UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Wireless Neural Interface Design

Permalink
https://escholarship.org/uc/item/5r00q0c5

Author
Yeager, Daniel

Publication Date
2014

Peer reviewed|Thesis/dissertation

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

Wireless Neural Interface Design

by

Daniel James Yeager

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

in

Engineering - Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jan Rabaey, Chair
Professor Elad Alon
Professor Robert Knight

Fall 2014

Wireless Neural Interface Design

Copyright 2014
by
Daniel James Yeager

Abstract

Wireless Neural Interface Design

by

Daniel James Yeager

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Science

University of California, Berkeley

Professor Jan Rabaey, Chair

Neural interfaces promise to radically change medicine. Currently, amputees and persons suffering from debilitating brain disorders lack a way to regain mobility and freedom. By recording and interpreting signals from the motor control regions of the brain, researchers have already demonstrated rudimentary control of robotic prosthetic arms in primate and human trials. Now, the next generation of neural interface electronics must provide the required advances in size and power consumption to enable long-term viability of complex, high degree-of-freedom prosthetic devices.

This dissertation presents two complete neural interface systems to address two key challenges: evading the brain's foreign body response to achieve long probe longevity, and scaling wireless, implantable systems to high channel counts. The first, a self-contained, 0.125 mm^2, 4-channel wireless recording system, achieves an unprecedented level of miniaturization. This opens the possibility of free-floating neural nodes in the brain tissue, which eliminates strain caused by transcranial wires. Ultimately, this may lead to probes that outsmart the brain's biological response, and provide stable, long-term recordings for chronic brain-machine interfaces. The second system achieves an unprecedented level of integration, combining 64 recording channels, 16 stimulation channels, and neural data compression onto a single 4.78 mm^2 IC. Furthermore, the IC achieves substantial improvements in power and area versus state-of-the-art. These improvements in performance and functionality enable neural recording systems that scale up to thousands of channels, or scale down to extremely compact, low weight, low area, wireless interfaces.

To My Parents

For their support of my inquisition and exploration, as far back as I can remember.

Contents

Contents **ii**

List of Figures **iv**

List of Tables **viii**

1 Introduction **1**
1.1 Applications of Neural Interfaces 5
1.2 Challenges 6
1.3 Neural Probes 9
1.4 Neural Recordings: the Signal and the Noise 9
1.5 Circuit Techniques for Neural Recording 11
1.6 Efficiency Metrics for Neural Amplifiers 13
1.7 Stimulation 14
1.8 Outline 14

2 Prior Work **16**
2.1 Low Noise Amplifiers 16
2.2 Analog to Digital Converters 17
2.3 Neural Interface Systems 18

3 A Fully-Integrated Wireless Neural Sensor **21**
3.1 Introduction 21
3.2 System Design 23
3.2.1 Frequency Selection 23
3.2.2 Communication Protocol 23
3.2.3 System Architecture 25
3.3 Communication Circuitry 26
3.3.1 Demodulator 26
3.3.2 Frequency Synthesis with FLL 28
3.4 Power Management 29
3.4.1 Wireless Power Transfer Optimization 29

3.4.2 Bias Current Generation . . . 30
3.4.3 Voltage Regulation . . . 32
3.4.4 Power-On Reset . . . 33
3.5 Data Acquisition . . . 33
3.5.1 Amplifier . . . 35
3.5.2 ADC and Mixed-Signal Feedback . . . 35
3.5.3 Calibration and Biasing . . . 36
3.5.4 Performance . . . 37
3.6 System Results . . . 37
3.6.1 Wireless Operational Range . . . 38
3.6.2 Single and Multi-Node Communication Tests . . . 38
3.6.3 Wirelessly-Powered Full System Test . . . 42
3.6.4 Wirelessly-Powered *In Vivo* Recording . . . 42
3.7 Conclusion . . . 43

4 A Fully-Integrated Neuromodulation SoC 47
4.1 System Vision . . . 47
4.2 Acquisition Channel Design . . . 49
4.2.1 Amplifier Design . . . 50
4.2.2 ADC Buffer . . . 53
4.2.3 10-bit SAR ADC Design . . . 53
4.2.4 Layout . . . 56
4.3 Characterization . . . 57
4.3.1 Amplifier Measurement Results . . . 57
4.3.2 ADC Measurement Results . . . 61
4.3.3 Performance Summary . . . 61
4.4 In Vivo Testing . . . 64
4.4.1 Measurement Setup . . . 64
4.4.2 In Vivo Measurements . . . 66
4.5 Conclusion . . . 68

5 Conclusion 71
5.1 Future Work . . . 72
5.2 Acknowledgements . . . 72

Bibliography 73

List of Figures

1.1 A neuron transmits action potentials through its axon to other neurons as well as organs such as muscles. A low-noise amplifier magnifies the electrical potential as measured by an extracellular electrode. Adapted from [2]. 1
1.2 An example raster plot showing detected spike locations as black lines. Average firing rates in a moving 100 ms window are overlaid in red. 3
1.3 Example tuning curves, which represent the average firing rate for different neurons vs. the direction of hand reach. For example *in vivo* data, see [16]. 4
1.4 Variation in spike shape from different neurons, normalized to a unit amplitude. Amplitudes also vary due to differences in distance between neurons and each electrode. 5
1.5 Future vision for a seamless brain-controlled prosthesis. 6
1.6 A conventional wired neural interface. Adapted from [24]. 7
1.7 Planar (left) and needle (right) neural probes. A rough scale for the size of a neuron can be seen in the planar probe, where the exposed electrodes are 30 μm in diameter. Probes range from 1-5 mm in length. Images adapted from [36] and [41]. 10
1.8 Typical components of a neural acquisition channel. 11
1.9 Widely-used amplifier topology proposed in 2002 by Harrison [48]. 12
1.10 Design tradeoffs for a neural acquisition channel. 13

3.1 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. 22
3.2 Theory of a Miller encoded communication scheme for multi-node interrogation showing miller encoded waveforms (2,4,6,8,10 MHz) for a data set. 24
3.3 Theory of a Miller encoded communication scheme for multi-node interrogation showing the resulting frequency spectrum from 5 nodes communicating simultaneously. 24
3.4 Theory of a Miller encoded communication scheme for multi-node interrogation showing the recovered raw waveform before and after bandpass filtering, and the recovered original transmitted M4 signal and equivalent data. 25

3.5 System diagram, subdivided into three primary circuit blocks: Power management, communication circuitry, and data acquisition. 26
3.6 Top: The implemented demodulator schematic utilizes two peak detectors and a replica detector to extract V_H, V_{env}, and V_L, which represent the high, instantaneous, and low envelopes, respectively. A switched capacitor filter calculates V_{mean}, the mean of the high and low envelopes. Bottom: A conceptual waveform showing the high and low envelopes which are recovered from the peak detector/replica and are averaged to generate slicing threshold for the sampling comparator. 27
3.7 Frequency-locked loop for subcarrier generation. Recovered beacons reset the accumulator (rst) and evaluate the comparator ($eval$). The comparator increments or decrements the second accumulator (up/dn), which in turn controls the DCO tuning bits (ctl). Finally, programmable dividers generate the output Miller subcarrier (f_{sc}) and internal 2 MHz clock (f_{2MHz}). 28
3.8 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. 31
3.9 Switched-capacitor bias current generation schematic, utilizing two-phase non-overlapping clocks. 31
3.10 Discrete-time LDO regulator schematic utilizing a comparator with capacitive offset cancelation and a charge pump loop filter. 32
3.11 The measured discrete-time LDO regulator supply rejection across frequency. . . 33
3.12 Power-on reset schematic. Switched-capacitor resistors pull up against pseudo-resistors, disabling the reset signal after several clock cycles. 34
3.13 Amplifier block diagram utilizing mixed signal feedback to cancel DC offset and LFP signal. 35
3.14 Measured amplifier frequency response over various high-pass corners, normalized to peak gain. 36
3.15 Amplifier front-end calibration, input and output common mode biasing, and cross-coupled gain boost. 37
3.16 Measured amplifier noise spectrum with and without the LFP loop enabled. . . 38
3.17 Left: Setup for wireless operational range testing. The IC is attached to a micromanipulator using double-sided tape. Right: Simulated path loss compared to the measured minimum TX power required for operation in air. 39
3.18 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). 40
3.19 Frequency spectrum of two wirelessly powered nodes, communicating simultaneously with the same interrogator at different sub-carrier frequencies. 40
3.20 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. 41

3.21 Simulated BER vs. SNR for 1-5 simultaneous nodes. 41
3.22 A wirelessly-powered system recording and transmitting a 1.6 kHz, 150 μV sine wave input from all channels simultaneously. 42
3.23 The setup for *in vivo* recording trials utilized a rat which was implanted with a microwire array. The 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. 43
3.24 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 500 Hz. 44
3.25 Chip layout and die photo of the full system, showing the input bonding pads, the RX coil and PGS (active area is underneath). 45

4.1 Berkeley wireless neural interface vision. 48
4.2 Berkeley wireless neural interface chip block diagram. 49
4.3 Block diagram of a recording channel row. Eight rows, each containing 8 amplifiers time-multiplexed onto a 10-bit SAR ADC, comprise the 64 channel array. . 50
4.4 LNA folded cascode core schematic. 51
4.5 VGA cascode core schematic. 51
4.6 LNA & VGA CMFB schematic. 52
4.7 LNA & VGA DC feedback level shifter schematic. 52
4.8 Bias current and 900mV cascode bias ("VBP900") generation. 52
4.9 ADC buffer core schematic. 53
4.10 Buffer feedback schematic. 53
4.11 Buffer and ADC timing diagram. 54
4.12 Simulated ADC buffer settling error versus input voltage (full scale input = 250 mV). 54
4.13 ADC top level schematic. 54
4.14 Bottom-plate sampled ADC capacitor array with sampling and evaluation switches. The array is composed of unit capacitors; half of the unit capacitors for each bit are sampled with respect to VDD and GND, respectively, which creates an effective 500 mV reference. 55
4.15 Layout of the amplifier and ADC. 57
4.16 Layout of a row of 8 amplifiers and 1 ADC. 57
4.17 Die photo of the 64-channel recording, dual stimulation and integrated compression IC. 58
4.18 Measured bode plot of LNA & VGA with gain set to 50 vs. selected high-pass corner settings. 59
4.19 Measured bode plot LNA & VGA with gain set to 500 vs. selected high-pass corner settings. 59
4.20 Measured VGA high-pass corner vs. control code at gain of 50 (blue) and 500 (red) V/V. 59

4.21 Measured (green) and simulated (blue) LNA & VGA gain vs. control code. . . . 59
4.22 Gain matching for the array, at the highest gain setting. 60
4.23 Measured amplifier input referred noise spectral plot. 60
4.24 Measured ADC DNL for 4 chips, with a peak of 1 bit. 61
4.25 Measured ADC DNL about midpoint for 4 chips, revealing correlation between chips. 61
4.26 Measured ADC INL for 4 chips, with a peak of approximately 1.5 bits. The average INL for the 4 samples is overlaid as a dark blue trace. 62
4.27 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). 64
4.28 A screen capture of the GUI with test signals present. The GUI is able to display waveforms in real time as well as log data to a file. 65
4.29 A screen capture of the GUI in overlay mode with test signals present. The GUI is able to display detected spike overlays using data from either the chip epochs or from full-streams using NEO or threshold-based software spike detection. . . 66
4.30 Time-aligned epochs recorded from one channel of *in vivo* neural data. 67
4.31 Raw streams, epochs, and firing rates of *in vivo* recorded data. 68

List of Tables

1.1 Neural Amplifier Input Signal Components 10

2.1 Prior work on low noise neural amplifiers 17
2.2 Survey of Neural Recording and Stimulation Systems 20

3.1 Comparison of neural recording systems with wireless telemertry. 44
3.2 Summary of System Performance 45

4.1 Component Performance Summary 63
4.2 Effective Per-Channel Power, Area, and Gain 63
4.3 System Comparison 69

Acknowledgments

I first want to thank my graduate advisors, Professors Jan Rabaey, Elad Alon, Brian Otis and Joshua Smith for their encouragement and guidance though my Master's and PhD studies. I truly understand how rare of an opportunity it was to work for advisors of this caliber. I also want to thank Will Biederman for his endless camaraderie and imagination. All of the systems presented in this dissertation were designed in direct collaboration with him and could not have been realized without him. Nathan Narevsky and Jackie Leverett were also fantastic teammates, and likewise, the systems could not have come together without them. Lastly, I'd like to thank Alanson Sample for leading me through my first research experiences, which provided a base for so many of my subsequent opportunities.

Chapter 1

Introduction

The human brain is perhaps the most energy-efficient computer that exists today. Of the roughly 100 watts that power a human, just 20% is required to enable our complex auditory and visual processing, planning, decision-making, and memory [1]. The planning, control, and visual / auditory processing capacity of a personal computer, which consumes 100-200 watts, pales in comparison. For example, the ability of a dedicated supercomputer to best a human at the game of chess is a recent breakthrough. How the brain can operate with such remarkable efficiency and robustness continues to inspire intensive study, from the biology of neurons to high level models of computation in the brain.

Neural activity in the brain is represented by electrical impulses, called action potentials (APs) or spikes (Fig. 1.1). These impulses carry information **from** sensory inputs (ex., ears), **between** neurons responsible for interpretation, decision-making, and memory, and

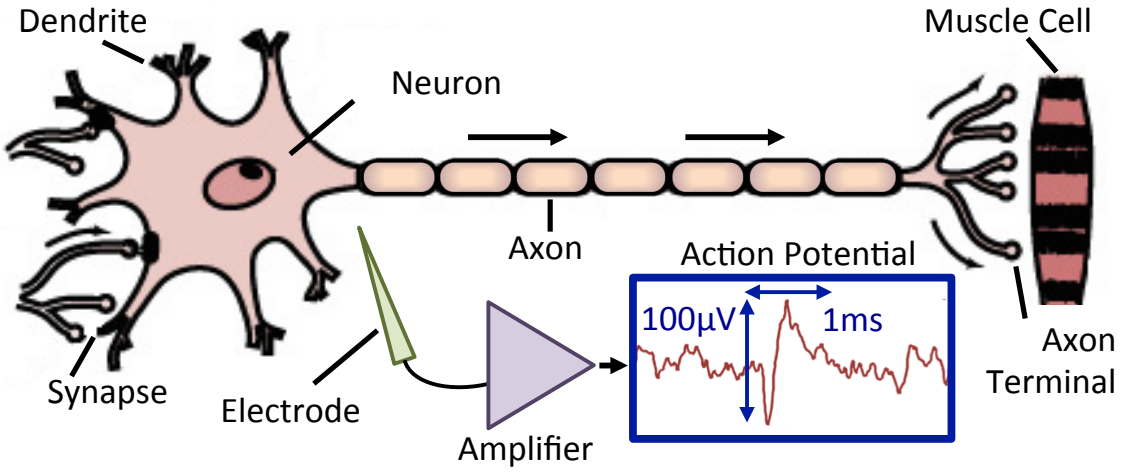


Figure 1.1: A neuron transmits action potentials through its axon to other neurons as well as organs such as muscles. A low-noise amplifier magnifies the electrical potential as measured by an extracellular electrode. Adapted from [2].

to muscles and other organs. Neurons have many inputs called dendrites, and many outputs called axon terminals. When there is sufficient stimulation of a neuron's inputs, an action potential is generated in the neuron, which travels through the axon to the terminals. The connections between terminals (outputs) and dendrites (inputs) are called synapses, which can be strengthened or weakened over time. This "plasticity" is an important mechanism in memory and learning [3, 4]. And, this plasticity is essential for robust control of a brain-controlled prosthetic [5].

Neuroscientists have gained a tremendous understanding of the brain through a spectrum of non-invasive and invasive methods, where the latter techniques require surgery. Examples of non-invasive methods include functional magnetic resonance imaging (fMRI) [6] and electroencephalography (EEG) [7]. fMRI uses magnetic resonance imaging to detect changes in blood flow in the brain, which relates to energy usage by brain cells [8]. This is useful for understand what areas of the brain are active, but provides relatively coarse spatial (2-3 mm) and temporal (1-5 s) resolution relative to the size and speed of a neuron. EEG employs electrodes on the scalp to detect electrical activity in the brain. Due to the filtering affect of the skull and scalp, as well as the distance of the probe from the brain tissue, spatial (1 - 10 cm) and temporal (DC - 100 Hz) resolution are also very constrained. Broad patterns such as phases of sleep cycles can be detected [9], and at least one study has linked AP activity with EEG activity [10]. One benefit of EEG over fMRI is that fMRI relies on a bulky MRI machine and stationary user, while EEG probes can be (conspicuously) worn without constricting the user.

Invasive methods involve placing electrically-conductive probes **in** and **on** the brain tissue. These two methods are called intracortical recording and electrocorticography (ECoG), respectively. Intracortical recording, shown in Fig. 1.1, places the probes in close proximity ($<$100 μm) to the neurons. This in turn provides a high-fidelity, fine-grained view of the behavior of individual neurons. ECoG achieves much better spatial and temporal resolution than EEG by placing the probes within 1-2 mm of the neurons, but still provides an aggregate, population level recording like EEG. Substantial progress has been achieved using ECoG to understand and predict speech decoding in the brain [11].

The remainder of this dissertation focuses on invasive techniques, and more specifically, on intracortical recording. The primary reasons for this focus are twofold. First, to date, the direct recording of action potentials (APs) is the only type of neural interface proven to provide enough temporal and spatial resolution to control complex robotic prostheses [12]. This is also evident in the approaches taken by recent human clinical trials [13, 14]. Second, intracortical probes can be used deliver optical or electrical impulses to "stimulate" neurons and thereby evoke neural activity. With these techniques, neuroscientists have been able to mimic sensory input to the brain [15]. This could enable sensory feedback from a prosthesis.

When neuroscientists record from an array of intracortical probes, a tremendous amount of data is generated. Even just 100 recording channels, sampled with 10-bit resolution and at 20×10^3 samples-per-second, produce 150 MB per minute of data. These raw traces hold little value on their own, so the data is next passed through a spike-detection algorithm. The algorithm attempts to discriminate between the higher signal-to-noise ratio (SNR) spikes of

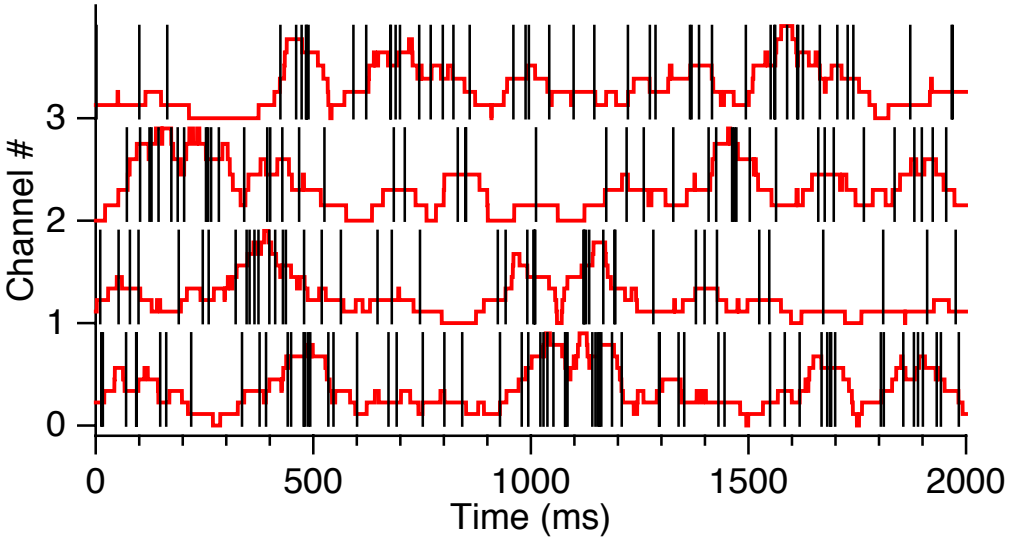


Figure 1.2: An example raster plot showing detected spike locations as black lines. Average firing rates in a moving 100 ms window are overlaid in red.

nearby neurons and the "biological interference" of many distant, low SNR neurons added to thermal noise of both the recording electronics and the electrode impedance. The temporal locations of detected spikes are then saved, often with samples of the detected spikes for later analysis and verification.

The detected spikes can be viewed in several ways. A raster plot shows activity versus time, with a vertical line for each detected spike as shown in Fig. 1.2. However, of even greater value is the average firing rate, in spikes per second, over a moving window between 50 ms and 500 ms (Fig. 1.2). This enables study of correlation between a neuron's firing rate and either motor control (ex., arm movement) or perception (ex., auditory stimuli). Fig. 1.3 depicts an idealized example study of firing rate vs. arm reach direction, sometimes called a tuning curve. This representation helps illustrate one way that neural activity encodes information; i.e., the firing rate can be modeled as a cosine or gaussian function of the arm reach direction, with different neurons tuned to different directions. Then, using these tuning curves to reverse decode the intended arm reach direction, a cursor on a computer screen or robotic arm can then be manipulated in real time via recordings from these neurons.

Algorithms that are based on neuron firing rates rely on an accurate spike detector. This detector takes as an input the noisy raw recorded waveform (Fig. 1.1) and outputs a binary decision to represent when spikes occur (Fig. 1.2). Many different spike detection algorithms have been proposed in the literature. The most simple algorithm detects spikes via threshold crossing, which may be set to some multiple of the RMS noise level. Moderate complexity algorithms attempt to filter the recorded waveform to increase detection accuracy. Examples include the nonlinear energy operator (NEO) [17], derivative operation, and high-pass filter [18]. A threshold detection may then be applied to these filtered waveforms. High complexity

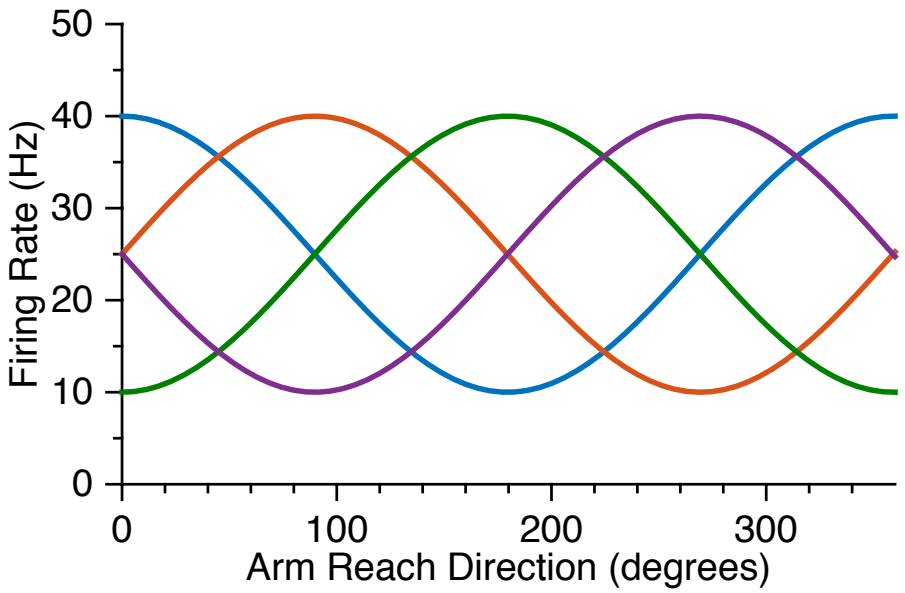


Figure 1.3: Example tuning curves, which represent the average firing rate for different neurons vs. the direction of hand reach. For example *in vivo* data, see [16].

algorithms typically involve correlation with characteristic spikes, basis functions, or average recorded waveforms [19]. Clustering algorithms may be applied initially to attribute recorded spikes to neurons and generate characteristic spikes for the correlation operation. Finally, a threshold operation may be used to detect spikes from the correlation output. These high complexity algorithms can discriminate between spikes that originate from different nearby neurons based on their amplitude, which represents distance [20], and shape [21] (Fig. 1.4). By sorting spikes rather than attributing all action potentials per electrode to one neuron, more information is extracted from the recording because this measures firing rates for more individual neurons.

Electrodes arrays generally lack precise control of electrode placement in the brain tissue, and electrode locations can shift over time due to many factors. When a neuron is very close to an electrode and thus has high relative SNR, it is called a single-unit recording. Here, a simple threshold detector may be sufficient to discriminate spikes from noise. More often, however, one or more neurons are moderately close to a neuron, resulting in low-SNR spikes. Here, a moderately-complex algorithm is necessary to accurately discriminate spikes from noise. When spikes from multiple neurons are present, called multi-unit activity, a high-complexity algorithm can be used to attribute spikes to individual neurons.

Spike sorting typically employs a computationally-intensive algorithm to correlate measured spike waveforms against the average waveform for each nearby neuron, which is at odds with the stringent power requirements that will be discussed in Section 1.2. Fortunately, researchers have also demonstrated reasonable performance without spike sorting [22]. These tradeoffs between power consumption and recording quality are a recurring theme through-

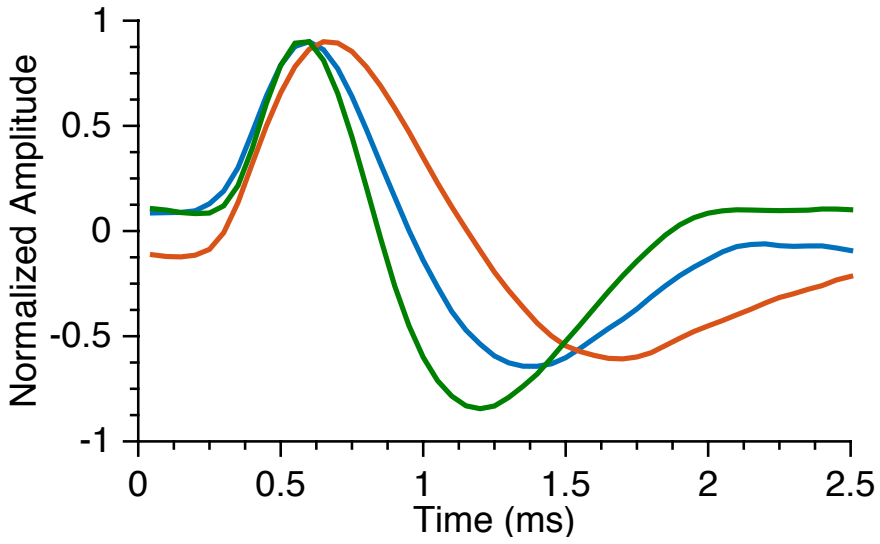


Figure 1.4: Variation in spike shape from different neurons, normalized to a unit amplitude. Amplitudes also vary due to differences in distance between neurons and each electrode.

out this dissertation.

The remainder of the introduction includes an overview of therapeutic and scientific applications for brain interfaces in Section 1.1, and a detailed discussion of challenges in Section 1.2. A background on neural recording probes, electrical characteristics of recorded signals, and signal acquisition electronics follows in Sections 1.3, 1.4, 1.5, 1.6, and 1.7. Finally, Section 1.8 provides an outline of this dissertation.

1.1 Applications of Neural Interfaces

A neural interface is the ability to record and/or evoke neural activity in the brain, which facilitates study of cognition, memory, perception, and motor control. Neural interfaces also enable important medical techniques. For example, open-loop deep brain stimulation is currently used to treat Parkinson's disease symptoms such as tremors, and ECoG is replacing ice water as a tool to guide surgical treatment of epilepsy and cancer. However, the most audacious goal of this research is to achieve a seamless, long-term brain-machine interface (BMI). This gateway into the brain may someday enable natural brain control of prosthetic limbs, computer interfaces, and unforeseen new applications. These advances will provide an incredibly powerful clinical tool to help amputees and those suffering from debilitating brain disorders, as pictured in Fig. 1.5.

Already, breakthrough research in both primate [23] and human [13, 14] studies have demonstrated the basic feasibility of brain-controlled prosthetic devices. However, there are a host of technical challenges to overcome before neural interfaces can achieve widespread clinical adoption. Solving these challenges will require broad interdisciplinary collaboration



Figure 1.5: Future vision for a seamless brain-controlled prosthesis.

ranging from the physical design of electrodes that reside in and on the brain tissue, to circuits that record and stimulate neurons via those electrodes, to RF links which provide power and retrieve data to/from the electronics through the skull. Techniques and designs to overcome some of these challenges represent the main contribution of this dissertation.

