SPI interface enabling and disabling the stimulator engines and programming their polarity (anodic or cathodic).

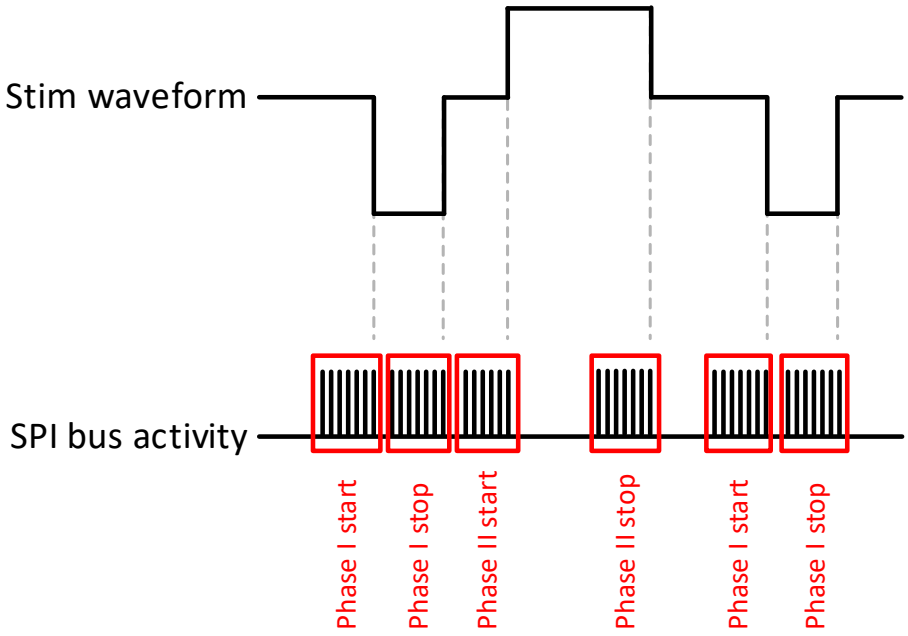

Figure 2.34 Stimulation waveform generation with RHS2116.

Thus, the SPI master is intrinsically aware of the stimulation waveform phases as it is the one generating them. This makes safe updates of stimulation parameters very easy. After the SPI master finishes generating the last phase of the stimulation pulse, it can take all the time to update the waveform parameters before it initiates the next stimulation pulse. Due to the absence of pulse generation timers, pulse generation step-size and precision can vary from pulse to pulse and are largely dependent on the SPI speed, and the number of sensing channels enabled for readout. The interrupts received by the processor controlling the SPI master peripheral, for servicing tasks other than constantly reading/writing from RHS2116, can also affect the aforementioned precision. From a safety perspective, if the SPI master firmware loses control after enabling a stimulation phase, neural tissue will be damaged by constantly receiving unbalanced stimulation. Lastly, stimulation waveform timing generation through constant SPI commands will require ample power, which will drain the battery quickly. This is due to two factors – 1) high

activity factor on the SPI bus, and 2) SPI bus IO voltage (e.g., 1.8V) is higher than the core voltage required for internal timers (e.g., 1V).

2.3.2.2 Cactus Semiconductor CSI021

Unlike RHS2116, Cactus semiconductor CSI021 target market includes implantable neural stimulators, and it has built-in timing generation for stimulation pulses. Usage of internal timers means that after enabling the stimulation, the SPI master will not know what is the current phase of the stimulation waveform to apply the new parameters safely. In order to identify the waveform phase for a given stimulation channel, the SPI master can poll the appropriate bit of the status register (Figure 2.35). Once the end of the stimulation pulse is detected, the SPI master can write the new configuration. This write should be completed before the next stimulation pulse; hence much higher SPI communication speed (1-10MHz) is required compared to the internal pulse generator clock (10-400kHz). Thus, although feasible, safe parameter update implementation with CSI021 will require some power overhead due to the energy spent on status register polling.

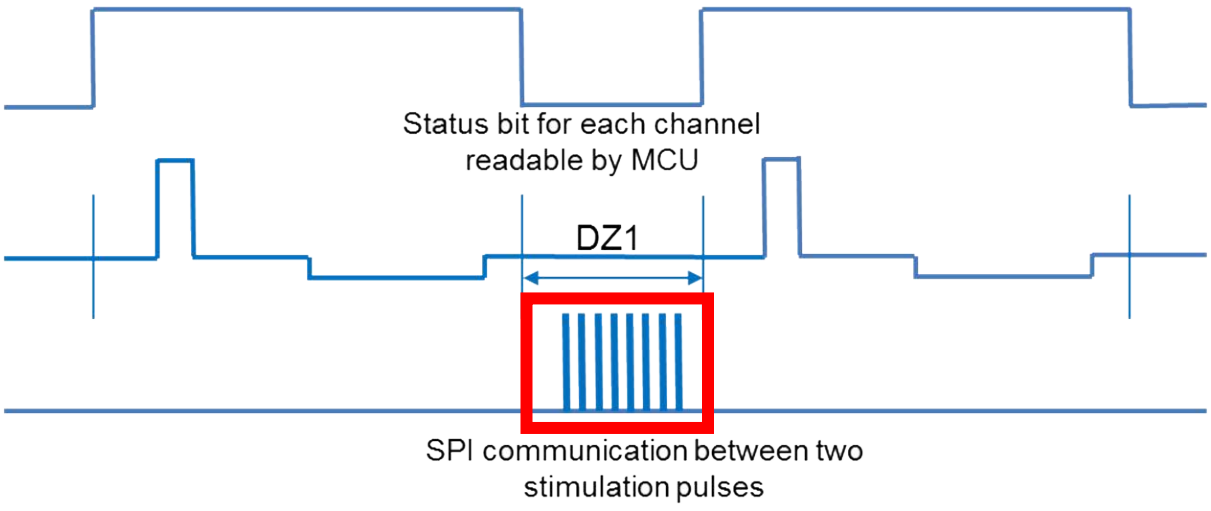

Figure 2.35 Safe parameter update sequence in CSI021.

2.3.3 The Proposed Safe Waveform Parameter Update Mechanism

The power overhead due to polling, described for CSI021, can be easily eliminated if the system can accommodate an additional interrupt line per stimulation channel/engine between the stimulation IC and the SPI master. At best, one interrupt line is required followed by one status register read to identify the interrupt source (stimulation channel/engine). We implemented a different scheme for safe parameter updates, which does not require status register polling or additional line(s) for waveform phase identification and does not require a high-speed SPI bus to complete the write operation before the next stimulation pulse.

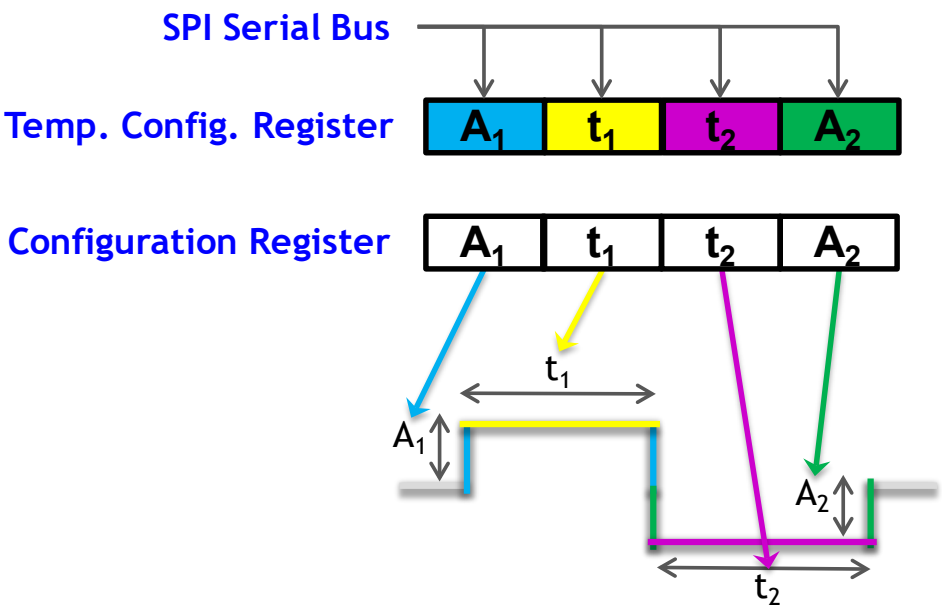

Figure 2.36 Stimulation IC double-buffered configuration register illustration.

To ensure the serial writes do not disturb the currently running stimulation pulses, we first decouple the serially updating configuration register from the configuration register attached to the waveform generation engine (Figure 2.36). When the serial updates to the temporary register are done, we can then transfer it in its entirety to the configuration register controlling the pulse

generation. Conceptually this is similar to double buffering in computer science to eliminate the flickering due to partial screen updates. This is achieved by swapping the "back" buffer with the "front" buffer when the "back" buffer writes are complete, and rendering the content of the "front" buffer as a single image. The key, in this case, is the time when the temporary register is applied to the configuration register. To indicate to the stimulation IC that we are done writing to the temporary register, we send a special SPI command. This command sets a "transfer" flag which stays active until the end of the current stimulation pulse. The temp registers are then transferred to the configuration registers controlling the pulse generation, and the "transfer" flag is cleared (Figure 2.37).

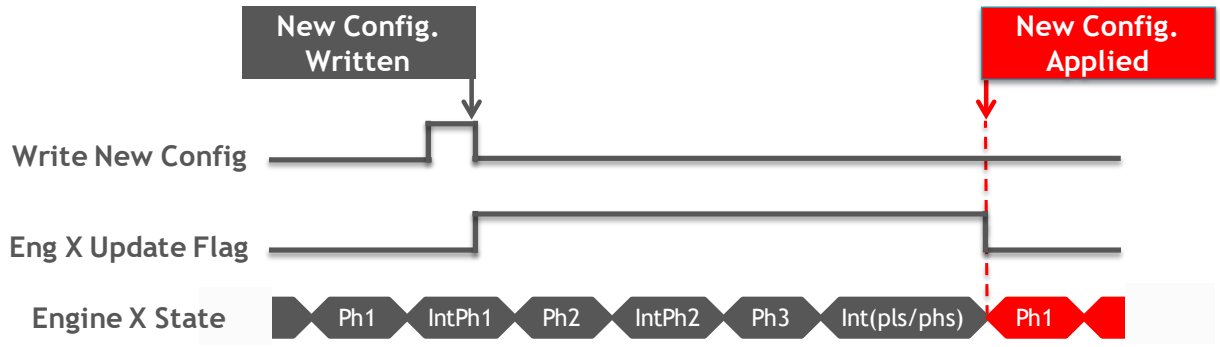

Figure 2.37 Temporary-to-configuration register transfer mechanism.

This approach is low power and very robust as writes can occur anytime and at any SPI speeds while preserving the atomicity of the parameter set describing the entire waveform.

2.3.4 Non-Rectangular Waveform Generation

According to recent studies, some non-rectangular waveforms can achieve the same neural activation levels with less energy. For generating a conventional rectangular pulse, two parameters need to be stored per each of the anodic and cathodic phases – the duration and the amplitude.

The duration is generated with some fixed temporal resolution and the amplitude stays constant during the entire pulse time. If we can also modify the amplitude at each of the time-steps of the pulse duration we can generate step-line-like non-rectangular waveforms (Figure 2.38).

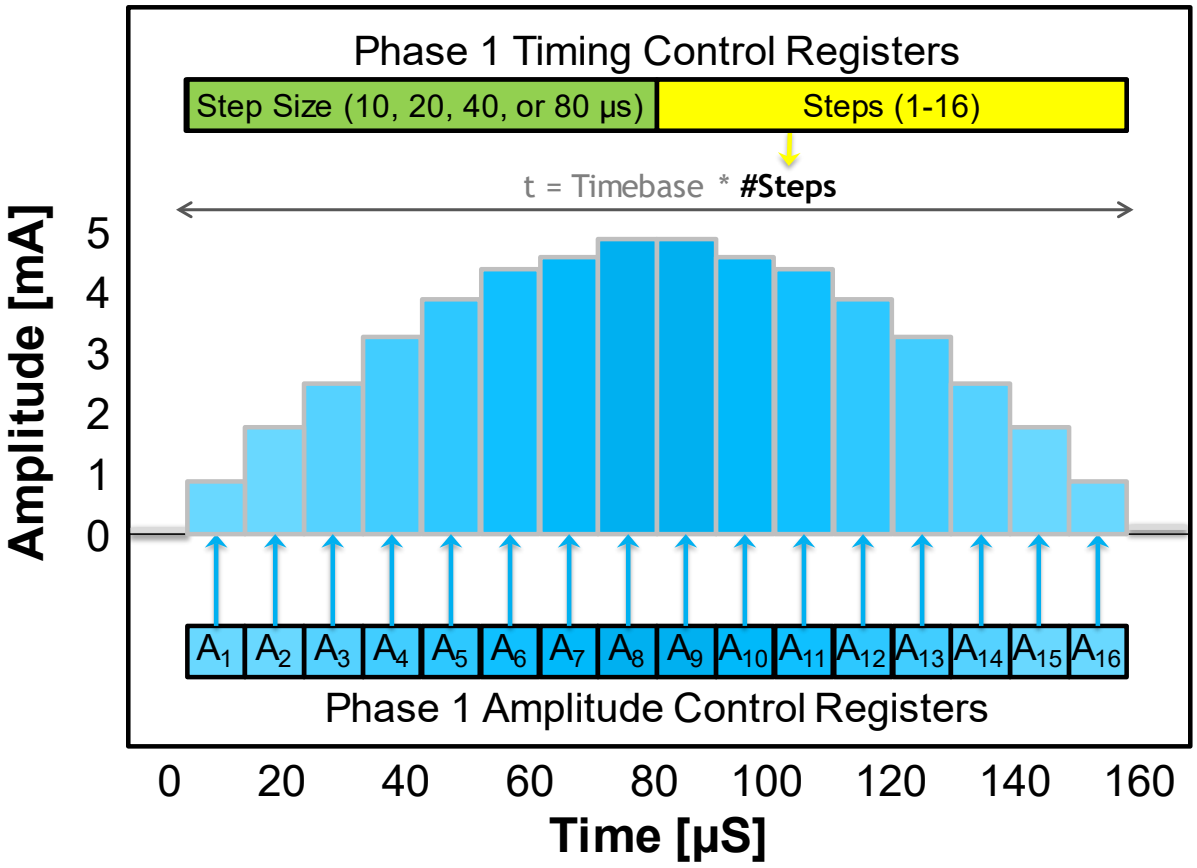

Figure 2.38 Generation of a custom, non-rectangular pulse. The amplitude is modified based on stored values at each time-step spanning the pulse. The time-step resolution is also programmable.

The price we pay for this flexibility is the additional memory for storing the time-step amplitudes. In our version of custom waveform generator, we also made the time-step size programmable for even more flexibility. Figure 2.39 shows some examples of non-rectangular waveforms generated by our 16-step, 8-bit amplitude stimulation IC.

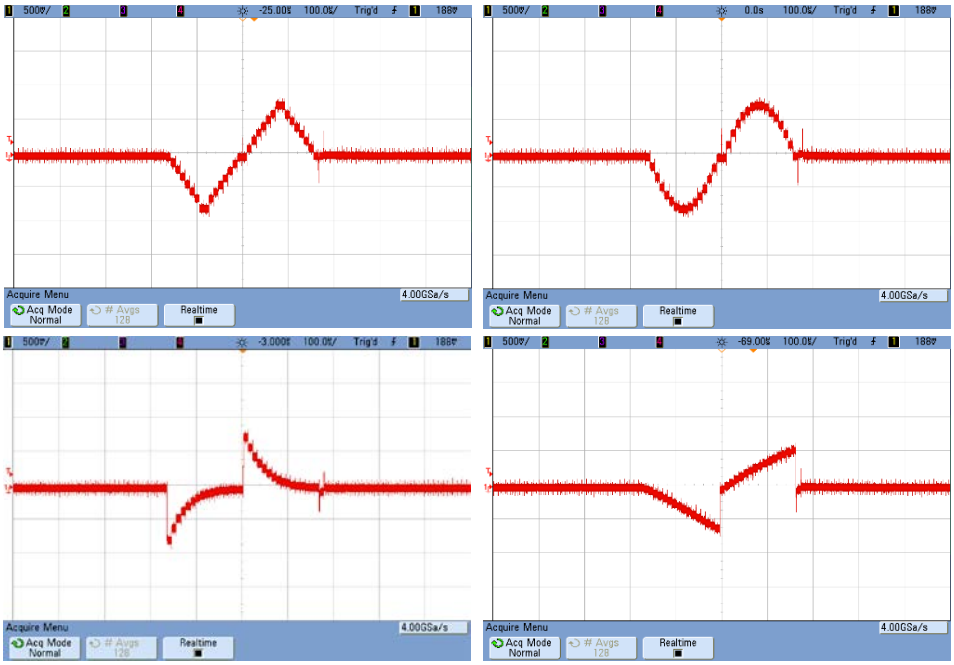

Figure 2.39 Examples of custom, non-rectangular waveforms generated by our stimulation IC.

2.3.5 Stimulation Waveform Programmability

Although anodic and cathodic pulse duration and amplitude are the most important parameters, stimulaiton IC pulse generator takes ten more parameters for shaping the entire waveform (Figure 2.40). These additional parameters add flexibility for application specific waveform generation. The timing control for each engine is independent of all the other engines. This means that each engine can run a pulse train with different pulse durations and frequencies. For generating a pulse train that repeats indefinitely, the "inter-pulse delay" should be adjusted for desired frequency, and the "pulse count" should be set to "0". Bursts of pulses are supported through the "inter-burst delay" and the "burst count" parameters. When instead of indefinitely repeating patterns one would like to generate a specific number of pulses and/or bursts, "pulse count" and "burst count" parameters should be set to non-zero values. Pulse generation engine also supports optional delays between "phase 1" and "phase 2" ("inter-phase delay 1"), and between "phase 2" and "phase 3"

("inter-phase delay 2"). "Phase 3" is the charge recovery phase and is used to remove residual charges built up on the electrode-electrolyte interface after anodic and cathodic phases if they do not happen to be perfectly balanced. Charge imbalance could arise even if the charge-balanced stimulation protocol is chosen due to variations in transistor characteristics across the chip. In the recovery phase, anodic and cathodic electrodes are briefly shorted to a ground potential.

Table 2.8 Stimulation waveform programmability.

