Power Conditioning and Stimulation for Wireless Neural Interface ICs

Will Biederman

Electrical Engineering and Computer Sciences
University of California at Berkeley

December 17, 2014

Power Conditioning and Stimulation for Wireless Neural Interface ICs

by

William James Biederman III

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jan M. Rabaey, Chair
Associate Professor Elad Alon
Professor Paul K. Wright

Fall 2014

Power Conditioning and Stimulation for Wireless Neural Interface ICs

Abstract

Power Conditioning and Stimulation for Wireless Neural Interface ICs

by

William James Biederman III

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Jan M. Rabaey, Chair

Brain machine interfaces have the potential to revolutionize our understanding of the brain, restore motor function, and improve the quality of life to patients with neurological conditions. In recent human trials, control of robotic prostheses has been demonstrated using micro-electrode arrays, which provide high spatio-temporal resolution and an electrical feedback path to the brain. However, after implantation, scar tissue degrades the recording signal-to-noise ratio and limits the useful lifetime of the array. This work presents two systems which utilize wireless techniques to mitigate this effect and create high-density, long-term interfaces with the human brain.

A wirelessly powered 0.125mm^2 65nm CMOS IC integrates four 1.5µW amplifiers (6.5µVrms input-referred noise with 10kHz bandwidth) with power conditioning and communication circuitry. Multiple nodes free-float in the brain and communicate via backscatter to a wireless interrogator using a frequency-domain multiple access communication scheme. The full system, verified with wirelessly powered in vivo recordings, consumes 10.5µW and operates at 1mm range in air with 50mW transmit power.

A 65nm CMOS 4.78mm^2 neuromodulation SoC integrates closed loop BMI functionality on a single IC which can be arrayed on a wireless sub-cranial platform. The IC consumes 348µA from an unregulated 1.2V supply while operating 64 acquisition channels with epoch compression (at an average firing rate of 50Hz) and engaging two stimulators (with a pulse width of 250µs/phase, differential current of 150µA, and a pulse frequency of 100Hz). Compared to the state of the art neural SoCs, this represents the lowest area and power for the highest integration complexity achieved to date.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to start by thanking my adviser, Professor Jan Rabaey for his wisdom, guidance, encouragement and advice throughout graduate school. I could not have wished for a better experience at UC Berkeley or advisor.

I would like to thank Professor Elad Alon for countless insightful discussions and his incredible teaching. Elad's seemingly endless knowledge and his ability to quickly understand and dissect complex problems continually amazes me.

I would like to thank Professor Jose Carmena for collaborating on many projects and serving on the committee for my qualifying exam. I would also like to thank Professor Paul Wright for serving on both my dissertation and qualifying exam committees and reading this thesis.

I would like to thank Dan Yeager for being a great friend, classmate and research partner. Working with a close friend made this journey and the many long nights at BWRC much more pleasant. I would also like to thank Nathan Narevsky and Jackie Leverett for their friendship, help and contributions to the systems presented in this thesis.

Last by not least, I would like to thank my parents for their infinite love and support throughout my undergraduate and graduate studies.

Chapter 1

Introduction to Brain Machine Interfaces

The burden of neurological disorders has a substantial global impact on society. In 2015, for every 100,000 individuals, 1186.3 years of healthy life will be lost due to neurological disorders [1], Figure 1.1. In total, neurological disorders make up 14.23% of all healthy years of life lost due to disability, resulting in approximately 85 million years of life lost world wide at the time of this writing (2014). In addition to the emotional and physical pain inflicted on patients and families, this massive loss of life has incalculable economic consequences, both direct (cost of patient care) and indirect (loss of human productivity). Over the past few decades there has been a dramatic increase in the use of electronics and medical implants to combat disorders for humans. Some of the most common applications include artificial pacemakers and cochlear implants. Until recently, many neurological disorders have not seen the benefits of technological advances in medicine, however, research has now shown that the quality of life of patients can be improved through a direct interface with the brain.

Brain Machine Interfaces (BMIs) create a direct interface between the human brain and a machine. This enables stimulation of brain regions to mitigate the effect of some neurological conditions (ex. Parkinsons) or allows an alternate method of interacting with the physical world. For example, a patient with spinal cord damage could use a BMI to perform voluntary motor actions using an artificial actuator in virtually the same way that we see, walk or grab an object with our own natural limbs.

The vision of restoring motor functionality to handicapped patients using BMIs has existed for many decades. In 1972 the first cochlear implants became commercially available, which paved the way for BMIs as a treatment for patients suffering from severe motor impairment such as amyotrophic lateral sclerosis (ALS), spinal cord injury, stroke, and cerebral palsy. BMI research began in 1970 at UCLA; however, it wasnt until 1999 when the first experimental demonstration showed neurons could be used to directly control a robotic arm [2]. The use of BMI to improve the quality of life for these individuals is promising, however, due to the a high level of invasiveness and risk BMI primarily remains a research field.

The modern vision of a closed loop BMI and the different components required for im-

Cause category/sequelae	2005		2015		2030	
	YLDs (100 000 population)	Percentage of total YLDs	YLDs (100 000 population)	Percentage of total YLDs	YLDs (100 000 population)	Percentage of total YLDs
Epilepsy	64.7	0.73	60.9	0.73	55.6	0.71
Alzheimer and other dementias	147.4	1.66	165.4	1.98	203.9	2.60
Parkinson's disease	17.7	0.20	17.3	0.21	17.1	0.22
Multiple sclerosis	20	0.23	19.3	0.23	18.4	0.23
Migraine	118.9	1.34	108.9	1.31	96	1.22
Cerebrovascular disease	176.8	2.00	174.9	2.10	177.8	2.27
Neuroinfections	98.4	1.11	71.8	0.86	45.6	0.58
Nutritional deficiencies and neuropathies	194.9	2.20	174.3	2.09	133.9	1.71
Neurological injuries	425.4	4.80	393.5	4.72	360.8	4.60
Total	1264.2	14.27	1186.3	14.23	1109.1	14.14

Figure 1.1: A World Health Organization table of years lost to disability (YLD), summarizing the number of healthy years of life lost from various neurological disorders [1].

plementation is shown in Figure 1.2. This figure depicts a large scale array of electrodes implanted in the brain used for both sensing and feedback. Neural signals are sensed and amplified for signal processing, where the information of interest is extracted. This information is analyzed in real-time, resulting in the implementation of an action (such as movement in a direction), which is relayed through a wireless interface to a multi-channel mechanical actuator. The actuators could control a limb with several degrees of freedom to perform the desired action. Sensors on the limb (e.g. pressure sensors) are used to provide feedback (e.g. tactile sensation) via electrical stimulation to the brain, resulting in a fully closed loop system.

The remainder of the chapter will cover the concepts and technological challenges in realizing a closed loop BMI for humans. Section 1.1 gives a brief overview of the different types of signals which can be measured in the brain. Section 1.2 presents the various types of recording methods and their advantages or disadvantages. Section 1.3 discusses the interface between the implanted electrode and the surrounding brain tissue, and how to model the impedance and noise. Finally, Section 1.4 examines the cell and tissue reaction to an implant and the gives an overview of prior work on mitigating this biological response.

1.1 Signals of the Brain

The neuron is the central element of the nervous system and transmits information through electrical and chemical signals. Like many other cells, neurons maintain a voltage gradient across their membranes by separating ions (e.g. sodium, potassium, chloride and calcium) using ion pumps. Unlike other cells, the membrane of a neuron is electrically active, allowing control of the ion gradients using voltage-gated ion channels, which are activated by
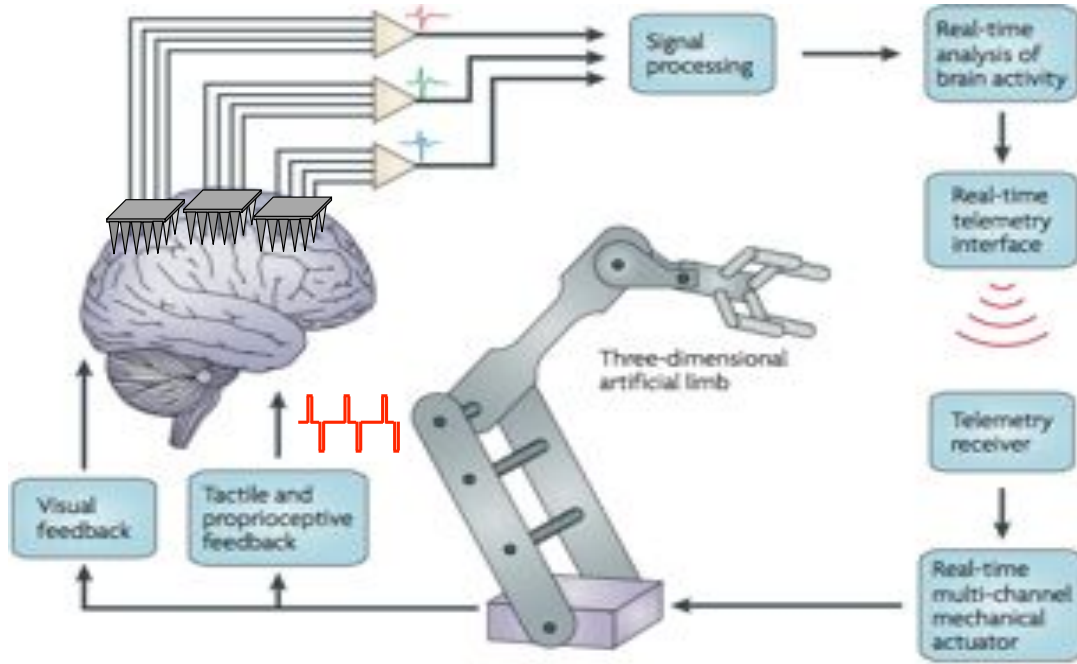
Figure 1.2: Conceptual diagram of a closed loop BMI system. Image adapted from [3].

changes in membrane potential, and chemically-gated ion channels, which are activated by interactions with chemicals in the extracellular fluid.

Neural function is enabled by the communication of neurons through the synaptic process. Typically a neuron's membrane potential rests around -65mV to -70mV, however, synaptic inputs may cause the potential to rise or fall. Once the potential surpasses a threshold, an action potential (AP) is triggered, resulting in rapid depolarization followed by repolarization. Figure 1.3 depicts the typical voltage waveform across a neuron membrane during an action potential, which typically last on the order of 1 ms.

It is possible to measure the action potential voltage in vivo by using implanted micro-electrodes. These electrodes can detect the membrane currents of neurons in the vicinity of the electrode area through conductive brain tissue. Brain tissue can be modeled as linear, resistive and homogeneous, therefore, the numerous signal sources in the corresponding area are assumed to combine linearly [4]. Consequently, using different size electrodes or placing the electrodes in different physiological locations, results in different measured signals. Some of the typical measured signals are summarized in Table 1.1. Low frequency signals are typically the superposition of neural activity over a large area of the brain, resulting in the average signal for a large population of neurons. Depending on the region of the brain, these
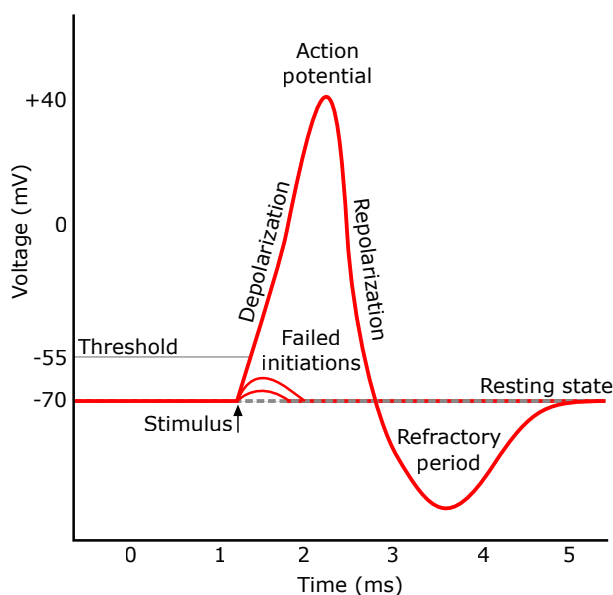
Figure 1.3: A typical action potential voltage waveform measured across a neuron's cell membrane.

Table 1.1: Typical neural recording methods and their resulting measurable signal bands.

Typical Recording Method	Signal Band	Frequency Range (Hz)
EEG	Delta	4
	Theta	4-7
	Mu	8-12
	Alpha	8-15
	Beta	16-31
	Gamma	31-100
ECoG	Local Field Potential	<300
Micro-electrode array	Single & Multi-Unit	300-10,000

signals have been grouped based on the frequency of the signal, such as the Delta, Theta, Alpha, Mu, Beta and Gamma bands. Averaging neural activity over a smaller region gives rise to signals typically less than 300Hz, and has been named the Local Field Potential (LFP). Finally, the limit of recording resolution is measuring the activity of single neurons adjacent to a recording electrode (Single & Multi-Unit), which can have frequency components up to 10kHz.

1.2 Neural Recording Techniques

There are various commonly used neural recording techniques, which can be categorized as invasive or non-invasive. The two most common non-invasive recording techniques are Electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI). EEG, shown in Figure 1.4, measures the changing electric field over time due to the fluctuating ion concentrations during neural activity (action potentials) using electrodes on the outside of the skull, approximately 2cm above the cortex. Consequently, the electrical field (E-field) generated by a single action potential is too weak to be detected through the very lossy channel (brain tissue, skull, skin etc.). Therefore, EEG is only capable of detecting the superposition of the E-fields from the synchronous activity of many neurons across an approximate 3 cm spatial extent, Figure 1.5.

fMRI is a type of MRI that measures changing blood flow related to neural activity in the brain. This imaging technique allows data to be gathered from all regions of the brain (unlike EEG, which is limited to the cortical surface), and has high spatial resolution (down to 1mm). However, fMRI has extremely poor temporal resolution, the Blood-oxygen-level dependence (BOLD) response takes over a second to become detectable. In addition, fMRI requires large machines which make it impractical for any mobile application. While fMRI and EEG are useful for specific applications in neural engineering, currently they do not offer both the spatial and temporal resolution necessary to control robotic prosthesis. Furthermore, neither fMRI or EEG is capable of closed loop BMI control systems using stimulation. In order to achieve the sensing resolution required to seamlessly operate a BMI and the ability to provide feedback, invasive techniques are required.

Invasive BMI techniques enable recordings over smaller cortical areas or the direct recording of action potentials (APs) from single neurons (single units) and small groups (multi-unit) of neurons. Invasive BMIs can detect firing patterns of neurons during the execution of specific and intricate motor functions. These patterns can vary greatly over time and from neuron to neuron, however, averaging across many trials yields fairly consistent patterns. There are several methods of invasive neural recording: intracranial EEG (iEEG), Electrocorticography (ECoG) and microelectrode arrays. Larger recording devices such as ECoG (Figure 1.4) or iEEG, have electrode spacing on the order of 1mm to 1cm and are placed on the surface of the brain. These devices sit directly above the cortex and average neural activity over an approximate 0.5cm range (Figre 1.5) and, therefore, are limited to detecting only Local Field Potentials (LFPs).

There are two types of microelectrode arrays: microwire and silicon micro machined arrays, shown in Figure 1.4. Micro wire arrays are the most frequently used in BMI research; they have only one recording location (at the tip of each wire) and are capable of recording deep in the cortex (up to 5mm). Silicon based probes are physically larger than the smallest microwire arrays but can allow multiple recording sites along the shank. The most common silicon probes are the Utah and the Michigan probes, both of which are manufactured on a silicon substrate using micro electromechanical mechanical systems (MEMS) processing
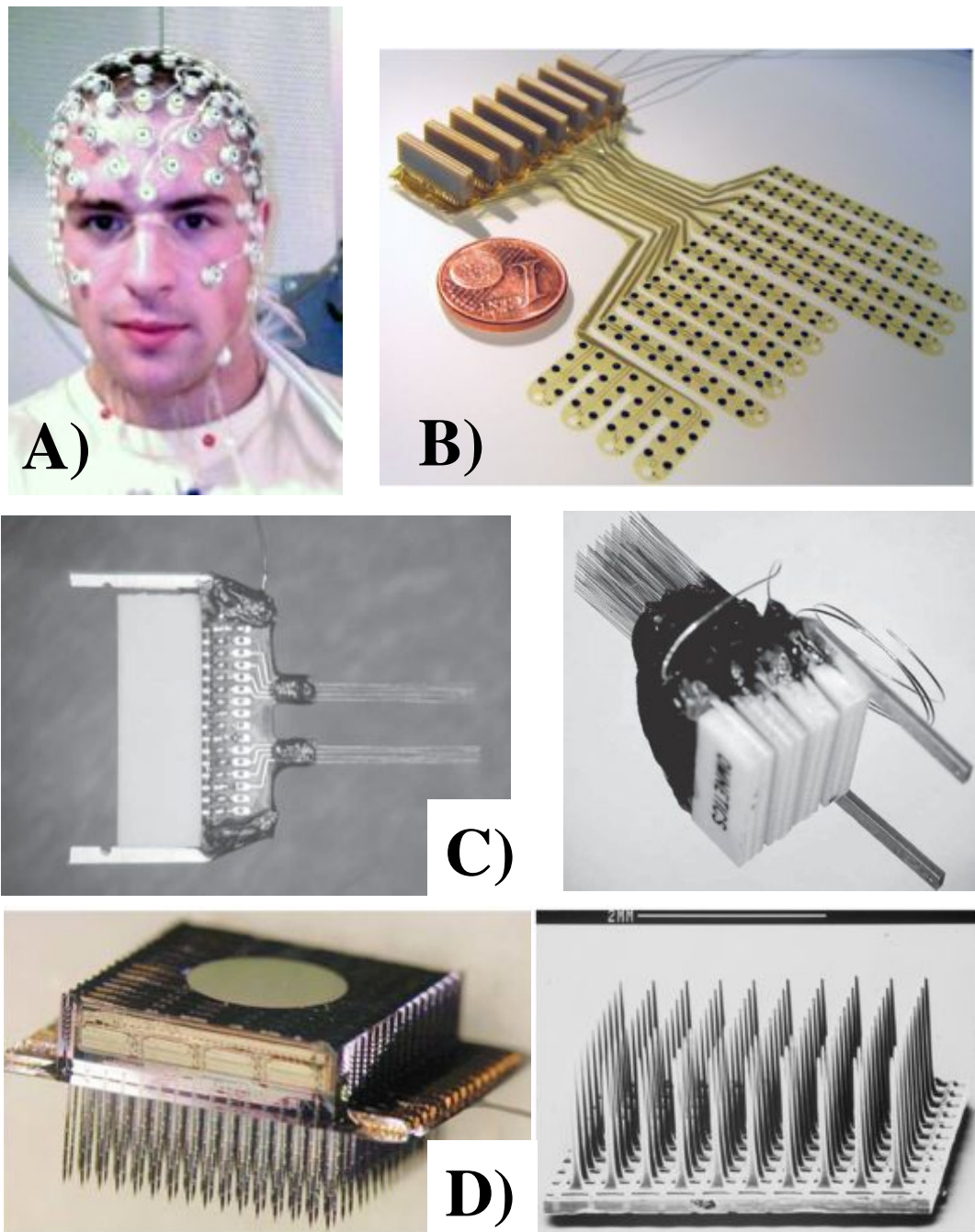
Figure 1.4: Examples of various neural recording techniques. A) Shows a patient wearing an EEG array during a clinical study, B) Shows a micro machined 256 channel ECoG array [5], C) A 16 and 128 channel micro wire array [6], and D) (left) A silicon "Michigan" probe array [7] (right) A silicon "Utah" probe array [8].
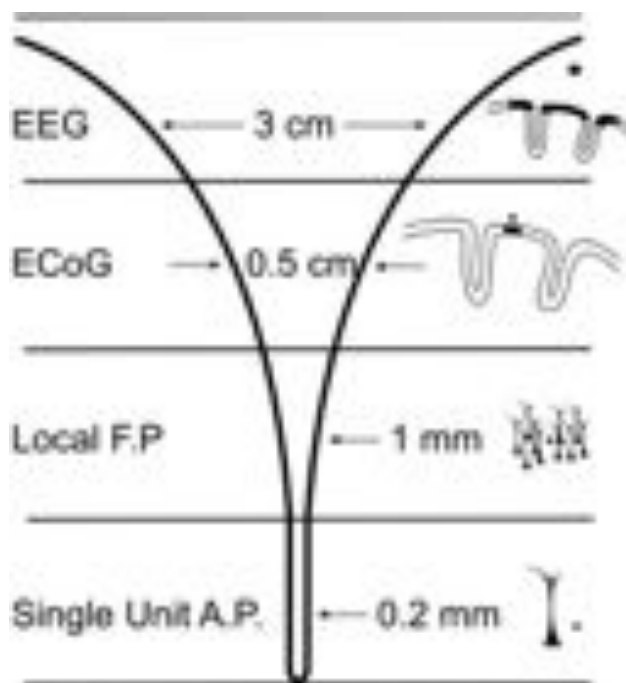
Figure 1.5: A comparison of the spatial resolution for different recording modalities. EEG, located on the skull, averages neural activity from approximately 3cm of cortical area. ECoG, located directly on the cortex, averages neural activity from approximately 0.5cm of cortical area. While LFP and single unit APs are recorded from smaller cortical areas within the brain. [9]

techniques. The Utah array is most similar to the microwire arrays, and only has one recording site at its tip. The Michigan probe has several electrodes positioned along the shank, allowing for recordings at multiple cortical depths.

To date, the only neural recording technique which has successfully demonstrated the control of a robotic prosthetic limb is the direct recording of APs using micro-electrode arrays. Non-invasive methods of sensing brain activity (ex. fMRI or EEG) lack the spatial or temporal resolution. Some research suggests (e.g. [10]) future neuroprosthetic devices could be controlled by ECoG or hybrid sensing solutions, electrodes implanted into the brain are required for stimulation in a closed loop BMI system. Although microelectrode arrays are very effective at providing high spatial and temporal resolution of neural activity, the recording SNR gradually decreases over months to years, eventually rendering the recording site useless. Obtaining a stable, long-term recording of large neurons from a large population of neurons across the brain is one of the key challenges faced by BMI researchers [11].
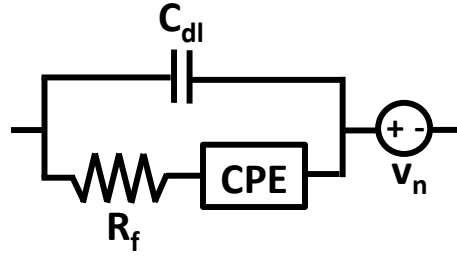
Figure 1.6: A lumped element impedance model of the electrode tissue interface. C_{dl} models the double layer capacitance, R_f represents Faradaic currents, CPE represents a constant phase delay and v_n represents a lumped noise model.

1.3 Electrode-Tissue Interface Model

Providing fine control of robotic prosthesis requires a large number of recording sites, and designing a system which doesn't physically constrain or burden a patient requires a high level of integration. These design constraints, coupled with CMOS scaling and other performance improvements have lead to an interest by the IC design community in the field of neuroscience and BMI. In the past decade there has been a large amount of work on integrated neural recording ICs, for example [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. These ICs are intended to interface directly with recording electrodes, therefore, modeling the eletrode-tissue interface and noise sources properly is critical to achieving the desired *in vivo* performance.

The impedance of the electrode tissue interface can be modeled using lump elements as shown in Figure 1.6. The capacitance, C_{dl}, models the double layer capacitance formed by the interface boundary of the conductive electrode and the brain tissue. The resistance, R_f, models the resistance due to Faradaic current generated through redox reactions at the electrode. Lastly, the constant phase element (CPE) models the signal phase shift resulting from ion diffusion limitations, electrode surface morphologies, and other non-idealities [4]. The voltage source, v_n, represents a lumped element noise source due to both electrical and biological sources other than the signal of interest.

There are two natural sources of cortical recording noise: thermal and biological. Thermal noise is generated by the recording electrode and tissue interface. Biological noise between 500Hz to 5kHz arises from asynchronous neural activity in close proximity to the recording site and is outside of the LFP band. Prior work has modeled thermal and biological noise during cortical recording using silicon microelectrodes and found that for a 450Hz-10kHz recording bandwidth, the recording noise floor is approximately 13.5μV (based on Section 4.2 and Table I from [26]). The total recording and amplifier input referred noise is equal to the sum of their variances, shown in Eqn. 1.1. With an amplifier input referred noise of 6.5μV, the total estimated recording input referred noise, σ_{Total}, is approximately 15μV. Many neural amplifiers target noise floors as low as 1-3μV (ex. [16, 17, 27]), significantly

Table 1.2: Typical Neural Recording Noise Components

Signal Component	Typical Amplitude
Thermal & Biological Noise	13.5µV [26]
Amplifier Noise	1.5 - 10µV
Total Noise	13.58 - 16.8µV

below this recording noise floor, which results in wasted power. Table 1.2 summarizes typical Thermal and Biological noise, amplifier noise and the total overall recording noise.

$$\sigma_{Total} = \sqrt{\sigma_{Amp}^2 + \sigma_{Therm}^2 + \sigma_{Bio}^2} \tag{1.1}$$

An amplifier design can modulate the power spent in the low noise amplifier (LNA) and subsequent amplifiers to set the input referred noise. As the power of the analog front end increases the power overhead of the analog to digital converter (ADC) and digital logic will constitute a lower fraction of the total system power, and the noise efficiency of the entire system will improve. However, this methodology should be used with caution as eventually increasing the amplifier power results in diminishing returns in the overall recorded noise, despite improvements in the amplifier noise performance.

There are several factors which effect the length of time that a chronic neural recording site can remain active. The initial recording sensitivity and selectivity is determined by the proximity of the electrode to the neuron of interest and the electrode size, which sets the baseline signal to noise ratio (SNR) of the recording. The recording stability determines the rate at which the SNR of the targeted neuron changes over time. The stability can be affected by degeneration, damage, or morphological changes to the neuron. In addition, a reactive biological response from implantation causes changes in the surrounding tissue properties and results in modifications of the lumped elements in the ETI model [4]. Understanding, and mitigating this biological response to implanted electrodes is one of the critical challenges to achieving a long term recording interface with the human brain.

1.4   Biological Response of a Neural Implant

The implantation of micro-electrode arrays to record APs causes scar tissue formation, severely degrading the recording signal-to-noise ratio (SNR) over time. The scar tissue effectively isolates the electrodes from the signal source they are trying to detect (the neurons), rendering them useless. Even after the acute inflammatory response declines, there is a chronic increase in the observed number of glial cells around the foreign object, which causes the formation of a glial sheath, as shown in Figure 1.7. Studies have shown that the glial scarring process is similar to the fibrotic encapsulation reaction that occurs with a foreign body in soft tissue areas of the body [28]. There are several different types of cell
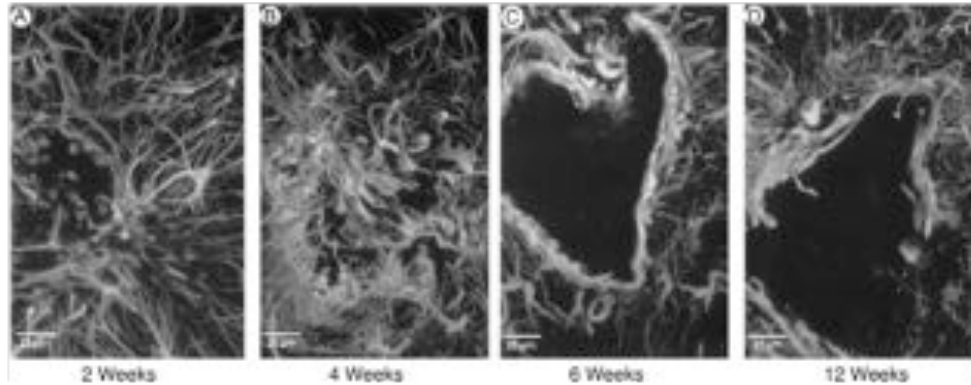
Figure 1.7: Timeline for the formation of glial scar tissue around an implanted neural probe over a 12 week period, [28].

involved in the biological response to foreign objects implanted in the brain, but microglia and astrocytes are the two which are central to the brains response to injury [29].

Microglia primarily act as cytotoxic cells that kill pathogenic organisms, and remain inactive until activated by injury mediated mechanisms. Upon activation, they begin to proliferate, become more compact, phagocytose foreign material, and upregulate the production of lytic enzymes to aid in foreign body degradation [29]. Additionally, since microglia are cytotoxic cells, they can produce neurotoxic factors, which can lead to neuronal death surrounding an activation site.

Astrocytes make up the majority of glial cells in the central nervous system and have many cellular extensions which envelope synapses made by neurons. They provide growth cues to neurons during central nervous system development, mechanically support the mature neuronal circuits, help control the chemical environment of the neurons, buffer the neurotransmitters and ions released during neuronal signaling, and can even modulate the firing activity of neurons [29]. During an injury, astrocytes enlarge in an activation process, transform into a reactive phenotype, and migrate toward the foreign body. Upon activation, astrocytes propagate a foreign body alert in the form of Ca2+ waves.

Reactive astrocytes are the major component of scaring in the central nervous system, and their ability to migrate and communicate to each other via Ca2+ waves ensures that the foreign body is encapsulated long term. Once an implanted electrode is encapsulated, the effective recording impedance to neighboring neurons increases significantly, making the signal undetectable. This process takes several weeks or months depending on the implanted subject. An example timeline of this effect is shown in Figure 1.7.

Improving the reliability of AP sensing using microelectrode arrays has been seen as a challenging long-term goal achievable only by preventing or permanently reducing the biological response to an implant. Prior work (e.g. [30], [31]) investigated the effect of size, shape, texture and insertion method on glial scar formation. Figure 1.8 shows a reduction in the acute reaction of an implant due to a smaller cross sectional area of an electrode site

compared to the insertion shank. However, glial staining revealed that although there were minor temporal differences (on the order of 13 weeks) in the time course of the scarring, at 6 and 12 weeks post-implantation the tissue response to all of these electrodes was essentially identical. Other work investigated bioactive coatings/films containing anti-inflamatory compounds, adhesion promotors or growth proteins with limited success ([32], [33], [34]).

To date, prior work demonstrated improvement in the acute reaction, but has been unsuccessful in preventing chronic scaring. Recording sites can remain active for up to several years [35, 36, 37], and it is hypothesized that they eventually fail after "micro-motion" causes continued aggravation. Micro-motion is the independent movement of the brain with respect to the electrode array, which causes agitation to the surrounding tissue. Studies indicate that reducing or eliminating the effects of micro-motion may be the key to improving implant longevity [38]. In order to mitigate micro-motion the wired interface cables through the skull that connect the implant to recording racks must be eliminated. This requires replacing the large bulky electronics with a recording solution directly integrated on an implanted array and utilizing a wireless link to transfer power and data through the skull. Furthermore, this integrated solution must be small and compliant, such that it can move freely with the brain.