1.2 Challenges

From the early work of Ken Wise in the 1970s to ongoing research efforts today, circuit designers have been working aggressively to make the science-fiction of neural interfaces a reality. Why have circuit designers spent so long on this problem, and why are neural interfaces a circuit design problem? The answer lies in the power and data link through the skull.

Fig. 1.6 shows a conventional wired neural implant. The probe is implanted in the brain tissue, typically reaching depths of 1-2 mm. An epoxy or dental acrylic fill attempts to seal the brain and tissue from bacterial infection, and wires carrying the neural signals extend through the fill to a connector. The high-impedance, low-SNR signals from the electrodes are prone to 60 Hz interference and electrostatic pickup, which introduces motion artifacts. To mitigate these problems, a buffer amplifier or acquisition IC is placed as close to the

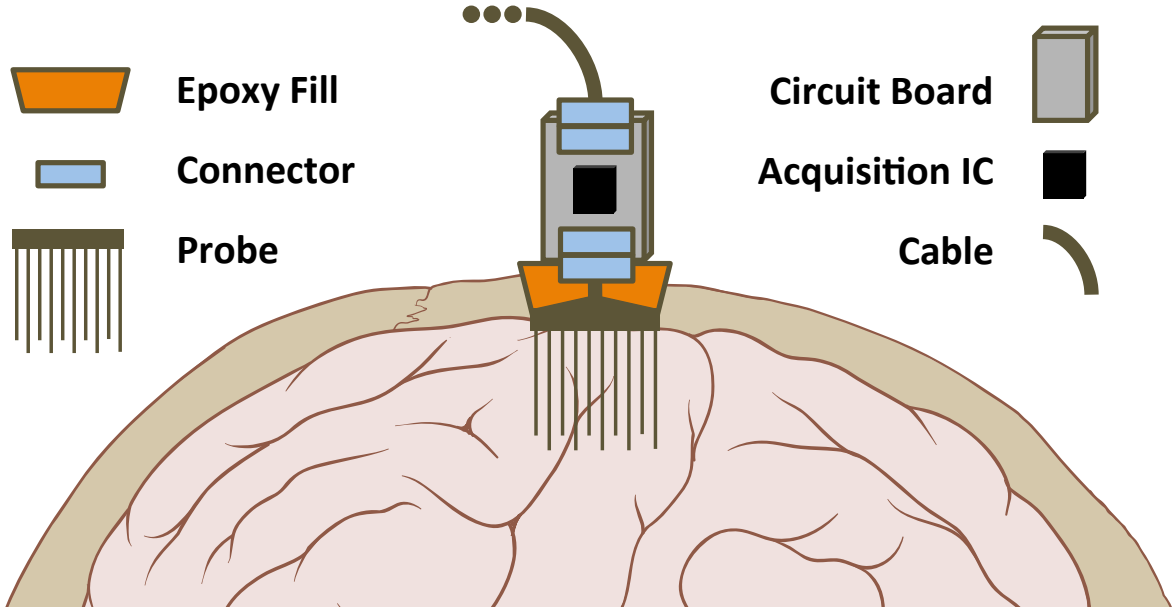


Figure 1.6: A conventional wired neural interface. Adapted from [24].

implant as possible. Finally, a wired tether carries the neural information to a computer for storage and processing.

Wires through the skull cause two problems. First, they present a chronic infection risk. Because the skin cannot bond to the epoxy fill to prevent bacterial influx, a continuous application of anti-bacterial medicine is required. However, this continued infection risk is particularly dangerous in the brain, where there is no immune system to ward off infection. Second, attaching a neural implant to wires leading through the skull interferes with the natural movement of the brain inside the skull. This creates chronic abrasion from the movement of the electrodes in the brain, which is theorized to contribute to the buildup of scar tissue. This scar tissue insulates the electrodes from the neural signals, thereby reducing the longevity of the implant. It is possible to eliminate these wires, but it places stringent demands on the implant design, and this is where the circuit designers come in.

To combat the problems with wires, power delivery and data transmission can be performed wirelessly. The power delivery link typically employs near-field, inductive coupling between two coils. For a transcranial link, the "primary" coil is positioned outside the scalp and the "secondary" coil lies beneath the skull on the implanted recording platform ("implant"). The primary coil is driven by a transmitter, typically at a single frequency in an unlicensed band such as 125 kHz or 13.56 MHz. This transmitted power couples onto the secondary, where a rectifier converts the sinusoidal excitation to DC. Power regulation circuits create a stable supply voltage from this DC rectifier output. Note that this technique is currently used in cochlear implants [25], electric toothbrushes [26], and most pertinently, near-field communication (NFC) radio-frequency identification (RFID) [27].

The wireless data link often employs backscatter or load modulation, also similar to RFID tags [27]. Because backscatter communication relies on reflection of the signal from the primary, very little power is expended by the implant. Some systems have proposed an active transmitter in the 433 MHz or 915 MHz unlicensed bands, which requires a substantial radio power budget [28].

Wireless, implantable neural recording systems face many interrelated challenges that all stem from two forms of miniaturization. First, to minimize the severity of the surgery, the implant's surface area should be kept below approximately 1 cm^2. Second, to limit the implant's displacement of the brain, the implant should be maximally planar with a thickness of a few millimeters at most.

Extreme miniaturization creates two specific electrical design constraints. First, the surface area constraint limits the antenna aperture, which in turn limits the amount of power that can be wirelessly delivered via inductive coupling through the skull. Prior work indicates that 10 mW can be delivered to a 1 cm diameter implant coil at 1 cm depth [29, 30, 31, 32]. Second, safety constraints limit the amount by which implanted medical devices may raise the surrounding tissue to 1 °C, or 40 mW/cm^2 [33]. For example, one study modeled a 6x6 mm^2 implant and found the limit to be about 10 mW [34]. Since the limit for thermal power for a 1 cm^2 implant is higher than the delivered power, an implanted storage mechanism could enable short-duration, high-power activities like stimulation. For example, a supercapacitor or rechargeable battery could be trickle-charged with surplus power while recording. During stimulation events, this storage device would be depleted. At any rate, this limited power budget translates into three main constraints on the continuous recording electronics that comprise the implant.

The first constraint on the electronics is the thermal noise of the neural recording channels. Thermal noise requirements place a lower bound on power consumption of the low noise amplifiers (LNA) that condition signals from the neural electrodes. In a system with 1000 amplifiers, small improvements in amplifier efficiency dramatically affect the system. Or, viewed another way, there is a direct tradeoff between the power consumption of the amplifiers and the number of neurons that can be observed.

The second constraint on the electronics is the radio power consumption, which limits the wireless data rate. Many implants utilize backscatter, similar to passive RFID tags. This consumes nearly zero power by the implant, but limits uplink (implant to reader) data rates to approximately 4 Mbps [35]. Each neural channel generates 200 kbps or more of raw data. A significant percentage of this data contains no neural information, but it does contain substantial background noise. A previously described, an accurate spike detector is is required to extract the neural information in the presence of this background noise. Then, a dataset tailed to the specific application (neuroscience vs. BMI) can be transmitted out of the skull.

The third constraint on the electronics is limited computational resources. Neuroscientists have built their understanding of the brain, their prototype BMI systems, and research programs around sophisticated algorithms running on powerful computers. Practical, unobtrusive BMI systems cannot rely on such computing power, and must compress the neural

data on the implant in order to satisfy the data rate constraints of the wireless link. Furthermore, this compression must be done within a per-channel power budget comparable to the amplifier. Fortunately, the neuroscience community is learning that brain-machine interfaces can be made robust without sophisticated spike sorting algorithms [22]. However, there is still tremendous room to innovate as neuroscientists are generally unaware of the circuit implementation power cost of their algorithms, and circuit designers generally lack access to means of validating that their algorithms work robustly in a clinical setting.

The challenges outlined above highlight the gap between clinical neuroscience and the attempts of circuit designers to create the next generation of neural recording tools. They also highlight the monumental effort required to bring a full-featured clinical device to market. This discussion has focused primarily on the neural recording path; ultimately, this neural data must be translated into control signals for a robotic prosthesis. These decoding algorithms, as well as motor control algorithms, electro-mechanical prosthesis design, and other technical challenges must also be overcome before a clinical prosthesis can be made.

1.3 Neural Probes

Fig. 1.1 depicts the conceptual setup for extracellular recording. However, Fig. 1.7 shows a more true-to-scale drawing of planar [36, 37, 38] and needle [39, 40] probes, which are widely used. As can be seen in Fig. 1.7, probes are designed with a limited surface area such that each electrode captures signals from a small number of neurons. This limited surface area translates into a high source impedance, and this in turn influences the design of the recording electronics.

A probe is typically constructed of metal or silicon and encased in a biocompatible insulator like parylene. The exposed electrode sites are often made of a porous or textured metal like platinum to increase surface area and thereby lower impedance. Electrode sites can vary from less than 100 μm^2 to over 1000 μm^2. A smooth metal surface results in 0.2 $pF/\mu m^2$, and rougher surfaces may have five times this capacitance [42, 43]. For example, a 36 μm diameter probe (1000 μm^2) may have a capacitance ranging from 200 pF to 2 nF. Probe manufacturers and researchers typically report the probe impedance at 1 kHz. For example, a 200 pF probe impedance translates to 0.8 MΩ at 1 kHz. For a comprehensive review of probe manufacturing techniques, prior work and current research efforts, see [44].

1.4 Neural Recordings: the Signal and the Noise

A typical extracellular neural recording consists of three signal components: an electrochemical offset, the low frequency component of the extracellularly-recorded neural activity, known as the Local Field Potential (LFP), and the high frequency APs. Table 1.1 compares the frequency range and approximate recording amplitude of these signal components to the approximate recording noise floor.

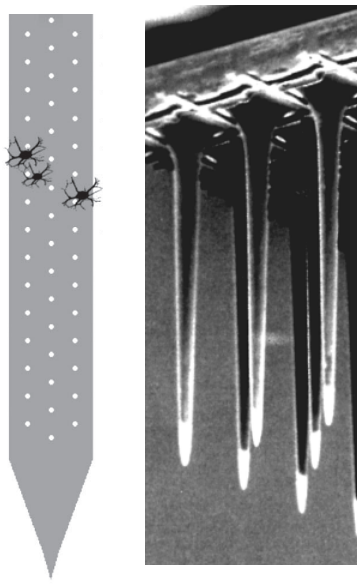


Figure 1.7: Planar (left) and needle (right) neural probes. A rough scale for the size of a neuron can be seen in the planar probe, where the exposed electrodes are 30 μm in diameter. Probes range from 1-5 mm in length. Images adapted from [36] and [41].

Table 1.1: Neural Amplifier Input Signal Components

Signal Component	**Frequency**	**Amplitude**
Electrochemical Offset	DC	±50 mV
Local Field Potentials (LFP)	10-300 Hz	5 mVpp
Action Potentials (AP)	500-3000 Hz	50-500 μVpp
Recording Noise Floor	450-10,000 Hz	13.5 μVrms [a]

There are two natural sources of cortical recording noise: thermal and biological. Thermal noise is generated by the resistive part of the electrode and tissue interface. The electrode impedance, and thus noise, is dependent on electrode size; a 1000 μm^2 probe contributes approximately 5 μVrms of thermal noise [45]. Biological noise is simply interference from neighboring neurons, which naturally falls in the same frequency bands as the desired AP signals. Prior work has modeled thermal and biological noise during cortical recording using silicon microelectrodes and found that for a 450 Hz to 10 kHz recording bandwidth, the recording noise floor is approximately 13.5 μV (based on Section 4.2 and Table I from [46]).

The resulting 70 dB (13.5 μV to 50 mV) input dynamic range requirement of the amplifier is typically reduced via AC-coupling, where the high-pass corner is set at approximately

[1] Based on Section 4.2 and Table I of [46].

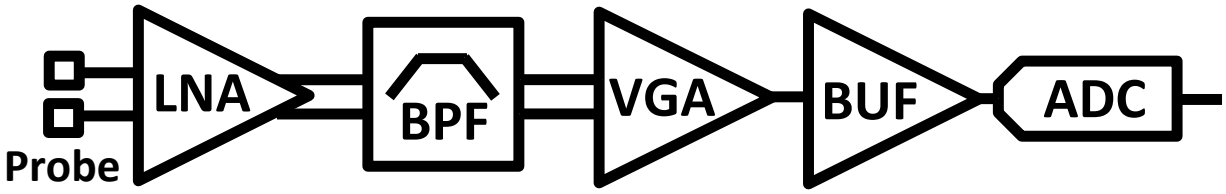


Figure 1.8: Typical components of a neural acquisition channel.

500 Hz to eliminate the offset and LFP. This reduces the dynamic range to approximately 32 dB (13.5 μV to 500 μV). Section 1.5 presents a more detailed introduction to circuit techniques.

1.5 Circuit Techniques for Neural Recording

The field of circuit design for neural recording dates back to the 1970's when Ken Wise pioneered the first efforts to integrate an amplifier with the neural probe [47]. Since those first efforts, a barrage of varying circuit design approaches have appeared in the literature. However, most systems follow an architecture similar to the form shown in FIg. 1.8, which consists of a low-noise amplifier (LNA), band-pass filter (BPF), variable-gain amplifier (VGA), buffer (BUF), and analog-to-digital converter (ADC).

The first amplifier (LNA) plays an important role in setting the power and noise performance of the system. Because the input-referred noise (IRN) of subsequent amplifiers are reduced by the gain (often 20-40 dB) of the preceding stages, it is the first stage that dominates the total IRN. This is why it is referred to as the low-noise amplifier (LNA). Consequently, the LNA consumes a large fraction of the total system power, and optimizing the noise efficiency (power consumed vs noise added to signal) is the focus of many research efforts. Because large transistor area is required to suppress flicker noise, the LNA tends to also dominate the area of the signal acquisition channels.

The purpose of the BPF is to provide a high-pass cutoff around 500 Hz to remove the DC offset and LFP as well as a low-pass cutoff around 7-10 kHz to remove high frequency noise which can alias in-band when sampled by the ADC. The BPF is often integrated into the LNA and VGA. For example, the capacitive feedback network of the LNA can perform the high-pass function. Similarly, limiting the bandwidth of the amplifiers can perform the low-pass function.

The VGA scales neural signal ranging from 50 μVpp to 1 mVpp to the full scale range of the ADC of around 1.0 V. The BUF provides a low output impedance to drive the sampling capacitance of the ADC. Finally, the ADC digitizes the neural signals. A resolution of 8-10 bits is common, ensuring that the quantization noise is minimal and allowing for some variation in the amplitude of neural spikes.

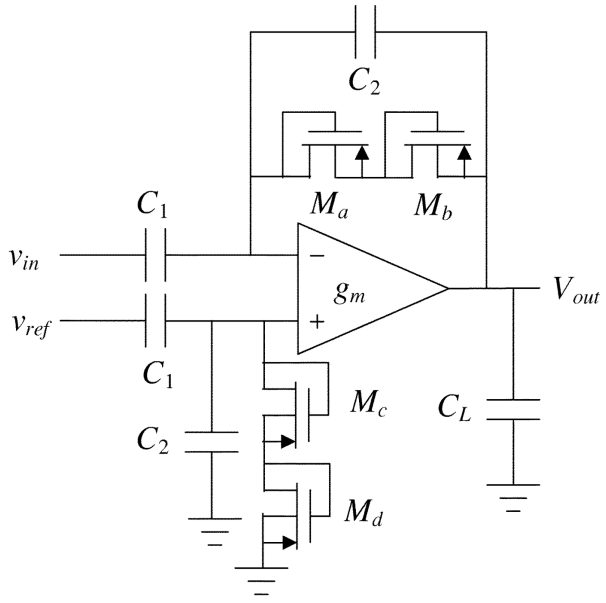


Figure 1.9: Widely-used amplifier topology proposed in 2002 by Harrison [48].

In 2002, Reid Harrison applied a technique to emulate a very large resistive impedance on-chip (10^{12} Ω) via MOS pseudo-resistors [48]. This permits small (10 pF) on-chip capacitors to be used for AC coupling of the LNA while maintaining a 1 Hz to 100 Hz high-pass corner, which is suitable for action potential recording [48]. Moving the AC coupling capacitors on-chip dramatically increases the number of channels that can be incorporated on an implantable recording system. This architecture, shown in Fig. 1.9, has been widely used by subsequent works, and efforts have focused on improving the noise efficiency of the active devices and the area consumption of the overall recording channel. As another example, there is a fundamental limit to noise efficiency, and therefore power and noise performance can be traded. This ultimately trades channel count against recording fidelity because recording systems operate on a fixed power budget. As a final example, mismatch grows as device area shrinks. This loss of robustness and yield can be recovered through complex and time-consuming calibration techniques.

Attempts to shrink area and improve area efficiency have exposed a number of tradeoffs, as illustrated in Fig. 1.10. For example, in 2012, Rikky Muller introduced a topology to cancel the DC electrode offset while removing the AC-coupling capacitors. This drastically reduced the size of the recording channel versus prior works. However, this area shrink came at an expense; the open-loop gain is unsuitable for some applications and the DC-coupled front end is considered unsafe for some medical applications.

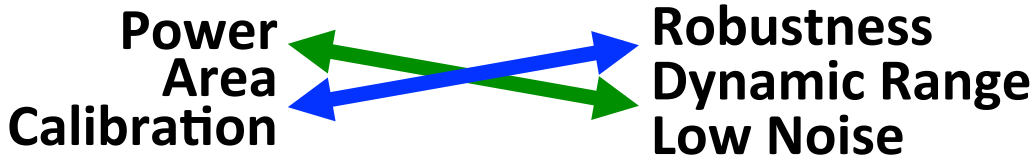


Figure 1.10: Design tradeoffs for a neural acquisition channel.

1.6 Efficiency Metrics for Neural Amplifiers

A longstanding metric used to evaluate how efficiently an amplifier uses its bias current to reduce noise is called the noise-efficiency factor (NEF) [49]. Eqn 1.1 defines the NEF, where k is Boltzmann's constant ($\approx 1.38 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}), T is the temperature in Kelvins (body temperature = 310 K), V_T is the thermal voltage (26.7 mV at body temperature), I_D is the transistor bias current, and BW is the -3 dB amplifier bandwidth. This metric compares an amplifier's noise and current consumption to a BJT amplifier with equal -3 dB bandwidth.

$$NEF = \sqrt{\frac{V_{ni,rms}^2 \cdot I_{total}}{V_{ni,rms,bjt}^2 \cdot I_{BJT}}} = V_{ni,rms}\sqrt{\frac{I_{total}}{4kT \cdot V_T \cdot \pi/2 \cdot BW}} \tag{1.1}$$

The minimum theoretical input-referred noise, $V_{ni,rms}$ of a differential CMOS amplifier has been analyzed by [50] and is given by Eqn. 1.2, where K is a process constant representing the subthreshold gate coupling coefficient (K $\approx$ 0.7). This leads to a minimum theoretical NEF of approximately 2.0, as given by Eqn. 1.3.

$$V_{ni,rms} = \sqrt{\frac{4kT \cdot V_T}{k^2 \cdot I_D}\frac{\pi}{2}BW} \tag{1.2}$$

$$NEF_{ideal} = \frac{\sqrt{2}}{K} \approx 2.0\Big|_{K=0.7} \tag{1.3}$$

The NEF metric has several limitations. First, insufficient and excessive bandwidths are not penalized. Neural signals contain a finite signal bandwidth. Insufficient amplifier bandwidth distorts and attenuates the neural signal, and excessive amplifier bandwidth adds noise. Second, the spectral characteristics of the noise are not considered in NEF. Flicker noise tends impact the neural signal band, and wide amplifiers bandwidths can make the overall NEF appear attractive while in-band noise performance is poor. Third, NEF compares current efficiency, not power efficiency. The initial reasoning was that noise, to first order, is not a function of supply voltage. However, an amplifier requires **power**, not current, to operate. The power efficiency factor (PEF) attempts to correct for this by comparing noise to a BJT amplifier operating at the same supply voltage and bias current level [51]. The equation for PEF is given by Eqn 1.4, where V_{DD} is the supply voltage.

$$PEF = \frac{V_{ni,rms}^2 \cdot V_{DD} I_{total}}{V_{ni,rms,bjt}^2 \cdot V_{DD} I_{BJT}} = \frac{V_{ni,rms}^2 \cdot V_{DD} I_{total}}{4kT \cdot V_T \cdot \pi/2 \cdot BW} = NEF^2 \cdot V_{DD} \tag{1.4}$$

Thus, the PEF captures differences in power efficiency between two amplifiers. For example, an ideal switching regulator would provide double the battery life for a 1.0 V amplifier as compared to a 2.0 V amplifier, if they each consume the same current from their respective supplies. Both the NEF and PEF metrics will be presented through this dissertation to illustrate how different circuit design techniques impact these two metrics.

It is important to consider the practical limitations of the PEF metric. A linear regulator is often required to reduce noise from a switching regulator, and the switching regulator conversion efficiency varies as a function of the conversion ratio. In other words, it may be difficult to directly compare a 1.2 V recording channel to a 1.0 V recording channel until the power supply is designed. Battery life is an excellent way to normalize the otherwise difficult-to-compare efficiencies and inefficiencies of varying approaches.

1.7 Stimulation

Another significant aspect of a neural interface the restoration of sensory feedback. Methods include electrical and optogenetic stimulation, where stimulation refers to the ability to excite or suppress the firing rate of nearby neurons. Electrical stimulation involves injection of a controlled amount of charge into the brain through an electrode, typically at a constant rate of current. Then, to prevent permanent tissue damage, the charge is removed. The charge and discharge are typically performed at the same rate of current flow. Optogenetic stimulation involves DNA modification, typically by virus, to make neurons sensitive to light. This allows fibre optics or micro-LEDs on an implanted probe to stimulate neurons via illumination. For more information on charge-based and ontogenetic stimulation, see [52] and [53], respectively.

1.8 Outline

Chapter 2 surveys prior works on neural amplifiers, data converters, and recording systems. A tremendous amount of work has been done on individual components, generally ignoring the integration challenges of very highly scaled system implementations. Yet, some full systems do exist, and their performance is discussed.

Next, this document details the design and measurement results from two neural interface implementations. First, Chapter 3 presents a tiny (0.125 mm^2) four-channel neural sensor with bidirectional wireless communication, which opens the possibility of free-floating neural nodes in the brain tissue. For reference, this system is smaller than most single neural amplifiers found in the literature. By eliminating wires to the sensor, applied mechanical forces due to wires are also eliminated, which originate from motion of the brain tissue

relative to the dura and skull. The end goal is to reduce the brain's biological response in which insulating scar tissue forms around electrodes and consequently degrades recording SNR.

Second, Chapter 4 presents a system to bridge the gap between long-term neuroscience research and today's biomedical needs by integrating 64 recording channels, 16 stimulation channels, and neural data compression onto a single 4.78 mm^2 IC. The low power and area of this chip can enable neural recording systems that scale up to thousands of channels, or scale down to extremely compact, low weight, low area, wireless interfaces. The IC achieves a level of integration not previously demonstrated in the literature or in commercial designs.

Chapter 2

Prior Work

An comprehensive technical review of integrated circuits for neural recording can be found in [45], including a history of the field, neural probes, and recording systems. Rather than provide a broad, comprehensive review, this chapter aims to highlight the forefront of the field and the benchmark for contributions.

2.1 Low Noise Amplifiers

Table 2.1 presents a summary of state-of-the-art neural amplifier designs where leading metrics are highlighted in bold. For a review of metrics to compare neural amplifiers, see Section 1.6. Rai [54] and Wattanapanitch [55] achieve excellent NEF, while Liew [56] achieves excellent PEF. Rai and Liew achieve low NEF through current-reuse topologies, while Wattanapanitch utilizes degeneration and current scaling in the transconductance vs. load devices to reduce noise.

Muller [57] achieves a very compact design, consuming 5-10x less area than competing designs. This low area comes at the cost of open-loop gain and a DC-coupled front end, which limits use in some applications. In particular, stimulation can induce voltages that may damage the recording IC. Also, safety precautions for medical devices often necessitate AC coupling to prevent large DC voltages from reaching the body in the case of a faulty or damaged recording IC [45]. Furthermore, use of active load devices cause poor noise performance. Several notable techniques are introduced: a VCO-based ADC both filters and digitizes the signal, and mixed-signal feedback improves dynamic range by canceling the DC electrode offset and LFP. An improved design is presented in Chapter 3 that improves noise efficiency while maintaining low area usage.

Because flicker noise decreases with frequency, it plays a less significant role as the amplifier bandwidth increases. Wattanapanitch [50] is significant for simultaneously achieving a low NEF and a low bandwidth, effectively mitigating flicker noise contributions. However, strict matching requirements incur an area penalty, and a large number of stacked devices results in a high supply voltage. In [55], both the supply voltage and area are significantly

Table 2.1: Prior work on low noise neural amplifiers

Author	Muller	Wattanapanitch	Rai	Liew
Reference	[57]	[50]	[54]	[56]
Year	2012	2007	2009	2011
Area (mm^2)	**0.013**	0.16	0.4	0.073
Process (nm)	65	500	130	130
Supply (V)	0.5	2.8	1.0	0.5
Power (μW)	5.04	7.56	12.5	0.86
HP (kHz)	0.3	0.045	0.023	0.3
LP (kHz)	10	**5.32**	11.5	7.5
BW (kHz)	9.7	5.275	11.5	7.2
Noise (μVrms)	4.9	3.06	1.95	5.32
NEF	6.09	**2.67**	**2.48**	3.17
PEF	18.55	19.95	6.14	**5.03**

improved, but NEF degrades to 4.4 and PEF degrades to 34.5. Improvements to this design are presented in Chapter 4, resulting in area, NEF, PEF and supply voltage improvements.

2.2 Analog to Digital Converters

A successive approximation register (SAR) ADC is a popular choice for neural acquisition channels due to their potential for low power and area. A resolution between 8 and 10 bits is often chosen to provide 50 to 60 dB of nominal dynamic range. The effective number of bits (ENOB) is often 1-2 bits below the resolution due to mismatch and noise, which results in about 40 to 50 dB of real dynamic range. ADC performance is often quantified by a figure of merit (FOM) equal to $Power/(2^{ENOB} f_s)$, which represents the average energy consumed per bit of the ADC conversion.

The ADC, more than any other component in the signal chain, places stringent constraints on the preceding buffer. This is due to the setting requirements of the sampling capacitance of the ADC. A large sampling capacitor can demand substantial power consumption in the buffer to settle within 1 LSB during each sampling window. Publications of stand-alone ADCs with record-breaking performance have no impact if they place crippling demands on the buffer. Finally, because the ADC itself tends to consume a relatively small fraction of system power, it is often optimized only to the point where the LNA and other components dominate. Some ADCs also require a precision $V_{DD}/2$ reference, which can then necessitate a second low impedance (high power consumption) buffer.

One example of a highly-optimized SAR can be found in [58]. Input capacitance is kept to 250 fF at 10 bit resolution (9.4 bit ENOB), supply voltage is a reasonable 1.1 V, and the design achieves a FOM of 6.5 fJ per conversion step. To digitize one neural channel, the

ADC consumes 100 nW at 20 kHz. Other components of the recording channel, such as the LNA, render this ADC power negligible. For low-voltage system applications, a follow-up paper achieved 2.2 fJ / step FOM performance at 600 mV supply voltage [59]. For neural recording applications, no further optimization is currently required.

2.3 Neural Interface Systems

A wide variety of neural interface systems have appeared in the literature. Table 2.2 presents a survey those systems leading in integration and/or performance. Some include an active or passive (backscatter, see Chapter 3) radios ("Radio"), and some include a wireless power delivery system ("RF Pwr."). The radio provides the data link through the scalp and/or skull. Wireless power allows long term operation of an implanted system and is often delivered by magnetic coupling, as described in Section 1.2.

There are many interesting points of comparison in Table 2.2. Despite two systems achieving excellent LNA NEF (< 3) and others achieving poor LNA NEF (> 5), the power consumption for full recording channels is relatively constant. All of the systems consume between 23 to 35 μW per recording channel. It is unclear if this reflects a focus on the LNA FOM for the sake of publication or a struggle with the system integration challenges. Regardless, it has now been demonstrated that a full neural acquisition channel can be implemented in 1.1 μW [56]. The next wave of system builders must meet these performance levels in order to achieve the channel count demanded by neuroscientists.