Parameter	Value
Phase 1 amplitude	20µA *(0 to 254)
Phase 2 amplitude	20µA *(0 to 254)
Sense-off delay	20µs * (1 to 16)
Phase 1 duration	(10, 20, 40 or 80) µs * (1 to 16)
Inter-phase delay 1	10µs * (0 to 15)
Phase 2 duration	(10, 20, 40 or 80) µs * (1 to 16)
Inter-phase delay 2	10µs * (0 to 15)
Phase 3 duration (charge recovery phase)	10µs + 640µs * (0 to 15)
Inter-pulse delay	10µs + 320µs * (0 to 255)
Pulse count	1 to 255 + 0 = continuous
Burst count	1 to 8191 + 0 = continuous
Inter-burst delay	81,920µs + 1280µs*(0 to 255)

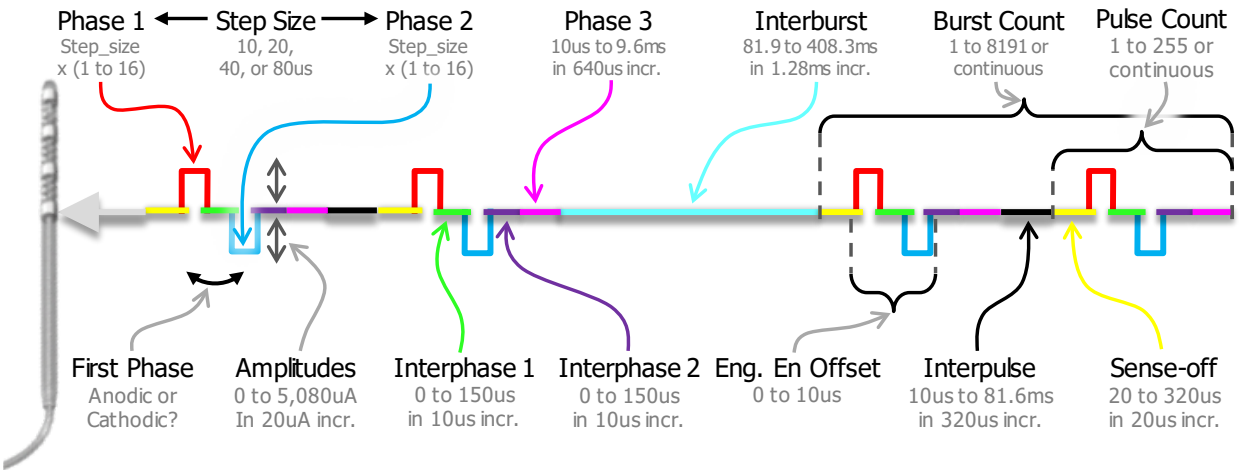

Figure 2.40 Stimulation waveform programmability.

2.3.6 Patient Safety Features Built into the IC Communication Protocol

In our DBS system architecture, smart leads communicate with the control module through relatively long cables (50cm). There could be cases when CM and NMU modules receive something other than what was transmitted due to influences of external interferes. If transmission errors go unnoticed, results can be devastating for patients' health. To mitigate this risk, three features are built into the stimulation IC communication protocol for bidirectional detection of transmission errors – 1) command and argument checks, 2) acknowledgment replies and 3) cyclic redundancy checks (CRC).

Communication protocol consists of command, argument, data, and CRC bytes. When an SPI master transmits the command and argument bytes of the packet, stimulation IC checks whether a valid command byte is received and whether argument falls within the allowable range. If an error is detected, the stimulation IC FSM enters into the GENERAL_ERROR state. If the RESET_ON_ERROR bit of system configuration register is set (default value after reset), the stimulation IC will reset on the GENERAL_ERROR event and turn off all active stimulation engines. If the RESET_ON_ERROR bit is not set, stimulation IC will ignore the rest of the packet bytes then return to the IDLE state when the SPI_SSEL line is de-asserted.

If the stimulation IC does not detect any errors with the command and argument bytes, it replies back to the SPI master with an acknowledgment byte (Figure 2.41). This acknowledgment byte is unique for each command, its arguments, data, and CRC bytes. This uniqueness helps the SPI master to identify whether the stimulation IC FSM is in a wrong internal state. For example, when the SPI master receives an acknowledgment byte for a data byte (DATA_ACK_BYTE)

instead of a CRC_ACK_BYTE it knows that the stimulation IC is in an erroneous state and can reset it.

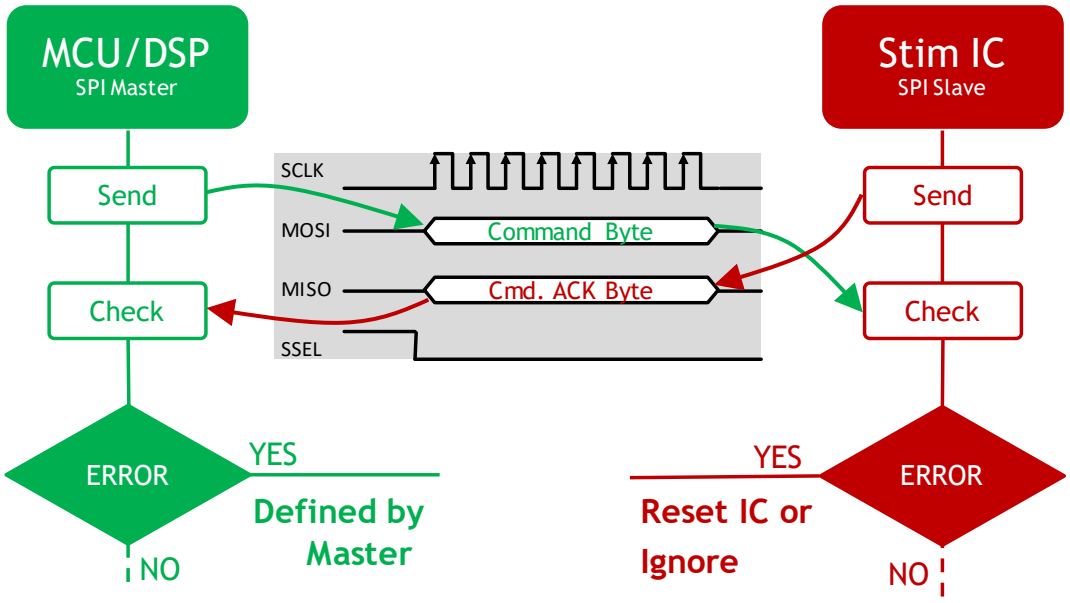

Figure 2.41 Bidirectional error detection enabled by command, argument, and acknowledgment bytes checks.

Commands and arguments are easy to check for errors because they belong to a limited set (there are only eight valid stimulation IC commands). To enable bidirectional error checking for data bytes during register read and write operations CRC checksums are employed. For the register write command, the SPI master transmits the data bytes followed by their two-byte CRC checksum. When receiving these data bytes, the stimulation IC computes an internal CRC checksum for the received bytes and compares these bytes with the received checksum and transmits the comparison result back to the SPI master (Figure 2.42). When comparison fails, the stimulation IC enters into the CRC_ERROR state after transmitting the comparison result. If the RESET_ON_CRC_ERROR bit of the system configuration register is set (the default value after reset), the stimulation IC will reset on the CRC_ERROR event and turn off all active

stimulation engines. If the RESET_ON_CRC_ERROR bit is not set, the stimulation IC will ignore the rest of the packet bytes then return to the IDLE state when the SPI_SSEL line is de-asserted. In the latter case, the SPI master can still act upon the CRC_ERROR condition because it receives the comparison result. For the register read command, the stimulation IC transmits the requested data followed by their two-byte CRC checksum. The SPI master can compute a local checksum of the received data and compare it with the received CRC checksum, then act upon CRC_ERROR conditions as dictated by the application logic.

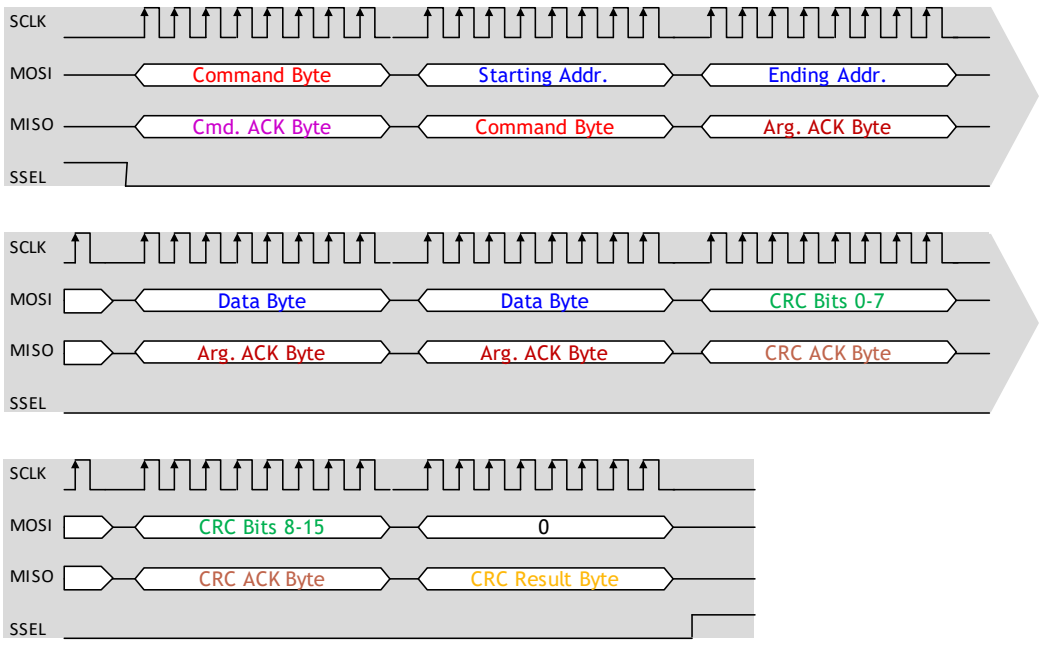

Figure 2.42 Packet structure for register write command. Data bytes are followed by two CRC bytes.

2.3.7 Sensing Switch-Matrix Control and Support for "Blanked" Sensing

The stimulation IC employs sensing and stimulation switch-matrices to give the stimulation engines and sensing IC access to the neural lead electrodes (Figure 2.43). Accessing electrodes through the sensing switch-matrix is important in case the biosignal recording IC cannot handle

the high voltages generated during the stimulation phases. The integrated sensing switch-matrix can also be used with the biosignal recording ICs that can handle the large stimulation voltage swings at their inputs, however, are configured to do a "blanked" recording (no recording during stimulation). The internal pulse generator generates the control signals for both the stimulation and sensing switch-matrices (stim_en, and sense_en). For sensing IC input protection, there is a programmable dead-time (20-320uS) between the stim_en and sense_en signals. The complement of the sense_en signal (sense_off) is routed to a pad for sensing IC consumption. This signal notifies the sensing IC when the sensing switch-matrix is off so that the sensing IC can mark samples invalid (Figure 2.29) or stop recording during the off-period. For simultaneous stimulation and sensing application, the sense_en signal can be forced to stay enabled during the stimulation phases by setting a configuration register bit.

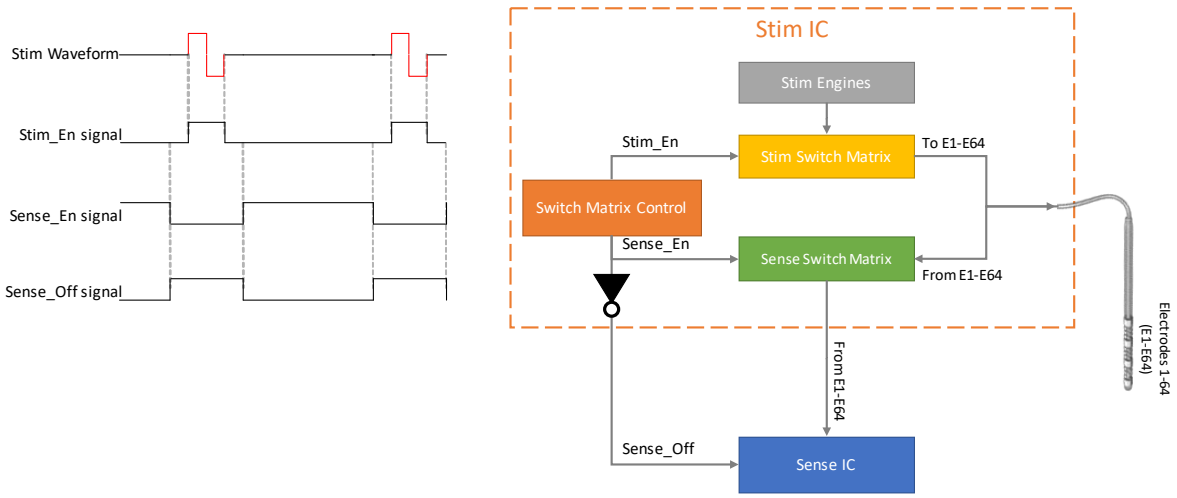

Figure 2.43 Sensing and stimulation switch-matrices integrated into the stimulation IC.

2.3.8 IC Micrographs and Low-Power Optimizations

The four versions of the stimulation IC were designed and fabricated in XFAB's 0.18µm XH018 high-voltage CMOS process. In the chip micrographs shown in Figure 2.44 the digital subsystem

discussed in this work is highlighted in orange, while the analog blocks in the black-and-white region were designed by colleagues Dejan Rozgic, and Ippei Akita.

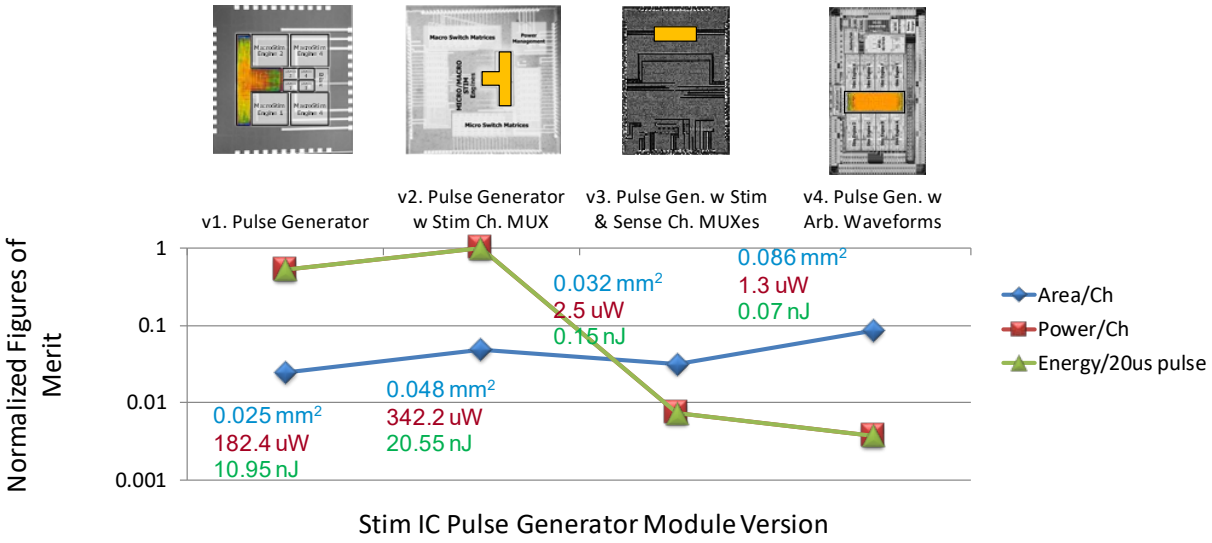

Figure 2.44 Micrographs of stimulation IC v1, v2, v3, and v4 (top-left to top-right) along with their power and area figures (bottom chart). Digital subsystem is highlighted in orange, the analog blocks in black-and-white are designed by Dejan Rozgic.

The high-power consumption in v1 and v2 is due to a high system clock speed (12.288MHz) which was chosen for ease of integration with sensing IC running at the same speed. The almost doubling of power consumption in v2 is due to added programming granularity and the addition of switch-matrix control block. In v3 the sensing-switch-matrix control block was added, and the system clock was reduced to 100kHz – the minimum required speed to generate delays with 10uS timing resolution. This clock speed optimization resulted in about 140x reduction in power consumption compared to v2. Non-rectangular waveform generation, power management control, and robust communication protocol discussed in Section 2.3.6 were introduced in v4 of the stimulation IC for half the power consumption of v3 thanks to the clock gating optimization.

3 System Verification

This chapter focuses on the verification of the designed closed-loop ready high-channel-count neuromodulation unit. Section 3.1 presents the NMU performance tests while Section 3.2 describes the hardware, firmware, and software designed for the NMU verification.

3.1 Performance Tests

When comparing proposed NMU with existing systems we claimed that it stands out due to 1) its ability to capture neural signals of interest and stimulation artifacts without saturation, 2) its high-channel-count, and 3) its ability to generate custom, non-rectangular waveforms. Since the channel-count of our system was discussed in Section 0, and our neurostimulator IC's ability to generate various custom shapes was demonstrated in Section 2.3.4, this section focuses on our high-dynamic-range VCO ADC performance.

3.1.1 VCO ADC Dynamic-Range and Linearity

High dynamic-range is important for simultaneous sensing and stimulation application and is the ratio of the largest (stimulation artifacts) and smallest (neural signals) signals that the sensing front-end can faithfully reproduce. In a perfectly linear ADC, each tone supplied at the input will produce an output at the same tone with the gain of the ADC. Nonlinearities in the ADC will lead to additional tones at the output (harmonics) occurring at multiples of the input tone. If these harmonics are buried in the noise floor of the ADC then the ADC dynamic range is limited by noise; otherwise, the nonlinearity of the ADC is the limiting factor of the dynamic range. Given the VCO based nature of our ADC, it is only expected that the dynamic range limitation is due

to nonlinearities rather than noise. One way to quantify the linearity of our sensing front end is through the spurious-free dynamic range (SFDR), which is the ratio of the fundamental tone to strongest spurious signal (harmonic) in the output. Figure 3.1 shows the spectrum of the VCO ADC in response to two different 50mV tones – 37Hz and 203Hz. These signals have been carefully chosen to eliminate the 60Hz interference in the calculations and result in 88dB and 79dB spurious free dynamic range at the output respectively. When integrating the noise and harmonics in the LFP band of interest (0-250Hz), about 12 bits of ENOB can be expected from the ADC.

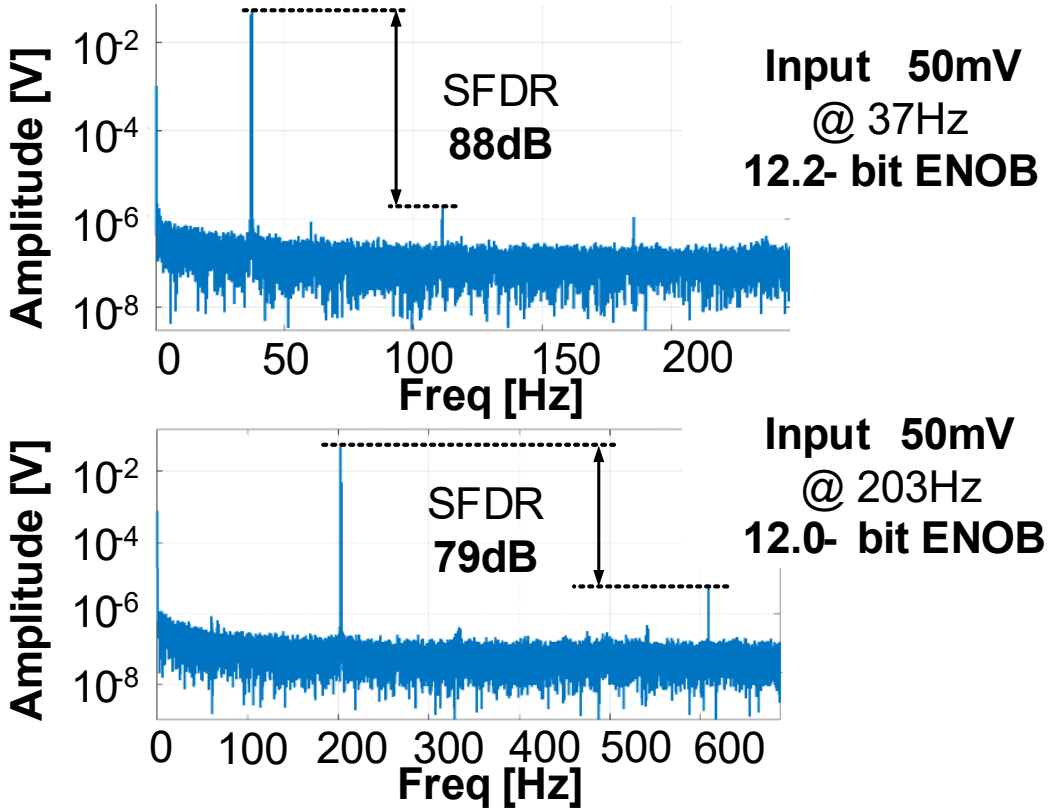

Figure 3.1 Single-tone test for linearity.