The remainder of this thesis will focus on the use of low power integrated-circuit (IC) techniques to design System-on-Chips (SoCs) to enable high density, long-term recording interfaces with the human brain by addressing the challenges of recording scale and micro motion. Chapter 2 presents the design of the smallest wireless neural sensor reported to date, which free-floats in the brain to improve recording longevity. Chapter 3 presents the highest complexity per-area neuromodulation SoC, combining neural recording, compression and stimulation to perform closed loop BMI integrated in a single IC. Finally, Chapter 4 will summarize the contributions of this thesis and discuss future work.
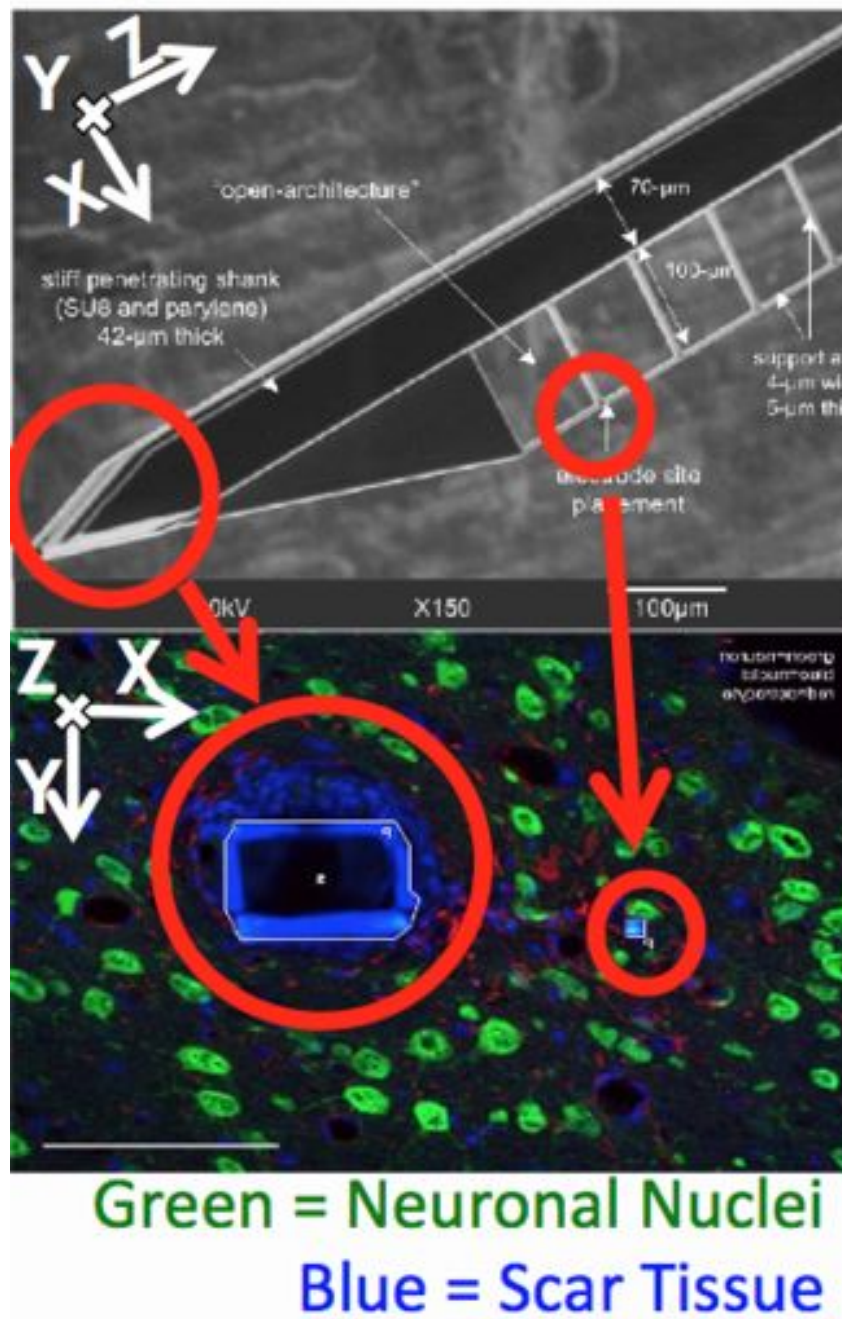
Figure 1.8: A reduction in the cross section area of an implant can mitigate an acute biological reaction. Image is is adapted from [31].

Chapter 2

Design of A Fully-Integrated, Miniaturized (0.125mm^2) 10.5μW Wireless Neural Sensor

To date, the direct recording of APs is the only type of BMI proven to provide enough temporal and spatial resolution to control complex robotic prostheses. However, the implantation of micro-electrode arrays to record APs causes scar tissue formation, severely degrading the recording SNR over time. Studies indicate that reducing the amount of tissue displaced by an implant and eliminating the long-term damage caused by 'micro-motion' effects may mitigate a biological response [38]. Micro-motion is the independent movement of an implant with respect to the brain, resulting in tissue abrasion. This effect can be reduced by eliminating the interface cables and utilizing a wireless link to transfer power and data. Furthermore, the implant should be sufficiently small and light to entirely free-float in brain tissue, eliminating friction with the dura or skull.

Prior work (e.g. [14, 18, 13]) has developed wirelessly powered neural interfaces that utilize large external antennas and bulky off-chip capacitors. To enable an electrode-sized implant to float in brain tissue, an SoC solution with an order of magnitude reduction in active circuit area is required. This reduction in area also reduces the available power, necessitating a similar reduction in power consumption of the circuits. This work achieves a 10x reduction in area and 58x reduction in power, per channel, compared to prior state-of-the-art wirelessly powered neural recording systems. This enables a fully-integrated wireless SoC without the use of any off-chip components.

The proposed system (Fig. 2.1) utilizes a subcranial interrogator to power and communicate with an array of implanted, free-floating AP sensors through the brain's dura. The dura is the outermost membrane surrounding the brain and performs an important biological role; therefore, it is desirable to re-close it after implantation. The sensor nodes are implanted lengthwise, allowing the 4 electrodes to extend deep enough to reach relevant neurons. Four data acquisition channels amplify and digitize the sensed neural potentials into an 800kbps data stream via four 10b, 20kHz ADCs. A single receive (RX) coil on the
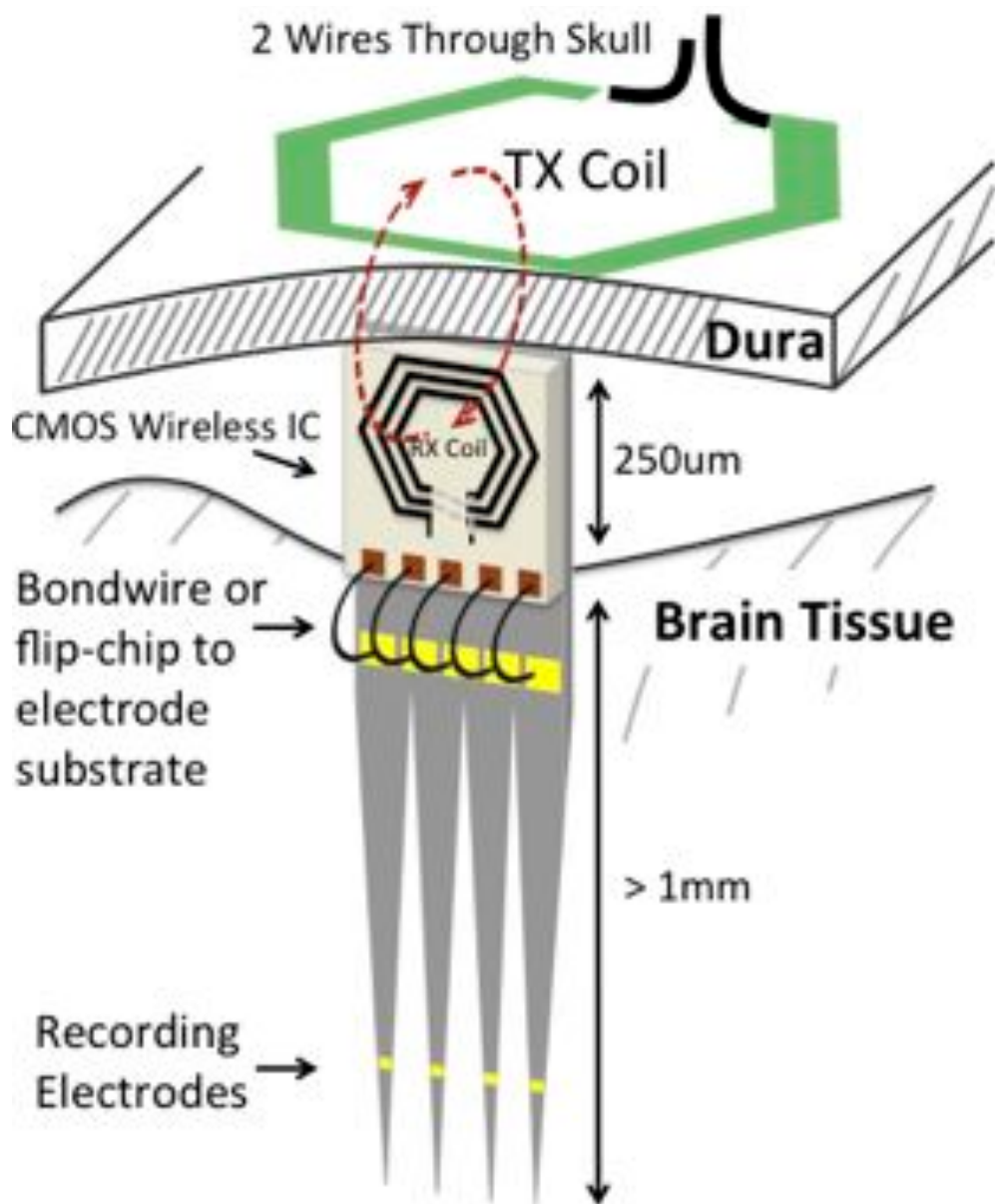
Figure 2.1: A conceptual diagram of the implementation of a BMI utilizing the developed
wireless neural sensor. The sensor free-floats under the dura, while receiving power from and
communicating to an interrogator beneath the skull.

sensor couples perpendicularly to a superdural transmit (TX) coil and achieves both power
and data transmission simultaneously. To further minimize the node's area/volume and
maximize the antenna size, the RX coil is placed on top of the active circuitry.
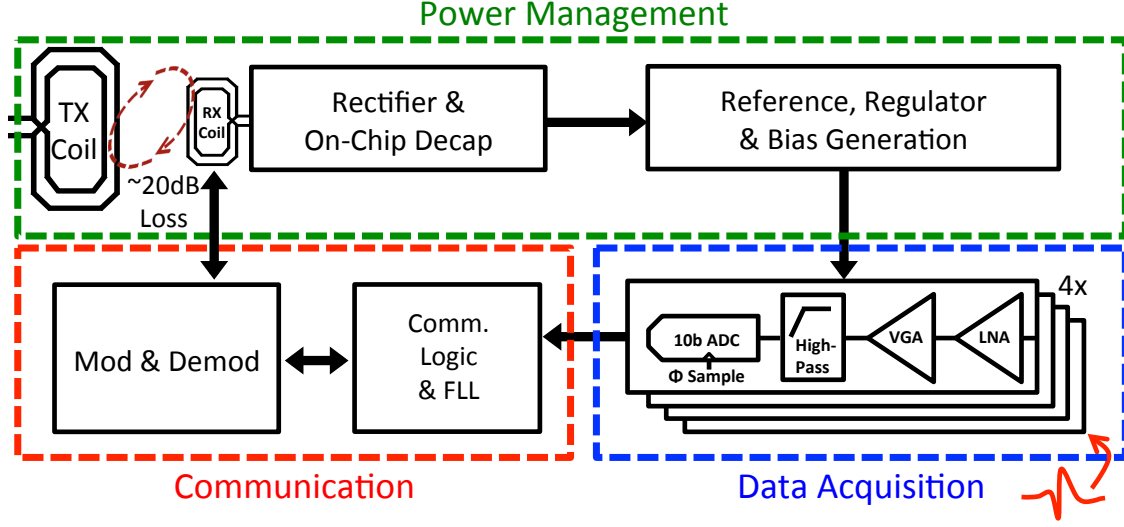
Figure 2.2: System diagram, subdivided into three primary circuit blocks: Power management, communication circuitry, and data acquisition.

2.1 System Design

The constrained node size in Fig. 2.1 places aggressive circuit power and area constraints, which are met through system architecture decisions as well as with the choice of key circuit block topologies. A block diagram of the system architecture is shown in Fig. 2.2. The system components are sub-divided into three categories: Communication Circuitry, Power Management, and Data Acquisition. The core communication circuits include a demodulator, which enables recovery of the low-duty-cycle beacons, and the frequency locked loop (FLL), which generates the Miller subcarrier clock. The power management circuits, consist of a rectifier, supply generation and bias circuitry. A switched-capacitor (SC) bandgap reference was utilized to minimize power and area consumption [39] and self calibration techniques were employed for automatic current and resistor trim. The data acquisition block consists of a multistage neural amplifier, and a 10b counter-based ADC. The system has a total of four amplifiers and input electrodes, which share a common reference electrode.

The system architecture, is ultimately limited by the channel loss for power transfer and the communication protocol data rate. The high data rate and need for a precise clock necessitate an interrogator-provided time base. To enable robust multi-node communication while providing a low-overhead reference clock to the nodes, the communication protocol is optimized for this application using miller-encoded backscatter (Section 2.1.1). The core communication circuits include a demodulator, which enables recovery of the low-duty-cycle beacons, and the frequency locked loop (FLL), which generates the Miller subcarrier clock. The lack of a battery or external antenna requires highly optimized wireless power delivery through careful selection of the node size and wireless transmission frequency, which is discussed in Section 2.1.2.

2.1.1 Communication Protocol

The proposed system enables a single interrogator to wirelessly power multiple implanted nodes. However, each node generates 800kbps (4ch x 10b x 20kHz) of neural data which it must continuously stream to the interrogator. Time interleaving the communication of N nodes reduces the energy per bit by a factor of N, requires N times the data rate per node, and incurs N timing overheads between the time-interleaved communication intervals. Instead, simultaneous transmission by all nodes in unique frequency bands is proposed. For this 5-node system, each node's backscatter is Miller-encoded at a programmable subcarrier frequency between 2MHz and 10MHz. Fig. 2.3 (top) shows conceptual time domain waveforms of 5 wireless packets with this system's possible subcarrier frequencies (2, 4, 6, 8, 10 MHz). Fig. 2.3 (middle) shows the frequency spectrum of 5 nodes transmitting simultaneously. Finally, Fig. 2.3 (bottom) shows a simulated time domain waveform received from 5 nodes (Raw), the band-pass filtered waveform (Filtered) isolating the Miller 4 node and the resulting data as modulation (M4) and raw bits (Data).

The Miller subcarrier frequency of each node must be precise enough such that the interrogator can filter the responses from each frequency channel. The nodes generate a precise local clock with the help of the interrogator, which sends a short downlink beacon pulse every 50μs. The nodes recover this 20kHz clock, which initiates the ADC conversions of neural potentials as well as communication of the 40-bit data packets containing the ADC output. The 2-10MHz Miller subcarrier clock is generated by a frequency-locked loop (FLL), which locks to a multiple of the 20kHz beacons.

To initiate downlink communication, the interrogator sends two consecutive beacons, followed by PPM data. The encoding format is similar to EPC Gen2 RFID [40]. After receiving the response from a unique ID query, the interrogator initializes each node with its unique subcarrier frequency. Downlink communication is only used for initialization of the nodes. Since the downlink configuration packets are infrequent, the node discards the ADC sample when being programmed.

2.1.2 SAR & Frequency Selection

The maximum power available to a node is limited by the the transmission medium, transmission distance, the frequency of operation and the specific absorption rate. For this application, the transmission medium is known and the minimum transmission distance is set by the thickness of the dura above the primary motor cortex (M1). In humans, the the 99.7th percentile ($\mu + 3\sigma$) for thickness of the dura measures 0.61mm [41]. This means that only the node size and transmission frequency are free variables. In biological media, operating at a frequency between 1-3GHz minimizes channel loss for edge-to-edge coupling [42] and reduces the RX coil size by several orders of magnitude compared to [14, 18, 13]. Thus, the transmission frequency for this system was selected to be 1.5GHz, trading a reduction in node size and channel loss for an increase in the specific absorption rate.
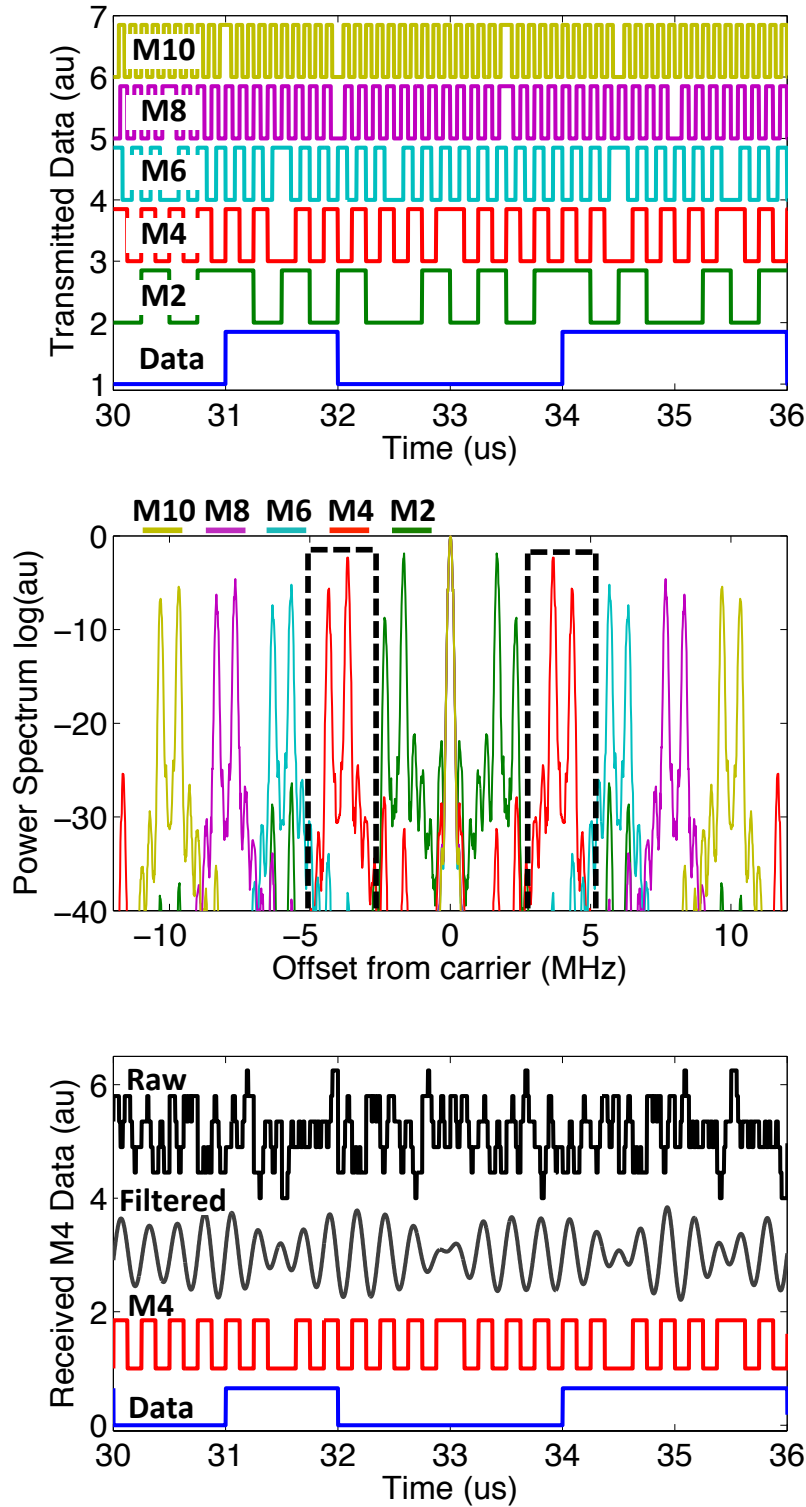
Figure 2.3: Theory of a Miller encoded communication scheme for multi-node interrogation. Top: Miller encoded waveforms (2,4,6,8,10 MHz) for a data set. Middle: The resulting frequency spectrum from 5 nodes communicating simultaneously. Bottom: The recovered raw waveform before and after bandpass filtering, and the recovered original transmitted M4 signal and equivalent data.

All wireless systems must ensure that they adhere to the FDA criteria, which restrict tissue heating to 1°C. Following the IEEE recommendations on Specific Absorption Rate (SAR), which determines how much power a volume of tissue absorbs, can give a rough guideline for meeting the FDA criteria. SAR is defined as the Joule heating per volume mass evoked by an E&M field, Equation 2.1.

$$SAR = (\sigma E^2)/2\rho \tag{2.1}$$

Where ρ is the density of biological tissue (kg/m^3), σ denotes the electric conductivity (S/m) of the tissue and E denotes the electric field strength (V/m). The only method of ensuring compliance to the FDA tissue heating regulation is through simulation and measurement. Furthermore, tissue heating can also result from power dissipation of the implanted electronic devices themselves. Prior work (e.g. [43]) has shown that a power density of 500 μW/mm^2 results in a 1°C increase in surrounding tissue tissue. Therefore, care should be taken to include the effects from both SAR and electronic power dissipation in the calculation of temperature change.

Given a target transmission frequency, a known medium and minimum distance, the channel loss for different size RX antennas can be calculated. The minimum node size achievable is then determined by the antenna design and resulting coupling factor. The antenna optimization is a critical element in overall power transfer efficiency and is discussed in detail in Section 2.2.

2.2 Power Management

The power management circuits convert the inductively-coupled RF power source into a stable DC supply voltage and bias currents for the system. Section 2.2.1 describes the co-optimization of the antenna coil and the rectifier, which convert the incident RF power into an unregulated DC supply. The voltage reference and regulator, described in Section 2.2.2, provide a stable 500mV supply for the digital core and data acquisition channels. Bias generation is discussed in Section 2.2.4, including a basic bias source for the other power management blocks as well as a precision bias generator for the data acquisition channels. Finally, the power-on reset circuit is used to sequence start-up and is described in Section 2.2.5.

2.2.1 Antenna Optimization

A carefully optimized wireless power link minimizes the required amount of transmit power, reducing tissue heating and power consumption of the interrogator. Due to the relative small transmission distance (approximately 1mm), the frequency of operation in this system, 1.5GHz ($\lambda = 154$mm in water), allows modeling the power link as near-field inductive coupling. Eqn. 2.2 approximates the power transfer efficiency, η, where Q' represents the loaded quality factor, Q, of the the transmit (T) and receive (R) inductors [44].

$$\eta = k^2 Q'_T Q'_R \qquad (2.2)$$

Since the amount of magnetic flux captured by the node is constrained by its physical size, the coupling, k, is fixed for a given coil separation. The receive coil quality factor, Q_R is determined by the geometry of the metal turns, as well as constants such as the loss tangent of the silicon substrate. As the number of turns increases, the quality factor decreases due to the required reduction in metal width for a given area constraint, as well as increased substrate losses.

In contrast to the coil Q, the rectifier efficiency improves with the number of turns (to first order) due to the increasing open circuit voltage of the coil. The open circuit voltage is given by Eqn. 2.3, where P_{tx} is the amount of transmitted power and R_p is the effective source impedance of the coil at resonance. R_p can be expressed in terms of the inductance and quality factor as shown in Eqn. 2.4. Improvements in rectifier efficiency must be weighed against losses in power transfer efficiency (η). Optimizations in MATLAB showed that 6 turns maximized the total power transfer efficiency of the link.

$$V_{oc} = \sqrt{\eta P_{tx} R_p} \qquad (2.3)$$

$$R_p = \omega L Q \qquad (2.4)$$

The rectifier is designed to source 10.5μW (15μA at 700mV) and 120pF of output capacitance reduces supply ripple during communication. A two-stage self-synchronous rectifier topology, shown in Fig. 2.4, was found to maximize RF to DC conversion efficiency in this operating region. The coil was designed in an extra-thick aluminum redistribution layer (RDL) with a patterned ground shield (PGS). It occupies almost 500μm x 250μm of area in the top metal layers above other circuits and achieves a quality factor and inductance of approximately 8 and 18nH, respectively. The resulting R_p is 1.36kΩ, yielding a simulated rectifier efficiency of 24%.

2.2.2   Voltage Regulation

Uplink and downlink backscatter communication induce unregulated supply ripple at the programmable subcarrier frequency ranging from 31.25mV at 2MHz to 6.25mV at 10MHz (assuming a 15μA load on the 120pF decoupling capacitor). A discrete time linear regulator, shown in Fig. 2.5, is used to provide a low noise supply for the neural data acquisition circuitry, as well as minimize the dynamic and leakage power of the digital communication logic. A comparator with capacitive offset cancellation (OS$_{pos}$, OS$_{neg}$) is used instead of a linear amplifier in order to provide a high gain-bandwidth with minimal power consumption. A charge pump based loop filter sets the bandwidth as well as output ripple while consuming minimal power and area. Native Vth NMOS power devices are used for both the analog (A$_{vdd}$) and replica digital (D$_{vdd}$) supplies. The regulator consumes less than 300nA at the maximum supply voltage and occupies 55μm x 54μm. Input and output capacitors, including
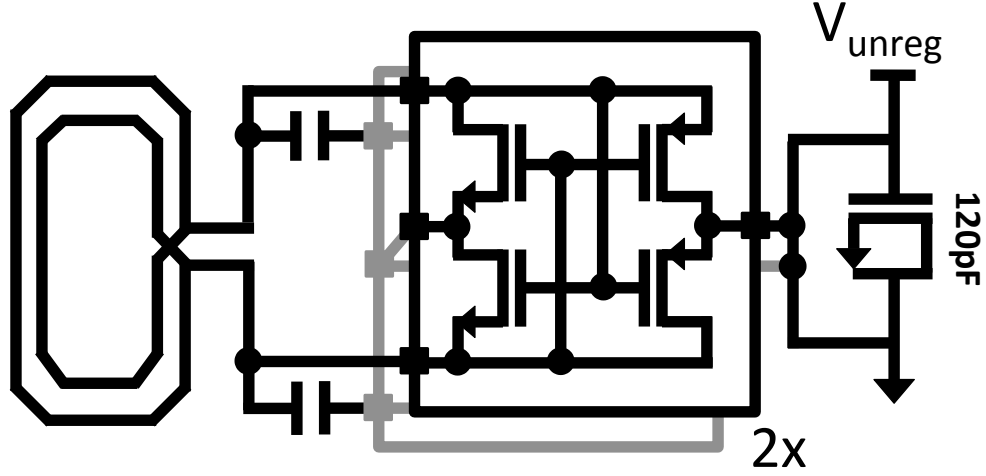
Figure 2.4: A high-R_p on-chip coil increases the open circuit voltage and maximizes the efficiency of the self-synchronous rectifier. 120 pF of on-chip decoupling capacitance is implemented with thick-oxide native devices.
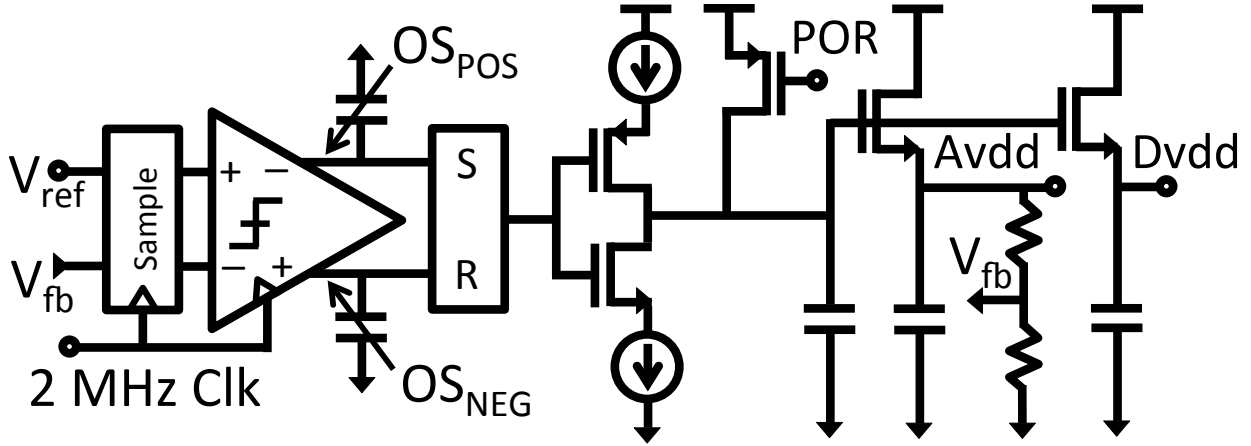
Figure 2.5: Discrete-time LDO regulator schematic utilizing a comparator with capacitive offset cancelation and a charge pump loop filter.

the 120pF decoupling capacitor for V_{unreg}, consume 450µm x 63µm. The measured PSRR across frequency is shown in Fig. 2.6. With a worst case PSRR of 27dB, communication-induced supply ripple is reduced to less than 1.5mV.

The regulator requires a robust precision voltage reference with low area and power consumption. By utilizing a SC bandgap architecture proposed in [39], the reference eliminates the use of resistors, op-amps, process-sensitive MOS V_{th} or leakage-based techniques. This bandgap topology provides drastic area and power savings over the previous state-of-the-art, and the design will be discussed in detail in Section 2.2.3.
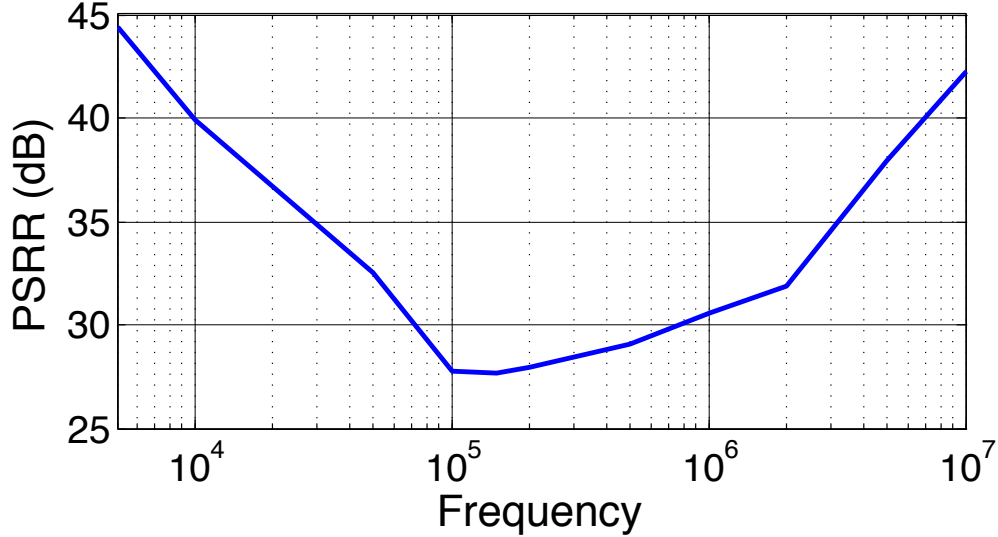
Figure 2.6: The measured discrete-time LDO regulator supply rejection across frequency.

2.2.3 Voltage Reference Generation

Ensuring system operation across a wide range of process corners, supply voltages, and temperatures (PVT) often necessitates the use of an on-chip bandgap reference. However, as process nodes scale to single nm gate lengths, the reduction of oxide thickness will cause the maximum supply voltage to scale far below the minimum operating voltage of conventional bandgap references. Furthermore, many ultra-low power wireless applications such as RFID tags and wireless sensor nodes have aggressive power and area requirements, necessitating improvements on these specifications in current state-of-the-art bandgap designs. This motivates exploration of circuit topologies that can continue to generate a bandgap-referenced voltage while operating from a sub-1V supply with minimum power and area consumption.

The traditional CMOS-compatible bandgap topology is shown in Fig. 2.7. The nominal output voltage of 1.2V requires a supply voltage that is incompatible with modern CMOS technologies. The voltage domain summation of **proportional** and **complementary** to absolute temperature (PTAT and CTAT) signals based on V_{BE} and ΔV_{BE} limits the supply to $V_{BG} + V_{DSAT}$. Prior works have reduced the required supply voltage by adding the PTAT and CTAT terms in the current domain [45, 46]. The drawback of this approach is an area/power tradeoff; mega-ohm resistors are needed to reduce bias currents to microamp levels. For example, the lowest power bandgap to date requires 4.6µW and consumes 0.1mm^2 [47].