Chae [60] demonstrates 128 recording channels integrated with an on-chip high-data-rate radio that can transmit the raw stream from all channels simultaneously. On-chip DSP provides spike detection and feature extraction (min / max) for one selected channel. Application of the DSP to all channels would triple the total power budget, at 100 μW per channel. Lack of a power delivery mechanism prohibits long term use as an implanted system. Finally, wireless range is not reported. Nevertheless, the system is noteworthy for successfully recording and transmitting 128 channels wirelessly with a 6 mW power budget.

Azin [18] and Rhew [61] successfully integrate recording, compression, and stimulation. Chen [19] utilizes a system-in-package approach to connect a 16-channel recording IC to a combined 16-channel DSP and 8-channel stimulation IC; quoted area and power are for the DSP / stimulation IC only. The recording channel count of these systems are insufficient for all but the most rudimentary prosthesis control. Lastly, the system by Rhew requires a 5 V battery for stimulation, which is not recharged by the wireless power delivery system. Thus, none of these systems are implantable.

Lee [62], Sodagar [28, 63], and Harrison [64] present recording systems without stimulation. Wireless data and power transmission are included, allowing long term implantation. Lee relies on direct transmission of the raw recordings, like Azin. Therefore this system cannot scale in channel count. The systems by Sodagar and Harrison provide spike detection and packetization of compressed spike events. At 75 μW and 35 μW and roughly 0.1 mm^2 per recording channel, significant reductions in power and area are required to scale up channels

counts. Furthermore, provisions for stimulation are necessary to provide sensory feedback in BMI applications.

In summary, levels of functionality and area consumption vary greatly between published systems. Total chip area is likely set by the silicon sponsor's generosity and not necessarily by economic or application constraints. Some systems lack data compression and/or stimulation, and the compression loss and data formats vary widely. No systems offer a firing-rate option, which compresses the data rate to the level where 1000's of channels could be wirelessly recorded. As the field of neural engineering matures, a more standard feature set, driven by proven application requirements, will emerge.

Table 2.2: Survey of Neural Recording and Stimulation Systems

System Specs.	Chae [60]	Azin [18]	Rhew [61]	Chen [19]	Lee [62]	Sodagar [28]	Harrison [64]	Units
System								
Technology	350	350	180	350	500	500	500	nm
VDD	3.3	1.5	1.8	5.0	3.0	1.8	3.3	V
Off-Chip	Ant	1 µF	DC-DC	Amp	Ant	Ant	Ant	
			Bat	ADC	Cap	Cap	Cap	
Power	6	0.375	0.468	10.46	5.85	14.4	8	mW
Area	63.4	10.9	4	28.3	16.2	217	25.4	mm^2
Radio	Yes	Yes	Yes	No	Yes	Yes	Yes	
RF Pwr.	No	No	Yes	No	Yes	Yes	Yes	
Amp/ADC[a]								
Channels	128	8	4	16	32	64	100	
Power	23.4	25.8	61.25	NA	25	75	35.2	µW
Area	0.039	0.3122	0.354	NA	0.162	0.072	0.16	mm^2
Gain	57-60	51-65.6	54	NA	66-78	59.7	60	dB
HP	0.1-100	1-525	700	NA	1	0.1-100	250	Hz
LP	2-20	5-12	6	NA	10	9.1	5	kHz
Noise	4.9	3.12	NR	NA	4.95	8.0	5.1	µVrms
NEF	2.7	2.9	NR	NA	5.6	21	9.3	
PEF	24.2	12.6	NR	NA	93	973	286	
Stimulation[a]								
Channels	0	8	8	8	0	0	0	
I_{max}	-	0.0945	2x 4.2	6.25	-	-	-	mA
			6x 0.116		-	-	-	
Area	-	0.038	0.05	0.7	-	-	-	mm^2
Compression[a]								
Power	100	3.28	34.5	256.9	-	NR	17.6	µW
Area	NR	0.0676	0.8	0.191	-	NR	NR	mm^2
Outputs								
Raw ADC	No	1 Ch.	Yes	Yes	Yes	1 Ch.	1 Ch.	
Epochs	Yes	No	Yes	Yes	No	No	No	
Events	No	Yes	No	Yes	No	Yes	Yes	
Firing Rates	No	No	No	No	No	No	No	
Other	P-P	-	LFP	PCA	-	-	-	

[a] Amp/ADC, Stimulation, and Compression power are given per recording / stimulation channel.

Chapter 3

A Fully-Integrated Wireless Neural Sensor

3.1 Introduction

To date, the direct recording of action potentials (APs) is the only type of brain-machine interface (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 signal-to-noise ratio (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 [65]. 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. [62, 28, 41]) 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, a system-on-chip (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 systems. This enables a fully-integrated wireless SoC without the use of any off-chip components.

The proposed system (Fig. 3.1) 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

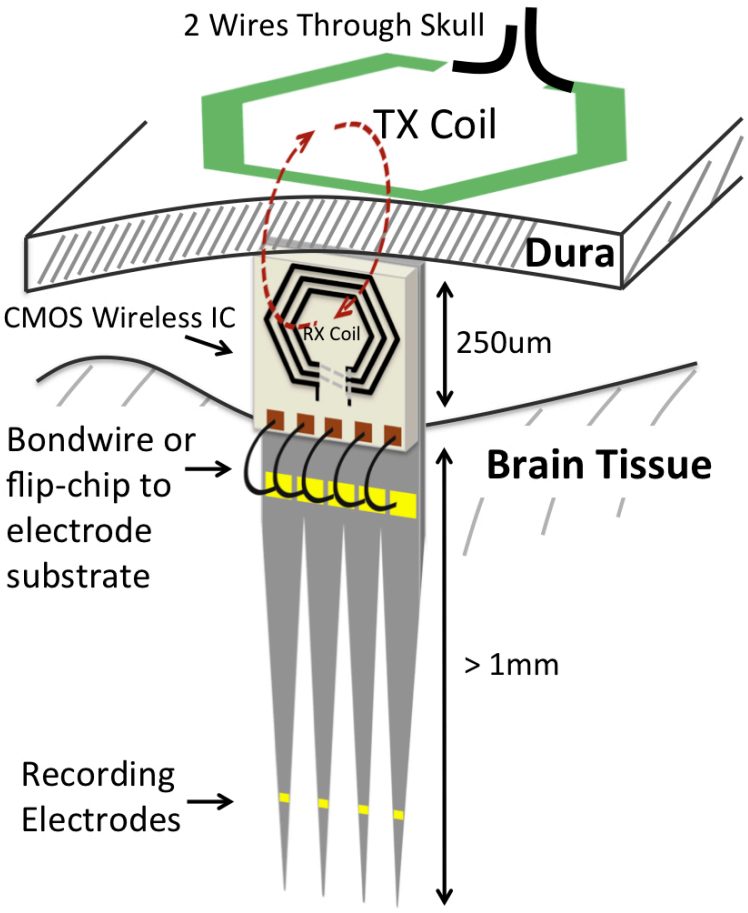


Figure 3.1: A conceptual diagram of the implementation of a BMI utilizing the developed wireless neural sensor. The sensor free-floats under the dura, while receiving power from and communicating to an interrogator beneath the skull.

data acquisition channels amplify and digitize the sensed neural potentials into an 800 kbps data stream via 10b, 20 kHz ADCs. A single receive (RX) coil on the 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.

This chapter was originally published as [66] and [67]. Broadly, the author's contributions focused on the RF front end, communication protocol, digital logic, and acquisition channel amplifiers. The first coauthor, William Biederman, focused on the wireless link, antenna design, system integration, as well as several circuit blocks including the bandgap reference, acquisition channel ADC, and SC current reference. The third coauthor, Nathan Narevsky, also contributed to system integration, verification of the acquisition channel mixed-signal feedback loop, and software for signal processing of the communication channel.

3.2 System Design

Realization of the system illustrated in Fig. 3.1 is limited by the available power at the node and the communication protocol data rate. Power delivered to the node is maximized by careful selection of the the wireless transmission frequency (Section 3.2.1). To enable robust multi-node communication while providing a low-overhead reference clock to the nodes, the communication protocol is optimized for this application (Section 3.2.2). Finally, the system architecture (Section 3.2.3) outlines the co-operation of circuit blocks, which ultimately enables low-noise neural potential recordings under a highly constrained power budget.

3.2.1 Frequency Selection

The maximum available power for a given node size is determined by the transmission distance and the frequency of operation. The minimum transmission distance for this system is determined by the thickness of the dura above the primary motor cortex (M1), which has a $\mu + 3\sigma$ thickness measuring 0.61 mm in humans [68]. In biological media, operating at a frequency between 1-3 GHz minimizes channel loss for edge-to-edge coupling [69] and reduces the RX coil size by several orders of magnitude compared to [62, 28, 41]. Thus, the transmission frequency for this system was selected to be 1.5 GHz, trading a reduction in node size and channel loss for an increase in the specific absorption rate (SAR). Based on simulations, the estimated channel loss through 0.6 mm of brain tissue is approximately 20 dB, which correlates well to measurements in air as discussed in Section 3.6.1.

3.2.2 Communication Protocol

The proposed system enables a single interrogator to wirelessly power multiple implanted nodes. However, each node generates 800 kbps 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, we propose simultaneous transmission by all nodes in unique frequency bands. For this 5-node system, each node's backscatter is Miller-encoded at a programmable subcarrier frequency between 2 MHz and 10 MHz. Fig. 3.2 shows conceptual time domain waveforms of 5 wireless packets with this system's possible subcarrier frequencies (2, 4, 6, 8, 10 MHz). Fig. 3.3 shows the frequency spectrum of 5 nodes transmitting simultaneously. Finally, Fig. 3.3 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 50 µs. The nodes recover this 20 kHz clock, which initiates the ADC conversions

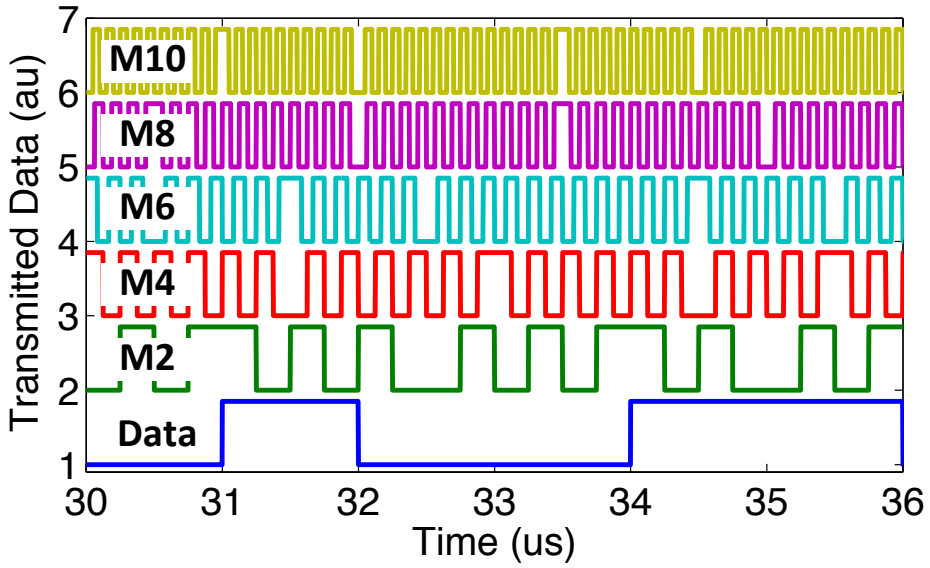


Figure 3.2: Theory of a Miller encoded communication scheme for multi-node interrogation showing miller encoded waveforms (2,4,6,8,10 MHz) for a data set.

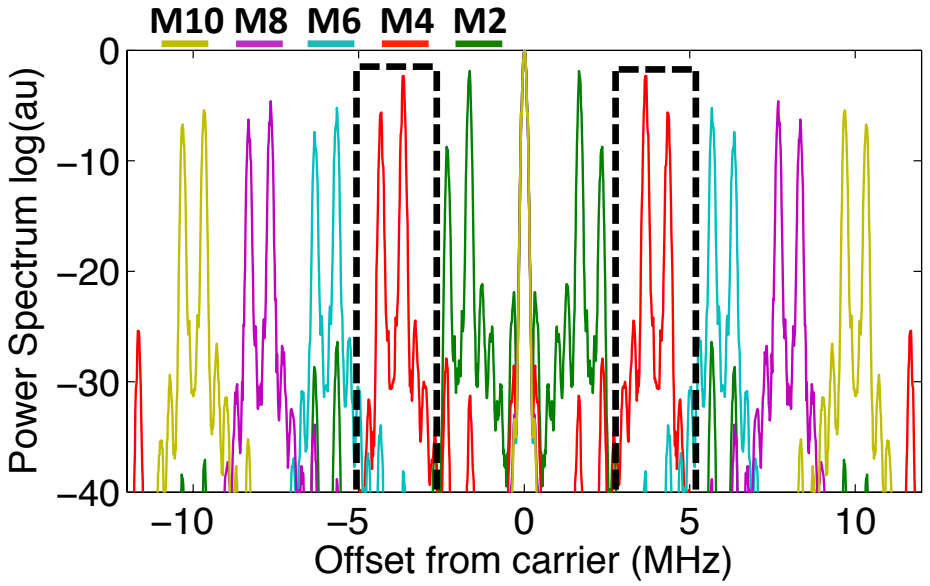


Figure 3.3: Theory of a Miller encoded communication scheme for multi-node interrogation showing the resulting frequency spectrum from 5 nodes communicating simultaneously.

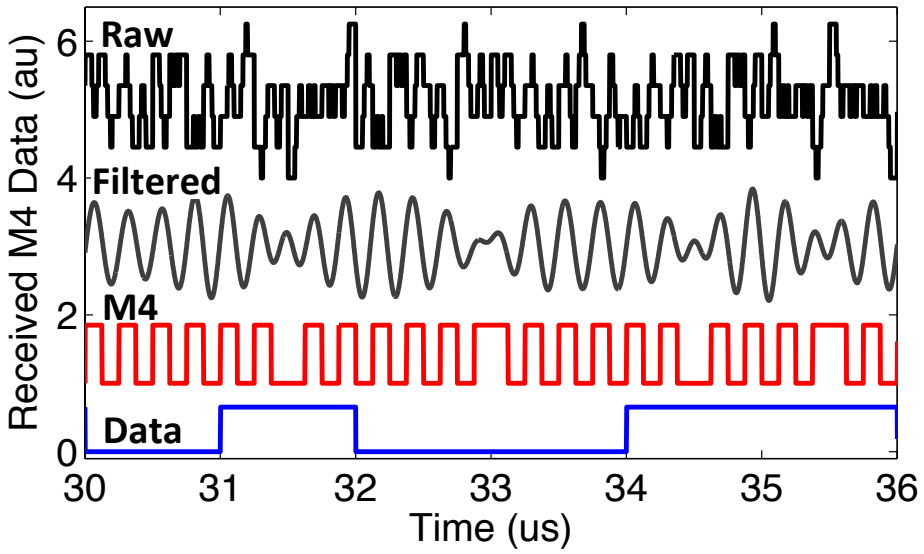


Figure 3.4: Theory of a Miller encoded communication scheme for multi-node interrogation showing the recovered raw waveform before and after bandpass filtering, and the recovered original transmitted M4 signal and equivalent data.

of neural potentials as well as communication of the 40-bit data packets containing the ADC output. The 2-10 MHz Miller subcarrier clock is generated by a frequency-locked loop (FLL), which locks to a multiple of the 20 kHz beacons. The circuit details are presented in Section 3.3.2.

To initiate downlink communication, the interrogator sends two consecutive beacons, followed by PPM data. The encoding format is similar to EPC Gen2 RFID [70]. 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.

3.2.3 System Architecture

The system architecture, shown in Fig. 3.5, is ultimately determined by the specifications of the neural potential recording channels as well as the limited area and power available for circuit design. The high data rate and need for a precise clock necessitate an interrogator-provided time base. Section 3.3 presents the 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. Section 3.4 describes the inductive link optimization as well as the rectifier, regulator, and bias generation. Lastly, the wide dynamic range of neural potentials necessitates an aggressive mixed-signal topology to achieve low input referred noise.

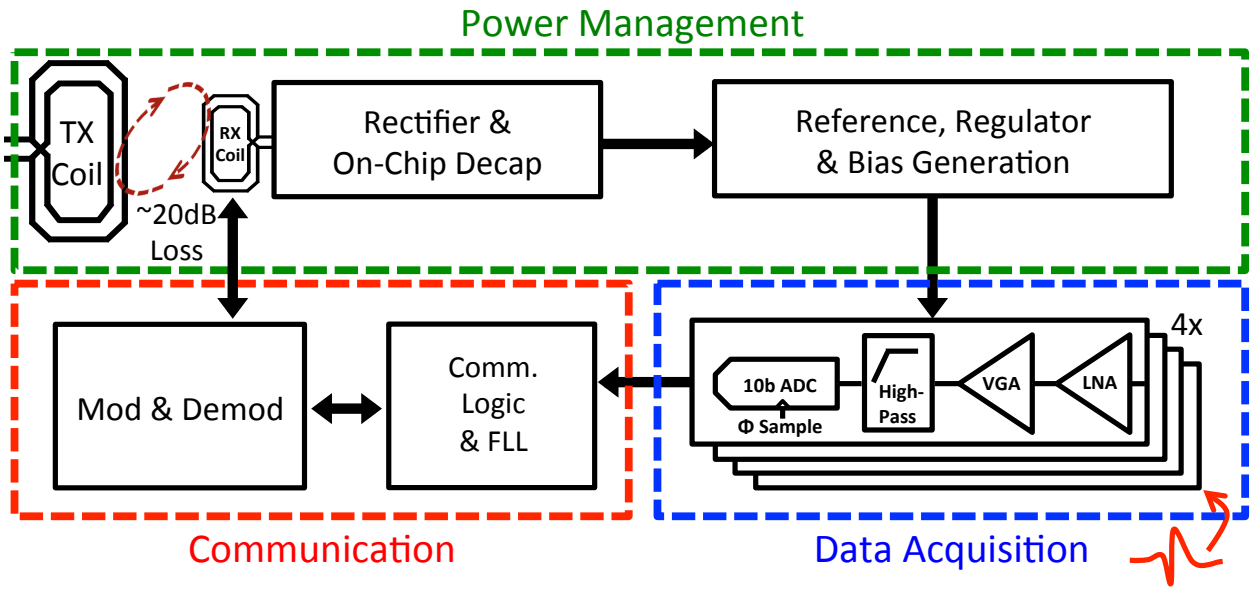


Figure 3.5: System diagram, subdivided into three primary circuit blocks: Power management, communication circuitry, and data acquisition.

Section 3.5 details the design as well as the measurement results of the data acquisition components. Section 3.6 presents the testing results of the fully integrated neural node and validates the system functionality *in vivo*. Finally, Section 3.7 compares these results to prior work.

3.3 Communication Circuitry

The communication circuitry must facilitate high date rate, multi-node communication under an extremely constrained power budget, without the use of an off-chip crystal reference. Two key circuit blocks enable the system's communication protocol: the demodulator and FLL. The proposed demodulator, described in Section 3.3.1, must recover low-duty cycle pulses, which is often impractical in conventional designs. This permits a low overhead clock recovery scheme, which provides a reference clock for the FLL. The FLL, which is described in Section 3.3.2, is then responsible for producing a Miller subcarrier modulation clock with minimal power overhead.

3.3.1 Demodulator

Conventional RFID demodulators use a dedicated rectifier to track the RF envelope, which is then low-pass filtered to generate the mean RF value. The original recovered RF envelope and the low-pass filtered output are used as comparator inputs to perform data slicing. In this system, it is desirable for the interrogator to send low (∼1%) duty cycle timing beacons

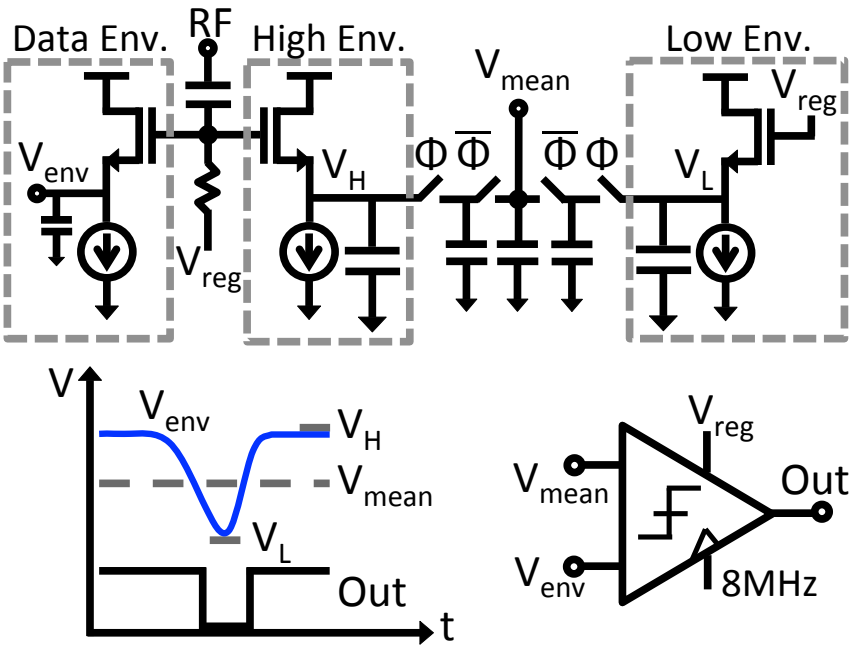


Figure 3.6: Top: The implemented demodulator schematic utilizes two peak detectors and a replica detector to extract V_H, V_{env}, and V_L, which represent the high, instantaneous, and low envelopes, respectively. A switched capacitor filter calculates V_{mean}, the mean of the high and low envelopes. Bottom: A conceptual waveform showing the high and low envelopes which are recovered from the peak detector/replica and are averaged to generate slicing threshold for the sampling comparator.

to minimize supply ripple and protocol overhead. Therefore, the low-pass filter used in a traditional demodulation scheme would provide a poor reference voltage for data slicing. Furthermore, the use of a dedicated rectifier for envelope detection loads the RF input and thereby decreases system efficiency.

Ideally, the data slicing threshold should be set equal to the mean of the high and low recovered RF envelopes, as shown in the graph in Fig. 3.6. In this work, peak detectors are used to recover the RF envelope (Fig. 3.6) instead of a dedicated rectifier. Because they do not significantly load the RF input, two separate peak detectors can be used to recover the instantaneous data envelope independently from the slicing reference. By decoupling the short decay time constant necessary to detect beacons from the long time constant necessary to preserve the slicing threshold during modulation, a highly asymmetric duty cycle can be used.

Fig. 3.6 shows the schematic of the proposed peak detector-based demodulation scheme. The data envelope peak detector has a small capacitor at its output, V_{env}, which enables tracking of short 250 ns OOK pulses. Conversely, the high envelope peak detector output, V_H, has a large output capacitance to retain the maximum envelope voltage while data modulation occurs. The low envelope, V_L, is generated by a replica circuit, and a switched

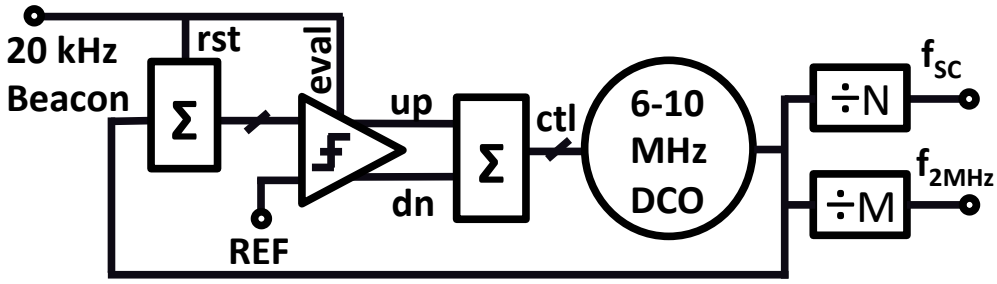


Figure 3.7: Frequency-locked loop for subcarrier generation. Recovered beacons reset the accumulator (rst) and evaluate the comparator ($eval$). The comparator increments or decrements the second accumulator (up/dn), which in turn controls the DCO tuning bits (ctl). Finally, programmable dividers generate the output Miller subcarrier (f_{sc}) and internal 2 MHz clock (f_{2MHz}).

capacitor network generates the mean voltage, V_{mean}, of V_H and V_L. Finally, a clocked comparator generates the decoded digital output by comparing V_{env} to the slicing threshold, V_{mean}.

3.3.2 Frequency Synthesis with FLL

Fig. 3.3 shows the spectrum of the communication channels, which are spaced apart by 2 MHz. If the center frequencies of the channels drift by more than approximately 10%, the spectrum from adjacent channels will begin to overlap and cause communication errors. Generating a sub-microwatt clock with better than 10% accuracy across mismatch and process corners typically requires trimming, which is undesirable in a tiny IC with no room for engineering pads or large metal lines for trimming. Furthermore, in this system, as well as in RFID tags and other battery-less sensors, a crystal reference is prohibitively bulky.

Many RFID tags divide down an uncertain local clock using a programmable clock divider. The divider value is calculated from reference timing that is communicated by the reader [70]. This incurs very little power overhead because of the low clock frequencies and data rates required in typical RFID applications. In this system, the nearly 1 Mbps data rate necessitates Miller subcarrier frequencies at or above 1 MHz, which would incur a large power overhead when synthesized using clock division. For example, in order to reduce the maximum clock divider residual error to 10% for a 10 MHz subcarrier, an input clock frequency of 50 MHz is required. Some systems have proposed injection locking to the received RF carrier [71]. However, the circuitry required to recover and divide down a gigahertz clock incurs a substantial power overhead of 14 μA in [71].

In this work, we use an FLL to generate Miller subcarriers ranging from 6-10 MHz, and lower subcarriers ranging from 2-4 MHz are generated via fixed, low-power dividers. The FLL reference is provided by the interrogator, which can afford space for a precision frequency reference. The schematic of the FLL is shown in Fig. 3.7. The 20 kHz beacons, sent by the

interrogator and recovered by the demodulator, provide a reference clock. A digital counter measures the DCO frequency by counting the number of DCO periods in each reference period. The comparison reference (REF) is set to the ratio f_{DCO}/ 20 kHz. A 2 MHz clock is also generated for the switched-capacitor (SC) current reference and for the $\Sigma\Delta$ DAC in the data acquisition channels. Finally, an up/down signal updates the DCO control bits stored in a 5-bit accumulator. This provides approximately 250 kHz resolution and sufficient range to span 6-10 MHz over process corners. This results in a maximum residual error of 4.2% at 6 Mhz.

Because the modulation switch creates input amplitude variations that are indistinguishable from downlink modulation, the node cannot both talk and listen without prohibitively complex and power-hungry circuitry. Thus, if the FLL frequency is initially set slightly lower than the target communication rate, the node will talk over every second beacon. The FLL will slow down the clock until it has locked at half of the target frequency because it has no way to detect that it is skipping beacons while it is talking. Therefore, to prevent errant locking, the node does not transmit until the FLL detects a lock condition. The communication interval in which the node transmits cannot drift because transmissions are initiated by the beacons, which effectively phase-locks the node to the beacons. Finally, the lock signal is reset when the node needs to change its Miller subcarrier frequency, since the FLL must change its output frequency.

3.4 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 3.4.1 describes the co-optimization of the antenna coil and the rectifier, which convert the incident RF power into an unregulated DC supply. Bias generation is discussed in Section 3.4.2, including a basic bias source for the other power management blocks as well as a precision bias generator for the data acquisition channels. The voltage reference and regulator, described in Section 3.4.3, provide a stable 500 mV supply for the digital core and data acquisition channels. Finally, the power-on reset circuit is used to sequence start-up and is described in Section 3.4.4.

The rectifier, regulator, $\Delta V_{GS}/R$ current reference, power-on-reset were designed by the author. The wireless link, inductive coils, bandgap reference, and SC bias current reference were designed by a coauthor, William Biederman.

3.4.1 Wireless Power Transfer Optimization

A carefully optimized wireless power link minimizes the required amount of transmit power, reducing tissue heating and power consumption of the interrogator. Eqn. 3.1 approximates the power transfer efficiency, η, where Q' represents the loaded quality factor, Q, of the the transmit (T) and receive (R) inductors [72].

$$\eta = k^2 Q'_T Q'_R \tag{3.1}$$

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. 3.2, 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. 3.3. 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.

$$V_{oc} = \sqrt{\eta P_{tx} R_p} \tag{3.2}$$

$$R_p = \omega L Q \tag{3.3}$$

The rectifier is designed to source 10.5 μW (15 μA at 700 mV) and 120 pF of output capacitance reduces supply ripple during communication. A two-stage self-synchronous rectifier topology, shown in Fig. 3.8, 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 500 μm x 250 μm of area in the top metal layers above other circuits and achieves a quality factor and inductance of approximately 8 and 18 nH, respectively. The resulting R_p is 1.36 kΩ, yielding a simulated rectifier efficiency of 24%.