Single tone tests are good for quantifying the linearity of the ADC, however, in practice more than a single tone is going to be present in the signal. In these cases, the output will be contaminated with harmonics of the sum or difference of these tones (intermodulation

components). Two-tone tests can be used to produce these intermodulation components. Figure 3.2 shows the output spectrum of the VCO ADC with NLC engine turned off (gray line) and on (red line). We can see that the NLC engine suppressed the distortion components (1.5µV) below the 4µV input referred noise figure of the ADC.

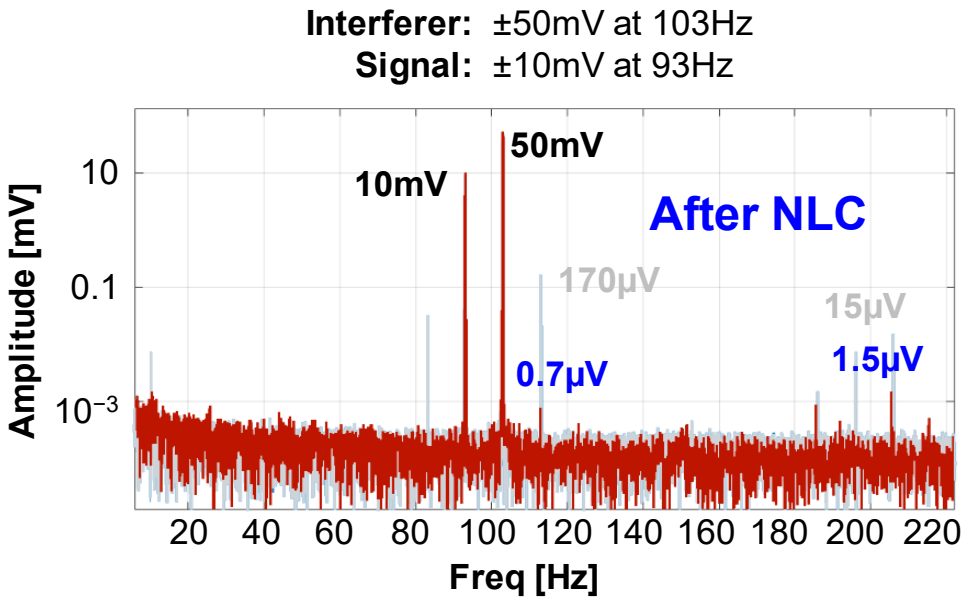

Figure 3.2 Two-tone test for linearity.

3.1.2 VCO Linear-range Stability Over Operational Temperature Range

As discussed in Section 2.2.3, the main contributor of the voltage-to-frequency mapping drift is the operating temperature variation. The use of our static foreground calibration technique was justified because of the regulated nature of the human body temperature. This test intended to demonstrate that our VCO ADC preserves it linear input range over the operational temperature drifts of ±2°C. The test setup is shown in Figure 3.3. VCO ADC testbed was housed inside "Test Equity Model 115" temperature chamber, NLC coefficients were calculated and programmed at 20°C, then the temperature was varied in 5°C increments. For each of the temperature points,

two test sine tones (3 Hz, 50mVp, 203 Hz, 50mVp) were fed to the sensing IC and the recordings were analyzed in Table 3.1. As we can see the ±5°C deviations from the calibration temperature result in no more than 3dB swings in the linear input range. Even at ±10°C, VCO ADC output still shows good linearity (73dB). These are more than acceptable results as the expected temperature swing is ±2°C.

Table 3.1 VCO ADC linearity drift over 0°C to +60°C temperature range.

Temp (°C)	0	5	10	15	20	25	30	35	40	45	50	55	60
3 Hz, 50mVp (dB SFDR)	64	67	73	82	79	80	74	69	69	65	63	62	60
203 Hz, 50mVp (dB SFDR)	66	69	76	85	85	85	74	69	66	64	63	62	60

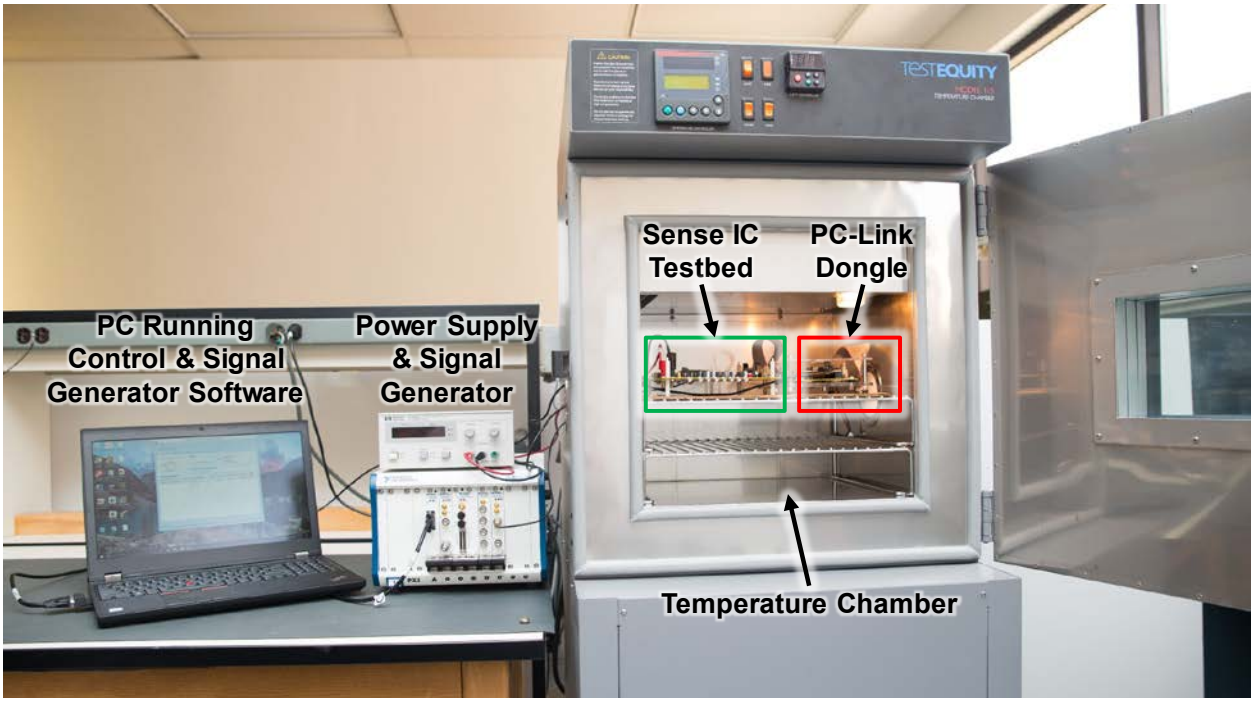

Figure 3.3 Temperature stability test setup.

3.1.3 Concurrent Stimulation and Sensing Test

This test intends to demonstrate the ability of our neuromodulation unit to record saturation-free neural data in the presence of therapeutic levels of stimulation. In-vitro test setup was used and instead of neural tissue, we used saline as it approximates the electrical properties of the cerebrospinal fluid (Figure 3.4).

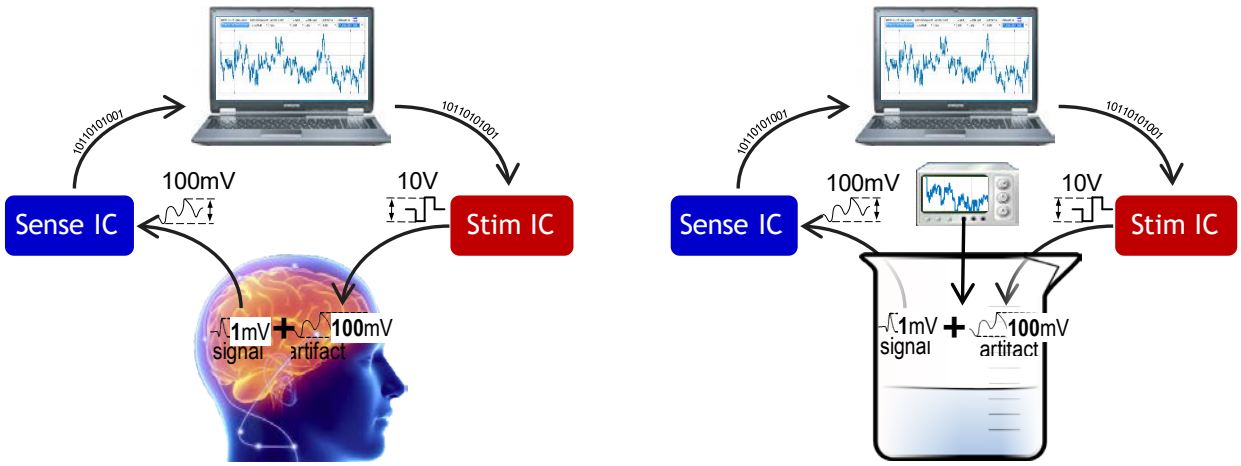

Figure 3.4 Illustration of in-vivo (A) and In-vitro (B) functional verification setups.

Since the saline does not produce neural signals as the live tissue does, a low-amplitude sine wave was injected to mimic neural signals. In the test setup illustrated below (Figure 3.5) a signal generator was used to simulate neural signals, power supplies delivered power to neuromodulation unit and its testbed, an oscilloscope was used to monitor stimulation output, and a PC running the control software allowed us to change stimulation parameters and log the sensed signals.

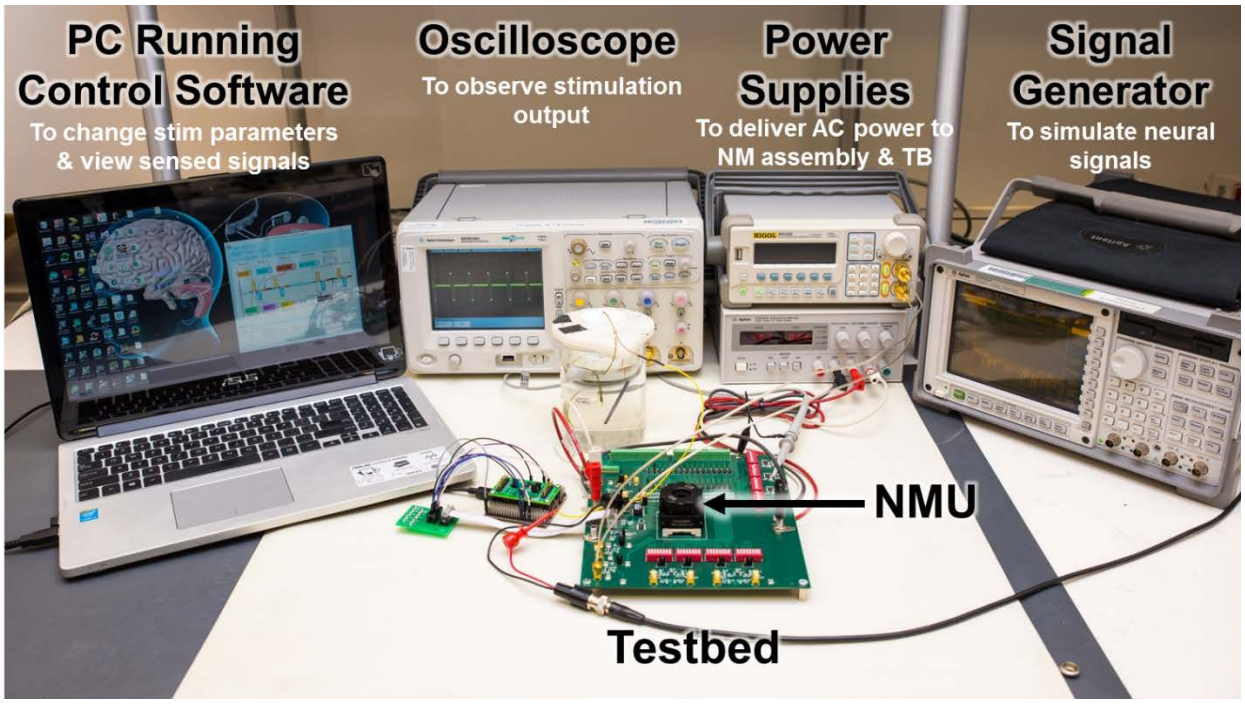

Figure 3.5 Concurrent sensing and stimulation test setup.

The neuromodulation unit was housed in a custom designed spring-loaded socket which ensured a good connection to NMU's feedthrough contacts (Figure 3.6).

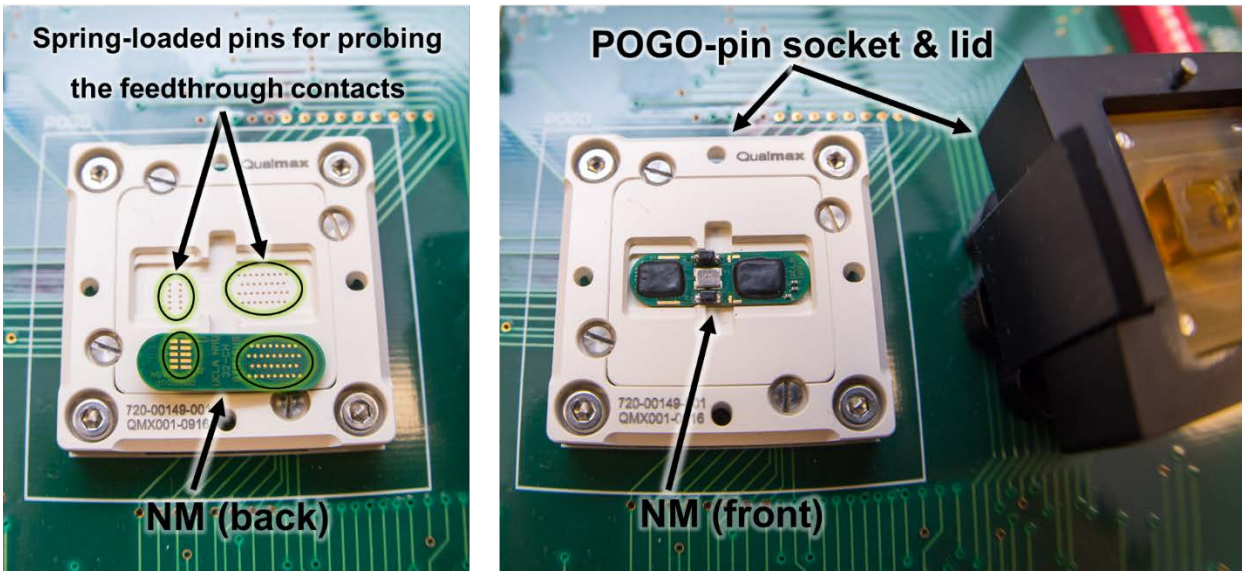

Figure 3.6 Spring-loaded test socket housing the NMU.

A 7Hz, +-2.8mV sine wave was injected into the saline beaker through the large probe on the right (Figure 3.7). The smaller probe on the left is the actual neural probe which has both the stimulation and sensing contacts. A 2mA constant current biphasic stimulation was on during the test. The summation of mimicked neural signal and the stimulation artifacts was picked up by an electrode connected to sensing front-end channel.

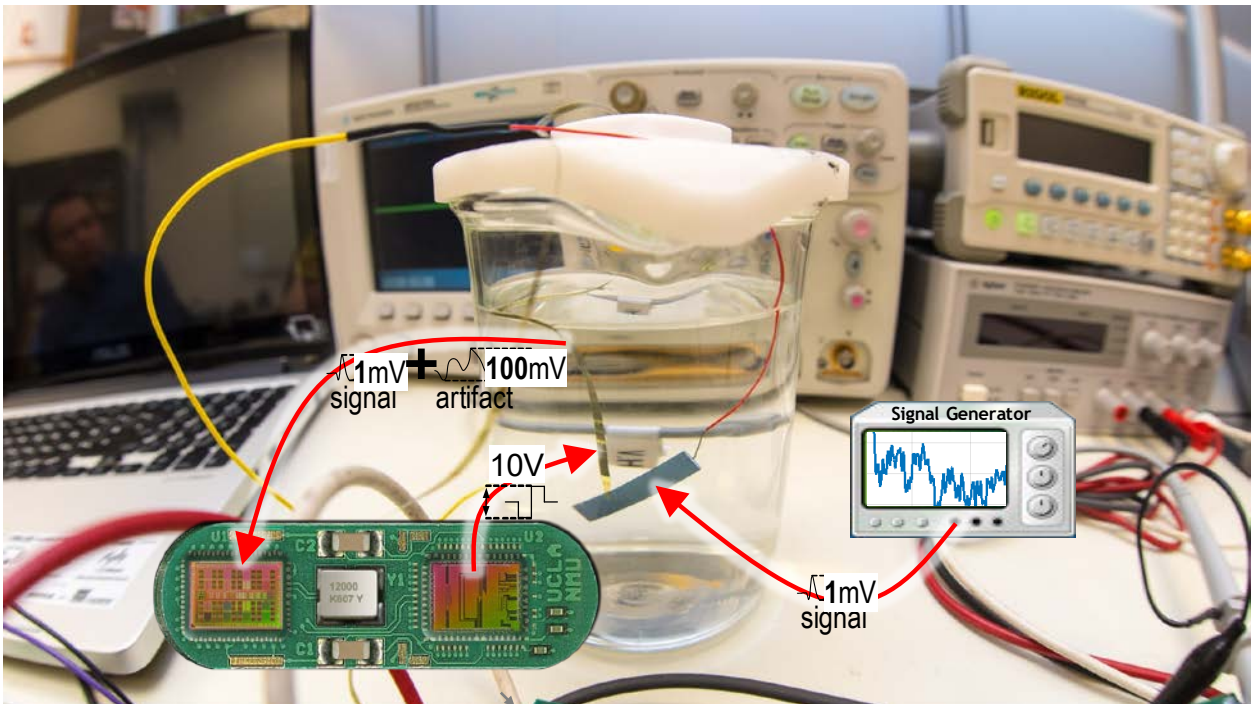

Figure 3.7 Saline beaker containing the neural lead as well as the electrode injecting mimicked neural signals.

Figure 3.8 shows the concurrent stimulation and sensing test results. The red line in the top figure mimics the neural signal, and the blue spikes are the stimulation artifacts. As we can see, our VCO-based sensing front-end does not saturate. The plot on the bottom shows the spectral composition of the signal. The mimicked neural signal tone is to the left, while the high-frequency stimulation artifact components are on the right. It is worth to note that the mimicked neural

signal choice is simplistic for illustration purposes. This choice provides easily identifiable envelope on the top signal plot and easily separable spectrum on the bottom plot.