The high turn on voltage for silicon diodes has motivated exploration of other voltage references. Prior works have explored the use of MOS V_{TH}, which can be less than half the value of V_{BE}. MOS V_{TH}-based references have reported power levels as low as 36nW [48], and a 2T MOS leakage-based reference reported 2.2pW [49]. In addition, MOS references typically consume much less area, in part due to the lack of diodes. The lowest area MOS
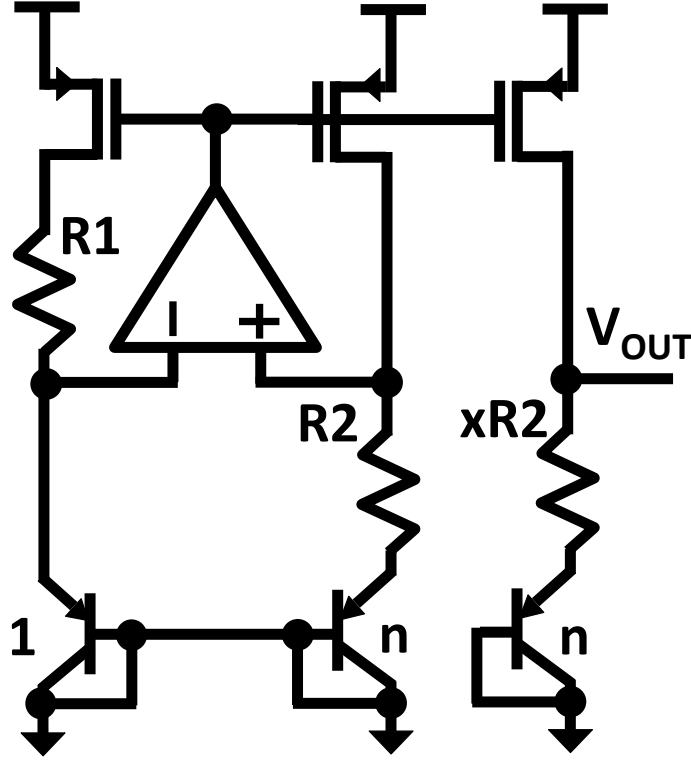
Figure 2.7: A typical bandgap architecture, requiring $V_{DD,min} \approx 1.4\text{V}$.

V_{TH}-based references have achieved areas less than 0.024mm^2 [50], while bipolar references have been reported with areas as low as 0.0445mm^2 [51]. However, because MOS V_{TH} and leakage-based reference voltages are a strong function of process parameters (unlike the V_{BE} of a BJT), bandgap references remain attractive for their robustness.

In this section, I present a method for generating a reference based on the bandgap of Si which improves on the state-of-the-art by approximately 10x in area and power, and is able to operate down to -35°C with a supply of 0.75V. The design utilizes a switched-capacitor (SC) technique inspired by [52] without the use of resistors, op-amps, or process-sensitive MOS V_{TH} or leakage-based techniques.

2.2.3.1 Architecture

A bandgap reference creates an output voltage independent of temperature by summing two voltages with opposite temperature coefficients. These CTAT and PTAT voltages are typically generated from V_{BE} and ΔV_{BE}, which have temperature coefficients (TC) of approximately -2mV/°C and 0.085mV/°C respectively. In the classical bandgap reference circuit shown in Fig. 2.7, the op-amp generates a voltage equal to ΔV_{BE} across R_1. The resulting output voltage is given by (2.5), and its respective TC is given by (2.6). Thus, for this
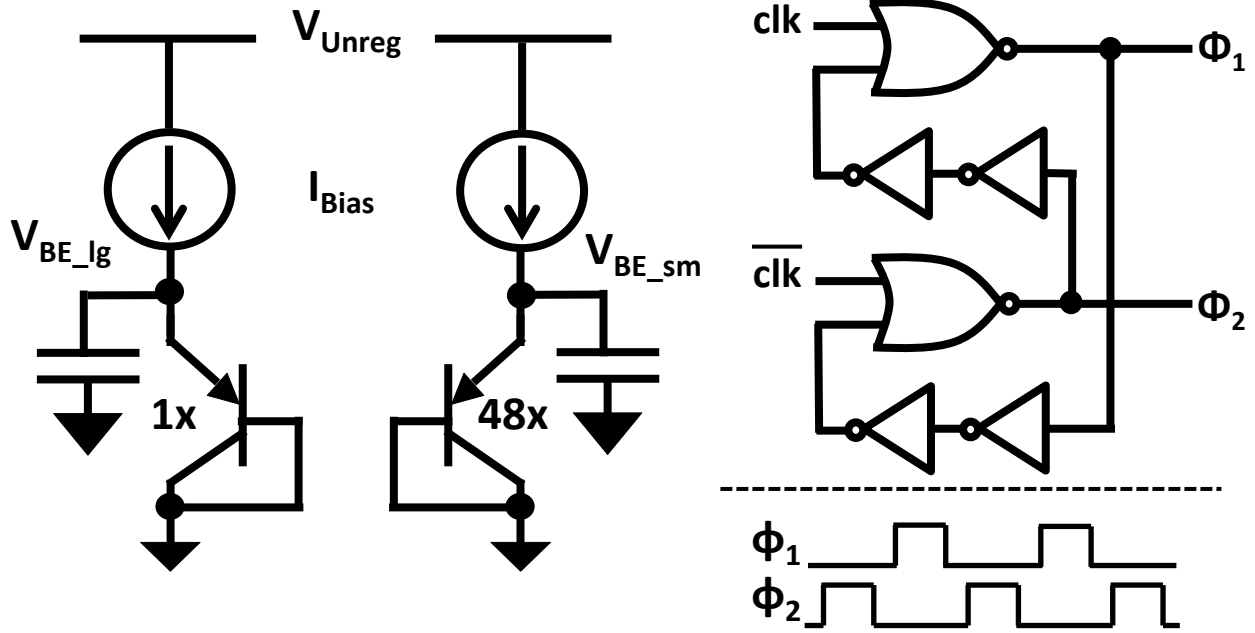
Figure 2.8: Schematics for generation of $V_{BE\_lg/sm}$ and non-overlapping clocks.

circuit to achieve a zero TC, $x \cdot ln(n)$ must equal 23.5 by properly choosing the diode (n) and resistor (x) size ratios.

$$V_{OUT} = V_{BE} + x \cdot ln(n) \cdot V_T \tag{2.5}$$

$$TC_{V_{OUT}} = -2\text{mV/}°\text{C} + x \cdot ln(n) \cdot 0.085\text{mV/}°\text{C} \tag{2.6}$$

While this traditional approach has been successful, the minimum supply voltage is limited by V_{OUT} (i.e. V_{BG}) plus the V_{DSAT} of the current source, typically resulting in a minimum V_{DD} of 1.4V or more. To break this tradeoff, a fractional bandgap reference can be generated by introducing a division term into the output equations shown in (2.7) and (2.8).

$$V_{OUT} = \frac{1}{D} \cdot V_{BE(on)} + M \cdot ln(n) \cdot V_T \tag{2.7}$$

$$TC_{V_{OUT}} = -2\text{mV/}°\text{C} \cdot \frac{1}{D} + M \cdot ln(n) \cdot 0.085\text{mV/}°\text{C} \tag{2.8}$$

Previous works have implemented this scaling with resistive subdivision [45, 46]. However, resistors impose an area/power tradeoff, which is undesirable for low power applications. Instead, a SC network can be used to achieve the same effect in the voltage domain by dividing V_{BE} and multiplying ΔV_{BE} (and consequently their TCs) to achieve a net TC of zero. Table 2.1 shows combinations of division/multiplication factors that produce a TC of

Table 2.1: Table of possible reference voltages and their required multiply and divide factors

V_{BE} Divide Ratio	ΔV_{BE} Multiply Ratio	Theoretical V_{REF}
6	1	196mV
4	1.5	293mV
3	2	391mV
2	3	587mV
1.5	4	782mV
1	**6**	**1174mV ≈ BG$_{Si}$**

approximately zero with a diode multiplier of 48 and their resulting reference voltage. Realistic ratios are shown that could be implemented practically using a SC network. For our application we chose division and multiplication factors of 3 and 2, respectively, which results in a theoretical output voltage of 391mV with a diode bias current of 25nA in this process. Schematics of V_{BE} generation and the implemented SC design are shown in Fig. 2.8 and 2.9, respectively.

When designing the SC network, MOM capacitors should be utilized instead of MOS capacitors for linearity and reduced leakage. Ideally, the output voltage is independent of capacitor size, leading to minimally sized capacitors to save power and area. However, a lower bound on the capacitor size is necessary to prevent error in the output voltage induced by switch leakage. With minimum sized switches in this process, 45fF was found to be sufficient.

In contrast to other voltage reference designs, this architecture requires the use of a clock. In many systems, a clock is readily available or can be generated by a simple ring oscillator. Such a clock may be used for this design despite the wide variability present in an open loop ring oscillator because the frequency has minimal impact on the output voltage. The measured sensitivity in our design from 500kHz to 2MHz is 0.25mV/MHz. Two non-overlapping clocks are required to prevent short circuit current flow in the SC network. Generation of these clocks can be achieved with a low power circuit similar to a SR latch with added delay elements in the feedback, shown in Fig. 2.8. Since the gates can be minimum-sized, this circuit consumes only 4nA at 1MHz in simulation for our process.

Finally, the last component necessary to operate this bandgap architecture is bias current generation. Minimizing the bias current enables both low power operation and a low operating voltage because the lower limit of the supply is equal to V_{BE} plus V_{DSAT}. However, losses due to parasitic capacitances and leakage through the switches in the SC network necessitate a minimum bias current on the order of 25nA based on simulation.

Many systems utilize a $\Delta V_{GS}/R$ reference in order to bias analog circuits such as amplifiers. This temperature dependent current can be used to bias the diodes with minimal impact on the temperature coefficient of the bandgap output voltage. Using a $\Delta V_{GS}/R$ bias may also be desirable to enable independent trimming of the bias current and thus output voltage. This trimming can be used to correct for process variation of V_{BE}. Alternatively, if
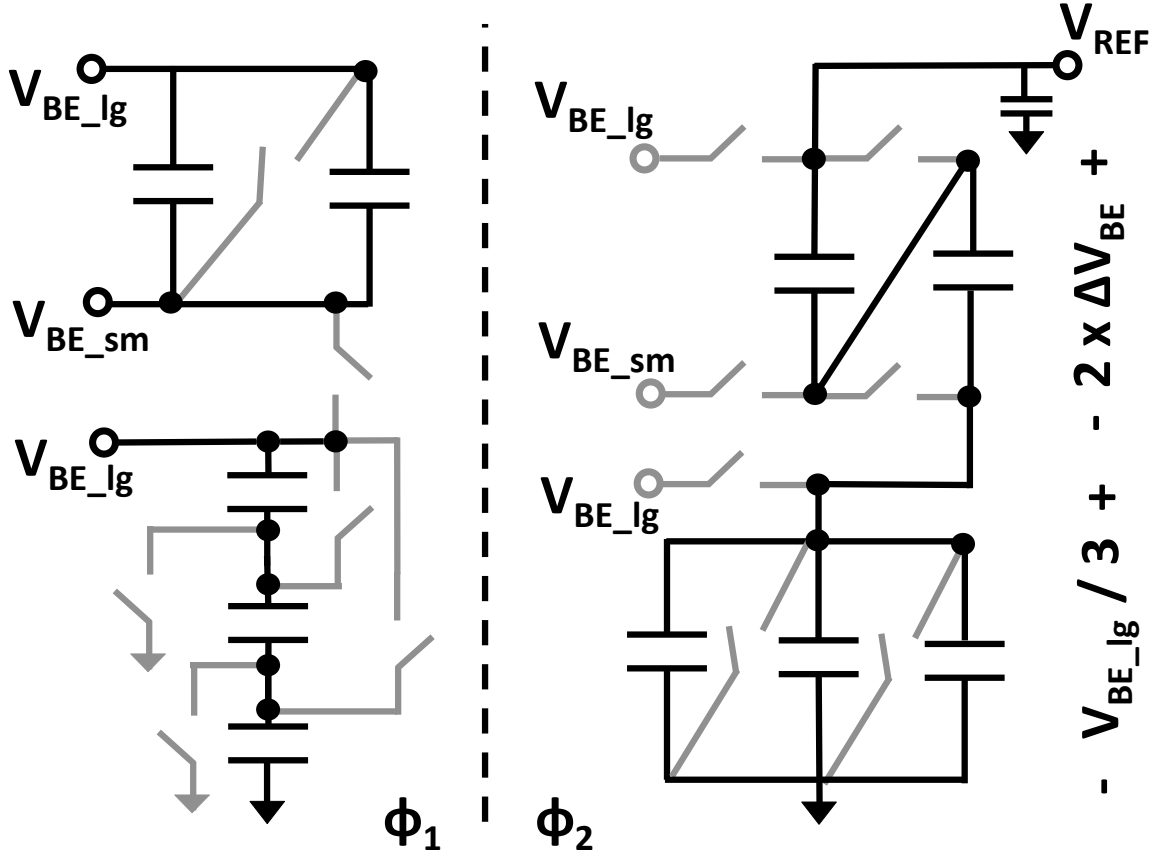
Figure 2.9: Schematic of the SC network with a divide/multiply ratio of 3/2.

a bias current generation circuit is not available, this design can be self-biased in the same way as the traditional design shown in Fig. 2.7 by adding a resistor and op-amp.

In both this design as well as conventional bandgap references, the output voltage is sensitive to the absolute value of the bias current. Simulations of a textbook $\Delta V_{GS}/R$ current source exhibit a σ of $\pm 27\%$ due to mismatch and variation of up to $\pm 26\%$ across process corners. This would induce a net error of up to 2% in the bandgap output voltage.

2.2.3.2 Error Analysis

Typical sources of variability in a traditional bandgap architecture have been analyzed in the literature. Three out of four of the largest output error sources are due to op-amps and resistors [53]. While these error sources are eliminated in this design, parasitic capacitance from the MOS switches, routing, and bottom/top plate of the MOM cap in the SC network induce a large source of error. Parasitic capacitors, with reference to ground, are pre-charged to either V_{BE_lg} or V_{BE_sm} in the first phase. Charge sharing then occurs during the second

Table 2.2: Table of output sensitivity to mismatch, process, and voltage

Circuit Block	Mismatch ($\sigma_\%$)	Process ($\pm\%$)	Supply (mV/V)
Current Sources	3.4	0.15	5.25
Diodes	Not Modeled	0.94	-
S.C. Network	0.35	1.89	9.2

phase. In this design, all of the node voltages during the second phase of the SC network are less than the values stored on the parasitic capacitors. Therefore, charge sharing from the parasitic capacitors always increases the node voltages and, consequently, the output voltage.

Careful layout and design help to minimize the effect of this parasitic charge sharing on the output voltage. For example, error is reduced by referencing the division capacitors to ground rather than to ΔV_{BE} in the second phase. This allows a larger amount of intermediate node parasitic capacitance to be discharged. Furthermore, the use of minimum device sizes for switches and the elimination of lower metal layers on MOM caps reduce parasitics.

An estimation of the parasitic effects can be obtained from simulations of the extracted layout. Compared to an ideal output from the SC circuit of 391mV, the simulated output is 411mV, closely matching a MATLAB model. The measured lot average of 423mV (shown in Fig. 2.12) is within 2.7% of the simulated value. Measured samples are from one lot; the shift in the mean falls within expected process variation.

The majority of mismatch-induced error in the output voltage is due to random dopant fluctuations causing V_{TH} variation in the diode current sources. Mismatch between the diode current sources affects ΔV_{BE}, while variation in the reference current affects V_{BE}. With $M = 2$ and $D = 3$, the output is 6x more sensitive to current source mismatch than to variability in bias current generation. This results in sensitivity coefficients, in percent, of V_{OUT} to $\sigma_{\Delta I_D}$ and σ_{I_D} of 0.43 and 0.072, respectively.

Pelgrom's model provides an equation describing variability in a current mirror: $\sigma^2_{\Delta I_D} = A_{V_{th}}^2/WL + (gm/I_D)^2 \cdot A_\beta^2/WL$. Since minimizing V_{DSAT} is necessary to minimize the supply voltage, this equation reveals that the only design variable available to reduce variability is the device area. However, in future designs, the tradeoff between device size and variability can be broken by chopping the diode current source loads, similar to the dynamic element matching technique.

Output variability is analyzed in simulation from each of the three principle components of the system: the diodes, the diode current sources, and the SC network. Table 2.2 summarizes the effects of mismatch, process, and supply variability on the output voltage. The total effect of process variability on the output voltage is approximately 3%, significantly less than MOS V_{TH} references, which can exceed 10% [48].
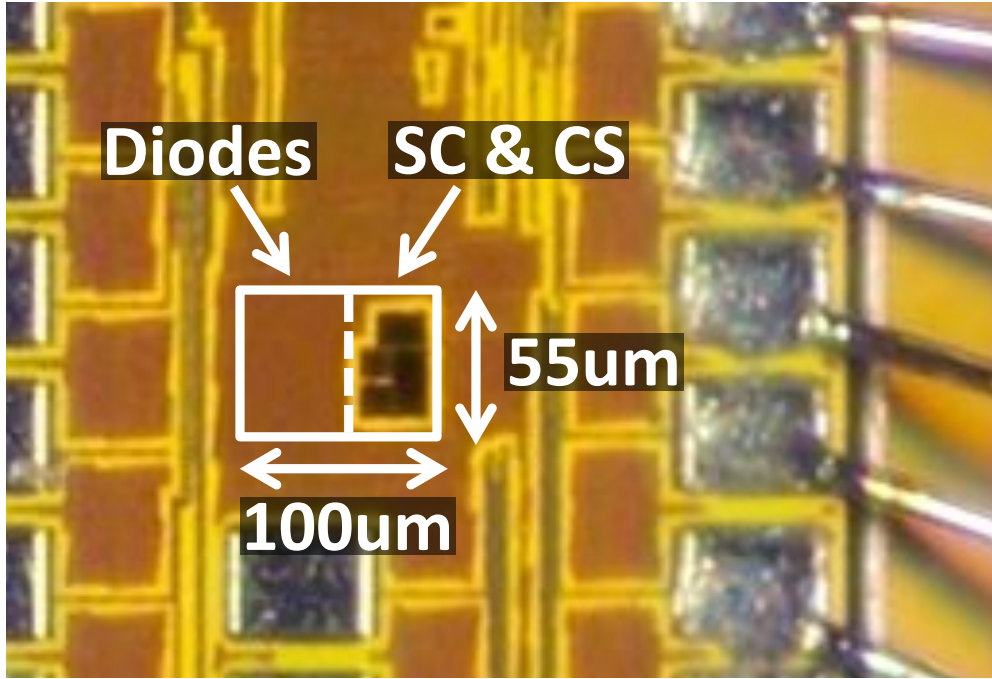
Figure 2.10: Die photo of the implemented design (0.0055mm^2 excluding pads).

2.2.3.3   Performance

The proposed reference was fabricated in a 65nm CMOS process, and the die photo is shown
in Fig 2.10. The measured output voltage from -35°C to 80°C is shown in Fig. 2.11, with a
peak sensitivity of 160ppm/°C. In this plot there are two unique features: at low temperature,
there is a rapid increase in temperature coefficient when the diode voltage becomes large
enough to force current source out of the saturation region. At high temperatures these is
another rapid change in temperature coefficient due to a design error where a power-on-reset
switch was implemented using an LVT device. At high temperatures this switch begins to
leak to the output from the supply, causing an increasing in the output voltage. Fig. 2.12
shows a histogram of measured output voltage variation from 15 die. The equivalent sigma
of 2.2% falls within the Monte Carlo results shown in Table 2.2.

The reference functions as low as -35°C at 750mV supply voltage, or -10°C at 700mV
supply voltage. The current consumption varies with clock frequency; at 1MHz the measured
current is 138nA. Table 2.5 summarizes the measured performance and compares this design
to existing MOS and bandgap-based references. This design reduces the area and current
consumption compared to prior bandgap-based references by 8.1x and 15.9x respectively.
Additionally, this reference functions at 100mV lower supply voltage than existing designs.
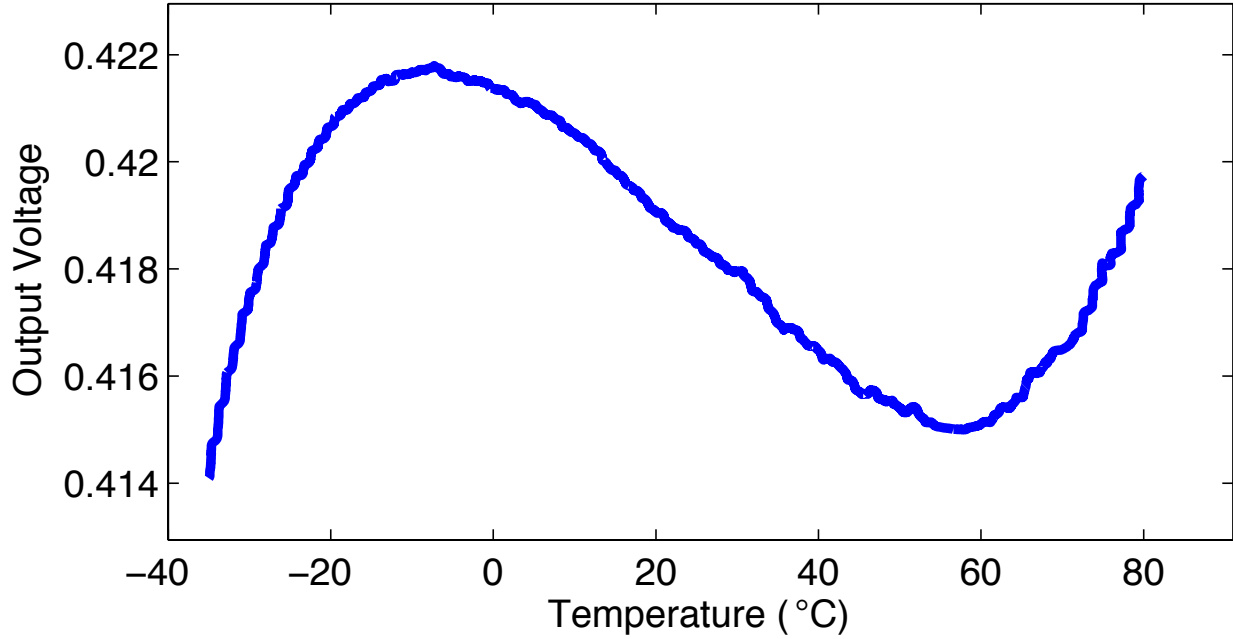
Figure 2.11: Measured output voltage versus temperature. The peak variation from from
-35°C to 80°C equates to a sensitivity of 160ppm/°C.

Table 2.3: Comparison of published (measured) Bandgap Reference performance

Author	Ref. Type	Ref. Voltage (mV)	Min. Supply (V)	Total Current (µA)	Max. TC (ppm/°C)	Area (mm^2)	Meas. σ (%)	Process (nm)
Gambini [54]	MOS	260	0.5	22	136	NR	NR	90
De Vita [48]	MOS	670	0.9	0.04	10	0.045	3.1 [a]	350
Huang [50]	MOS	221	0.85	3.3	323 [b]	0.0238	NR	600
Wang [55] [c]	MOS	400	0.56	4.8	80	0.045	2.5 [a]	180
Ker [56]	Diode	238	0.85	28	58.1	NR	NR	1200
Boni [57]	Diode	493	1	10	22.6 [b]	NR	0.86	350
Ng [51]	Diode	235	0.95	28	34	0.0445	NR	500
Banba [47]	Diode	515	2.1	2.2	58.25	0.1	0.97	400
This Work	**Diode**	**423**	**0.75**	**0.138**	**160**	**0.0055**	**2.2** [a]	**65**

[a] Single lot measurement   [b] Numerical value extracted from plot   [c] Architecture originally proposed by [58]

2.2.3.4   Conclusions

The growing interest in ultra-low power devices and the continuation of Moore's Law have
generated a demand for a low voltage, power, and area bandgap reference. This architecture
uses a SC network for division, multiplication and summation to create an ideally zero TC
reference. Utilizing a SC network to create and sum voltages with opposite TCs has enabled
the lowest current, voltage and area bandgap-based reference reported to date. The design
occupies an area of 100µm x 55µm and consumes 138nA from a 750mV supply. The area
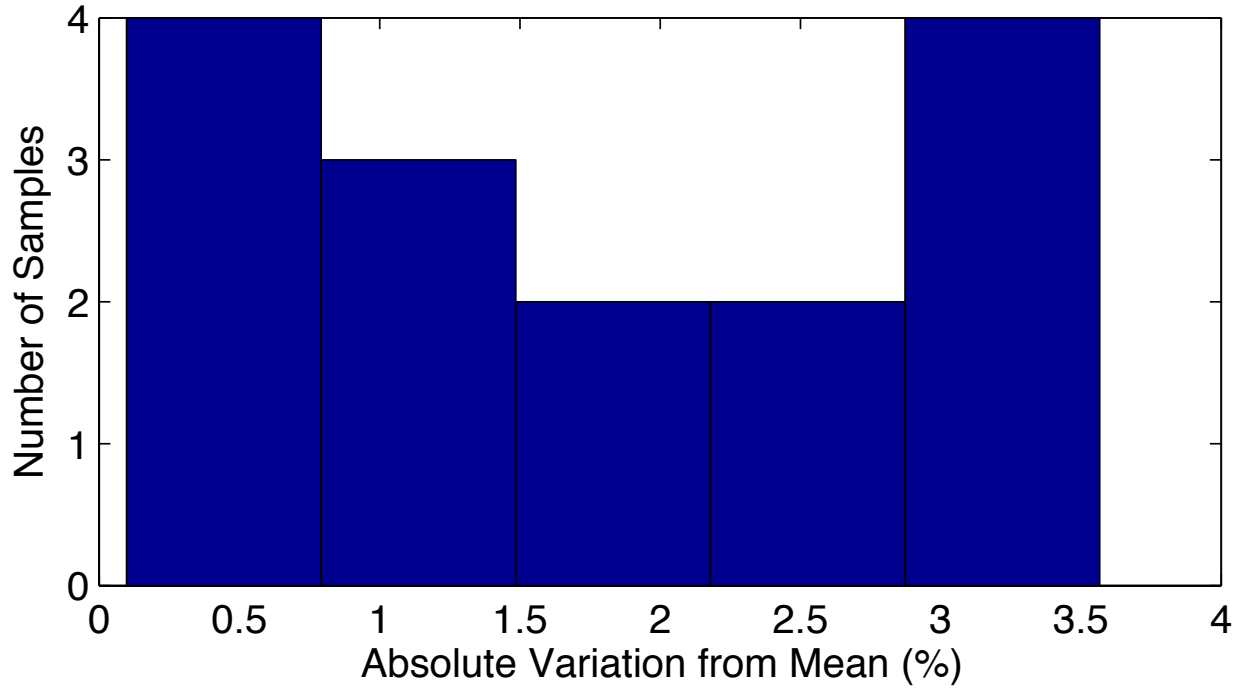
Figure 2.12: Histogram of measured absolute output voltage variation from 15 samples, with a σ of 2.2%. The lot average is 423mV.

and power consumption are comparable to MOS V_{TH}-based references, while utilizing an architecture that retains the lower process sensitivity of V_{BE} compared to V_{TH}.

2.2.4   Bias Current Generation

Bias current generation is divided into two parts to prevent circular dependencies. The first part provides bias current to the regulator, DCO, and demodulator, which require current source that is independent of the clock or regulator. A standard $\Delta V_{GS}/R$ current reference, powered from the unregulated supply, biases these circuits.

The second part, provides bias current to the data acquisition blocks such as the amplifiers and ADCs, can remain off until the system has powered on. However, supply rejection is critical to prevent modulation of the amplifier gain and ADC conversion gain. A precision current reference, shown in Fig. 2.13, forces 300mV across a resistor. The accuracy of poly resistors is dependent on the poly width, thus creating an area/variability tradeoff. Since the interrogator provides a reliable frequency reference, a SC resistor was used to break this tradeoff. The equivalent resistance of an SC resistor is $1/(fC)$, and thus a small capacitance can be utilized to generate a nA current reference instead of a large resistor. This allows substantial area savings and reduces variability in our process. The SC resistor utilizes non-overlapping clocks to minimize error and the 300mV op-amp reference voltage is generated

Figure 2.13: Switched-capacitor bias current generation schematic, utilizing two-phase non-overlapping clocks.

using a pseudo-resistor voltage divider from the regulated supply.

2.2.5 Power-On Reset

Both the regulator and the bandgap reference require a clock to function and the oscillator requires a regulated supply to provide a stable clock frequency. Thus, a power-on reset (POR) signal is needed to transfer the oscillator from an unregulated to a regulated supply, and ensure that all circuits power on successfully.

In steady state, the loop gain of the system is less than unity and, therefore, the system is stable. However, before the oscillator starts, the regulator output is stuck at an unknown, unregulated voltage. Hence, the primary goal of the POR circuit is to assert the reset signal until the clock has been established.

The POR circuit, shown in Fig. 2.14, utilizes a complementary pair of SC resistors that overpower the MOS pseudo resistors when clocked. A standard level shifter is used to convert the internal analog voltages to a digital output. When the node initially powers on, the capacitors pull the internal nodes into the reset state. This pulls up the regulator and bandgap outputs and enables the oscillator to start. The oscillator clocks the SC resistors and turns off the POR. Due to the large-valued pseudo resistors and the absence of amplifiers or other analog circuits, the POR consumes only 10nW in steady state (simulated) and occupies 225μm^2.

Figure 2.14: Power-on reset schematic. Switched-capacitor resistors pull up against pseudo-resistors, disabling the reset signal after several clock cycles.

2.3 System Results

Although most circuit blocks were verified independently, much of the challenge in this system design is to maintain consistent performance when all components are integrated together and operating over a wireless link. Therefore, to verify functional and robust system operation, several full system tests were also performed. This section discusses the testing to verify the wireless operation range (Section 2.3.1), multi-node communication (Section 2.3.2), simultaneous channel recordings (Section 2.3.3), and operation of the system *in vivo* (Section 2.3.4).