3.4.2 Bias Current Generation

The task of bias current generation is divided between two groups of circuits. The first group, including biases the regulator, DCO, and demodulator, requires an independent current source that is not dependent on the clock or regulator. A standard $\Delta V_{GS}/R$ current reference, powered from the unregulated supply, biases these circuits.

The second group, which includes 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. 3.9, forces 300 mV 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

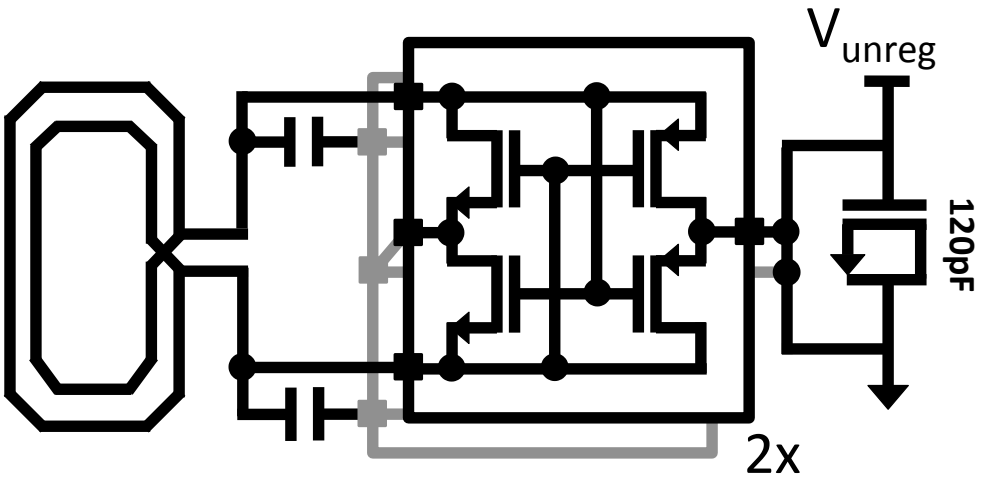


Figure 3.8: A high-R_p on-chip coil increases the open circuit voltage and maximizes the efficiency of the self-synchronous rectifier. 120 pF of on-chip decoupling capacitance is implemented with thick-oxide native devices.

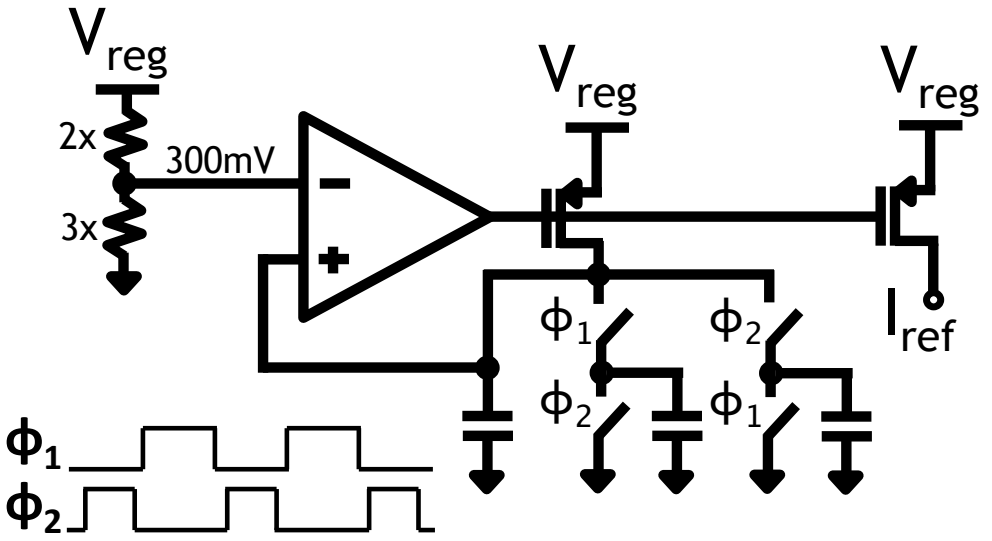


Figure 3.9: Switched-capacitor bias current generation schematic, utilizing two-phase non-overlapping clocks.

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 300 mV op-amp reference voltage is generated using a pseudo-resistor voltage divider from the regulated supply.

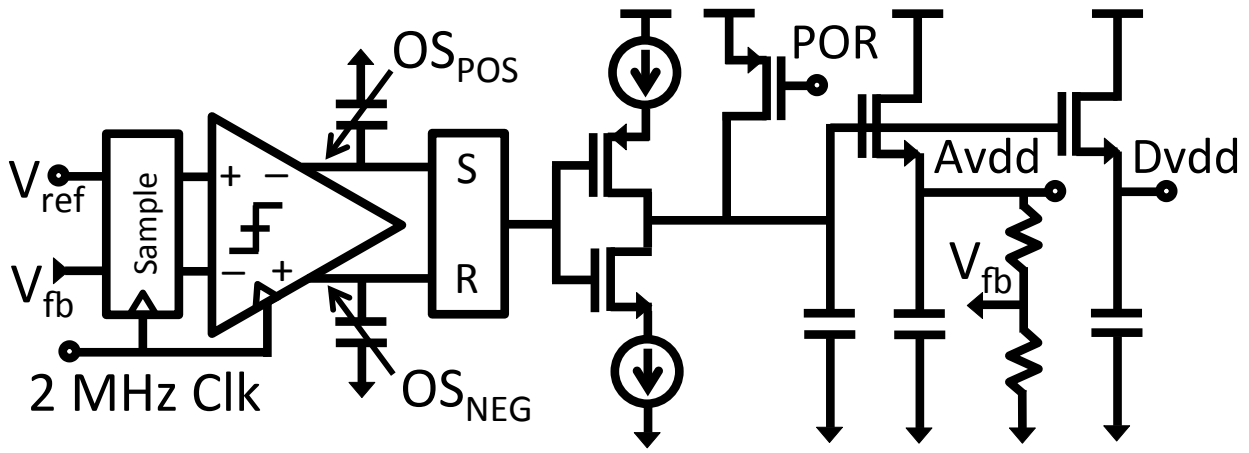


Figure 3.10: Discrete-time LDO regulator schematic utilizing a comparator with capacitive offset cancelation and a charge pump loop filter.

3.4.3 Voltage Regulation

Uplink and downlink backscatter communication induce unregulated supply ripple at the programmable subcarrier frequency ranging from 31.25 mV at 2 MHz to 6.25 mV at 10 MHz (assuming a 15 μA load on the 120 pF decoupling capacitor). A discrete time linear regulator, shown in Fig. 3.10, 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 strong-arm comparator [73] 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 300 nA at the maximum supply voltage and occupies 55 μm x 54 μm. Input and output capacitors, including the 120 pF decoupling capacitor for V_{unreg}, consume 450 μm x 63 μm. The measured PSRR across frequency is shown in Fig. 3.11. With a worst case PSRR of 27 dB, communication-induced supply ripple is reduced to less than 1.5 mV.

The regulator requires a robust precision voltage reference with low area and power consumption. By utilizing a SC bandgap architecture proposed in [74], 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, consuming only 138 nA of current and 0.0055 mm^2 of area. Furthermore, the design enables operation with an unregulated supply voltage as low as 650 mV at 37°C, which helps minimize the dropout voltage of the system.

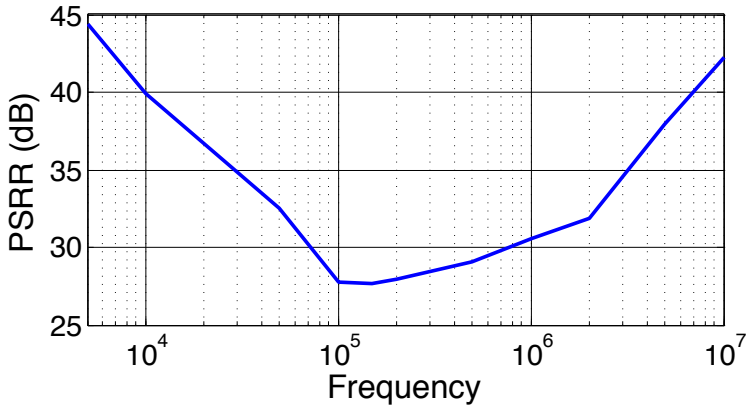


Figure 3.11: The measured discrete-time LDO regulator supply rejection across frequency.

3.4.4 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 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. 3.12, 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 10 nW in steady state (simulated) and occupies 225 μm^2.

3.5 Data Acquisition

As discussed in Section 1.4, a typical neural potential recording consists of three components: DC offset, LFP and APs. The 70 dB (13.5 μV to 50mV) of input dynamic range is typically reduced via AC-coupling, where the high-pass corner is set at approximately 500 Hz to eliminate the offset and LFP. Synthesizing this filter corner frequency on chip with passives can consume substantial die area, which scales poorly with the process node. Instead, this work utilizes a DC-coupled, mixed-signal data acquisition architecture, shown in Fig. 3.13

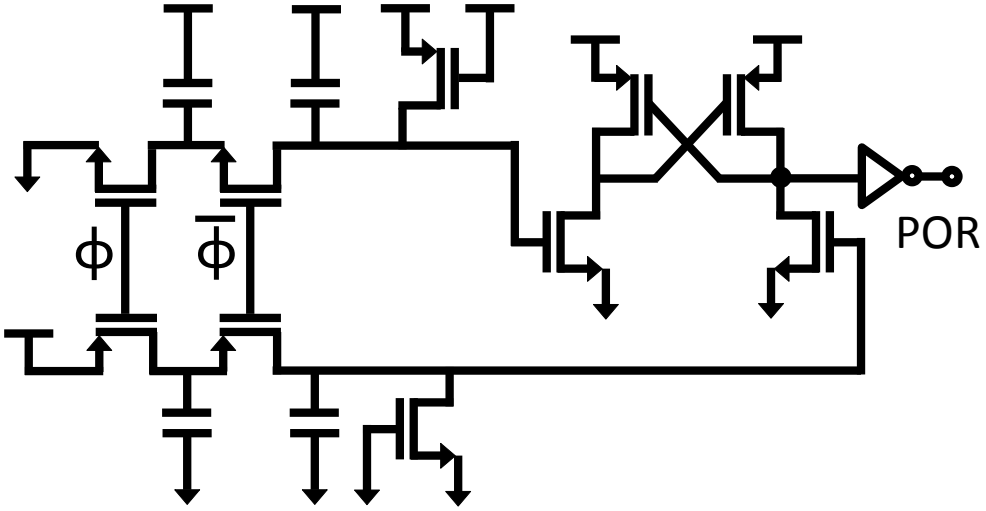


Figure 3.12: Power-on reset schematic. Switched-capacitor resistors pull up against pseudo-resistors, disabling the reset signal after several clock cycles.

and originally proposed in [57]. The offset and LFP are canceled in the analog domain to alleviate the dynamic range constraints, and filters are synthesized in the digital domain, thereby eliminating bulky passive components.

Also discussed in Section 1.4, the input signal from the electrode contains noise from both the biological background chatter and also the thermal noise of the electrode interface. This varies with electrode material and dimensions; we modeled it to be approximately 13.5 μV (based on Section 4.2 and Table I from [46]). Many neural amplifiers target noise floors as low as 1-3 μV (ex. [50, 75, 76]), significantly below this recording noise floor, which results in wasted power. Consequently, in this work we targeted a comparatively large amplifier input referred noise of 6-7 μV, allowing power minimization without significantly contributing to recording noise. The total recording and amplifier input referred noise is equal to the sum of their variances, shown in Eqn. 3.4. With an amplifier input referred noise of 6.5 μV, the total estimated recording input referred noise, σ_{Total}, is approximately 15 μV.

$$\sigma_{Total} = \sqrt{\sigma_{Amp}^2 + \sigma_{Therm}^2 + \sigma_{Bio}^2} \tag{3.4}$$

For applications requiring a lower noise floor, the same design procedure and circuit topology can be applied. As more power is spent in the LNA and subsequent amplifiers to reduce noise, the power overhead from the ADC and digital logic will constitute a lower fraction of the total power, and the noise efficiency of the system will improve.

The following subsections describe the design of the LNA, VGA, DAC, and ADC. Calibration and biasing techniques are also discussed, which ensure robust operation over process

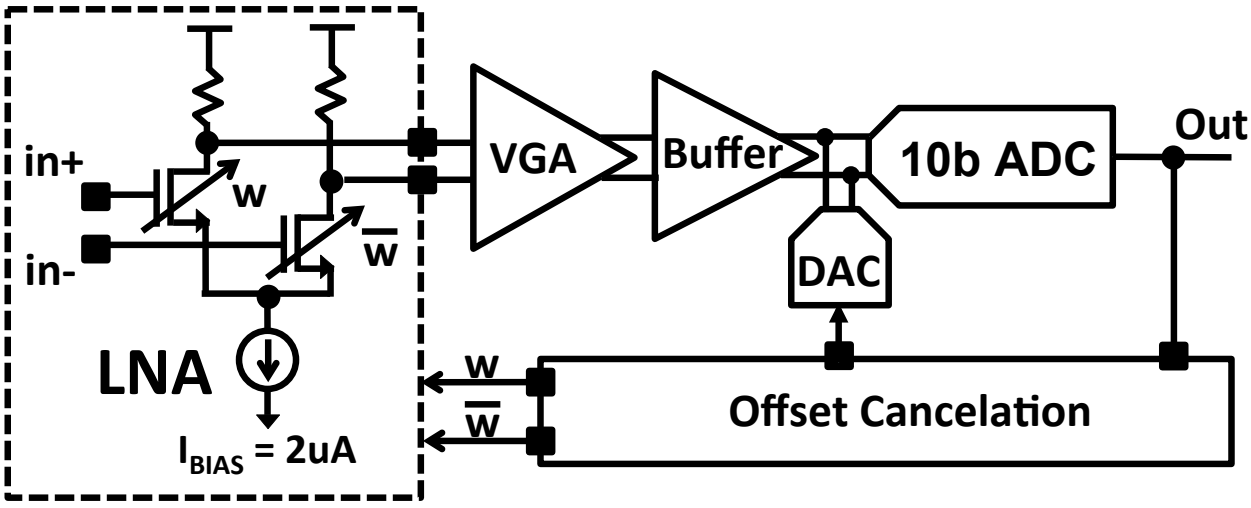


Figure 3.13: Amplifier block diagram utilizing mixed signal feedback to cancel DC offset and LFP signal.

corners at a 500 mV supply voltage. Finally, a performance summary is given. The ADC was designed by a coauthor, William Biederman. The remainder of this chapter was the work of the author.

3.5.1 Amplifier

The amplifier consists of three open-loop gain stages, shown in Fig. 3.13. The entire acquisition channel is powered from the 500 mV supply. The input capacitance of the LNA was limited to 250 fF differential / 1 pF common-mode to enable compatibility with a wide range of electrodes. The LNA and VGA utilize weak positive feedback to achieve maximum gains of 15 dB and 20 dB, respectively. The VGA gain is modulated by shunting the differential resistor load, allowing gain ranging from approximately 0-20 dB.

The LNA gain is sufficiently large that its noise dominates the total system noise, which is 6.5 μV_{rms}. The LNA consumes 2.2 μA and occupies 54 μm x 72 μm. The low noise floor, power, and area are enabled via the mixed signal architecture as well as the use of calibrated poly resistor loads. To prevent the resistors from dominating the chip area, trim and auto-calibration are used.

3.5.2 ADC and Mixed-Signal Feedback

A fully-differential VCO-based ADC digitizes the neural signal, and mixed-signal feedback removes the LFP signal via a current DAC and the offset via modulation of the LNA input device width [57]. The ADC topology provides a compact layout of 640 μm^2, low power

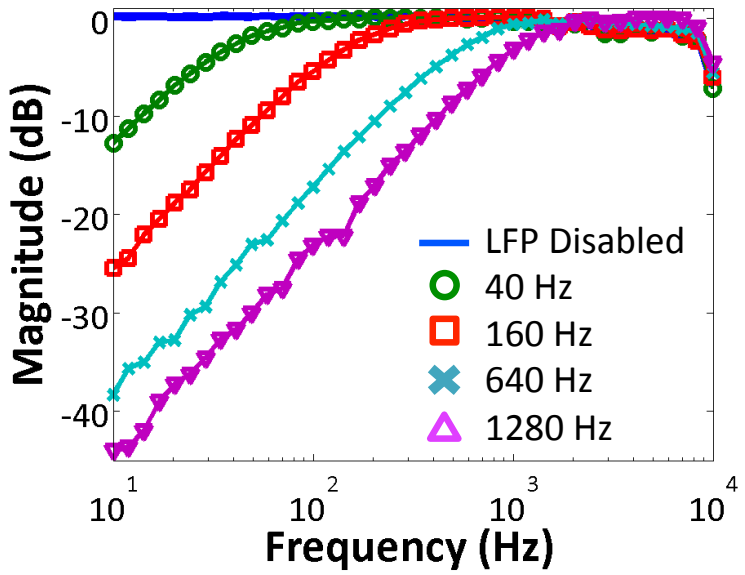


Figure 3.14: Measured amplifier frequency response over various high-pass corners, normalized to peak gain.

consumption of 290 nW including integrating counters, and obviates the need for an anti-aliasing filter due to its sinc transfer function.

A 10-bit ΣΔ-modulated current DAC cancels the LFP signal in the analog domain. A digital accumulator in the feedback path drives the LFP DAC, creating a high-pass corner. The corner frequency is wirelessly programmable by adjusting the feedback gain (Fig. 3.14). The input referred noise of the LFP DAC is well below the system noise floor.

The offset cancellation DAC is implemented by modulating the LNA input device width. This enables offset cancellation with a very minimal increase in input noise across the DAC range and eliminates the input AC coupling capacitors. An overflow condition in the LFP integrator is used to adjust the offset DAC, or a wireless programming command can manually set the control bits.

3.5.3 Calibration and Biasing

The combination of aggressive power minimization, low voltage operation, and process variation in 65nm necessitate compensating calibration circuitry. In this work, an active input biasing circuit for the LNA (Fig. 3.15) provides a precise headroom of 100 mV across the tail current source despite substantial transistor V_{TH} variations that could otherwise pinch off either the input devices or the tail device.

There is a tradeoff between the process variability of poly resistors and the resistor width, which translates to area. However, doubling the device width requires doubling the length, and thus quadrupling the area, in order to achieve the same resistance. Here, the 200 kΩ LNA

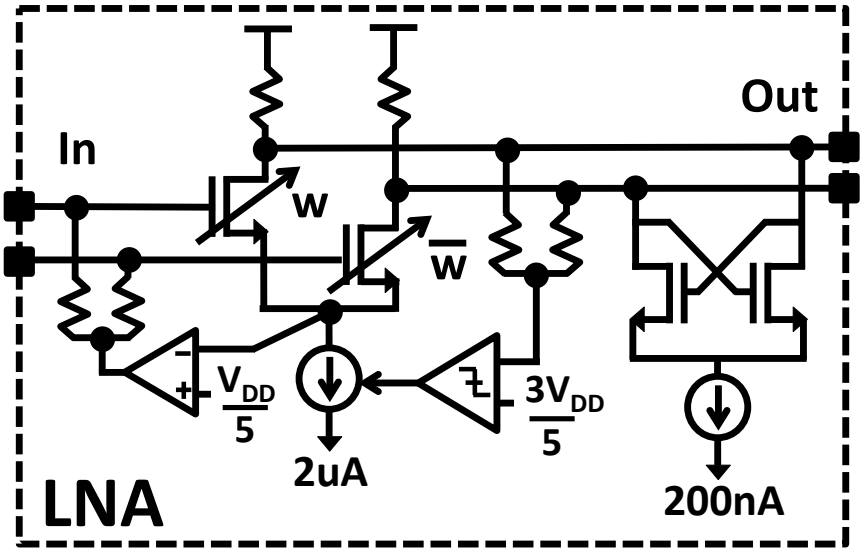


Figure 3.15: Amplifier front-end calibration, input and output common mode biasing, and cross-coupled gain boost.

load resistors are upsized in both length and width to keep variation to a manageable range. However, mega-ohm resistors in the VGA and DAC utilize near-minimum poly widths and consequently require trimming. The bias and resistor calibration takes place in two discrete steps. First, the global amplifier bias current is adjusted to produce a 300 mV LNA output common mode. Then, a global resistor calibration circuit forces an output common mode of 300 mV on a replica amplifier stage.

3.5.4 Performance

At the highest gain setting, the data acquisition channels provide a 1 μV LSB and achieve an input referred noise of 6.5 μV_{rms}. The input referred noise was measured while wirelessly powered and using all on-chip voltage references, regulator, and bias currents. The measured noise spectrum with and without the LFP cancelation loop is shown in Fig. 3.16. The full acquisition channel consumes 3 μA at 0.5 V, yielding an NEF / NEF^2VDD of 4.34 / 9.42, and improves on the prior state-of-the-art NEF^2VDD of 17.96 in [57] by nearly 2x.

3.6 System Results

Although most circuit blocks were broken out and verified individually, 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

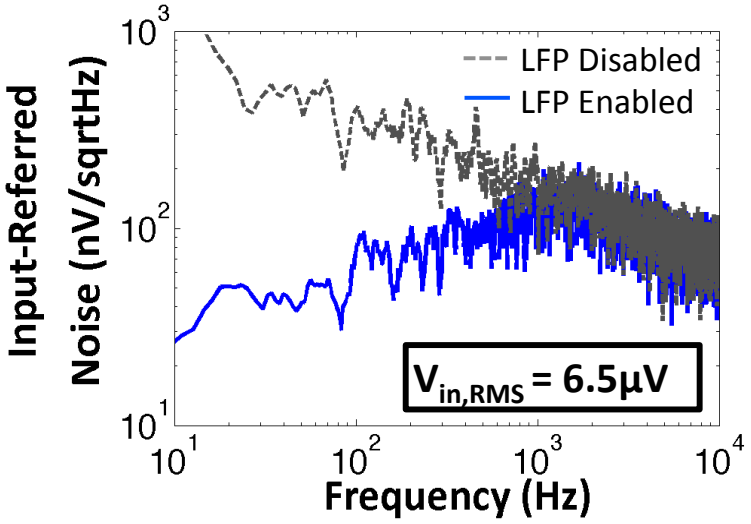


Figure 3.16: Measured amplifier noise spectrum with and without the LFP loop enabled.

the testing to verify the wireless operation range (Section 3.6.1), multi-node communication (Section 3.6.2), simultaneous channel recordings (Section 3.6.3), and operation of the system *in vivo* (Section 3.6.4).

3.6.1 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. 3.17. 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. 3.17 with fitted trend lines. A transmission distance of 1 mm in air is achievable with approximately 50 mW of transmit power. The path loss in the brain was simulated to be approximately 6 dB larger than in air, yielding an equivalent transmission distance of 0.6 mm *in vivo*.

3.6.2 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 4 MHz subcarrier is shown in Fig. 3.18, with 2% duty cycle

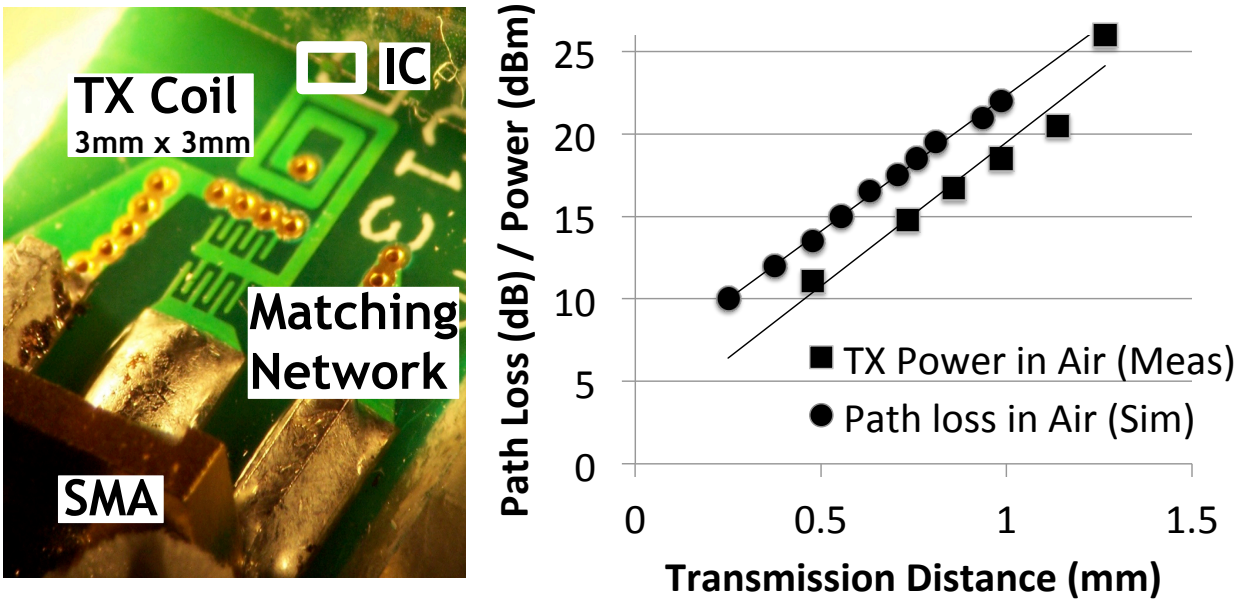


Figure 3.17: Left: Setup for wireless operational range testing. The IC is attached to a micro-manipulator using double-sided tape. Right: Simulated path loss compared to the measured minimum TX power required for operation in air.

interrogator beacons visible at 0 µs and 50 µs. 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. 3.19. The corresponding simultaneous 4 MHz and 8 MHz backscatter can be filtered into independent data streams for decoding, as demonstrated by Fig. 3.4. The multi-node time domain backscatter from Fig. 3.19 is shown in Fig. 3.20 after filtering to isolate the Miller 4 node. The ideal (simulated) waveform is also shown for comparison. 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.

Simulations of the bit error rate (BER) were performed in MATLAB for various numbers of nodes and the results are shown in Fig. 3.21. Initially, in all simulations, the sensitivity improves with increasing SNR. However, above 10 dB 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 10 dB SNR provides an acceptable BER for this application.

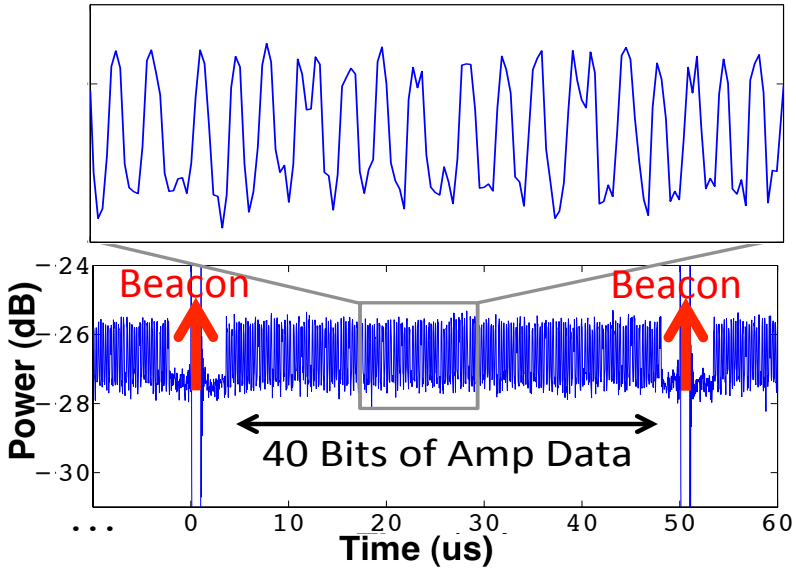


Figure 3.18: A wireless packet encoded with Miller modulation. Backscatter communication of 40 bits of signal acquisition data is initiated by a pulse from the transmitter (seen on either side of the packet).