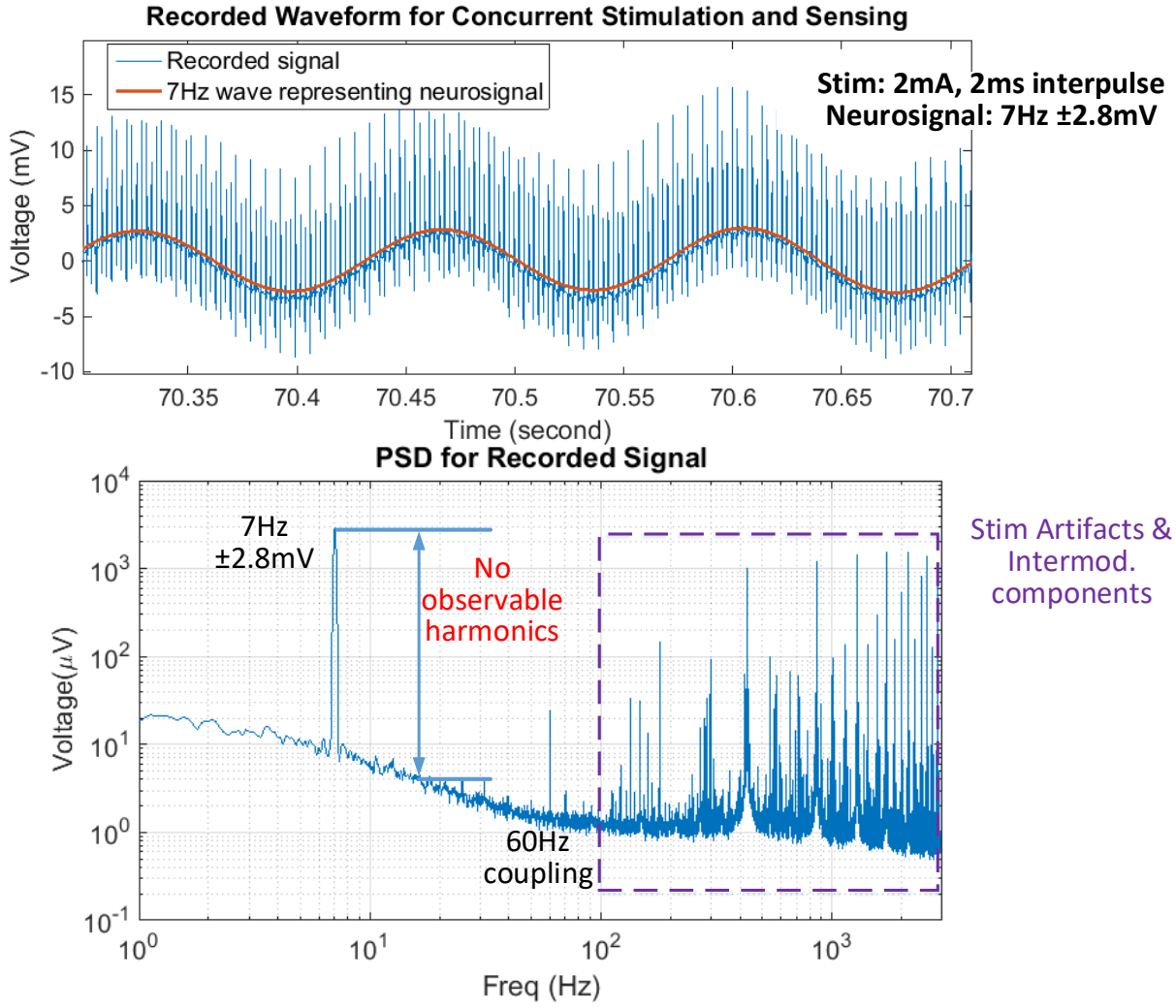

Figure 3.8 Sensing in the presence of large interferers such as stimulation artifacts.

3.2 Verification Platform

Due to the custom electronics and communication protocol, no standard test equipment or measurement instrumentation could have been used for testing the functionality and performance of the proposed NMU. For this reason, a custom test platform was built. This section describes the hardware, firmware, and software designed for that platform.

3.2.1 PC-link Dongle

NMU uses a three-wire SPI for communication. An SPI master is needed to configure and control the NMU and to read the sensed neural signals. Since SPI is not a communication interface available on personal computers, an adapter device is needed to facilitate data exchange between the PC and the NMU. Being a ubiquitous communication interface, Universal Serial Bus (USB) is used on the PC side of the dongle. Any controller supporting both – USB and SPI interface can serve as an adapter between the NMU and the PC (Figure 3.9). In our case, an off-the-shelf field programmable gate array (FPGA) board from Numato Lab called Saturn was used. The SPI master was implemented inside the FPGA, while for USB communication an external USB to FIFO controller was used (FTDI 2232H).

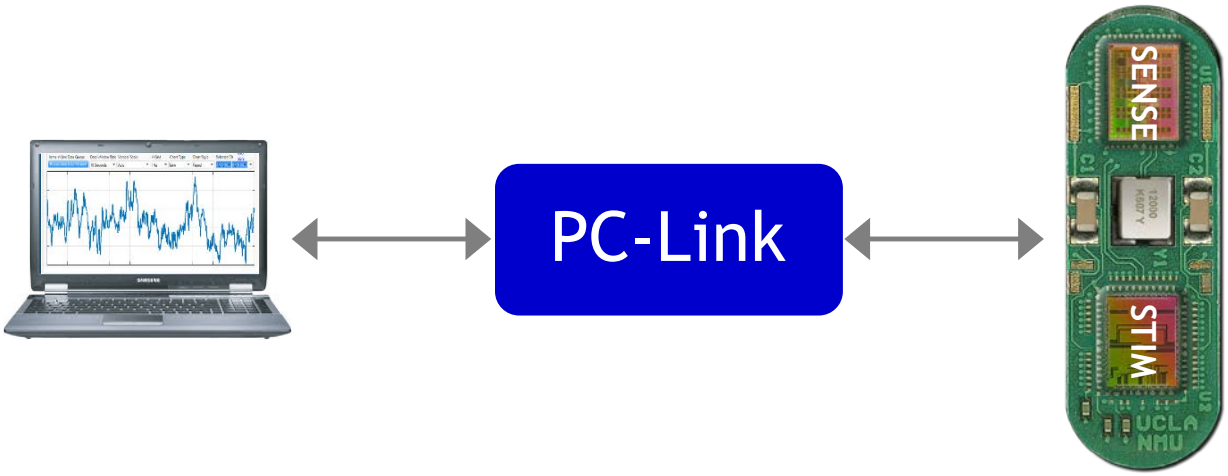

Figure 3.9 PC-link dongle for PC to NMU communication.

3.2.1.1 PC to PC-link Communication Packet Structure

On the FPGA side, the FTDI2232H has a one-byte FIFO interface (empty, full, data). On the PC side, data can be sent and received with one-byte chunks using either virtual communication port (VCP) or FTDI D2XX direct drivers. NMU communication protocol requires two (2) to 391

byte exchanges with SPI master (PC-link). To accommodate these exchanges between the NMU and PC-link a variable length packet structure was devised for the PC to PC-link communication (Figure 3.10). Each request (from PC to PC-link) and response (PC-link to PC) packet consists of COMMAND, ARGUMENT_LENGTH, PAYLOAD, ERROR, and CRC bytes. For request packets, the COMMAND byte tells the PC-link what command to execute, the ARGUMENT_LENGTH bytes denote the following payload length, the PAYLOAD bytes store the command arguments, and the CRC byte contains the payload checksum. The ERROR byte is not used for request packets. For response packets, the COMMAND byte tells the PC application what command was executed, the ARGUMENT_LENGTH bytes denote the length of the following payload length, the PAYLOAD bytes store the reply of the NMU and PC-link to the executed command, and the CRC byte contains the checksum of the PAYLOAD. If errors occurred during the command execution, the error code is returned in the ERROR byte of the response packet.

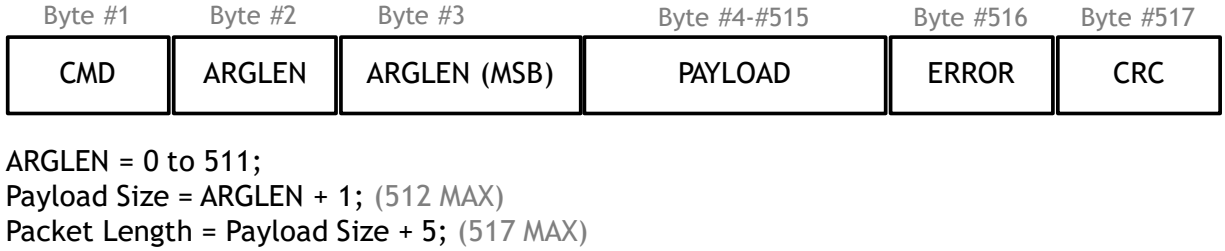

Figure 3.10 PC to PC-link variable-length communication packet structure.

3.2.1.2 Firmware

The PC-link FPGA firmware is implemented as a Mealy finite state machine (FSM) – a simplified state diagram of which is depicted in Figure 3.11. After the initialization sequence, the FSM polls

the FTDI2232H FIFO interface for incoming bytes. Once a full command packet (Figure 3.10) is received, the execution passes to the packet processing state. In this state, stimulation and sensing IC specific commands are sent to NMU through the SPI master. The reply of the stimulation and sensing ICs is packaged into the payload of the response packet and is transmitted back to PC application. After the response packet transmission, the FSM returns to the IDLE state where it waits for new commands unless the command being executed is the ADC_READ command. For ADC_READ command, after a sample is transmitted to PC, the FSM will enter the packet processing state indefinitely unless interrupted with a special command. This allows for continuous streaming of VCO ADC samples to PC.

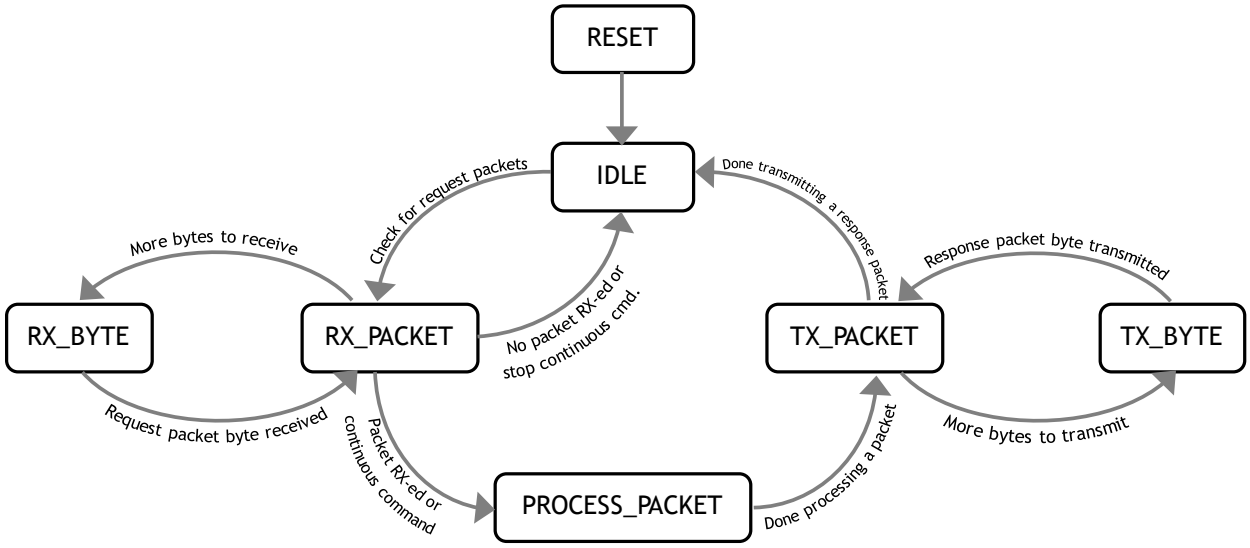

Figure 3.11 Simplified state machine diagram of the PC-link firmware.

3.2.2 Custom PC GUI

The custom graphical user interface (GUI) designed for the NMU verification allows users to exchange commands, status information, and ADC samples with the stimulation and sensing ICs. On the PC application side, the data to/from the PC-Link dongle is being produced and consumed

through FTDI's direct D2XX drivers. In comparison with the standard virtual COM port (VCP) drivers, D2XX accesses the USB device directly without additional OS layers (Ports device setup class and the Serial function driver). D2XX also allows for direct enumeration and detection of FTDI2232H devices using serial numbers. This means that the PC-Link dongle can be reliably detected every time and on every system. When using VCP drivers the device detection is not straightforward since the PC-link dongle COM port number will depend on the order and number of other connected devices. When using VCP mode, COM port detection would require opening all available COM ports, sending a ping, and waiting for a correct reply. This would not be an elegant approach since the ping requests can cause unknown behavior or malfunctions in other connected devices.

The GUI consists of three main sections – 1) PC-link verification screen, 2) stimulating IC control screen and 3) sensing IC control and readout screen. Functionality covered in these GUI screens is presented in the sections below.

3.2.2.1 Synchronization With PC-link

The data between the PC GUI and the PC-link dongle is being exchanged in one-byte chunks. Since the PC-link communication packet can contain a sequence of up to 517 bytes (Figure 3.10) and physical disruptions in the USB connection can occur any time (loose cable or device removal/insertion) a handshaking mechanism ensuring that the GUI is synchronized with the PC-link is necessary. When powered up, the PC-link dongle will enter into the SYNC mode and will increment and send back any byte sent by the GUI. Incremented bytes will indicate to the GUI that the PC-link is in the SYNC mode. The GUI can end the SYNC mode by sending the

END_OF_SYNC command byte. After the END_OF_SYNC byte, the GUI will be synchronized with the PC-link and the next byte will be interpreted as the CMD byte of a new packet. The above-outlined portion of the handshaking mechanism ensures synchronization when the PC-link is physically attached or re-attached (intentionally or due to a loose cable) to a USB port. In case the PC-link is already attached and synchronized, the PC-link and the GUI can still be thrown out of sync due to a GUI crash in the middle of a packet transmission. To initiate re-synchronization, the GUI starts sending a stream of invalid commands at startup. When the PC-link receives an invalid command, the SYNC mode is reactivated as outlined previously. Since the GUI can be thrown out of synch at any point in the communication packet transmission, the number of the invalid commands sent during the GUI startup is more than the maximum length (517) of the PC-link communication packet to ensure the SYNC mode activation.

3.2.2.2 PC-link Verification Screen

When configuring the NMU through the GUI or acquiring samples, communication errors could occur on the PC to PC-link dongle path (USB cable) or between the PC-link and NMU (three-wire SPI). To isolate the error source, the GUI has a PC-link test screen (Figure 3.12). This screen allows the user to run bidirectional and unidirectional data streaming tests. When running the bidirectional streaming test, packets with a pseudo-random byte sequence payload are sent from the PC to the FPGA. On the FPGA side, the payload bytes are incremented and the packet is transferred back to the PC. On the PC side received payload bytes are checked for their expected value, and then error statistics and data transfer speeds are updated. In bidirectional streaming test about 2.88 Mbytes/S uplink and downlink speeds can be achieved. Unidirectional data

streaming tests transfer pseudorandom data only in one direction (PC to FPGA, or FPGA to PC) and can reach transfer speeds of up to 5.76 Mbytes/S. Although the PC-link is designed for one (1) NMU verification, the attained data bandwidth will allow simultaneous streaming from up to 10 NMU modules at their maximum data rates which are reached when sensing from all 32 channels simultaneously with three-byte resolution and 6kHz sampling rate.

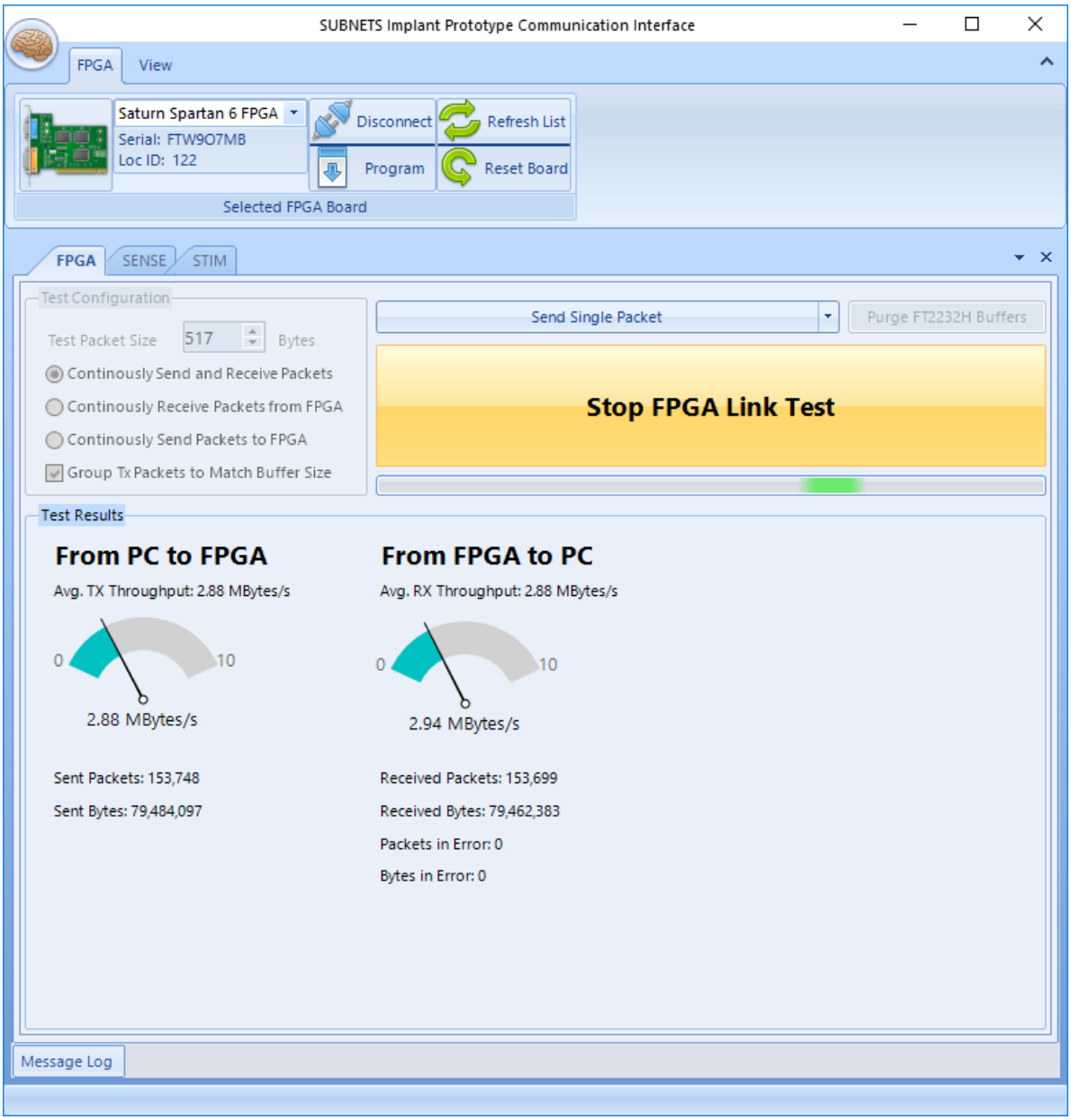

Figure 3.12 The PC-link verification screen of the GUI.

3.2.2.3 The Stimulation IC Control Screen

The stimulation IC control screen allows users to configure stimulation timing and amplitude, sensing and stimulation switch-matrices, power subsystem, start and stop stimulation, and read stimulation and battery status registers. The stimulation timing and amplitude screen is common for all eight (8) stimulation engines (named A to H). Once the stimulation engine is selected in the drop-down list on the top left corner, timing and amplitude parameters can then be set for that engine (Figure 3.13).