2.3.1 Wireless Operational Range

To measure the wireless transmission distance, a node was attached to a micro-manipulator oriented for perpendicular (edge-to-edge) coupling with the TX coil. A photograph of the testing setup is shown in Fig. 2.15. Using the micro-manipulator, the node was moved along the Z-axis of the TX coil while the TX power was swept to find the minimum operating value at a given distance. Ansys HFSS simulations show that the estimated path loss for our system in air matches the measured minimum transmitter power (accounting for rectification and modulation losses) and the comparison is shown in Fig. 2.15 with fitted trend lines. A transmission distance of 1mm in air is achievable with approximately 50mW of transmit power. The path loss in the brain was simulated to be approximately 6dB larger than in air, yielding an equivalent transmission distance of 0.6mm *in vivo*.
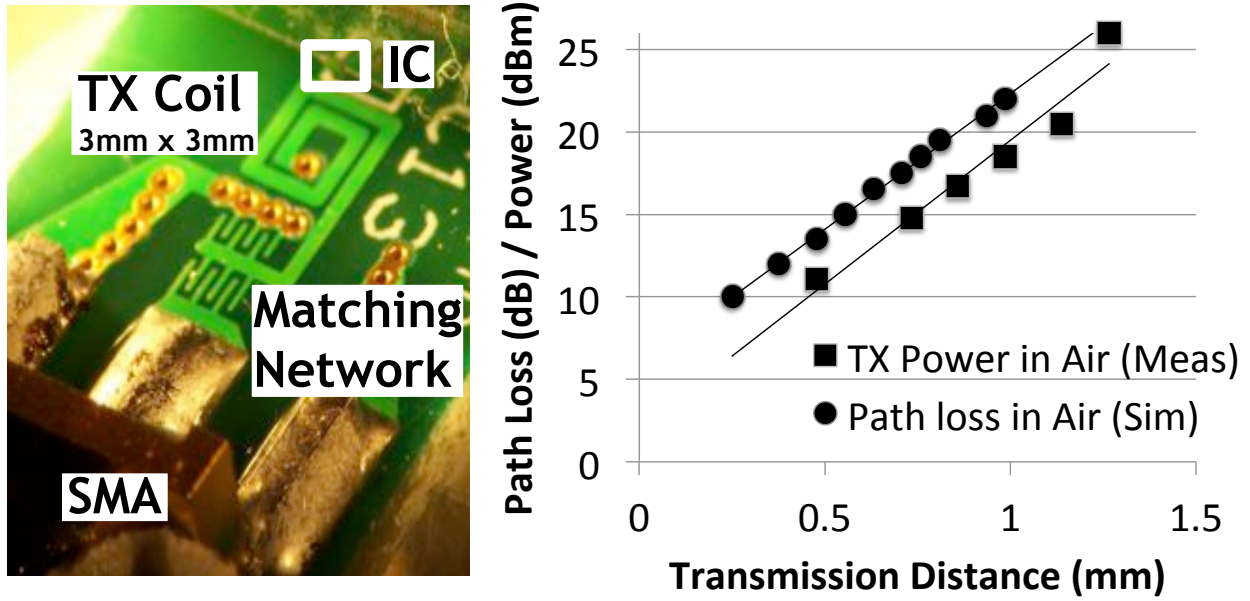
Figure 2.15: Left: Setup for wireless operational range testing. The IC is attached to a micro-manipulator using double-sided tape. Right: Simulated path loss compared to the measured minimum TX power required for operation in air.

2.3.2 Single and Multi-Node Communication Tests

To verify communication functionality, commands with a known response (e.g. changing the subcarrier frequency) were issued and the correct responses were validated. The on-chip digital communication output is connected to the modulation switch for wireless backscatter and also to a direct buffered output for wired verification. Wireless communication tests were performed using a spectrum analyzer in conjunction with COTS components. A measured wireless data packet with a 4MHz subcarrier is shown in Fig. 2.16, with 2% duty cycle interrogator beacons visible at 0µs and 50µs. This packet was measured using a spectrum analyzer and shows power (in dBm) reflected from the node during backscatter.

The use of a FDMA communication scheme allows interrogation of multiple wireless nodes simultaneously from a single antenna. Two sensor nodes were wirelessly programmed to have different subcarrier frequencies using the same antenna. A spectrum analyzer was used to observe the frequency spectrum, and the measured output is shown in Fig. 2.17. The corresponding simultaneous 4MHz and 8MHz backscatter can be filtered into independent data streams for decoding, as demonstrated by Fig. 2.3. The multi-node time domain backscatter from Fig. 2.17 is shown in Fig. 2.18 after filtering to isolate the Miller 4 node. The ideal (simulated) waveform is also shown for comparison and shows excellent consistency with measurements. Small differences in the waveforms are due to the fact that the exact interference from other nodes is a function of the random, uncorrelated data that each node is transmitting.
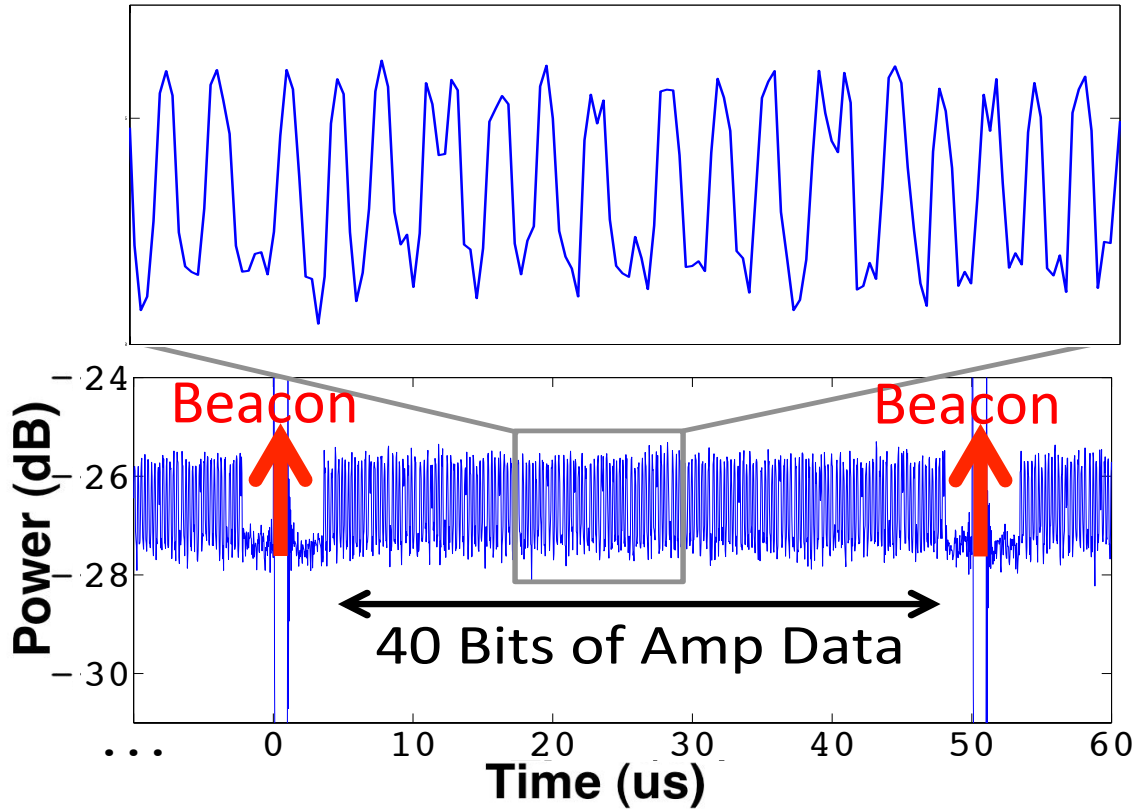
Figure 2.16: A wireless packet encoded with Miller modulation. Backscatter communication of 40 bits of signal acquisition data is initiated by a pulse from the transmitter (seen on either side of the packet).

Simulations of the bit error rate (BER) were performed in MATLAB for various numbers of nodes and the results are shown in Fig. 2.19. Initially, in all simulations, the sensitivity improves with increasing SNR. However, above 10dB SNR, the BER becomes limited by interference (as opposed to thermal noise) in environments with 4 or more nodes. With any number of nodes, a 10dB SNR provides an acceptable BER for this application.

2.3.3 Wirelessly-Powered Full System Test

Verification of the complete system functionality with simultaneous recordings from all four input channels was performed on bench-top. The system was die-attached to a PCB above a TX coil and inputs were bonded out to facilitate easier testing. A 1.6kHz, 150µV sine wave was applied to all four inputs while the system was powered wirelessly through the PCB inductive link. Fig. 2.20 shows the decoded output of all four channels recorded during a
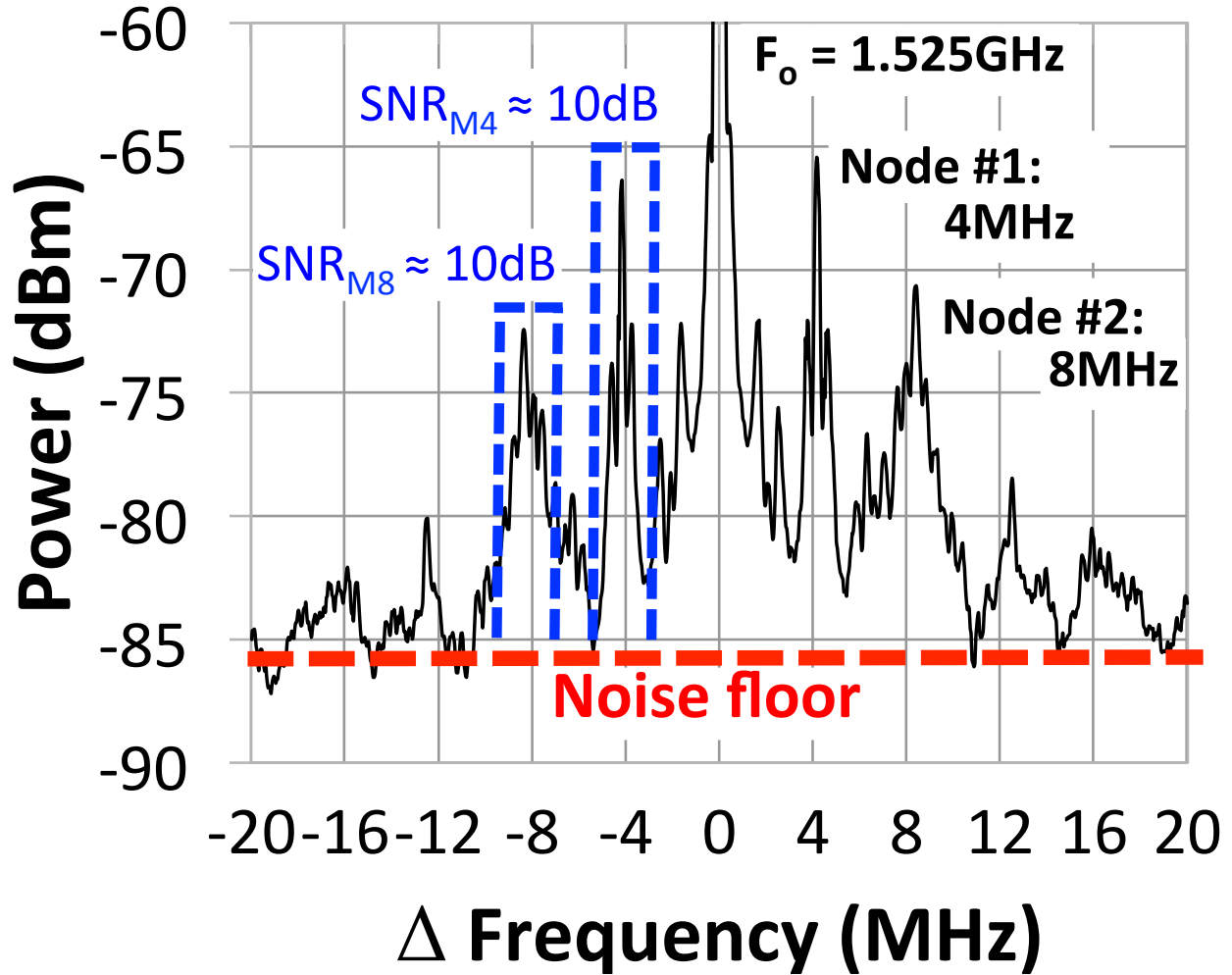
Figure 2.17: Frequency spectrum of two wirelessly powered nodes, communicating simultaneously with the same interrogator at different sub-carrier frequencies.

testing trial. The outputs show ADC and amplifier performance consistent with the results of stand-alone measurements of 6.5µVrms. The digitally-encoded modulation waveform was connected to the modulation switch and buffered directly off chip to an FPGA, which was used to gather long data streams.

2.3.4 Wirelessly-Powered *In Vivo* Recording

The system was tested *in vivo* to verify performance with a realistic signal source. To reduce testing overhead and measurement uncertainty, the system was wirelessly powered outside the animal and a single channel was connected to a pre-implanted microwire array, which could also be connected to a standard rack-mount recording system for validation of recordings. Fig. 2.21 shows a diagram of the testing setup used to obtain *in vivo* recordings.
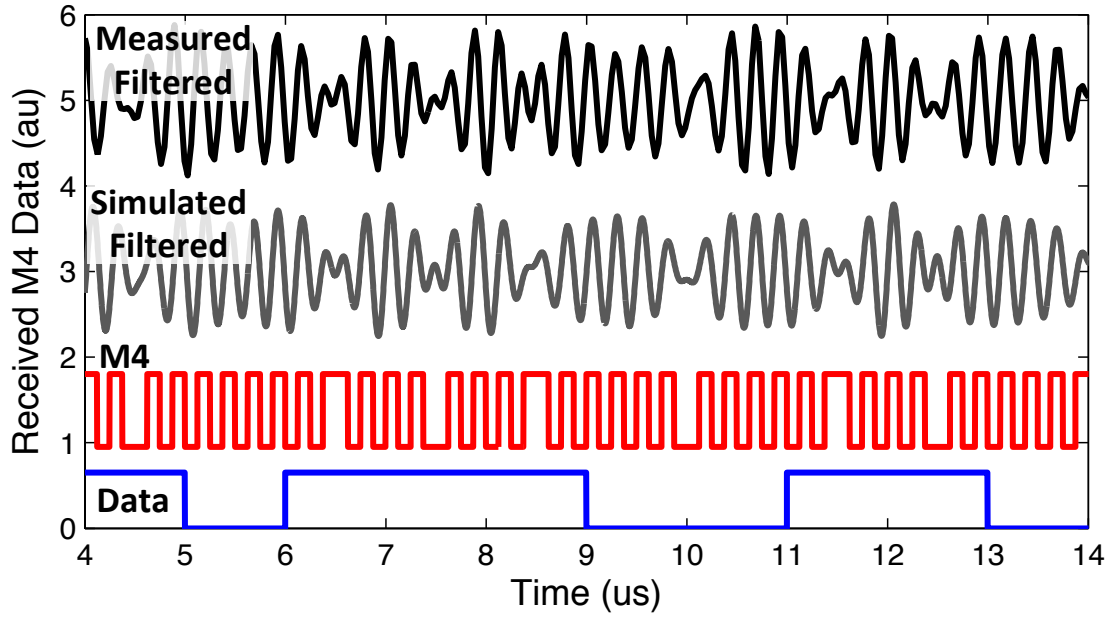
Figure 2.18: The measured time domain waveform after filtering of M4 during a multi-node interrogation test. Results are compared to an ideal filtered waveform, and the equivalent Miller waveform with decoded data is shown.
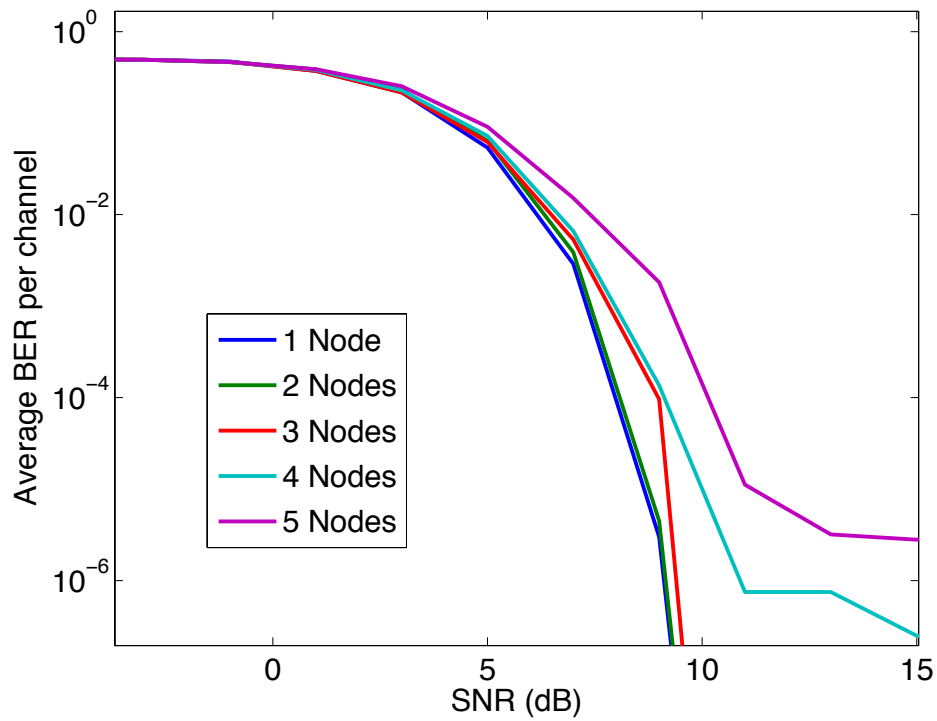
Figure 2.19: Simulated BER vs. SNR for 1-5 simultaneous nodes.
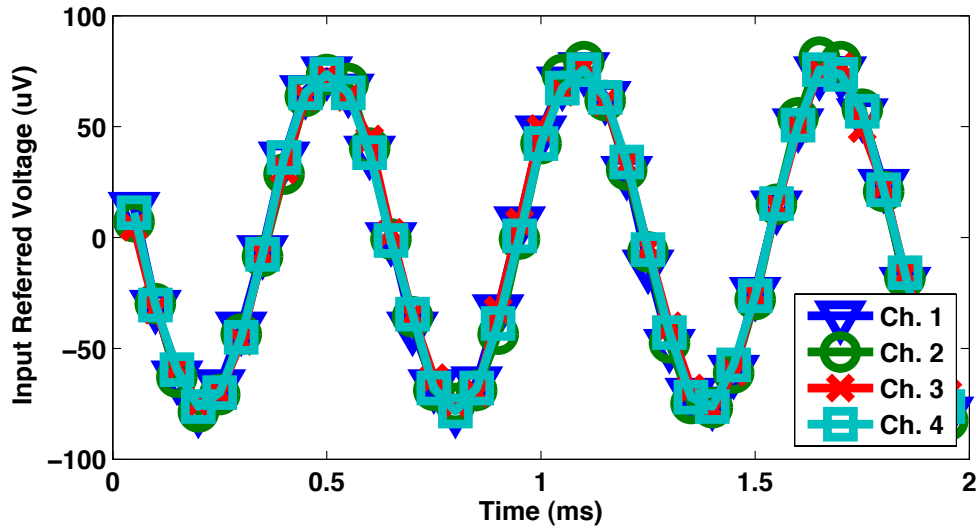
Figure 2.20: A wirelessly-powered system recording and transmitting a 1.6kHz, 150μV sine wave input from all channels simultaneously.

One adult male Long-Evans rat was chronically implanted with microwire arrays bilaterally in the primary motor cortex (M1). Arrays consisted of teflon-coated tungsten microwires (35 μm diameter, 250 μm electrode spacing, 250 μm row spacing; Innovative Neurophysiology, Inc., Durham, NC, USA). The array in the right hemisphere contained 32 recording channels (8x4 configuration), while the array in the left hemisphere contained 16 recording channels (8x2 configuration). All animal procedures were approved by the UC Berkeley Animal Care and Use Committee.

Extracellular recordings were performed for several consecutive days, more than one month after the surgery. Clearly identified waveforms with a high signal-to-noise ratio were chosen for further investigation as single unit responses. Putative single units were validated based on waveform shape, reproducibility, amplitude, and duration. We also verified that the characteristics of the inter-spike interval distributions were close to Poisson and exhibited a clear absolute refractory period.

Fig. 2.22 shows the recorded waveform from one trial capturing multiple APs. The amplifier gain was set to its maximum, and the LFP feedback cancelation high-pass corner was set to be approximately 500Hz. Recorded noise levels varied between recording sites from 15μV to 20μV. These noise measurements agree with expectations of the biological noise level as described in [23].

2.4 Conclusion

The wireless neural recording sensor was fabricated in a 65nm LP CMOS process with all electronics and wireless interface integrated into an area of 0.125mm². The top-level lay-
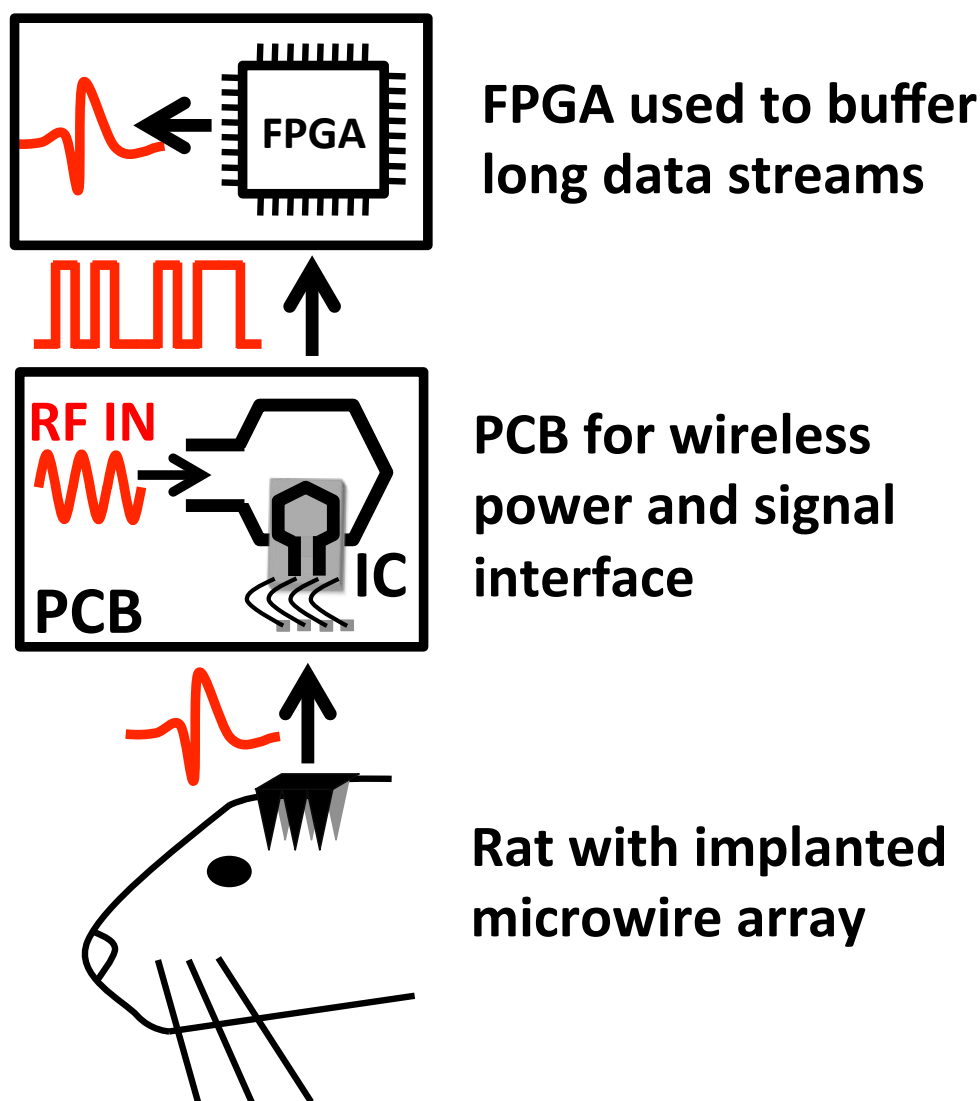
Figure 2.21: The setup for *in vivo* recording trials utilized a rat which was implanted with a microwire array. Our system was die-attached to a PCB to facilitate wireless powering and signal interfacing. In order to gather long data streams, a FPGA was used to buffer the on-chip Miller-encoded neural data.
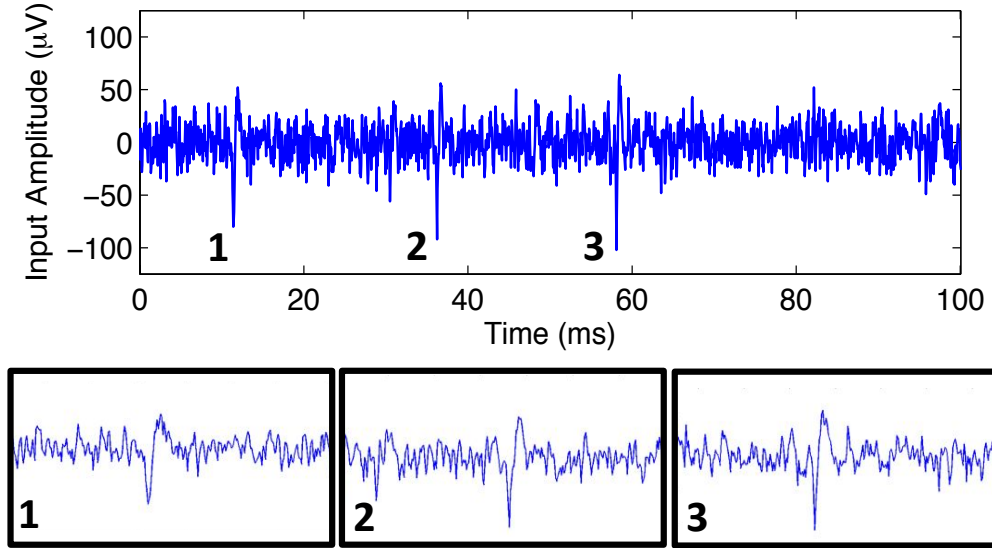
Figure 2.22: One example trial of wirelessly powered *in vivo* neural data from a live rat.
The LFP feedback cancelation high-pass corner was set and measured to be approximately
500Hz.

out is shown in Fig. 2.23a, and the die photo is shown in Fig. 2.23b, with the 4 inputs,
power/communication coil and PGS visible. The node was wirelessly powered and interro-
gated using a custom PCB antenna and COTS components on bench-top and in an *in vivo*
setting. Table 2.4 summarizes the performance of the system. The complete sensor has no
off-chip components and consumes 15μA from an unregulated voltage source of 700mV, for
a total power consumption of 10.5μW (2.6μW/channel) in under 500μm x 250μm.

This SoC reduces the average power per channel by 18x compared to [15] and 58x com-
pared to [13]. Although [13] used a larger (500nm) process, passives used to build the analog
filters consume substantial area even in modern processes. Compared to [13], this work
reduces the average area per channel by 10x, and decreases the amplifier and ADC area to
110μm x 100μm, compared to 400μm x 400μm (for an amplifier, comparator and DAC).
Table 2.5 compares this system to prior neural recording systems with wireless telemetry.

In this system my key contributions centered around the techniques and system design
elements for power delivery and management. The modeling and optimization of the power
transfer link, including the antenna design enabled the smallest fully-integrated wireless
neural sensor reported to date. In addition, I presented a novel low power and area technique
for generating on-chip reference voltages and currents. The system architecture was designed
to leverage the power train performance and relax design constraints on the neural amplifier
design, enabling further power reductions.

There were two lessons from this work which helped direct my subsequent research.
First, in order achieve aggressive area minimization, fundamental tradeoffs were made be-

Table 2.4: Summary of System Performance

Power / Signaling Frequency	1.5GHz
Uplink Comm. Frequencies (MHz)	2, 4, 6, 8, 10
Downlink / Uplink Data Rate	1Mbit (Half-Duplex)
Unregulated / Regulated Supply Voltage	700mV / 500mV
Rectifier $V_{in,min}$ (15μA @ 700mV load)	1.07V
Regulator PSRR / Dropout	27dB / 50mV
Neural Signal Amplifier Gain	46dB (1-10kHz)
Input Referred Noise	6.5μV
Single Amp Bias Current	3μA (1.5μW)
ADC Sampling Rate	20kHz
Number of Channels	4
Total Chip Area	**0.125mm²**
Total Chip Power	**10.5μW**
	(2.625μW/Ch)

Table 2.5: Comparison of neural recording systems with wireless telemertry.

Author	Off-Chip (Ant, Cap)	Wireless Power?	*In Vivo* Results	Ch. (#)	Avg. Pwr (mW/Ch)	Total Area (mm²/Ch)	Amp. Noise (μVrms)	Process (nm)
Chae [15]	Y, N	N	No	128	0.047	63.36	4.9 (Wired)	350
Lee [14]	Y, Y	Y	No	32	0.183	16.2	4.95 (Wireless)	500
Sodagar [18]	Y, Y	Y	Yes	64	0.22	217 [a]	8.0 (Wireless)	500
Harrison [13]	Y, Y	Y [b]	Yes	88 [c]	0.153	27.3	5.1 (Wired [b])	500
This Work	*N, N*	*Y*	*Yes*	*4*	*0.0026*	*0.125*	*6.5 (Wireless)*	*65*

[a] Incl. off-chip   [b] *In vivo* tests and noise measurements used wired power   [c] Not incl. REF channels

tween transistor sizes and mismatch in many circuits. Although the performance degradation of individual circuit blocks was acceptable, the cascading interaction between many blocks resulted in low overall system yield. Second, the goal of creating an ultra-small implantable sensor node was successful, however, the technological methods to perform *in vivo* implantation of arrays of sensor did not exist. This limitation prevented implantation and measurement of a sensor array during the time frame of the project. These lessons and the desire to integrate stimulation with neural recording to create a fully-integrated closed-loop implantable BMI system motivated the work in Chapter 3.
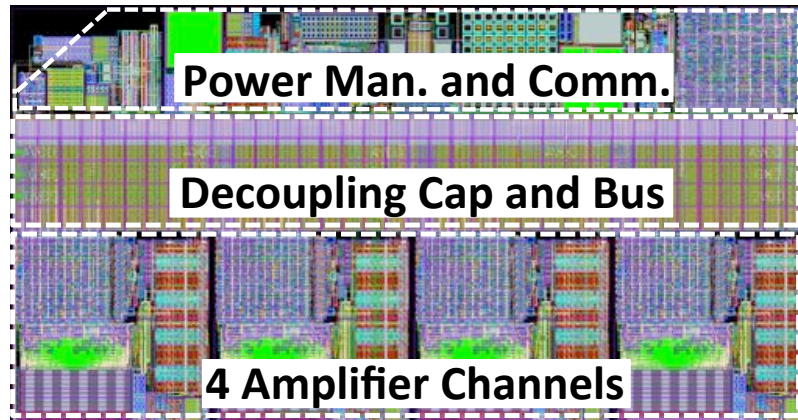
(a) Chip layout.
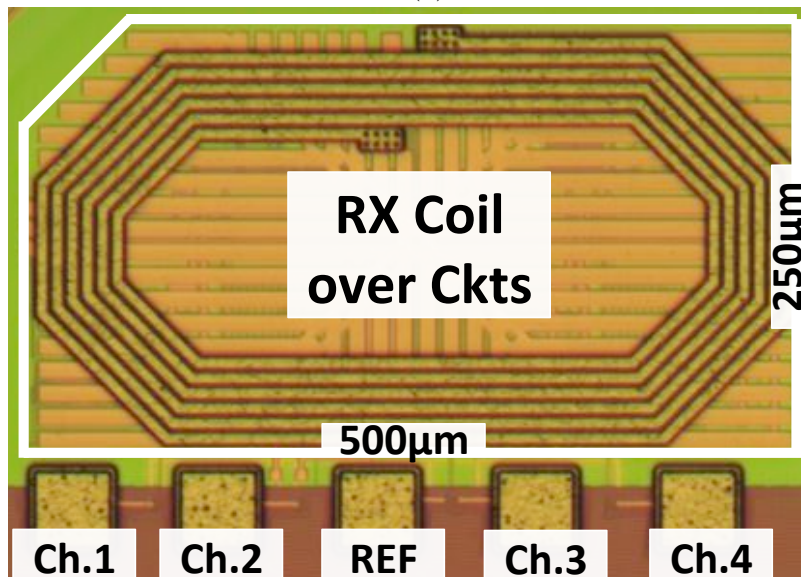
(b) Die photo.

Figure 2.23: Die photo of the full system, showing the input bonding pads, the RX coil and PGS. The active circuit area is underneath the PGS, and is depicted by an image of the chip layout.

Chapter 3

Design of a 4.78mm^2 Neuromodulation SoC Combining 64 Acquisition Channels with Digital Compression and Simultaneous Dual Stimulation

To fully restore limb mobility, a neural interface must achieve long term AP recordings from a large population of neurons (thousands) in multiple brain regions [11]; however, the standard rack mount electronics and large cables typically used for experimentation prohibit this scaling. Furthermore, a wired interface through the skull introduces a persistent infection risk for patients, and space constraints prohibit significant energy storage beneath the skull. Consequently, next generation neural interfaces must be powered and communicate wirelessly, with the ability to scale to thousands of channels. In order to realize this level of scaling with implantation techniques available today, a new system must be designed compared to Chapter 2. Furthermore, to create a truly closed-loop BMI system, stimulation must be integrated into the implanted electronics.