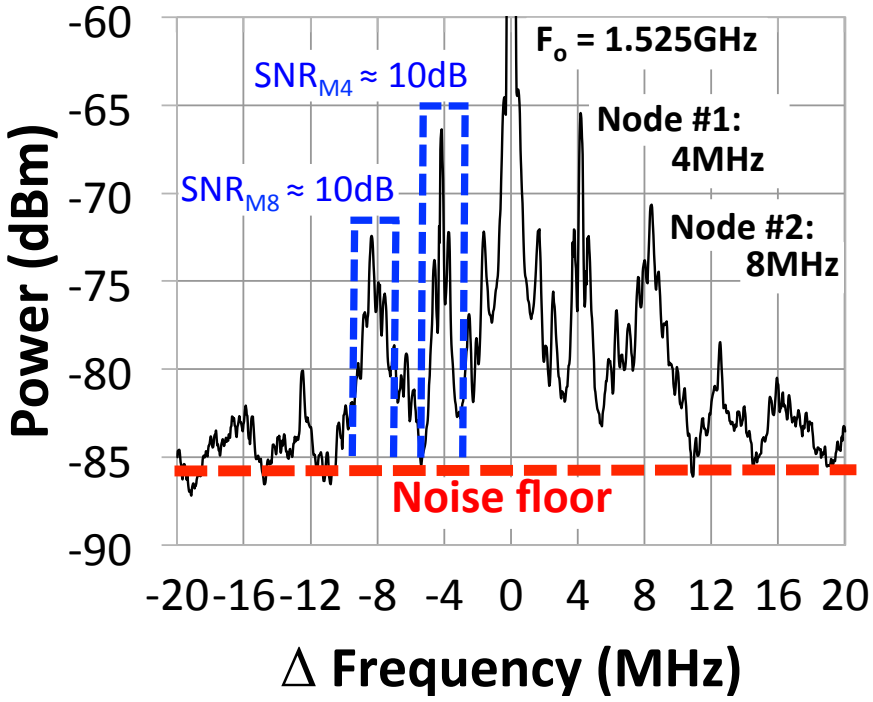


Figure 3.19: Frequency spectrum of two wirelessly powered nodes, communicating simultaneously with the same interrogator at different sub-carrier frequencies.

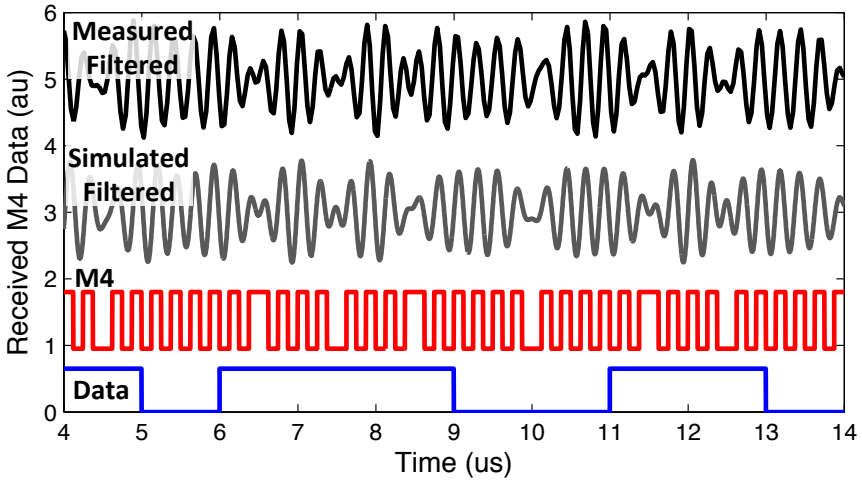


Figure 3.20: The measured time domain waveform after filtering of M4 during a multi-node interrogation test. Results are compared to an ideal filtered waveform, and the equivalent Miller waveform with decoded data is shown.

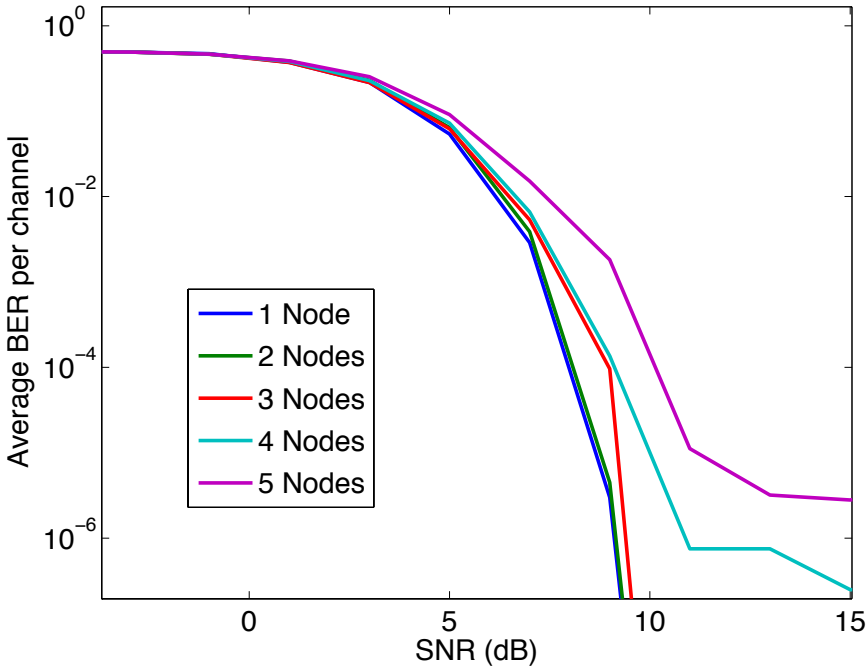


Figure 3.21: Simulated BER vs. SNR for 1-5 simultaneous nodes.

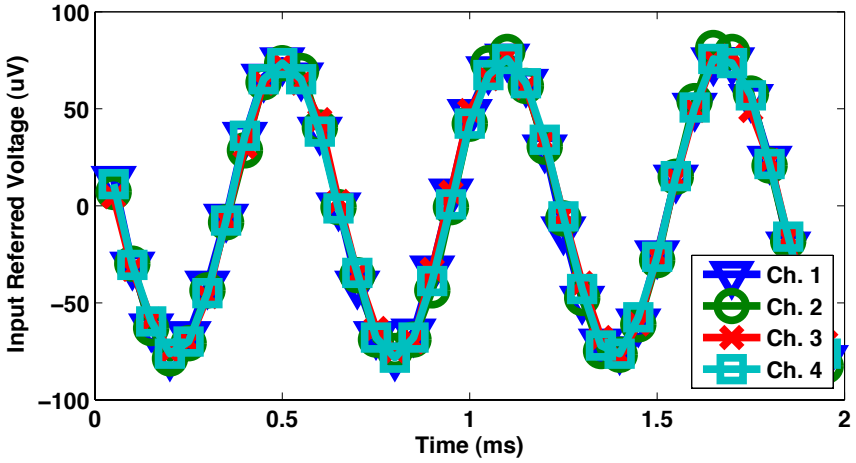


Figure 3.22: A wirelessly-powered system recording and transmitting a 1.6 kHz, 150 µV sine wave input from all channels simultaneously.

3.6.3 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 1.6 kHz, 150 µV sine wave was applied to all four inputs while the system was powered wirelessly through the PCB inductive link. Fig. 3.22 shows the decoded output of all four channels recorded during a testing trial. The outputs show ADC and amplifier performance consistent with the results of stand-alone measurements performed in Section 3.5. 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.

3.6.4 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. 3.23 shows a diagram of the testing setup used to obtain *in vivo* recordings.

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.

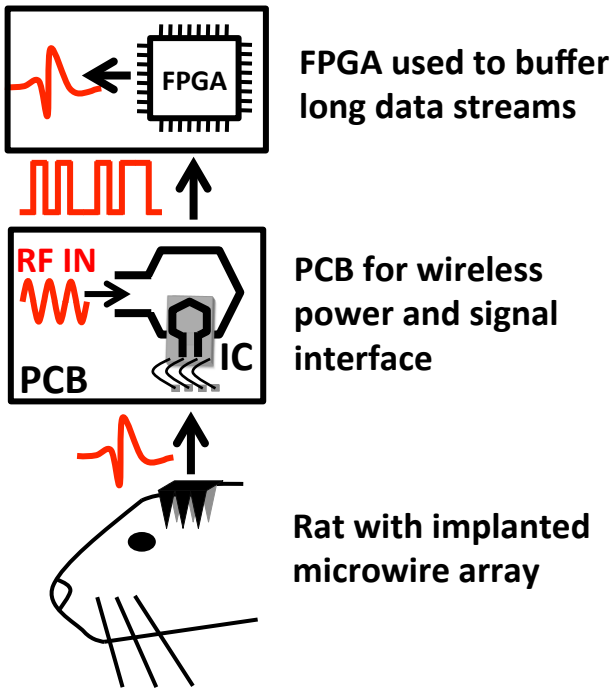


Figure 3.23: The setup for *in vivo* recording trials utilized a rat which was implanted with a microwire array. The system was die-attached to a PCB to facilitate wireless powering and signal interfacing. In order to gather long data streams, a FPGA was used to buffer the on-chip Miller-encoded neural data.

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. 3.24 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 500 Hz. Recorded noise levels varied between recording sites from 15 μV to 20 μV. These noise measurements agree with expectations of the biological noise level as described in Section 3.5.

3.7 Conclusion

This system was fabricated in a 65 nm LP CMOS process with 4 recording channels, power conditioning, communication logic, RF front end, and antenna integrated into an area of

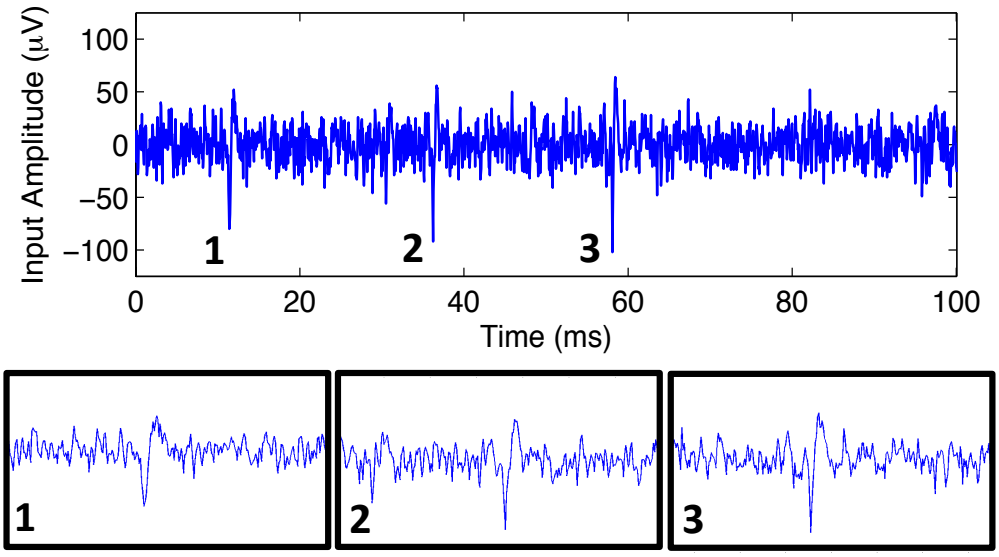


Figure 3.24: One example trial of wirelessly powered *in vivo* neural data from a live rat. The LFP feedback cancelation high-pass corner was set and measured to be approximately 500 Hz.

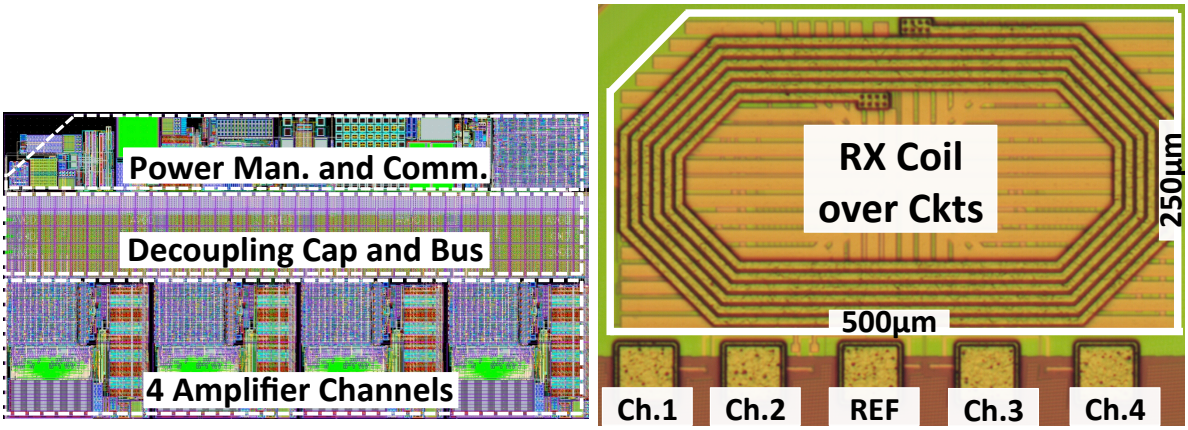
Table 3.1: Comparison of neural recording systems with wireless telemertry.

Author	Chae [60]	Lee [62]	Sodagar [28]	Harrison [41]	*This Work*
Off-Chip (Ant, Cap)	Ant	Ant, Cap	Ant, Cap	Ant, Cap	*None*
Wirelessly Powered	No	Yes	Yes	Yes [b]	*Yes*
In Vivo Results	No	No	Yes	Yes	*Yes*
# Channels	128	32	64	88 [c]	*4*
Total Power (mW)	6	5.85	14.4	13.5	*0.0105*
Avg. Power (μW/Ch)	47	183	220	153	*2.6*
Total Area (mm^2)	63.36	16.2	217	27.3	*0.125*
Avg Area (mm^2/Ch)	0.495	0.506	3.39 [a]	0.310	*0.031*
Amp. Noise (μVrms)	4.9	4.95	8.0	5.1	*6.5*
Noise Meas.	Wired	Wireless	Wireless	Wired [b]	*Wireless*
Process (nm)	350	500	500	500	*65*

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

Table 3.2: Summary of System Performance

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



(a) Chip layout. (b) Die photo.

Figure 3.25: Chip layout and die photo of the full system, showing the input bonding pads, the RX coil and PGS (active area is underneath).

0.125 mm^2. The top-level layout is shown in Fig. 3.25a, and the die photo is shown in Fig. 3.25b, with the 4 inputs, power/communication coil and PGS visible. The node was wirelessly powered and interrogated using a custom PCB antenna and COTS components in a bench-top and in an *in vivo* setting. Table 3.2 summarizes the performance of the system. The complete sensor has no off-chip components and consumes 15 µA from an unregulated voltage source of 700 mV, for a total power consumption of 10.5 µW (2.6 µW/channel) in under 500 µm x 250 µm.

Table 3.1 compares this system to prior neural recording systems with wireless telemetry. This work reduces the average power per channel by 18x compared to [60] and 58x compared

to [41]. Although [41] used a larger (500 nm) process, passives used to build the analog filters consume substantial area even in modern processes. Compared to [41], this work reduces the average area per channel by 10x, and decreases the amplifier and ADC area to 110 μm x 100 μm, compared to 400 μm x 400 μm (for an amplifier, comparator and DAC).

Probe longevity is one of the primary challenges that limit clinical deployment of brain-controlled prosthetic devices today. Repeated surgeries to replace neural probes is unrealistic given the risks and potential complications of each surgery. The main contribution of this chapter is a complete wireless system that is small enough to integrate onto a free-floating neural probe. As discussed in Section 3.1, elimination of micro-motion from a wired tether may help mitigate the brain's biological response, which encapsulates neural probes and reduces the recording SNR over time.

Several key contributions enabled this system. First, the co-optimized communication protocol and modulation / demodulation circuitry provide the required power delivery and data transmission at low power and area overhead. Second, optimizations of the recording channel architecture, originally proposed in [57], provide a nearly 2x improvement in PEF. This was important, as the recording channels account for 40% of the system power even after this optimization. Finally, careful design of the supporting circuitry, including the regulator, reference, and biasing, contribute to the low power and area of the fully-integrated system.

At minimum, this system allows study of how neural probe design affects probe longevity without the confounding factors introduced by wired tethers to the probe. This may lead to improved probe designs, improved tethers, and a reference point for probe longevity. Note that a complete, implantable neural interface that is based on this system also requires a second wireless power link to deliver power and data through the scalp (as in [77], skull, or chest. When the high required channel counts of BMI applications are not required, the losses of two wireless links can be tolerated. However, for clinical BMI solutions, the combined losses of two wireless power links prohibit scaling to thousands of channels. The next chapter presents a system-on-chip solution to address the requirements of BMI applications.

Chapter 4

A Fully-Integrated Neuromodulation SoC

The previous chapter explored the limits of miniaturization and energy efficiency for a wireless implantable recording device. The work presented in this chapter aims to translate those advances into a useful, scalable tool for BMI and neuroscience research. Controlling a high degree-of-freedom (DOF) prosthesis (human hand = 27, human arm = 7) requires a tremendous number of neurons. It is estimated that fast, robust control of a simple 6-DOF arm requires upwards of 1000 neurons [12, 78]. Achieving this channel count in a wirelessly-powered implantable recording device requires the advancements in power and area that were achieved in the previous chapter, while retaining the precision and calibration-free nature of traditional recording channels. Moreover, integrated spike detection and compression are necessary to reduce data rates to a practical level for a wireless link. Finally, an integrated stimulator is required to provide the sensory feedback in a closed-loop BMI.

This chapter presents a fully-integrated SoC with the highest complexity and lowest power/area per recording and stimulation site reported to date. This SoC achieves significant improvements in area, power and signal compression over current state of the art (e.g. [18, 61, 19]). When arrayed across the brain, 16 ICs provide 1024 recording and 128 stimulation sites. This would require 6.67 mW and 320 kbps, which can be delivered through the skull as shown by [64, 79].

4.1 System Vision

Fig. 4.1 shows our vision for a wireless implantable recording and stimulation platform. A base station outside the skull transmits power to the sub-cranial platform, which lies between the skull and the brain tissue. Compliant cables tether probes implanted in the brain tissue to recording electronics on the implanted platform. There were several motivations for this architecture. First, the compliant tethers between the probes and the implant eliminate the power losses of the wireless link in Chapter 3. This is critical when scaling to high channel

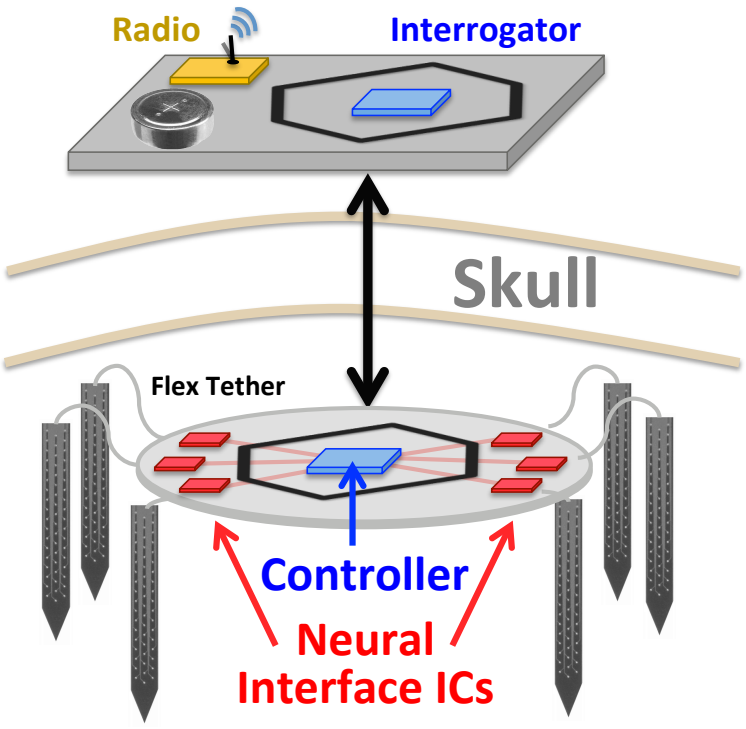


Figure 4.1: Berkeley wireless neural interface vision.

counts. Second, by placing the implant beneath the skull (and potentially beneath the dura), the implant can move with the brain relative to skull. This eliminates strain caused by wires leading through the skull as in [28, 77]. Third, the implant can support a large number of neural interface ICs, and the implant substrate can accommodate fast, low power data links to the aggregator. The aggregator can be customized to support an RFID data link to the interrogator, a subdermal wire to a prosthesis, or a closed-loop system for treatment of Parkinson's. Lastly, the implant can be sized to provide a sufficiently large inductive coil for wireless power delivery and data transfer. The flexible tethers decouple the implant size and rigidity from the probes, and also allow researcher to continue development of probes and tethers independent of the recording electronics. Note that the recording ICs can by bonded to the probes to reduce the wire count in the tether to 5 conductors (power, ground, data in, data out, clock) versus 64 conductors (one per electrode); this is left to the system designer.

The IC architecture, shown in Fig. 4.2, 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. This work was originally published as [80]. The design of the recording channels, which is the author's contribution, is presented here. The compression and stimulation circuits were designed by coauthors Nathan Narevsky and William Biederman, respectively, and design details may be found in

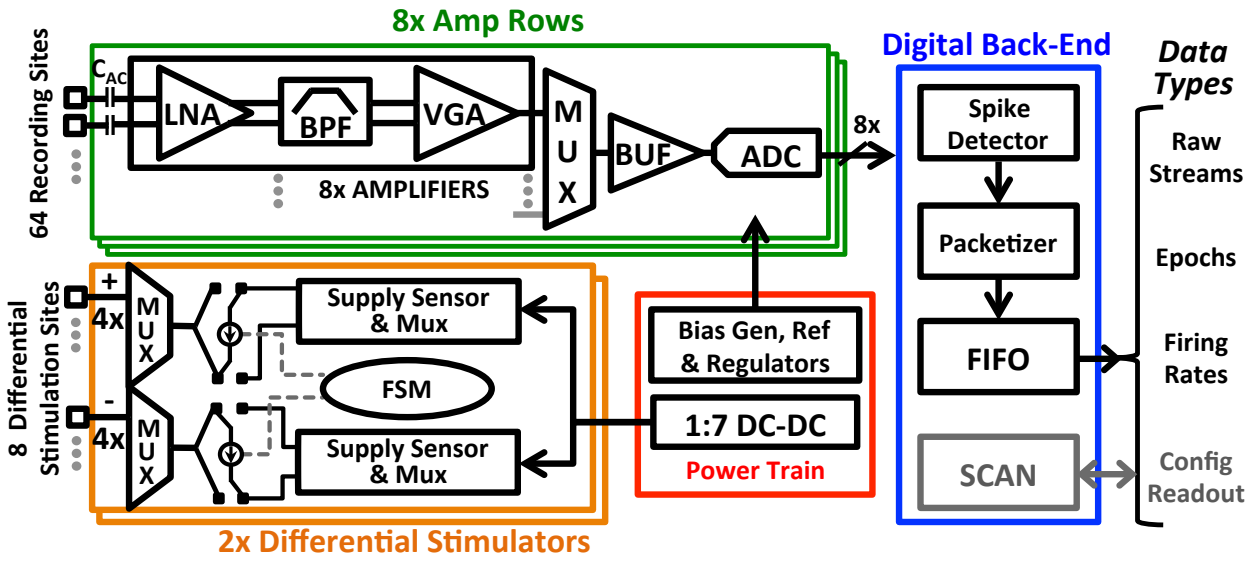


Figure 4.2: Berkeley wireless neural interface chip block diagram.

[81]. System integration was largely the work of Jaclyn Leverett, and [82] presents a wireless headstage for animal recordings that is based on this IC. Section 4.2 presents the design and Section 4.3 presents the measurement results of the data acquisition components. Finally, Section 4.4 presents the *in vivo* testing results of the fully integrated system.

4.2 Acquisition Channel Design

The requirements for the acquisition channels are significantly different than those presented in Chapter 3. First, this system is designed to be compatible with large (2 - 3 V) stimulation common-mode voltages. Second, the chip should be safe even in the case of gate-oxide failures. Tissue damage may occur if the supply voltage becomes connected to an electrode [45]. The first and second requirements necessitate AC coupling. Third, the input-referred gain of the recording channels must be both precise and accurate, as defined by [83]. This necessitates either closed-loop gain or sophisticated, automated, on-chip gain calibration. Because the closed-loop gain approach is less complex, and initial studies suggested that the traditional area penalties of a closed-loop design could be mitigated, that approach is taken here. Finally, substantial area reduction is required versus state-of-the-art AC coupled recording channels. Architectural and circuit design techniques are applied to mitigate this concern.

Fig. 4.3 presents the top-level block diagram for the acquisition channel. The inputs are AC-coupled by on-chip 10 pF capacitors. Eight rows, each consisting of 8 amplifiers and an ADC, form the 64 recording channels. A time-multiplexed switched-capacitor ADC driver utilizes separate sampling capacitors for each of its 8 input channels, thereby decoupling the fast settling requirements of the ADC from the low bandwidth requirements of the input

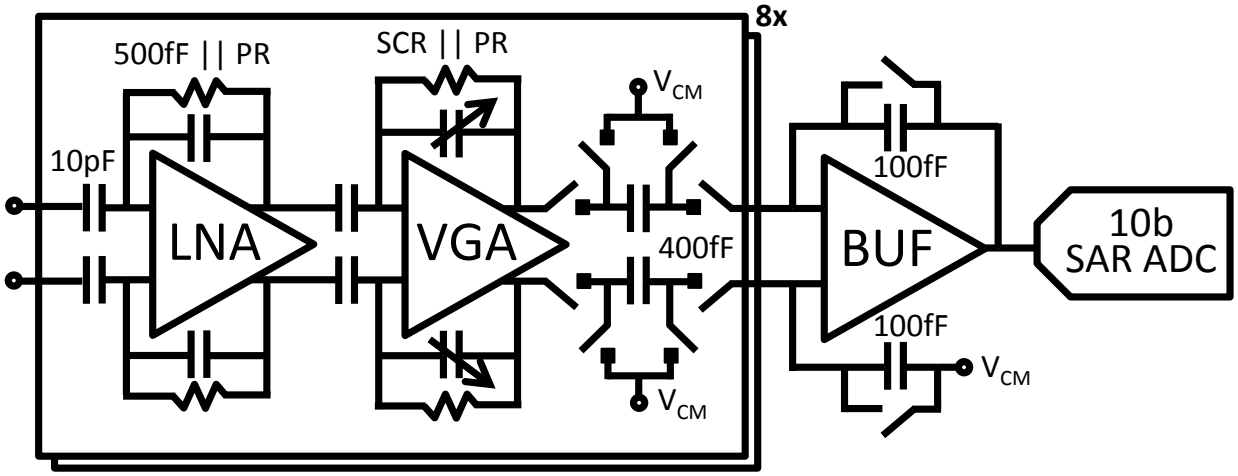


Figure 4.3: Block diagram of a recording channel row. Eight rows, each containing 8 amplifiers time-multiplexed onto a 10-bit SAR ADC, comprise the 64 channel array.

signal. Finally, a 10-bit SAR ADC digitizes the amplified signals.

The input-referred gain is set through closed loop feedback to provide accurate, calibration-free operation. The LNA, VGA, and BUF provide gains of 26 dB, 6-30 dB, and 12 dB, respectively, with a total of 44-68 dB of gain. This results in an input-referred full-scale at the 10-bit 1.0 V ADC of approximately 200 μVpp to 6.3 mVpp. The gain, bandwidth and bias current (noise performance) are individually adjustable on a per-channel basis, enabling power savings on high SNR electrodes. However, as is critical for a 1000 channel implantable system, the default settings are sufficient for most applications.

4.2.1 Amplifier Design

The amplifier consists of three AC-coupled closed-loop gain stages, shown in Fig. 4.3. The LNA must provide low input-referred thermal and flicker noise down to approximately 500 Hz, adequate gain to suppress noise of succeeding stages, AC coupling to reject input DC offset, and sufficiently high input impedance (>10 MΩ at 10 kHz). The VGA must provide variable gain, low-pass filtering to reduce noise aliasing at the ADC, and sufficiently-low noise to avoid degrading SNR from the LNA. Finally, the buffer (BUF) must provide rail-to-rail swing at the ADC input, settling within 1 LSB, and minimal loading on the VGA. By decoupling these design requirements into 3 separate stages, each can be optimized independently to maximize their performance.

