Ultra-low Power Circuit and System Design for Deployable Body-worn Devices

Fan Zhang

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:
Brian Otis, Chair
Jacques C. Rudell
Yasunori Miyahara

Program Authorized to Offer Degree:
Department of Electrical Engineering
University of Washington

Abstract

Ultra-low Power Circuit and System Design for Deployable Body-worn Devices

Fan Zhang

Chair of the Supervisory Committee:
Associate Professor Brian Otis
Department of Electrical Engineering

Recent advances in ultra-low power chip design techniques, many originally targeting wireless sensor networks, will enable a new generation of body-worn devices for health monitoring. Off-the-shelf chips often consume too much power and are usually too bulky for body-worn applications. In this thesis, I have contributed in advancing the state-of-the-art ultra-low power circuit design techniques, including low-noise bioamplifiers for ExG (e.g. ECG, EMG, EEG, etc.) and neural applications, analog signal processors targeted for but not limited to ECoG applications, transmitter and receiver for short-range wireless. In addition, I have also helped realizing the next-generation wireless sensing system ICs that aim at prolonging battery life, even eliminating the battery leveraging emerging energy-harvesting solutions. Lastly, a few works presented in the thesis (such as the “Bumblebe”, “SoCWISP”) have not only been published as literatures, but also deployed at various labs assisting scientists to perform experiments that are difficult if not impossible to be done otherwise.
# TABLE OF CONTENTS

List of Figures .................................................. iii

Glossary ............................................................. ix

Chapter 1:  Introduction: enabling next-generation wireless sensors ............. 1

Chapter 2:  Analog circuit innovations for biosignal acquisition ................. 6
  2.1 State-of-the-art biopotential amplifiers .................................. 7
  2.2 Design of low-noise biopotential amplifiers ............................ 8
  2.3 Measurement results of the biopotential amplifiers .................... 18
  2.4 System-Level Implementation ........................................... 25
  2.5 in-vivo testing ....................................................... 29
  2.6 Discussion .................................................................. 34
  2.7 Chopper-Stabilized Low-Noise Amplifiers ............................. 36
  2.8 Conclusion ................................................................ 40

Chapter 3:  Analog signal processing on ECoG signals ............................. 41
  3.1 Background on ECoG signals ........................................... 41
  3.2 System Architecture ..................................................... 43
  3.3 Circuit Design .......................................................... 48
  3.4 Experimental results ................................................... 58
  3.5 in-vivo experiment results .............................................. 64
  3.6 Discussion ................................................................ 67
  3.7 Conclusion ................................................................ 72

Chapter 4:  Wafer-scale packaging for FBAR-based Oscillators .................. 73
  4.1 The first “FMOS” ....................................................... 74
  4.2 A low jitter fully-differential “FMOS” oscillator ....................... 81
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>2012 US budget proposal diagram.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>A hypothetical wireless body-area network (WBAN).</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>A generic block diagram for a biopotential-recording system.</td>
<td>3</td>
</tr>
<tr>
<td>1.4</td>
<td>A wireless sensing platform made from commercial-of-the-shelf (COTS) parts.</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Schematics of three low-noise biopotential amplifier designs.</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>a) Differential-mode gain path; b) CMFB gain path.</td>
<td>16</td>
</tr>
<tr>
<td>2.3</td>
<td>Closed-loop amplifier start-up concern alleviated by adding a leakage path</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>through a diode-connected transistor at the first-stage output.</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>Microphotographs of BPA1, BPA2, and BPA3.</td>
<td>18</td>
</tr>
<tr>
<td>2.5</td>
<td>Top: Bode magnitude comparison of the three BPAs; Bottom: Measured</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>input-referred noise comparison of the three BPAs.</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>Top: PSRR bode magnitude comparison of the three BPAs; bottom: CMRR</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>magnitude comparison of BPA1 and BPA3.</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>Performance comparison of state-of-the-art biopotential amplifiers.</td>
<td>25</td>
</tr>
<tr>
<td>2.8</td>
<td>Top-level block diagram of the 500 µW neural interface system IC.</td>
<td>26</td>
</tr>
<tr>
<td>2.9</td>
<td>a) The closed-loop schematic of the VGA and the Gm cell; b) VGA double-</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>folded cascode OTA schematic.</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>a) The Bumblebee, a miniaturized, ultra lightweight, low power wireless</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>sensor platform; b) the functional block diagram of the Bumblebee; c) the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>block diagram of the custom receiver.</td>
<td></td>
</tr>
<tr>
<td>2.11</td>
<td>BPA3 tested in vivo in rat motor cortex. a) Recorded rat spike; b) two</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>classes of spikes sorted by post-processing programs.</td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>Neural recording from a mouse visual cortex through two adjacent electrodes.</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>The top is recorded through a rack-mount wired setup, while the bottom</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reconstructed from the Bumblebee in real time.</td>
<td></td>
</tr>
<tr>
<td>2.13</td>
<td>a) EMG traces from human flexor and extensor regions simultaneously</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>captured, wirelessly transmitted by two Bumblebee chips, and reconstructed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>in real-time; b) human ECG captured, wirelessly transmitted by a Bumblebee,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and reconstructed.</td>
<td></td>
</tr>
<tr>
<td>2.14</td>
<td>Schematic of fully-differential chopper-stabilized low-noise amplifier</td>
<td>37</td>
</tr>
</tbody>
</table>
2.15 Chopper amplifier architecture ........................................ 37
2.16 Measured transfer function and noise plot of the low-noise chopper-stabilized biosignal amplifier ................................. 39

3.1 Common recording modalities and their respective signal characteristics ............................................................. 42
3.2 System architecture of an implantable EC0G/EEG recording unit to be used in BCI applications. It includes the proposed EC0G/EEG processing IC (EPIC), A/D converter, data and power telemetry ................................................................. 44
3.3 Architecture of one band of the proposed energy extractor ......................................................................................... 48
3.4 a) Schematic of a closed-loop chopper-stabilized amplifier and the OTA [141]; b) schematic of the transconductor; c) schematic of the VGA ............................................................................................. 49
3.5 Schematic of the 2nd-order SC biquad .................................................................................................................... 51
3.6 Schematic of the multiplier and the SC integrator ..................................................................................................... 53
3.7 a) The clock generation architecture. b) The design of a quadrature phase shifter .......................................................... 54
3.8 Microphotograph of the chip ......................................................................................................................................... 58
3.9 Measured frequency responses of one biquad: a) coarse $f_0$ tuning when $f_c$ is varied; b) fine $f_0$ tuning when $C_3$ is changed; c) Q tuning when $C_1$ is changed .................................................. 60
3.10 Top: measured frequency response (solid) compared to theoretical (dotted) of BPF tuned to $\alpha$, $\beta$, low-$\gamma$ and $\gamma$ bands; bottom: measured input-referred noise plots of LNA, VGA and BPF tuned to these four bands ................................................................. 61
3.11 a) Increased activity in 70-100 Hz range correlates to the onset of movement; b) theoretical (dashed) and measured (solid) energy profile in a) signal, $\alpha$, $\beta$, $\gamma$, d) Low-$\gamma$, c) $\gamma$ band for a 6-second ECoG waveform [80] ............................................................................................. 63
3.12 In-vivo experiment setup ........................................................................................................................................... 65
3.13 4-second in-vivo ECoG recording with the chip (red) and the commercial ADC (blue), and the error between the two. a) Segment I: $\beta$/low-$\gamma$ activity (top); error in time (bottom); b) Segment II: baseline activity (top); error in time (bottom); c) Error power in segment I; d) Error power in segment II ........................................ 66
3.14 PSD of segments I and II recorded with EPIC (solid) and the commercial ADC (dotted) ........................................................................................................................................... 68
3.15 An area/power efficient multi-channel ECoG system: a) raw-signal mode; b) sub-banded energy mode; c) proposed analog energy extractor ........................................................................................................ 69
3.16 Power law of ECoG signals ......................................................................................................................................... 70

4.1 Photograph of disassembled hermetic oscillator package containing A) FBAR die; B) lid with integrated active circuitry ....................................................................................................................... 74
4.2 Schematic description of normal FBAR/micro-cap process flow .................................................................................. 75
4.3 A) Schematic circuit lid flow; B) micro-capped oscillator die; C) cross-section SEM of FBAR/bipolar process showing completed die. .......................................................... 77
4.4 A) Schematic of Pierce oscillator with divider; B) Differential Colpitts oscillator with $g_m$-boosting. .......................................................... 78
4.5 A) Measured signal at 1.5 GHz RF output (top) and 23.4 MHz divide-by-64 output (bottom); B) Measured temperature response as measured by the ZDR frequency response, an integrated diode thermometer, and a reference diode placed 10 mm away from the oscillator. .......................................................... 79
4.6 A) Measured tuning curve of FBAR oscillator; B) Phase noise of the Pierce oscillator and a differential oscillator measured with an Agilent 5052 signal source analyzer. The 1/f corner for the Pierce is 800 Hz, while the differential Colpitts is 3 kHz. The phase noise of the divide-by-64 output indicates a 20-log(64) improvement in phase noise. .......................................................... 80
4.7 Wafer map showing functional oscillators (green) from one of the first chip-scale packaged oscillator wafers. .......................................................... 81
4.8 Schematic of the oscillator and buffer. .......................................................... 82
4.9 Measured oscillator current (left) and output power (right). ............ 83
4.10 A) Measured oscillator frequency drift as a function of temperature; (B) Measured oscillator frequency variation as a function of supply voltage. ............... 84
4.11 Phase noise of the oscillators with 3 resonators of 1.7, 2.6, and 3.4 GHz respectively, measured with an Agilent 5052 signal source analyzer. ............... 85
4.12 Measured tuning curve of ZDR oscillator. .......................................................... 86

5.1 High-level system architecture. Passive tags are remotely powered and perform sensing. Data is transmitted at a 1/3 frequency carrier derived from the RF power through injection locking. .......................................................... 89
5.2 Block diagram of the proposed transmitter tag architecture. .......................... 90
5.3 Schematic of the injection-locked divide-by-3 circuit. ................................................ 91
5.4 Phase relationship between the current and voltage components and small-signal equivalent circuit of the ILFD when in a) free running and b) injection-locked condition. .......................................................... 92
5.5 a) CAD drawing of the custom loop antenna and its dual-band matching network; b) Measured antenna input return loss. ................................................ 93
5.6 a) The chip micrograph of the transmitter tag; b) Measured input return loss. ................. 94
5.7 a) Measured rectifier efficiency vs. input power; b) Measured regulator output voltage vs. input voltage. .......................................................... 95
5.8 a) Spectrum of the free-running and injection-locked ILFD; b) Measured injection locking range. .......................................................... 96
5.9 a) Spectrum b) transient waveform of OOK-modulation at 4 Mbps. ................. 97

v
5.10 a) Wireless link test setup; b) Baseband OOK-modulation signal and received/demodulated signal at 20 kbps. ................................. 97

6.1 $V_{TH}$ as a function of the bulk-source forward biasing voltage $V_{BS}$. ................................. 101

6.2 Design metrics as a function of a low-$V_{TH}$ nFET transistor length for 1 mA and 0.5 mA $I_d$: a) $V_{TH}$ and $V_{gs}$; b) normalized transit frequency $f_T/f_0$; c) $g_m/I_d \cdot f_T/f_0$; d) intrinsic gain $g_m/g_{ds}$. ................................. 102

6.3 Design metrics as a function of a low-$V_{TH}$ pFET transistor length for: a) $V_{TH}$ and $V_{gs}$; b) $f_T/f_0$; c) $g_m/I_d \cdot f_T/f_0$; d) $g_m/g_{ds}$. ................................. 103

6.4 Design metrics as a function of the inversion coefficient for a low-$V_{TH}$ nFET transistor: a) $V_{TH}$ and $V_{gs}$; b) normalized transit frequency $f_T/f_0$; c) $g_m/I_d \cdot f_T/f_0$; d) intrinsic gain $g_m/g_{ds}$. ................................. 105

6.5 System block diagram of the proposed Low-IF receiver using transformer-coupled techniques. ................................. 107

6.6 a) The layout of a symmetrical bifilar planar transformer; b) A simplified transformer circuit model. ................................. 111

6.7 Schematic of the RF front-end (folded LNA, transformer-feedback LC VCO, and quasi-passive mixer). Also shown is the microphotograph of an on-chip transformer, used throughout the front-end. ................................. 113

6.8 a) Folded LNA using pFET cascode; b) proposed transformer-coupled folded LNA. ................................. 115

6.9 a) Conventional cross-coupled VCO; b) transformer-coupled VCO. ................................. 118

6.10 a) Active Gilbert-cell mixer; b) active switched-$g_m$ mixer; c) passive mixer; d) proposed quasi-passive mixer. ................................. 120

6.11 Calculated and simulated $R_{in}$ vs. duty-cycle ................................. 123

6.12 Calculated and simulated $R_{in}$ vs. duty-cycle ................................. 124

6.13 VCO loop gain and differential zero-to-peak voltage vs. duty-cycle ................................. 125

6.14 Schematic of the IF amplifier and filter chain. ................................. 127

6.15 Schematic of the IF amplifier and filter chain. ................................. 130

6.16 Measured performance of the LNA: a) $-S_{11}$; b) gain and noise figure; c) IIP3. ................................. 133

6.17 Measured VCO performance: a) tuning curve; b) phase noise (free-run and injection-locked). ................................. 134

6.18 Measured performance at the mixer output: a) gain and noise figure; b) IIP3. ................................. 134

6.19 Measured gain and NF performance at the IF output. ................................. 136

6.20 Measured filter response normalized to the IF frequency. ................................. 136

6.21 Test structures used for de-embedding: a) open; b) short. ................................. 137

6.22 The simulated and measured parameters of the transformers: a) L; b) Q; c) k. ................................. 138
6.23 Receiver system test: a) far-out selectivity; b) BER at 200 kbps; c) wireless transmitted and received bit sequence. ........................................ 140

6.24 Effect of mixer bias voltage on the measured VCO: a) average current; b) output power (after buffer); c) phase noise at 10 kHz frequency offset; d) phase noise at 1 MHz frequency offset. ........................................ 143

6.25 Effect of mixer bias voltage on the measured performance at the mixer output:
a) gain and noise figure; b) IIP3. ........................................ 145

6.26 Measured RF front-end performance comparison with the active mixer and quasi-passive mixer: a) gain; b) noise figure; c) IIP3; d) VCO phase noise. . . . 146

7.1 Neural spike streaming system block diagram. ........................................ 152

7.2 a) Low-noise analog front-end schematic; b) Transmitter edge combiner schematic details and the output match. ........................................ 153

7.3 a) Measured gain and noise response of LNA and VGA; b) Transmitter spectrum with FSK symbols overlaid. ........................................ 155

7.4 a) The microphotograph; b) Reconstructed measured data and system Performance summary. ........................................ 157

7.5 The block diagram of the ”Bumblebee” PCB. ........................................ 158

7.6 Companion receiver block diagram. ........................................ 159

7.7 Front and back of the Bumblebee board on a quarter. A thin 30 gauge monopole wire antenna a few cm long allows >10m range. ........................................ 160

7.8 Conceptual diagrams of the conventional BAN (top) and the proposed solution (bottom) with energy harvesting, integrated power management, and ultra-low power flexible DSP architecture [148]. ........................................ 163

7.9 System block diagram for the proposed chip comprising the energy harvesting/supply regulation, AFE, subthreshold DSP, and TX subsystems [148]. . . 164

7.10 Measured results from an on-body thermal-harvesting experiment with 4x4 cm$^2$ COTS TEG. ........................................ 165

7.11 a) Schematic of the energy harvesting section; b) Measured startup sequence for the chip. A RF pulse kick-starts $V_{BOOST}$, allowing the boost converter to turn on. ........................................ 167

7.12 a) Top level diagram of the supply regulation subsystem; b) Diagram of the switched-capacitor variable DC-DC converter. ........................................ 169

7.13 Diagram of the $V_{BOOST}$ monitoring path. ........................................ 170
7.14 a) An example that makes use of the ‘stoplight’ scheme. Here, the green threshold is set to $V_{CAP_{DIG}} = 205$. A subroutine in the yellow mode ensures timely processing of stored data when the chip recuperates from red mode. b) Measured closed-loop power management ‘stoplight’ scheme. $V_{BOOST}$ is sampled at 256Hz. The node automatically updates its ‘stoplight’ status according to the $V_{BOOST}$ value. ........................................ 172

7.15 a) Summary of DPM energy saving when compared to our general purpose MCU; b) diagram of the $V_{BOOST}$ monitoring path; c) Summary of the DPM ISA. .......................................................... 174

7.16 a) Block diagram for the proposed subthreshold data processing subsystem; b) Measured energy-delay curves for MCU, RR+AFib accelerator, and 30-tap, 1-channel FIR. ....................................................... 175

7.17 Integration of DPM and GPP MCU. DPM and MCU share the same memory, and instructions are steered to the appropriate processor. The MCU is power-gated when idle. Both efficient control of the node and generic processing are achieved. ......................................................... 177

7.18 Measured waveforms showing correct data reconstruction from wirelessly streamed data. ....................................................... 181

7.19 a) Measured system response in the R-R extraction mode, where the R-R interval information is wireless transmitted every 5s. Also shown is the measured accuracy of the results. b) Measured system AFib demo experiment using R-R extractor and AFib accelerator. Normal and AFib ECG waveforms are from MIT-BIH database. The last 8 beats of raw ECG are stored in DMEM and streamed over TX if AFib is detected. .............. 182

7.20 a) Current breakdown for R-R extraction mode; b) chip die photo. .............. 183

7.21 a) Measured performance summary; b) Performance comparison with state-of-the-art BAN nodes. ........................................ 184

8.1 Research summary showing the die photos of the chips that I taped out during my Ph.D. program. ........................................... 187
GLOSSARY

ADC: analog-to-digital converter.

AIN: aluminum nitride.

FBAR: thin-film bulk acoustic wave resonator.

FBAR VCO: voltage-controlled FBAR oscillator.

FOM: figure-of-merit.

ILFD: injection locking frequency divider.

IC: integrated circuit.

LC: inductor and capacitor.

LO: local oscillator.

MBVD: modified Butterworth Van-Dyke model.

MEMS: microelectromechanical systems.

PLL: phase-locked loop.

PVT: process, voltage, and temperature.

Q: quality factor.

RLC: resistor, inductor, and capacitor.
RF: radio frequency.

VCO: voltage-controlled oscillator.

XTAL: crystal.

ZDR: zero-drift resonator.

APE: analog front-end.

BASN: body area sensor network

ECOG: Electrocorticography

BCI: Brain-computer interface

BAN: Body-area network

BSN: Body sensor node

EEG: Electroencephalography

ECG: Electrocardiography

EMG: Electromyography

ECOG: Electrocorticography

EXG: A general term for EEG, ECG, EMG and ECG

AFIB: Atrial Fibrillation

TX: transmitter
RX: receiver

TRX: transceiver

XFMR: transformer

TF: transformer-feedback
ACKNOWLEDGMENTS

I still can’t believe that I’m at the end of my Ph.D. program. I was never sure that I wanted to do a Ph.D., let alone those times that I just wasn’t sure if I could make it. But here I am, a few days before my defense. I have to say that I can never make it if it’s not because of a few very special people in my life.

First, I’d like to acknowledge my adviser Brian Otis. He’s taught me so many things, from circuit design, to dealing with people, to philosophy on life. We share so many colorful memories together, toughing through tapeouts and testings, or even hanging out like friends. I vividly remember that during my first tapeout, he stayed late with us in the lab helping us drawing layouts and debugging LVS errors. Just like many people during tapeouts, I was extremely exhausted and stressed and before I know it, warm tear drops streamed down my cheeks. He pulled me aside, and told me gently but sincerely that everything would be okay. And he patted on my back and said, “let’s beat that XX guy!” (XX being our competitor at that time). Such simple words carried so much power and gave me so much energy. Little did I know, that was the beginning of my 11 tapeouts. Every tapeout was a struggle, but made memorable with his incessant support and encouragement.

There was another time that I failed to meet a tapeout deadline despite of my working around the clock for two weeks. I had excuses: the chip was too big, people gave me their blocks too late to integrate, etc. But I was devastated. I failed meeting a goal and I disappointed myself and other people including Brian. He asked me for a walk outside, in the beautiful summer afternoon filled with generous sunshine to which I paid no interest. “If it were a company, it’d go bankrupt; if it were a marriage, it’d be a divorce..” he said. After listing the serious repercussions that I could’ve incurred if this were not just a school tapeout, his tone became much softer. He started to reflect on the things that we could’ve done better. And even more, we started making plans for how to proceed with this project
and do a good job on the next run. In the end, he looked at me in the eyes and told me that I did not disappoint him, and he still believed in me. To me, that’s just what I needed from him.

Like many Ph.D. students, there’s a period when I lost my directions. I couldn’t see where my research was going, not mentioning graduation. I really wanted to explore RF circuits after spending three to four years working on analog circuits. However, change is never easy. I felt it’s too late into my program to make such a sharp turn in my research topic. The stake were too high, at least so I thought. Brian spent hours brainstorming with me about potential topics and offering suggestions. After hearing my concern about switching topic, he supported and encouraged me, “you don’t want to spend another two years doing things you don’t like.” Yes, passion for what he does—that’s what makes him tick. And he taught me that lesson.

I can go on and on about the anecdotes I had with Brian. Ultimately, he’s been much more than just an adviser to me. He’s a father figure, he’s a great friend, and so much more. I can’t express my gratitude enough for what opportunities he’s brought me, how much he’s taught me, and in what ways he’s helped me grow. He’s been, and always will be an inspiration for his passion, work ethics, warm heart and cool head.