This chapter presents the design of an SoC solution capable of closed loop BMI, which achieves significant improvements in area, power and signal compression over current state of the art (e.g. [20, 22, 59]) and can be arrayed across the brain to achieve thousands of recording sites. Section 3.1 describes the system architecture and gives an overview of the neural amplification circuitry and digital compression. Section 3.2 describes the implementation of the power management circuitry, Section 3.3 details the stimulator architecture, design and measurements, and Section 3.4 presents the *in vivo* testing results of the fully integrated system. Finally, Section 3.5 compares these results to prior work.
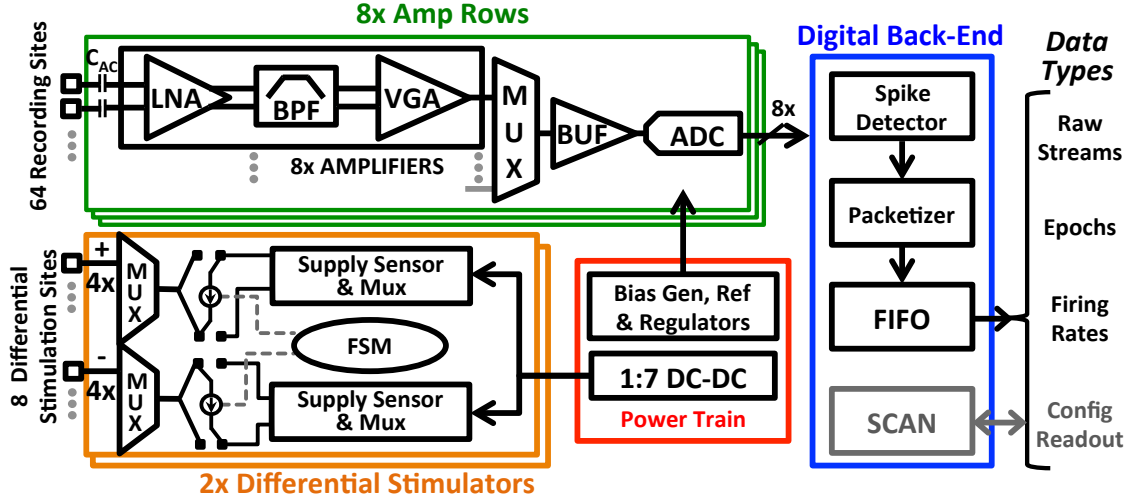
Figure 3.1: The system architecture, subdivided into the primary circuit functions: Amplification, Stimuation, Digital Compression and Power Train.

3.1   System

The IC architecture, Fig. 3.1, combines 64 channels of real-time neural recording with on-chip compression and dual stimulation on 8 selectable channels without any off-chip components, paving the way for closed-loop neuromodulation. The fully-integrated SoC achieves the highest complexity and lowest power and area per recording and stimulation site reported to date. When arrayed across the brain, 16 ICs provide 1024 recording and 128 stimulation sites. This would require 6.67mW and 320kbps (20kbps/IC for firing rates x 16 ICs), which can be delivered through the skull as shown by [60, 61].

The recording channels (Fig. 3.2) are designed to consume minimal power and area, without sacrificing noise performance. The channel gain is set through closed loop feedback to provide accurate, calibration-free operation. The gain, bandwidth and bias current (noise performance) are individually adjustable on a per-channel basis, enabling power savings on high SNR electrodes. A time-multiplexed switched-capacitor ADC driver utilizes separate sampling capacitors for each of its 8 input channels, thereby minimizing settling speed requirements of the VGAs. Finally, the inputs are AC-coupled (using 10pF capacitors), providing compatibility with large stimulation common-mode voltages.

Fig. 3.3 shows the block diagram of the implemented digital back-end, including all signal interfaces and I/Os. A spike detection algorithm based on the nonlinear energy operator (NEO [62]) extracts spike events, enabling data reduction by only sending a 2.1ms time window of data around an event (epochs), and/or spike counts in a 2.4-50ms programmable window. Sending epochs, spike rates, and uncompressed data (streams) can be enabled on a per channel basis via scan instructions. Finally, all packets are put into a clock domain crossing FIFO, which allows the system clock to operate at a different frequency than the
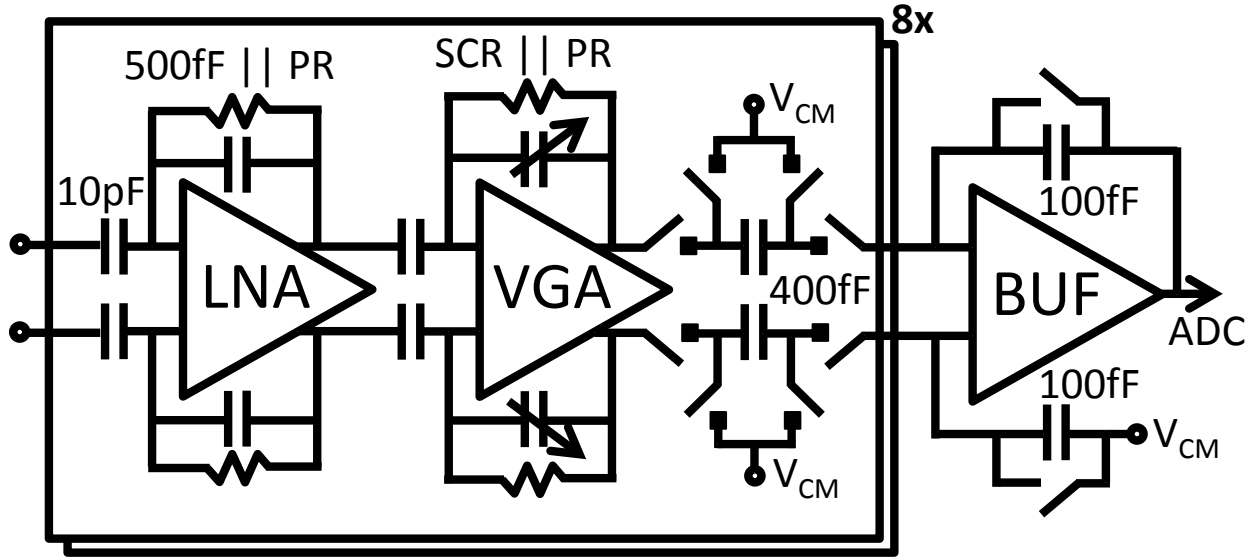
Figure 3.2: Amplifier block diagram showing the LNA, VGA and ADC buffer.

Data Types	64 Ch Power 64 Ch Data Rate Compression Ratio
Raw Stream	238µW 13.653Mbps 1x
Epochs 50Hz/Ch	113.6µW 1.6384Mbps 8.3x
Firing Rates	77.63µW 20kpbs 700x
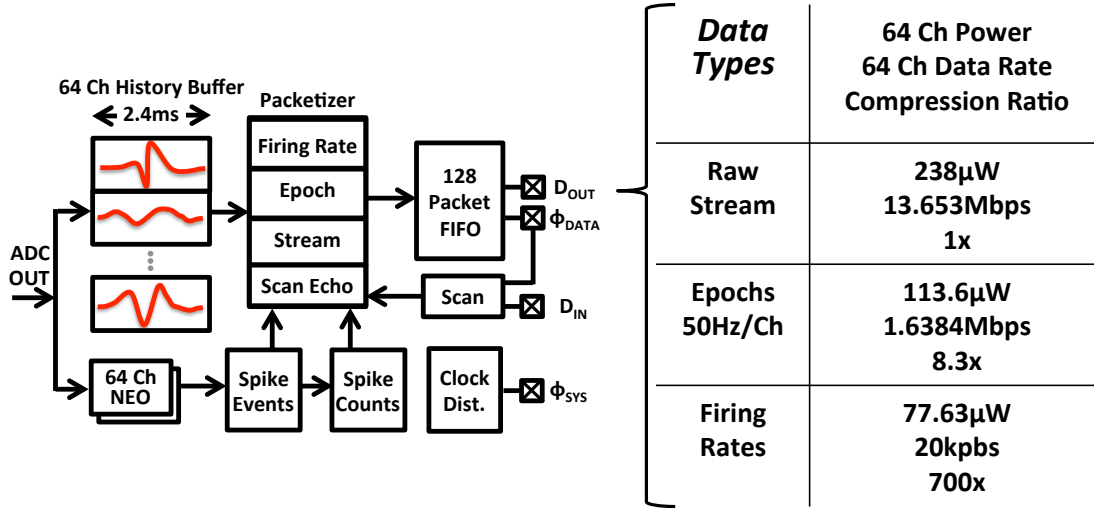
Figure 3.3: Compression block diagram and the power consumption, data rates, and compression ratio for different output modes.

output data rate, resulting in further power savings and system optimization. With an average firing rate of 50Hz per channel the total digital power is 77.63µW for firing rates and 113.6µW for epochs. Section 3.4 presents *In vivo* data that illustrates the different data representations and operation of the compression algorithm.
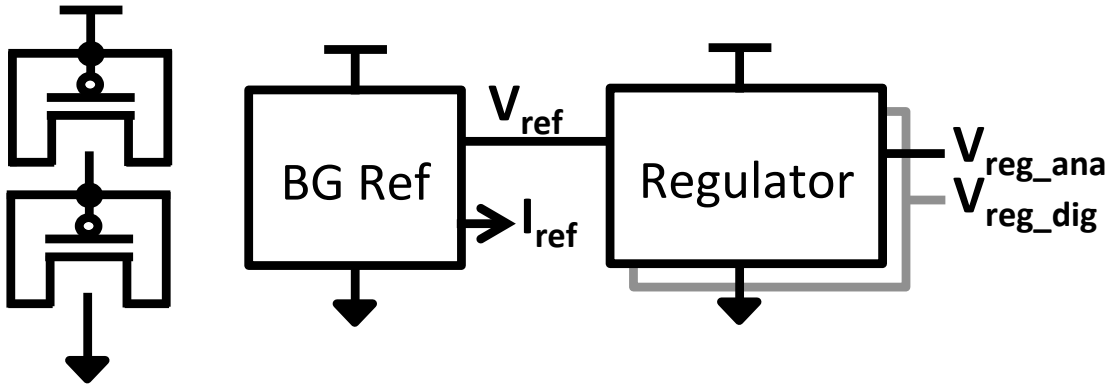
Figure 3.4: A simplified block diagram depicting the components of the power management sub-system: a supply clamp, band gap reference and two regulators.

3.2 Power Management

The architecture of the power train is shown in Figure 3.4 and consists of a voltage limiting supply clamp, a band gap reference and two voltage regulators. While area and power minimization are still critical for the power train, the bulk of the system power and area consumption is dominated by the arrayed amplifier. In this system, there are 64 recording channels, and 8 stimulation channels compared to only 4 recording channels for the system in Chapter 2. The 64 amplifiers result in over 1350μm x 1250μm of area and the digital logic occupies 1700μm x 500μm. To simplify startup and system complexities resulting from the use of a discrete time power train, continuous time topologies for the band gap reference (3.2.1) and regulators (3.2.2) were chosen rather than the topologies presented in 2.2.3 and 2.2.2. Although this results in an absolute area and power increase of these blocks, in this system it is irrelevant (ultimate the power train only occupies approximately 0.5% of the total system area and less than 3.4% of total power) compared to other system components and results in a worthwhile tradeoff.

The SoC is designed to operate from wireless power for a fully implanted system or directly from a battery for a head mounted system. Due to their small size and high energy density, Zinc air batteries would be the preferred power source to meet the size, wight and operating life specifications for a neural head stage. A typical output voltage for a zinc air battery can be between 1.4V and 1.15V when loaded, therefore, the power train is designed to provide a stable supply with as small as a 1.1V unregulated voltage. When wirelessly powered from a rectifier, the unregulated voltage can vary wildly depending on the input power to the rectifier. A passive diode clamp provides over voltage protection and is designed to have minimal leakage in FF, hot corners during normal operating conditions while providing sufficient clamping above 2.5V (the thick-oxide breakdown).
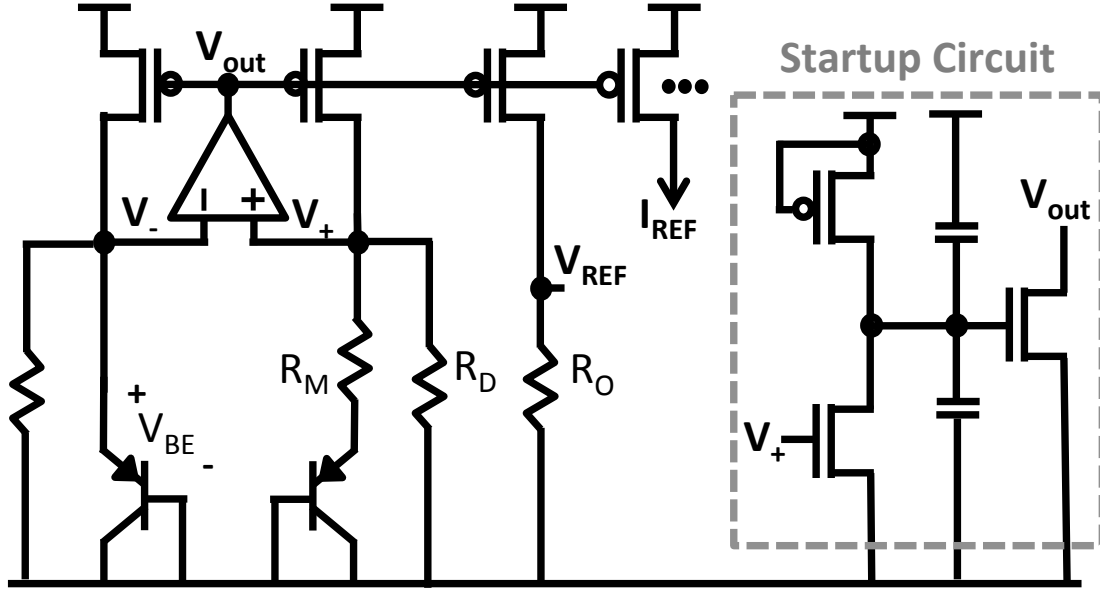
Figure 3.5: Schematic of the resistive sub-divisionn band gap reference and startup circuit.

3.2.1 Bandgap Reference

Although temperature independence is not critical for an implanted system (since the temperature of the brain is well regulated to 37°C) the SoC is designed to be compatable with a wireless head stage. Furthermore, the bandgap voltage reference is largely process independent as opposed to other voltage reference techniques and this topology provides drastic power savings over other designs. The bandgap reference implements a resistive subdivision technique proposed by [47] to achieve an output voltage at a fraction of the band gap of Si. Conceptually, a fractional bandgap reference can be generated by introducing a division term D into the traditional bandgap output equations as shown in Equations 2.7 and 2.8. The architecture, shown in Fig. 3.5, uses an opamp to force V_+ and V_- to be equal. Resistors R_M and R_D produce currents $I_{\Delta V_{BE}}$ and $I_{V_{BE}}$ respectively which are summed at the V_+ node. The resulting current, I_{REF}, is temperature independent and can be mirrored and converted to a fractional band gap voltage via R_O, resulting in Equation 3.1. In addition, I_{REF} is used as a bias current for other blocks. A startup circuit employs a DC leakage path and AC coupling capacitors to pull the op-amp output low at startup.

$$V_{OUT} = R_O \cdot (\frac{V_{BE}}{R_D} + \frac{\Delta V_{BE}}{R_M}) \tag{3.1}$$

The total active area is 75μm x 100μm and the circuit has a measured current consumption of approximately 750nA. The measured line regulation and temperature independence are shown in Figures 3.6 and 3.7 respectively. This design is able to achieve a stable output voltage with an unregulated supply of less than 1V and achieve a temperature coefficient of approximately 180pp/°C.
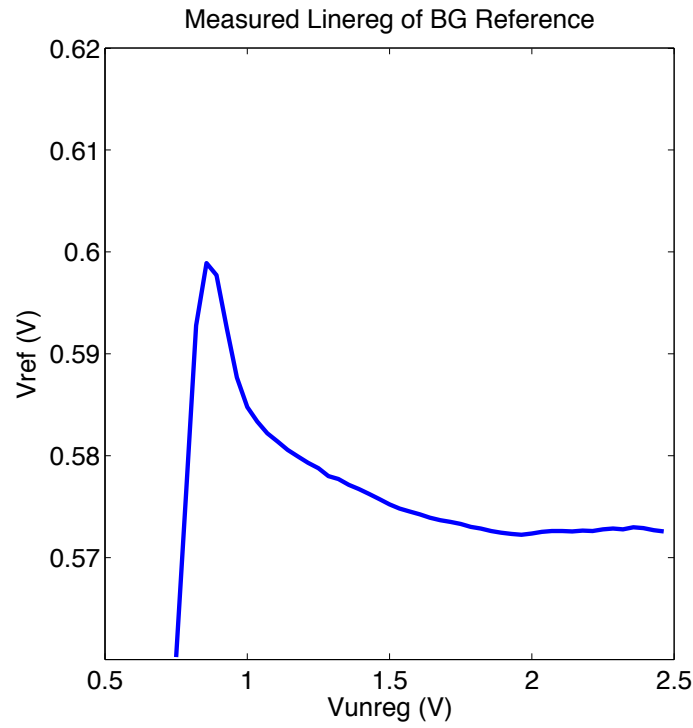
Figure 3.6: The measured line regulation of the band gap reference.
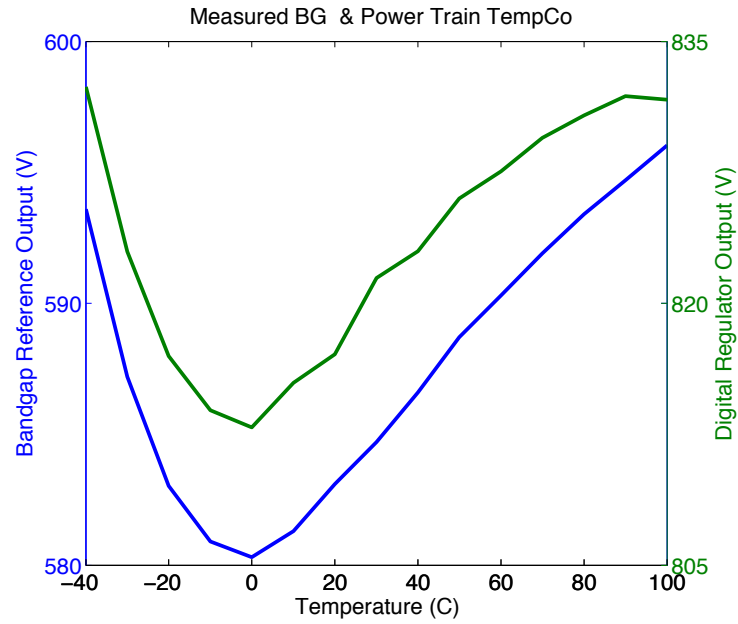
Figure 3.7: The measured temperature sensitivity of the band gap reference and digital supply voltage. The temperature coefficient is approximately 180ppm/C.
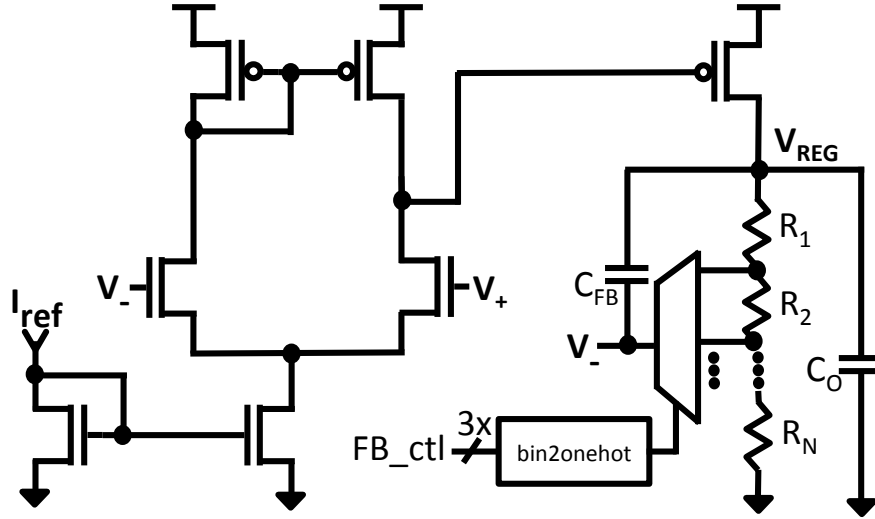
Figure 3.8: The schematic of the regulator with a 3b tunable output voltage and output compensation.

3.2.2 Regulators

To reduce the effect of digital supply noise on the sensitive analog neural amplifiers, two separate regulators are implemented on-chip. The regulator design, shown in Figure 3.8, is output compensated to allow operation of additional off-chip peripherals such as a data aggregation/wireless comm IC. An additional feedback capacitor, C_{FB}, creates a zero to compensate for the pole created from C_{gs} of the input device and the feedback resistors. The output compensation cap for each regulator is distributed beneath the I/O bond pads allowing utilization of otherwise wasted area. The total available area under the bond pads allows for up to 3.3nF of capacitance from a stack of thick oxide MOS, plus 3 layers of MOM cap.

The output voltage of each regulator has 3 bits of programmability, with the default voltage set to maximum at startup. A binary-to-one-hot converter selects from a resistor string to provide an output voltage between approximately 600mV and 1.2V. The regulator bias current is tunable from 500nA to 5uA to maintain stability over various load conditions while minimizing power overhead. The measured line and load regulation is shown in Figure 3.9 and the measured PSRR of the power train is shown in Figure 3.10. The total active area of the regulators is 50um x 150um, not including the capacitors under bond pads. The total power train current consumption was measured to be 14μA while supporting the maximum system load current (64 amplifier channels enabled and while streaming raw data from the digital).
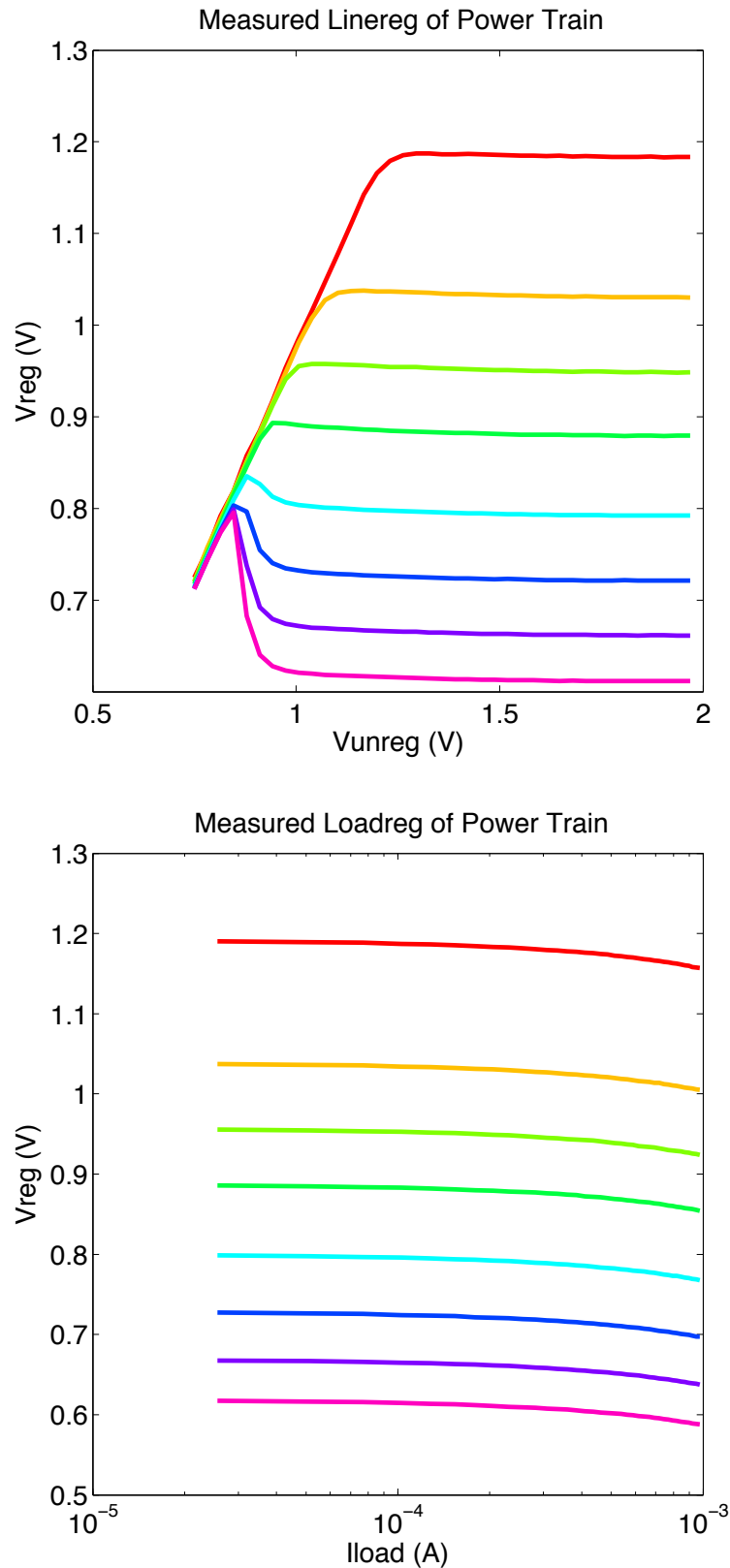
Figure 3.9: The measured line (top) and load (bottom) regulation for all 3 bits of output tuning.
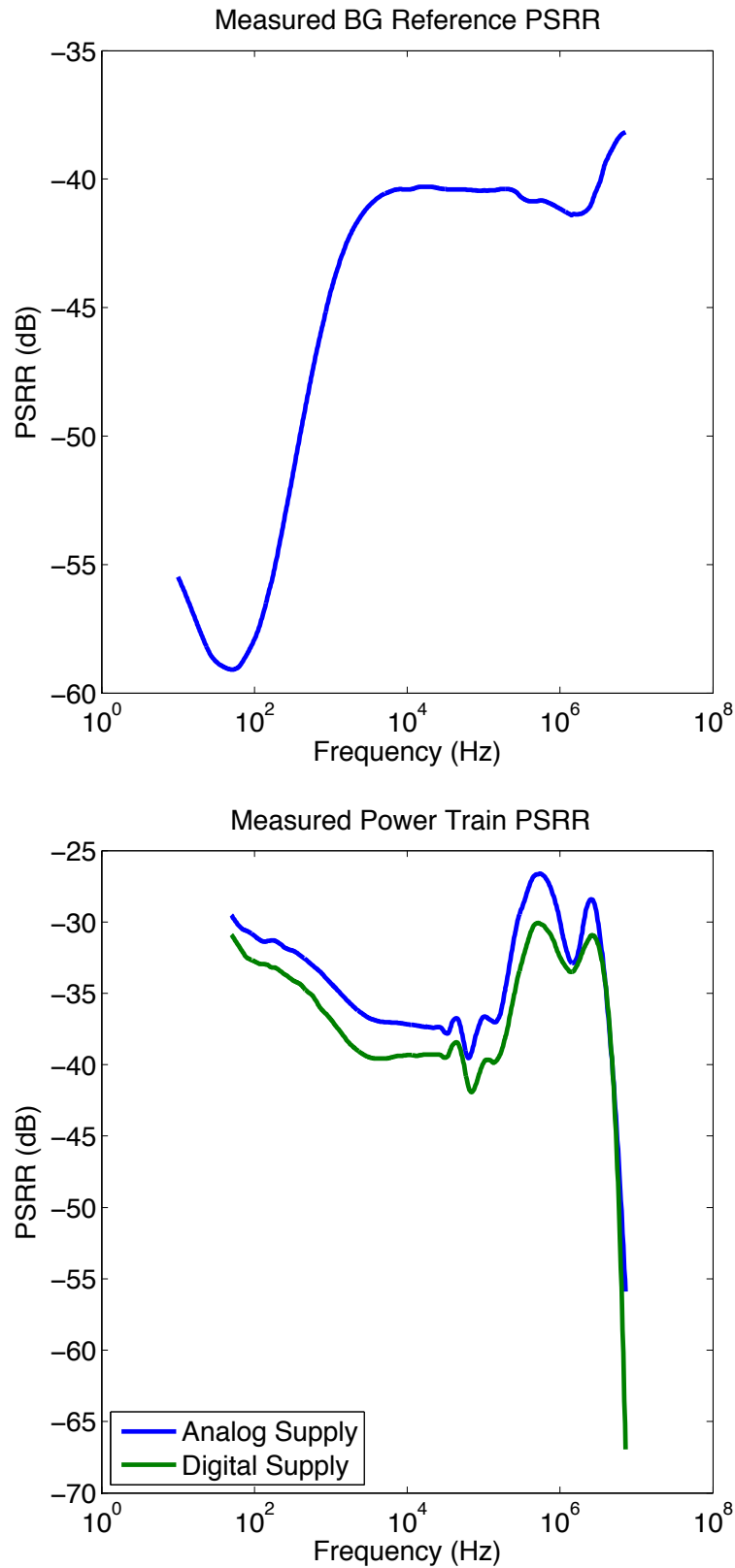
Figure 3.10: The measured PSRR of the band gap (top) and entire power train (bottom).
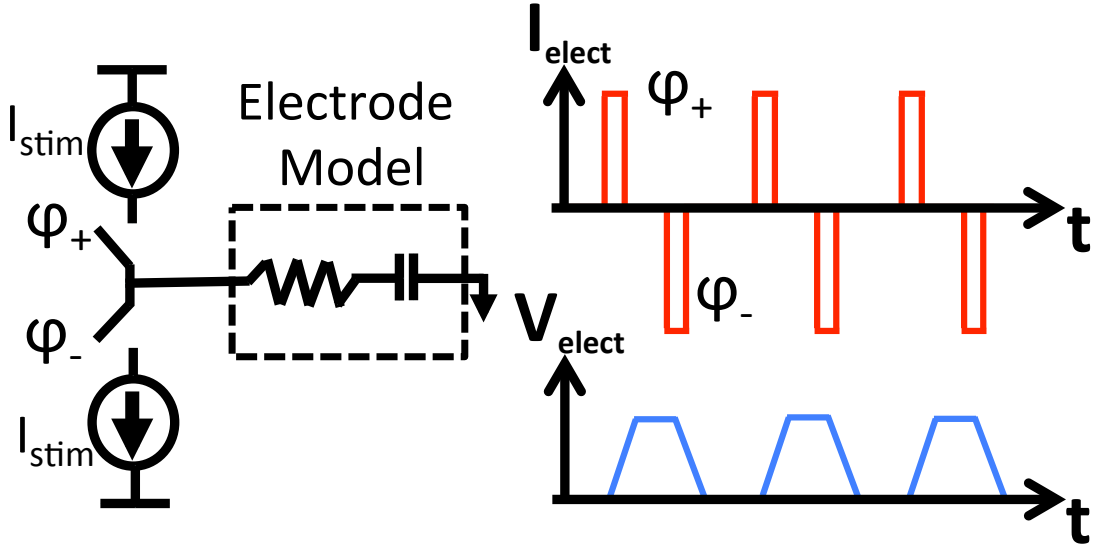
Figure 3.11: A traditional charge balanced, bi-phasic current stimulation architecture and conceptual stimulation waveforms.