The fully-differential LNA and VGA gain stages are shown in Figs. 4.4 and 4.5. The LNA is based on [50], which heavily degenerates the input-stage current sources to nearly eliminate their noise contribution. Simultaneously, the output stage employs substantially lower bias current, which drastically reduces the noise contribution of the output stage. This enables high gain without the noise penalty of transistor load devices. Both the LNA and

VGA utilize thick-oxide input devices; this prevents contention with the pseudo-resistor DC feedback and maintains compatibility with large stimulator transient voltages.

In this work, the LNA utilizes precision replica biasing that enables a substantial supply reduction to 1.0 V while retaining the low-flicker-noise poly degeneration resistors. Next, a fully differential architecture relaxes the stringent matching requirements in [50] because the common-mode feedback (CMFB) serves to absorb mismatch between the PMOS current sources. This would otherwise require consume substantial area; this LNA requires 13.4x less area than [50]. Finally, a level shifter decouples input and output common-mode levels for the DC-feedback path (Fig. 4.7). This satisfies the input common mode requirements of the thick oxide input devices. The capacitive feedback is fed back directly from the LNA output, eliminating drive requirements of the level shifter.

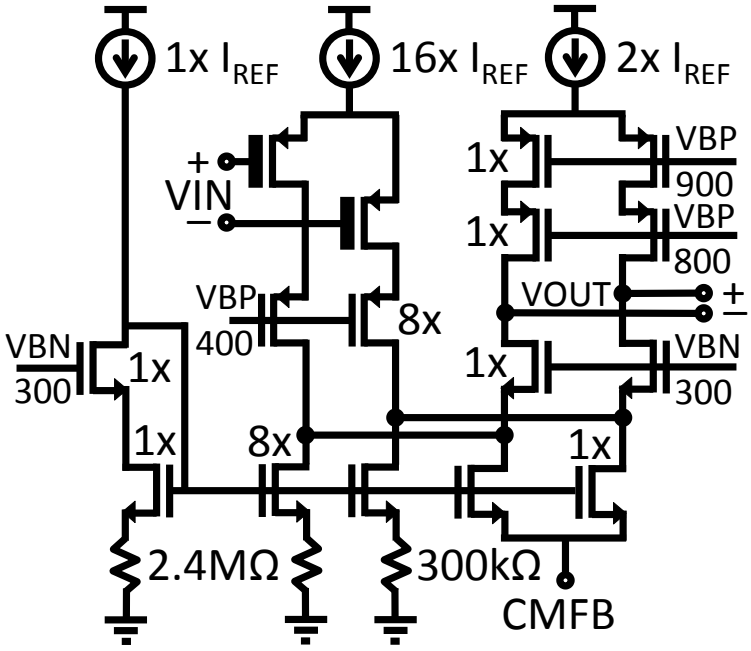


Figure 4.4: LNA folded cascode core schematic.

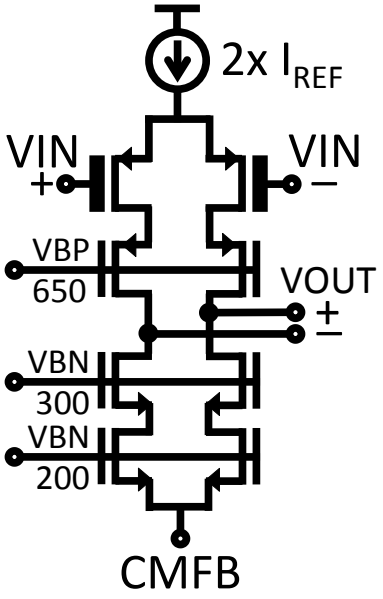


Figure 4.5: VGA cascode core schematic.

The LNA utilizes vertical metal-insulator-metal (MIM) capacitors for AC-coupling and feedback, which enables placement of the capacitors over the active circuits in order to minimize die area. The VGA instead uses small, low-overhead 50 fF lateral metal-oxide-metal (MOM) unit capacitors to adjust the gain from 2x to 32x, where the feedback capacitance ranges from 16x 50 fF to 1x 50 fF, and the series capacitance is fixed at 32x 50 fF. By decreasing the effective load capacitance as the closed-loop gain increases, the variation in the VGA bandwidth across gain settings is minimized. Switched capacitor feedback resistors tune the high-pass corner, achieving approximately 100 Hz to 1 kHz high-pass corner range for any gain setting. If the SC resistors are not clocked, the MOS pseudo resistors set the high-pass corner at approximately 10 Hz, subject to process variation. The layout is presented in Section 4.2.4.

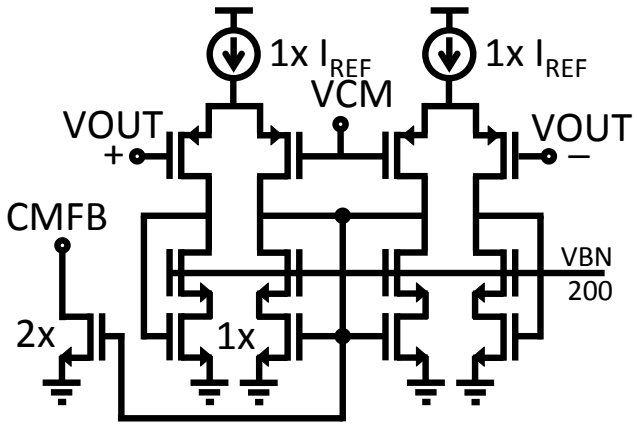


Figure 4.6: LNA & VGA CMFB schematic.

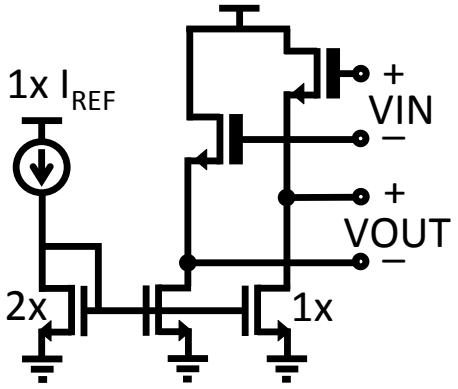


Figure 4.7: LNA & VGA DC feedback level shifter schematic.

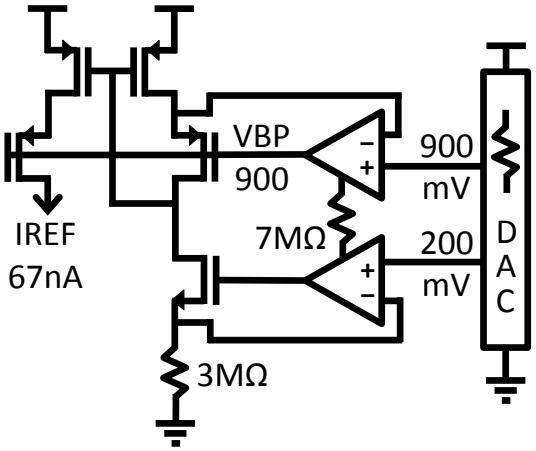


Figure 4.8: Bias current and 900mV cascode bias ("VBP900") generation.

Fig. 4.8 shows the unit current reference generation. A supply-referenced resistor string voltage DAC sets the target bias voltages. An opamp forces 200 mV across a 3 MΩ poly resistor, creating 67 nA reference currents for each row. This reference current is divided by 6 to distribute a unit current of 11 nA to each bias DAC for each LNA, VGA and BUF. For example, most testing was performed with a DAC setting of 5, which results in about 56 nA I_{REF}.

Fig. 4.8 also illustrates the first cascode reference generation replica stage. To provide 100 mV of headroom for the PMOS current source, the gate of the PMOS cascode is set by an opamp to provide at 900 mV cascode's source. The remaining cascode bias voltages are generated by similar replica topologies, except that the 67 nA reference current biases subsequent opamps instead of a resistor. The entire bias generation circuit occupies 0.03 mm^2 and consumes 2.2 μA, which serves all 64 channels.

4.2.2 ADC Buffer

Each ADC buffer (BUF) must serially settle 8 channels onto its respective ADC's 260 fF ADC input capacitance at 160 kHz (8 x 20 kHz). As shown in Fig. 4.3, each VGA drives a separate 400 fF BUF sampling capacitor to allow maximum settling time for the VGA. The settling time for the VGA and evaluation time for the BUF are shown in Fig. 4.11. The BUF feedback capacitors are reset after evaluation. Settling for the next VGA evaluation therefore always starts from zero, effectively mitigating crosstalk between amplifiers.

To minimize power consumption, the conversion time of 6.25 µs (1/160 kHz) is divided unevenly as depicted in Fig. 4.11. The 6.25 µs is divided between 10 ADC conversion cycles and N sampling cycles. N=22 resulted in a convenient 32x clock (5.12 MHz), which is 1/4 of the 20.475 MHz system clock. Consequently, the BUF must settle within 22/32 of the 6.25 µs ADC conversion window, or 4.3 µs. This settling time (T_o) is divided between slewing and gm/C settling. The BUF requires 1.2 µA to settle at 8x 20 kHz, equivalently, 150 nA per 20 kHz amplifier.

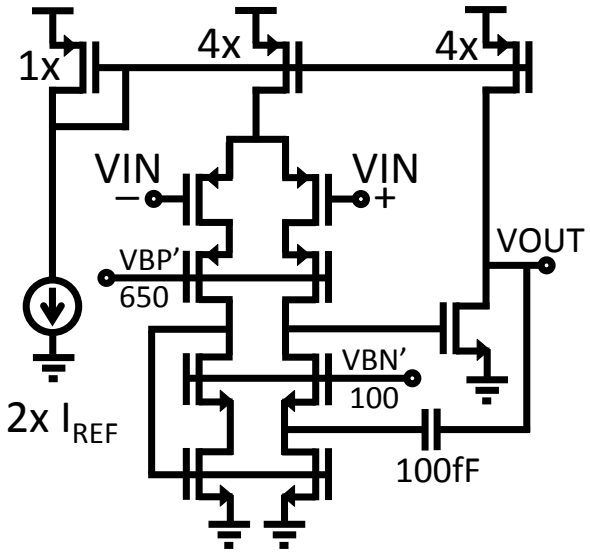


Figure 4.9: ADC buffer core schematic.

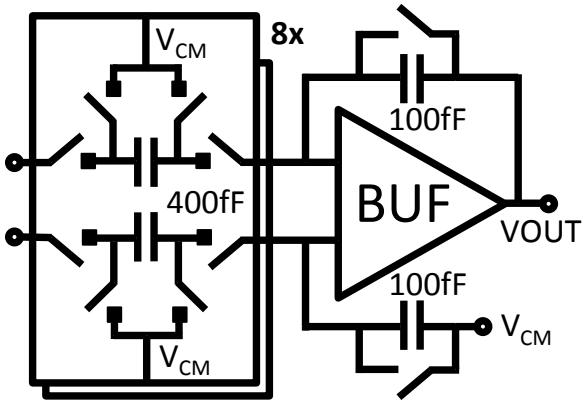


Figure 4.10: Buffer feedback schematic.

4.2.3 10-bit SAR ADC Design

The 10b SAR ADC schematic is shown in Fig. 4.13, and the bottom-plated sampled ADC capacitor DAC is shown in Fig. 4.14. Custom 260 aF MOM unit capacitors allow for a compact, low power ADC. The two capacitor DACs consume approximately 100x70 µm. The capacitor array has 4 thermometer bits to limit DNL, and the remaining 6 bits are binary-switched.

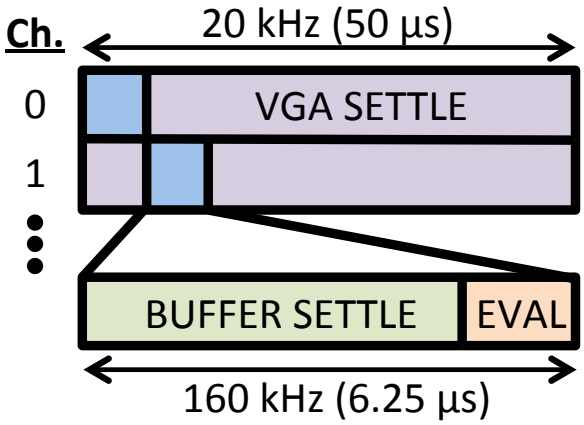


Figure 4.11: Buffer and ADC timing diagram.

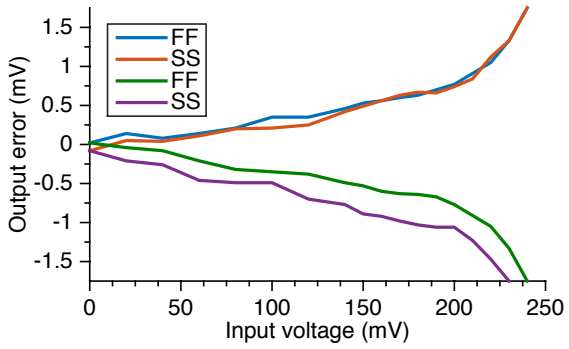


Figure 4.12: Simulated ADC buffer settling error versus input voltage (full scale input = 250 mV).

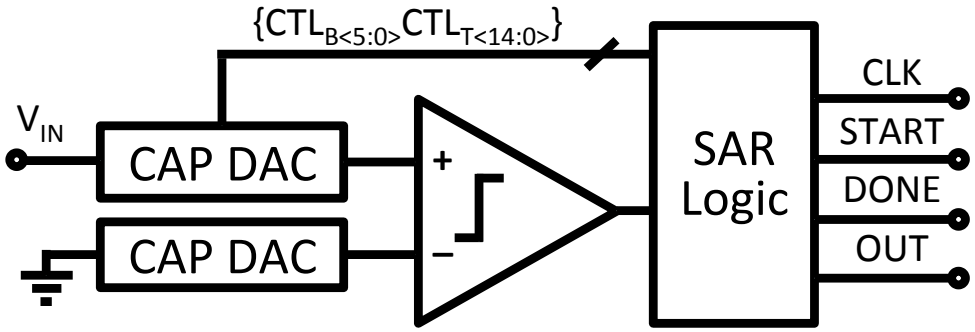


Figure 4.13: ADC top level schematic.

To avoid charge pumps and swings beyond the rails or a power-hungry VDD/2 reference, half of the top plates are charged to VDD and the other half are charged to GND during the sampling phase, effectively creating a VDD/2 reference. No area overhead is incurred because ADCs are constructed with unit capacitors, and half of the unit capacitors for each bit are allocated to the VDD and GND arrays. The LSB and the fixed capacitor effectively balance out, with one on each side of the array.

The required sampling capacitance of a SAR ADCs given thermal noise constraints is highly sensitive to the required resolution. The following equations provide the thermal noise and matching requirements for the capacitors.

$$N_q \approx \frac{\Delta^2}{12} \tag{4.1}$$

$$N_t = \frac{k_B T}{C} \tag{4.2}$$

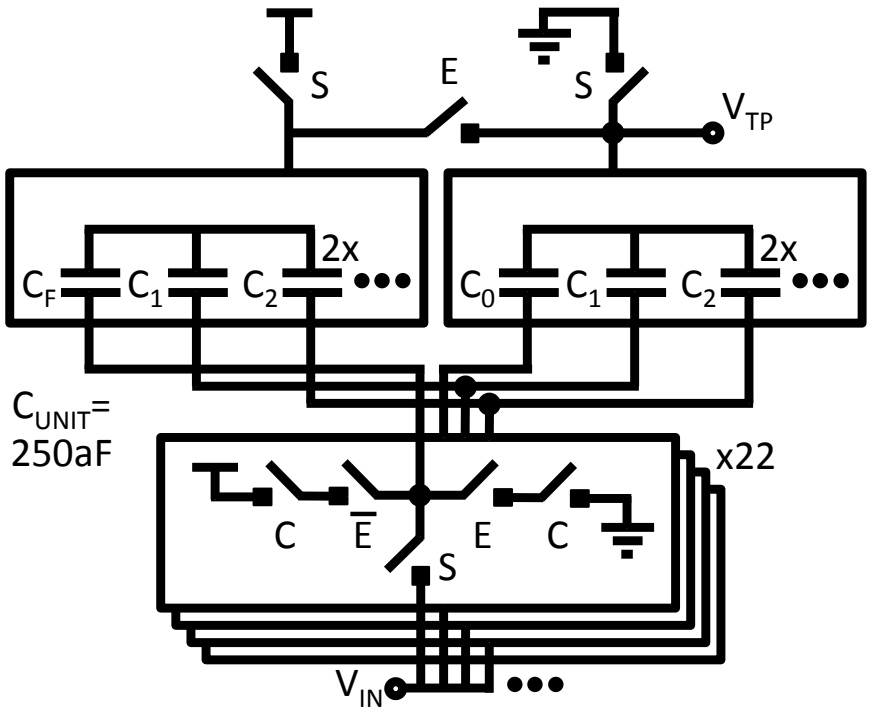


Figure 4.14: Bottom-plate sampled ADC capacitor array with sampling and evaluation switches. The array is composed of unit capacitors; half of the unit capacitors for each bit are sampled with respect to VDD and GND, respectively, which creates an effective 500 mV reference.

$$\frac{k_B T}{C} \leq \frac{\Delta^2}{12} \tag{4.3}$$

$$\Delta = \frac{V_{FS}}{2^B - 1} \tag{4.4}$$

$$C \geq 12 k_B T \frac{2^{2B}}{V_{FS}^2} \approx 52fF \mid_{B=10}^{V_{FS}=1V} \tag{4.5}$$

$$C_{UNIT} \geq 12 k_B T \frac{2^{B}}{V_{FS}^2} \approx 51aF \mid_{B=10}^{V_{FS}=1V} \tag{4.6}$$

For a 10-bit ADC constructed with unit capacitors, the minimum unit capacitor is approximately 51 aF if the thermal noise is to be less than the quantization noise. However, most CMOS technologies do not provide standard capacitors below approximately 1 fF, and moreover, achieving 10-bit resolution typically requires larger unit capacitors. By analyzing mismatch statistics of metal-oxide-metal (MOM) lateral finger capacitors ranging from 400 aF to 100 fF, a regression for 1σ mismatch was estimated as $\frac{1}{2.3\sqrt{C}}$ for C in fF which

should result in 0.87% unit mismatch for 250 aF unit capacitors. Equations 4.7 and 4.8 give the mismatch for a 6 bit binary (B_B), 4 bit thermometer (B_T) array.

$$\sigma_{INL} \approx 2^{B/2-1}\sigma_\epsilon = 16\sigma_\epsilon \mid_{B=10} \tag{4.7}$$

$$\sigma_{DNL} \approx 2^{(B_B+1)/2}\sigma_\epsilon = 11.3\sigma_\epsilon \mid_{B_B=6} \tag{4.8}$$

Careful efforts were taken to minimize systematic mismatch. First, a common-centroid layout cancels gradients and differences between the left and right capacitor arrays (Fig. 4.14). Second, two dummy cells are placed on each side of the array to mitigate edge effects. Third, the LSB capacitors are centered in the array to prevent gradient effects that cannot be canceled out to the small number of unit capacitors. To a large degree, these efforts are successful. Results are presented in Section 4.3.2.

4.2.4 Layout

Layout of a 64 channel amplifier array requires careful planning to avoid crosstalk. For example, parasitic feedback paths could cause oscillation, and crosstalk between clock and input could introduce noise. To prevent potential issues, the input signals are shielded from the pads to the LNA. Bias voltages from the bias generator are also shielded; NMOS voltages are ground-shielded, and PMOS bias voltages are VDD-shielded. This means that supply noise perturbs the gate and source of transistors in the same way. Bias currents are routed as currents, not voltages, until as close to the amplifier as possible. This keeps signal impedances low during long routes, helping reduce susceptibility to disturbance. Finally, cascode bias voltages and bias currents are decoupled to their respective supplies via thick-oxide MOS capacitors (to minimize leakage) inside the amplifiers.

The layouts of the amplifier and ADC are shown in Fig. 4.15. The large LNA MIM capacitors and small VGA MOM unit capacitors are clearly visible. Routing of input and output signals occurs along the bottom. Bias currents are routed along the top; they come from a mirror in the center of the row. Bias voltages are routed in a vertical shielded channel through the middle of the AMP. Finally, supply and ground rings around each pixel automatically connect when pixels are tiled.

The layout of a row, which containing 8 amplifiers and 1 ADC, is shown in Fig. 4.16. The LNA inputs route automatically to the left and right edges of the row, while the VGA outputs route automatically to the BUF/ADC in the middle. Bias voltages pass vertically, which necessitates connection by a horizontal strap at the array level. This horizontal strap then connects to the bias generator. The die photo, presented in Section 4.3, elucidates the floorpan, and required routing, at the top level.

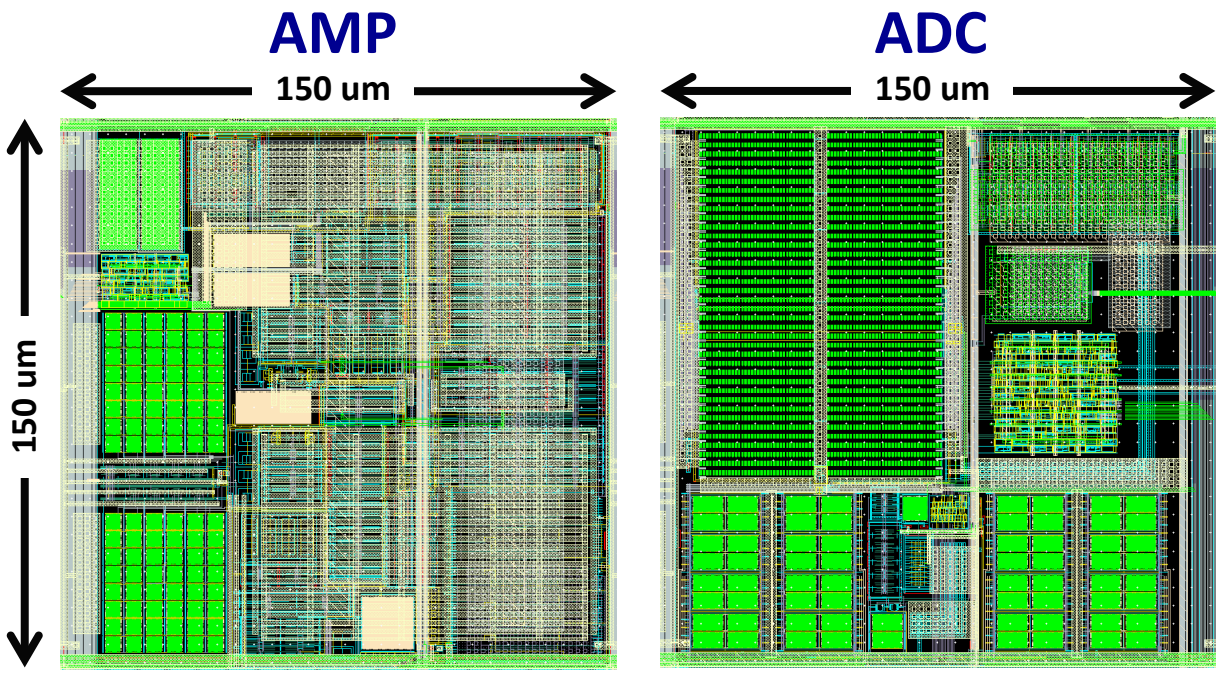


Figure 4.15: Layout of the amplifier and ADC.

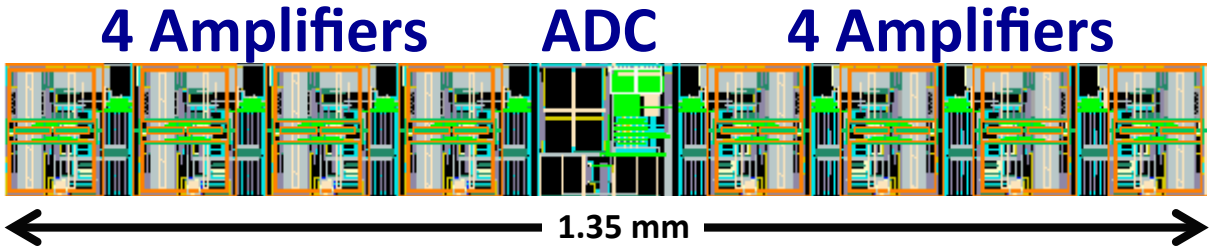


Figure 4.16: Layout of a row of 8 amplifiers and 1 ADC.

4.3 Characterization

The system was fabricated in TSMC 65nm LP CMOS and occupies 4.78 mm^2 including pads. A die photo is shown in Figure 4.17 with key circuit blocks outlined and annotated.

4.3.1 Amplifier Measurement Results

The high-pass tuning range was characterized for a low (46 dB) and high (66 dB) gain setting. Results are shown in Figs. 4.18 and 4.19, respectively. The resultant corner frequencies are then plotted in Fig. 4.20. Approximately 100 Hz to 1 kHz high-pass corner frequencies are achievable across the gain settings. The theoretical and measured range of gain are plotted in 4.21, demonstrating the gain tuning range of 44-66 dB.

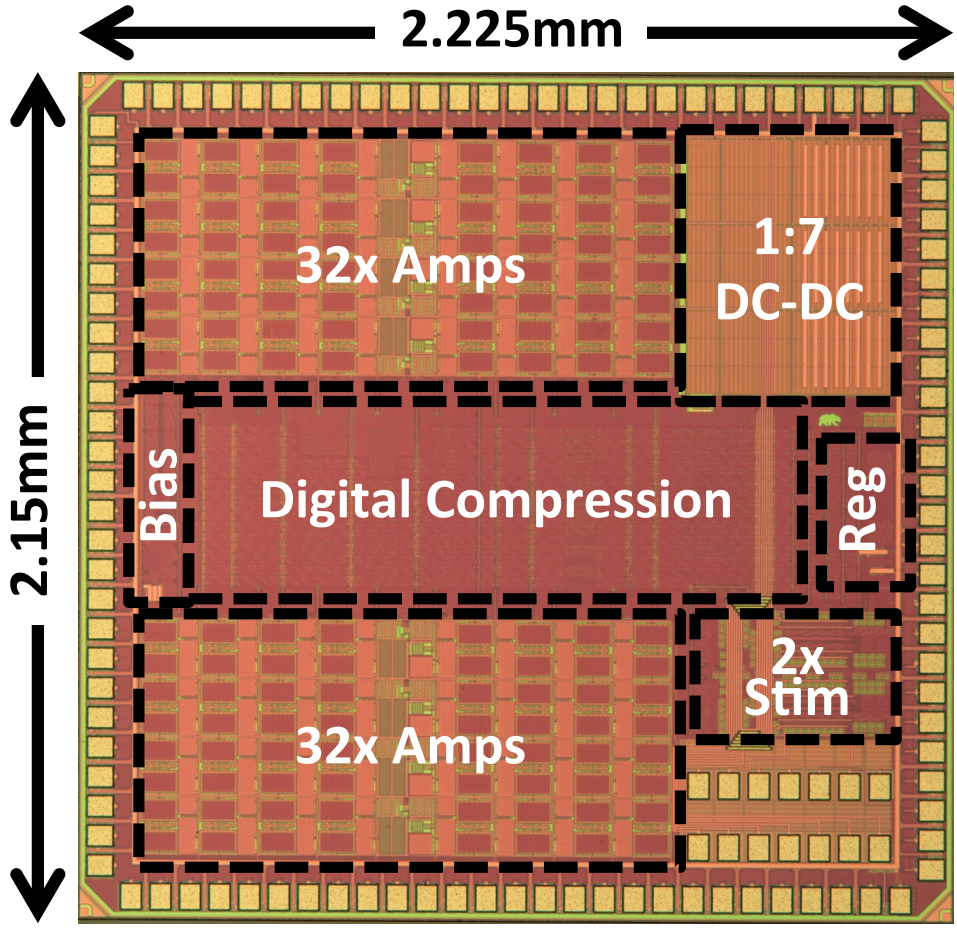


Figure 4.17: Die photo of the 64-channel recording, dual stimulation and integrated compression IC.

Closed loop gain varies the most at the highest gain setting, where the variation in open-loop amplifier gain begins to affect the closed loop gain. Fig. 4.22 shows a transient recording for a full array with a 100 μVrms input at 1 kHz. Peak to peak gain variation at the highest gain setting was measured at 0.9 dB and a standard deviation of 2.2% and is shown in Fig. 4.22,. At the second highest gain setting, peak variation reduced to 0.7 dB, or 1.7% standard deviation.