Secondly, I’d like to acknowledge my fiance Cheng. It’s amazing how much he understood and supported me in the past five and half years. Although we’re long-distance, he’s always been committed and involved in my life and my work. Being a circuit design graduate student himself, he knows what a tapeout means. In addition to moral support, he in fact gave technical support as well. I remember so well how he stayed up all night with me helping me with my first few tapeouts. He even put aside his own work and helped me draw layouts, synthesize digital logics and place and route. My labmates made fun of how many of my blocks are named after him, as he either modified it or did it himself. It’s fair to say that, I would probably not have made those tapeouts without his help. I always know that I could count on him when I needed another pair of hands for help, or a pair of ears to hear me out.
Not only did Cheng help me, he’s also helped my labmates. He’s helped people set up synthesis flow, help people debug their circuits. Whenever he came to visit me in Seattle, he’d not go to lab with my in vain. He’d almost always help my labmates with digital-related questions as we analog designers don’t often have a fair grasp on digital circuit design knowledge. I really appreciate him spending time helping my labmates although he barely knows them.

I’d also like to acknowledge my labmates and the people from the SOC and FAST labs, for I could not make it to today without their help. I’ve almost worked with everybody in the lab on various projects. I’ve learned so much from them. I had the pleasure to get to know some of them well, and really enjoyed one another’s company while working hard. They really helped make my days at the lab more fun.

Lastly, I’d like to acknowledge my parents for their understanding and support in the past few years. Although my father is in China, he always called to check on me, mostly regarding to my work. He’s always there when I needed advice making decisions or sometimes dealing with people. My mom has lived with me for the past three years and had been extremely attentive caring for me. She’d rarely ask for things from me, mostly just give. I owe a whole lot to my parents for they’ve not only gave me life but also taught me the lesson of love.
DEDICATION

to my gracious God, and those who I love.
Chapter 1

INTRODUCTION: ENABLING NEXT-GENERATION WIRELESS SENSORS

According to 2012 US budget proposal, health care is allocated the largest share of budget, leading social security, national defense, income security and other government programs (Fig. 1.1). Escalating health care costs and the aging population calls for inexpensive home-monitoring of patient health.

Wireless body-area networks (WBAN) promise to provide significant benefits to the healthcare domain by enabling continuous monitoring, actuation, and logging of patient bio-signal data, which can help medical personnel to diagnose, prevent, and respond to various illnesses such as diabetes, asthma, and heart attacks [137]. In a WBAN, biophysiological data (e.g., ECG, EMG, and blood pressure) is first measured with on-body sensors, the data is then conditioned and digitized with body-worn sensor nodes (BSNs) and wirelessly transmitted to computers or hand-held devices for further processing. Assessment and treatment information is fed back to the patient (Fig. 1.2). In such a system, frequent battery replacement for the BSNs is not convenient or desirable. Therefore, we aim to develop Ultra-low-power (ULP) BSNs to effectively acquire, process, and transmit biosignals to the base station, as well as receive information from the base station.

In addition to body-worn applications, implantable systems for chronic use also require ultra-low power operation to minimize heat dissipation, avoid frequent battery replacement, and enable operation from wirelessly-delivered or harvested energy (e.g. thermoelectric generation). For instance, rapid advances in ultra-low power microelectronic design have given rise to a new paradigm in neuroscience and clinical research for treating brain disorders. Electrical engineers and neuroscientists have made dramatic progress in designing neural-electronic interfaces to operate as miniature, lightweight, chronically implanted wireless systems. Recently, technology further bridges engineering and neuroscience toward realizing
Figure 1.1: 2012 US budget proposal diagram.

Figure 1.2: A hypothetical wireless body-area network (WBAN).
brain computer interfaces (BCIs). BCIs can enable people with severe paralysis to communicate and/or manipulate objects, significantly improving their quality of life [68][39].

An example of a BSN system is illustrated in Fig. 1.3. The acquisition of biosignals on the order of μVs to mVs require amplification and signal conditioning. The amplified signals may be processed to extract the most salient information and reduce the data rate either in the digital or the analog domain. The signal is then transmitted to an external device, where the information is used to diagnose physiological disorders or infer biostate, for example. In addition to the BSN system show in the figure, their companion receiver systems are also of interest as they complete the feedback loop between the patient and the base station. The power consumption requirement of the circuitry at the base station is much more relaxed as they can be plugged in to the wall. Therefore, for the remaining of the thesis, I'll be focusing on the ULP circuit and system designs for the BSNs.

How do we tackle the problem today? Well, we often integrate commercial-of-the-shelf (COTS) components onto a wireless sensing platform. They usually exhibit high power consumption. Fig. 1.4 shows one of the most miniaturized platforms that were implemented for small animal monitoring. However, such a system consumes on average more than 20 mW during continuous transmission. A 33 mAh coin-cell battery can only sustain less than 2 hours of operation. This is not acceptable for our vision. During continuous health monitoring, the batteries can not be replaced this frequently on all the sensing nodes. A
closer look reveals the power breakdown of such a system: COTS amplifier (INA333) and ADC consume ≈ 100 µW, the microcontroller MSP430 consumes ≈ 500 µW (in low-power mode), the transmitter (CC1101) consumes 30 mW. This opens up opportunities to make innovations in both circuit blocks and systems to realize ULP operations.

This thesis is organized as following: Chapter II and III discuss analog circuit innovations for biosignal acquisition and processing. Chapter IV to VI present communication circuit blocks for short-range wireless, and Chapter VII present novel system-building for on-body wireless sensor nodes. In Chapter II, I’ll present a few low-noise amplifier designs and compare/contrast their performances. An ECoG processing IC intended for emerging BCI applications is presented in Chapter III. This chip conditions ECoG signals and simultaneously extracts energy in four fully-programmable frequency bands. Chapter IV presents a technique for integrating active circuitry into the lid of the wafer-scale hermetic Film Acoustic Bulk Resonator (FBAR) package while the FBAR resonator resides on the base wafer. We focus on the circuit design of these wafer-scale FBAR oscillators that exhibit either low-power or low-jitter. In Chapter V, we propose a tag architecture that employs an active transmitter to avoid the “self-jamming” problem present in RFID systems. re
ducing the complexity of reader design. In Chapter VI, we present a 2.4 GHz receiver that operates from a supply voltage of only 300 mV, the lowest report to date. It allows direct powering from various energy harvesting sources, enabling battery-free operations. In Chapter VII, two WBAN systems are presented with the focus on their system integration and deploy-ability. The "Bumblebee" is a self-contained miniaturized wireless sensing platform that continuously streams biosignal waveforms for as long as 3 days on a 33 mAh coin-cell battery. To further extend the lifetime, we also present an energy-harvesting WBAN system operating on body heat enabled by various ULP circuit blocks and architecture-level innovations.
Chapter 2

ANALOG CIRCUIT INNOVATIONS FOR BIOSIGNAL ACQUISITION

Monolithic amplifiers have been used for electrophysiological recording signals for decades [124][92][91]. The large time constants inherent in the amplifier dynamics typically preclude timesharing of a single amplifier between multiple electrodes [43]. Therefore, multi-channel systems typically use one amplifier per channel, imposing severe power constraints on the amplifier design.

Table 2.1 shows a few examples of the electrical characteristics of some electrophysiological signals that are commonly of interest in neuroscience. Single-unit recordings provide the finest spatial resolution of the brain, but they typically incur relatively high power consumption due to the wide amplifier bandwidth required and high resulting datarate [116]. EEG, on the other hand, is non-invasive and has modest amplifier bandwidth constraints at the cost of low spatial resolution. ECoG is an invasive modality that uses non-penetrating electrodes and offers a compromise that is receiving increased attention in the neuroscience community. Though this paper focuses on BPAs for single unit recording, the concepts presented here can be adapted to systems requiring different bandwidths and noise performance.

Here we outline the basic design requirements of a spike-based recording amplifier. Typical extracellular action potentials, or spikes, have amplitudes up to 500 µV, with much of

<table>
<thead>
<tr>
<th>Table 2.1: Characteristics of Electrophysiological signals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Single-Unit</td>
</tr>
<tr>
<td>LFP (local field potential)</td>
</tr>
<tr>
<td>ECoG (electrocorticography)</td>
</tr>
<tr>
<td>EEG (electroencephalography)</td>
</tr>
</tbody>
</table>
the signal energy in the 100 Hz - 7 kHz band. Low-frequency local field potentials (LFPs) have amplitudes as high as 5 mV and may contain signal energy below 1 Hz [52][68]. The low amplitudes of the spikes require BPA gain of around 100× up to a few kHz. Large DC-offsets at the electrode-tissue interface require offset cancellation or AC coupling. The input impedance of the amplifiers should be high enough (a few MΩs at 1 kHz) to limit signal attenuation from the electrode-tissue interface. The input-referred noise of the BPAs should be kept below the background noise of the recording site (< 10 μV). The power dissipation should be kept as low as possible (∝ 100 μW/channel) in a multi-channel system to contain the heating of the tissue within a few degrees Celsius [44][50]. Sufficient common-mode and power-supply rejection should also be ensured to reject the inevitable interference and supply noise. Lastly, the amplifiers should occupy small silicon areas to allow multiple instantiations in a larger system.

These requirements form the initial condition for our proposed design evolution. This chapter is organized as follows: section 2.1 gives an overview of the present state of the art. Section 2.2 presents an evolution of BPA designs that exhibit low noise while consuming minimal power, including a closed-loop telescopic differential BPA, an open-loop complementary-input single-ended BPA, and a closed-loop complementary-input differential BPA design. Section 2.3 compares and contrasts the measurement results of the three amplifier designs. Section 2.4 presents an example system-level implementation using BPA3. Section 2.5 presents the in-vivo measurement results recorded from a rat motor cortex and a mouse visual cortex. Section 2.6 provides a detailed analysis on issues involved in using the proposed amplifiers in a multi-channel system. Finally, section 2.7 presents the design and implementation of an alternative amplifier topology, the chopper-stabilized amplifier.

2.1 State-of-the-art biopotential amplifiers

The pioneering power-efficient neural recording ICs [124][91][44][19][95] typically consumed about 100 μW to achieve < 5 μVrms of noise for 5-10 kHz bandwidth. Recently published amplifiers [130][131][18] have featured reduced power consumption in an effort to enable large recording arrays. Below we present a brief review of some representative techniques used in state-of-the-art low-power biopotential amplifier designs. More comprehensive reviews can
be found in [43][50][93].

The oft-cited [44] presents many useful techniques for designing biopotential-recording amplifiers, including the use of capacitive feedback and pseudo-resistors around an open-loop transconductance amplifier (OTA).

By reducing the number of current branches, the one-stage self-biased preamplifier in [18] achieves $4.9 \mu V_{rms}$ input-referred noise while drawing only 2 $\mu A$ under $\pm 1.65$ V supply.

The fully-differential folded-cascode preamplifier in [143] includes multiple adjustable parameters for various biopotential recording applications. While achieving $3.6 \mu V_{rms}$ input-referred noise over 20 Hz to 10 kHz, the current consumption is 8 $\mu A$ at $\pm 1.7$ V supply. The extra current branches in the folded-cascode topology result in sub-optimal power-noise tradeoff.

The design in [130] employs a modified folded-cascode topology with severely scaled currents in the input and folded branches and source degeneration to reduce the power and noise contribution from the folded branches. However, the 4.5 M$\Omega$ degeneration resistors increase the area and voltage head-room required of the amplifier, requiring a 2.8 V supply voltage.

Other recent advances in the field include mixed-signal front-end for filtering and digitization [86][88][38]. With the trend towards integrating analog and digital subsystems on a single die, it has become increasingly important for analog circuitry to operate from the 1 V supply typical of modern digital CMOS. This paper presents a power-efficient amplifier topology with a 1V supply.

2.2 Design of low-noise biopotential amplifiers

In this section, we show a logical progression of biopotential amplifiers by comparing and contrasting their design strategies. First, a closed-loop fully-differential telescopic amplifier (BPA1) is included as a “baseline” design to compare against the other designs. Then, an open-loop single-ended complementary-input amplifier (BPA2) serves as an instructional example to demonstrate a design technique with the potential to surpass the theoretical limit of the power-noise tradeoff of a conventional amplifier at the expense of PSRR. Lastly, combining the salient features of BPA1 and BPA2, we designed a closed-loop fully-differential
complementary-input amplifier (BPA3) with excellent power-noise performance, sufficient linearity and power-supply rejection performances.

2.2.1 Design of a closed-loop fully-differential telescopic-cascode amplifier

The schematic of a typical closed-loop telescopic BPA is shown in Fig. 2.1(a). The input signals are AC-coupled into the amplifier to reject large DC offsets from the electrode-tissue interface. $C_s$ must be made small enough to avoid attenuation of the input signal from the electrode, but large enough to avoid attenuation from the capacitive divider it forms with the OTA input capacitance, which will increase input-referred noise, as shown in Eqn. 2.1. Let $v_{mi}\text{OTA}^2$ and $\overline{v_{mi}}^2$ represent the input-referred noise of the OTA and the BPA, respectively.

$$\overline{v_{mi}}^2 = \left(\frac{C_s + C_f + C_{in}}{C_s}\right)^2 \overline{v_{mi}\text{OTA}^2}$$

(2.1)

The ratio $C_s/C_f$ sets the mid-band gain of the amplifier to roughly 40 dB. We chose 180 IF for $C_f$ to ensure sufficient mid-band gain while limiting the increase in the input-referred noise due to the input capacitive divider (Eqn. 2.1) to 12%. Pseudoresistors are used here as an area-efficient approach to bias the input transistors and form a sub-Hz high-pass frequency corner with $C_f$ to accommodate EEG/LFP signals. Thick-oxide MOS transistors are used at the input to reduce gate leakage current, which could result in significant DC-offsets. The input-referred noise from the feedback pseudoresistors is:

$$\overline{v_{mi,R}}^2 = \left(\frac{V_{n,R}}{1 + j\omega RC_f}\right) \left(\frac{1}{A_{CL}^2}\right)$$

(2.2)

Let $A_{CL}$ represent the closed-loop gain of the BPA. Scaled by $A_{CL}^2$ and attenuated at 20 dB/dec after the sub-Hz frequency corner $\frac{1}{2\pi RC_f}$, the noise contribution from the pseudoresistors at frequencies of interest is negligible compared to flicker or thermal noise. More detailed analysis on the noise contributed by pseudoresistors can be found in [43].

In order to lower the power consumption and ease integration with complex digital subsystems, the amplifier operates from a supply as low as 1 V. A two-stage fully-differential design was chosen to provide sufficient gain, signal swing, and supply rejection while operating from a 1 V supply. We simulated an open-loop gain of 69 dB. In addition, the
Figure 2.1: Schematics of three low-noise biopotential amplifier designs.
fully-differential topology provides higher CMRR and PSRR when compared to its single-ended counterpart, critical for low-voltage supply conditions.

The input stage employs a telescopic cascode rather than a folded cascode because of its reduced number of active branches and because the small input amplitude precludes the need for a wide input swing in the first stage. We used NMOS input transistors because their higher $g_m/I_D$ compared to PMOS transistors results in lower thermal noise, which dominates over flicker noise in this design.

To reduce noise, the transistors are carefully sized for appropriate inversion coefficient (IC) [34] as shown in Table 2.2.

The input-referred thermal noise can be approximated as:

$$v_{n,th}^2 = \frac{16kT}{3g_{m1}} \left(1 + \frac{g_{m5}}{g_{m1}}\right) \Delta f.$$

(2.3)

The input-referred 1/f noise can be calculated as:

$$v_{n,1/f}^2 = \frac{1}{C_{ox} \Delta f} \cdot \left(\frac{K_n}{(WL)_1} + \frac{K_p g_{m5}^2}{(WL)_5 g_{m1}^2}\right).$$

(2.4)

Here, $C_{ox}$ represents the gate dielectric capacitance while $K_n$ and $K_p$ denote the nFET and pFET flicker noise constants. The values are process-dependent and are on the order of $10^{-13}$ V^2-pF [105] with $K_n > K_p$. All the transistors, particularly the input pair, use large gate areas to reduce the 1/f noise. Similar to the $v_{n,th}$, a low $g_{m5}/g_{m1}$ ratio also reduces $v_{n,1/f}$. 

| $W/L$ ($\mu$m) | $L_d$ ($\mu$A) | Inv. Coeff | $g_m/I_d$ (V^{-1}) | $|V_{GS}| - |V_I|$ (mV) |
|----------------|---------------|-------------|---------------------|---------------------|
| M1.2           | 616/2         | 3           | 0.023               | 27.56               | -154                |
| M3.4           | 12/5          | 3           | 0.27                | 22                  | 102                 |
| M5.6           | 12.2/13       | 3           | 11.3                | 6.9                 | 258                 |
| M0             | 100.8/8       | 6           | 0.54                | 19.4                | 7                   |
| M7.8           | 12/4          | 2.2         | 0.96                | 16.82               | 76                  |
| M9.10          | 8.6/3         | 2.2         | 1.92                | 13.64               | 103                 |
2.2.2 Design of an open-loop complementary-input amplifier

Although BPA1 achieves good power-noise performance, extra current was consumed in the differential input branches and the second stage. For a given bias current, open-loop amplifiers achieve superior noise performance at the expense of linearity, imprecise gain control and reduced power-supply rejection. In this section, we will present the design of an open-loop complementary-input single-ended BPA [46] that achieves an excellent power-noise tradeoff at the expense of PSRR.

Our open-loop design serves as an instructional extreme case that minimizes noise and power consumption. This design philosophy is motivated by the characteristics of the signals targeted by this amplifier. The small amplitude (\(\sim 100 \, \mu V\)) of neural signals relaxes the linearity requirement compared to more general-purpose amplifiers. Gain precision is also not critical; the absolute signal amplitude is not as important as preservation of relative amplitudes. In addition, a stable power supply is also possible with careful system design and deployment.

Leveraging these design tradeoffs, we introduce the schematic shown in Fig. 2.1(b). The input signals are AC-coupled into the amplifier and the gates of the input transistors are biased using pseudoresistors. Diode-connected transistors \(M_3\)−\(5\) provide a means to vary the output conductance and thereby the gain and bandwidth.

By driving the gates of both \(M_1\) and \(M_2\), the total transconductance efficiency of the amplifier (which determines noise efficiency) is \(2g_{m1}/I_d\), theoretically allowing a noise efficiency factor (NEF) less than unity. Assuming the two transistors have equal transconductance, the amplifier’s transconductance \(G_m\) is effectively doubled. The input-referred thermal noise can be expressed as:

\[
v_{\text{noise}}^2 = \left( \frac{8kT}{3 \cdot (g_{m1} + g_{m2})} \right) \Delta f \tag{2.5}\]

If \(g_{m1}=g_{m2}\), then

A common-source amplifier with current source load will have more thermal noise than the topology proposed here by a factor ranging from \(\sqrt{2}\) for a noiseless current source (i.e. \(g_m=0\)) to 2 for a current source with \(g_m\) equal to that of the amplifying transistor. The minimum practical \(g_m\) for the current source load is dictated by supply headroom and the
required overdrive voltage.

Because the positive supply is connected to the source of $M_2$, supply noise is coupled to the output through $M_2$. The power-supply rejection ratio can be expected to be approximately $\frac{g_{m2}}{g_{m1}+g_{m2}}$, or about 6 dB if $M_{1,2}$ have equal transconductance. Therefore, in order to realize the potential noise performance of BPA2, a stable supply with noise comparable to the amplifier’s input-referred noise must be provided.

2.2.3 Design of a closed-loop fully-differential complementary-input amplifier

Leveraging BPA2’s power-efficient complementary-input topology, we further improved the PSRR, linearity, and precision of gain/bandwidth control in a closed-loop fully-differential amplifier design (BPA3) [100]. The schematic of BPA3 is shown in Fig. 2.1(c). BPA3 has the same bias current and mid-band gain as BPA1. We also retain the two-stage fully-differential design with similar compensation and CMFB schemes.

Similar to BPA2, we drive both the NMOS and PMOS input pairs to allow a significant reduction in the input-referred noise. Two small-signal gain paths from the NMOS and PMOS inputs necessitate two capacitive feedback paths from the output to the gates of both input transistors. The open-loop gain is simulated to be $> 70$ dB. The increase in the input-referred noise from the capacitive divider (Eqn. 2.1) at both the NMOS and PMOS inputs are found to be approximately 10%. The complementary-input strategy doubles the amplifier’s effective transconductance. As a result, the input-referred noise voltage is reduced by a factor of $\sqrt{2}$, similar to its open-loop counterpart. The input-referred thermal noise power is twice that of BPA2 because differential branches double the output noise. If $g_{m1} = g_{m3}$, it can be expressed as:

$$v_{n_{i,th}}^2 = \left(\frac{16kT}{3g_{m1} \cdot 2}\right) \Delta f$$

(2.6)

Similarly, Eqn. 2.3 shows the input-referred thermal noise for BPA1. If $g_{m5} = 0$, then Eqn. 2.3 reduces to

$$v_{n_{i,th}}^2 = \left(\frac{16kT}{3g_{m1}}\right) \Delta f$$

(2.7)

A comparison between Eqn. 2.6 and Eqn. 2.7 reveals that the input-referred noise voltage
of BPA3 is approximately \(1/\sqrt{2}\) that of BPA1. Table 2.3 shows the parameters and operating conditions of each transistor in the OTA. The input-referred flicker noise from the input devices can be approximated by

\[
\eta_{ni,1/f}^2 = \frac{1}{C_{ox} \Delta f} \cdot \left( \frac{K_{n1}g_{m1}}{(WL)_1} + \frac{K_{p2}g_{m3}}{(WL)_3} \right) \cdot \frac{1}{(g_{m1} + g_{m3})^2}
\]  

(2.8)

If \(g_{m1} = g_{m3}\), then

\[
\eta_{ni,1/f}^2 = \frac{1}{4C_{ox} \Delta f} \cdot \left( \frac{K_n}{(WL)_1} + \frac{K_p}{(WL)_3} \right)
\]

(2.9)

We would also like to compare the 1/f noise of BPA1 with that of BPA3. In BPA1, if \(g_{m5} = 0\), Eqn. 2.4 reduces to:

\[
\eta_{ni,1/f}^2 = \frac{1}{C_{ox} \Delta f} \cdot \left( \frac{K_n}{(WL)_1} \right)
\]

(2.10)

If we assume \(\frac{K_n}{(WL)_1} = \frac{K_p}{(WL)_3}\) in Eqn. 2.9 for simplicity, then the net input-referred flicker noise voltage of BPA3 is \(1/\sqrt{2}\) that of BPA1. Notice that the complementary-input strategy has similar influence on the input-referred thermal and flicker noise.