3.3 Neural Stimulation

In a BMI system, neural stimulation is used to provide feedback to the patient. Charge is deposited onto an electrode until the resulting electric field becomes strong enough to trigger a response from neighboring neurons. Afterward, the charge must be removed from the electrode to prevent build up; failure to remove all deposited charge can result in permanent tissue damage. Traditionally, stimulation is performed using two independent current sources, as shown in Figure 3.11. There are three problems faced by this topology: first, layout induced mismatch between the two current sources necessitates calibration prior to use. Second, this topology consumes excess power because the supplies are statically set to the compliance voltage of the stimulator, which is the maximum electrode voltage reached during stimulation. This results in the stimulation current, I_{elect}, constantly being consumed from the compliance voltage, independent of the intermediary or final values of V_{elect}. In addition, when discharging the electrode, the current is commonly discharge into ground, resulting in wasted power during this half of the stimulation cycle. Third, stimulation patterns are typically spatio-temporal, requiring large IC area consumption to support one stimulator per electrode. In this section, I will introduce a new stimulation architecture that addresses these issues.

3.3.1 Stimulator Operation

The proposed topology is implemented with differential electrodes to create localized electric fields, minimizing recording artifacts and amplifier recovery on adjacent electrodes after
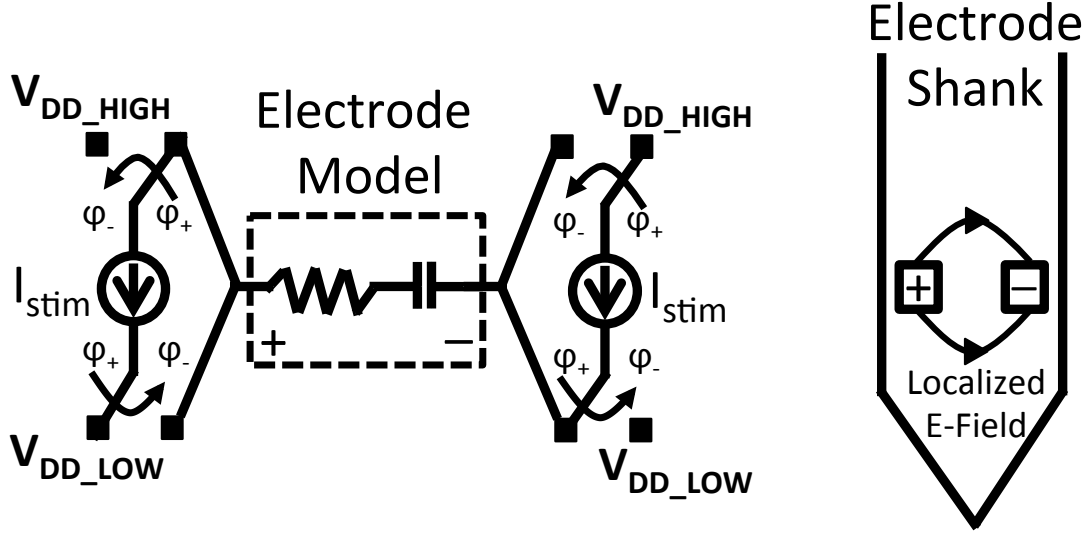
Figure 3.12: The proposed differential stimulation topology which utilizes a single current source for each electrode. The current source switches between a dynamically adjustable high and low supplies to minimize waster power and maximize recovered power respectively. The electrodes are placed adjacently on an electrode shank, localizing the electric field and reducing stimulation artifacts on recording channels.

stimulation, as shown in Figure 3.12. Each stimulator utilizes a single current source for both positive and negative stimulation phases, which reduces the effect of current mismatch between phases and eliminates the need for calibration. The system instantiates two differential, bi-phasic current stimulators which can be independently multiplexed onto four electrode pairs for a total of 8 unique stimulation sites (Figure 3.13). Time multiplexing a single stimulator onto multiple electrodes minimizes the IC area overhead while maintaining a high number of stimulation sites.

The current source is implemented using a thick oxide PMOS 6b binary weighted current DAC with a 7μA LSB, allowing a maximum differential stimulation current of 900μA. PMOS switches are used to mux the current source terminals, and their control signal voltage levels are set to $V_{DD, HIGH}$ and $V_{DD, LOW}$ by using AC coupled level shifters. At the end of each stimulation cycle, the electrodes are shorted together to remove any small amount of residual charge. A 1.25MHz clock and 9b counter are used to configure pulse length, inter-phasic delay, and the stimulation period.

To minimize power consumption, the stimulator implements an adiabatic, charge-recycling architecture without utilizing off-chip components, enabling a fully integrated system as opposed to prior state-of-the-art [63]. During the charging phase of an electrode, the $V_{DD, HIGH}$ supply is dynamically increased to minimize wasted power while keeping the current source in the saturation region. During the discharge phase of an electrode, current is discharged into a bi-directional DC-DC converter. In addition, the $V_{DD, LOW}$ supply is dynamically

lowed to recycle the maximum charge.

To facilitate the dynamic supply switching, a supply voltage detection circuit (Figure 3.13) uses a thin oxide comparator and a switched capacitor network to sample the electrode voltage and a reference voltage generated from the current selected supply voltage. A state machine uses the output of the voltage detectors to select one of 7 supply voltages from the DC-DC to minimize the power consumed (and maximize the power recovered) throughout the stimulation cycle.

Finally, the stimulator architecture was designed to support both electrical and optical (i.e. optogenetics) stimulation. By utilizing a differential output pair, the stimulator can drive current to illuminate an LED and control neurons that have been genetically sensitized to light. LED on/off time, repetition rate and drive current can be set and controlled with the same precision as bi-phasic current stimulation mode. Functionality was demonstrated with bench top measurements using off-the-shelf LEDs as shown in the photograph in Figure 3.14.

The layout of the stimulator is shown in 3.15 and occupies approximately 150um x 450um of area. A common centroid layout was used for the 6b current DAC to minimize mismatch induced non-linearity. Each device in the DAC has a W/L of 4um/2um and the DAC consumes 45um x 137um of total area. Due to the capacitors required for the AC coupled level shifters and the relatively large number of control signals required for the muxes in the stimulator, the AC level shifters and current DAC consume the majority of the layout area.

3.3.2 DC-DC Design

The stimulator output common mode is set at mid-rail from a single fully-integrated switched-capacitor DC-DC converter. The DC-DC is implemented using a Dickson ladder topology (Figure 3.16) with a tunable input voltage of up to approximately 1.3V. A maximum conversion ratio of 1:7 was implemented and was limited by the NWELL/PSUB breakdown voltage of the process. Level shifted control signals for the switches allow bi-directional charge flow and current recycling. The switches are implemented using floating well PMOS devices, and the control signals are bootstrapped off of subsequent stages using AC coupled level shifters (Figure 3.16). Additional output switches are added to each stage to store the voltage on 100pF on-chip capacitors. The switching frequency of the DC-DC can be set between 160kHz to 20MHz to maximize overall efficiency depending on the chosen stimulation current.

3.3.3 Performance

The INL and DNL of the current DAC was measured to be 0.082 and 0.039 LSB respectively with a unit device size of 4μm/2μm and is shown in Figure 3.18. The difference between sourcing and sinking the maximum DAC output current (448μA) was measured to be less than 2μA; this results in less than 200pC (20mV on a 10nF electrode) of residual charge remaining on the electrodes after a typical 100μs stimulation cycle. After the cycle, the two electrodes are shorted together, dissipating the small residual charge and returning the
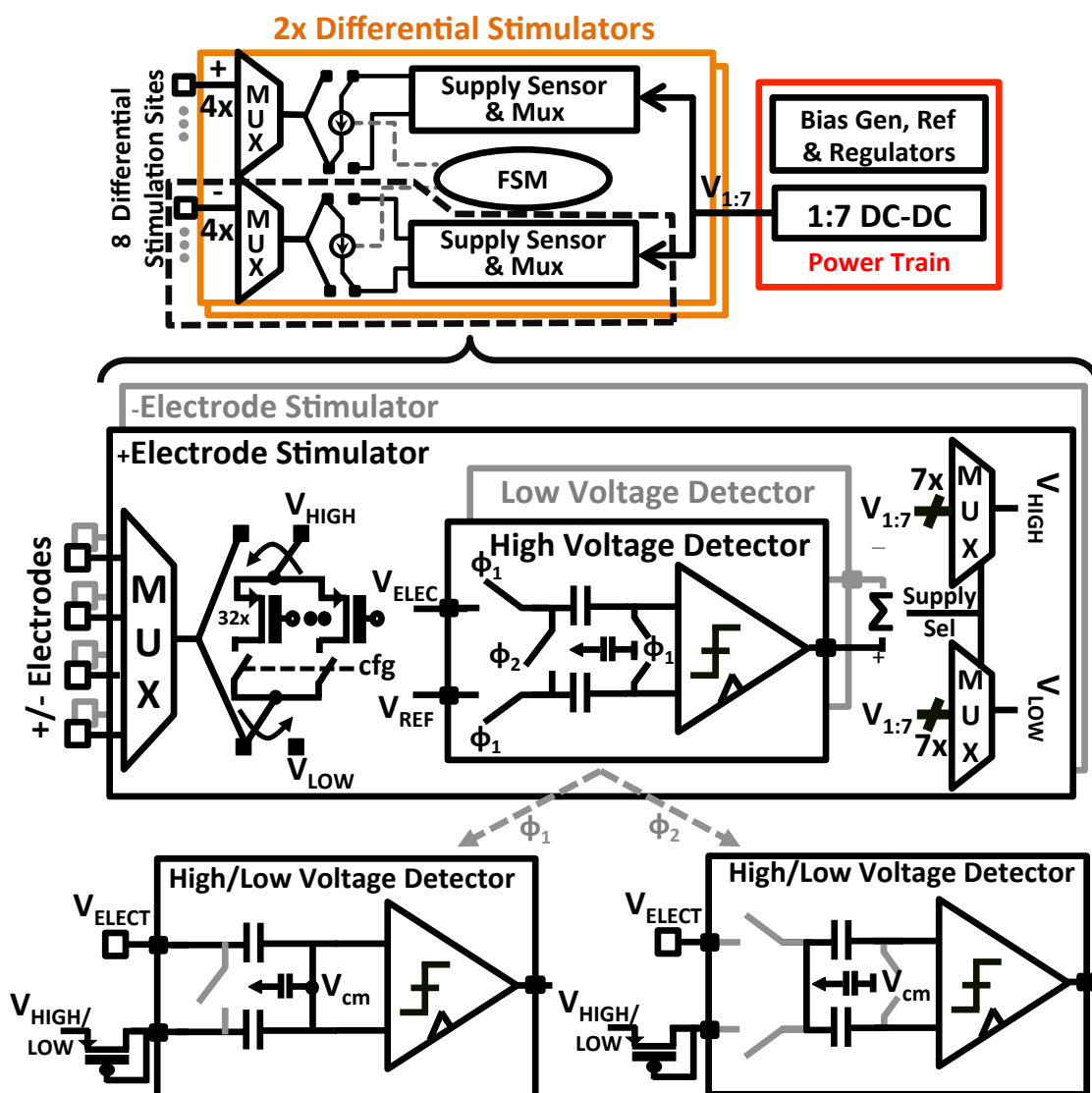
Figure 3.13: The stimulator block diagram, illustrating the 6b current source, supply sensor and supply mux operation.
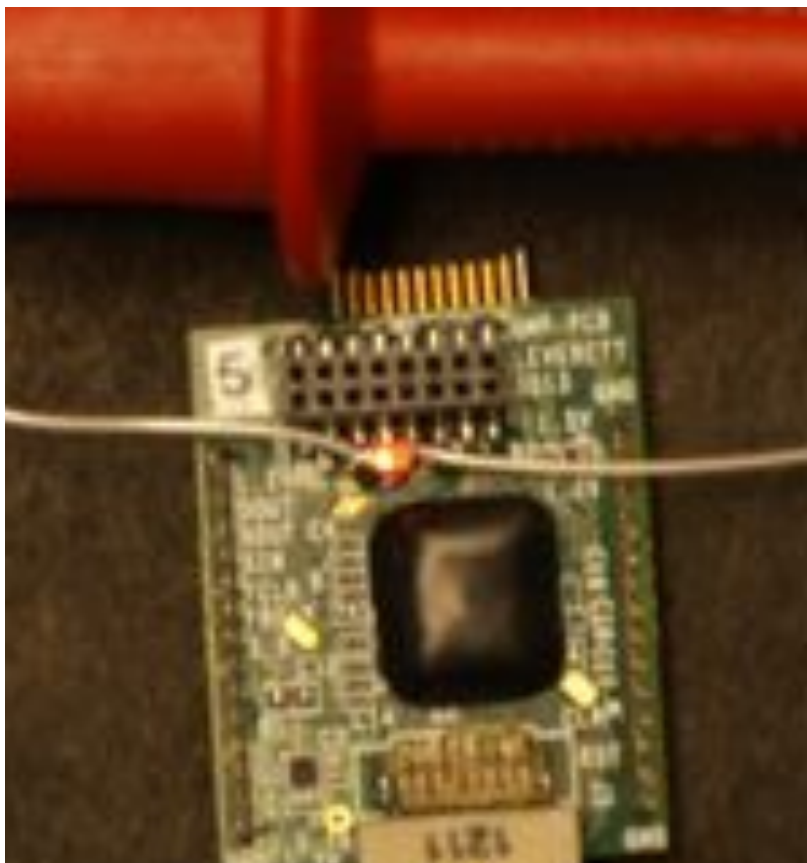
Figure 3.14: A photograph of an LED being illuminated by the output of the stimulator from the wired head stage.
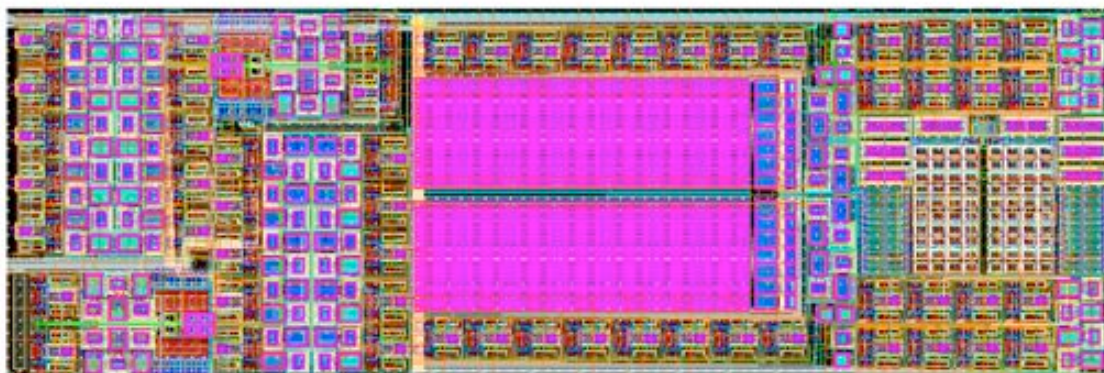
Figure 3.15: The layout of the differential stimulator consumes approximately 150um x 450um of area.
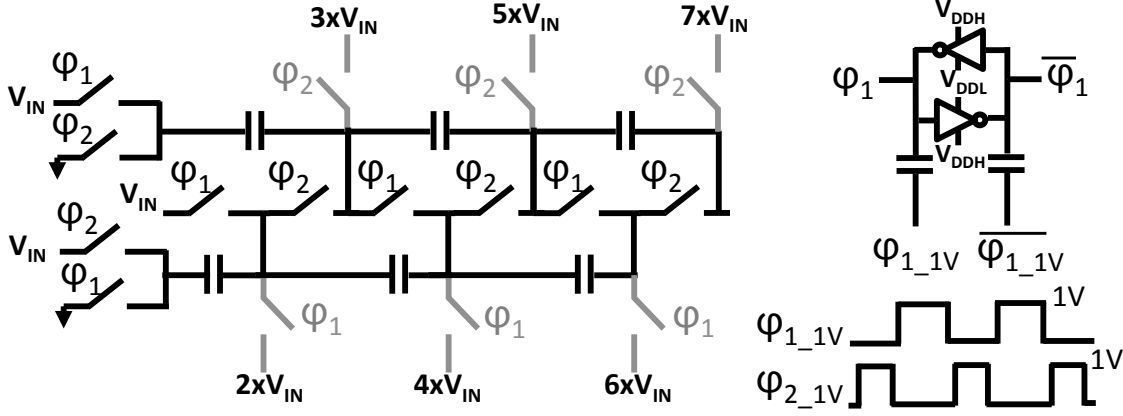
Figure 3.16: Schematic of the actively switched, 1:7 switched capacitor DC-DC and schematic of the AC coupled level shifters.

electrodes to the common mode voltage. A maximum output voltage of 8.7V was measured from the DC-DC with an input voltage of 1.3V. Figure 3.17 shows the measured efficiency of the DC-DC across different output load currents with varying input frequencies, with a maximum efficiency of 68%.

A typical stimulation pattern of 300µA differential current with 150µs per phase was performed on bench-top with a 1kOhm/10nF RC electrode model; the electrode and supply voltages are shown in Figure 3.19. As the positive electrode voltage increases during the positive stimulation phase, V_{HIGH} and V_{LOW} for the positive electrode track the voltage, minimizing the current drawn from V_{unreg} through the DC-DC. Similarly, the supplies track the electrode voltage during the negative phase, maximizing the current recycled through the DC-DC from the positive electrode, which can be used concurrently to drive the negative electrode back to the common mode voltage. The resulting measured input referred current supplied by the DC-DC over one cycle of operation is also shown in Figure 3.20. In addition, Figure 3.21 shows a series of stimulation current pulses and voltages measured from an electrode pair with a 4ms period.
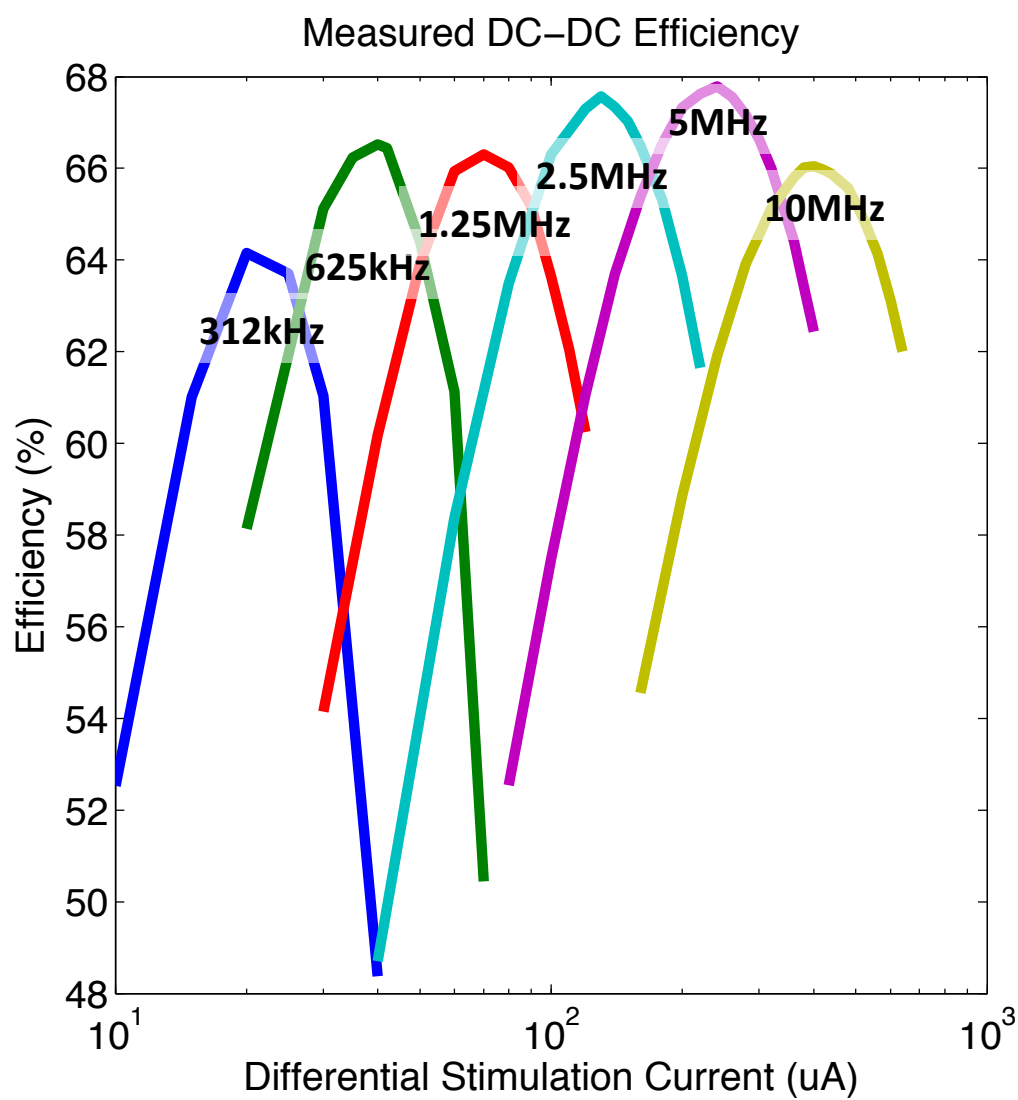
Figure 3.17: The efficiency of the DC-DC measured across load current and switching frequencies.
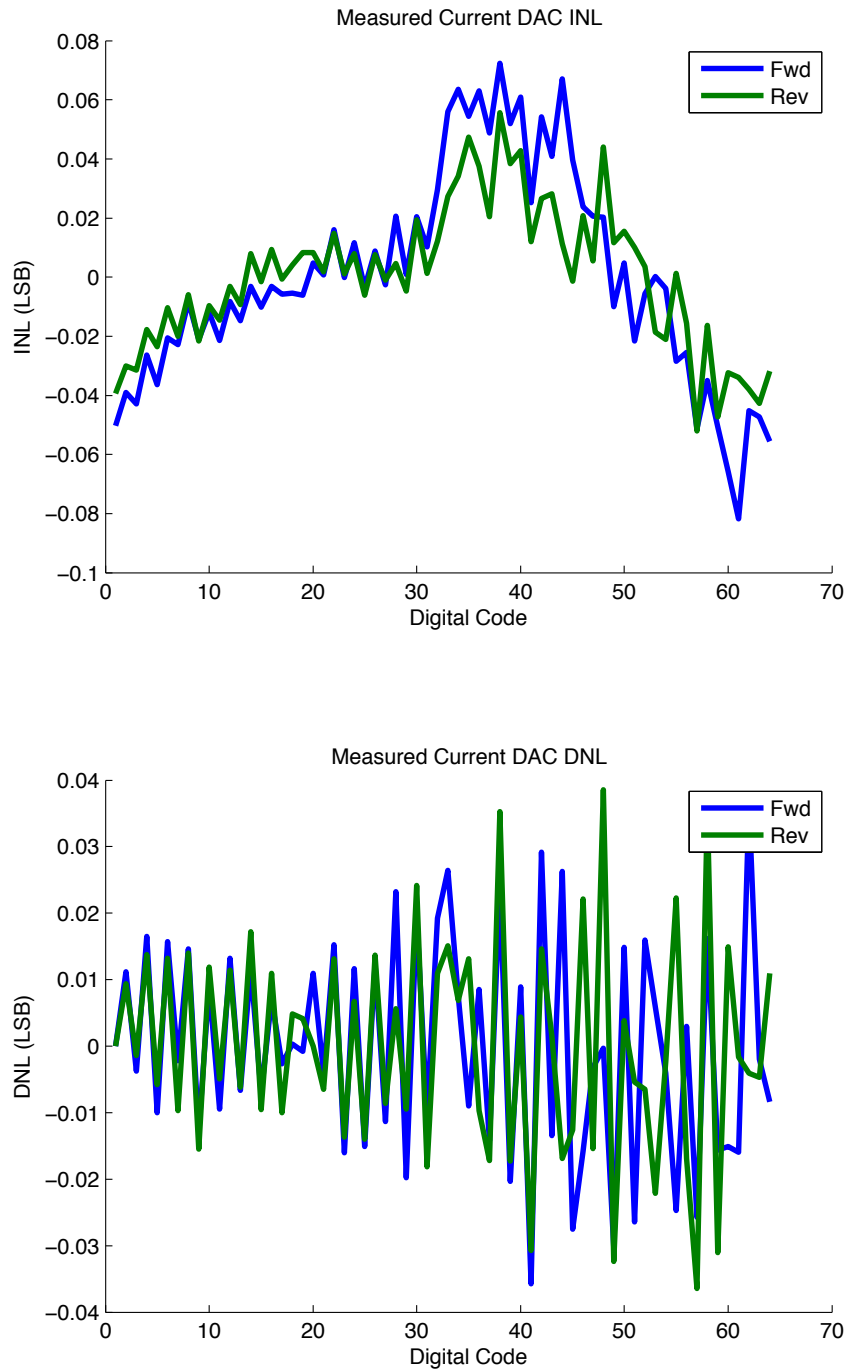
Figure 3.18: The measured DNL and INL of the 6b binary weighted current DAC.
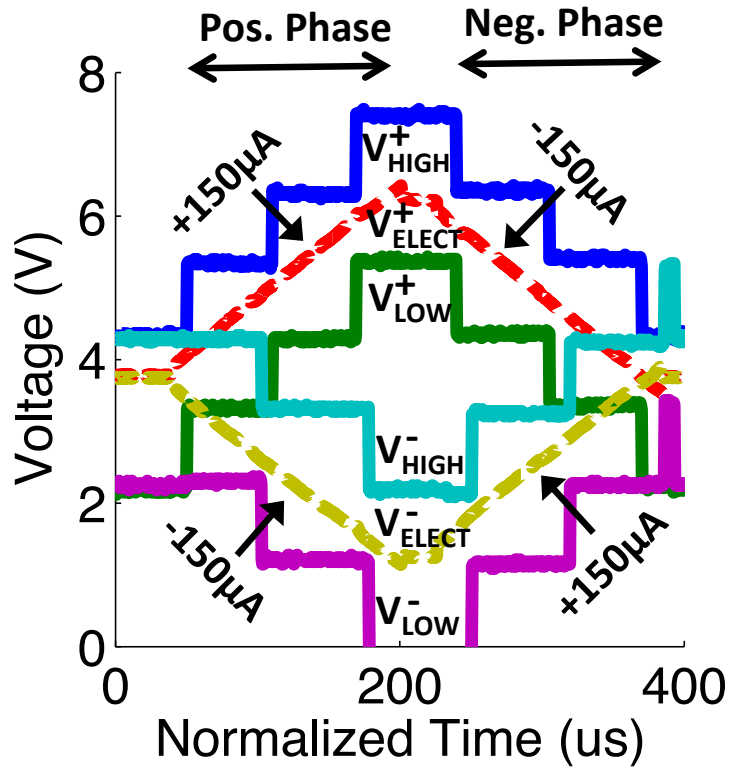
Figure 3.19: The electrode voltage and dynamically switching stimulator supply voltages during a typical stimulation pattern of 300μA differential current with 150μs per phase.
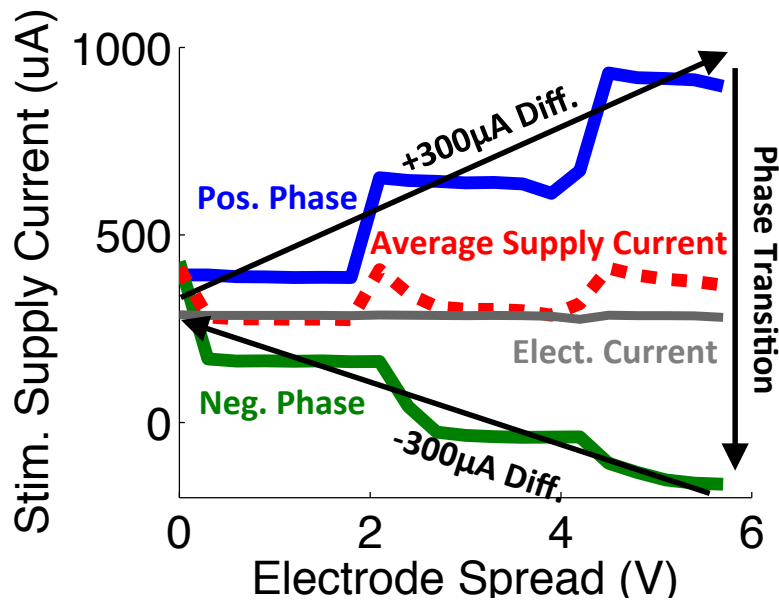
Figure 3.20: The current supplied by V_{unreg} via the DC-DC over one 300μA (differential) stimulation cycle.
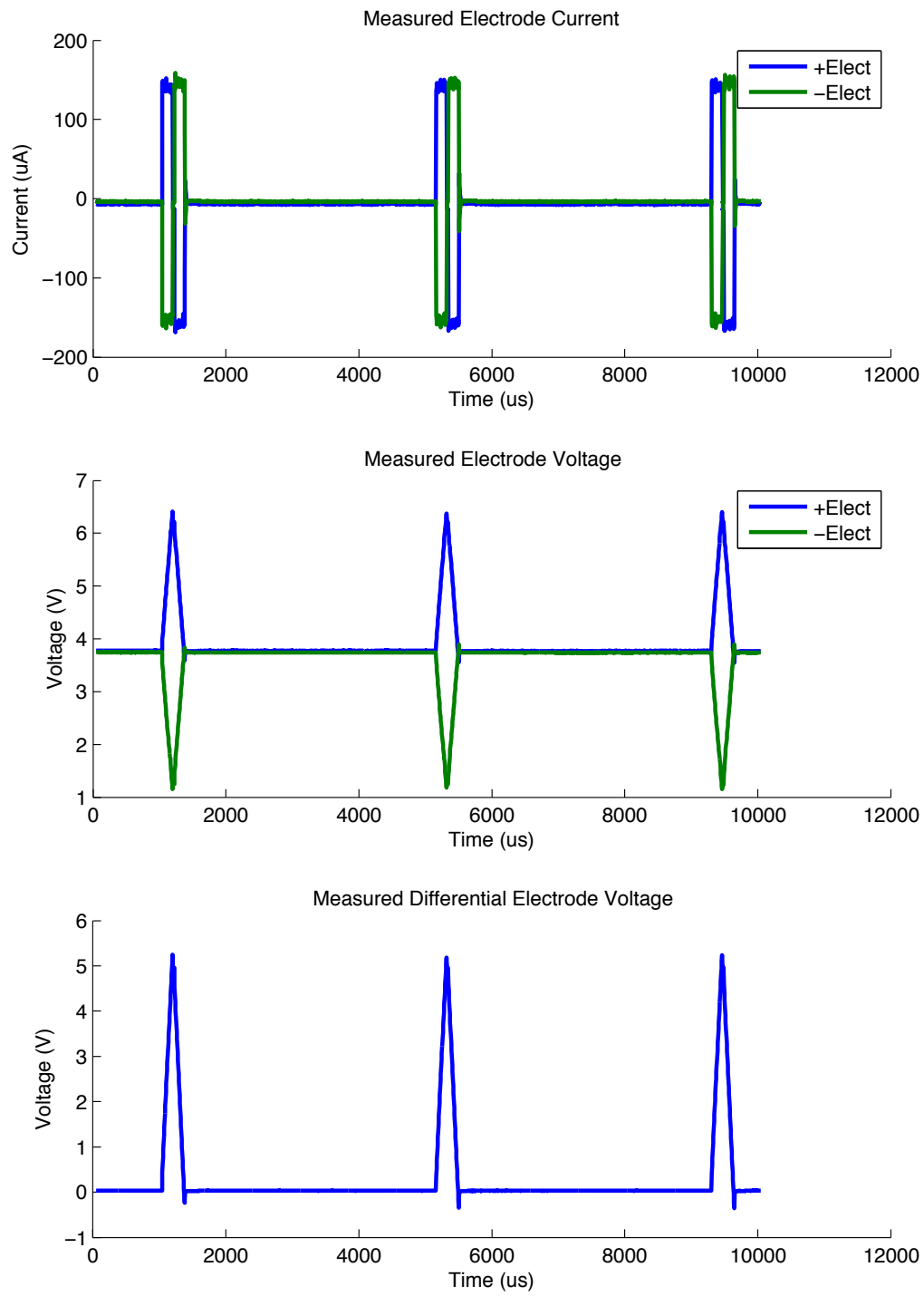
Figure 3.21: A train of stimulation current pulses and voltages measured from an electrode pair with a 4ms period.