The measured amplifier input referred noise spectrum is plotted in Fig. 4.23. The noise data was taken with an input signal present, which is used to calculated the input-referred noise. The input signal fundamental was blanked (set the average noise of its neighbors), but harmonics were not blanked and are seen in the noise spectral plot. These harmonics minimally affect the total integrated noise and were thus ignored. The final input-referred integrated noise, as measured through the on-chip ADC, was 7.5 μVrms.

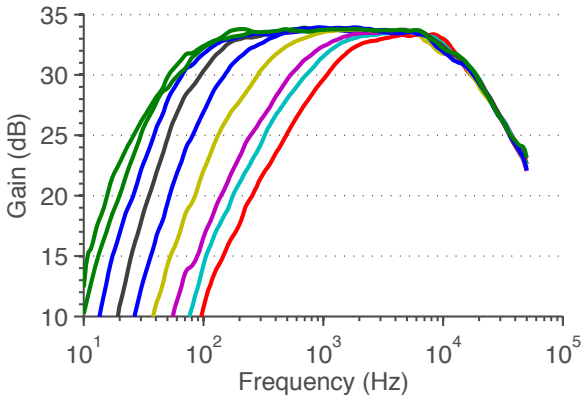


Figure 4.18: Measured bode plot of LNA & VGA with gain set to 50 vs. selected high-pass corner settings.

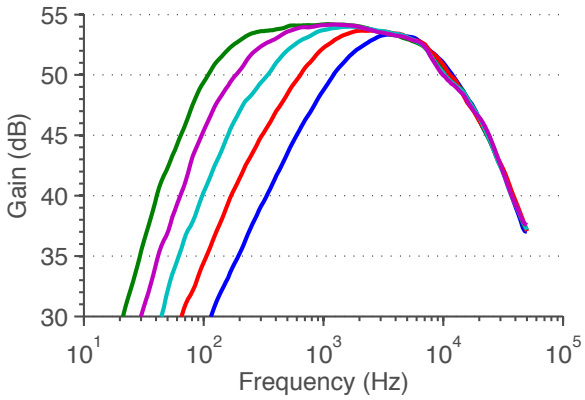


Figure 4.19: Measured bode plot LNA & VGA with gain set to 500 vs. selected high-pass corner settings.

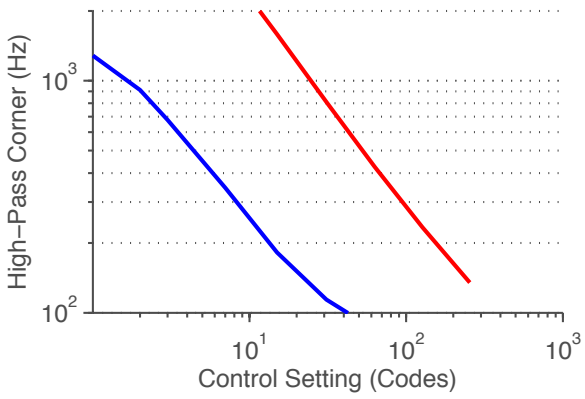


Figure 4.20: Measured VGA high-pass corner vs. control code at gain of 50 (blue) and 500 (red) V/V.

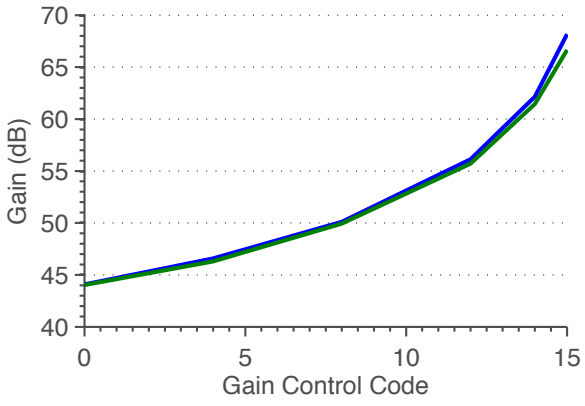


Figure 4.21: Measured (green) and simulated (blue) LNA & VGA gain vs. control code.

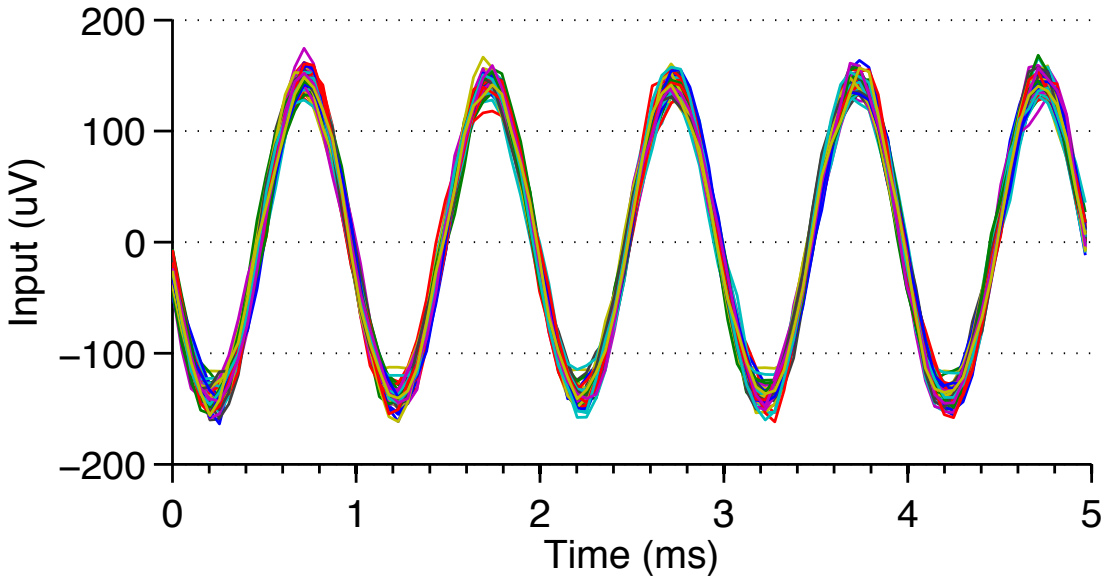


Figure 4.22: Gain matching for the array, at the highest gain setting.

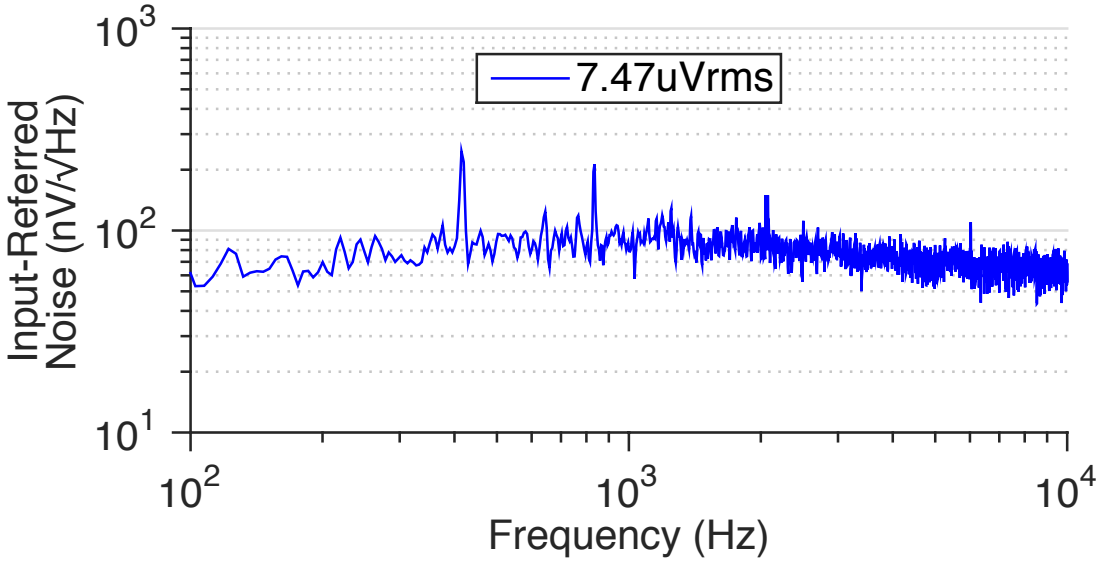


Figure 4.23: Measured amplifier input referred noise spectral plot.

4.3.2 ADC Measurement Results

The measured DNL and INL are shown in Figs. 4.24, 4.25, and 4.24. The peak DNL is approximately 1 LSB, and the peak INL is approximately 1.5 LSB. This results in an ENOB ranging from 8.2 to 8.6 bits for the 4 measured samples. ADC achieves an ENOB of at least 8.2 bits across the four samples.

However, Fig. 4.25 shows that the layout clearly suffers from systematic mismatch. In order to preserve symmetry, the first and second LSB capacitors and the fixed capacitor are centered in the array, and the remainder of those rows are filled with dummy capacitors. The DNL plot shows that the first and second LSBs are more correlated to each other than to the third LSB. This suggests that the dummy capacitor induces mismatch which was not captured in extracted simulation.

The mismatch may be due to mask corrections such as OPC which are not always modeled. The use of a 2.5D extraction tool may also have contributed, as it fails to fully capture fringing fields. Lastly, lithographic, etching, or density gradients may have affected the manufacturing.

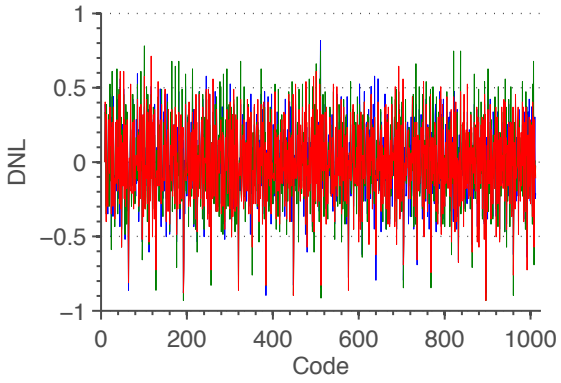


Figure 4.24: Measured ADC DNL for 4 chips, with a peak of 1 bit.

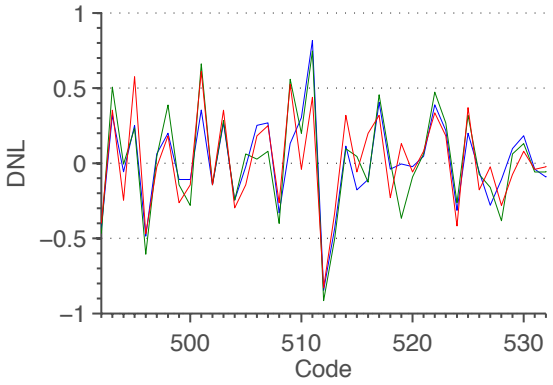


Figure 4.25: Measured ADC DNL about midpoint for 4 chips, revealing correlation between chips.

4.3.3 Performance Summary

Table 4.1 lists key performance metrics for components of the signal acquisition channels. This work achieves the lowest area per channel (0.0258 mm^2 including biasing and digitization) for a closed-loop, AC-coupled design by 3x versus [56] and 6x versus [55]. This area efficiency enables a substantial increase in number of recording channels per chip, as well as integration of on-chip compression and stimulation. Note that this design consumes 2.3x the area of our previous open-loop design [67]. However, that design required per-channel gain calibration that we considered unsuitable for deployment in complex systems with hundreds to thousands of channels.

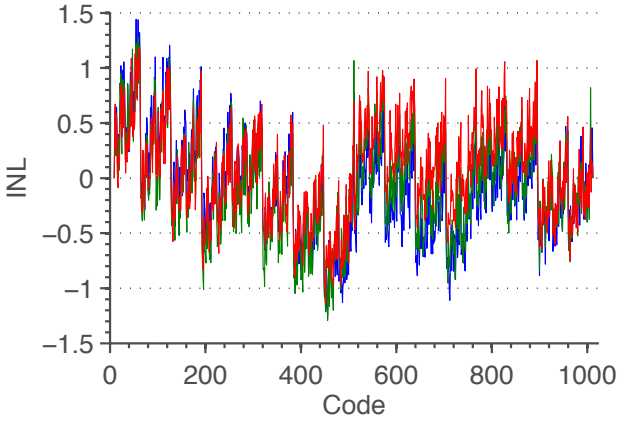


Figure 4.26: Measured ADC INL for 4 chips, with a peak of approximately 1.5 bits. The average INL for the 4 samples is overlaid as a dark blue trace.

Wireless systems demand simultaneous low noise and low power, which is characterized by the noise / current efficiency (NEF) and noise / power efficiency (PEF=NEF^{2}VDD). For comparison with other works, the LNA alone consumes 1.2 μW excluding biasing, and has an NEF and PEF of 3.6 and 12.9, respectively. However, this alone is meaningless; the entire recording channel's power consumption must be considered. Here, each channel consumes 1.84 μW including biasing and digitization, resulting in system NEF and PEF of 4.45 and 19.8, respectively. This compares favorably with [55] (NEF=5.4, PEF=52.5). Excellent PEF is achieved by [56] (NEF=3.6, PEF=6.6), which may be partially due to a lower-flicker-noise process. Nevertheless, there may be opportunities to further optimize this design for better noise efficiency. In particular, the LNA performance should be able to approach [50] because the poly-resistor degeneration was not foregone as in [55]. Excessive measured flicker noise in this design did not match foundry models, and further investigation is required to uncover the culprit.

Table 4.2 lists the per-channel power, area, and gain, broken down by component. The power and area of the BUF and ADC are shared among 8 channels, and are divided accordingly. The bias generation is shared by all 64 channels, and is divided by 64 to give per-channel cost. The bulk (68%) of the power in this design is spent to achieve low noise in the LNA. The VGA requires significantly less power than the LNA at 18% of the total budget. Finally, the time-multiplexed BUF and ADC combined only require of 11% of the power per channel.

Table 4.1: Component Performance Summary

LNA	Gain	26 dB
	NEF	3.6
	PEF	12.9
VGA	Gain	6-27 dB
	Swing	±250 mV
BUF	Gain	12 dB
	Swing	±1 V
ADC	ENOB	8.2 bits
	Conv. Rate	160 ksps
	DNL	0.84 LSB
	INL	1.5 LSB
	Power	800 nW
	Area	9200 μm^2
Bias Gen	Power	2.18 µW
	Area	0.03 mm^2
Channel	Gain	44-65 dB
	IRN	7.5 µVrms
	Power	1.84 µW
	NEF	4.45
	PEF	19.8

Table 4.2: Effective Per-Channel Power, Area, and Gain

Component	Power (µW)	Power (%)	Area (μm^2)	Area (%)
LNA	1.26	68.4	11,900	46.2
VGA	0.33	17.9	6,700	26.0
BUF	0.12	6.5	1150	4.5
ADC	0.10	5.4	823	3.2
Bias Gen	0.034	1.8	470	1.8
Routing	-	-	4738	18.3
TOTAL	1.84	100	25,781	100

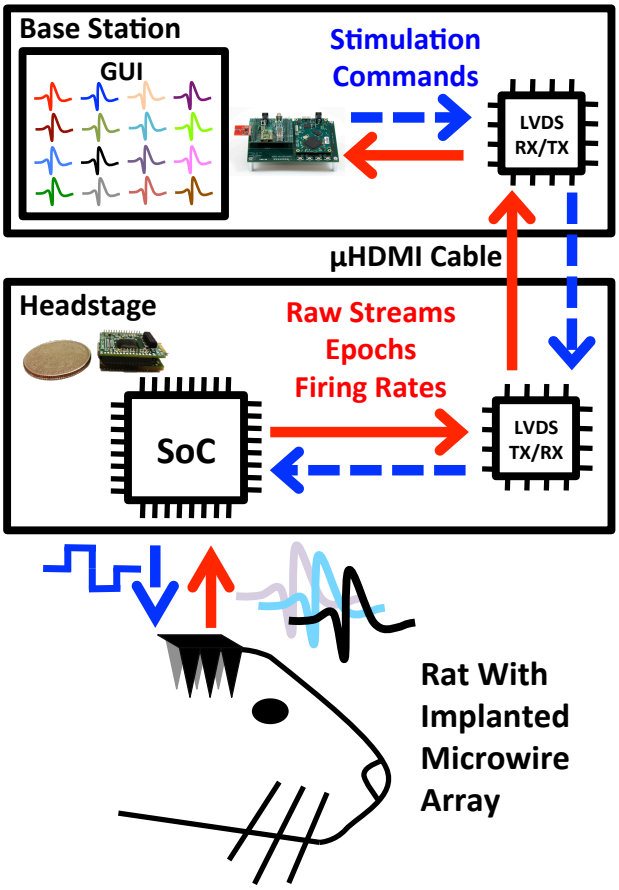


Figure 4.27: The *in vivo* neuromodulation test system is composed of a microwire implanted array, a compact headstage containing the SoC, a base station, and a Graphical User Interface (GUI).

4.4 In Vivo Testing

4.4.1 Measurement Setup

A diagram of the testing system designed to seamlessly obtain *in vivo* data is displayed in Figure 4.27, which includes a compact 0.65" x 0.8" headstage, a base station, and a Graphical User Interface (GUI) [84]. 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. A screen capture of the GUI recording test signals through the ASIC is presented in Fig. 4.28.

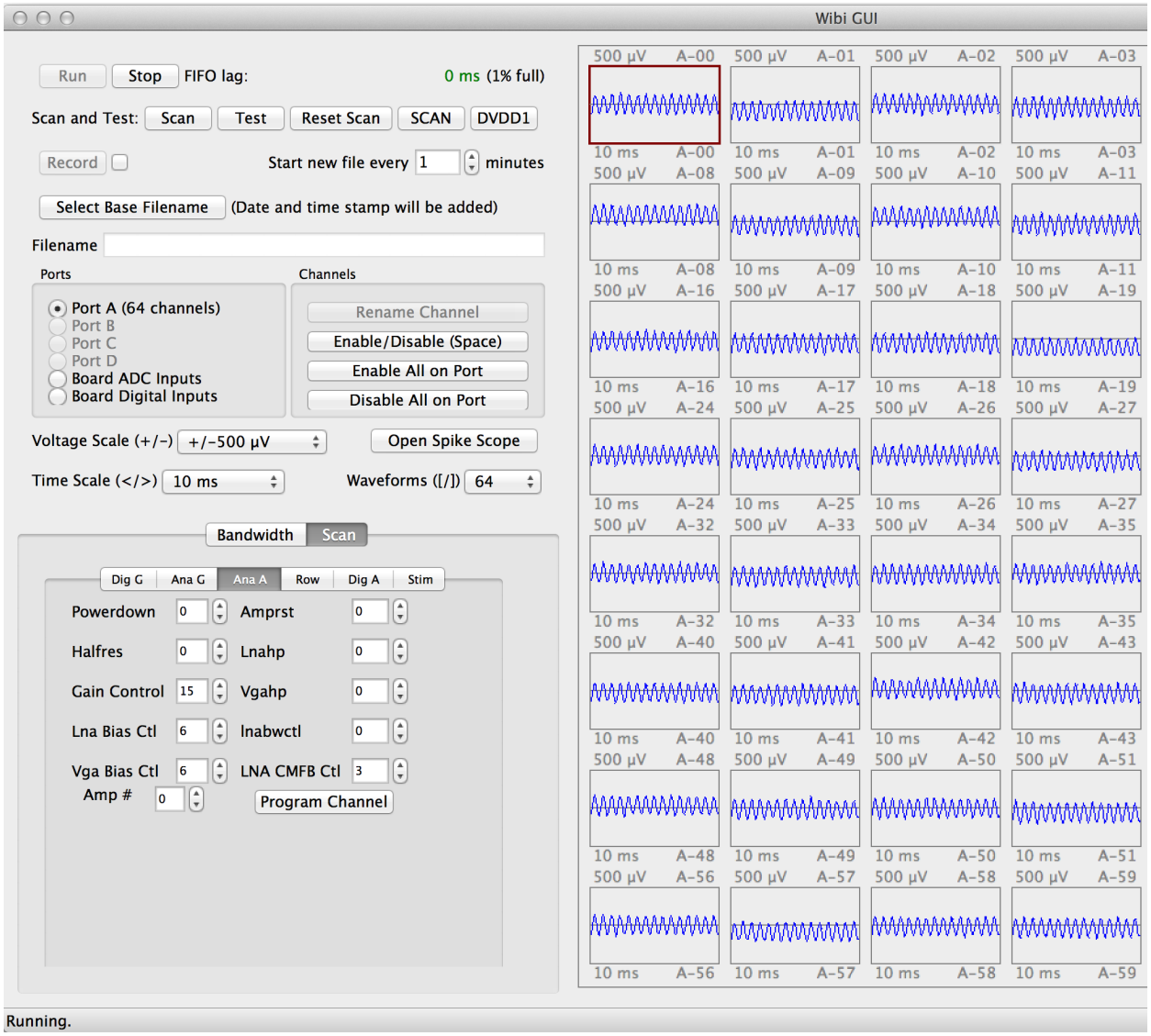


Figure 4.28: A screen capture of the GUI with test signals present. The GUI is able to display waveforms in real time as well as log data to a file.

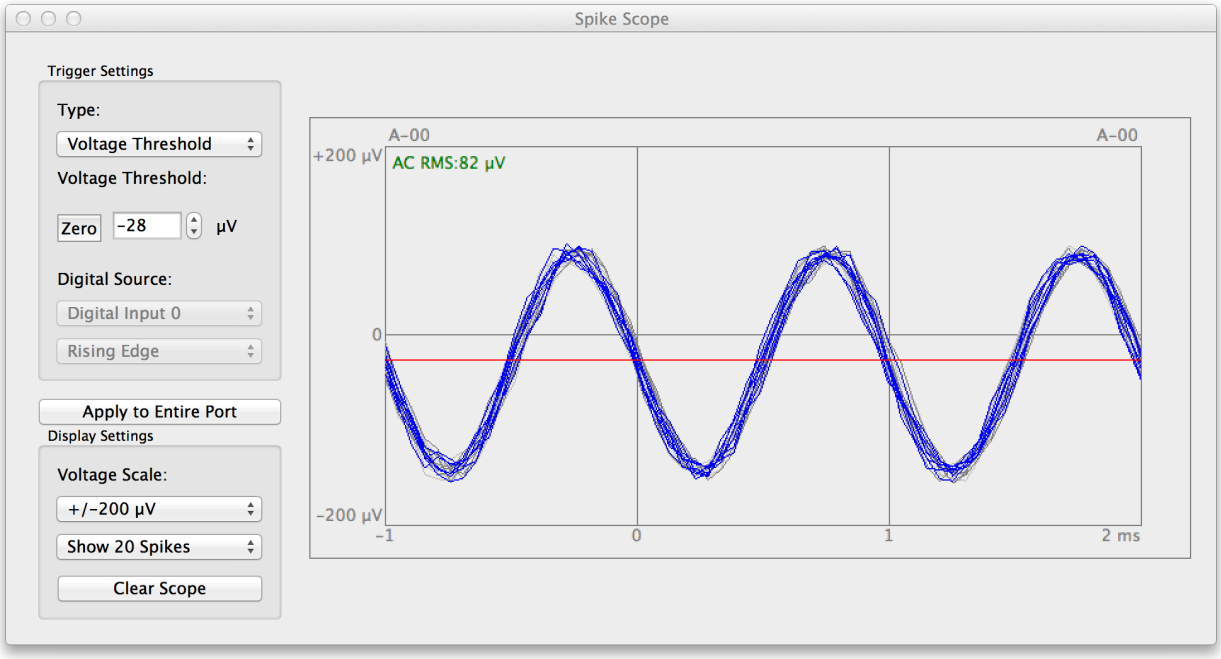


Figure 4.29: A screen capture of the GUI in overlay mode with test signals present. The GUI is able to display detected spike overlays using data from either the chip epochs or from full-streams using NEO or threshold-based software spike detection.

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. We also verified that the characteristics of the inter-spike interval distributions were close to Poisson and exhibited a clear absolute refractory period.

4.4.2 In Vivo Measurements

A typical subset of recorded *in vivo* data is shown in Fig. 4.30, 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 aligned in time, as displayed in Fig. 4.31. 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. As

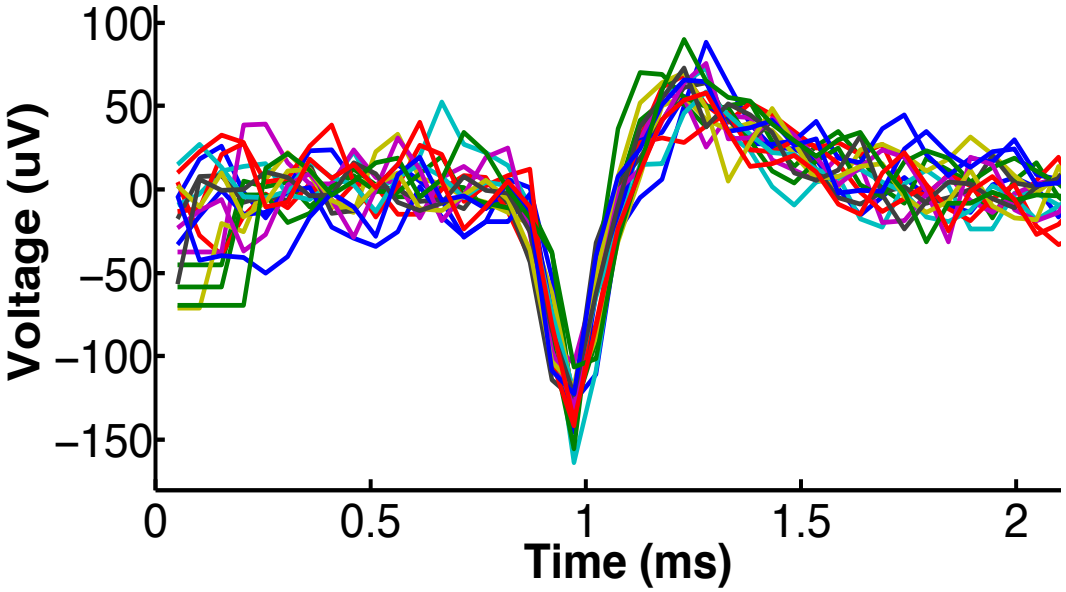


Figure 4.30: Time-aligned epochs recorded from one channel of *in vivo* neural data.

described previously, the SoC computes firing rates over a programmable window of time, which in this case was 26.2ms.

Fig. 4.31 illustrates the dramatic differences in data rate between the compression modes. With a conservatively-high average firing rate of 50 Hz on each channel, firing rates and epochs provides compression ratios of 700x and 8.3x, respectively. With an average firing rate of 50Hz per channel, the total digital power is 77.63 µW for firing rates and 113.6 µW for epochs. These results are significant because firing rates have been demonstrated to be sufficient for BMI control [85].

With a challenging 4.3 dB SNR, the detector achieves a 93.5% true positive (TP) rate and a 0.1% false positive (FP) rate. At 20 kHz sampling rate, this FP rate creates an additional, false 20 Hz background firing rate. This FP rate can be further reduced by sacrificing TP detection rate. If epoch compression mode is used, software post-processing can eliminate these false detections. See [74] for an illustration of this SNR, and more details on the detection algorithm and results.