Because the input also drives the PMOS transistor pair \(M_{3,4}\), the transconductances of \(M_{3,4}\) not only contributes to the differential gain, but also the common-mode gain. To ensure high CMRR/PSRR, we use dual tail current sources in the first stage to degenerate the common-mode transconductance, thus reducing the common-mode gain.

The common-mode gain \((A_{cm})\) and the gain of power-supply interference \((A_{ps})\) can be expressed as

\[
A_{cm} = \frac{V_{out}}{V_{in,cm}} \approx \frac{(g_{05} + g_{06})g_{m8}/(g_{01}g_{02})}{1 + sC_c/(g_{05} + g_{06})}
\]

(2.11)

Any variation in the supply is attenuated by approximately \(\frac{g_{m6}}{(g_{m3} + g_{m4})} \cdot (1 - \frac{V_{ref}}{V_{dd}})\) before being amplified by the \(g_{m}\) mismatches in \(M_{3,4}\) (Eqn. 2.13). Let \(g_{01,2}\) denote the output conductance of the first and second stage, \(g_{05,6}\) denote the output conductance of current source transistors \(M_{5,6}\), \(g_{m8}\) denotes the transconductance of the second stage, \(\Delta g_{m}\) denotes
Table 2.3: Device parameters of complementary-input closed-loop biopotential amplifiers

|    | W/L (μm) | I_d (μA) | Inv. Coeff | \( \frac{g_{m}}{I_d} (V^{-1}) \) | \( |V_{G2S} - |V_{S}\) (mV) |
|----|----------|----------|------------|---------------------------------|-----------------|
| M1.2 | 552/2    | 3        | 0.022      | 27.53                           | -152            |
| M3.4 | 552/2    | 3        | 0.12       | 24.23                           | -106            |
| M5   | 110.4/8  | 6        | 0.54       | 19.4                            | 5               |
| M6   | 73.2/8   | 6        | 1.98       | 13.5                            | 103             |
| M7.8 | 12/4     | 2.1      | 0.93       | 16.96                           | 30              |
| M9.10| 8.6/3    | 2.1      | 1.93       | 13.62                           | 103             |

the \( g_m \) mismatch in \( M_{3,4} \), and \( C_c \) denotes the compensation capacitor.

\[
A_{ps} = \frac{V_{out}}{V_{in, supply}} \approx \frac{\Delta g_m \gamma g_m}{1 + sC_c/\Delta g_m} \mid (2.12)
\]

\[
\gamma = \frac{g_m}{g_m + g_m} \cdot (1 - \frac{V_{dd}}{V_{dd}})
\]

The design of the second stage is focused on ensuring sufficient output swing (differential peak-to-peak over 1 V) while achieving a reasonable gain (20dB).

We employed continuous-time CMFB, with the output common-mode voltage sensed using two large resistors. The output of the CMFB amplifier controls the gate voltage of \( M_6 \) to adjust the common-mode voltage of the first stage.

The signal propagation of the common-mode feedback path of BPA3 consists of two parts: from the average output \( V_{cmout} \) to the feedback control \( V_{ctrl} \), and from \( V_{ctrl} \) to the amplifier output. The first part of the CMFB path has a wide bandwidth and small DC gain (\( \approx 1 \)); the second part determines the CMFB frequency response as illustrated in Fig. 2.2. Let \( g_{m1-4,6-8} \) denote the transconductance of the corresponding transistors, \( C_c \) and \( C_L \) denote the compensation and load capacitors, \( g_{o1} \) and \( g_{o2} \) denote the total output conductances of stage 1 and 2. Then

\[
A_{cmfb} = \frac{V_{out, CM}}{V_{ctrl}} = \frac{-sg_{m6}C_c + g_{m6}g_{m7,8}}{s^2C_cC_L + sC_cg_{m7,8} + g_{o1}g_{o2}}
\]  

(2.13)
Figure 2.2: a) Differential-mode gain path; b) CMFB gain path.

\[
A_{dm} = \frac{V_{out,DM}}{V_{in,dm}} = \frac{-s(g_{m1,2} + g_{m3,4})C_c + (g_{m1,2} + g_{m3,4})g_{m7,8}}{s^2C_cC_L + sC_cg_{m7,8} + g_{o1}g_{o2}}
\]

(2.14)

Both the differential and common-mode gains share the compensation capacitor \(C_c\) and \(g_{m7,8}\) stage. The similarity of the transfer functions leads to a stable CMFB path if the differential-mode path is unity-gain stable. This CMFB topology achieves both high gain and bandwidth, while saving power by sharing one CMFB circuit between both the first and second stages. A brief analysis of the common mode circuit shows that the closed loop common-mode gain of the amplifier (including the effect of CMFB) is \(V_{oc}/V_{ic} = A_{cm}/(1+A_{cm,fb})\) where \(A_{cm}\) is the amplifier’s common-mode gain, and \(A_{cm,fb}\) is the combined gain of the common-mode detection circuit and the common-mode control (i.e. \(A_{cm,fb} = g_{m6}/(g_{o1}+g_{o2})\) if the gain of the block labeled “CMFB” in Fig. 2 is unity.) A high gain-bandwidth product in the CMFB loop reduces the common-mode voltage gain while leaving
Figure 2.3: Closed-loop amplifier start-up concern alleviated by adding a leakage path through a diode-connected transistor at the first-stage output.

the differential-mode voltage gain unaffected. Thus, increasing the CMFB loop bandwidth will improve CMRR at higher frequencies.

In order to provide DC feedback and bias the input transistors, the outputs are fed back through pseudoresistors to bias the gates of the NMOS input transistors $M_{1,2}$. However, this feedback inevitably forms a positive feedback loop at low frequencies. As shown in Fig. 2.3, this is particularly problematic when the output common-mode voltage is initially low. In this case, the pull-down paths are turned off as the gates of $M_{1,2}$ are low. At the same time, the common-mode feedback control voltage rises, which also turns off the pull-up paths, leaving the first stage output in a high-impedance state. Notice that if the outputs are fed back to both the nFET and pFET input transistors $M_{1,2,3,4}$, the gain of the positive feedback loop will be doubled, which requires a higher CMFB gain and increases the likelihood for initial DC latch-up. To ensure reliable start-up, we added a pair of diode-connected transistors at the output of the first stage connecting to ground. This scheme provides additional current paths through the diode-connected transistors when both the pull-up and pull-down paths are initially turned off. The resulting leakage currents are
negligible during normal operation of the amplifier.

2.3 Measurement results of the biopotential amplifiers

The telescopic-cascade (BPA1) and complementary-input (BPA3) closed-loop biopotential amplifiers were both fabricated in 0.13 \( \mu m \) process. We used Metal Insulator Metal (MIM) capacitors for their high density, good linearity, and low substrate capacitance. The complementary-input open-loop biopotential amplifier (BPA2) was fabricated in 0.5 \( \mu m \) SOI-BiCMOS process and employed CMOS devices exclusively. The amplifier occupies 33,000 \( \mu m^2 \). The circuit operates from a supply between 1 V and 5 V. The telescopic cascade amplifier (BPA1) occupies 46,800 \( \mu m^2 \) with 57.8% used for capacitors. The complementary-input closed-loop amplifier (BPA3) occupies 71,750 \( \mu m^2 \), 67.1% of which is occupied by capacitors. A die photo containing BPA1 and BPA3 is shown in Fig. 2.4(a) and Fig. 2.4(b).

The measurements of all the three amplifiers presented here were taken with a 1 V supply.

Fig. 2.5(a) compares the measured frequency response of the three BPA designs. The mid-band gains of BPA1 and BPA3 are 40.5 dB and 40 dB, respectively. The different high-pass corner frequencies of BPA1 and BPA3 are due to different lengths of PMOS transistors.
Figure 2.5: Top: Bode magnitude comparison of the three BPAs; Bottom: Measured input-referred noise comparison of the three BPAs.
used to realize the pseudoresistors. Specifically, four cascaded PMOS transistors of 40 $\mu$m length are used in BPA3, compared with two cascaded PMOS transistors of 50 $\mu$m length in BPA1, resulting in higher effective resistance. A larger effective resistance results in a lower high-pass corner at the cost of a higher thermal noise (the larger device sizes in the pseudoresistors result in larger input capacitive divider (eqn. 2.1)). A comprehensive analysis of pseudoresistor noise contribution is given in [43]. The -3 dB low-pass corners occur at approximately 8 kHz for BPA1, and 10 kHz for BPA3. The difference is attributed to the larger effective transconductance $G_m$ of BPA3 compared with that of BPA1. The compensation capacitor used in BPA3 is larger than BPA1. The measured mid-band gains of BPA2 are 36 and 44 dB, and the -3 dB low-pass corners are 4.7 kHz and 1.9 kHz for minimum and maximum gain settings, respectively. The gain variation measured across 10 chips was 2 dB. For the remainder of the discussion regarding BPA2, we will primarily focus on the lowest gain setting (shown in Fig. 2.5(a)) because it provides comparable bandwidth to the other two amplifiers.

Fig. 2.5(b) compares the measured input-referred noise spectrum of the three BPAs. Consistent with our analysis, the noise of BPA1 is higher than that of BPA3 due to the higher effective $G_m$ of the first stage of BPA3. The noise spectrum of BPA2 is higher than that of the other amplifiers because of a lower bias current resulting in a higher thermal noise in BPA2. The measured input-referred noise of BPA1, BPA2 and BPA3 integrated from 0.1 Hz to 25 kHz are 3.1 $\mu$V, 3.5 $\mu$V and 2 $\mu$V respectively. Although flicker noise corners are high (800 Hz-1 kHz) in all three cases, the measured flicker noise contributes approximately 20% of the total integrated noise. Flicker noise can be further reduced by employing chopper-stabilization techniques [141][28]. We have extended our measured noise profile based on a conservative one-pole (20 dB) roll-off to 10 times the bandwidth of the amplifiers to allow consistent comparison with the literature. The input-referred noise integrated from 0.1 Hz to 105 kHz is respectively 3.2 $\mu$V, 3.6 $\mu$V, and 2.2 $\mu$V.

Fig. 2.6(a) compares the PSRR of the three BPAs. The PSRR for BPA1 is approximately 20 dB lower than that of BPA3. The supply coupling of BPA3 is attenuated by the ratio $\frac{g_{m6}}{(g_{m3}+g_{m4})} \cdot (1 - \frac{V_{dd}}{V_{dd}})$ before amplified by the mismatches in $M_3$ and $M_4$, consistent with the analysis in Eqn. 2.13. Due to the single-ended nature of BPA2, the positive and negative
Figure 2.6: Top: PSRR bode magnitude comparison of the three BPAs; bottom: CMRR magnitude comparison of BPA1 and BPA3.
supplies directly modulate the pFET and nFETs, respectively. Therefore, we expect that
the gain from the power supply to the output will be approximately half the gain from input
to output. This results in an expected PSRR of 6 dB, consistent with the measured PSRR
of 5.5 dB at low frequencies.

Fig. 2.6(b) compares the CMRR of BPA1 and BPA3. The CMRR for BPA1 has an
average value of 60 dB, compared with 80 dB for BPA3. The larger devices in BPA3 should
result in smaller expected values of CMRR due to reduced mismatch.

Finally, the linearity of the amplifiers are examined. Many papers use total harmonic
distortion (THD) to describe linearity. However, in our experience, the main concern for
spike-recording applications is gain compression due to interferers such as electromagnetic
interference or low frequency local field potentials that can result in time-varying gain.
Therefore, we posit that it is more useful to characterize the -1dB gain compression point
(approximately 89% of voltage gain) than THD for these amplifiers. This metric also pro-
vides a simple “max signal level” specification that facilitates matching amplifiers with var-
ious applications. We will thus evaluate the linearity performance by comparing their -1dB
gain compression input voltage. The -1dB gain compression point occurs at input level of 3
mV for BPA1, 1.8 mV for BPA2, and 4 mV for BPA3. As expected, the open-loop amplifier
exhibits more nonlinearity than the closed-loop amplifiers. The difference between the two
closed-loop amplifiers BPA1 and BPA3 can be attributed to the complementary-input topo-
logy employed in BPA3. Since linearity performance is enhanced through feedback, a high
loop gain, or a large open-loop gain given the same feedback ratio is desired. Because the
input drives both the nFETs and pFETs, a larger signal swing is allowed at the input before
significantly attenuating the open-loop gain and exacerbating the linearity performance.

To compare our noise and power performance to other amplifiers, we use the noise
efficiency factor (NEF) [124]:

\[
NEF = V_{rms, in} \sqrt{\frac{2 \cdot I_{total}}{\pi \cdot U_T \cdot 4kT \cdot BW}}
\]  \hspace{1cm} (2.15)

Where \( I_{total} \) is the total amplifier supply current, \( U_T \) is the thermal voltage \( kT/q \), \( BW \)
is the amplifier bandwidth, and \( V_{nk, rms} \) is the amplifier’s input-referred RMS voltage noise.
Table 2.4: Performance summary of the presented biopotential amplifiers

<table>
<thead>
<tr>
<th></th>
<th>Spec</th>
<th>BPA1</th>
<th>BPA2</th>
<th>BPA3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>≥ 40 dB</td>
<td>40.5 dB</td>
<td>36.1 dB</td>
<td>40 dB</td>
</tr>
<tr>
<td>$I_{\text{Amp}}$</td>
<td>minimize</td>
<td>12.5 µA</td>
<td>805 nA</td>
<td>12.1 µA</td>
</tr>
<tr>
<td>NEF</td>
<td>minimize</td>
<td>4.5</td>
<td>1.9</td>
<td>2.9</td>
</tr>
<tr>
<td>$v_{n,\text{rms}}$</td>
<td>&lt; 10 µV</td>
<td>3.2 µV</td>
<td>3.0 µV</td>
<td>2.2 µV</td>
</tr>
<tr>
<td>THD(@ input)</td>
<td>minimize</td>
<td>1.5% @ 1 mVpp</td>
<td>7.1% @ 1 mVpp</td>
<td>1% @ 1 mVpp</td>
</tr>
<tr>
<td>PSRR</td>
<td>≥ 60 dB</td>
<td>≥ 60 dB</td>
<td>5.5 dB</td>
<td>≥ 80 dB</td>
</tr>
<tr>
<td>CMRR</td>
<td>≥ 60 dB</td>
<td>60 dB</td>
<td>-</td>
<td>80 dB</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>0.5 Hz - 7 kHz</td>
<td>4 Hz - 8.5 kHz</td>
<td>3 Hz - 4.7 kHz</td>
<td>0.5 Hz - 10.5 kHz</td>
</tr>
<tr>
<td>$Z_{in} @$1 kHz</td>
<td>a few MΩ</td>
<td>8 MΩ</td>
<td>22 MΩ</td>
<td>4 MΩ</td>
</tr>
<tr>
<td>Area</td>
<td>minimize</td>
<td>0.047 mm²</td>
<td>0.046 mm²</td>
<td>0.072 mm²</td>
</tr>
<tr>
<td>Tech</td>
<td>-</td>
<td>0.13 µm</td>
<td>0.5 µm</td>
<td>0.13 µm</td>
</tr>
</tbody>
</table>

$V_{n,\text{rms}}$ used here for the NEF calculation is integrated up to 105 kHz. This FOM normalizes the input-referred rms noise voltage to that of an ideal single-transistor bipolar amplifier with equal current consumption and bandwidth.

The measured performance of BPA1, BPA2, and BPA3 are summarized in Table 2.4. The open-loop complementary-input amplifier (BPA2) design achieves the best NEF (1.9). However, its poor PSRR performance (5.5 dB) offloads extremely stringent noise and power-supply rejection requirements to the voltage regulation circuitry, potentially increasing the design complexity and power consumption of the system. Although BPA2 has limited application in realistic recording scenarios, it served as a stepping-stone to the design of BPA3. By employing fully-differential closed-loop architecture, BPA3 achieves favorable power-noise tradeoff as well as sufficient PSRR (≥ 80 dB) and linearity performances (1% THD at 1 mV peak-to-peak input voltage) for practical recording use.

The three amplifiers are compared with a few other state-of-the-art neural amplifiers in Table 2.5. BPA2 achieves the best NEF (1.9) with a compromise in linearity and PSRR. BPA3 has better NEF performance (2.9) than all the referenced closed-loop amplifiers except [130]. However, the low-frequency high-pass corner in [130] is three orders of magnitude higher than BPA3, significantly filtering out the 1/f noise. The BPA3’s input-referred noise can be reduced and NEF can be improved by increasing the high-pass corner. In addition to sufficient gain and linearity, BPA3 also achieves comparable PSRR and CMRR perfor-
Table 2.5: Performance comparison of biopotential amplifiers

<table>
<thead>
<tr>
<th></th>
<th>BPA1</th>
<th>BPA2</th>
<th>BPA3</th>
<th>[44]</th>
<th>[28]</th>
<th>[130]</th>
<th>[131]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vdd (V)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>+/-2.5</td>
<td>3.3</td>
<td>2.8</td>
<td>0.8</td>
</tr>
<tr>
<td>I_Amp (µA)</td>
<td>0.5</td>
<td>0.8</td>
<td>12.1</td>
<td>16</td>
<td>12</td>
<td>27</td>
<td>0.33</td>
</tr>
<tr>
<td>NEF</td>
<td>4.5</td>
<td>1.9</td>
<td>2.9</td>
<td>4.0</td>
<td>4.9</td>
<td>2.67</td>
<td>3.8</td>
</tr>
<tr>
<td>NEF²/Vdd</td>
<td>20.3</td>
<td>3.6</td>
<td>8.4</td>
<td>80</td>
<td>43.2</td>
<td>20</td>
<td>11.6</td>
</tr>
<tr>
<td>Gain (dB)</td>
<td>40.5</td>
<td>36</td>
<td>40</td>
<td>39.5</td>
<td>45.5</td>
<td>30.8</td>
<td>40.2</td>
</tr>
<tr>
<td>1 dB comp (@ Input) (mV)</td>
<td>3</td>
<td>1.7</td>
<td>4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>V_{in,RMS} (µV)</td>
<td>3.2</td>
<td>3.6</td>
<td>22</td>
<td>2.2</td>
<td>0.93</td>
<td>3.06</td>
<td>2.7</td>
</tr>
<tr>
<td>PSRR (dB)</td>
<td>≥60</td>
<td>5.5</td>
<td>≥80</td>
<td>≥85</td>
<td>—</td>
<td>75</td>
<td>62-63</td>
</tr>
<tr>
<td>Bandwidth (Hz)</td>
<td>4-8.5k</td>
<td>3-4.7k</td>
<td>0.5-10.5k</td>
<td>0.025-7.2k</td>
<td>0.180</td>
<td>45-53k</td>
<td>3m-245</td>
</tr>
<tr>
<td>Area (mm²)</td>
<td>.047</td>
<td>.046</td>
<td>.072</td>
<td>.16</td>
<td>—</td>
<td>.16</td>
<td>1</td>
</tr>
<tr>
<td>Technology (µm)</td>
<td>.13</td>
<td>.5</td>
<td>.13</td>
<td>1.5</td>
<td>.8</td>
<td>.5</td>
<td>.35</td>
</tr>
</tbody>
</table>

mance compared to the other amplifiers despite its low supply voltage, enabling realistic recording scenarios accompanied by power supply noise, electromagnetic (EM) interference and crosstalk. While it is difficult to compare the area due to different technologies used, it is desirable to reduce the size of the amplifiers when scaled up to a multi-channel system. Area can be further reduced by stacking metal-insulator-metal (MIM) capacitors above the amplifier transistors. These modifications would result in a 50% reduction in area.

Although NEF contains current, noise, and bandwidth information, all of which are important, our system-level integration efforts (next section) have revealed that a low (1 V) supply voltage is desirable. This allows integration of complex systems into fine-line (0.13 µm and below) CMOS processes that nominally use a 1.2 V supply. These processes allow a) integration of very low power synthesized logic, b) high f_T allowing low power RFIC design on the same chip, and c) MIM capacitors with high-density and precision that we heavily utilize in our amplifiers. Because the conventional NEF does not consider the supply voltage, a modified metric NEF²Vdd [88] is included in Table 2.5. We can see that BPA3 achieves the best modified NEF compared to other state-of-the-art amplifiers. Fig. 2.7 also highlights the superior NEF performance of BPA3 compared to other amplifiers operating at similar supply voltages ([131] and [28]), or significantly lower supply voltage for the amplifier with comparable NEF performance ([130]).
2.4 System-Level Implementation

We have demonstrated the practical implementation of BPA3 in a neural interface that wirelessly streams a digitized neural waveform over 10m [100]. As shown in Fig. 2.8, the neural interface comprises an analog front-end (AFE) with variable gain from 42 to 78 dB, an 8b successive approximation (SAR) ADC, and a Medical Implant Communication Service (MICS)-band (402 to 405 MHz) FSK transmitter. Continuous spike data is sampled at 8b, 9.1 kS/s with 3b of interleaved synchronization packet headers, then transmitted at 100 kb/s. We target the maximum allowable MICS-band EIRP of -16 dBm (25 µW) to maximize the wireless communication range.

The low-noise AFE includes the BPA3 presented above AC-coupled to a variable-gain amplifier (VGA) with 6 variable gain and 7 variable high-pass corner settings (Fig. 2.9(a)). The VGA consists of a complementary rail-to-rail folded cascode core to improve the input signal swing (Fig. 2.9(b)). The closed-loop gain can be adjusted from 0-38 dB by selecting different feedback capacitors. The variable low frequency corners are set from 1 Hz to 300 Hz by programming the feedback transconductor bias current. The feedback transconductors
Figure 2.8: Top-level block diagram of the 500 μW neural interface system IC.

are linearized $g_m$ cells with bias current as small as a few nAs. Alternatively, pseudoresistor feedback can be selected to obtain a 1-Hz low-frequency corner. This is helpful in amplifying signals with frequency content below 10 Hz (e.g., LFP, ECoG or electromyography (EMG)). The transmitter achieves high power efficiency (16%) by directly modulating the reference oscillator at 44.545 MHz and driving a $9 \times$ frequency multiplying power amplifier. FSK modulation is realized by directly pulling a 4 pF on-chip capacitor, creating a 16.5 kHz $\Delta f_{ref}$ and 148 kHz $\Delta f_{rf}$ at 400 MHz. An edge-combiner uses 9 equally-spaced edges generated by a DLL. Operating the entire DLL at $f_{ref}$ (instead of $f_{rf}$) enables significant power savings while ensuring quartz stability.