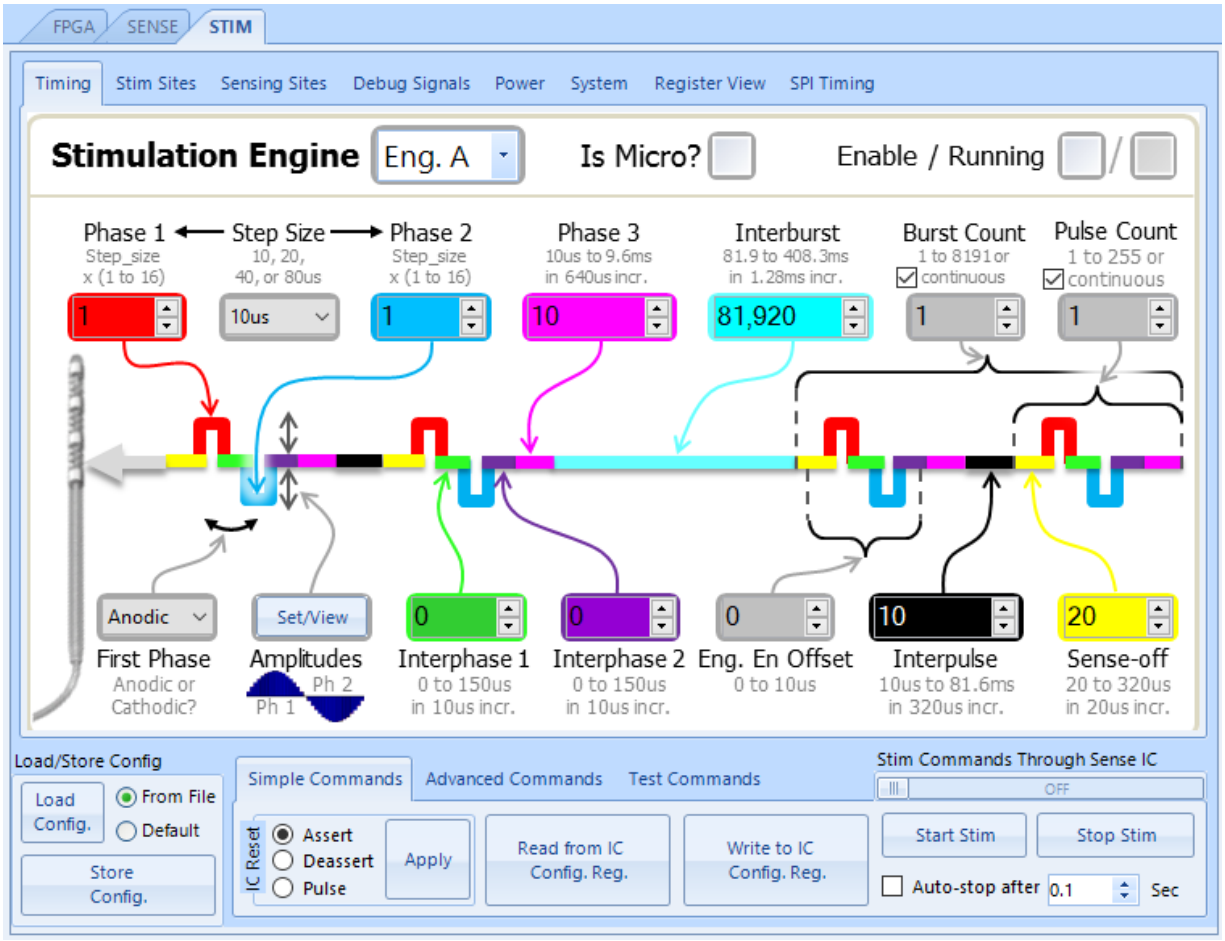

Figure 3.13 Stimulation timing configuration screen.

The three main timing parameters are the “phase 1” and “phase 2” pulse durations, “inter-phase 1 delay”, and “inter-pulse delay”. These parameters allow generation of conventional continuous pulse trains. The burst stimulation paradigm can be programmed by setting the pulse count (1 to 255) and unchecking the “continuous” checkbox underneath. The interburst parameter will then control the duration between the groups of pulses (bursts).

Anodic and cathodic phase amplitudes can be set through the stimulation amplitude configuration screen (Figure 3.14). This screen allows users to define stimulation phase amplitudes either by selecting from one of the pre-defined shapes (rectangular, exponential, sine, sawtooth, triangular) or by drawing a custom shape using the mouse.

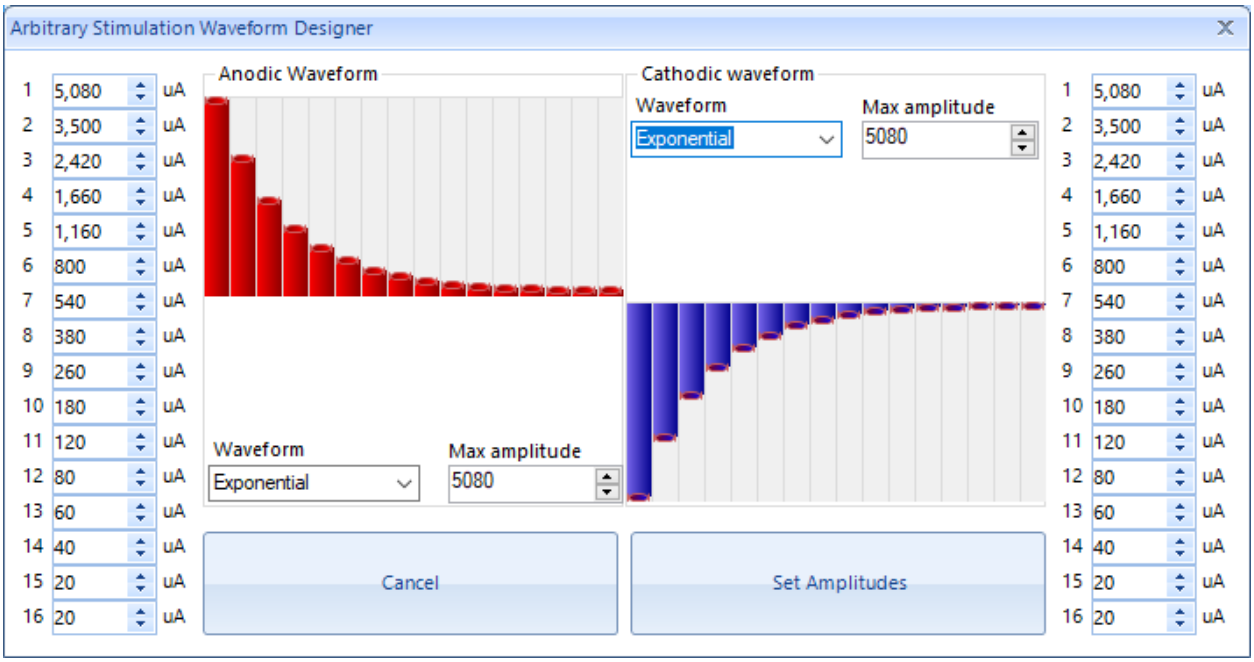

Figure 3.14 Stimulation amplitude configuration screen.

The eight (8) stimulation engines are mapped to 64 contacts through stimulation switch-matrix. Each engine can be mapped to 10 contact pairs. The stimulation switch-matrix configurator

(Figure 3.15) screen allows the user to select and visualize the sites on neural probe where the stimulation is going to be applied.

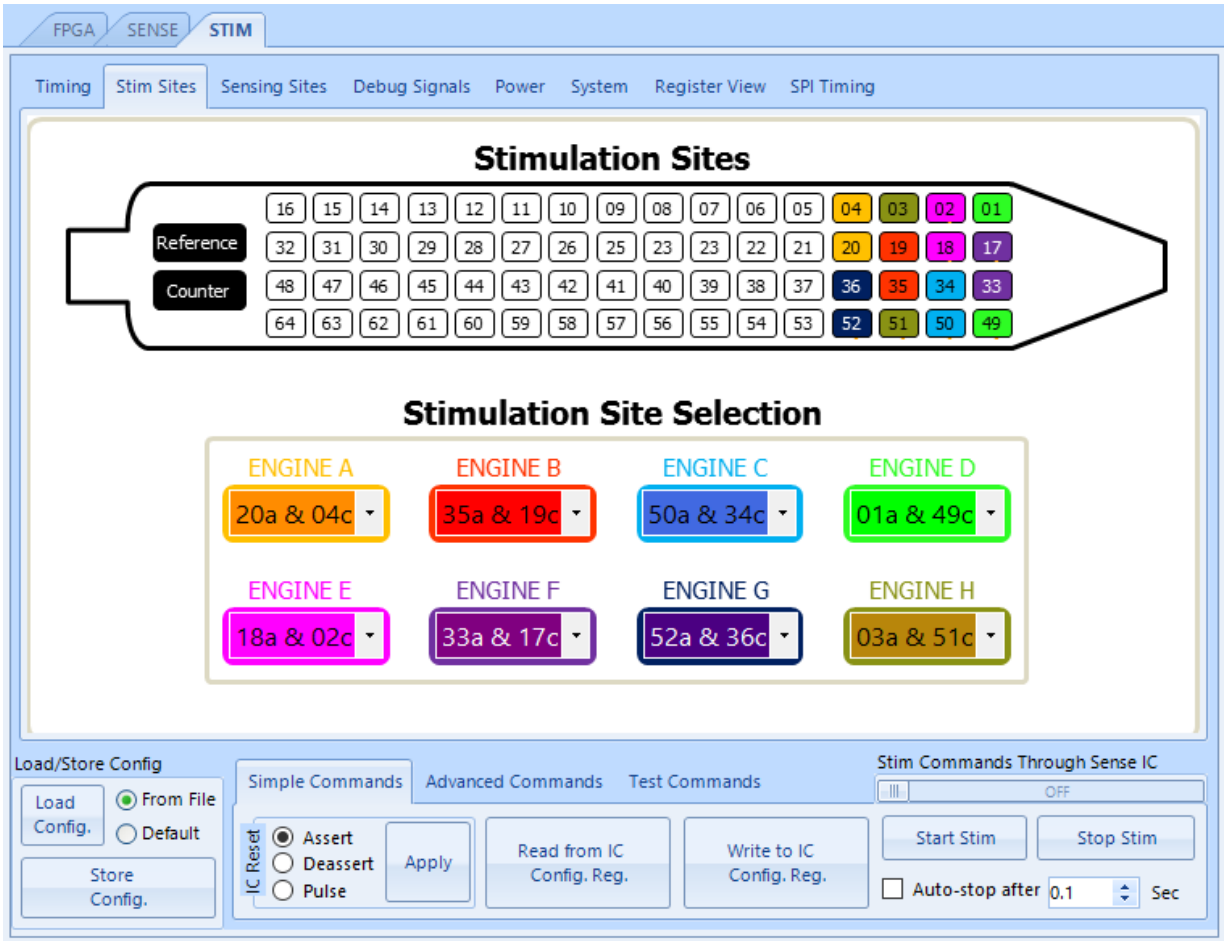

Figure 3.15 Stimulation switch-matrix configurator.

When high stimulation currents are selected, the induced voltages at stimulation sites will exceed the sensing IC IO pad maximum voltage which can damage IC's internal electronics. To protect the sensing IC and also to provide "blanking" functionality at sensing contacts connected to other than stimulation sites a sensing switch-matrix is employed. This matrix is highly configurable and can be controlled at individual contact level. When a sensing site is enabled it will be connected to the neural probe when no stimulation is active. If a sensing site is also configured to be enabled

during stimulation it will be connected to neural probe even when stimulation is active on other sites. A sensing site can also be forced to be connected to neural probe when there is an active stimulation on that site. All of these options can be set from the sensing site selection screen (Figure 3.16).

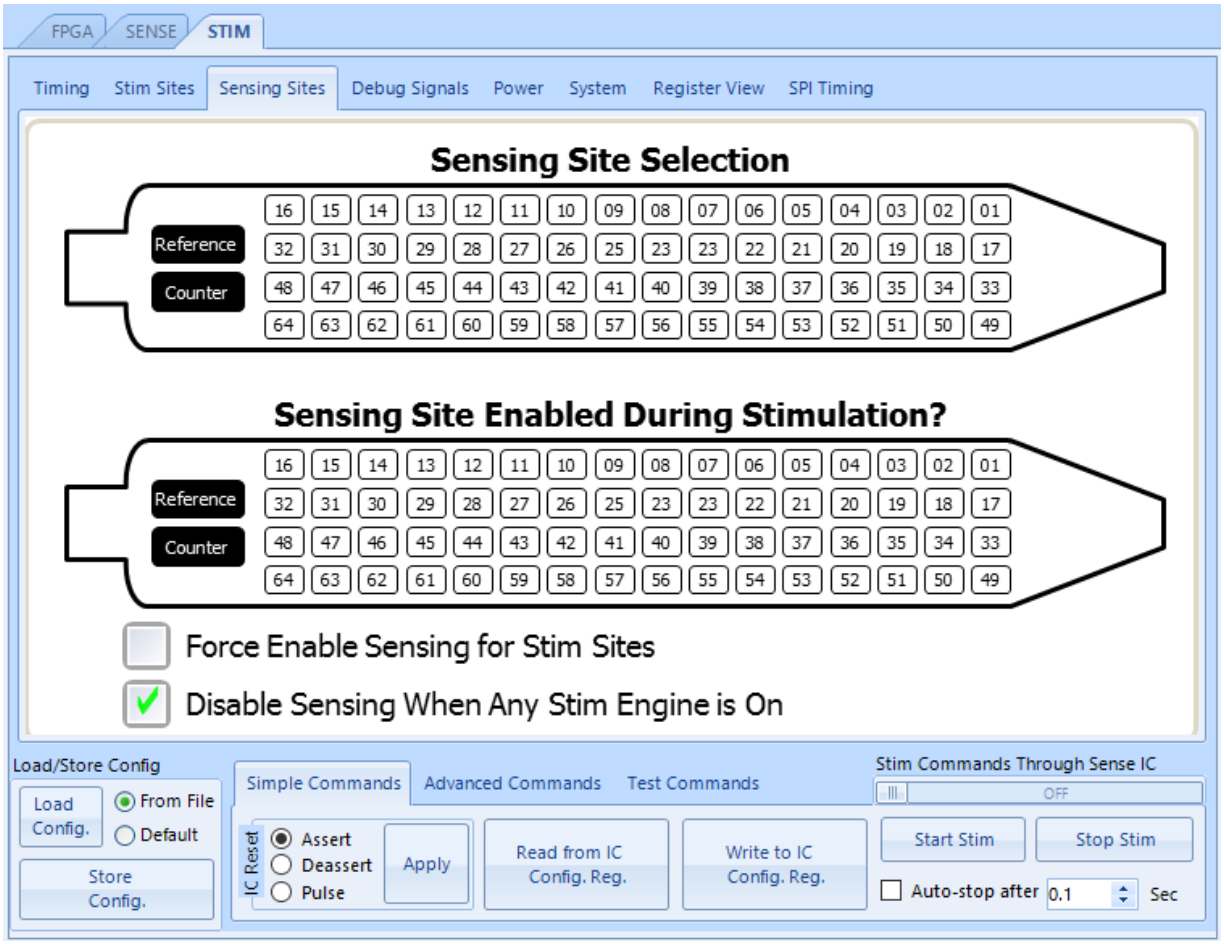

Figure 3.16 Sensing site selection screen.

The power management configuration screen (Figure 3.17) controls the behavior of the rectifier, battery charger, and charge pump circuits of the stimulation IC. The operation mode controls configure the stimulation IC for either wireless (inductive link) or wired power delivery. If the system is equipped with a single-cell Li-Ion battery, battery charging and status readout circuits

can be enabled as well. Stimulation IC is designed to deliver constant current stimulation pulses. This means that the output voltage of the DAC is automatically adjusted (up to the compliance voltage) based on the electrode impedance to deliver the constant amount of current. For increasing the efficiency of the stimulation engine, the compliance voltage can be adjusted given a specific electrode and desired current combination. This can be achieved through the compliance voltage drop-down control.

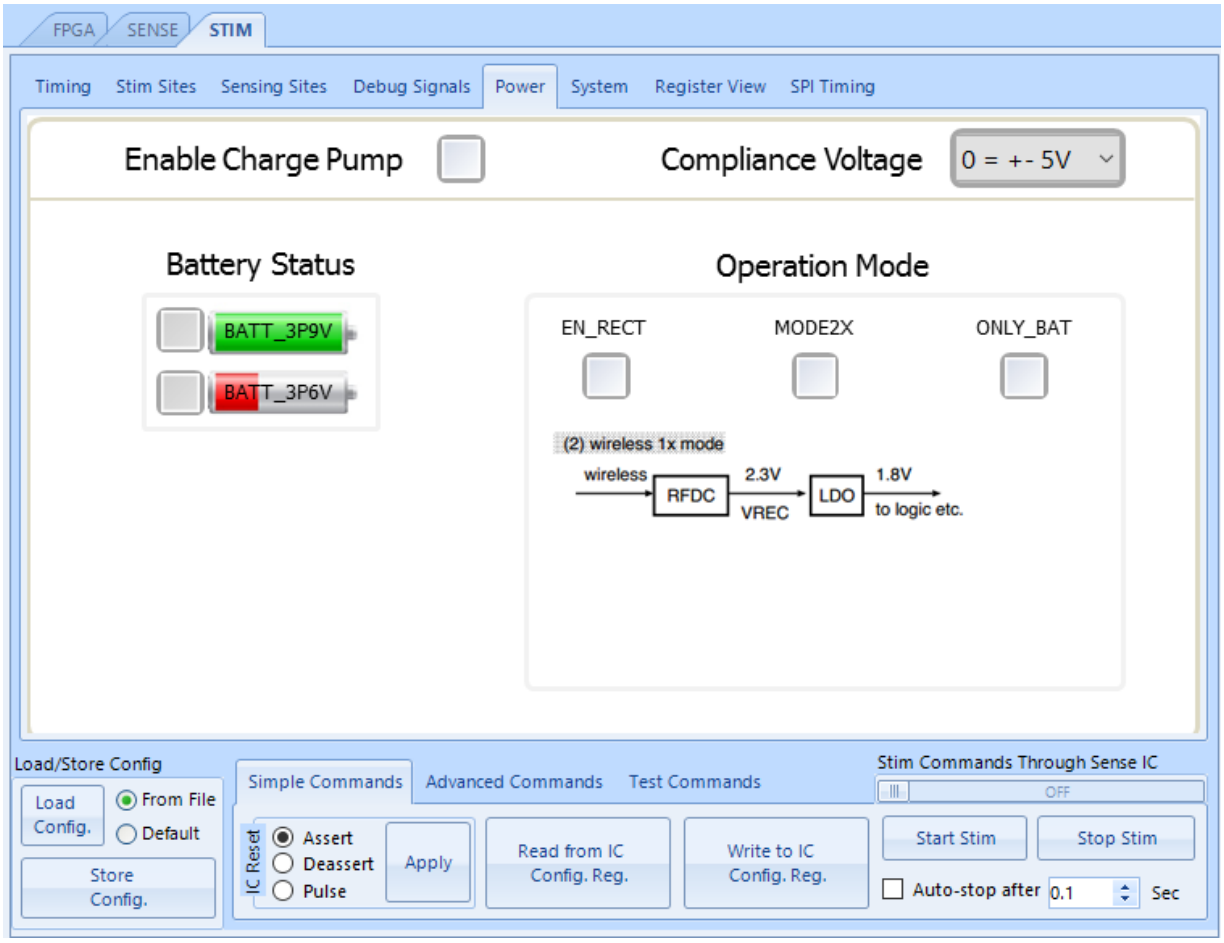

Figure 3.17 Stimulation IC power management configuration screen.