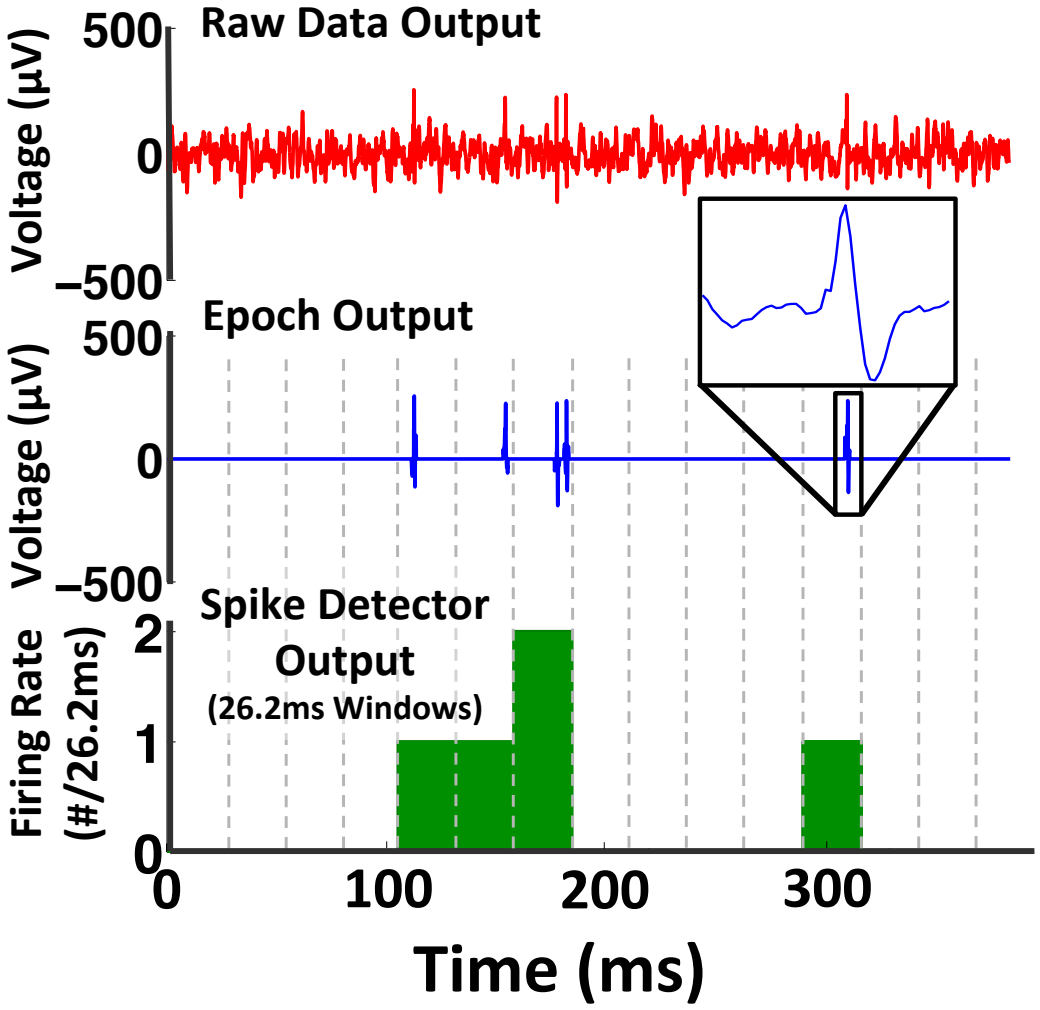


Figure 4.31: Raw streams, epochs, and firing rates of *in vivo* recorded data.

4.5 Conclusion

The key metrics of the design are summarized in Table 4.3 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 [18] while achieving comparable NEF and PEF. The key enabling factors in achieving low area included use of MIM capacitors over the active circuits to implement the LNA AC coupling and careful design and optimization of the remaining feedback capacitors using low overhead, low-valued MOM capacitors. The key factor in achieving low supply voltage, and thus low power consumption, was a precise replica biasing approach. Lastly, current consumption was optimized through several approaches. The LNA, which dominates the total power budget, was designed based on a very efficient existing design [50]. Other topologies in the literature, especially current-reuse [54, 60, 56], were

Table 4.3: System Comparison

System Specs.	This Work	Azin [18]	Rhew [61]	Chen [19]	Units
Technology	65	350	180	350	nm
VDD	1.2[a]	1.5	1.8	5.0	V
Off-Chip	None	1 µF	DC-DC	Amp IC	
Amp/ADC					
Channels	64	8	4	16	
Power[b]	1.81	25.8	61.25	N/A	µW
Area[b]	0.0258	0.3122	0.354	N/A	mm^2
Gain	45-65	51-65.6	54	N/A	dB
LP	10-1000	1-525	700	N/A	Hz
HP	3-8	5-12	6	N/A	kHz
Noise	7.5	3.12	N/R	N/A	µVrms
NEF	3.6	2.9	N/R	N/A	
PEF	12.9	12.6	N/R	N/A	
Stimulation					
Channels	8	8	8	8	
I_{max}[c]	0.9	0.0945	2x 4.2 6x 0.116	6.25	mA
Area[c]	0.0169	0.038	0.05	0.7	mm^2
Digital					
Power[b]	1.21[d]	3.28	34.5	256.9	µW
Area[d]	0.0105	0.0676	0.8	0.191	mm^2
Outputs					
Raw ADC	Yes	1 Ch.	Yes	Yes	
Epochs	Yes	No	Yes	Yes	
Events	No	Yes[e]	No	Yes	
Firing Rates	Yes	No	No	No	
Other	-		LFP	PCA[f]	

[a] 1.2 V unregulated input voltage; 1.0 V analog supply, 0.8 V digital supply.
[b] Per recording channel.
[c] Per stimulation site.
[d] 1.21 µW (Firing Rates) & 1.775 µW (Epochs).
[e] 1 raw channel and its discrimination events, or 8 channels of discrimination events.
[f] Outputs can include detected events, the extracted features, or the classification results.

found to have excessive flicker noise contributions in our process given a 10 pF constraint on the AC coupling capacitors (which limits the input transistor sizes). Other important techniques to reduce current consumption included distributed buffer sampling capacitors and optimized ADC unit capacitors which result in low area and low sampling capacitance of 260 fF. This in turn reduces the required drive power by the preceding circuits.

This SoC also integrates two fully-on-chip current stimulators with 8 V compliance voltage and 8 selectable, differential stimulation sites. This design reduces the total stimulator area per site by 2.25x compared to [18]. A power saving adiabatic architecture recycles charge to efficiently drive the high capacitance of typical stimulation probes. The stimulator can also drive LEDs for ontogenetic stimulation, but charge recycling benefits are lost. A coauthor, William Biederman, was responsible for the stimulation circuit, and further details are found in [67].

Finally, the SoC also integrates custom logic for 64-channel spike detection and compression. This block consumes 2.7x less power and 6.4x less area per channel compared to [18] while implementing more features. Three operating modes are selectable on a per-channel basis, including raw streaming data, epochs, and firing rates. An on-chip FIFO eases buffering requirements on an aggregator or radio. A coauthor, Nathan Narevsky, was responsible for the compression logic, and further details are found in [67].

Finally, the SoC was integrated onto a low power wired headstage for evaluation. System integration and facilitation of animal recordings was largely the work of a coauthor, Jaclyn Leverett. Successful capture of *in vivo* measurements from an awake rodent require a robust, tightly-integrated PCB, both mechanically and electrically. Two problems are common when attempting live recordings: interference from 60 Hz and 120 Hz sources and motion artifacts. Interference can originate from computers, lighting, and power supplies, which couple onto high impedance signal lines. Use of battery power (as opposed to 60 Hz power supplies) and careful wiring to avoid ground loops are helpful to avoid pickup of interference. Motion artifacts are signals induced from capacitive coupling onto high-impedance signal wires - typically between the probes and recording channels. To minimize motion artifacts, a small headstage that avoids creating torque on the implant connector, as well as a flexible, compliant cable for wired testing, are key.

Lastly, and perhaps most importantly, strong teamwork and leadership were essential in the realization of this SoC. Three state-of-the-art subsystems (recording, compression, stimulation) were integrated together, requiring collaboration, communication, cross-validation, co-simulation, and testing. Furthermore, there is a substantial delta in work between a single amplifier and a full neuromodulation SoC. Amortization of non-glamorous layout, system infrastructure, and testing hardware/software helped to lessen the burden of system integration. Finally, a shared vision for the system and its potential impact also helped drive the project. The high integration level, low power consumption, and compact size of this system, provide the next step in enabling fully-implanted, wirelessly-powered, high-density neural interfaces in the human body.

Chapter 5

Conclusion

The first generation of BMI, including cochlear implants and deep-brain stimulation, have radically improved the lives of the deaf and those suffering from Parkinson's disease. Clinical treatments for diseases and injuries resulting in paralysis represent the second generation of BMI. Neuroscientists have now demonstrated working, brain-controlled robotic prostheses [13, 14]. Transforming these research efforts into deployable clinical tools requires contributions in many fields, including neural recording electronics. This dissertation presents two neural interface systems to address two key challenges: evading the brain's foreign body response to achieve long probe longevity, and scaling wireless, implantable systems to high channel counts.

The first system, presented in Chapter 3, explores the limits of miniaturization to increase probe longevity. Wireless power delivery, bidirectional communication, 4 acquisition channels, and an antenna are squeezed into 0.125 mm^2. This enables free-floating neural probes, which are free of micro-motion effects from wired tethers. Due to the high wireless power delivery losses, this system aims to provide neuroscientists with a limited number of high-longevity neural channels. However, the combined losses of two wireless power links (one through the skull, one through the dura) will prohibit use in BMI systems with 1000's of channels. Furthermore, the open-loop gain and DC-coupled front end do not lend themselves to neuroscience research or clinical applications for reasons discussed in Chapter 4.

The second system, discussed in Chapter 4, targets the high channel counts required to control complex, high-DOF robotic prostheses from a wireless neural implant. This in turn requires low area, low power, and low data rate per neural recording channel. To meet these demands, 64 recording channels, neural data compression, bias generation, and power conditioning are integrated onto a single 4.78 mm^2 IC in under 500 μW of power. This enables systems that scale up to thousands of channels distributed across the brain, or scale down to extremely compact, low weight, low area, wireless interfaces.

5.1 Future Work

Substantial work remains before this research can impact the medical domain. This productization may be better left to industry as the complexity and scope reach beyond what a graduate student can readily tackle. For example, an aggregation IC is needed to collect firing rates from an array of recording ICs and to coordinate stimulation. Integration of a microprocessor, such as an ARM M0, on the aggregation IC will enable customization of the BMI solution for each patient. It will also provide neuroscientists a programmable platform to experiment with closed-loop algorithms to treat Parkinson's disease and potentially other diseases such as epilepsy. Furthermore, either a subdermal wire or a wireless link to an externally-worn interrogator / base station (see Fig. 4.1) is required to transfer control signals to the prosthesis from the aggregator. Current deep-brain stimulators wire power from a battery in the chest cavity, through the neck, to the stimulation probe in the brain tissue; a similar technique may provide the fastest, most near-term data link for clinical applications.

In addition to an aggregator, further work on neural probes is required to reach the an acceptable implant working life. In general, a brain surgery to replace neural probes is impractical; ideally, the probes will outlast the patient. Probe failure modes have been studied extensively [86]. Flexible tethers [87], miniaturized probes [88, 89], biocompatible materials [90], and stimulation [91] hold promise in the race to evade the biological response.

5.2 Acknowledgements

This work was supported by BWRC sponsors, MuSyC, NSF GRFP and NDSEG. IC fabrication was generously donated by the TSMC University Shuttle Program. Simulation and CAD were supported by Cadence and BDA.

I wish to thank the coauthors of "A Fully-Integrated, Miniaturized ($0.125mm^2$) $10.5\mu W$ Wireless Neural Sensor," William Biederman, Nathan Narevsky, Jackie Leverett, Aaron Koralek, Jose Carmena, Elad Alon, and Jan Rabaey. Similarly, I wish to thank the coauthors of "A $4.78mm^2$ Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels with Digital Compression and Simultaneous Dual Stimulation," William Biederman, Nathan Narevsky, Jackie Leverett, Ryan Neely, Jose Carmena, Elad Alon, and Jan Rabaey. Finally, I wish to thank Andreas Mehmanna, Prof. Jeremy Holleman, and Rikky Muller for their contributions and feedback.

Bibliography

[1] GJ Siegel et al. *Basic Neurochemistry: Molecular, Cellular and Medical Aspects*. 2006.

[2] Liebgott. *A, motor neuron*. 2011. URL: http://content.answcdn.com/main/content/img/elsevier/dental/f0429-01.jpg.

[3] Fabio Benfenati. "Synaptic plasticity and the neurobiology of learning and memory". eng. In: *Acta Bio-Medica: Atenei Parmensis* 78 Suppl 1 (2007), pp. 58–66. ISSN: 0392-4203.

[4] Alvaro Pascual-Leone et al. "The Plastic Human Brain Cortex". In: *Annual Review of Neuroscience* 28.1 (2005), pp. 377–401. DOI: 10.1146/annurev.neuro.27.070203.144216.

[5] R. Heliot et al. "Learning in Closed-Loop Brain-Machine Interfaces: Modeling and Experimental Validation". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 40.5 (Oct. 2010), pp. 1387–1397. ISSN: 1083-4419. DOI: 10.1109/TSMCB.2009.2036931.

[6] S. Ogawa et al. "Brain magnetic resonance imaging with contrast dependent on blood oxygenation". en. In: *Proceedings of the National Academy of Sciences* 87.24 (Dec. 1990), pp. 9868–9872. ISSN: 0027-8424, 1091-6490.

[7] L F Haas. "Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography". In: *Journal of Neurology, Neurosurgery & Psychiatry* 74.1 (Jan. 2003), pp. 9–9. ISSN: 00223050. DOI: 10.1136/jnnp.74.1.9.

[8] Scott A. Huettel. *Functional magnetic resonance imaging*. 2nd ed. Sunderland, Mass: Sinauer Associates, 2008. ISBN: 9780878932863.

[9] I. Feinberg and I. G. Campbell. "Longitudinal sleep EEG trajectories indicate complex patterns of adolescent brain maturation". en. In: *AJP: Regulatory, Integrative and Comparative Physiology* 304.4 (Feb. 2013), R296–R303. ISSN: 0363-6119, 1522-1490. DOI: 10.1152/ajpregu.00422.2012.

[10] Kevin Whittingstall and Nikos K. Logothetis. "Frequency-Band Coupling in Surface EEG Reflects Spiking Activity in Monkey Visual Cortex". In: *Neuron* 64.2 (Oct. 2009), pp. 281–289. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2009.08.016.

[11] Brian N. Pasley et al. "Reconstructing Speech from Human Auditory Cortex". In: *PLoS Biol* 10.1 (Jan. 2012), e1001251. DOI: 10.1371/journal.pbio.1001251.

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

[13] Leigh R Hochberg et al. "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm". In: *Nature* 485.7398 (2012), pp. 372–375.

[14] Jennifer L Collinger et al. "High-performance neuroprosthetic control by an individual with tetraplegia". In: *The Lancet* 381.9866 (2013), pp. 557–564.

[15] S. Venkatraman et al. "A System for Neural Recording and Closed-Loop Intracortical Microstimulation in Awake Rodents". In: *IEEE Transactions on Biomedical Engineering* 56.1 (Jan. 2009), pp. 15–22. ISSN: 0018-9294. DOI: 10.1109/TBME.2008.2005944.

[16] Ian H. Stevenson et al. "Statistical assessment of the stability of neural movement representations". en. In: *Journal of Neurophysiology* 106.2 (Aug. 2011), pp. 764–774. DOI: 10.1152/jn.00626.2010. URL: http://jn.physiology.org/content/106/2/764 (visited on 05/02/2014).

[17] 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.

[18] 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.

[19] 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.

[20] U. Frey et al. "Depth recording capabilities of planar high-density microelectrode arrays". In: *4th International IEEE/EMBS Conference on Neural Engineering, 2009. NER '09*. Apr. 2009, pp. 207–210. DOI: 10.1109/NER.2009.5109270.

[21] Bruce P. Bean. "The action potential in mammalian central neurons". en. In: *Nature Reviews Neuroscience* 8.6 (June 2007), pp. 451–465. ISSN: 1471-003X. DOI: 10.1038/nrn2148.

[22] George W Fraser et al. "Control of a brain-computer interface without spike sorting". eng. In: *Journal of neural engineering* 6.5 (Oct. 2009). PMID: 19721186, p. 055004. ISSN: 1741-2552. DOI: 10.1088/1741-2560/6/5/055004.

[23] Peter J. Ifft et al. "A Brain-Machine Interface Enables Bimanual Arm Movements in Monkeys". en. In: *Science Translational Medicine* 5.210 (Nov. 2013). PMID: 24197735, 210ra154–210ra154. ISSN: 1946-6234, 1946-6242. DOI: 10.1126/scitranslmed.3006159.

[24] Patrick J. Lynch. *Skull and brain normal human diagram*. Creative Commons Attribution 2.5 License, Dec. 2006. URL: http://creativecommons.org/licenses/by/2.5/.

[25] James M. Harrison and Peter M. Seligman. "Power transfer for implanted prostheses". US4741339 A. May 1988. (Visited on 12/10/2014).

[26] J. Petrovic and J. Trenary. "Rechargeable electric portable appliance". US4374354 A. Feb. 1983. (Visited on 12/10/2014).

[27] R. Want. "An introduction to RFID technology". In: *IEEE Pervasive Computing* 5.1 (Jan. 2006), pp. 25–33. ISSN: 1536-1268. DOI: 10.1109/MPRV.2006.2.

[28] 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.

[29] M. Zargham and P.G. Gulak. "Maximum Achievable Efficiency in Near-Field Coupled Power-Transfer Systems". In: *IEEE Transactions on Biomedical Circuits and Systems* 6.3 (June 2012), pp. 228–245. ISSN: 1932-4545. DOI: 10.1109/TBCAS.2011.2174794.

[30] M.W. Baker and R. Sarpeshkar. "Feedback Analysis and Design of RF Power Links for Low-Power Bionic Systems". In: *IEEE Transactions on Biomedical Circuits and Systems* 1.1 (Mar. 2007), pp. 28–38. ISSN: 1932-4545. DOI: 10.1109/TBCAS.2007.893180.

[31] C. Sauer et al. "Power harvesting and telemetry in CMOS for implanted devices". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 52.12 (Dec. 2005), pp. 2605–2613. ISSN: 1549-8328. DOI: 10.1109/TCSI.2005.858183.

[32] S. Kim et al. "Polymer based thin film coils as a power module of wireless neural interfaces". In: *2006 IEEE Workshop on Microelectronics and Electron Devices, 2006. WMED '06*. 2006, pages. DOI: 10.1109/WMED.2006.1678285.

[33] Patrick D. Wolf. "Thermal Considerations for the Design of an Implanted Cortical Brain-Machine Interface (BMI)". eng. In: *Indwelling Neural Implants: Strategies for Contending with the In Vivo Environment*. Ed. by William M Reichert. Boca Raton (FL): CRC Press, 2008. ISBN: 9780849393624.

[34] S. Kim et al. "Thermal Impact of an Active 3-D Microelectrode Array Implanted in the Brain". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 15.4 (Dec. 2007), pages. ISSN: 1534-4320. DOI: 10.1109/TNSRE.2007.908429.

[35] S. Mandal and R. Sarpeshkar. "Power-Efficient Impedance-Modulation Wireless Data Links for Biomedical Implants". In: *IEEE Transactions on Biomedical Circuits and Systems* 2.4 (Dec. 2008), pp. 301–315. ISSN: 1932-4545. DOI: 10.1109/TBCAS.2008.2005295.

[36] Timothy J Blanche et al. "Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording". eng. In: *Journal of neurophysiology* 93.5 (May 2005), pp. 2987–3000. ISSN: 0022-3077. DOI: 10.1152/jn.01023.2004.

[37] G.E. Perlin and K.D. Wise. "An Ultra Compact Integrated Front End for Wireless Neural Recording Microsystems". In: *Journal of Microelectromechanical Systems* 19.6 (Dec. 2010), pp. 1409–1421. ISSN: 1057-7157. DOI: 10.1109/JMEMS.2010.2082496.

[38] R.J. Vetter et al. "Development of a Microscale Implantable Neural Interface (MINI) Probe System". In: *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. Jan. 2005, pp. 7341–7344. DOI: 10.1109/IEMBS.2005.1616207.

[39] Chenyang Xu, William Lemon, and Chang Liu. "Design and fabrication of a high-density metal microelectrode array for neural recording". In: *Sensors and Actuators A: Physical* 96.1 (Jan. 2002), pp. 78–85. ISSN: 0924-4247. DOI: 10.1016/S0924-4247(01)00766-X.

[40] P.K. Campbell et al. "A silicon-based, three-dimensional neural interface: manufacturing processes for an intracortical electrode array". In: *IEEE Transactions on Biomedical Engineering* 38.8 (Aug. 1991), pp. 758–768. ISSN: 0018-9294. DOI: 10.1109/10.83588.

[41] 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.

[42] M. Schuettler. "Electrochemical Properties of Platinum Electrodes in Vitro: Comparison of Six Different Surface Qualities". In: *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS 2007*. Aug. 2007, pp. 186–189. DOI: 10.1109/IEMBS.2007.4352254.

[43] Otto J. Prohaska et al. "Thin-Film Multiple Electrode Probes: Possibilities and Limitations". In: *IEEE Transactions on Biomedical Engineering* BME-33.2 (Feb. 1986), pp. 223–229. ISSN: 0018-9294. DOI: 10.1109/TBME.1986.325894.

[44] Carolina Mora López. "Design of Low-Noise Integrated Circuits for In Vivo Neural Interfaces". nl. PhD thesis. KU Leuven, Oct. 2012. URL: https://lirias.kuleuven.be/handle/123456789/359241.

[45] Thomas Jochum, Timothy Denison, and Patrick Wolf. "Integrated circuit amplifiers for multi-electrode intracortical recording". In: *Journal of Neural Engineering* 6.1 (Feb. 2009), p. 012001. ISSN: 1741-2560, 1741-2552.

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

[47] K. Wise and J. Angell. "A microprobe with integrated amplifiers for neurophysiology". In: *Solid-State Circuits Conference. Digest of Technical Papers. 1971 IEEE International*. Vol. XIV. Feb. 1971, pp. 100–101.

[48] R.R. Harrison. "A low-power, low-noise CMOS amplifier for neural recording applications". In: *IEEE International Symposium on Circuits and Systems, 2002. ISCAS 2002*. Vol. 5. 2002, pp. 197–200. DOI: 10.1109/ISCAS.2002.1010674.

[49] M. S. J. Steyart and W. M. C Sansen. "A micro power low-noise monolithic instrumentation amplifier for medical purposes". In: *Solid-State Circuits, IEEE Journal of* 22.6 (1987), pp. 1163–1168.

[50] 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.

[51] R. Muller, S. Gambini, and J.M. Rabaey. "A 0.013mm2 5 μW DC-coupled neural signal acquisition IC with 0.5V supply". In: *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*. Feb. 2011, pp. 302–304.

[52] Stuart F. Cogan. "Neural stimulation and recording electrodes". eng. In: *Annual Review of Biomedical Engineering* 10 (2008), pp. 275–309. ISSN: 1523-9829. DOI: 10.1146/annurev.bioeng.10.061807.160518.

[53] T Knöpfel and E Boyden. "A comprehensive concept of optogenetics". In: *Optogenetics: Tools for Controlling and Monitoring Neuronal Activity* 196 (2012), p. 1.

[54] Shailesh Rai et al. "A 500μW neural tag with 2μVrms AFE and frequency-multiplying MICS/ISM FSK transmitter". In: *Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*. IEEE. 2009, pp. 212–213.

[55] W. Wattanapanitch and R. Sarpeshkar. "A Low-Power 32-Channel Digitally Programmable Neural Recording Integrated Circuit". In: *IEEE Transactions on Biomedical Circuits and Systems* 5.6 (Dec. 2011), pp. 592–602. ISSN: 1932-4545. DOI: 10.1109/TBCAS.2011.2163404.

[56] 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.

[57] 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.

[58] P. Harpe et al. "A 7-to-10b 0-to-4MS/s flexible SAR ADC with 6.5-to-16fJ/conversion-step". In: *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*. Feb. 2012, pp. 472–474. DOI: 10.1109/ISSCC.2012.6177096.

[59] P. Harpe, E. Cantatore, and A. van Roermund. "A 10b/12b 40 kS/s SAR ADC With Data-Driven Noise Reduction Achieving up to 10.1b ENOB at 2.2 fJ/Conversion-Step". In: *IEEE Journal of Solid-State Circuits* 48.12 (Dec. 2013), pp. 3011–3018. ISSN: 0018-9200. DOI: 10.1109/JSSC.2013.2278471.

[60] 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.

[61] 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.

[62] 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.

[63] A.M. Sodagar, K.D. Wise, and K. Najafi. "A fully-integrated mixed-signal neural processing module for implantable multi-channel cortical recording". In: *IEEE Biomedical Circuits and Systems Conference, 2006. BioCAS 2006.* Nov. 2006, pp. 37–40. DOI: 10.1109/BIOCAS.2006.4600302.

[64] 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.

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

[66] D Yeager et al. "A fully-integrated 10.5μW miniaturized (0.125 mm^2) wireless neural sensor". In: *VLSI Circuits, 2012 Symposium on.* IEEE. 2012, pp. 72–73.

[67] W. Biederman et al. "A Fully-Integrated, Miniaturized (0.125 mm^2) 10.5 μW Wireless Neural Sensor". In: vol. 48. 4. Apr. 2013, pp. 960–970.

[68] 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.

[69] 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.

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

[71] L. Lincoln, K. Leung, and H.C. Luong. "A 7-μW clock generator in 0.18-μm CMOS for passive UHF RFID EPC G2 tags". In: *Solid State Circuits Conference, 2007. ESSCIRC 2007. 33rd European.* Sept. 2007, pp. 412–415.

[72] 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.

[73] J. Montanaro et al. "A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor". In: *IEEE Journal of Solid-State Circuits* 31.11 (Nov. 1996), pp. 1703–1714. ISSN: 0018-9200. DOI: 10.1109/JSSC.1996.542315.

[74] 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, pp. 1–4.

[75] 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.

[76] 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.

[77] David A. Borton et al. "An implantable wireless neural interface for recording cortical circuit dynamics in moving primates". en. In: *Journal of Neural Engineering* 10.2 (Apr. 2013), p. 026010. ISSN: 1741-2552. DOI: 10.1088/1741-2560/10/2/026010.

[78] Julie Dethier et al. "Design and validation of a real-time spiking-neural-network decoder for brain-machine interfaces". eng. In: *Journal of neural engineering* 10.3 (June 2013), p. 036008. ISSN: 1741-2552. DOI: 10.1088/1741-2560/10/3/036008.

[79] 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.

[80] D Yeager et al. "A 4.78mm^2 Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels with Digital Compression and Simultaneous Dual Stimulation". In: *VLSI Circuits, 2014 Symposium on*. IEEE. 2014.

[81] W. Biederman et al. "A 4.78mm^2 Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels with Digital Compression and Simultaneous Dual Stimulation". In: *Solid-State Circuits, IEEE Journal of* (To Appear).

[82] Jaclyn Leverett. "A Low-Power, Lightweight, Wireless Neural Recording and Stimulating Headstage for Brain Machine Interfaces". In: *U.C. Berkeley Master's Report* (2014).

[83] Anna Balazs. "International Vocabulary of Metrology - Basic and General Concepts and Associated Terms". In: *CHEMISTRY International* (2008).

[84] Intan Technologies. *RHD2000 Open Source Evaluation Software*. Accessed: 2014-09-30. URL: http://www.intantech.com/RHD2000_evaluation_system.html.

[85] Johan Wessberg et al. "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates". In: *Nature* 408.6810 (2000), pp. 361–365.

[86] Abhishek Prasad et al. "Comprehensive characterization and failure modes of tungsten microwire arrays in chronic neural implants". eng. In: *Journal of Neural Engineering* 9.5 (Oct. 2012), p. 056015. ISSN: 1741-2552. DOI: 10.1088/1741-2560/9/5/056015.

[87] Harbaljit S. Sohal et al. "The sinusoidal probe: a new approach to improve electrode longevity". In: *Frontiers in Neuroengineering* 7 (2014), p. 10. DOI: 10.3389/fneng.2014.00010.

[88] John P. Seymour and Daryl R. Kipke. "Fabrication of polymer neural probes with sub-cellular features for reduced tissue encapsulation". eng. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 1 (2006), pp. 4606–4609. ISSN: 1557-170X. DOI: 10.1109/IEMBS.2006.260528.

[89] John P. Seymour and Daryl R. Kipke. "Neural probe design for reduced tissue encapsulation in CNS". eng. In: *Biomaterials* 28.25 (Sept. 2007), pp. 3594–3607. ISSN: 0142-9612. DOI: 10.1016/j.biomaterials.2007.03.024.

[90] Amarjit Singh et al. "Glial cell and fibroblast cytotoxicity study on plasma-deposited diamond-like carbon coatings". eng. In: *Biomaterials* 24.28 (Dec. 2003), pp. 5083–5089. ISSN: 0142-9612.

[91] Kevin J. Otto, Matthew D. Johnson, and Daryl R. Kipke. "Voltage pulses change neural interface properties and improve unit recordings with chronically implanted microelectrodes". eng. In: *IEEE transactions on bio-medical engineering* 53.2 (Feb. 2006), pp. 333–340. ISSN: 0018-9294. DOI: 10.1109/TBME.2005.862530.