This system occupies a die area of 2.5×1 mm$^2$ using 0.13 μm process. The only off-chip components used in this system are two quartz resonators and 5 passive components used for impedance matching and system clock generation. The system consumes a total of 500 μW with a 30 μW AFE, sub-μW ADC and a 400 μW transmitter. This ultra-low-power sensing platform enables continuous monitoring of electrophysiological activity in untethered animals, providing unprecedented opportunities for neuroscience and medical research.

Miniaturization of wireless neural interface systems enables more profound discoveries in neuroscience and clinical research. A reduced board size of 7.6 × 8.7 mm$^2$ was used for deployment, resulting in a system we call the “Bumblebee” [87]. As shown in Fig. 2.10(a)
Figure 2.9: a) The closed-loop schematic of the VGA and the Gm cell; b) VGA double-folled cascode OTA schematic.
Figure 2.10: a) The Bumblebee, a miniaturized, ultra lightweight, low power wireless sensor platform; b) the functional block diagram of the Bumblebee; c) the block diagram of the custom receiver.
and Fig. 2.10(b), This miniaturized board contains a custom IC, all necessary passive components, and a coin-cell battery. The entire system is powered from the single battery, which is subsequently regulated to 1 V. Since there is no non-volatile memory on-chip, an external programmer is required to load appropriate settings for the chip upon powerup. Bumblebee is entirely self-contained and ideal for a variety of sensing/recording applications (EMG, spike-based, audio). We have also developed a low-cost companion receiver for the Bumblebee. The receiver is not power constrained because it can sit at a base station. Fig. 2.10(c) shows the block diagram of the receiver. The receiver reads the data wirelessly transmitted from the Bumblebee, performs clock and data recovery, and reconstructs the analog signal. Simultaneously, it sends digital sample data over a standard USB link to a PC for further processing.

To power the entire system on a single coin-size battery, traditional sensor network Zigbee-type radios are unsuitable because they consume >20 mW during transmission as well as high average power dissipation. Since the energy density of low mass batteries is extremely limited, sub-mW power dissipation of the IC is mandatory for a reasonable lifespan.

We regularly utilize the Bumblebee to transmit biosignal data to a custom-designed USB-compatible 433 MHz receiver dongle. The chip runs for approximately 3 days continuously on one 0.17 g hearing aid battery.

2.5 in-vivo testing

We have tested the neural interface system described above in various realistic electrophysiological signal-recording scenarios, including two separate in-vivo neural-recording experiments from a rat and a mouse, EMG and electrocardiography (ECG) signal acquisition from humans. All these experiments have successfully verified the system’s compatibility with the high source impedance of neural electrodes. Because extreme care has been taken in the experiment setup (e.g. short input wire length, proper grounding, effective input high-pass filtering, etc.), no significant 50/60 Hz line noise was observed throughout the various experiments. BPA3 has proved to be extremely robust in terms of adequate ESD protection, immunity to ambient line noise, RF interference coupled from the on-chip/board
Figure 2.11: BPA3 tested in vivo in rat motor cortex. a) Recorded rat spike; b) two classes of spikes sorted by post-processing programs.
Figure 2.12: Neural recording from a mouse visual cortex through two adjacent electrodes. The top is recorded through a rack-mount wired setup, while the bottom reconstructed from the Bumblebee in real time.

supplies and substrate, and reliable start-up.

2.5.1 Spike-recording from a rat

In the first experiment, we recorded from a rat motor cortex using neural electrodes from NeuroNexus Technology. The tungsten electrodes are insulated with teflon, and have an equivalent impedance of 100-500 kΩ measured at 1 kHz [57]. After having identified active spiking cells from traditional rack-mounted instrumentation, we then began recording data directly from the VGA output. By keeping the input wires short and setting the high-pass corner to 200 Hz, no significant low-frequency interference was observed during the experiment. Fig. 2.11(a) shows a single recorded rat spike. Fig. 2.11(b) shows sorted spikes recorded through our prototype amplifiers. The two types of spikes line up well, demonstrating extremely high fidelity neural recording achieved from a 1 V supply with less than 28 μW power consumption per channel.

2.5.2 Spike-recording from a mouse

In the second experiment, the wireless Bumblebee board was used to record from a mouse visual cortex using NeuroNexus A-16 Series probes. The probes were implanted approx-
imately 800 μm into the brain. As the stimulus, a white moving bar on the computer monitor was presented to the mouse, simultaneous recordings were conducted through the Bumblebee and a conventional rack-mounted wired setup from two adjacent electrodes. The extra-cellular action potentials were adequately amplified, filtered, digitized and transmitted using the Bumblebee. Fig. 2.12 shows the neural recording from the wired setup (top) and the reconstructed neural signal received from the Bumblebee 3 meters away (bottom). Detected spike occurrences are also marked in the figure. With the two electrodes spaced apart only by 200 μm, close resemblance of neural activity can be observed between the two. The detected spikes from the Bumblebee’s recording have been used to successfully map the visual receptive field of mouse neurons in the visual cortex.

2.5.3 EMG-recording from humans

In addition to recording neural signals, we also verified usability of our prototype Bumblebee in other recording scenarios. For instance, we measured EMG signals from self-adhesive surface electrodes attached to the flexor and extensor muscles of a human arm. Fig. 2.13(a) shows two EMG traces recorded from a human flexing his arm in quick succession. We have incorporated this miniaturized wireless EMG recording device into a system that provides real-time feedback to subjects for rehabilitation therapy.

2.5.4 ECG-recording from humans

Similarly, We measured ECG signals from self-adhesive surface electrodes attached to a human chest. Fig. 2.13(b) shows the recorded ECG trace. Thanks to the VGA, we can adjust the AFE gain based on the input amplitude in various recording applications to maximize the signal-to-noise ratio (SNR) without incurring saturation. We have successfully demonstrated the usability of the system for signals with amplitudes as small as tens of μVs (spikes), and as large as a few mVs (EMG, ECG).
Figure 2.13: a) EMG traces from human flexor and extensor regions simultaneously captured, wirelessly transmitted by two Bumblebee chips, and reconstructed in real-time; b) human ECG captured, wirelessly transmitted by a Bumblebee, and reconstructed.
2.6 Discussion

Although this paper intends to focus on the circuit design of low-power BPAs, the analysis should also be considered in the context of a multi-channel system. In addition to increased area and power, the larger system also poses additional design constraints on the output impedance of the amplifiers interfacing with the ADC, transmit data rate, input impedance of the amplifiers, and crosstalk. To conserve area and power dedicated to the ADC, DSP, and transmitter circuitry, the outputs of front-end amplifiers are typically multiplexed before digitization.

The circuitry interfacing with the ADC needs to have low enough output impedance such that the sample-and-hold at the ADC input can settle within the acquisition time of the ADC (a few ADC clock periods) to an error smaller than the ADC resolution. We use a VGA between BPA3 and the ADC to enable the recording of a variety of electrophysiological signals. It also decouples the ADC drive requirement from the power consumption of the BPA. Thanks to a more relaxed noise specification, the VGA's current can be concentrated in the second stage, reducing the output impedance. The relationship between the (SAR) ADC acquisition time ($T_{acq}$) and the VGA output impedance ($R_{out}$) can be summarized in Eqn. 2.16, where $N_{acq}$ and $N_b$ represent the number of ADC clock cycles dedicated to acquisition and conversion, and $N_{ch}$, $f_s$, $C_{sample}$ represent the number of channels, sample rate per channel, and the size of the sampling capacitor at the ADC input.

$$T_{acq} = \frac{N_{acq}}{N_{ch} \cdot f_s} \cdot \frac{N_{acq}}{(N_{acq} + N_b)} = \ln (2^{N_b}) \cdot R_{out} \cdot C_{sample} \quad (2.16)$$

For instance, if 16 channels sampling at 10 kSamples/s/channel are muxed before an 8-bit SAR ADC with a 3-cycle acquisition time and a 2 pF input capacitance, the output impedance of the VGA is required to be < 400 kΩ. In the Bumblebee, for example, a 5 μW VGA was used to provide an output impedance that varies from 700 Ω to 50 kΩ as the gain varies from 0 to 38 dB, contributing sufficient drive strength. As discussed later, channel crosstalk also requires low output impedance of the amplifiers muxed prior to the ADC.

The input impedance of the front-end BPAs is mostly capacitive, and should be significantly higher than the impedance of the electrodes to minimize the input signal attenuation.
from the capacitive divider formed at the electrode-tissue interface. This is especially important in a multi-channel system, where one signal electrode is used per channel while one reference electrode is shared among all the channels. The size of the reference electrode is usually a few orders of magnitude larger than the signal electrode. As a result, a common interference signal (e.g. EM or power supply noise) presented to a signal electrode and the reference electrode is converted to differential signals at the BPA inputs as it experiences different amount of attenuation from the capacitive dividers [122]. When a closed-loop amplifier with capacitive feedback is used (e.g. BPA1 and BPA3), the input capacitance ($C_i$) within the signal bandwidth approximately equals the input DC-blocking capacitors ($C_s$). While the potential divider effect at the electrode-chip interface discussed above sets the upper bound for $C_s$, the capacitive divider at the OTA input sets the lower bound for $C_s$ to minimize the increase in the input-referred noise (Eqn. 2.1). The $C_i$ in BPA1, BPA2 and BPA3 provides impedances of ≈ 8 MΩ, 22 MΩ and 4 MΩ at 1 kHz respectively, high enough to interface with electrodes, while ensuring a sufficiently low input-referred noise.

Crosstalk is also an issue in a multi-channel system. First, there is crosstalk from the capacitive coupling between electrode probes. As feature sizes reduce, the electrical crosstalk due to electrical coupling also increases [91]. In our in-vivo experiments, we used tungsten electrodes with 200 µm spacing on a silicon substrate, which resulted in an approximately < 1 pF coupling capacitance ($C_c$). The impedance magnitude of the electrode ($C_e$) is roughly 0.5 MΩ at 1 kHz. Although we cannot measure crosstalk with our single-channel prototype, we can approximate the first-order crosstalk from the capacitive divider $C_c / (C_c + C_e + C_i)$ to be 0.3%, or -50 dB, where $C_c = 1$ pF, $C_e = 318$ pF, $C_i = 40$ pF. Secondly, there is crosstalk due to substrate (or supply) coupling. In the audio frequency range of interest here, these sources of coupling can be minimized through careful layout. Sufficient PSRR also ensures adequate rejection of the noise coupled in the supply. Additionally, there is crosstalk in the circuit block that mixes all the channels into the ADC. The off resistance of the switches forms a resistive divider with the input impedance of the ADC in parallel with the impedance of the other channels, leading to signal crosstalk. High off resistance of the switches and low output impedance of the mixed amplifiers are required to minimize the signal crosstalk. For our VGA/MUX/ADC, simulations reveal a crosstalk of < -80 dB,
reducing the input-referred mux crosstalk to below the noise floor.

2.7 Chopper-Stabilized Low-Noise Amplifiers

2.7.1 Circuit design

In addition to neural and body-area-sensing applications, there are a variety of other sensor interfaces that demand an extremely low noise floor (<2 μVrms input referred) under a relatively low bandwidth (<1 kHz). For instance, these applications include biosignal detection, thermocouple readout, and gas detection. When the signals of interest fall below a few hundred Hz, the dominating circuit noises shift from the thermal noise to 1/f and popcorn noise [28]. Excess low-frequency noise can undermine the system’s signal-to-noise ratio (SNR) and cause errors in the measurement. As a result, we use a chopper-stabilized topology to suppress 1/f noise and offset that plague sub-micron CMOS processes.

Closed-loop chopper-stabilization has been adopted recently [74] [139] [28] to suppress gain and sensitivity errors, as well as to prevent saturation due to amplifier offset. Among the recent implementations, [28] provides the best figure-of-merit so far. AC feedback is employed to ensure all signals entering the amplifier are well above 1/f noise corner. However, separate active input-biasing circuitry is used, and higher supply voltage is required due to single-ended approach.

As shown in Fig. 2.14 and Fig. 2.15, we employ a fully-differential closed-loop architecture to ensure sufficient linearity and supply rejection. Operating transistors in the subthreshold region enables the use of a power-efficient telescopic-cascode op-amp topology under low supply voltages. Signal up-conversion occurs at the gate of the input transistors, which are biased in weak inversion to maximize the transconductance. We introduce a novel dual-feedback technique to simultaneously set the mid-band gain of the amplifier through $C_{fb}$, and bias the amplifier’s input node through high-resistance pseudo-resistors. Chopper switches are included in both the signal and biasing paths to not only guarantee negative feedback around the amplifier, but also avoid using additional input-biasing circuitry as in [28]. We realize the chopper modulator with a minimally-sized CMOS transmission gate to minimize charge injection. The input capacitance ($C_{in}$) is 15 pF. When modulated with
Figure 2.14: Schematic of fully-differential chopper-stabilized low-noise amplifier

Figure 2.15: Chopper amplifier architecture
a 10 kHz chopper clock, the input impedance (1.06 MΩ) is high enough to avoid loading
the electrodes for biomedical applications. The ratio of $C_m$ and $C_{fb}$ establishes a 40 dB
mid-band gain. $C_{fb}$ is sized slightly smaller (140 fF) to take into account the addition of
parasitic and switch capacitances.

Two additional sets of chopper switches are added in the first stage of the amplifier: one
set of switches is placed at the drains of the input transistors to demodulate the ac signal
down to baseband and modulate the input offsets up to the chopper frequency; another
pair is placed at the drains of the PMOS current source to modulate their flicker noise
up to higher frequency. At the output of the amplifier, the signal returns to baseband
while the offsets and flicker noise are modulated up to high frequency and then filtered by
the amplifier’s 2nd-stage. The 2nd-stage is implemented as common-source to increase the
output swing under low supply voltages. The output is then fed back to the summing node
at the input of the amplifier after being modulated up to the chopper frequency. In order to
avoid large passive devices, we implemented continuous-time tunable Gm-C filters to reduce
ripple at the output of the amplifier. The input-referred noise from the ripple filter (Gm-C
filters) is designed to be negligible. The six achievable bandwidths of the Gm-C filters are
logarithmically spread between 150 Hz to 400 Hz. The tunability of the filters is realized
through digital control of the transconductor current.

2.7.2 Measurement Results

Fig. 2.16(a) plots the gain magnitude with chopping on and off. The mid-band gain for both
cases is approximately 38.5 dB. When the chopper clock is off, the amplifier operates as a
conventional AC-coupled amplifier and has a high-pass corner of 0.2 Hz. Amplification is
preserved down to DC when chopper-stabilization is enabled. The tunable low-pass corner
is set to 230 Hz in this measurement.

Fig. 2.16(b) illustrates the input-referred noise of the amplifier with chopping on and off.
Low frequency spot noise is reduced by more than a decade when the chopper is enabled.
The measured integrated noise from 0.05 Hz to 100 Hz is 1.25 µV when the chopper switches
are on, compared to 4.46 µV when the chopper switches are off.
Figure 2.16: Measured transfer function and noise plot of the low-noise chopper-stabilized biosignal amplifier.
2.8 Conclusion

Reducing the power consumption of BPAs while ensuring sufficiently low noise is essential in reducing the power consumption of a biopotential-recording system. In this chapter, we discussed the progression of three BPA designs: a closed-loop fully-differential telescopic-cascode amplifier (BPA1), an open-loop complementary-input (BPA2), and a closed-loop fully-differential complementary-input amplifier (BPA3) that leverages the salient design techniques of the first two amplifiers. The three BPAs exhibit low input-referred integrated noise of 3.2 $\mu$V, 3.6 $\mu$V and 2.2 $\mu$V while consuming 12 $\mu$W, 0.8 $\mu$W, and 12 $\mu$W respectively under 1 V supply. Consistent with theory, BPA1 using conventional architecture has an NEF comparable to the prior work. BPA2 and BPA3 achieve significantly better power-noise performance than BPA1 due to the power-efficient complementary-input topology. While the practical use of BPA2 is still an active area of research, BPA3 not only achieves superior power-noise tradeoff under low power supply conditions, but also ensures robust performance in realistic recording scenarios. We have demonstrated the robustness of BPA3 in a low power wireless system on chip that has been deployed in many in-vivo experiments. Lastly, we also present the design and implementation of a chopper-stabilized amplifier that exhibits lower flicker noise than the continuous-time BPAs. The scalability of the amplifier is also analyzed in the context of multi-channel biopotential-recording applications.
Chapter 3

ANALOG SIGNAL PROCESSING ON ECOG SIGNALS

3.1 Background on ECoG signals

Electrocorticography (ECoG) is a method of recording electrical brain activity using planar electrodes on the surface of the brain. The energy content generally concentrate below 200 Hz with amplitudes < 100 μV, although higher frequencies are also interesting to some neuroscientists [68][82]. ECoG uses non-penetrating electrodes, which offer an intriguing compromise between invasiveness and resolution that is receiving increased attention by the neuroscience community. Fig. 3.1 shows the characteristics of common bio-signals. ECoG electrodes may be more suitable for long-term recordings than intracortical single-unit electrodes, as ECoG recording quality may be less affected by electrode movement and tissue response [68]. In addition, the higher spatial resolution afforded by ECoG electrodes provides more specific information than noninvasive electroencephalogram (EEG) recordings and are less affected by muscular or external artifacts [57].

Table 3.1: Correlation between sub-banded spectral changes in EEG/ECoG and Motor events.

<table>
<thead>
<tr>
<th>Author</th>
<th>Frequency range</th>
<th>Spectral changes</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murthy [89]</td>
<td>25-35 Hz</td>
<td>fast oscillations</td>
<td>exploratory behavior</td>
</tr>
<tr>
<td>Donoghue [30]</td>
<td>15-50 Hz</td>
<td>increased oscillations</td>
<td>movement preparation</td>
</tr>
<tr>
<td></td>
<td>15-50 Hz</td>
<td>decreased oscillations</td>
<td>movement execution</td>
</tr>
<tr>
<td>Baker [10]</td>
<td>20-30 Hz</td>
<td>increased oscillations</td>
<td>precise grip</td>
</tr>
<tr>
<td>Aoki [7]</td>
<td>30-90 Hz</td>
<td>increased amplitude</td>
<td>movement execution</td>
</tr>
<tr>
<td>Pfurtscheller [99]</td>
<td>15-30 Hz</td>
<td>decreased amplitude</td>
<td>movement execution</td>
</tr>
<tr>
<td></td>
<td>60-90 Hz</td>
<td>increased amplitude</td>
<td>movement execution</td>
</tr>
<tr>
<td>Leuthardt [68]</td>
<td>40-180 Hz</td>
<td>direction-dependent increased amplitude</td>
<td>joystick movements</td>
</tr>
<tr>
<td>Röckert [107]</td>
<td>16-42 Hz</td>
<td>decreased amplitude</td>
<td>movement execution</td>
</tr>
<tr>
<td></td>
<td>63-200 Hz</td>
<td>direction-dependent increased amplitude</td>
<td>movement execution</td>
</tr>
<tr>
<td>Crone [27]</td>
<td>75-100 Hz</td>
<td>increased activity</td>
<td>movement execution</td>
</tr>
</tbody>
</table>
Brain Computer Interfaces (BCIs) can enable people with severe paralysis to communicate and/or manipulate objects, significantly improving their quality of life [68][39]. Several considerations have recently led researchers to consider using ECoG as the brain signal source for BCIs. First, the decoding of ECoG signals can be performed in the frequency domain on raw signals, in contrast to the spike sorting required for the decoding of single-unit activity. Furthermore, ECoG recordings are routinely made in human patients with medically-refractory epilepsy to identify the brain regions causing the seizures [132]. As such, the surgical expertise to implant ECoG electrodes is readily available and their safety in human patients has been verified. For instance, studies using ECoG data have shown two-class (tongue and hand) classification accuracies upwards of 90% [68][118], and five-class (individual finger) accuracies of 23% [117], with less than half an hour of training. In another study conducted at the University of Washington (UW), locations and frequency bands of ECoG activities associated with movements (or imagined movements) were identified in epileptic patients. The patients were then trained to use these ECoG activities to control cursors [68].

Spectral ECoG characteristics have been used to identify, classify, and analyze sensorimo-
tor activities. A number of studies in monkeys and humans have investigated fast oscillations of motor cortical field potential activity during movements. Table 3.1 shows the correlation between spectral changes of motor cortical areas in specific frequency ranges and movement intentions or executions. The usefulness of dividing spectral energy into multiple frequency sub-bands has been demonstrated. For instance, [107] suggests that high-frequency oscillations have to be divided into at least two functionally different regimes: one ≈30 Hz and one > 60 Hz. In [118][84], band power features in 11-40 Hz, 71-100 Hz and 101-150 Hz are used to distinguish between individual finger movements of one hand using ECoG signals [117].

These initial studies have been performed on bulky rack-mounted instrumentation and PC-based signal analysis. However, for the next generation of implantable BCIs, dedicated hardware and real-time signal processing is required. Because of the important role spectral content plays in ECoG research, a single-chip recording system that not only amplifies and conditions the signal, but also extracts sub-banded energy in multiple frequency bands would be useful for ECoG-based BCI applications. This real-time ECoG signal processing platform would need to dissipate less than 10 μW/channel to allow completely wireless powering from an inductive link. Biomedical implantable micro-stimulation devices have also been developed recently [64][65], which can be used in conjunction with the recording systems to close the loop in BCI applications. As previously presented in [145], an EEG/ECoG processing integrated circuit (EPIC) with 6.4 μW core power dissipation is primarily designed for, but not limited to, ECoG-based BCI applications. In this chapter, we provide detailed background information, discussion on circuit design and future directions. We will first focus on the system architecture and circuit design details, then present the testing results and in-vivo experiment data.