When the stimulation IC is paired with the sensing IC, the sense_off signal from the stimulation IC notifies the sensing IC when the sensing switch-matrix is disabled. This signal is used by the

sensing IC to mark the samples recorded during this time as invalid because they were recorded when no neural signal was present (high impedance). The stimulation system configuration screen allows users to select the stimulation engine which is connected to the sense_off signal. Sense_off can optionally be sourced from the logical OR of all stimulation engines sense_off signals. The stimulation system configuration screen also allows users to select whether or not to reset the stimulation IC when general or CRC errors occur.

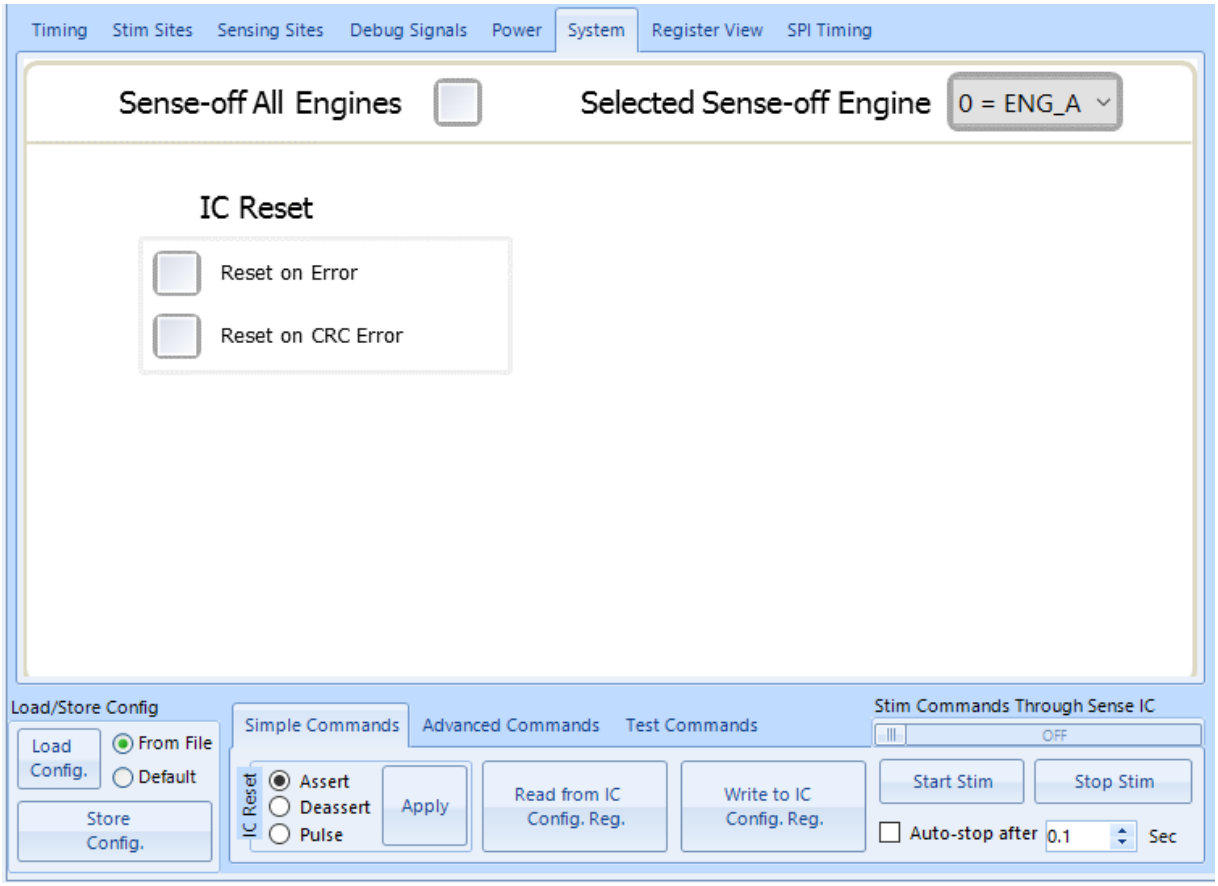

Figure 3.18 Stimulation IC system configuration screen.

The stimulation IC SPI timing configuration screen allows testing of the stimulation IC at various SCK speeds and interbyte delays. When "Auto-adjust SCK frequency" and/or "Auto-adjust interbyte delay" checkboxes are activated, the SCK speed and/or interbyte delay will be selected

based on the datasheet recommendations of the stimulation IC. For ensuring the viability of the selected SCK speed/delay combination, an automated loopback test can be run from the "Test Commands" section of the stimulation UI.

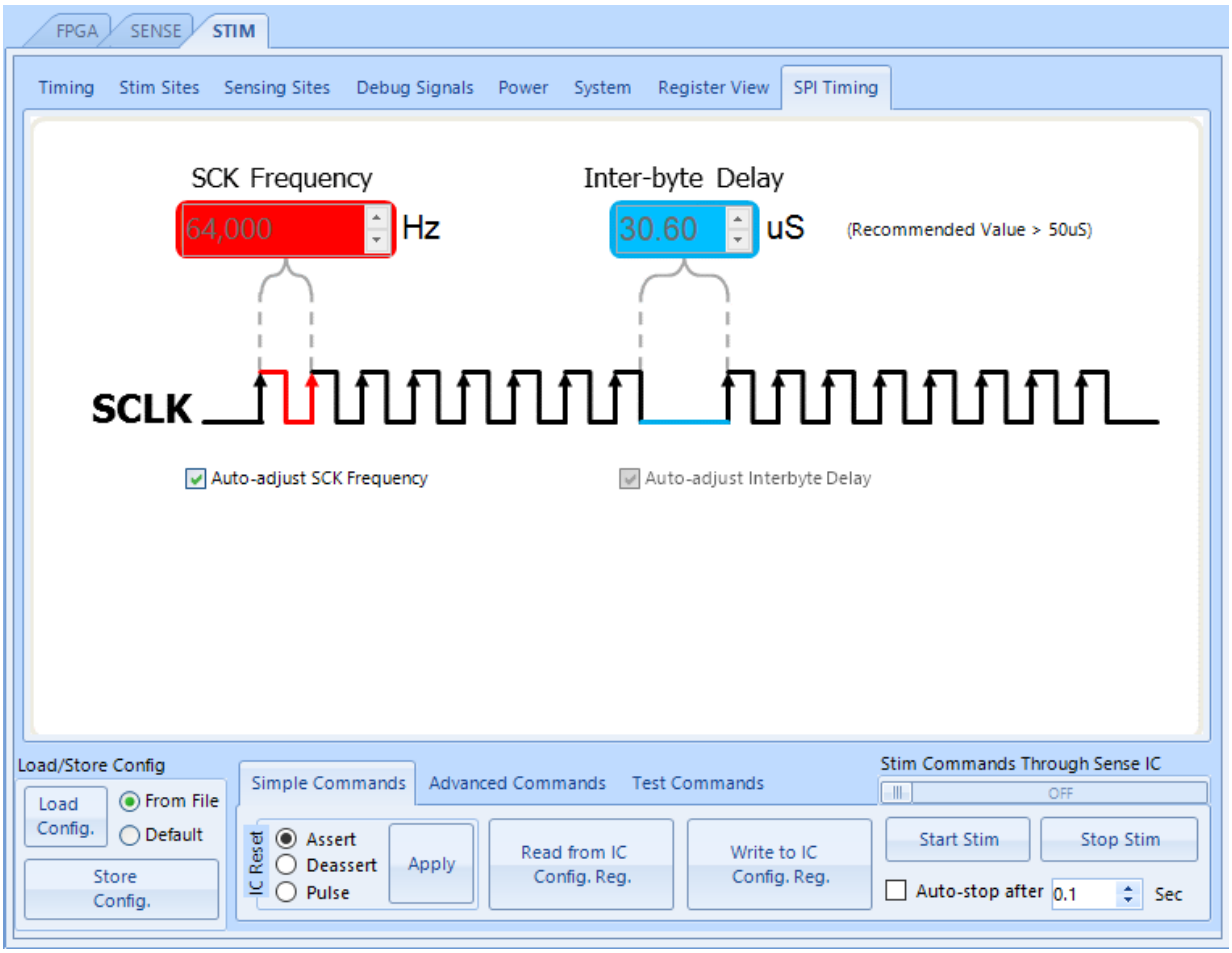

Figure 3.19 Stimulation IC SPI timing configuration screen.

The debug signal selection screen allows users to output some of the internal signals to the stimulation IC pads for probing/debugging purposes. The source signals include the "phase 1", "phase 2", and "phase 3" switch control signals as well as the DAC enable signals for all eight (8) stimulation engines.

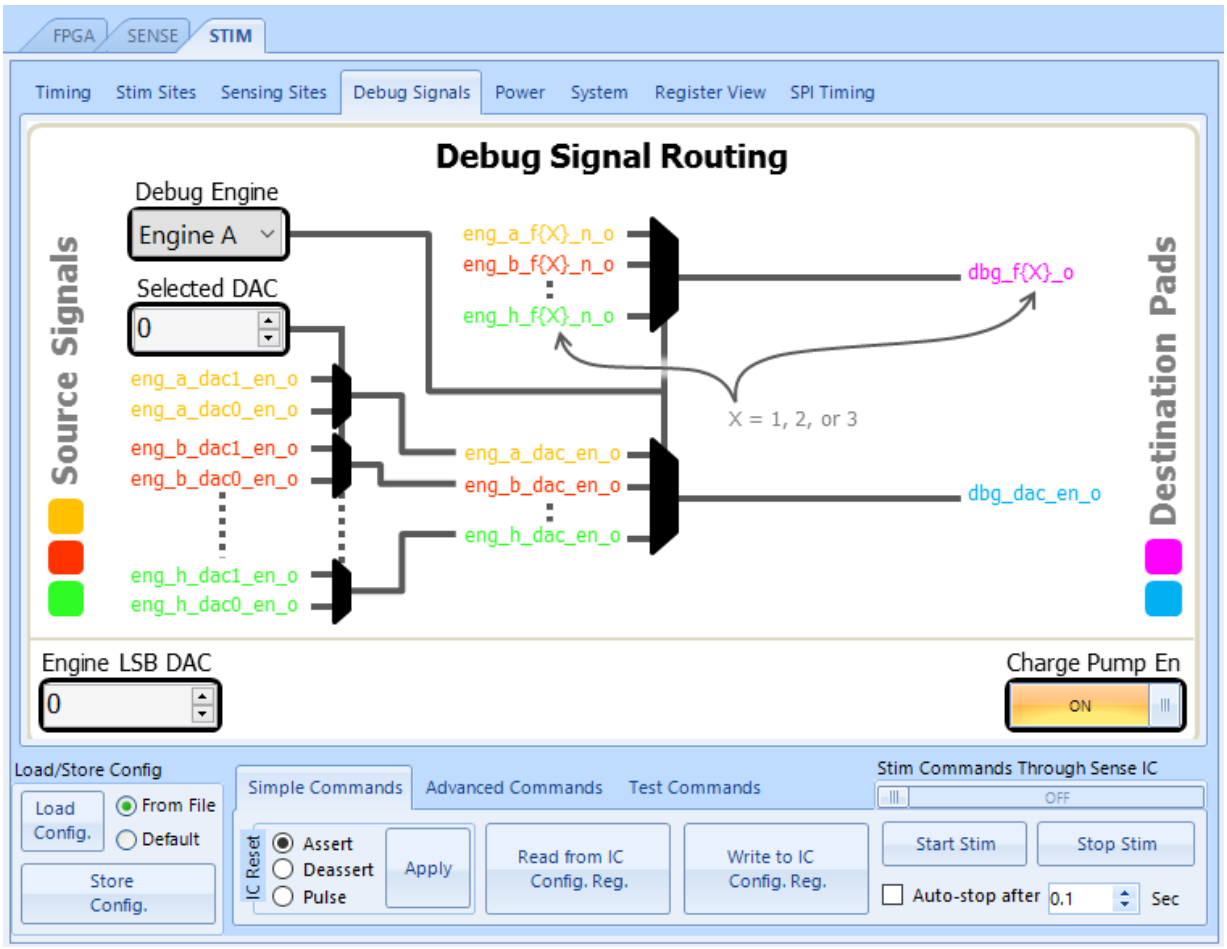

Figure 3.20 Stimulation IC debug signal selection screen.

The stimulation IC register view screen allows users to view and modify all previously discussed stimulation IC configuration parameters in a simple matrix form mimicking the register map of the IC. Each row in this view represents a stimulation IC register. Bits can be set/unset by clicking on their checkbox representations. The rest of the stimulation IC screens will update automatically when the configuration is changed from the register view. Similarly, when the stimulation IC configuration is changed from other UI screens, the register view will update as well. This gives the user two different ways of configuring the stimulation IC from the GUI.

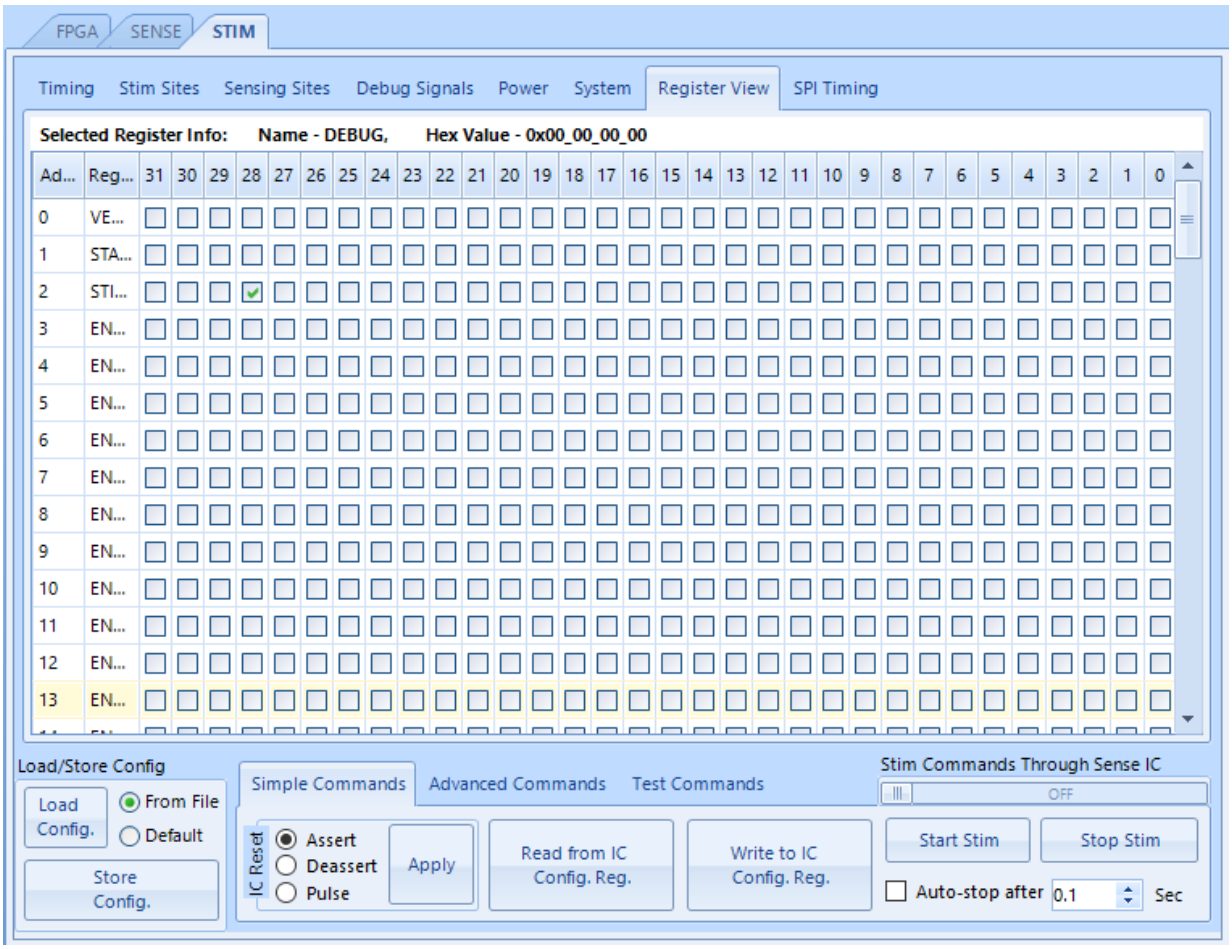

Figure 3.21 Stimulation IC register view.

3.2.2.4 The Sensing IC Control and Readout Screen

The sensing IC control UI allows users to configure VCO ADC sampling parameters, program and test the NLC, ASAR, and decimation DSP engines, and display/log the streamed ADC samples. Sensing IC configuration memory consists of two sections – 1) the NLC coefficient memory bank, and 2) the IC register bank. These configuration memories are read and written using different SPI commands. The NLC coefficient memory bank can be programmed through the sensing IC NLC coefficient configuration screen (Figure 3.22). Coefficients can be imported and exported from the screen using the comma separated value (CSV) file format. The "Compare

Coefficients to ..." button allows comparison of the coefficients already loaded into the application to external CSV files. This can be useful when determining whether the correct coefficient set is loaded into the IC before acquiring data or for running a single loopback test. The latter, however, is not the preferred way of verifying the communication with the sensing IC. An automated loopback test, available in the "Test Commands" section of the sensing UI, should be used for that purpose instead.