3.4   In Vivo System Measurements

A diagram of the testing system designed to seamlessly obtain *in vivo* data is displayed in Figure 3.22, which includes a compact 0.65" x 0.8" headstage, a base station, and a Graphical User Interface (GUI). A photograph of the measurement setup is shown in 3.23. The SoC was incorporated onto the headstage, which was created to sit atop a small animal's head and connect to an implant in the brain. Information is transferred between the headstage and the base station via a 2.6mm diameter μHDMI cable using Low Voltage Differential Signaling (LVDS) for high speed communication. The base station serves as an intermediary between the headstage and the computer's GUI. From the GUI, the user can select which channel(s) to record, as well as send stimulation commands and adjust compression levels on a per channel basis.

In vivo stimulation was performed in the rat's visual cortex with a 210μA differential current for 125μs per phase. Stimulation artifacts are shown across multiple channels in Figure 3.24, and the relative amplitude correlates with proximity to the stimulation site. It is important to note that the recorded artifact was a differential measurement referenced to the labeled "Ref" electrode. In this orientation, the electrodes closest to the stimulation sites (CH1, CH2, and CH11) should exhibit the highest amplitude artifact while the furthest channels (CH3, CH4, and CH12) exhibit minimal artifact due to the localized nature of the differential stimulation. Figure 3.24 also shows that the time it takes for neighboring recording channels to recover from perturbation during stimulation is less than 1ms after stimulation ends.

Extracellular recordings were performed using a 16-channel microwire array implanted in the visual cortex of an adult Long-Evans rat. Arrays consisted of teflon-coated tungsten microwires (35μm diameter, 250μm electrode spacing, 250μm row spacing; Innovative Neurophysiology, Inc., Durham, NC, USA). All animal procedures were approved by the UC Berkeley Animal Care and Use Committee. Extracellular recordings were performed for several consecutive days, more than one month after the surgery. Clearly identified waveforms with a high signal-to-noise ratio were chosen for further investigation as single unit responses. Putative single units were validated based on waveform shape, reproducibility, amplitude, and duration. The characteristics of the inter-spike interval distributions were close to Poisson and exhibited a clear absolute refractory period.

A typical subset of recorded *in vivo* data is shown in Figure 3.25, which displays time-aligned epochs recorded from one channel. In order to verify *in vivo* compression accuracy, all three forms of the SoC's outputs were recorded and aligned in time, as displayed in Figure 3.26. Each epoch data packet includes a time stamp, which allows for spike detection confirmation when superimposed onto the raw data stream. In addition, accurate firing rate calculations were verified by ensuring that the firing rate counter incremented with each spike event. The SoC computes firing rates over a specified window of time, which in this case was 26.2ms.
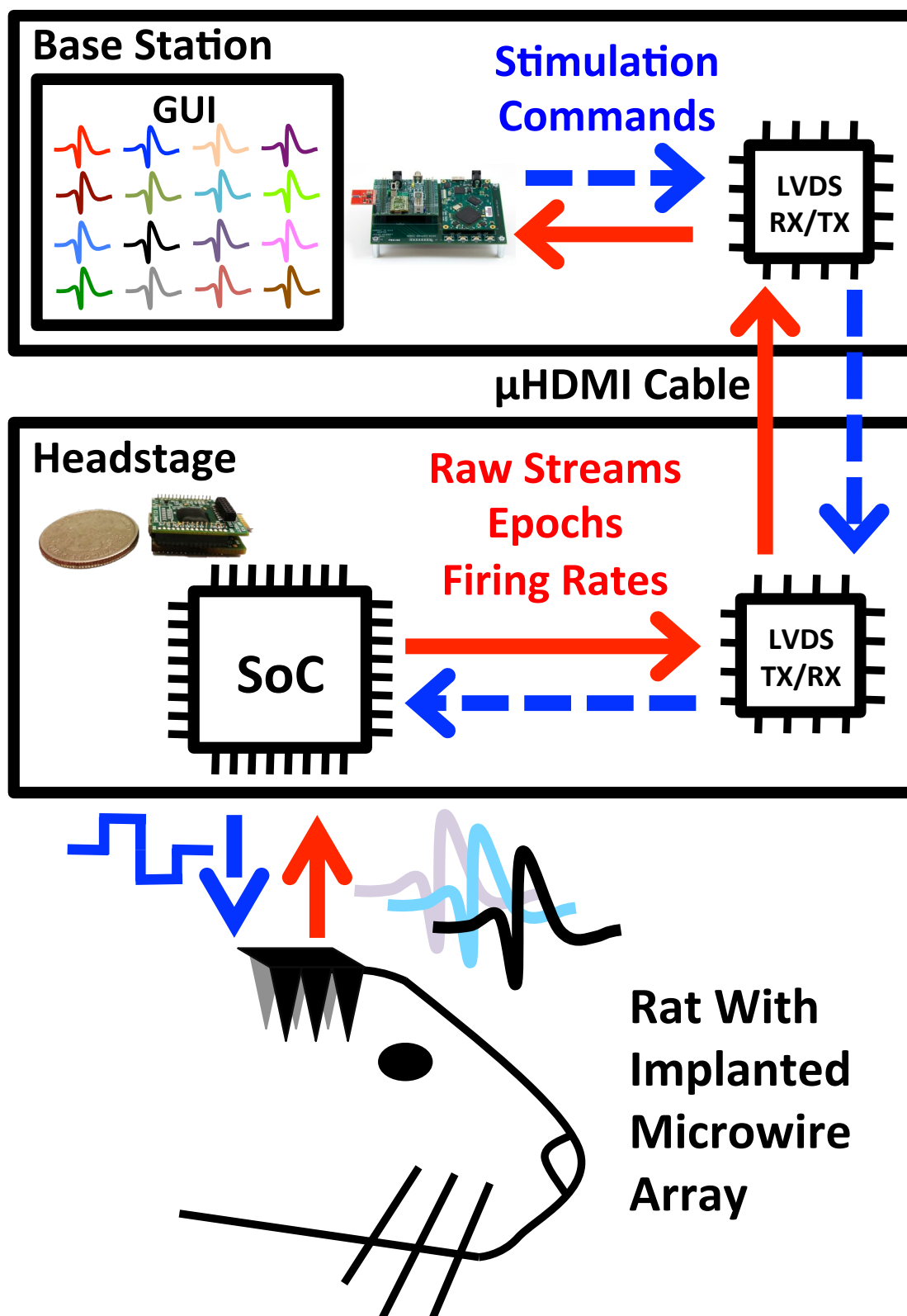
Figure 3.22: The *in vivo* neuromodulation test system is composed of a microwire implanted array, a compact headstage containing the SoC, a base station, and a Graphical User Interface (GUI).
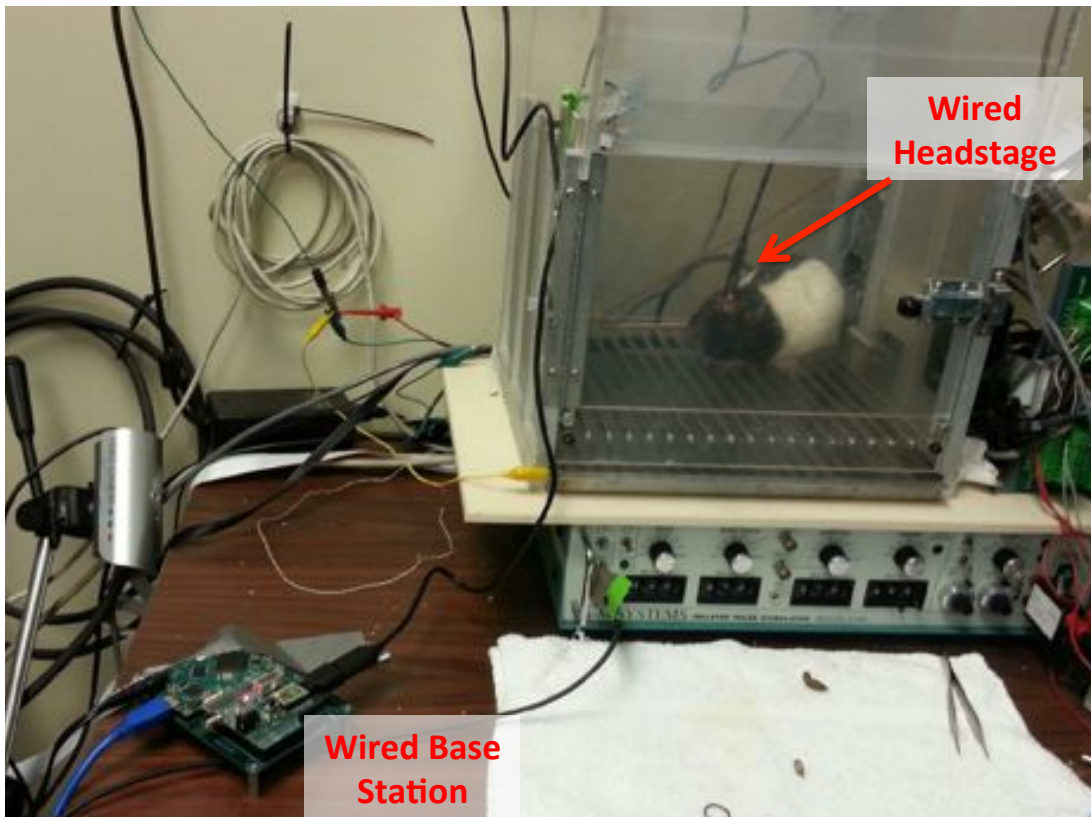
Figure 3.23: A photograph of the *in vivo* measurement setup as depicted in 3.22

3.5 Conclusion

For this system, I architected and designed the power management subsystem and a novel fully integrated stimulation architecture. The power management circuitry occupies less than 0.5% of the total IC area and consumes only 3.4% of the total IC power. It is fully tunable to adapt to a variety of system power conditions while maintaining a worst case 30dB PSRR. This work implements 2 stimulators, multiplexed onto 8 stimulation sites and reduces the total stimulator area per site by 2.25x compared to [20]. The stimulator utilizes an adiabatic architecture reducing power consumption by 3x less current on average compared to a traditional non-adiabatic topology operating from the same DC-DC. The DC-DC is implemented using a Dickson ladder topology with a tunable input voltage of up to approximately 1.3V and a conversion ratio of 1:7. The entire architecture was integrated on a single SoC without any off chip components.

The principal challenges I encountered in this design involved the use of high voltages (greater than the gate oxide breakdown of the MOS devices) and properly modeling the DC-DC. The largest thick oxide breakdown voltage in this process was 3.3V and since the
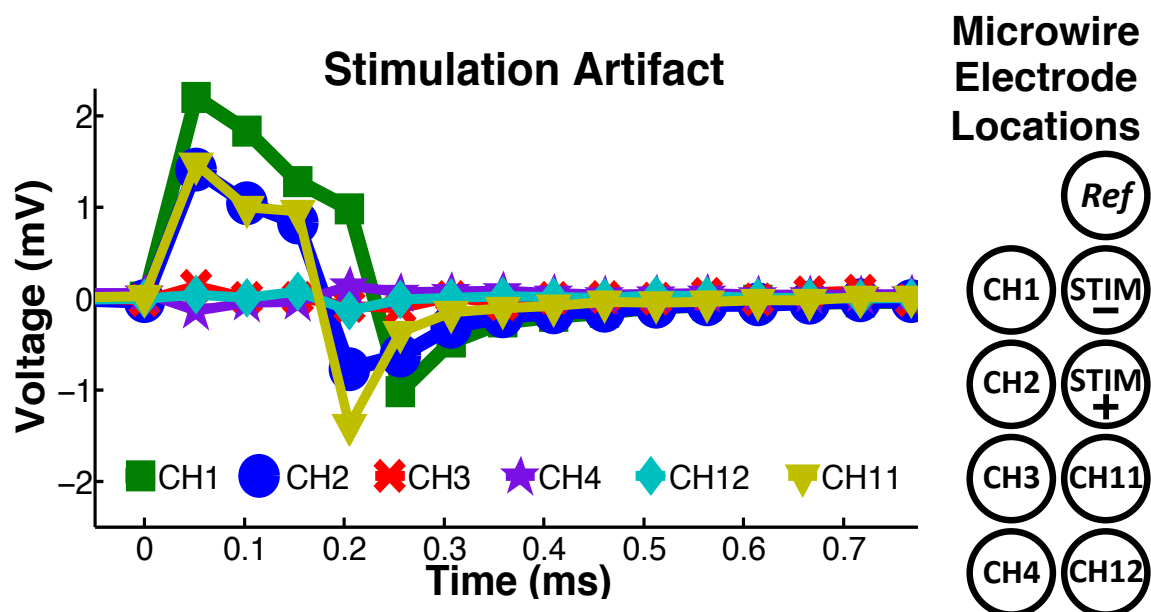
Figure 3.24: *In vivo* stimulation artifact measured by neighboring amplifier channels.
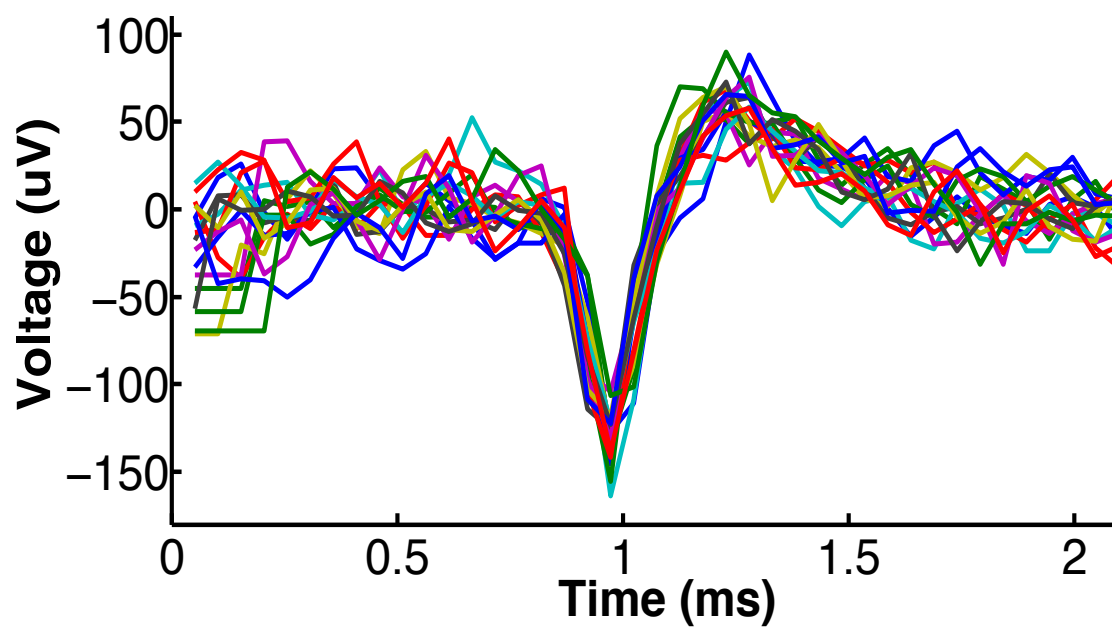
Figure 3.25: Time-aligned epochs recorded from one channel of *in vivo* neural data.
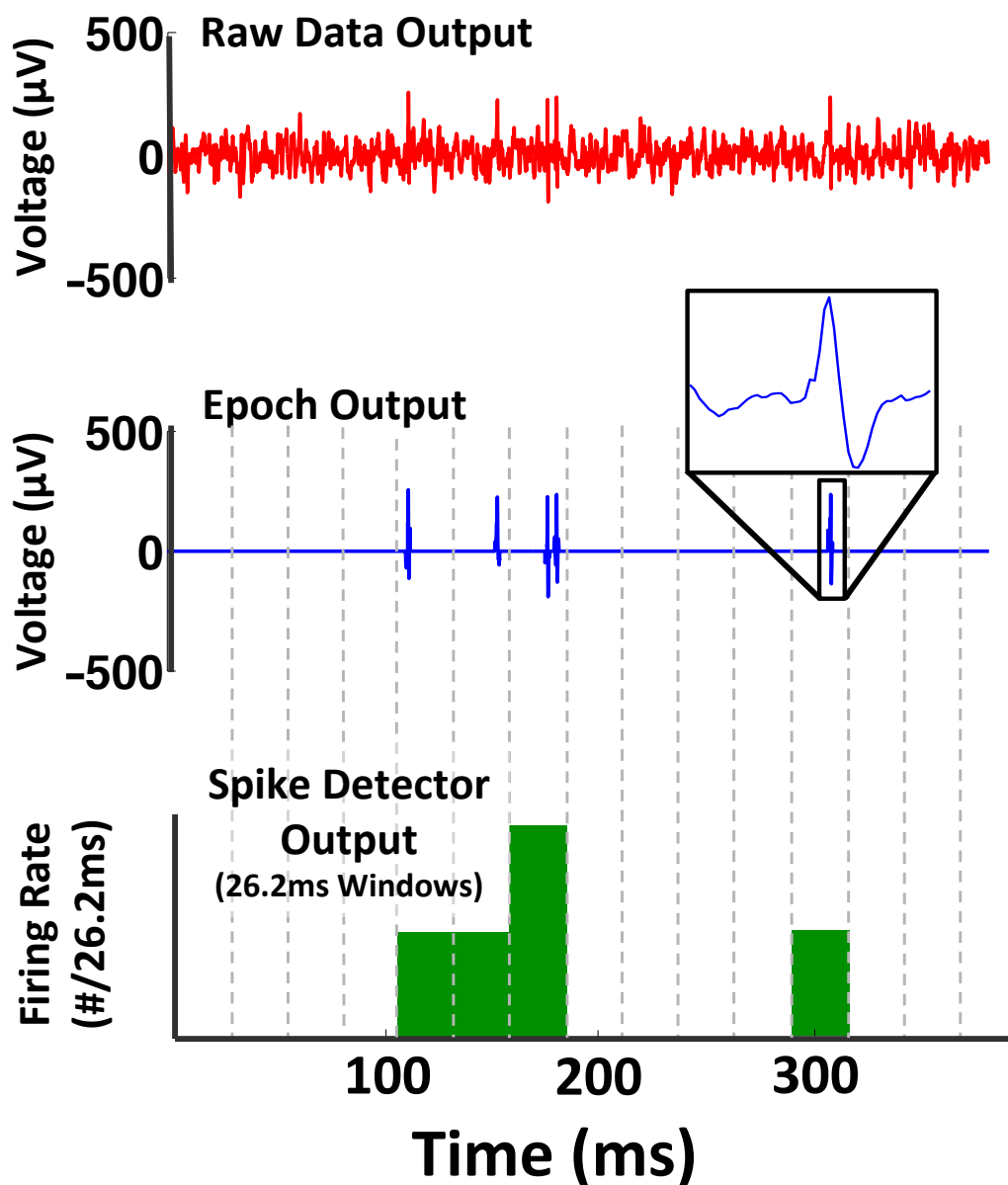
Figure 3.26: There are three possible digital outputs from the neuromodulation IC with varying levels of compression as shown for a typical *in vivo* recording. Raw streaming data (top) has no compression, this consumes 13.653 Mbps for 64 channels. Epoch data (middle) only sends a 2 ms window of data around a detected spike event, consuming 1.6384 Mbps (@50Hz firing rate) for 64 channels. Firing rates (bottom) only sends the count of detected spike events in a 26.2ms window and consumes 20kbps for 64 channels.

application requires generating voltages much higher than this on-chip, the majority of devices in the stimulator and DC-DC utilize floating wells. This creates complications for otherwise trivial circuit blocks, such as creating a current DAC, which then requires AC coupled level shifters with dynamically changing high and low voltage levels to be used for the control bits. These level shifters do not have a DC path to ground and therefore, require an input initialization pulse at startup for the output to take a known state. Achieving high DC-DC efficiency with a large conversion ration and high output voltage was also a challenging task. Care in modeling the NWELL and other parasitics as well as switch bootstrapping was necessary to ensure proper operation and startup. Despite the overall success, the DC-DC efficiency was approximately 9% less efficient than final simulations estimate.

The IC was fabricated in TSMC 65nm LP CMOS and occupies 4.78mm² of area including pads. A die photo is shown in Figure 3.27. The key metrics of the design are summarized in Figure 3.28 and compared with the state of the art. This work reduces the average amplifier power per channel by 14x and area per channel by 12x compared to [20] while achieving comparable NEF and PEF. Finally, the compression block consumes 2.7x less power and 6.4x less area per channel compared to [20] while implementing more features. The high integration level in addition to the low power and area consumption of this system provides the next step in enabling high-density, fully-implanted, wirelessly-powered neural interfaces in the human body.

2.225mm

2.15mm

32x Neural Amplifiers
1350um x 625um

1:7 DC-DC
550um x
700um

Digital Compression
1700um x 500um

Power Train
75um x 350um

32x Neural Amplifiers
1350um x 625um

2x Neural
Stimulators
450um x 300um
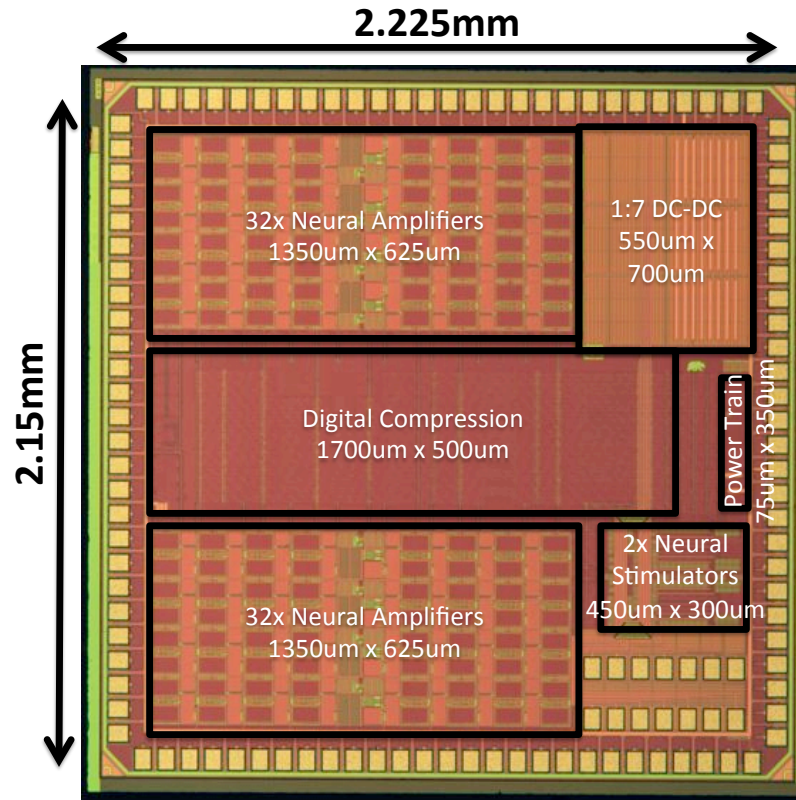
Figure 3.27: A die photo of the full system, with annotations for the primary circuit blocks and dimensions.

System Specs.	This Work	[2]	[3]	[4]
Technology / VDD	65nm / 1.0V, 0.8V (Ana, Dig)	0.35um / 1.5V	180nm / 1.8V	0.35um / 5.0V
Off Chip Req?	None	1uF Capacitor	DC-DC	16-Ch. Recording IC
# Amp / Stim Ch.	64 / 8	8 / 8	4 / 8	16 (Off-Chip) / 8
Amp & ADC Power / Area per Ch.	1.81µW / 0.0258mm²	25.8 µW / 0.3122mm²	61.25µW / 0.354mm²	N/A
Gain / LP / HP	45-65dB / 10-1k / 3k-8k	51-65.6dB / 1-525Hz / 5-12kHz	54dB / 700hz / 6khz	N/A
Noise / NEF/ PEF	7.5µVrms / 3.6 / 12.9	3.12µVrms / 2.9 (5.1kHz) / 12.6	Not Reported	N/A
Stim Imax / Area per Ch.	>500µA / 0.0675 mm²	94.5µA / 0.038mm²	4.2mA, 116µA / 0.05mm²	6.25mA / 0.7mm²*
Compression or DSP Type	Raw Data, Epoch, Firing Rate (any combination, per-ch.)	8 Spike Detector Outputs or 1 Ch. Raw	Log-DSP for LFP Energy, Output Mode: 4Ch Raw	Spike Detections, Classification, PCA
Digital Power / Area per Ch.	1.21µW (FR) & 1.775µW (Epoch) / 0.0105mm²	3.28µW / 0.0676mm²	34.5µW / 0.8mm²*	256.875µW / 0.191mm² *

*Area estimated from die photo

Figure 3.28: System summary and comparison table.

Chapter 4

Conclusion

This thesis presented two wireless, fully-integrated ICs designed to address the challenges that are preventing the realization of BMIs as a mainstream treatment for neurological disabilities. The free floating sensor node presented in Chapter 2 utilizes a subcranial interrogator to power and communicate with an array of implanted, free-floating AP sensors through the brain's dura. The wireless sensor was fabricated in a 65nm LP CMOS process with all electronics and the wireless interface integrated into an area of 0.125mm^2, without any additional off-chip components. This wireless neural sensor, the smallest reported to date, is small and light enough to free-float in brain tissue, reducing the effect of micro motion. The sensor consumes 15µA from an unregulated voltage source of 700mV, for a total power consumption of 10.5µW (2.6µW/channel) and 450µm x 250µm of area. This work reduces the average power per channel by 18x compared to [15] and 58x compared to [13]. Compared to [13], this work reduces the average area per channel by 10x, and decreases the amplifier and ADC area to 110µm x 100µm, compared to 400µm x 400µm (for an amplifier, comparator and DAC).

The transmission frequency for this system was selected to be 1.5GHz, trading a reduction in node size and channel loss for an increase in SAR. An integrated on-chip antenna was designed to optimize the wireless power link and minimize the required amount of transmit power, reducing tissue heating and power consumption of the interrogator. A transmission distance of 1mm in air was achieved with approximately 50mW of transmit power. The path loss in the brain was simulated to be approximately 6dB larger than in air, yielding an estimated transmission distance of 0.6mm *in vivo*. The power management circuits convert the inductively-coupled RF power source into a stable DC supply voltage and bias currents for the system. These circuits utilize novel discrete time techniques to achieve state-of-the-art area and power consumptions. A novel band gap reference was presented utilizing a SC network to create and sum voltages with opposite TCs, which enabled the lowest current, voltage and area bandgap-based reference reported to date. The design occupies an area of 100µm x 55µm and consumes 138nA from a 750mV supply. The area and power consumption are comparable to MOS V_{TH}-based references, while utilizing an architecture that retains lower process sensitivity.

The neuromodulation SoC presented in Chapter 3 is designed to enable closed loop BMI, and achieves significant improvements in area, power and signal compression over state-of-the-art systems with similar functionality (e.g. [20, 22, 59]). The IC was fabricated in TSMC 65nm LP CMOS and occupies 4.78mm^2 including pads, reducing the average amplifier power per channel by 14x and area per channel by 12x compared to [20] while achieving comparable amplifier NEF and PEF. This work implements 8 stimulation sites and reduces the total stimulator area per site by 2.25x compared to [20], while implementing a power saving, adiabatic architecture. The compression block consumes 2.7x less power and 6.4x less area per channel compared to [20] while implementing more features. When arrayed across the brain, 16 ICs provide 1024 recording and 128 stimulation sites, which would consume 6.67mW and require a data rate of 320kbps. The high integration level, in addition to the low power and area consumption of this system, provides the next step in enabling fully-implanted, wirelessly-powered, high-density neural interfaces.

Two differential, bi-phasic current stimulators are independently time multiplexed onto four electrode pairs for a total of 8 unique stimulation sites, minimizing the IC area overhead while maintaining a high number of stimulation sites. Each stimulator utilizes a single current source for both positive and negative stimulation phases which reduces the effect of the current mismatch between phases and eliminates the need for calibration. To minimize power consumption, the stimulator implements an adiabatic, charge-recycling architecture without utilizing off-chip components, enabling a fully integrated system as opposed to prior state-of-the-art [63]. A FSM dynamically selects one of 7 supply voltages from a switched-capacitor DC-DC converter to minimize the power consumed (and maximize the power recovered) throughout the stimulation cycle. The DC-DC is implemented using a Dickson ladder topology with a tunable input voltage of up to approximately 1.3V and a conversion ratio of 1:7. A maximum output voltage of 8.7V was measured from the DC-DC with an input voltage of 1.3V and measured peak efficiency of 68%. A typical stimulation pattern of 300μA differential current with 150μs per phase was performed on bench-top with a 1kOhm/10nF RC electrode model and consumes approximately 3x less current on average compared to a traditional non-adiabatic topology operating from the same DC-DC.

The free floating recording node and the neuromodulation IC each take different approaches to achieving the goal of a high density long-term interface with the human brain. The free floating recording node, presented in Chapter 2, focused on creating the smallest possible free floating active neural recording sensor to directly address the longevity challenge faced by BMI. By integrating all components, including the antenna, into a 450μm x 250μm die size, the maximum transmission distance is reduced and relies on a secondary system with a larger antenna on the surface of the brain to relay information through the skull. Furthermore, to keep the active circuit area to a minimum, the node only implements AP recording and defers other complexities, such as data compression, to the relay. On the other hand, the neuromodulation IC, presented in Chapter 3, focuses on integrating all aspects necessary for closed loop BMI onto a single integrated IC. Although the IC area is much larger, it was designed to be arrayed on a platform sitting on the surface of the brain to achieve very high recording and stimulation densities. Therefore, the design relies

on creating flexible and compliant probes to address longevity issues.

4.1 Future Work

4.1.1 Integration of a Floating Recording Node IC with Electrodes for *In Vivo* Studies

An SoC was presented which achieved more than an order of magnitude reduction in active circuit area compared to previous work, creating a miniaturized implant that can float in brain tissue. By eliminating the interface cables and creating a node small enough to free-float in brain tissue, the hypothesis is that micro-motion will be mitigated and the recording lifetime will increase. In order to study the effect on recording longevity, the IC needs to be integrated with recording electrodes. The IC was designed to attach to commercially available probes, such as a NeuroNexus probe (Figure 4.1). For small scale studies, the IC can be die attached, wire bonded and encapsulated with these NeuroNexus probes and studied side by side with a wired version of the probes as an experimental control.

Creating a high density recording array will require deploying thousands of these fully integrated nodes within the brain. Fabricating extremely large quantities of these nodes by laser dicing, or bonding to individual substrates, is burdensome. A scalable alternative is to directly etch the neural probe, including active circuitry, directly from a foundry wafer. The concept of a CMOS active neural probe has been demonstrated in [19] and the process flow that was used is shown in Figure 4.2. In their post-processing, they start with a five metal layer Al back-end-of-line (BEOL) TSMC wafer, without final passivation, resulting in exposed Al. First, a low stress layer of SiO_2 is deposited using alternating PECVD and CMP steps to achieve a target thickness. Second, vias are created through the SiO_2 to the Al pads to form the electrodes by RIE. Ti/TiN is deposited using PVD and patterned by RIE to form circular electrode sites. Fourth, SiO_2 is removed over bond pad locations by the same method as in previous steps. Fifth, RIE is used to create a trench to the final depth around the outline of the probe. Finally, the sixth and seventh step involves back grinding to release the dies and transfer to UV expansion tape.