3.2 System Architecture

Various EEG experiments implement a general-purpose BCI platform (BCI2000) to process the conditioned signals with different linear spatial filters (e.g., Laplacian derivation, common average, etc.) and temporal filters (a slow wave filter, autoregressive spectral estimation, etc.) operations [77][73][104][61]. [81] measured the total integrated power for each electrode in various frequency bands, and processed the spectral power data of both
Figure 3.2: System architecture of an implantable ECoG/EEG recording unit to be used in BCI applications. It includes the proposed ECoG/EEG processing IC (EPIC), A/D converter, data and power telemetry.
movement and rest using a SVM classifier and sixfold nested cross-validation. The proposed chip implements a spectral decomposition method where integrated power is calculated in multiple frequency bands, compatible with the technique presented in [81]. Fig. 3.2 shows the system architecture of an implantable recording unit intended for EEG/ECOG-based brain computer interface (BCI) applications. This system includes the proposed EPIC chip that amplifies the signal and simultaneously extracts energy from four independently tunable bands. The outputs from all EPIC channels will be sequentially digitized and salient features can be extracted using digital signal processing. To avoid risks of infection, data should be wirelessly transmitted to a body-worn device or a remote station for further processing. Classifiers can then be applied to the transmitted spectral information to decode movement intentions. This recording unit can either be battery- or wirelessly-powered depending on the application.

As explained earlier, signal energy in key frequency bands is necessary to decode neuronal activities. In a wireless system, on-chip spectral decomposition would also reduce the amount of data that needs to be transmitted out of the recording chip, resulting in reduced power dissipation. This is particularly important if many channels are processed on chip. We partitioned the signal processing to combine the strengths of analog and digital signal processing techniques. Analog spectral decomposition reduces the signal bandwidth prior to the ADC to minimize power. After the signal is digitized, further processing can be done in a microprocessor to allow flexible programming. This work focuses on the design of the analog signal conditioning and signal processing circuitry, with an emphasis on the ultra-low power real-time analog spectral decomposition of four programmable frequency bands. A four-band prototype is fabricated as a proof of concept.

3.2.1 Design Specifications

The design specifications of the chip are dictated by the input signal/supply characteristics and the application requirements.

- With the trend towards integrating analog and digital subsystems on a single die, it has become increasingly important for analog circuitry to operate from the low supply
voltages (i.e. 1.2 V) typical of digital CMOS.

- The gain of the analog front-end (AFE) circuitry should be configurable (i.e. 40-80 dB) to accommodate a variety of EEG/ECOG amplitudes.

- The bandwidth should be $\geq 200$ Hz to pass the frequency of interest while filtering out high-frequency noise/interference.

- The input-referred noise of the sub-banded system should be much smaller than the signal (on the order of a few $\mu V_{rms}$ over the signal band of interest, e.g. 0.1-100 Hz).

- Sufficient common-mode rejection ratio (CMRR) and power supply rejection ratio (PSRR) (i.e. $\geq 80$ dB) to minimize coupling from common-mode interference (i.e. EM or power supply noise).

- To minimize attenuation of signals at the electrode-IC interface and crosstalk from neighboring electrodes, input impedance of the amplifiers should be high enough (a few M$\Omega$s to hundreds of M$\Omega$s depending on the electrode impedance).

- Considerations at the electrode-IC interface, including DC current blocking and ESD protection.

- Spectral decomposition must be reconfigurable over multiple key frequency bands. The center frequency ($f_c$) and bandwidth should be tunable to comply with a variety of neuroscience applications.

- Update rate of the energy extractors should be variable, and slow enough to reduce the total data throughput, but fast enough to capture changes in the energy profile (i.e. on the order of tens to hundreds of ms). [81][113]

In addition, there are also practical constraints on the design of the chip. First, the power consumption of the circuitry needs to be minimized to enable efficient wireless powering schemes and to limit tissue heating to within a few degrees Celsius. Secondly, DC leakage
current flowing through the electrode-IC interface should be minimized to avoid corrosion and charge imbalance in the tissue [125].

3.2.2 Brief Literature Review

We will present a brief literature review of a few state-of-the-art spectral decomposition system-on-chips (SoCs). Previous implementations of energy estimators have been reported in [9][43][128]. In [9], heterodyning dual-nested chopping architecture enables selectability of the center frequency ($f_c$) and bandwidth (BW), while eliminating offsets and low-frequency noise. However, the heterodyne chopping architecture requires two relatively power-hungry front-end low noise amplifiers to extract spectral power from one frequency band. In addition, the $f_o$ trim step size (5 Hz) is too coarse to realize stagger tuning [43] especially in $\alpha$ and $\beta$ bands.

In [43], Gm-C filters are used to realize a stagger-tuned fourth-order bandpass filter, followed by a squaring circuit and a leaky integrator. However, the $f_o$ and $Q$ of the bandpass filter (BPF) can not be independently tuned. In addition, the precise location of $f_o$ requires trimming due to process variation in the bias current.

In [128], local digital-processing extracts detection features, reducing the transmission-rate by 43×. Extraction of spectral energy from each EEG channel is implemented using a modulated filter formed by seven FIR filters, followed by a magnitude accumulator to drive the bin energy. The system generates EEG feature vectors every two seconds, consuming 9 $\mu$J per feature vector. The paper also demonstrated a complete EEG acquisition, feature-vector extraction, and final classification by configuring a one-channel system to detect relaxed eyes-closed and eyes-open states.

For completeness, it is worth mentioning that other efficient data compression techniques developed recently include performing signal activity dependent deterministic sampling by way of an asynchronous ADC [114][126]. Non-uniformly spaced samples in time result in the average sample rate lower than that of a conventional Nyquist ADC, creating the potential for large data compression and power reduction.
3.2.3 Spectral Decomposition Architecture

Fig. 3.3 shows the architecture of the energy extractor for one frequency band. It is composed of a 4th-order bandpass filter (BPF) that selects the frequency band of interest, a multiplier that squares the signal, and an integrator that sums the power within the band. The Butterworth response of the 4th-order BPF is realized by cascading 2nd-order BPF biquad sections. The two biquad sections are stagger-tuned to realize flatter frequency response [43].

To accommodate human subject variation and various classification algorithms, the frequency band selection, integration gain, integration duration should be fully programmable. Switched-capacitor (SC) techniques are used here to realize the fully-tunable band-pass filter and integrator. Very precise SC frequency responses are obtained without requiring external components or trimming because frequency response depends on monolithic capacitor ratios and clock frequency [4]. The continuous multiplier is employed also as an anti-aliasing filter between the two SC blocks. The two SC clock frequencies $f_{c1}, f_{c2}$ for the BPF and integrator should be carefully chosen to avoid aliasing.

3.3 Circuit Design

EPIC functions as a computational interface between the brain electrodes and A/D converter in a neural-recording system. As shown in Fig. 3.2, the system comprises four main blocks: a low-noise AFE that amplifies the input waveforms, four configurable band energy extractors,
clock generation, and control logic. All four band energy extractors are power- and clock-gated so that any or all of the bands can be enabled at a time. Fully-differential design is used throughout the chip to ensure sufficient output swing and improve the CMRR/PSRR for low supply voltages (1.2 V).

3.3.1 Analog Front-End (AFE)

![Figure 3.4](image)

Figure 3.4: a) Schematic of a closed-loop chopper-stabilized amplifier and the OTA [141]; b) schematic of the transconductor; c) schematic of the VGA.

ECoG/EEG-recording applications demand a low noise floor (< a few μV<sub>rms</sub> input-referred) and a relatively low bandwidth (< 200Hz). We employ a chopper-stabilized topology to suppress offsets and 1/f noise that dominate in low-frequency designs. Among these recently published chopper-stabilized amplifiers, [28] provides the best performance with the input-referred noise of 0.95 μV<sub>rms</sub> integrated over 0.05 to 100 Hz while consuming 1
µA. More in-depth comparison with these amplifiers are provided in [141]. Inspired by [28], the amplifier design in this work further simplifies the input-biasing circuitry and reduces the supply voltage to 1.2 V. Fig. 3.4(a) shows the closed-loop schematic of the chopper-stabilized amplifier. A 20 kHz chopper clock frequency is chosen to be significantly higher than the flicker-noise corner of the OTA (500 Hz) for effective flicker-noise reduction from the up/down-modulation of the desired signal. A fully-differential closed-loop topology is used to provide 40 dB of gain while ensuring sufficient supply rejection and signal swing under a 1.2 V supply [141]. A programmable Gm-C filter is used to reduce the switching ripple to below the noise floor. The schematic of the Gm cell is shown in Fig. 3.4(b). A detailed description of the LNA and Gm-C filter is provided in [141].

Input chopper switches are placed before the input capacitors $C_s$. Compared to a topology where the chopper switches come after $C_s$, our arrangement reduces the amplification of any OTA offsets that might saturate its output. In addition, mismatch in $C_s$ results in common-mode to differential-mode gain [128]. However, large DC offsets at the electrode interface will likely saturate the amplifier, as the chopper-stabilized amplifier is effectively DC-coupled. This problem becomes even more severe under the limited headroom afforded by a 1.2 V supply. To block any DC offset voltage at the interface, an off-chip 100 nF capacitor and a 10 MΩ resistor are used to form a high-pass filter between the electrode and the IC input. One drawback of this amplifier topology is that the input impedance is relatively low as the chopper switch conductance loads the amplifier input. Periodic steady-state simulations reveal that the input impedance magnitude with 15 pF $C_s$ is 4 MΩ (up to the -3dB bandwidth). When commercially-available ECoG electrodes are used, a higher input impedance is desired to further reduce loading of the electrodes. As described in [128], the order of the input capacitor and chopper switches can be interchanged to increase the input impedance and embedding the high-pass filtering on-chip, at the cost of degraded CMRR. However, this drawback can be alleviated by means of feedback [134]. In [140], the noise of the feedback paths is further reduced and the input impedance is increased by using a coarse-fine dual servo loop in an AC-coupled chopper-stabilized amplifier. Other low noise amplifiers achieved input resistance of 300 MΩ by using JFETs with low 1/f and channel thermal noise [69].
The output of the chopper amplifiers are fed into a fully-differential variable-gain amplifier (VGA). As shown in Fig. 3.4(c), the 6 programmable gains (0 to 34 dB) are set through the feedback capacitor $C_f$, and 7 programmable high-pass corners (< 1 Hz to > 300 Hz at the 0 dB gain setting) are set through the feedback transconductor.

3.3.2 Band Energy Extractor

Two stagger-tuned 2nd-order SC biquad sections are used to realize a programmable 4th-order BPF. Fig. 3.5 shows the simplified schematic of one 2nd-order SC biquad section. The biquad uses a noninverting integrator (switches $S_1$-$S_8$ and OTA1) and an inverting integrator (switches $S_9$-$S_{16}$ and OTA2). $\phi_{1,2}$ represent two non-overlapping phases. A dual-switch configuration (i.e. $S_{1,2,5,6}$) is used to reduce stray capacitance errors and charge
injection [35]. Specifically, parasitic capacitances at the left and right side of $C_3$ are either charged to the output of the OTAs or shorted to ground / virtual ground. As a result, the charge contribution to the integrator theoretically zero and the only charge transfer takes place via $C_3$. In order to minimize charge injection, minimum-sized transistors are used in all the switches. All the switches connecting to the virtual ground nodes are realized with NMOS-only switches, whereas the rest are realized as transmission gates. In addition, the switches connecting to the virtual ground nodes near the OTA inputs ($S_{5-8,9-12}$) are turned off first during $\phi_{1a,2a}$, so that the charge injected is the same from one clock cycle to the next and can be treated as a DC offset [51].

$f_o$ and $Q$ are determined by the capacitor ratios and the clock frequency Eqn. 3.1, Eqn. 3.2.

\[
f_o = \frac{C_3}{C_2} \frac{f_c}{2\pi}
\]

\[
Q = \frac{C_2}{C_1}
\]

To achieve a wide tuning range and fine tuning steps for $f_o$, coarse tuning adjusts the center frequency $f_o$ by varying the divided clock frequency $f_c$ in octave steps using an on-chip programmable clock generator. Fine tuning varies the $f_o$ within each coarse frequency step by controlling the capacitor banks $C_3$. Fine tuning provides 16 center frequencies spread logarithmically over each octave. This technique simultaneously provides a wide tuning range and fine frequency steps, which is consistent with the common frequency bands used in ECoG/EEG applications, where smaller frequency steps are required in the low frequencies to resolve consecutive bands within a few Hz of each other.

By adjusting the $C_1$ capacitor banks, the $Q$ is tunable from 0.5 to 8 over 16 linear steps. To save area, we use linear combinations of 7 capacitors to realize 16 $Q$ values (Eqn. 3.2). For instance, the combination of the $C_0$ and $C_1$ results in the eighth capacitance value. In $C_1$ capacitor banks, unused capacitors are connected to the output on one end, and virtual ground on the other to reduce switching glitches in the output [4]. However, in the $C_3$ capacitor banks, there is no need to ground unused capacitance because the capacitors are discharged every clock cycle [25].
The OTAs in the biquads are realized with fully-differential telescopic two-stage op-amps. To overcome op-amp output swing limitations and avoid resistive output loading of the op amp, switched-capacitor common-mode feedback (SC CMFB) is used here as a convenient and power-efficient approach to ensure a stable and well-controlled common-mode operation. Separate SC CMFB circuits are added to both the outputs of the first and the second stage to define their output common mode voltages. The OTA’s bandwidth should be at least 5 times the clock frequency $f_c$ to reduce the effects of finite OTA bandwidth on the switch-capacitor operation [76]. As a result, when $f_c$ is lowered to tune the biquads to lower frequencies, the bias currents in the OTAs are automatically reduced appropriately to save power.

Fig. 3.6 shows the multiplier circuit that squares the signals from the BPF. The BPF outputs are AC-coupled to reject DC offsets and enable independent DC biasing at the gates of $M_{12}$ and $M_{3-6}$. This block consists of a subthreshold CMOS Gilbert multiplier core ($M_{0-8}$) that squares the input voltage, a transimpedance stage ($M_{9-12}$) that converts current into voltage, and an output stage ($M_{13-16}$) to drive the SC integrator. The currents
Figure 3.7: a) The clock generation architecture. b) The design of a quadrature phase shifter.
in the transimpedance and output stages are mirrored from and scaled to 0.5× and 2× the current in the multiplier core. To ensure common-mode stability, 1.5 pF capacitors are added at the output of the transimpedance stage to form the dominant pole. Output common mode is sensed through two high-resistance pseudoresistors to avoid loading the output. Although the common-mode voltage is set through a negative common-mode feedback (CMFB) loop, the differential DC voltages at the output transimpedance and output stages might be different due to random transistor mismatches. For instance, the common-mode voltage can be set to 600 mV, whereas the differential outputs sit at 550 mV and 650 mV. A 3-b current DAC is added at the output of the transimpedance stage to trim the differential DC offsets. Since this is a DC trimming step, it does not affect the CMFB loop. The choice of current DACs are verified with Monte Carlo simulations to ensure proper CMFB operations.

Although a continuous-time running-average is an effective way to integrate signal energy, it has a few disadvantages. First, it either requires large area of passive components to realize the low-pass filter [9], or requires external trimming to obtain a precise time constant if active Gm-C filter is used [43]. Second, a running average is a lossy implementation that only approximately integrates the energy. Lastly, the output requires an additional sample-and-hold circuit to interface with the ADC. As a result, we used a SC technique to realize a lossless integrator with a precise and adjustable integration window and gain. An area-efficient approach [90] is employed to realize large integration time constants (Fig. 3.6). This technique effectively attenuates the input voltage by a factor $C_3/C_2$, then integrates it onto $C_2$ through $C_1$. The unity gain frequency of the integrator and the integration time constant are approximately given by:

$$f_u = \frac{1}{2\pi} \frac{C_1}{C_1 + C_2} \frac{C_3}{C_2} f_c$$

(3.3)

$$\tau = \frac{1}{2\pi f_u}$$

(3.4)

The integrator output is reset periodically by a variable reset clock, derived from the integrator SC clock. Because the gain of the integrator ($A_{int}$) can be approximated as the
ratio of the reset time window and the time constant \(T_{rst}/\tau\), \(C_2\) can be programmed to vary \(\tau\), and the integrator gain. \(C_1\) and \(C_3\) are chosen to be 200 fF and 686 fF respectively. Table 3.2 shows the settings of \(f_c\) and \(C_2\) that realizes various \(f_u\) and \(A_{int}\) with \(T_{rst}\) set 25 ms.

Table 3.2: SC Integrator frequency and gain settings

<table>
<thead>
<tr>
<th>(C_2) (pF) / code</th>
<th>(f_u) (Hz)</th>
<th>(A_{int})</th>
<th>(f_u) (Hz)</th>
<th>(A_{int})</th>
<th>(f_u) (Hz)</th>
<th>(A_{int})</th>
<th>(f_u) (Hz)</th>
<th>(A_{int})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.44 / 0001</td>
<td>23.92</td>
<td>3.76</td>
<td>47.83</td>
<td>7.51</td>
<td>96.67</td>
<td>15.03</td>
<td>191.33</td>
<td>30.05</td>
</tr>
<tr>
<td>0.67 / 0010</td>
<td>11.73</td>
<td>1.84</td>
<td>23.46</td>
<td>3.69</td>
<td>46.93</td>
<td>7.37</td>
<td>93.86</td>
<td>14.74</td>
</tr>
<tr>
<td>1.10 / 0011</td>
<td>4.75</td>
<td>0.75</td>
<td>9.5</td>
<td>1.49</td>
<td>19</td>
<td>2.98</td>
<td>38</td>
<td>5.97</td>
</tr>
<tr>
<td>0.96 / 0100</td>
<td>6.09</td>
<td>0.96</td>
<td>12.19</td>
<td>1.91</td>
<td>24.38</td>
<td>3.83</td>
<td>48.76</td>
<td>7.66</td>
</tr>
<tr>
<td>1.40 / 0101</td>
<td>3.66</td>
<td>0.48</td>
<td>6.12</td>
<td>0.96</td>
<td>12.24</td>
<td>1.92</td>
<td>24.49</td>
<td>3.25</td>
</tr>
<tr>
<td>1.62 / 0110</td>
<td>2.31</td>
<td>0.36</td>
<td>4.62</td>
<td>0.73</td>
<td>9.23</td>
<td>1.45</td>
<td>18.47</td>
<td>2.9</td>
</tr>
<tr>
<td>2.06 / 0111</td>
<td>1.47</td>
<td>0.23</td>
<td>2.94</td>
<td>0.46</td>
<td>5.89</td>
<td>0.92</td>
<td>11.77</td>
<td>1.85</td>
</tr>
<tr>
<td>1.36 / 1000</td>
<td>3.22</td>
<td>0.5</td>
<td>6.45</td>
<td>1.01</td>
<td>12.89</td>
<td>2.03</td>
<td>25.79</td>
<td>4.05</td>
</tr>
<tr>
<td>1.79 / 1010</td>
<td>1.91</td>
<td>0.3</td>
<td>3.82</td>
<td>0.6</td>
<td>7.65</td>
<td>1.2</td>
<td>15.29</td>
<td>2.4</td>
</tr>
<tr>
<td>2.02 / 1011</td>
<td>1.52</td>
<td>0.24</td>
<td>3.05</td>
<td>0.48</td>
<td>6.1</td>
<td>0.96</td>
<td>12.2</td>
<td>1.92</td>
</tr>
<tr>
<td>2.45 / 1011</td>
<td>1.06</td>
<td>0.16</td>
<td>2.1</td>
<td>0.33</td>
<td>4.2</td>
<td>0.66</td>
<td>8.4</td>
<td>1.22</td>
</tr>
<tr>
<td>2.31 / 1100</td>
<td>1.18</td>
<td>0.18</td>
<td>2.35</td>
<td>0.37</td>
<td>4.7</td>
<td>0.74</td>
<td>9.4</td>
<td>1.48</td>
</tr>
<tr>
<td>2.75 / 1101</td>
<td>0.84</td>
<td>0.13</td>
<td>1.69</td>
<td>0.27</td>
<td>3.37</td>
<td>0.53</td>
<td>6.75</td>
<td>1.06</td>
</tr>
<tr>
<td>2.97 / 1110</td>
<td>0.72</td>
<td>0.11</td>
<td>1.45</td>
<td>0.22</td>
<td>2.9</td>
<td>0.46</td>
<td>5.8</td>
<td>0.91</td>
</tr>
<tr>
<td>3.40 / 1111</td>
<td>0.56</td>
<td>0.09</td>
<td>1.11</td>
<td>0.17</td>
<td>2.22</td>
<td>0.35</td>
<td>4.45</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Properly choosing SC clock frequencies \(f_{c1}, f_{c2}\) for the BPF and integrator is important for avoiding aliasing and capacitor droop. Although the continuous-time multiplier is able to filter out the high-frequency harmonics of the BPF spectra, the output contains energy at \(2 \times f_o\) as well as the desired signal at DC. The high-frequency side-band at \(2 \times f_o\) is subject to aliasing when processed by another sampled-data system with clock frequency \(f_{c2}\). In general, \(f_{c2}\) should be greater than \(f_{c1}\) to avoid aliasing. However, because \(f_o\) and BW of the SC BPF and \(f_u\) of the SC integrator are proportional to \(f_{c1}\) and \(f_{c2}\) respectively in our implementation, the selection of high \(f_o\) and low \(f_u\) often result in \(f_{c1} > f_{c2}\). In this case, to the first order, \(f_{c2}\) needs to satisfy Eqn. 3.5

\[2f_o + BW/2 \leq f_{c2} - f_u\]  \(\text{(3.5)}\)
For instance, for a $f_0$ of 400 Hz, a bandwidth of 200, and $f_u \ll f_{c2}$, $f_{c2}$ should be higher than 900 Hz. Note that care should especially be taken at a high $f_0$, low Q and slow integration rate to avoid aliasing.