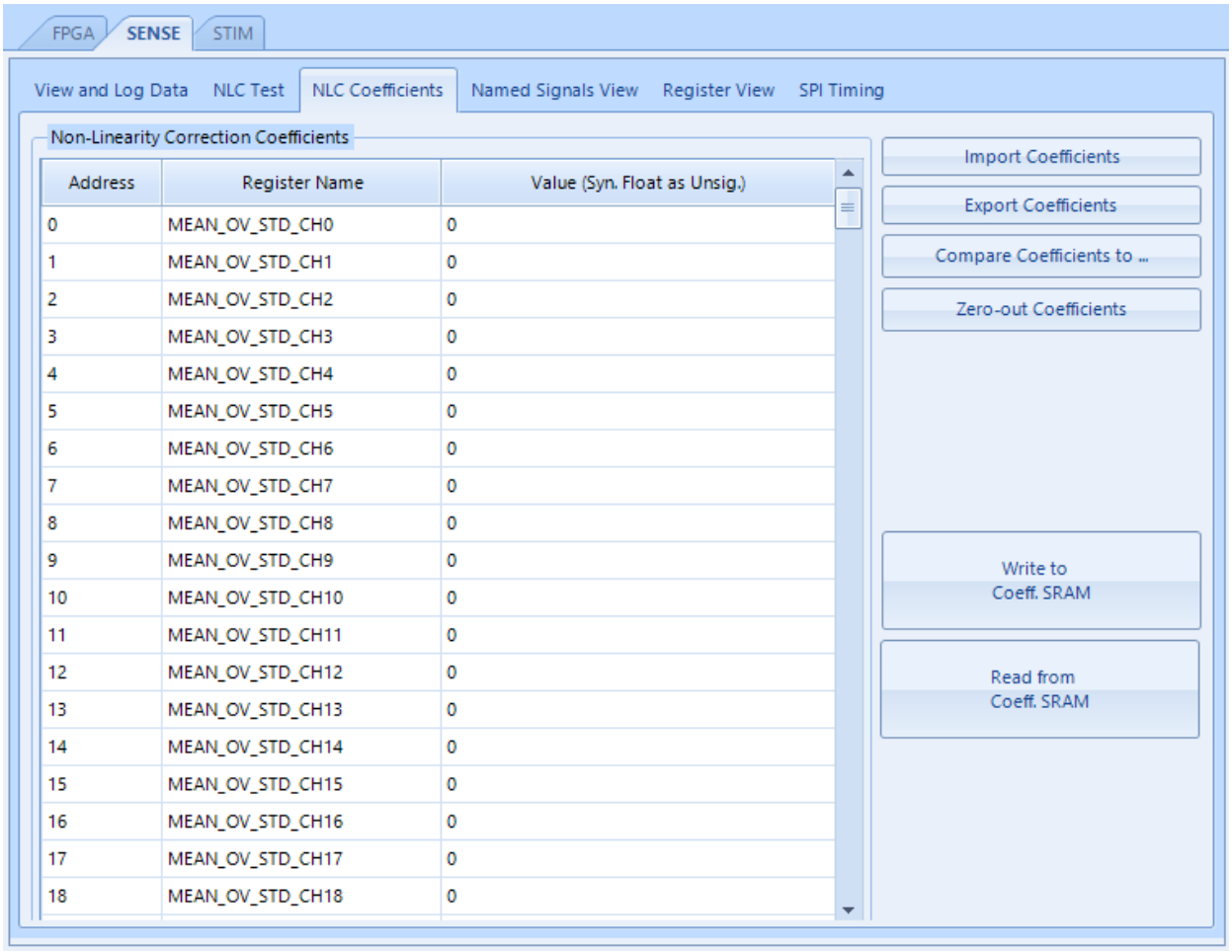

Figure 3.22 Sensing IC NLC coefficient configuration screen.

Once the NLC coefficients are loaded into the sensing IC, the NLC and ASAR tests can be run from the "NLC Test" screen of the sensing UI (Figure 3.23).

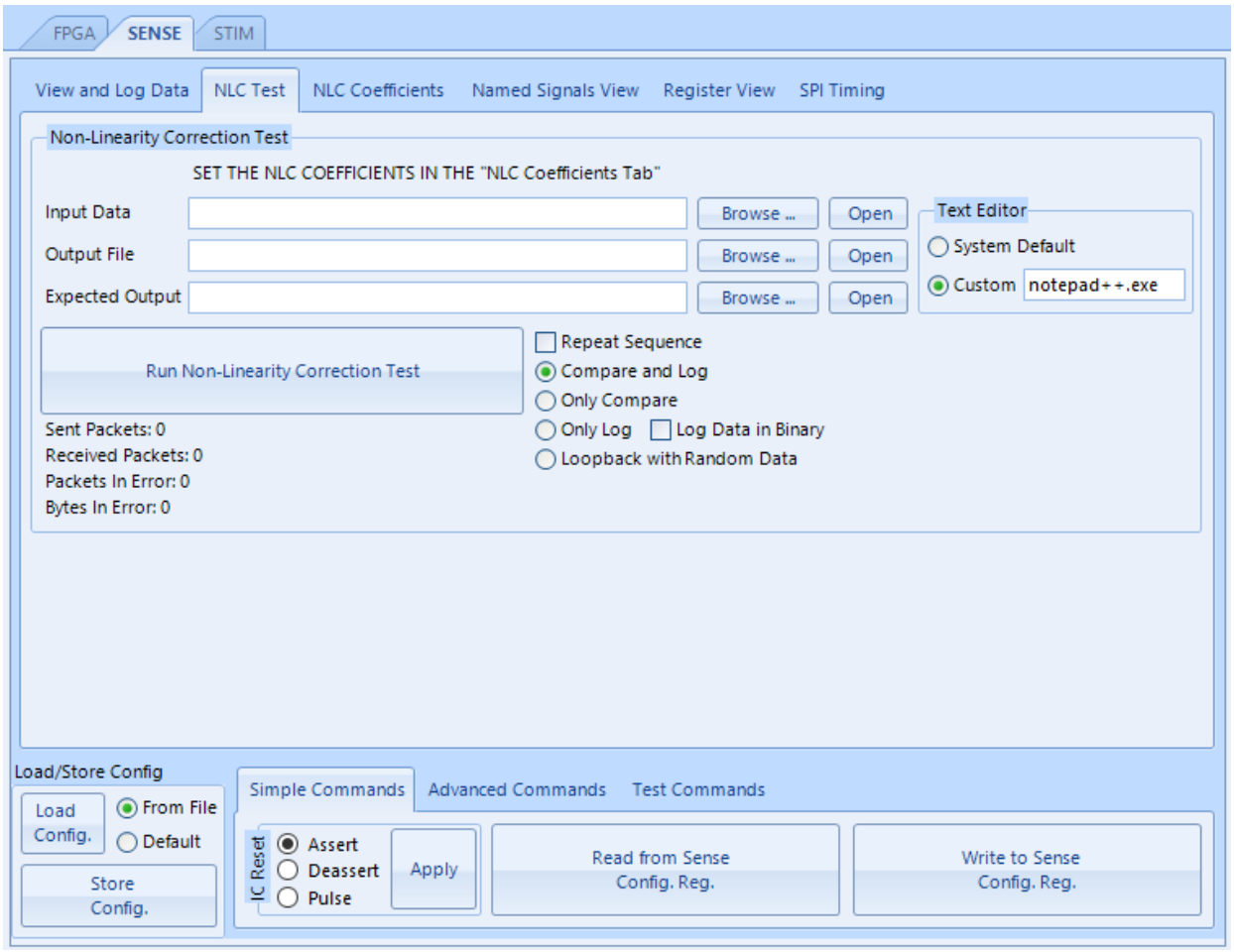

Figure 3.23 Sensing IC screen for testing the NLC, ASAR, and decimation DSP blocks.

Input vectors are fed to the DSP blocks externally, bypassing the VCO ADC (Figure 3.24).

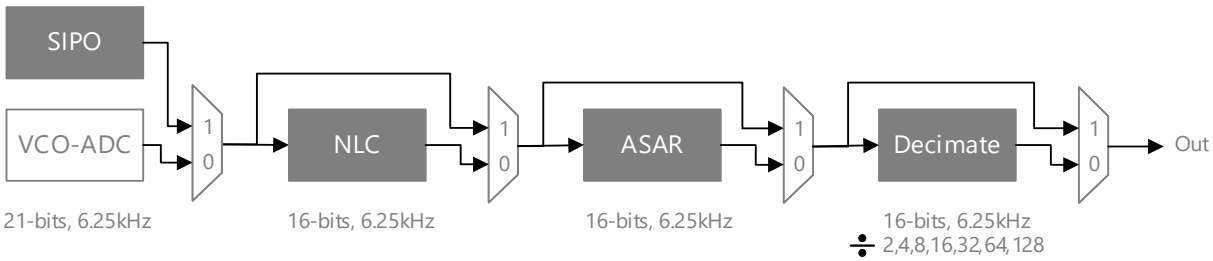

Figure 3.24 Bypassing VCO-ADC for testing NLC, ASAR, and decimation DSP blocks.

Depending on the system configuration bits, NLC, ASAR, and decimation block can be enabled or bypassed during this test. The output of this signal chain is then compared to the expected

output vector and results are logged. For average power measurement purposes, test vectors can be repeatedly fed into sensing IC in an infinite loop.

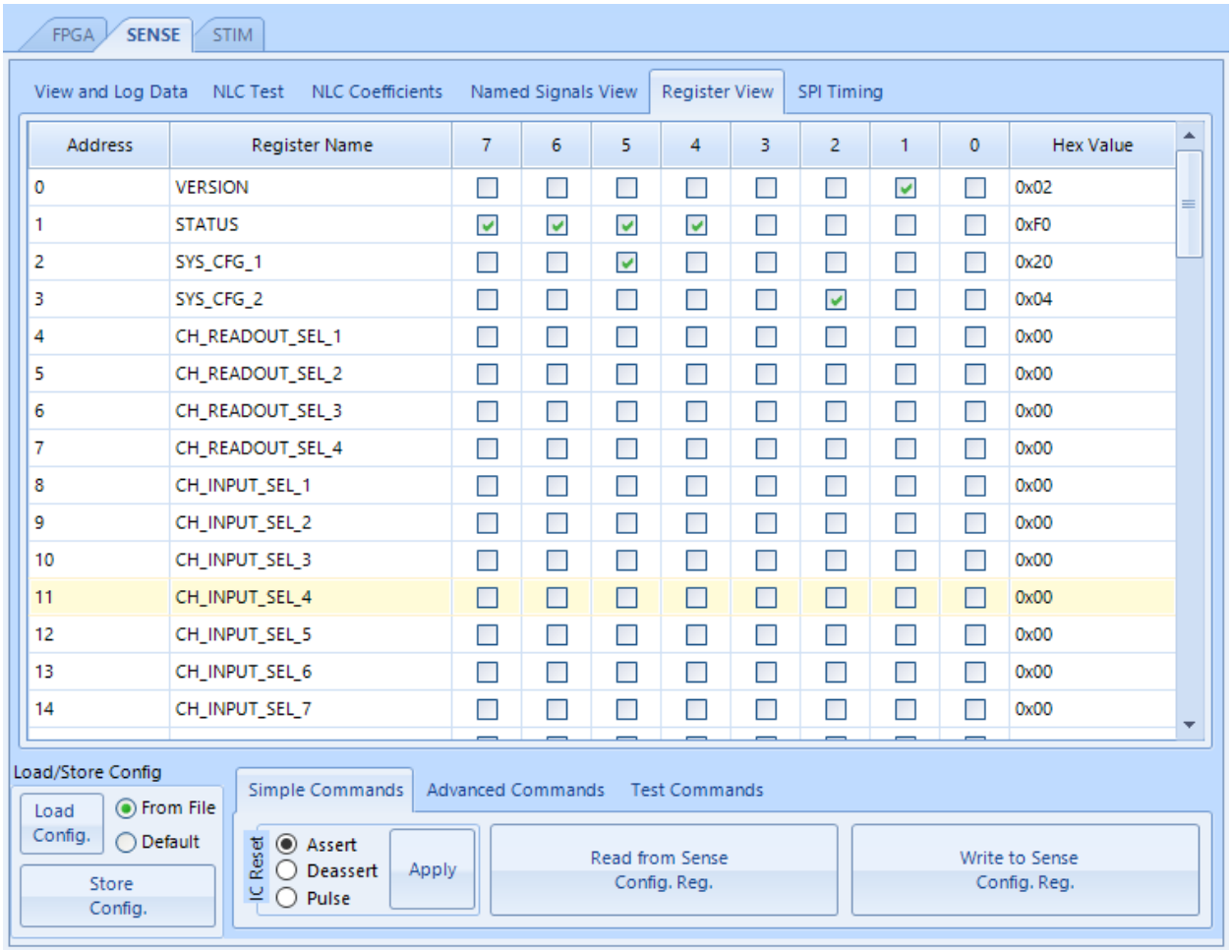

Figure 3.25 Sensing IC register view screen.

The sensing IC UI has two screens for reading/writing the IC configuration register bank. The sensing IC register view screen (Figure 3.25) presents the configuration bits in a matrix form where each row represents a sensing IC register. The user needs to be familiar with a register layout in order to make meaningful changes or decode read configuration. This view is useful when debugging IC communication issues, more specifically when serial SPI data-stream is being

observed for concordance with the UI content. In other cases, it is more user-friendly to modify or decode read IC configuration from the "Named Signals View" screen.

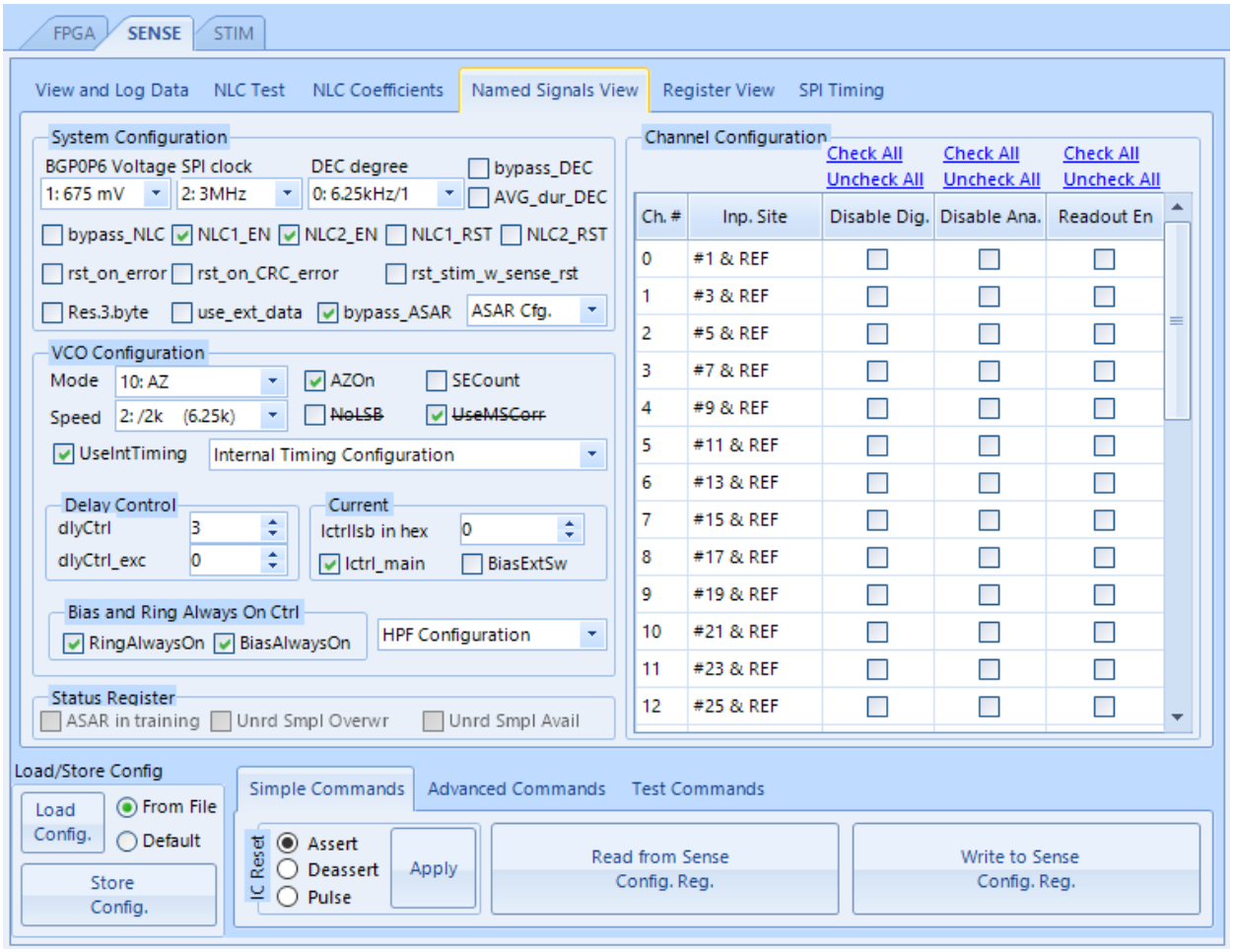

Figure 3.26 Sensing IC named-signals view screen.

The sensing IC named signals view screen (Figure 3.26) can be used for modifying or decoding read IC configuration even when the user is not familiar with the register map of the sensing IC. In this view, all of the configuration signals are laid-out by their respective names and are grouped by their function. Controls in the "VCO Configuration" section of the screen allow users to set parameters related to sampling rates, internal timing, adjust the VCO bias current, etc. "Channel Configuration" section has controls that let users enable/disable the readout from certain channels,

configure the 2x1 sensing channel mux, and enable analog and digital circuits of a specific channel. "Status Register" section allows the decoding of the status register and can tell when ASAR is in training, or whether unread samples are available or have been overwritten because they were not read fast enough. All other system level configuration bits can be controlled from the "System Configuration" section. To fit all UI controls in the "Named Signals View" screen, some controls are grouped into pop-up panels ("ASAR Cfg.", "Internal Timing Configuration", and "HPF Configuration").

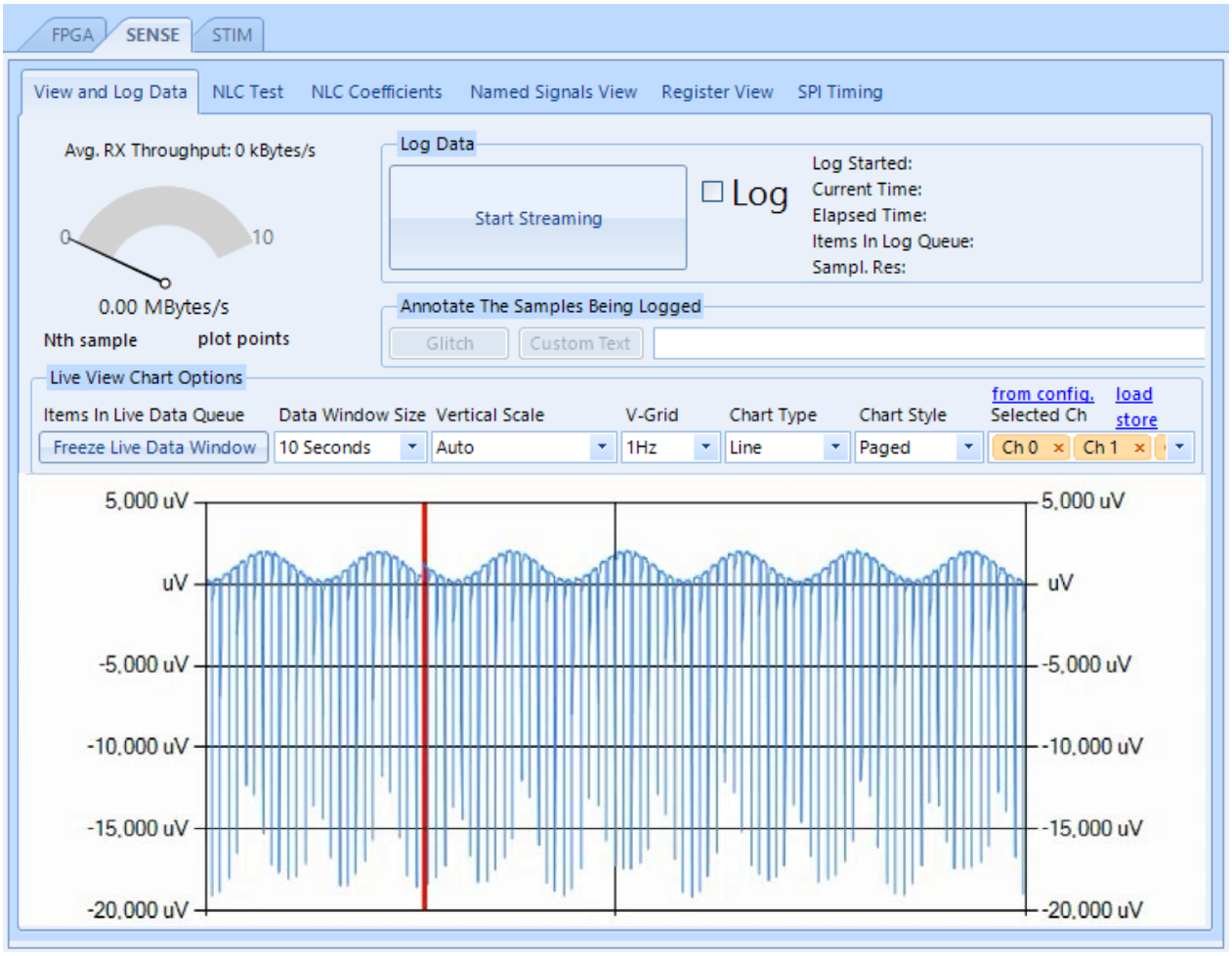

Figure 3.27 Sensing IC screen for viewing and logging ADC samples.