4.1.2 Design of a Wireless Stimulation Headstage

Wireless head stages are a critical tool in the research of BMIs and general neuroscience studies that require free-moving animals. To date, prior work (e.g. [24], [25], [64]) is large, heavy (for a typical mouse or rat), has a battery life of hours and does not implement any stimulation channels. Furthermore, no wireless head stage developed to date is capable of optogentic stimulation, a technique which is growing in popularity and importance. A performance summary of recent work is shown in Table 4.1.

A prototype head stage was developed by Jaclyn Leverett and Daniel Yeager, Figure 4.3. The head stage was designed using the neuromodulation IC, presented in Chapter 3, plus
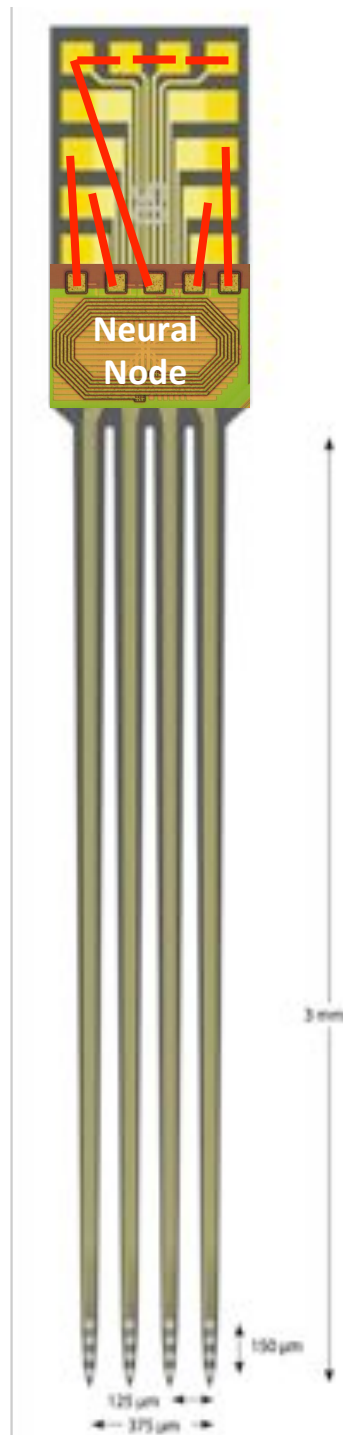
Figure 4.1: An implantable neural recording probe offered by NeuroNexus, the head of the probe is approximate 450μm wide. A die photo of the neural node is oriented in the approximate location for die attach and a potential bonding diagram is shown in red, allowing for 1 recording per shank. A large area reference electrode through the is created by stitch bonding multiple recording sites.
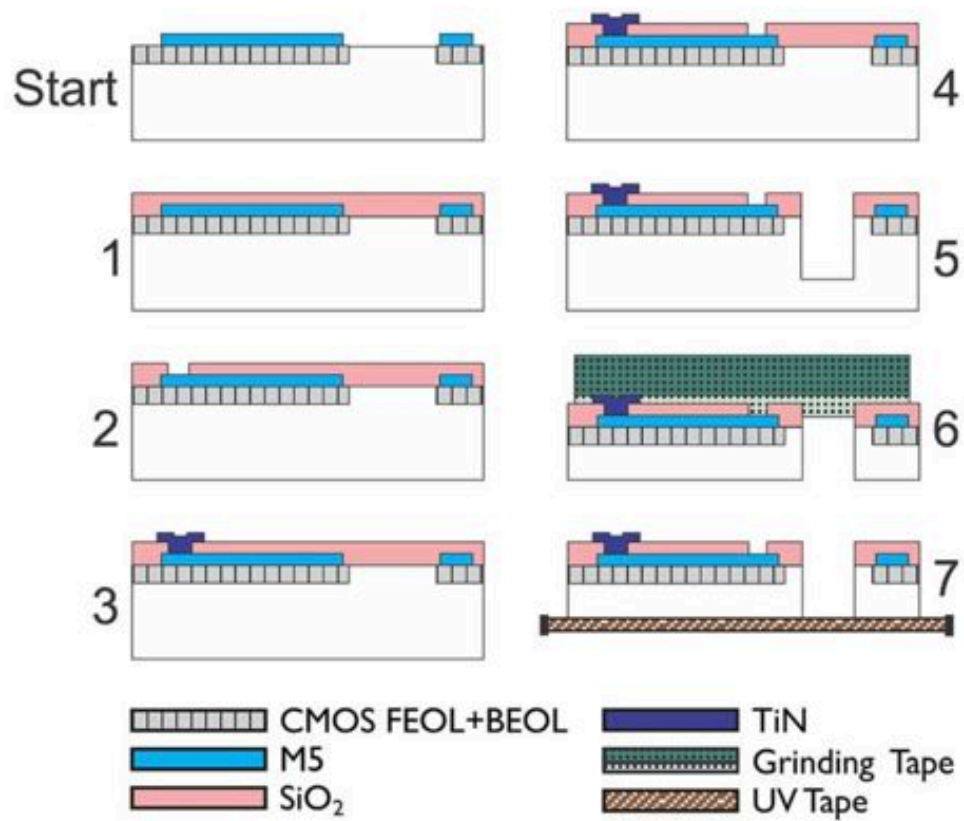
Figure 4.2: Example post processing steps for directly etching an active neural probe from a CMOS wafer, [19].

Table 4.1: A comparison of prior work on neural recording and stimulation wireless head stages.

Author	Recording/Stim (Channels)	Power Consumption (mW)	Battery Life (hours)	Size (cm³)	Weight (g)
Szuts [24]	64 / 0	345	6	150 [a]	67
Miranda [25]	32 / 0	142	33	73.64	60 [a]
Hampson [64]	16 / 0	1400 [a]	5	16 [a]	60.3

[a] Information not provided, estimate made from publication content.

off the shelf components, including a nordic radio and microcontroller. The board measures approximately 19mm x 25mm and utilizes two 32 channel Omnetics connectors for 64 total channels of amplification and one 16 channel Omnetics connector for 8 channels of differential stimulation. The battery used is a 110mAh, LiPo battery measuring 5.7x12x28mm (1.91 cm³) and weighing 2.65g. The battery is attached directly on top of other components on the board, resulting in a total thickness of less than 10mm and a weight of less than 5g. The power consumption is dominated by the nordic radio which consumes 7.5mA from the battery when transmitting at -12dBm, resulting in a battery life of more than 10 hours. The neuromodulation IC has enabled integration of 64 recording channels, and 8 electrical or optical stimulation channels into a form factor which is less than 10x the volume and weight compared to previous work which had less functionality and similar battery life.

4.1.3 Transcranial Link for a Free Floating BMI Platform

The architecture of the neuromodulation SoC was designed such that the IC can be arrayed on a customizable, scalable platform as shown in Figure 4.4. In this system, electrode shanks are connected to a flexible platform on the surface of the brain via a thin, flexible, tether in order to imitate a free floating probe and mitigate the effects of micro motion. The platform can have any number of neuromodulation ICs as required for the targeted application, which all communicate with power and data aggregation circuitry. An antenna is integrated into the platform and enables communication through the skull to a battery powered interrogator. The small and light weight interrogator acts as a relay for the data transmitted through the skull to a wireless base station.

Realization of the system illustrated in Figure 4.4 is limited by the available power at the platform. The maximum power transfer for a given platform size is determined by the transmission distance and the frequency of operation. In this case, the platform size is determined by the number of recording or stimulation channels desired; when arrayed on a platform, 16 ICs provide 1024 recording sites and 128 stimulation sites. The platform would occupy approximately 2.5 cm x 2.5 cm including peripheral circuitry, as shown in Figure 4.5. To estimate the channel loss characteristics, a skull thickness of 10mm and a scalp thickness of 2mm was used with Matlab modeling software developed in Chapter 2. The RX coil size was varied (up to the size required for 1k recording channels) and the minimal channel loss was reported in conjunction with the optimal frequency. RX coil sizes of 5mm, 10mm and
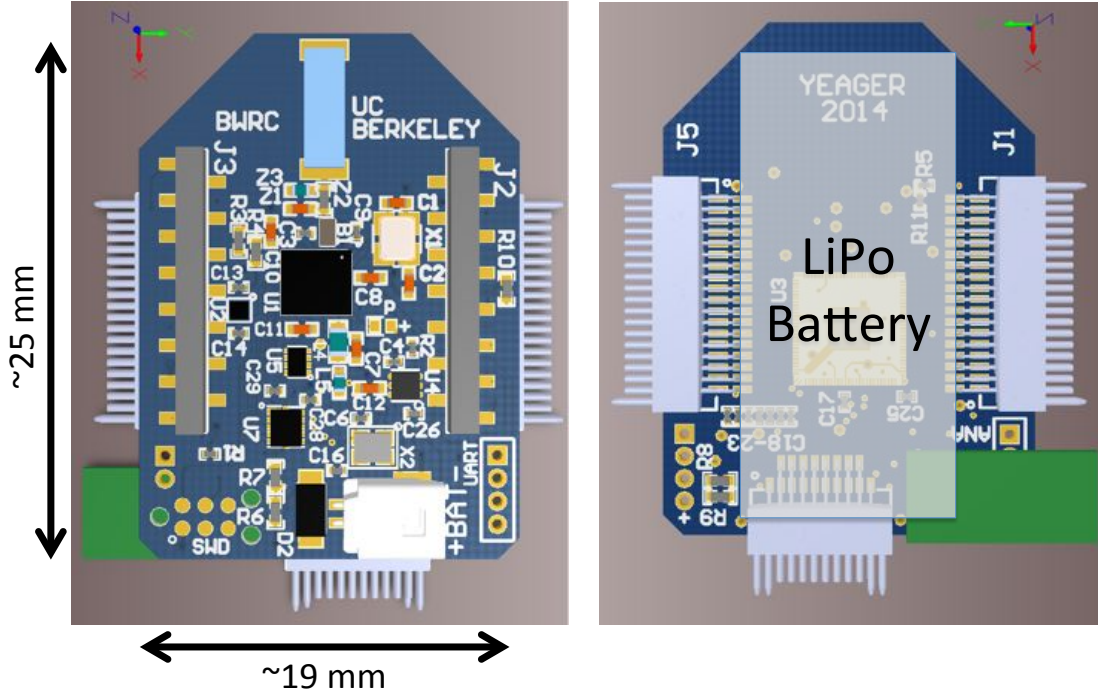
Figure 4.3: A wireless headstage designed by Jaclyn Leverett and Daniel Yeager, measuring 19mm x 25mm with 64 channels of neural recording and 8 channels of stimulation.

25mm were simulated in HFSS and found to match closely to calculation as shown in Figure 4.6. Based on simulations, the optimal transmission frequency for a 2.5cm x 2.5cm platform containing 16 neuromodulation ICs is approximately 13.56MHz and results in about 1dB of channel loss. This frequency is convenient because it corresponds to HF RFID, which will reduce the effort required to prototype and develop a reader.

4.2 Final Thoughts

Brain machine interfaces have the potential to revolutionize our understanding of the brain and restore motor function to amputees and patients suffering from paralysis. The use of micro-electrodes is the only method of both sensing neural signals and providing neural stimulation for the control of complex robotic prostheses in a closed loop BMI system. However, the success of BMI has been limited due to poor recording stability and ultimate chronic failure of recording sites caused by the reactive tissue response after implantation. In this thesis, I have presented novel techniques for power delivery, power management and
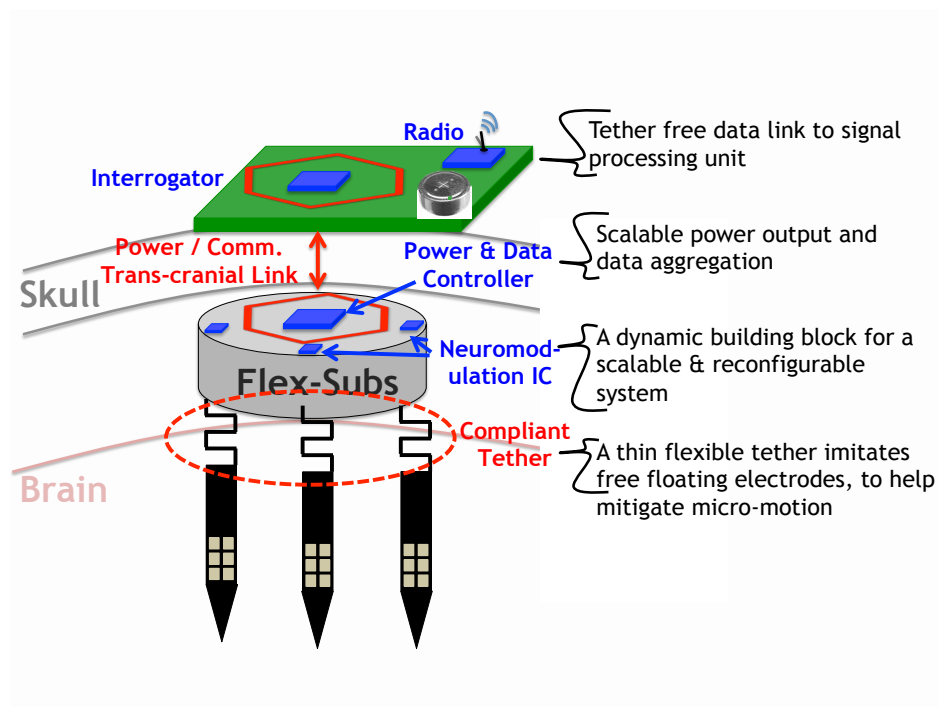
Figure 4.4: Conceptual diagram of a scalable platform with compliant tethers communicating to a wireless head stage through a trans-cranial link.

stimulation for wireless neural sensors and systems. These techniques have enabled extreme size and power reduction over present state-of-the-art BMI systems and ICs. The resulting systems represent significant progress towards realizing a fully implantable sub-cranial closed loop BMI system, creating new opportunities for high density, long term interfaces with the human brain.
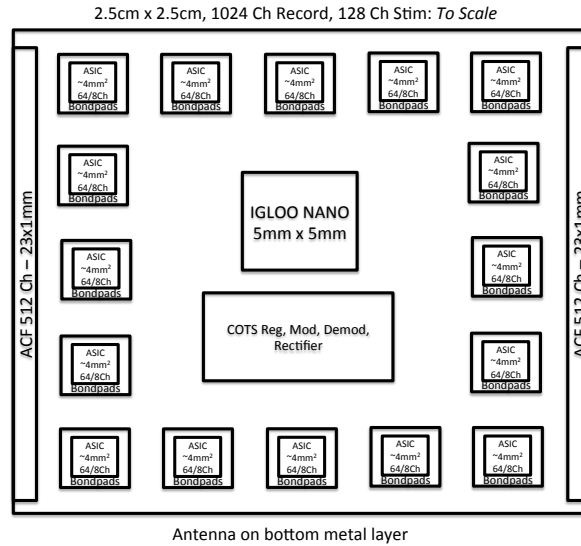
Figure 4.5: A floor plan diagram showing the size and layout of a 1024 channel implantable platform, with a total size of 2.5cm x 2.5cm.
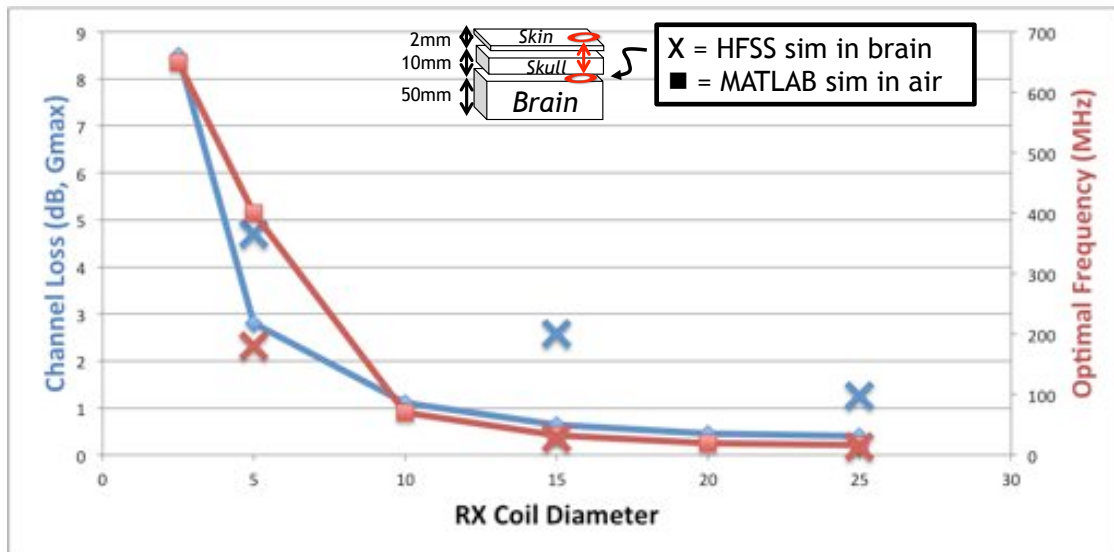
Figure 4.6: The channel loss and optimal frequency as a function of the RX coil diameter. For a 1024 channel platform, the maximum coil diameter is 25mm, resulting in an optimal frequency of approximately 13.56MHz and an estimated channel loss of 1-2dB.

Bibliography

[1]   World Health Organization. *Neurological disorders: public health challenges.* World Health Organization, 2006.

[2]   John K Chapin et al. "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex". In: *Nature neuroscience* 2.7 (1999), pp. 664–670.

[3]   Miguel AL Nicolelis and Mikhail A Lebedev. "Principles of neural ensemble physiology underlying the operation of brain–machine interfaces". In: *Nature Reviews Neuroscience* 10.7 (2009), pp. 530–540.

[4]   Jonathan Wolpaw and Elizabeth Winter Wolpaw. *Brain-computer interfaces: principles and practice.* Oxford University Press, 2011.

[5]   Birthe Rubehn et al. "A MEMS-based flexible multichannel ECoG-electrode array". In: *Journal of neural engineering* 6.3 (2009), p. 036003.

[6]   Gary Lehew and Miguel AL Nicolelis. "State-of-the-art microwire array design for chronic neural recordings in behaving animals". In: *Methods for neural ensemble recordings* 2 (2008), pp. 361–371.

[7]   KD Wise et al. "Wireless implantable microsystems: high-density electronic interfaces to the nervous system". In: *Proceedings of the IEEE* 92.1 (2004), pp. 76–97.

[8]   Richard A Normann et al. "A neural interface for a cortical vision prosthesis". In: *Vision research* 39.15 (1999), pp. 2577–2587.

[9]   Andrew B Schwartz et al. "Brain-controlled interfaces: movement restoration with neural prosthetics". In: *Neuron* 52.1 (2006), pp. 205–220.

[10]   Tobias Pistohl et al. "Prediction of arm movement trajectories from ECoG-recordings in humans". In: *Journal of neuroscience methods* 167.1 (2008), pp. 105–114.

[11]   Mikhail A Lebedev and Miguel AL Nicolelis. "Brain–machine interfaces: past, present and future". In: *TRENDS in Neurosciences* 29.9 (2006), pp. 536–546.

[12]   R. Muller, S. Gambini, and J.M. Rabaey. "A 0.013mm^2, 5μW, DC-Coupled Neural Signal Acquisition IC With 0.5V Supply". In: *Solid-State Circuits, IEEE Journal of* 47.1 (Jan. 2012), pp. 232–243. ISSN: 0018-9200. DOI: 10.1109/JSSC.2011.2163552.

[13]   R. R. Harrison et al. "A Low-Power Integrated Circuit for a Wireless 100-Electrode Neural Recording System". In: *Solid-State Circuits, IEEE Journal of* 42.1 (Jan. 2007), pp. 123–133. ISSN: 0018-9200. DOI: 10.1109/JSSC.2006.886567.

[14]   Seung Bae Lee et al. "An Inductively Powered Scalable 32-Channel Wireless Neural Recording System-on-a-Chip for Neuroscience Applications". In: *Biomedical Circuits and Systems, IEEE Transactions on* 4.6 (Dec. 2010), pp. 360–371. ISSN: 1932-4545. DOI: 10.1109/TBCAS.2010.2078814.

[15]   Moosung Chae et al. "A 128-Channel 6mW Wireless Neural Recording IC with On-the-Fly Spike Sorting and UWB Tansmitter". In: *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*. Feb. 2008, pp. 146–603. DOI: 10.1109/ISSCC.2008.4523099.

[16]   W. Wattanapanitch, M. Fee, and R. Sarpeshkar. "An energy-efficient micropower neural recording amplifier". In: *Biomedical Circuits and Systems, IEEE Transactions on* 1.2 (2007), pp. 136–147.

[17]   R.R. Harrison and C. Charles. "A low-power low-noise CMOS amplifier for neural recording applications". In: *Solid-State Circuits, IEEE Journal of* 38.6 (2003), pp. 958–965.

[18]   A.M. Sodagar et al. "An implantable 64-channel wireless microsystem for single-unit neural recording". In: *Solid-State Circuits, IEEE Journal of* 44.9 (2009), pp. 2591–2604.

[19]   C.M. Lopez et al. "An Implantable 455-Active-Electrode 52-Channel CMOS Neural Probe". In: *Solid-State Circuits, IEEE Journal of* 49.1 (Jan. 2014), pp. 248–261.

[20]   Meysam Azin et al. "A battery-powered activity-dependent intracortical microstimulation IC for brain-machine-brain interface". In: *Solid-State Circuits, IEEE Journal of* 46.4 (2011), pp. 731–745.

[21]   Wen-Sin Liew, Xiaodan Zou, and Yong Lian. "A 0.5-V 1.13-μW/channel neural recording interface with digital multiplexing scheme". In: *ESSCIRC (ESSCIRC), 2011 Proceedings of the*. IEEE. 2011, pp. 219–222.

[22]   Hyo-Gyuem Rhew et al. "A wirelessly powered log-based closed-loop deep brain stimulation SoC with two-way wireless telemetry for treatment of neurological disorders". In: *VLSI Circuits (VLSIC), 2012 Symposium on*. IEEE. 2012, pp. 70–71.

[23]   W. Biederman et al. "A Fully-Integrated, Miniaturized (0.125 mm^2) 10.5 ?W Wireless Neural Sensor". In: *Solid-State Circuits, IEEE Journal of* 48.4 (Apr. 2013), pp. 960–970. ISSN: 0018-9200. DOI: 10.1109/JSSC.2013.2238994.

[24]   Tobi A Szuts et al. "A wireless multi-channel neural amplifier for freely moving animals". In: *Nature neuroscience* 14.2 (2011), pp. 263–269.

[25] Henrique Miranda et al. "HermesD: A high-rate long-range wireless transmission system for simultaneous multichannel neural recording applications". In: *Biomedical Circuits and Systems, IEEE Transactions on* 4.3 (2010), pp. 181–191.

[26] S.F. Lempka et al. "Theoretical analysis of intracortical microelectrode recordings". In: *Journal of Neural Engineering* 8 (2011), p. 045006.

[27] P. Samsukha and S.L. Garverick. "A Monolithic Bandpass Amplifier for Neural Signal Processing with 25-Hz Low-Frequency Cutoff". In: *Custom Integrated Circuits Conference, 2007. CICC'07. IEEE*. IEEE. 2007, pp. 177–180.

[28] JN Turner et al. "Cerebral astrocyte response to micromachined silicon implants". In: *Experimental neurology* 156.1 (1999), pp. 33–49.

[29] Vadim S Polikov, Patrick A Tresco, and William M Reichert. "Response of brain tissue to chronically implanted neural electrodes". In: *Journal of neuroscience methods* 148.1 (2005), pp. 1–18.

[30] DH Szarowski et al. "Brain responses to micro-machined silicon devices". In: *Brain research* 983.1 (2003), pp. 23–35.

[31] John P Seymour and Daryl R Kipke. "Fabrication of polymer neural probes with sub-cellular features for reduced tissue encapsulation". In: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE. 2006, pp. 4606–4609.

[32] Amarjit Singh et al. "Glial cell and fibroblast cytotoxicity study on plasma-deposited diamond-like carbon coatings". In: *Biomaterials* 24.28 (2003), pp. 5083–5089.

[33] Yinghui Zhong and Ravi V Bellamkonda. "Dexamethasone-coated neural probes elicit attenuated inflammatory response and neuronal loss compared to uncoated neural probes". In: *Brain research* 1148 (2007), pp. 15–27.

[34] L Kam et al. "Selective adhesion of astrocytes to surfaces modified with immobilized peptides". In: *Biomaterials* 23.2 (2002), pp. 511–515.

[35] Cynthia A Chestek et al. "Long-term stability of neural prosthetic control signals from silicon cortical arrays in rhesus macaque motor cortex". In: *Journal of neural engineering* 8.4 (2011), p. 045005.

[36] Selim Suner et al. "Reliability of signals from a chronically implanted, silicon-based electrode array in non-human primate primary motor cortex". In: *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 13.4 (2005), pp. 524–541.

[37] Miguel AL Nicolelis et al. "Chronic, multisite, multielectrode recordings in macaque monkeys". In: *Proceedings of the National Academy of Sciences* 100.19 (2003), pp. 11041–11046.

[38] Hyunjung Lee et al. "Biomechanical analysis of silicon microelectrode-induced strain in the brain". In: *Journal of Neural Engineering* 2.4 (2005), p. 81.

[39] W. Biederman et al. "A CMOS Switched-Capacitor Fractional Bandgap Reference". In: *Custom Integrated Circuits Conference, 2012. CICC 2012. Digest of Technical Papers. IEEE International*. Sept. 2012.

[40] EPCglobal. *Class 1 Generation 2 UHF Air Interface Protocol Standard*. 2009.

[41] DJ Willatt, MW Yung, and TR Helliwell. "A correlation of the surgical anatomy of the dura to head and neck surgery". In: *European Archives of Oto-Rhino-Laryngology* 243.6 (1987), pp. 403–406.

[42] A.S.Y. Poon, S. O'Driscoll, and T.H. Meng. "Optimal Operating Frequency in Wireless Power Transmission for Implantable Devices". In: *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. Aug. 2007, pp. 5673–5678. DOI: 10.1109/IEMBS.2007.4353634.

[43] Sohee Kim et al. "Thermal impact of an active 3-D microelectrode array implanted in the brain". In: *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 15.4 (2007), pp. 493–501.

[44] R.R. Harrison. "Designing efficient inductive power links for implantable devices". In: *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*. IEEE. 2007, pp. 2080–2083.

[45] James T. Doyle et al. "Implementation of a 1 volt supply voltage CMOS subbandgap reference circuit". In: *Proc. IEEE International SOC Conference*. 2003, pp. 323–326. DOI: 10.1109/SOC.2003.1241535.

[46] P. Malcovati et al. "Curvature-compensated BiCMOS bandgap with 1-V supply voltage". In: *IEEE J. Solid-State Circuits* 36.7 (July 2001), pp. 1076–1081. ISSN: 0018-9200. DOI: 10.1109/4.933463.

[47] H. Banba et al. "A CMOS bandgap reference circuit with sub-1-V operation". In: *IEEE J. Solid-State Circuits* 34.5 (May 1999). ISSN: 0018-9200. DOI: 10.1109/4.760378.

[48] G. De Vita and G. Iannaccone. "A Sub-1-V, 10 ppm/°C, Nanopower Voltage Reference Generator". In: *IEEE J. Solid-State Circuits* 42.7 (2007), pp. 1536–1542. ISSN: 0018-9200. DOI: 10.1109/JSSC.2007.899077.

[49] Mingoo Seok et al. "A 0.5V 2.2pW 2-transistor voltage reference". In: *Proc. IEEE Custom Integrated Circuits Conference (CICC)*. Sept. 2009, pp. 577–580. ISBN: 978-1-4244-4071-9. DOI: 10.1109/CICC.2009.5280773.

[50] Po-Hsuan Huang, Hongchin Lin, and Yen-Tai Lin. "A Simple Subthreshold CMOS Voltage Reference Circuit With Channel- Length Modulation Compensation". In: *IEEE TCAS II: Express Briefs* 53.9 (Sept. 2006), pp. 882–885. ISSN: 1549-7747. DOI: 10.1109/TCSII.2006.881813.

[51] D. C. W. Ng, D. K. K. Kwong, and N. Wong. "A Sub-1V, 26µW, Low-Output-Impedance CMOS Bandgap Reference With a Low Dropout or Source Follower Mode". In: *IEEE Trans. on VLSI Systems* (Apr. 2010). ISSN: 1063-8210. DOI: 10.1109/TVLSI.2010.2046658.

[52] Barrie Gilbert and Shao-Feng Shu. *Switching bandgap voltage reference.* US Patent 5,563,504. Oct. 1996.

[53] Guang Ge et al. "A Single-Trim CMOS Bandgap Reference With a Inaccuracy of 0.15% From 40 °C to 125 °C". In: *IEEE J. Solid-State Circuits* 46.11 (Nov. 2011). ISSN: 0018-9200. DOI: 10.1109/JSSC.2011.2165235.

[54] S. Gambini and J. Rabaey. "Low-Power Successive Approximation Converter With 0.5 V Supply in 90 nm CMOS". In: *IEEE J. Solid-State Circuits* 42.11 (2007), pp. 2348–2356. ISSN: 0018-9200. DOI: 10.1109/JSSC.2007.906210.

[55] Han Wang and Qing Ye. "0.5-V operational transconductance amplifier for CMOS bandgap reference application". In: *Solid-State and Integrated Circuit Technology (IC-SICT).* 2006, pp. 1705–1707. DOI: 10.1109/ICSICT.2006.306400.

[56] Ming-Dou Ker, Jung-Sheng Chen, and Ching-Yun Chu. "A CMOS bandgap reference circuit for sub-1-V operation without using extra low-threshold-voltage device". In: *Proc. IEEE Intl. Symposium on Circuits and Systems.* May 2004. ISBN: 0-7803-8251-X. DOI: 10.1109/ISCAS.2004.1328126.

[57] A. Boni. "Op-amps and startup circuits for CMOS bandgap references with near 1-V supply". In: *IEEE J. Solid-State Circuits* 37.10 (Oct. 2002), pp. 1339–1343. ISSN: 0018-9200. DOI: 10.1109/JSSC.2002.803055.

[58] T. Ytterdal. "CMOS bandgap voltage reference circuit for supply voltages down to 0.6 V". In: *Electronics Letters* 39.20 (Oct. 2003), pp. 1427–1428. ISSN: 0013-5194. DOI: 10.1049/el:20030937.

[59] Tung-Chien Chen et al. "A biomedical multiprocessor SoC for closed-loop neuroprosthetic applications". In: *Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International.* IEEE. 2009, pp. 434–435.

[60] Reid R Harrison et al. "Wireless neural recording with single low-power integrated circuit". In: *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 17.4 (2009), pp. 322–329.

[61] Soumyajit Mandal and Rahul Sarpeshkar. "Power-efficient impedance-modulation wireless data links for biomedical implants". In: *Biomedical Circuits and Systems, IEEE Transactions on* 2.4 (2008), pp. 301–315.

[62] Vaibhav Karkare, Sarah Gibson, and Dejan Markovic. "A 130-μW, 64-channel spike-sorting DSP chip". In: *Solid-State Circuits Conference, 2009. A-SSCC 2009. IEEE Asian.* IEEE. 2009, pp. 289–292.

[63]   Scott K Arfin and Rahul Sarpeshkar. "An energy-efficient, adiabatic electrode stimulator with inductive energy recycling and feedback current regulation". In: *Biomedical Circuits and Systems, IEEE Transactions on* 6.1 (2012), pp. 1–14.

[64]   Robert E Hampson, Vernell Collins, and Sam A Deadwyler. "A wireless recording system that utilizes Bluetooth technology to transmit neural activity in freely moving animals". In: *Journal of neuroscience methods* 182.2 (2009), pp. 195–204.