3.3.3 Clock Generation

As shown in Fig. 3.7(a), four sets of independently-programmable BPF and integrator clocks are derived from a 20 kHz clock, divided down from a 160 kHz on-chip crystal oscillator. The crystal oscillator employs a Pierce topology, where the crystal is connected to the input and output of a self-biased inverter. 10 pF capacitors are also connected to the crystal to ensure proper oscillator start-up. Since the oscillating frequency is low (160 kHz), the power consumption can be kept below 1 $\mu$A. Full integration can be achieved after a few modifications. Specifically, the PVT-compensated ring oscillator with calibration could replace the reference generation using an external crystal oscillator since neither jitter nor absolute frequency accuracy is critical. A 20 kHz clock is chosen to maintain compatibility of the chopper-stabilized amplifier and all the SC blocks. We chose 20 kHz as a good compromise between dynamic power dissipation and capacitor voltage droop during sample-and-hold operations.

Because discrete-time SC operations are sensitive to sampling instants, the SC integrator clocks are shifted by 90 degrees to avoid sampling transients from the SC BPF. Fig. 3.7(b) shows the design of the quadrature phase clock generator. The exact phase errors of the two clocks don't matter as long as the two phases are apart enough to avoid sampling transients. Therefore, the I/Q generation circuits do not require calibration because the phase errors are within the expected PVT variations. Programmable SC BPF clocks are derived from the I-phase of the 20 kHz clock, while programmable SC integrator clocks are derived from the Q-phase, both through a series of divide-by-2 blocks. The four SC BPF and SC integrator clocks in the four bands can be independently controlled to realize different frequency responses. To ensure proper SC operation, non-overlapping phases $\phi_{1,2}$ and $\phi_{1a,2a}$ (early $\phi_{1,2}$ phases) are derived for each SC BPF and integrator.
3.3.4 Control Logic

On-chip control logic synchronizes the multiplexing of the four integrator outputs and supplies the integrator reset pulses. The same integration window is chosen for all four integrators to allow their outputs to be multiplexed in a round-robin fashion. The max sequentially cycles through all four bands once per integration window. The synchronization pulses that denote the end of the integration periods are also multiplexed before the A/D converter. In each band, the integrator reset pulse occurs one clock cycle after the corresponding synchronization pulse. Both the synchronization and reset pulses are clocked by the \( \phi_1 \) SC integrator clock.

3.1 Experimental results

![Image of a chip](image-url)

Figure 3.8: Microphotograph of the chip.

Our prototype chip was fabricated in a 0.13 \( \mu m \) CMOS process. The total current draw is roughly 9 \( \mu A \) from a 1.2 V supply: 2.7\( \mu A \) from the comparator amplifier/VGA, 2.6 \( \mu A \) from the four energy extractors combined, 2.7\( \mu A \) for bias generation, and approximately 1 \( \mu A \) for everything else (crystal oscillator, digital clock generation and logic). The die photo is shown in Fig. 3.8. The total active area is 1400 \( \mu m \) by 2000 \( \mu m \), and each energy...
extractor occupies 0.46 mm\(^2\). The system-level performance is summarized in Table 3.3. The characterization of the front-end amplifiers can be found in [141]. Simulated PSRR is better than 60 dB over the signal frequency bands including mismatch effects. Furthermore, large supply decoupling capacitors (up to a few nF's) guarantee sufficiently low supply ripple.

3.4.1 Characterization of the BPF

The tuning and programmability of the center frequency \(f_o\) and Q of the 2nd-order biquad sections were verified by changing the clock frequency \(f_c\) and capacitor ratio. All programming is done through a two-pin serial interface. In Fig. 3.9(a), the lowest \(f_o\) in each coarse frequency band ranging from 3.125 Hz to 200 Hz are shown. This plot shows the coarse tuning of \(f_o\) by stepping up \(f_c\) in octave steps while setting Q to 8. Similarly, the fine tuning capability is demonstrated in Fig. 3.9(b), where the \(f_o\) is changed through selecting different \(C_3\) values in the corresponding capacitor banks. The 16 BPF transfer functions are plotted in the 200-400 Hz frequency band with Q set to 8. Notice that the coarse and fine center frequencies are both spaced logarithmically, tailored to the characteristics of EEG/ECoG signals. This way we can achieve finer frequency steps in the \(\alpha / \beta\) bands, and wider steps in the higher frequency ranges. Lastly, the Q tuning is verified by programming capacitor \(C_1\) values in the corresponding capacitor banks with \(f_o\) set to 200 Hz (Fig. 3.9(c)). As the Q increases linearly from 0.5 to 8, the bandwidth decreases harmonically from 400 to 25 Hz.

3.4.2 Multi-band Spectral Analysis

The \(f_o\) and Q of the 2nd-order biquad sections are tuned to realize an overall 4th-order Butterworth BPF responses. BPF responses in four common physiological frequency bands used in ECoG research (\(\alpha\) or 8-12 Hz, \(\beta\) or 18-26 Hz, low-\(\gamma\) or 30-50 Hz, and \(\gamma\) or 70-100 Hz) are synthesized by configuring the two biquads to slightly different frequencies (Fig. 3.10(top)). Close resemblance between the measured (solid line) and the theoretical (dotted line) BPF responses were observed.

Fig. 3.10(bottom) illustrates the noise spectrum at the output of the VGA and BPF
Figure 3.9: Measured frequency responses of one biquad: a) coarse $f_o$ tuning when $f_c$ is varied; b) fine $f_o$ tuning when $C_3$ is changed; c) Q tuning when $C_1$ is changed.
Figure 3.11: Top: measured frequency response (solid) compared to theoretical (dotted) of BPF tuned to $\alpha$, $\beta$, low-$\gamma$ and $\gamma$ bands; bottom: measured input-referred noise plots of LNA, VGA and BPF tuned to these four bands.
when referred to the LNA input. The BPFs are tuned to the four bands described above. The AFE is set to the maximum gain in order to characterize the noise performance with the weakest input signals. Flicker noise and lower bias current used in the $\alpha$ and $\beta$ bands results in slightly higher noise. The input-referred noise when the BPF is tuned to $\alpha$, $\beta$, and the two $\gamma$ bands is integrated well above and below the high and low 3dB bandwidth of the BPF. The rms noise voltages are respectively $0.35 \mu V_{rms}$ (1 - 50 Hz), $0.36 \mu V_{rms}$ (5 - 55 Hz), $0.37 \mu V_{rms}$ (20 - 120 Hz) and $0.41 \mu V_{rms}$ (40 - 240 Hz). We have extrapolated our measure noise profile to the bandwidth of the LNA to obtain more realistic rms noise voltages. The input-referred noise integrated from 1 - 400 Hz is respectively $0.36 \mu V_{rms}$, $0.46 \mu V_{rms}$, $0.4 \mu V_{rms}$, $0.42 \mu V_{rms}$.

To verify the functionality of the energy extraction chain, we tested the chip with a 6-second pre-recorded human EEG waveform corresponding to an epileptic patient moving fingers [80]. The signal was recorded from a subdural electrode array placed on the surface of the brain. BCI2000 [115] provided the patient with a visual stimuli, acquired brain signals from the Synamps2 amplifiers (Neuroscan), and also recorded the flexion of fingers using a data glove (Fifth Dimension Technologies). Predictions of finger movements were made based on various characteristics of the recorded waveform, including the spectral changes in the relevant frequency bands.

The four SC BPFs were configured to realize 4th-order Butterworth responses in the salient frequency bands described above: 8-12 Hz, 18-26 Hz, 30-50 Hz, and 70-100 Hz. The integration window of the SC integrator was set to 25 ms – long enough to reduce the data rate (160 Hz for 4 bands), and short enough to capture rapid changes in the narrowband energy profile. Note that the update rate is slightly higher compared to a recent experiment that processed the spectral power data by taking the FFT with 100 ms step size (10 Hz/channel) [81] or one that updates the cursor movement once every 40 ms [68]. The output from the four SC integrators were then multiplexed in a round-robin fashion. The energy extraction chain with the same BPF response, multiplier gain, and integration window settings is also modeled using Matlab. The output of the software model serves as a reference to compare against the silicon measurement results.

Fig. 3.11(a) shows the raw waveform (top) and the integrated power in the 70-100 Hz
Figure 3.11: a) Increased activity in 70-100 Hz range correlates to the onset of movement; b) theoretical (dashed) and measured (solid) energy profile in a) signal, b) $\alpha$, c) $\beta$, d) Low $\gamma$, e) $\gamma$ band for a 6-second ECoG waveform [80].
gamma band measured from our chip (bottom). The measured (solid) and the modeled (dashed) responses show good matching. The chip output shows a significant spectral energy increase that correlates well with the patient moving their fingers. We have also measured the integrated power in the other salient frequency bands. Fig. 3.11(b) shows the measured (solid) along with the modeled (dashed) energy-extractor responses from four simultaneously recorded frequency bands. Close resemblance between the two responses are observed in all four bands. The matching in Alpha band is compromised somewhat by the strong presence of flicker noise. The worse matching in the $\gamma$ band can be attributed to a lower signal level as the attenuation through the electrode-tissue interface is more severe at higher frequencies. The energy increase in the $\beta$ and low-$\gamma$ bands correlates to the patient’s preparation for movement after given a visual cue. Note that the relationship of the spectral changes in different frequency bands and the onset of movement is in accordance with the findings in various neuroscience studies [30][7][99][107].

3.5 in-vivo experiment results

An in-vivo experiment was conducted at the National Primate Research Center located at University of Washington, Seattle. An in-vivo ECoG recording was performed from the primary motor cortex of a pigtailed macaque monkey (Macaca nemestrina), through a 32-electrode sub-dural array (PMT Corp.). The electrodes were platinum, had an exposed diameter of 0.075 mm and impedances of 50-100 k$\Omega$s measured at 1 kHz. In a sterile procedure with the monkey under general anesthesia, insulated leads connected to each of the 32 electrodes of the PMT array were routed up through the dura and skull and crimped to pins in a connector that was secured with acrylic to the skull. The entire implant area was then sealed with acrylic. The signal from each electrode was accessed via this connector in subsequent recording sessions with the awake animal by means of short insulated wires with mating sockets crimped on one end.

Fig. 3.12 shows the setup for the in-vivo experiment. To assess the fidelity of the signals recorded using our chip, we used a 24-bit commercial precision ADC (gUSBamp, Guger Tech) to simultaneously digitize the energy extractor outputs from the chip and the signals recorded directly from the electrodes. We split the signal coming from the electrode to
In-vivo Experimental Setup

Figure 3.12: in-vivo experiment setup.
the EPIC and the gUSBamp. The input impedances of the gUSBamp (>100 MΩ) is high enough to avoid loading of the electrodes. Better recording quality can be achieved if an IC with a higher input impedance (i.e., >100 MΩ) were used.

The ECoG electrode was connected to our chip through a 0.16 Hz high-pass filter (similar to the one used in [81]) formed by an off-chip 100 nF capacitor and a 10 MΩ resistor. The negative input was grounded. The chip ground was connected to the monkey skull screw.

Fig. 3.13 shows two 4-second segments of in-vivo ECoG data from an awake monkey simultaneously recorded with the chip (EPIC) and the commercial ADC (gUSBamp),
Fig. 3.13(a)(top) shows activity in the $\beta$ / low-$\gamma$ band (segment I); Fig. 3.13(b)(top) shows ECoG activity with no increased frequency-specific components (baseline, segment II). The amplitudes in both plots are normalized to account for different gains in the two paths. Also shown in Fig. 3.13(a)(bottom) and Fig. 3.13(b)(bottom) are the difference in the recordings from EPIC (red) and gUSBamp (blue). Although we could see that the amplitude differences between the two paths are smaller than the signals by more than an order of magnitude, the difference can be better understood in the frequency domain. Fig. 3.13(c) and Fig. 3.13(d) illustrate the power spectral density (PSD) of the amplitude differences in the two segments respectively. We can observe that in both cases, the error power is less than -190 dB across all frequencies of interest (5-300 Hz), indicating insignificant degradation in signal fidelity. The error power magnitude in frequencies lower than 45 Hz is larger in segment I compared to segment II because higher frequency-specific activity is observed in segment I. However, the error power magnitude in frequencies higher than 45 Hz demonstrate similar profile in the two segments due to the relatively lower signal power in the higher frequency bands.

Fig. 3.14 further illustrates the PSD of the two segments recorded from EPIC and gUSBamp. The increased power in the 15-45 Hz range is pronounced in the PSD of segment I compared to segment II, indicating $\beta$ / low-$\gamma$ band oscillatory activity in segment I. The first segment was taken during movement preparation – right before the monkey reached out his hand for food. This power increase corresponds well with the correlation between spectral change and movement intention explained earlier.

### 3.6 Discussion

#### 3.6.1 Multi-channel design considerations

The prototype chip performed as expected, both on the lab bench and in a primate-lab setting. However, since the prototype is single-channel, we should also analyze its scalability in the context of a multi-channel system. Instantiating multiple channels (Fig. 3.2) incurs a significant area / power penalty and increased configuration complexity. A more area/power-efficient architecture is shown in Fig. 3.15, where the wide-band raw amplified
signals from all the channels can be multiplexed before directly passed to a fast ADC (mode a), and the signal power from 16 programmable bands is averaged and multiplexed before digitized by a relatively slow ADC (mode b). Since only a few channels are used at the same time for a specific task in an experiment, any in channels with any n bands of energy calculated (max 16) can be flexibly selected.

The circuitry driving the ADC (the VGA or integrator) needs to have a low enough output impedance such that the sample-and-hold at the ADC input can settle within the acquisition time of the ADC to an error smaller than the ADC resolution. The requirement will be more relaxed when the ADC clock is lowered digitizing band energy outputs. As a result, the current in the VGA can be reduced while maintaining proper sample-and-hold settling at the ADC input. In EPIC, for example, a 1 μW VGA was used to provide an output impedance that varies from 3.5 kΩ to 250 kΩ as the gain varies from 0 to 32 dB, contributing sufficient drive strength for both mode (a) and (b).
3.6.2 Digital energy extractor

In general, it is more feasible to incorporate flexible computational capabilities in the digital domain. For this application, however, a very specific operation is needed (spectral decomposition). Thus, it is feasible to perform the computation using dedicated analog circuitry. To make a fair comparison, we have to take area, power and design complexity into consideration.

A good example of a digital implementation of an energy extractor is given in [128]. It uses a modulated filter bank formed by seven FIR filters (with fixed band centers from 2-20 Hz). Each filter is followed by a magnitude accumulator. The design is fabricated in a 0.18μm CMOS process. The EEG feature vectors are generated once every 2 seconds, consuming 2.1 μW at 600 Hz ADC sampling rate. Although this digital implementation has lower power consumption when compared to the proposed analog implementation (2.6 μW for four bands), we should take into account the power consumption of other circuit blocks. In a multi-channel system, the power consumption of the ADC and ADC drivers
becomes significant as the ADC sampling rate increases.

Miller et al derived a mathematical model based on empirical observations of ECoG data indicating a power-law characteristic [82]. Sub-band the spectrum into four sub-bands before digitization reduces the dynamic range of the signal power, therefore reduces the resolution requirements of the ADC. Using the four identified physiological frequency bands discussed earlier as example, the SNR difference between the highest and lowest frequency bands of interest (70-100 Hz and 8-12 Hz) is $< 20$ dB (Fig. 3.16). This difference in SNR approximately translates into 3 bits in a data converter [3]. Assuming the resolution requirements of the highest frequency band are fixed, digitizing the sub-banded energy saves approximately three bits compared to directly digitizing the raw signal as a result of a lower dynamic range. A lower number of bits used in the data-converter and the additional digital circuitry leads to a significant savings in the area and power consumption of a SAR ADC. 

In addition, because data reduction occurs prior to the ADC in the case of analog energy extraction, the sample rate of the ADC and additional digital signal processors can also be lowered, saving dynamic power. Additionally, power in the circuitry prior to the ADC can also be significantly lowered due to the relaxed drive strength requirement at a lower ADC clock frequency. The considerations given above suggest that a significant reduction in
Table 3.3: System performance comparison

<table>
<thead>
<tr>
<th>Topology</th>
<th>Supply</th>
<th>FeEx, Total Power/CH</th>
<th>Total Gain</th>
<th>Input-referred noise</th>
<th>CMRR</th>
<th>PSRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>This work</td>
<td>ASP</td>
<td>1.2</td>
<td>3.1 µW, 10.8 µW</td>
<td>40-74 dB</td>
<td>0.46 µVrms</td>
<td>&gt;80 dB</td>
</tr>
<tr>
<td>[28]</td>
<td>ASP&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.3-3.3</td>
<td>4 µW, 45 µW</td>
<td>54-80 dB</td>
<td>0.5 µVrms</td>
<td>&gt;80 dB</td>
</tr>
<tr>
<td>[138]</td>
<td>ASP&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2</td>
<td>10.6 µW, 26.5 µW</td>
<td>58-70 dB</td>
<td>85 nVrms/√Hz</td>
<td>&gt;105 dB</td>
</tr>
<tr>
<td>[128]</td>
<td>DSP&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
<td>2.1 µW, 9 µl</td>
<td>72 dB</td>
<td>1.3 µVrms</td>
<td>&gt;60 dB</td>
</tr>
<tr>
<td>[22]</td>
<td>DSP&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1</td>
<td>1.4 µW, 14.4 µW</td>
<td>46-80 dB</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<sup>a</sup> 4 CHs of 1-band energy extractor with programmable $f_o$ from DC-500 Hz in 101 linear steps and $BW$ from 5-23/3-15 in 6/7 linear steps

<sup>b</sup> 1-band energy extractor with activity-based sampling and electrode-tissue impedance monitoring

<sup>c</sup> 7 FIR filters with programmable $f_o$ from 2-20 Hz followed by magnitude accumulators

<sup>d</sup> 16 CHs of temporal/spatial pre-filters, followed by 4 types of temporal/spatial feature extractors and classifier

Power dissipation can be achieved using an analog implementation at the expense of larger chip area.

3.6.3 Performance comparison

Table 3.3 shows the system performance, along with a few similar work published recently, among which [9][138] use analog approaches, and [128][22] use digital approaches to achieve power-efficient feature extractions (FeEx). For the reasons given previously, we can observe that the digital implementations in general have lower power in the FeEx circuitry but higher power in the rest of the system when compared with the analog implementations. However, due to the powerful computational capability of DSP blocks, more complex operations are achieved in the digital implementations. When compare this work with the analog approaches, this work simultaneously extracts the energy profile in four independently selectable bands, whereas both [9] and [138] provide energy profile in only one band at a time. Compared with [9], our chip implements logarithmic tuning of $f_o$, which is more suitable for achieving fine steps in the low-frequency physiological bands (i.e. $\alpha$, $\beta$). This chip offers more programmability with lower energy/band and smaller active area. By employ-
ing new circuit techniques such as activity-based sampling and low-power electrode-tissue impedance monitoring, [138] achieves further power saving while improving the robustness against motion artifacts. However, it requires a supply voltage of 2 V, compared with 1.2 V used in our work.

3.7 Conclusion

ECoG implants, already used extensively in the clinical diagnosis of epilepsy, have recently demonstrated promising results foreshadowing potential use in BCIs. ECoG recording potentially provides a more robust and less invasive signal than single cell activity while providing more specific information than noninvasive EEG. Neuroscientists have found a strong correlation between spectral changes of ECoG in certain physiological frequency bands and movement or movement intentions.

Based on these considerations, we designed a four-band ECoG signal conditioning and processing platform with 6.4 \( \mu W \) power dissipation and 0.46 \( \mu V_{rms} \) input-referred noise floor integrated over ECoG/EEG-relevant frequency bands. This chip conditions the signal and simultaneously extracts energy from four programmable bands, potentially saving power over a similar algorithm performed in the digital domain. Detailed discussion on multi-channel design considerations was provided. Bench testing and in-vivo experimental results confirm the scientifically established correlation between spectral changes and movement intentions.
Chapter 4

WAFER-SCALE PACKAGING FOR FBAR-BASED OSCILLATORS

For over 20 years, FBAR or SMR-BAW (Solidly-Mounted Resonator Bulk Acoustic Wave) with its high fQ product has been proposed as an attractive alternative to conventional approaches for timing applications [31][54]. Like other mm-scale resonators, FBAR requires a hermetic package in order to maintain frequency stability. The FBAR is also not readily integrated with circuit technologies for the simple reason that the processing that produces the FBAR—high resistivity, high impedance electrodes; high dielectric constant for the piezoelectric material; and highly variable thicknesses depending on the frequency target—does not produce a viable electronic device. Although integrating a Bulk Acoustic resonator or filter on the same substrate as an IC has been demonstrated [33][32][2], there are three problems with this approach: 1) compromises between technologies must be made to eliminate interference between processing steps (e.g., temperature limitations), 2) the economics for integrating the two technologies on the same wafer is unfavorable [1], and 3) none of the previous work addressed the issue of how to create an open-air cavity around the BAW device while simultaneously providing a hermetic and robust package.

The previous work demonstrated the potential value of an FBAR [109], a wafer scale hermetic package [108], and a modified FBAR demonstrating temperature stability comparable to quartz (referred to as Zero Drift Resonators, or ZDR [98]). Low noise sub-mW oscillators using FBAR have also been demonstrated [101]. We now present a technique for integrating active circuitry into the lid of the wafer-scale hermetic FBAR package while the FBAR resonator, ZDR, or filter resides on the base wafer. Both sides of an opened and unfolded module are shown in Fig. 4.1. The 285μm×330μm active circuit area is clearly visible, as is the location for the FBAR resonator. This strategy is IC technology agnostic. Besides an area limitation, there are no special circuit design rules required for our technique, and process modification for the lid/IC wafer is minimal compared to a full SOC.
(system-on-a chip) implementation.