Streamed ADC samples can be viewed in real-time and optionally logged into a CSV format file using the "View and Log Data" screen of the sensing UI (Figure 3.27). The viewing section of the screen provides users many options for displaying the samples. The live stream can be displayed in continuous or paged modes. In continuous view mode, a new sample is added to the left of the live screen and entire plot moves to the right by one sample. In paged view mode, new samples flow from left to right with a cursor. At the end of the screen, the cursor jumps to the leftmost side again and the process repeats. This view is commonly used in commercial equipment for displaying electroencephalograms. A live fast Fourier transform view of the incoming samples can also be displayed in the data window. The type of the chart (line, points, step-line), the number of displayed channels, data-window size, and the vertical scale can be modified through the on-screen controls. When data is being logged, users can insert text annotations that corresponding to external events.

The sensing IC SPI timing configuration screen can be used to adjust the SPI SCK timing and the delay between consecutive byte accesses. Manual control of these parameters requires familiarity with the sensing IC datasheet and is useful for debugging SPI interface level issues. For normal use, it is recommended to have the "Auto-adjust SCK Frequency" and "Auto-adjust Interbyte Delay" checkboxes checked. This will populate the SCK frequency and interbyte delay text-boxes with the datasheet recommended values. To test the robustness of the selected SCK speed/delay combination an automated loopback test should be run from the "Test Commands" section of the sensing UI.

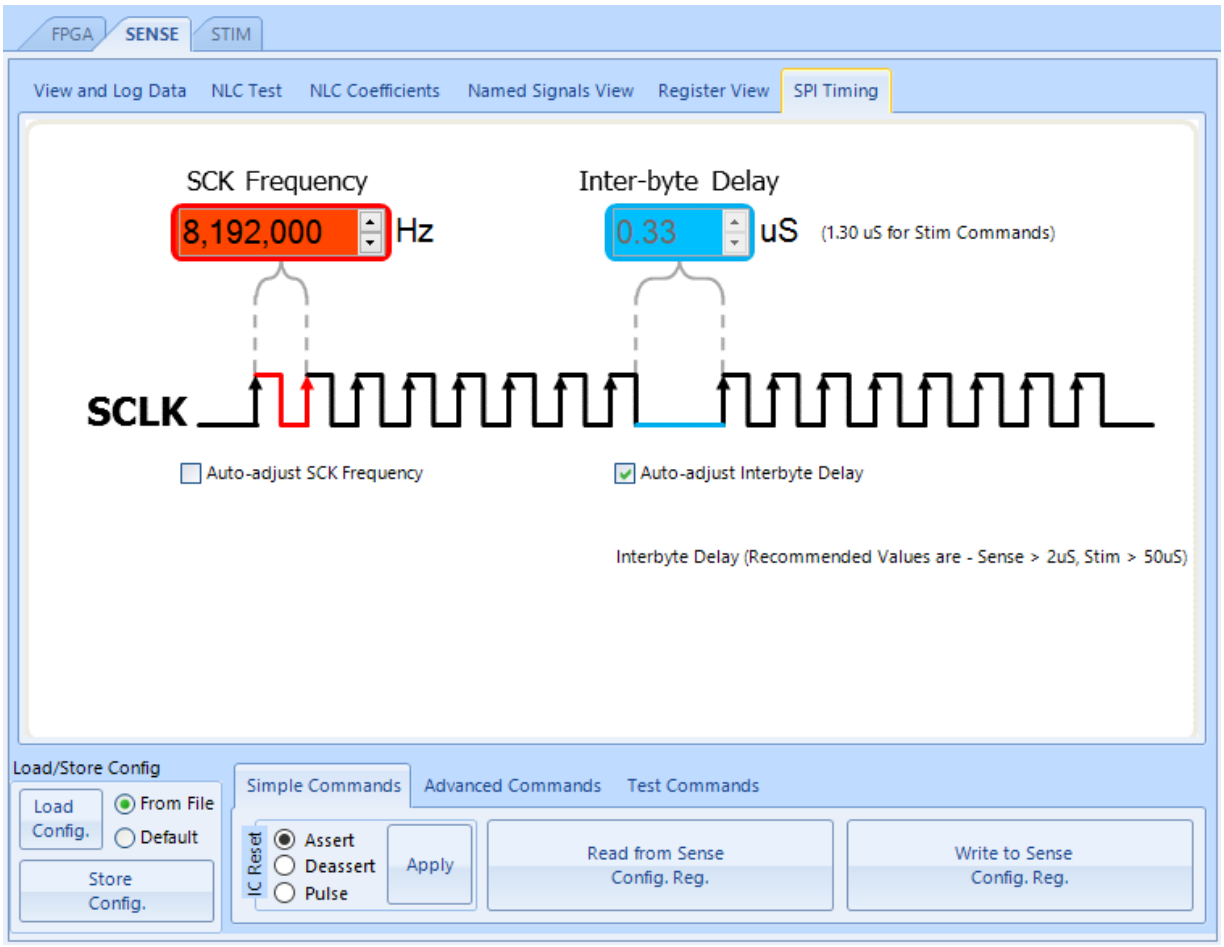

Figure 3.28 Sense IC SPI timing configuration screen.

4 Conclusion

While being compatible with traditional DBS therapy, the proposed neuromodulation system seeks to address its shortcomings – low-specificity and manual programming procedure (for initial setup or for periodic re-adjustments to combat habituation). These (re)programming sessions can particularly be lengthy for certain disorders with long response times (e.g., dystonia). Besides the patient discomfort and downtime, current practice is also inefficient in terms of cost and finding the optimal therapy program. Our system's multichannel nature (64 electrodes vs. the 4 for standard DBS leads) addresses the low-specificity aspect of the DBS therapy, while the high-dynamic-range sensing capability allows for auto-adjustment of stimulation parameters based on the disease state, medication status, and side effects. This closed-loop operation is contingent on neural disorder state decoding, which is a difficult problem and is not addressed in this work. In essence, we are faced with a chicken-and-egg problem, since in order to be able to decode the neural state, clinicians need high-quality neural recordings from relevant targets. On the other hand, in order to design an implant for acquiring high-quality neural data, engineers need to have a quantified need from clinicians to justify the added power overhead. Our goal in this work is to provide the best "chicken" possible with the power budget of an implant.

In addition to having high spatial resolution and being able to sense during stimulation, the proposed system can generate custom, non-rectangular waveforms, and has a modular and distributed architecture. While the custom waveforms can help to increase our system's energy efficiency, the modular and distributed architecture can help treating neuropsychiatric disorders

which are manifested due to dysfunction of several neural systems rather than one distinct anatomical region of the brain - thus opening new avenues for addressing more indications.

4.1 Summary of Research Contributions

The mission of this research is to provide a complete, feature-rich, user-friendly, fully-functional, and safe neuromodulation system, which can address the shortcomings of traditional DBS therapy and enable researchers to explore the biomarker space for neuropsychiatric disorders - thus laying the foundation for making closed-loop therapy possible for many clinical indications. In support of its mission this, work targeted the following aspects of neural implants.

1. Invasiveness
2. Patient safety
3. High-channel-count
4. Ability to sense during stimulation
5. Ability to generate custom, non-rectangular waveforms

This dissertation made the following key contributions to achieve the targets mentioned above.

- Evaluated the nonlinearity of the voltage-to-frequency transfer curve of a VCO-based ADC proposed by Vaibhav Karkare [22] and later improved by Wenlong Jiang [23]. Conducted a benchmarking study to compare the power and area of the traditional lookup table approach for correcting our VCO ADC nonlinearity. The benchmarking results rendered lookup table approach unsuitable for a high resolution and a multichannel system like ours.

- Proposed a new polynomial-based nonlinearity correction mechanism for correcting the VCO ADC nonlinearity. Evaluated the hardware complexity of the straight implementation of the proposed polynomial correction approach. Recognizing that the ADC sampling speed is much slower than the system clock speed, reformulated straight polynomial implementation into an iterative approach using Horner's formula. This area saving reformulation resulted in a much more energy efficient implementation due to leakage-dominated nature of the original design.
- Prototyped the iterative polynomial nonlinearity correction engine in TSMC40GP IC fabrication process and verified its functionality for a four (4) channel system. To scale the design to 32-channels, explored area-saving architectural transformation techniques such as data stream interleaving. Analyzed the hardware complexity of possible interleaving levels for optimal circuit architecture in terms of power and area and selected the 2 x 16-level architecture for the final design.
- In addition to the NLC block, developed a custom communication protocol and implemented the SPI slave, the register map, the communication controller, and the decimation blocks of the sensing IC, thus making it a complete, 32-channel, fully functional implant-scale neural sensing system with 10-times higher linear input range compared to prior art.
- Identified the frequent stimulation parameter updates in closed-loop neurostimulators as a potential source of permanent tissue damage due to possible oxidation-reduction reactions at the electrode-tissue interface. Surveyed the available commercial-off-the-shelf

component datasheets for recommended approaches ensuring charge-balance during parameter updates. Proposed and implemented a new waveform parameter update mechanism which is safe, power efficient, and is independent of the update request's timing and the serial interface speed.

- Implemented a custom-stimulation-waveform shaping engine to take advantage of potential energy savings of some non-rectangular waveforms [6],[7]. Since strength-duration curves and charge injection capacities of neural electrodes are currently defined for rectangular pulses, this engine will enable researchers to characterize strength-duration curves and charge injection capacities of neural electrodes for custom waveform shapes.
- Designed and implemented a custom communication protocol which ensures bi-directional error detection through 1) command and argument checks, 2) unique acknowledgment replies, and 3) cyclic redundancy checks. In case of error events, stimulation IC can be programmed to automatically reset itself and turn off stimulation for patient safety considerations.
- Prototyped a highly integrated yet capable neuromodulation unit by incorporating the above-mentioned stimulation and sensing ICs, along with several passive components on a custom designed printed circuit board. To reduce the number of wires running from AM's SPI master to both SPI slave ICs residing on NMU, a unique mechanism was devised which allowed for accessing two SPI slaves using only three wires compared to required five for conventional SPI interconnection. The NMU was used to demonstrate the saturation-free concurrent stimulation and sensing capability of our hardware, which will

allow neuroscientists to explore the dynamics of neural networks during the stimulation since the recordings are not going to be "blanked."

- Built a custom test platform for verifying the functionality and performance of the proposed NMU. This platform consists of an FPGA-based PC-link dongle housing a custom communication controller, which allows command, status, and data exchange between the NMU and our custom developed the user-friendly graphical interface. The latter gives users access to all aspects of the neuromodulation unit and along with the NMU and PC-link dongle can be deployed in research lab settings.

4.2 Future Work

When building upon this work, it will be worthwhile to consider the following points.

- Current NLC engine operates in foreground calibration mode – the coefficients are calculated in Matlab using the "polyfit" function then transferred to sensing IC's coefficient configuration memory. While due to reasons discussed in Section 2.2.3 the usage of foreground calibration is justified for implantable applications, this method introduces an additional manual calibration step. Switching to background calibration will eliminate this step, although it will introduce a need for an energy efficient coefficient derivation algorithm and an on-chip digital-to-analog converter (DAC). An added benefit of this background calibration could be the inclusion of the tissue-electrode interface nonlinearity when calculating the coefficients.
- While the current stimulation waveform programming interface is intuitive and user-friendly, it is humanly impossible to visualize the neural tissue activation volume generated

by the programmed parameter set. This is a common shortcoming among neurostimulator programming interfaces; however, it is especially important for a high-channel-count device such as ours. One way to address this shortcoming will be to come up with a finite element model that can render a given coefficient set into a three-dimensional geometry. This rendered neural activation volume geometry can then be overlayed on an actual brain scan and give clinicians a clear picture of which neural circuits are going to be activated by a given stimulation parameter set. A better solution will be for a clinician to draw a 3D geometry on a brain scan and let the neurostimulator user interface to convert that geometry into a parameter set(s) that can create such neural activation volume.

- While providing higher spatial resolution, the high-channel-count devices put an immense burden on clinicians who need to select effective stimulation parameters from a much larger set. Having an automated scan-and-select mode for finding the optimal therapy would be a great relief for clinicians, although due to yet absent neuroscientific framework this automation idea seems far-fetched.
- In the current iteration of the system, the closed-loop algorithm runs on a microcontroller residing in the CM. This is appropriate since the improved biomarker extraction algorithms can be applied to the system through firmware updates. When these algorithms mature, it will be more energy efficient to "bake" these feature extraction algorithms into silicon as hardware acceleration blocks.

References

[1] "WHO | Neurological disorders affect millions globally: WHO report," *WHO*. [Online]. Available: http://www.who.int/mediacentre/news/releases/2007/pr04/en/. [Accessed: 26-Feb-2016].

[2] "NIMH » U.S. Leading Categories of Diseases/Disorders." [Online]. Available: https://www.nimh.nih.gov/health/statistics/disability/us-leading-categories-of-diseases-disorders.shtml. [Accessed: 27-Aug-2017].

[3] "Worldwide Neuroscience Initiatives," *Society for Neuroscience*. [Online]. Available: http://www.sfn.org/advocacy/neuroscience-funding/worldwide-neuroscience-initiatives.

[4] P. Krack *et al.*, "Five-Year Follow-up of Bilateral Stimulation of the Subthalamic Nucleus in Advanced Parkinson's Disease," *N. Engl. J. Med.*, vol. 349, no. 20, pp. 1925–1934, Nov. 2003.

[5] P. J. Rossi *et al.*, "Proceedings of the Third Annual Deep Brain Stimulation Think Tank: A Review of Emerging Issues and Technologies," *Front. Neurosci.*, vol. 10, Apr. 2016.

[6] A. Wongsarnpigoon and W. M. Grill, "Energy-efficient waveform shapes for neural stimulation revealed with a genetic algorithm," *J. Neural Eng.*, vol. 7, no. 4, p. 046009, 2010.

[7] M. Sahin and Y. Tie, "Non-rectangular waveforms for neural stimulation with practical electrodes," *J. Neural Eng.*, vol. 4, no. 3, p. 227, 2007.

[8] "Remarks by the President on the BRAIN Initiative and American Innovation," *whitehouse.gov*. [Online]. Available: https://www.whitehouse.gov/the-press-office/2013/04/02/remarks-president-brain-initiative-and-american-innovation. [Accessed: 19-Oct-2015].

[9] "DARPA and the Brain Initiative." [Online]. Available: https://www.darpa.mil/program/our-research/darpa-and-the-brain-initiative. [Accessed: 03-Aug-2017].

[10] M. K. Lyons, "Deep brain stimulation: current and future clinical applications," *Mayo Clin. Proc.*, vol. 86, no. 7, pp. 662–672, Jul. 2011.

[11] "RNS System User Manual," Mar-2017. [Online]. Available: https://www.neuropace.com/manuals/RNS_System_User_Manual.pdf. [Accessed: 02-Aug-2017].

[12] W. M. Chen *et al.*, "A Fully Integrated 8-Channel Closed-Loop Neural-Prosthetic SoC for Real-Time Epileptic Seizure Control," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 286–287.

[13] Y. P. Lin *et al.*, "A Battery-Less, Implantable Neuro-Electronic Interface for Studying the Mechanisms of Deep Brain Stimulation in Rat Models," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 1, pp. 98–112, Feb. 2016.

[14] Y. K. Lo *et al.*, "A 176-Channel 0.5cm3 0.7g Wireless Implant for Motor Function Recovery after Spinal Cord Injury," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 382–383.

[15] S. Stanslaski *et al.*, "Design and Validation of a Fully Implantable, Chronic, Closed-Loop Neuromodulation Device with Concurrent Sensing and Stimulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 410–421, Jul. 2012.

[16] "How the RNS® System Works | NeuroPace, Inc.," *NeuroPace, Inc.* [Online]. Available: http://www.neuropace.com/the-rns-system/. [Accessed: 03-Aug-2017].

[17] "Press Release | Newsroom | Medtronic | New Medtronic Deep Brain Stimulation System the First to Sense and Record Brain Activity While Delivering Therapy." [Online]. Available: http://newsroom.medtronic.com/phoenix.zhtml?c=251324&p=irol-newsArticle&id=1845602. [Accessed: 03-Aug-2017].

[18] "Brief Functional Descriptions of Medtronic Investigational Research Tools for Consideration in Support of the NIH BRAIN Initiative." [Online]. Available: https://www.braininitiative.nih.gov/pdf/Medtronic_Device_Exhibits_0915_508C.pdf.

[19] R. Muller *et al.*, "24.1 A miniaturized 64-channel 225 µW wireless electrocorticographic neural sensor," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 412–413.

[20] R. Muller, S. Gambini, and J. M. Rabaey, "A 0.013 , 5 , DC-Coupled Neural Signal Acquisition IC With 0.5 V Supply," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 232–243, Jan. 2012.

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

[22] V. Karkare, H. Chandrakumar, D. Rozgić, and D. Marković, "Robust, reconfigurable, and power-efficient biosignal recording systems," in *Proceedings of the IEEE 2014 Custom Integrated Circuits Conference*, 2014, pp. 1–8.

[23] 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 J. Solid-State Circuits*, vol. 52, no. 1, pp. 173–184, Jan. 2017.

[24] J. F. Ewen and M. Soyuer, "VCO bias circuit with low supply and temperature sensitivity," US5440277 A, 08-Aug-1995.

[25] M. Z. Straayer and M. H. Perrott, "A 12-Bit, 10-MHz Bandwidth, Continuous-Time ADC With a 5-Bit, 950-MS/s VCO-Based Quantizer," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 805–814, Apr. 2008.

[26] G. Taylor and I. Galton, "A Mostly-Digital Variable-Rate Continuous-Time Delta-Sigma Modulator ADC," *IEEE J. Solid-State Circuits*, vol. 45, no. 12, pp. 2634–2646, Dec. 2010.

[27] J. Daniels, W. Dehaene, M. Steyaert, and A. Wiesbauer, "A 0.02mm2 65nm CMOS 30MHz BW all-digital differential VCO-based ADC with 64dB SNDR," in *2010 Symposium on VLSI Circuits*, 2010, pp. 155–156.

[28] J. Kim, T. K. Jang, Y. G. Yoon, and S. Cho, "Analysis and Design of Voltage-Controlled Oscillator Based Analog-to-Digital Converter," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 57, no. 1, pp. 18–30, Jan. 2010.

[29] L. Chang *et al.*, "Stable SRAM cell design for the 32 nm node and beyond," in *2005 Symposium on VLSI Technology, 2005. Digest of Technical Papers*, 2005, pp. 128–129.

[30] H. Kassiri *et al.*, "Battery-less Tri-band-Radio Neuro-monitor and Responsive Neurostimulator for Diagnostics and Treatment of Neurological Disorders," *IEEE J. Solid-State Circuits*, vol. 51, no. 5, pp. 1274–1289, May 2016.