4.1 The first “FMOS”

4.1.1 Process description

The wafer fabrication uses existing processes where possible, which has the advantages both of simplicity and of maintaining native device performance. In the Avago microcap process, the FBAR is fabricated on one wafer while a second lid wafer contains through-wafer vias, Au pads, sealing structures, and a recessed air cavity above the FBAR.

Fig. 4.2 contains a simplified FBAR/micro-cap process flow. On the FBAR wafer side, the cavity under the FBAR (swimming pool) is defined and filled with a sacrificial oxide, the resonators are processed, followed by the patterning of Au interconnect and sealing material and, finally, the removal of the sacrificial oxide. The Si micro-cap lid is then Au-diffusion
bonded to the FBAR wafer and pads patterned on top. The lid wafer is manufactured with standard micromachining processes and through-wafer vias are etched to make external pad connections. Both the FBAR and lid wafers are high-impedance Si to minimize cross-talk and capacitive losses.

Fig. 4.3(a) contains a schematic flow of the circuit containing lid wafer. The process flow proceeds as normal with minor modifications for pattern density. Inner-layer dielectric is left in the field until the devices are complete, after which they are removed down to Si. At that point the lid micro-machining can be completed. The cavity etch that leaves a gap above the FBAR removes any remaining epi Si between contact pads. The final oscillator die with integrated circuitry is shown in Fig. 4.3(b). We retained the Au-Au wafer bond that has proven to provide a robust hermetic seal T. The circuitry on the lid aligns with a corresponding depression in the FBAR wafer to provide clearance for wafer bonding. Since the depression is patterned along with the depression under the FBAR, no additional masking steps are required. Once the lid is bonded to the FBAR wafer, the lid Au metallization forms the interconnect between the circuitry and the FBAR. Fig. 4.3(c) contains a SEM cross section of a bonded die showing the relative position of the FBAR.
and the lid circuitry (reversed from the schematic). For this demonstration, our circuits used Avago's HP25 silicon bipolar process.

4.1.2 Circuit Implementation

The concept of integrating active circuitry into the lid has the economic advantage of reusing the lid wafer area for the electronics' pads and from eliminating external connections between the FBAR and the circuit. The size can be further reduced by placing the FBAR directly above the circuitry, which we hope to demonstrate in future iterations. Further economic advantage is gleaned by forcing the circuit size and the resonator (or filter) size to be comparable. The number of external connections is limited by size; a reasonable via count would be ten or less. In this work, we have six external pads and two internal connections from the lid electronics to the FBAR resonator.

Fig. 4.4A) shows the Pierce oscillator and divider that are integrated into the lid. All bias circuitry is integrated and derived internally. The oscillator and bias consume 150μA from a 2V supply. Emitter/Base diode varactors are used for tuning, and a temperature-sensing diode is included for on-die temperature monitoring. The oscillator is AC-coupled to an on-chip divide-by-64 circuit comprising six cascaded stages of divide-by-two bipolar current mode logic (CML). The oscillators operate at 1.5 GHz, yielding a divided clock source of 23.4 MHz. Each divide-by-two circuit includes two cross-coupled D-latches, with the clock input of each latch driven by a level shifter to set the proper DC bias of the input transistors. The maximum operating frequency of the divider-by-two circuit is a function of the D-latch delay and level-shifter delay. The bias currents of the latch and of the level shifter are set accordingly. Because each stage works at half the frequency of the previous one, the bias current of subsequent stages are scaled to reduce power consumption without degrading maximum operating frequency.

A fully-differential Colpitts oscillator was also designed and fabricated in this process. Fig. 4.4B) shows the simplified schematic. The Colpitts oscillator has superior cyclostationary noise properties and is thus a good candidate for low jitter reference clocks. However, the Colpitts oscillator requires a higher initial loop gain for oscillator startup. We utilize
Figure 1.3: A) Schematic circuit lid flow; B) micro-capped oscillator die; C) cross-section STEM of FBAR/bipolar process showing completed die.
a $g_m$-boosting technique to reduce power consumption by improving the oscillator startup characteristic. As shown in Fig. 4.4B), the FBAR resonator is connected between the bases of the top transistor pair. The center point of the FBAR resonator can be viewed as a virtual ground. Differential operation is sustained when the two sides oscillate at opposite phases. By cross-coupling the oscillator output and AC-coupling the signals to the base of the foot transistor pair, the effective loop gain of the oscillator is increased to reduce the bias current requirement. Frequency-dependent emitter degeneration is used to suppress a parasitic mode of oscillation that is not present in typical LC versions of this topology. The differential architecture minimizes common-mode noise introduced power supply and substrate coupling, and thus can further reduce the overall phase noise of the oscillator.

4.1.3 Experimental Results

Fig. 4.5(a) shows the measured transient waveforms from the oscillator at the RF carrier and the divided output. The measured oscillator phase noise at the 1.5 GHz carrier is -118 dBc/Hz @ 100 kHz offset. Although we typically achieve quality factors (Q) of over 2000
Figure 4.5: A) Measured signal at 1.5 GHz RF output (top) and 23.4 MHz divide-by-64 output (bottom); B) Measured temperature response as measured by the ZDR frequency response, an integrated diode thermometer, and a reference diode placed 10 mm away from the oscillator.

for temperature compensated resonators [108], our first devices had a fixable design error limiting the Q of these devices to about 500 to 800.

These chips integrate temperature compensated FBARs (zero drift resonators, or ZDR) which remove the linear temperature drift around a fixed turn-over temperature (TOT), yielding a residual 2nd order temperature dependence on the order of -15 ppb/°C². The ZDR in these devices have a TOT of approximately -120°C, which will be moved to 40°C in subsequent runs. To improve oscillator temperature stability to below 100ppm (over a 100°C temperature range), a feedback loop measuring temperature is necessary. The on-die temp sensing diode provides a nearly instantaneous read-out of the temperature that closely matches the FBAR temperature. Fig. 4.5(b) shows the temperature and frequency as measured by the integrated diode thermometer and a reference sensor placed 10 mm away from the oscillator. A 2W resistive heater on the board was activated for roughly 1 minute to simulate a PA or other heat source. Clearly, the integrated diode will greatly shorten the feedback cycle time and improve temperature compensation accuracy. To test the aging of the wafer-scale packaged resonator, the oscillator frequency shift was measured after 168 hours at 125°C and again after 336 hours at 125°C. Less than 1ppm of frequency drift was observed.

The on-die varactor is used to pull the resonant frequency of the oscillator to compensate
Figure 4.6: A) Measured tuning curve of FBAR oscillator; B) Phase noise of the Pierce oscillator and a differential oscillator measured with an Agilent 5052 signal source analyzer. The 1/f corner for the Pierce is 800 Hz, while the differential Colpitts is 3 kHz. The phase noise of the divide-by-64 output indicates a $20 \log(64)$ improvement in phase noise.

for temperature drift as well as any frequency variation due to processing uncertainty. Fig. 4.6(a) shows the measured pulling range of the oscillator using the integrated varactor ($\pm 500$ ppm).

We used bipolar devices in our first demonstration of this process for three reasons. First, this process has a higher $f_T$ than comparable legacy inexpensive CMOS processes. Secondly, the lower 1/f noise of BJTs provides a significant improvement (6-8dB) in close-in phase noise. Finally, the $g_m/I_d$ ratio is 2-3× higher than a MOSFET biased for high $f_T$. The properties of this bipolar process thus complement the CMOS circuitry of any
Figure 4.7: Wafer map showing functional oscillators (green) from one of the first chip-scale packaged oscillator wafers.

accompanying chip. Different applications of this technology will lead to different optimal partitioning. For example, a high performance PLL could use the FBAR/bipolar chip as a low noise VCO while the digital functionality were integrated onto a separate CMOS chip. Fig. 4.6(b) shows the measured phase noise profile of both oscillators.

One key benefit of the wafer-scale packaging, as opposed to 2-chip, approach to making FBAR oscillators is the ability to perform wafer-level testing on completed oscillators. Resonator frequency testing with a network analyzer is limited to 5-10 ppm accuracy due to limitations of the measurement and even to parasitic inductance introduced by small variations in probe placement. Oscillator testing reproducibility is <0.2 ppm in initial testing, making it suitable for evaluation of aging effects as well as die screening. Fig. 4.7 shows a map of one of the first wafers made, with the functional 64 (23.4 MHz) die shown in green.

4.2 A low jitter fully-differential “FMOS” oscillator

4.2.1 Circuit implementation

Fig. 4.8 shows the fully-differential oscillator and buffer circuits that are integrated into the lid. All bias circuitry is integrated. Cross coupled NPN transistors N1,2 generate a
negative resistance with an equivalent value of \(-2/g_m\). At resonance, this negative resistance cancels out the parallel resistance from the load tank circuit formed by the ZDR and the total capacitance at the oscillator output nodes. The output common-mode is first sensed through two resistors, level shifted through replica transistor N5, and then fed into the base of the follower transistors to form a negative-feedback stabilization loop. At low frequencies, the FBAR resonator is capacitive. The loop gain at low frequencies is therefore high, potentially allowing oscillation at undesired frequencies. To prevent parasitic modes of oscillation, we have inserted a capacitor C1 at the emitters of N1,2. At low frequencies, the capacitor acts as an open circuit. The loop gain is lowered as the transconductance of the cross-coupled pair is degenerated by N3,4 [110]. At high frequencies, the capacitor acts as a short circuit. The cross-coupled pair exhibit high transconductance as their emitters see a virtual ground. As a result, a high loop gain ensures proper oscillation. The capacitance value is chosen to minimize loop gain at low frequencies and maximize loop gain at the frequency of interest.

In another variant, varactors are added at the outputs to provide tuning capabilities.

The outputs of the oscillator are AC coupled to the buffer, designed to drive a 100Ohm
differential output load. The buffer uses a modified emitter-follower topology, where the signal drives the follower transistors $N_{8,9}$ as well as the current source transistors $N_{10,11}$. This current-reuse technique effectively doubles the transconductance and gain of the buffer for the same current consumption. AC-coupling through $C_{4,5}$ separates the DC bias voltages for $N_{8,9}$ and $N_{10,11}$. High current in the buffer is required to drive 1000Ωs; however, this reduces the base resistance at the input of the buffer. To avoid loading down the resonant tank, we insert a Darlington current buffer stage between the oscillator and the buffer. By scaling down the current in the $N_{6,7}$, we increase the base resistance and minimize de-Q-ing of the resonant tank.

4.2.2 Experimental results

The completed chip-scale oscillator is flip-chip bonded to a PCB. External components include a decoupling capacitor between power and ground, as well as DC-blocking capacitors at the two VCO outputs. For measurement purposes, a balun is also employed at the output for differential-to-single-ended conversion. The total bias current from the oscillator core and the buffer increases from 7.5 to 12.5 mA, while the output power increases from $-11$ to $-3.8$ dBm when the supply varies from 2.75 to 4.25 V (Fig. 4.9). The simulated oscillator current ranges from 1.2 to 2 mA as the supply increases from 2.75 to 4.25 V.
Relative to a conventional FBAR, the ZDR device removes the linear temperature drift around a fixed turn-over temperature (TOT), yielding a residual second order temperature dependence on the order of 15 to 30 ppm/°C². The ZDR in these devices have a TOT of approximately 10°C. Fig. 4.10(a) shows the frequency drift of the oscillator attached to a 2.6 GHz ZDR resonator when the temperature varies between -10 to 110°C. To improve oscillator temperature stability to below 100ppm (over a 100°C temperature range), we could include a feedback loop to measure temperature similar to the one described in [119]. Fig. 4.10(b) shows the frequency change as a function of supply voltage. Approximately 250 ppm of frequency variation is observed over a ±10% change in the supply voltage (at 3.3V). Frequency sensitivity over a 2:1 change in load VSWR is on the order of 15 ppm.

For the oscillator, we experimented with two kinds of ZDR resonators: one with a high $R_p$ (lower $Q$) acoustic stack and one with a lower $R_p$ (higher $Q$) acoustic stack. The latter stack in the oscillator gave 3 to 8 dB better measured phase noise (1kHz to 1 MHz) and the measured jitter was 25 fs. However, the output power was approximately 6 dB lower.

Different applications of this technology will lead to different optimal partitioning. For example, a high performance PLL could use the ZDR/bipolar chip as a low noise VCO with
the digital functionality integrated onto a separate CMOS chip. We used bipolar devices for three reasons: first, this process has a higher $f_T$ than comparable legacy inexpensive CMOS processes. Secondly, the lower 1/f noise of BJTs provides a significant improvement (6-8dB) in close-in phase noise. Finally, the $g_m/I_d$ ratio is 2-3x higher than a MOSFET biased in strong inversion for high $f_T$.

Fig. 4.11 shows the measured phase noise profile of the free-running oscillator when attached to three FBAR resonators of different frequencies, all with a high Rp stack (1.7, 2.6, and 3.4 GHz). We achieved better than -113 dBc/Hz phase noise at 1 MHz and an integrated jitter of 41 fs across all oscillators. As expected, the phase noise is degraded by approximately 6 dB when frequency doubles (from 1.7 GHz to 3.4 GHz). The on-chip varactor is used to pull the resonant frequency of the oscillator to compensate for temperature drift as well as any frequency variation due to processing uncertainty. An oscillator pulling range of 900 ppm is measured with the voltage on the integrated varactor varying from 0.5 to 1.8 V (Fig. 4.12). We measure less than 2000 ppm spread in the oscillator frequency across
Figure 4.12: Measured tuning curve of ZDR oscillator.

Table 4.1: Design and Performance Summary

<table>
<thead>
<tr>
<th>Integrated FBAR</th>
<th>This Work</th>
<th>IFCS'11 [119]</th>
<th>ISSCC'06 [115]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harmonic</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Frequency (GHz)</td>
<td>1.7</td>
<td>1.5</td>
<td>5.4</td>
</tr>
<tr>
<td>Ice ($\mu$A)</td>
<td>1580</td>
<td>3000</td>
<td>1700</td>
</tr>
<tr>
<td>PN(1 kHz) dBc/Hz</td>
<td>81.2</td>
<td>39</td>
<td>63</td>
</tr>
<tr>
<td>PN(10 kHz) dBc/Hz</td>
<td>-107.5</td>
<td>-104</td>
<td>-93</td>
</tr>
<tr>
<td>PN(100 kHz) dBc/Hz</td>
<td>130.4</td>
<td>124</td>
<td>118</td>
</tr>
<tr>
<td>PN(1 MHz) dBc/Hz</td>
<td>149.9</td>
<td>138</td>
<td>N/A</td>
</tr>
</tbody>
</table>

wafer.

Table 4.1 summarizes the performance of the 1.7 GHz oscillator implemented in this process and compares with related state-of-the-art. Since we are not targeting low-power applications, the current consumption of this work is higher than the prior work [119]. However, the far-off phase noise is better than reported previously.

We have presented a high-frequency, low-jitter frequency source using a ZDR-tuned oscillator implemented in a single-chip ZDR/bipolar process. A fully-differential oscillator allows large differential signal swings and achieves jitter as low as 25 fs. The ZDR resonator is fabricated at the wafer level with the circuitry, which provides an extremely small form
factor and (0.24 mm$^3$) minimal parasitics in a robust chip-scale package. Avago Technologies has tested hundreds of thousands parts and is currently in the process of commercializing this work.
Chapter 5

RF-POWERED TRANSMITTER FOR SHORT-RANGE WIRELESS

5.1 Background

Recently, there has been a growing demand for the integration of sensing and telemetry in diverse applications, including biosignal monitoring and smart buildings. For unobtrusive body-worn applications, both wireless data and power transmission are necessary because real-time biomedical data is otherwise inaccessible, and replacing batteries is undesirable.

Several power supply solutions have been proposed, including using a battery or an inductive link. Unfortunately, battery-powered sensors suffer from limited lifespan while inductive-coupling suffers from short (on the order of cm) wireless ranges. Others have used passive radio frequency identification (RFID) technologies to reduce the power consumption of the data transmission circuitry, thus extending the range of wireless power transfer [141][142]. However, the reader design is complex, partially because it must detect faint backscattered signals at the same frequency power is transmitted. Many commercial readers implement the Gen2 RFID protocol, which leads to a significant increase in tag complexity and transmit overhead. We introduce a passive tag architecture that decouples the power and data transmission, thus reducing the complexity of the reader.

Fig. 5.1 illustrates the proposed system architecture. The proposed tag architecture has the same wireless power transfer scheme as RFID systems, but actively transmits at 1/3 of the input frequency to avoid the “self-jamming” problem presented to conventional RFID readers. We use subharmonic injection-locking to avoid complex LO-generation circuitry while eliminating the need for quartz crystals. Though the downlink range (power delivery) is comparable to existing RFID systems, we demonstrate a far greater uplink range (sensor data transmission), thus improving receiver mobility.
Figure 5.1: High level system architecture. Passive tags are remotely powered and perform sensing. Data is transmitted at a 1/3 frequency carrier derived from the RF power through injection locking.

5.2 Literature Overview

Existing active transmitter topologies operating at low carrier frequencies (<100 MHz) typically consume less power and experience less tissue absorption than those operating at high frequencies [72]. However, in applications where small antenna sizes, high data rates (on the order of a few hundred kbps), and longer range are required, it is desirable to operate at higher frequencies (> 100 MHz). To avoid RF heating of the tissue, the FCC has mandated maximum permissible exposure (MPE) levels, which are especially stringent at high frequencies due to increased tissue absorption. It is therefore critical to reduce the power consumption of the tag, particularly in the active transmitter. Published implantable transceiver designs usually operate LC or ring oscillators in open loop [42][120]. LC oscillators consume relatively high power especially when low-Q on-chip inductors are employed. Ring oscillators consume less power, but are plagued with high phase noise and poor frequency stability.

In 59, the input RF signal powers the chip and injection-locks a frequency-doubling LC oscillator to generate an accurate system clock. The LC oscillator and choice of transmit frequency resulted in a high current consumption of 2 mA. Therefore, the system needs be
heavily duty-cycled, precluding its use in continuous streaming applications.

5.3 Proposed Architecture and Circuit Design

![Block Diagram of Proposed Transmitter Tag Architecture](image)

Figure 5.2: Block diagram of the proposed transmitter tag architecture.

Fig. 5.2 illustrates the proposed system block diagram. By transmitting in a different band than the RF power source, a conventional ASK narrowband receiver can be used instead of a complex RFID reader. Without having to comply with RFID protocols, the complexity of the proposed tag and the transmit overhead can also be reduced. The downlink (918 MHz) and uplink (306 MHz) frequencies are harmonically related and were chosen to fall into established frequency bands with sufficiently high radiated power permitted by FCC.

We designed a dual-band antenna to allow simultaneous uplink and downlink transmission. After impedance transformation by an L-match network, the incident RF signal splits into two paths. In one path, the input voltage is rectified and subsequently regulated to 1
V. The energy is stored in an off-chip 10 µF capacitor. In the other path, the input RF signal injection-locks a subharmonic ring oscillator, performing divide-by-3 functionality. The divider output is then AC-coupled to an open-drain driver. Binary data is OOK-modulated by power-cycling the driver. We chose OOK modulation here to conserve power during transmission of the “1” symbols. The driver’s load impedance is transformed through a tapped capacitor network to match to the 50 Ω antenna. Using this scheme, a stable low power high-frequency reference is obtained.

5.3.1 Injection-locked Frequency Divider (ILFD)

Injection-locked frequency dividers (ILFDs) usually consume less power than conventional flip-flop-based dividers. Compared to injection-locked LC-based dividers, ring-oscillator (RO) ILFDs offer smaller areas and larger locking range. The notoriously poor phase noise of ring oscillators is also improved dramatically in the injection locking process.

![Schematic of the injection-locked divide-by-3 circuit.](image)

Fig. 5.3 shows the schematic of the proposed 3-stage ILFD. Separation of the signal and bias paths through AC-coupling allows the input to simultaneously inject into $M_0$ and $M_1$, producing injection currents $i_{inj}$ at the $N_{0,1}$. The transistors $M_2$-$7$ function as sub-harmonic
mixers that translate the harmonics of \( i_{\text{inj}} \) and \( i_{\text{osc}} \) into the sum and difference frequency components \( m \cdot f_{\text{inj}} \pm n \cdot f_{\text{osc}} \) at the inverter outputs. When injecting a sufficiently strong signal within the locking range of the ILFD, the ring oscillator will lock to the frequency component closest to its \( f_o \) (1/3\( f_{\text{inj}} \) here).

Fig. 5.4 graphically depicts the injection-locking process. When the oscillator free-runs, the three identical stages equally distribute \( 180^\circ + n \cdot 360^\circ \) phase shift around the loop, and each stage has a phase shift of \( 240^\circ \). Injection of \( i_{\text{inj}} \cdot ( = K \cdot i_{\text{inj}}, K \) being the mixer conversion gain) shifts \( i_{\text{load}} \) by \( \phi \). The circuit is forced to oscillate at a different frequency such that each stage will contribute additional delay \( \theta \) to cancel \( \phi \) and the total phase shift around the loop remains \( 180^\circ + n \cdot 360^\circ \).

The locking range is directly proportional to \( \frac{|i_{\text{inj}}|}{|i_{\text{osc}}|} \). Injecting complementary signals into the top and bottom tails of multiple inverter stages doubles the injection strength, thus widening the locking range. In addition, the complementary injection drives both the rising and falling propagation delays, compared to the conventional tail-injection schemes where either rising or falling propagation delay are varied \([71]\). Consequently, the oscillation frequency and duty cycle can be better controlled.

In the proposed architecture, the power of the incident RF signal that injection-locks the ILFD and powers the chip is not well-controlled. Locking under high input power conditions is limited by the undesired non-linearity of the ILFD. To mitigate this locking deficiency at high input power, the tail transistors use low-\( V_T \) devices with small aspect ratios to improve
device linearity. This phenomenon could also be prevented by limiting the input power injected into the ILFD.

5.3.2 Power Management

The RF rectifier uses a 6-stage voltage-doubling charge-pump topology [141]. We chose zero-$V_T$ devices with low forward voltage drop and tolerable back-leakage to provide good sensitivity and efficiency. The output is clamped at 3 V with an off-chip Zener diode for over-voltage protection.

Bias currents for the chip are generated by a $V_{gs}/R$ reference that maintains a stable output current of 45 nA from 0.6 V to 3.6 V [141]. A bandgap reference provides a temperature-independent reference voltage that is stable to within 4 mV (of the nominal 1.2 V) across 0-100°C. A low-drop-out (LDO) linear regulator provides stable 1 V supply for the ILFD and driver.
Figure 5.6: a) The chip micrograph of the transmitter tag. b) Measured input return loss.

5.3.3 Antenna Design

The loop antenna is sized such that the circumference of the loop is $\sim \lambda/4$ at a frequency that lies between the input (918 MHz) and output frequency (306 MHz). The reactance looking into the loop antenna is inductive at 918 MHz and capacitive at 306 MHz. Since the input impedance at both frequencies is low, a dual-band matching network is used to provide relatively independent control of impedance matching at both frequencies. A drawing of the small loop antenna with dual band UHF matching network is shown in Fig. 5.5(a). $L_1$ and $C_2$ are used to match to 918 MHz, while $L_2$ and $C_1$ are used to match to 306 MHz. The fabricated antenna is populated with surface mount components that comprise the matching network. The return loss response of the assembled antenna was measured using an Agilent 8720 VNA as shown in Fig. 5.5(b).

5.4 Measurement results

The 918MHz/300MHz tag prototype was fabricated in a 0.13 µm CMOS process. The total current consumption of the chip varies from 19 to 23 µA as the input power increases from -10 to -6 dBm. The die photo is shown in Fig. 5.6(a). The total active area is 150 × 650 µm². Fig. 5.6(b) shows the measured input return loss $|s_{11}| < -28$ dB for downlink port.

Fig. 5.7(a) shows the measured rectifier efficiency as a function of input power. The
efficiency is 20-30\% for typical RF input power at output > 1 V. Fig. 5.7(b) shows the regulator output voltage as a function of input unregulated voltage. The minimum unregulated voltage required for regulation (dropout voltage) is 1.3 V.

Fig. 5.8(a) shows the overlaid spectrum of free-running and injection-locked ILFD. The output power when injection-locked is higher due to the increase in the output swing. The close-in phase noise of the ILFD inherits the phase noise of the RF power source when injection-locked. The measured input power range as a function of operating frequency for locking is shown in Fig. 5.8(b). The measured maximum locking range is 51\% (575 MHz to 969 MHz) when powered with a battery (dashed) and 39\% (673 to 995 MHz) when RF-powered (solid), ensuring injection-locking across realistic PVT variations. In the latter case, the range of input power for locking is limited on the low side by the minimum power required to supply to the circuitry and on the high side by the maximum power before the ILDF non-linearity prevent proper injection-locking.

OOK data modulation is verified by externally supplying a digital modulation signal. Fig. 5.9(a) shows the transmitter output spectrum corresponding to the driver power-cycled at 2 MHz (4 Mbps data rate). Fig. 5.9(b) shows the transient output waveform when OOK-modulated by a 2 MHz square-wave. The high on/off contrast ratio in the modulated output
relaxes the requirements on the OOK receiver. A start-up time of < 100 ns is achieved, allowing aggressive transmitter duty-cycling.

Fig. 5.10(a) shows the experimental setup that verifies the functionality of the wireless power and data transmission links. This experiment was performed in a laboratory environment. Complying with the FCC regulation, an Agilent E8254A signal generator transmitted a +35 dBm EIRP continuous wave at 918 MHz to the chip through a horn antenna. The input and output of chip are both connected to the custom loop antenna. An MSP430 microcontroller supplied the data stream to the chip. The loop antenna simultaneously receives power at 918 MHz from the horn antenna located 1.3 m away and transmits the 306 MHz OOK signal to a commercial Melexis TH7122 receiver. Fig. 5.10(b) shows the binary data (top) from the MSP430 and the data faithfully demodulated by the receiver located 6 meters away (bottom). The 20 kpbs data rate chosen here is limited by the maximum OOK data rate of the receiver.

Table 5.1 summarizes the performance of the proposed chip with a few recently published transmitters used in biomedical systems. The $P_{out}$ is measured at the distance reported. At -6 dBm input power, our transmitter consumes 23 μA (at 2.2 V) and achieves 10 pJ/bit, which is 42× more efficient than [59], while at a 9 dB lower output power.
Figure 5.9: a) Spectrum b) transient waveform of OOK modulation at 4 Mbps.

Figure 5.10: a) Wireless link test setup; b) Baseband OOK modulation signal and received/demodulated signal at 20 Mbps.
Table 5.1: Performance comparison of the proposed transmitter and existing implantable transmitters

<table>
<thead>
<tr>
<th></th>
<th>[120]</th>
<th>[42]</th>
<th>[72]</th>
<th>[59]</th>
<th>This Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Source</td>
<td>Ind.</td>
<td>Ind.</td>
<td>RF</td>
<td>RF</td>
<td>RF</td>
</tr>
<tr>
<td>Topology</td>
<td>FRRO</td>
<td>FRLC</td>
<td>FRLC</td>
<td>ILLC</td>
<td>ILRO</td>
</tr>
<tr>
<td>Modulation</td>
<td>OOK</td>
<td>FSK</td>
<td>FSK</td>
<td>BPSK</td>
<td>OOK</td>
</tr>
<tr>
<td>Uplink (MHz)</td>
<td>4 / 8</td>
<td>264</td>
<td>0.125</td>
<td>450</td>
<td>918</td>
</tr>
<tr>
<td>Downlink (MHz)</td>
<td>70-200</td>
<td>433</td>
<td>27.3</td>
<td>900</td>
<td>306</td>
</tr>
<tr>
<td>Current/Power</td>
<td>–</td>
<td>6.75 mW</td>
<td>1 mW</td>
<td>2 mA</td>
<td>.023 mA</td>
</tr>
<tr>
<td>Datarate (kbps)</td>
<td>2000</td>
<td>330</td>
<td>80</td>
<td>7000</td>
<td>4000</td>
</tr>
<tr>
<td>Min. Supply (V)</td>
<td>1.8</td>
<td>3.55</td>
<td>–</td>
<td>1.5</td>
<td>1.3</td>
</tr>
<tr>
<td>$P_{out}$ (dBm/cm)</td>
<td>–</td>
<td>-85.13</td>
<td>-70.10</td>
<td>-24.0</td>
<td>-33.0</td>
</tr>
<tr>
<td>Process (um)</td>
<td>1.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.25</td>
<td>0.13</td>
</tr>
</tbody>
</table>

5.5 Conclusion

An ultra-low power tag architecture for battery-free continuous streaming biomedical applications has been presented. The incident RF signal simultaneously powers the chip and injection-locks a 3× frequency divider to regenerate an LO frequency for data transmission. The injection-locking process provides a stable LO frequency, high data rate, and >6 meters of data uplink transmission range while consuming 23 μA, the lowest current consumption compared to similar transmitters reported to date.
Chapter 6

RECEIVERS FOR SHORT-RANGE WIRELESS

Low power, short-range ISM band transceivers have become extremely important in enabling new applications and have become highly successful commercially. New energy harvesting methodologies will open further applications but demand lower power and supply voltages than currently available chipsets can provide. For example, on-chip solar cells allow only a 200 mV to 900 mV supply; thermoelectric generators can be used for on-body generation, but exhibit low supply voltages (50-300 mV) [12].

Although boost converters can be employed, their limited efficiency (40% – 75%) often incurs a significant power penalty [49]. For instance, in a recent work, we demonstrated the feasibility of powering a wireless biosignal acquisition chip solely from a thermoelectric harvester and/or RF power [148]. The chip contains integrated supply regulation, analog front-end, DSP and transmitter (TX). A closed-loop power management monitors the harvested energy and duty-cycles the TX accordingly. Because an on-chip boost converter boosts the TEG voltage from 30 mV to 1.35 V at 38% efficiency, the system requires at least 50µW power to be harvested from the TEG although the active circuits only draw 19µW.

The goal of this work is a 2.4 GHz receiver that operates with a supply voltage of 300 mV allowing direct powering from various energy harvesting sources. Our target application is body-worn wireless devices requiring high datarates (wirelessly tethered hearing aids, for example). This application allows access to multiple energy harvesting power sources, like thermoelectric power harvested from body heat and photovoltaic.

Operating from a 300 mV supply voltage is challenging for a few reasons. First, the transit frequency $f_T$ is reduced, limiting the maximum gain and minimum noise figure achievable in LNAs and mixers. It also results in higher power consumption in VCOs and high-frequency dividers. Secondly, the limited voltage headroom also precludes stacking
of transistors, severely limiting the gain and bandwidth in analog/RF amplifiers. It could also degrade reverse isolation in LNAs and feed-through in mixers. Furthermore, limited voltage swing degrades VCO output power and phase noise, and in turn also reduces mixer conversion gain. Lastly, the ultra-low supply voltage limits the speed at which the digital logic can operate, restricting the baseband processing capabilities.

In this chapter, we present the design of a 2.4 GHz low-IF receiver consuming 1.6 mW from a single 300 mV supply voltage. It achieves -91.5 dBm sensitivity for a BER of 10^{-3} when tested with 2^23 bits of pseudo-random number sequence (PRNS). Section II presents a thorough device-level feasibility study. In Section III, we discuss the receiver architecture and the transformer-coupled technique pivotal for 300 mV RF operation. We follow by detailed circuit description of the RF front-end and the baseband in section IV and V. We present the measurement results on the receiver IC in section VI. In section VII, we offer discussion on the receiver performance at the presence of supply variation, bias voltage variation and choice for the mixer architecture.

6.1 Concept validation

To verify the feasibility of a 300 mV radio design, we need to study the DC and RF characteristics of the active and passive devices. We chose the TSMC CMN65GP 1p9m process option for this feasibility study. This process offers a compelling tradeoff between $f_T$ and flicker noise, both increase at smaller technology nodes. In addition, a thick metal layer with low sheet resistance is also available for high Q, low-loss passive components.

6.1.1 Transistor sizes and bias conditions

There are multiple $V_{TH}$ options in this process, including natural, low, nominal and high-$V_{TH}$. Because of the ultra-low supply voltage constraint, we focused on the performance of low-$V_{TH}$ (LVT) nFET and pFET transistors. The $V_{TH}$ of a lvtfet device is around 320 mV. By forward-biasing the bulk-source junction voltage of the nFET by 300mV, we lower its $V_{TH}$ by 35 mV (10%) (Fig. 6.1).

We first investigated how to choose an optimal length. We studied the effect of length have on $V_{gs}$, $V_{TH}$, $g_m$, $I_d$ and $f_T$ using a LVT nFET with fixed width (80 μm) and with
Figure 6.1: $V_{TH}$ as a function of the bulk source forward biasing voltage $V_{BB}$.

A constant $W/L$ ratio ($W/L = 133$). Fig. 6.2(a) shows that $V_{TH}$ decreases as the length increases due to the "halo" implant layer in the deep sub-micron process to reduce leakage current. We should therefore avoid using minimum length for lower $V_{TH}$. $V_{TH}$ of the device with a fixed width of 80 $\mu$m stays relatively constant around 300 mV for 1 mA of current, while that of a device with constant $W/L$ ratio scales with $V_{TH}$. Therefore, it is feasible to bias a transistor of realistic size and current slightly above $V_{TH}$.

For narrow band applications, we're interested in the ratio of $f_T$ to the operating frequency $f_o$. As previously discussed, a high $f_T/f_o$ for optimal analog/RF performance. Fig. 6.2(b) indicates that a smaller length results in a larger $f_T$. For low-power operations, we would like to maximize $g_{m}$ given a power budget. Current efficiency $g_{m}/I_d$ is a common metric to indicate the power saving capabilities. A high $g_{m}/I_d$ is often achieved at the expense of reduced $f_T/f_o$. A good figure-of-merit (FOM) to trade-off these two design metrics is their product $(g_{m}/I_d \cdot f_T/f_o)$. Fig. 6.2(c) shows that smaller length achieves better FOM. We were also interested in the intrinsic gain $(g_{m}/g_{ds})$ of the device, which dictates voltage gain of circuits like LNAs or IF amplifiers. Fig. 6.2(d) shows that a length between 100 $\mu$m and 250 $\mu$m achieves maximal gain.

We also conducted similar study on LVT pFET devices and compared to their nFET counterparts. They require larger width (200 $\mu$m) to achieve similar bias conditions (Fig. 6.3(a),...
Figure 6.2: Design metrics as a function of a low-$V_T$ nFET transistor length for 1 mA and 0.5 mA $I_d$: a) $V_{th}$ and $V_{gs}$; b) normalized transit frequency $f_t / f_{0.5}$; c) $g_{m} / L_{d} \cdot f_t / f_{0.5}$; d) intrinsic gain $g_{m} / g_{ds}$. 
resulting in larger parasitic capacitance. As a result, they have lower \( f_T/f_\alpha \) than their nFET counterparts (Fig. 6.3(b)). Fig. 6.3(c) shows that a smaller length also results in better FOM, consistent with the LVT nFETs. However, the magnitude of the FOM is 2-3 times lower. Fig. 6.3(d) shows that the maximum \( g_m/g_{ds} \) is half that of a LVT nFET device biased similarly, and can be achieved with a length between 200 nm and 300 nm. Similar to LVT nFETs, a 300 mV \( V_{th} \) voltage results in \( \approx 10\% \) decrease in the threshold voltage. However, the effects of forward bulk-source biasing on \( f_L \), FOM and \( g_m/g_{ds} \) are negligible because \( g_m \approx h_{m0} g_{ds} \), are not affected significantly. As a result, we should try to use LVT nFET in designs because they have better DC/RF performance than LVT pFET biased similarly.
6.1.2 Inversion Coefficient

After choosing the size of the transistor, we should find the optimal region of operation to bias the transistor. We used the EKV model because it is valid in all regions of inversion [34]. At the onset of inversion \( V_{GS} = V_{TH} \), the current can be expressed as:

\[
I_S = 2 \cdot n \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot \phi_t^2
\]  

(6.1)

Here \( n \) represents the subthreshold slope factor, \( \mu \) the mobility, \( C_{ox} \) the oxide capacitance and \( \phi_t = kT/q \) the thermal voltage. The current in the vicinity of \( V_{TH} \) can be modeled as:

\[
I_{DS} = \frac{I_S \cdot IC}{k_{fd}}
\]  

(6.2)

Here, \( IC \) represents the inversion coefficient, \( k_{fd} \) is a model-fitting parameter. The inversion coefficient \( (IC) \) represents the degree of inversion of the transistor, and can be expressed as:

\[
IC = (\ln(\frac{e^{(1+\sigma)V_{gs}-V_{TH}}}{2\sigma V_{gs}} + 1))^2
\]  

(6.3)

where \( \sigma \) represents the drain-induced barrier lowering (DIBL) factor. Generally, \( IC = 1 \) occurs when \( V_{gs} = V_{TH} \). To find the relationship between \( V_{gs} \) and \( IC \), we need to solve for the technology-dependent parameters \( \text{"n"} \) and \( \text{"\sigma"} \). We can use two data points in the leakage current equation (at \( V_{gs} = 0 \)) [75]:

\[
I_{Leakage} = I_S \cdot e^{\frac{\sigma V_{gs} - V_{TH}}{n \phi_t}}
\]  

(6.4)

After finding \( \text{"n"} \) and \( \text{"\sigma"} \) from Eqn. 6.4, we can find IC corresponding to \( V_{gs} \) using Eqn. 6.3 and thus develop simple yet relatively accurate models. The remaining discussion will focus on design optimization using inversion coefficient.

As illustrated in Fig. 6.2(a), \( V_{gs} \) is slightly above \( V_{TH} \) for a small length, allowing us to bias the nFETs in weak or moderate inversion. Strong inversion is not realistic given our ultra-low voltage operation. We chose LVE nFET (Nchv) and natural nFET (Nchna)
Figure 6.1: Design metrics as a function of the inversion coefficient for a low-$V_{TH}$ nFET transistor: a) $V_{TH}$ and $V_{gs}$; b) normalized transit frequency $f_t/f_c$; c) $g_m/I_D \cdot f_t/f_c$; d) intrinsic gain $h_{uu}/g_{ds}$. 
with similar aspect ratio (80/0.17 \( \mu m/\mu m \) and 80/0.2 \( \mu m/\mu m \) respectively) and simulated the key design metrics across the inversion coefficient from weak to moderate inversion. Nemo has lower \( V_{TH} \) than Nemoht, however, a larger minimum length (200 nm). We sized the two nFETs to have similar bias conditions (similar current and voltage), resulting in similar \( g_m/I_d \) ratio (Fig. 6.4(a)). As shown in Fig. 6.4(b), the stronger inversion, the higher the \( f_T \). In addition, decreasing the voltage headroom \( V_{ds} \) also degrades \( f_T \), re-iterating the challenge for RF operations under ultra-low supply voltage. Low-power operation motivates biasing transistors in weak inversion to achieve high current efficiency (\( g_m/I_d \)). However, the transistors should be biased in stronger inversion to achieve higher \( f_T \), critical for RF performance. To achieve an optimal trade-off of these two design metrics, we should bias the transistors in moderate inversion (Fig. 6.4(c)). Fig. 6.4(d) also illustrate the need to bias transistors in moderate inversion to achieve maximal intrinsic gain.

### 6.2 Architecture

Among the state-of-the-art ultra-low voltage receivers, Cook et al. achieves the lowest supply voltage of 0.4 V using mixer-first architecture [24]. Although it only consumes 330 \( \mu W \) and achieves high linearity, there is no LNA and the gain is not characterized. They relied on external high-Q matching network to achieve passive gain and attenuate the input-referred noise from the mixer and baseband. Among the receivers that include LNAs, Stanic et al. achieves the lowest supply voltage of 0.5 V using a sliding-IF architectures [123]. Although DC offsets are reduced, this architecture requires image-rejection filters inserted before each down-conversion stage. Other works that employ low/zero-IF receiver architecture all use voltages higher than 0.6 V and power consumption higher than 8 mW for 2.4 GHz operations [11][15]. To operate under 0.3 V, we need to employ the circuit design techniques discussed previously as well as a new architecture.

#### 6.2.1 Proposed system architecture

We chose low-IF direct down-conversion architecture suitable for low-power operation. Although low-IF architecture does not have the problems exist in zero-IF (e.g. DC offsets, even-order distortion, signal quality degradation by flicker noise, etc.), there is requirement
Figure 6.5: System block diagram of the proposed low-IF receiver using transformer-coupled techniques.

...for image rejection. The simple single phase down conversion architecture can be modified to include a QVCO and I/Q down-conversion mixers to reject the image.

Fig. 6.5 shows the proposed architecture. A single-ended 2.4 GHz RF input is amplified and converted to a differential signal through an intrinsic bahn. A quasi passive mixer then down-converts the differential RF signal to a low-IF of 1-10 MHz. A low IF frequency is desired to minimize the required bandwidth in the IF amplifiers and save power. At the same time, the lower bound of IF is set above the flicker noise corner of the mixer and baseband circuitry. The LO is also transformer coupled to the mixer. As discussed in the next section, the secondary coil is conveniently separated into two sections. One section is used for the inductive feedback between the drain and source of the cross-coupled pair in the VCO to increase swing and improve the phase noise, while another section is used to couple the LO to the mixer. A chain of IF amplifiers and narrowband filters are intercaved to perform programmable channel selection. The overall linearity of the baseband is limited by that of the IF amplifiers. It is therefore desirable to place the filter before the amplifiers, necessitating the filter noise to be very low and thus increasing the...
power consumption. Interleaving filter sections and amplifiers optimizes the noise-linearity trade-off for low-power operations.

6.2.2 On-chip transformers

The performance of on-chip transformers is critical to our topology. The TSMC p-cell libraries offer a variety of passives, including inductors, capacitors, varactors and diodes. However, it does not have transformers, useful for magnetically coupling signals in RF applications. Therefore, we need to custom-make transformers. The key parameters to design for are self-inductance (L), quality factor (Q), the coupling coefficient (k) and self-resonance (f_{SR}) given area constraints. For a certain L (determined by the operating frequency and total capacitance), we want to maximize Q and k while ensuring sufficiently high f_{SR}. In RFIC designs, a high Q leads to a high R_p, and is crucial for low power and low noise operations in amplifiers and oscillators. Maximizing k results in higher effective inductance per area and reduces signal loss when magnetically coupled from the primary to the secondary coil. In addition, a high k (i.e. > 0.7) is required to faithfully reflect the impedance of the secondary coil at the primary, important for RF input matching. Although f_{SR} ought to be sufficiently high to ensure the desired behaviors of the transformers, one that is much higher than the operating frequency f_o is not necessary.

There are two main transformer topologies: planar and stacked. Planar structure where both the primary and secondary coils are drawn on the same metal layer, has medium Q, k, but high f_{SR} and area. Conversely, stacked structure consisted of two metal layers, exhibits high k, L, but low Q and f_{SR}. There are many variations of these two main structures. One popular one is interleaved planar structure, where each metal trace is split into multiple fingers to increase the coupling coefficient. The various transformer topologies
Table 6.2: The effect of increasing transformer geometries have on key design metrics

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>Q at $f_o$</th>
<th>$f_{SR}$</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turns (3-8)</td>
<td>↑</td>
<td>↓</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Radius (40-90 μm)</td>
<td>↑</td>
<td>↑</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Width (5-20 μm)</td>
<td>↓</td>
<td>↑</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Spacing (2-4 μm)</td>
<td>↓</td>
<td>↑</td>
<td>↑</td>
<td>↓</td>
</tr>
</tbody>
</table>

are summarized in Table 6.1. The choice of the transformer structure depends on the application. Because we’re interested in using transformers to magnetically couple signals or as RF loads, our objective is to achieve high Q and k, and sufficient $f_{SR}$ for a given L. In our TSMC CMN65GP option, there is only one thick metal layer (M9) with low sheet resistance (0.005 Ω/sq). The next thickest metals (M8) has 4× higher sheet resistance, resulting in 4× lower Q. For our application where maximizing Q is crucial, we decided to employ a planar structure.

The geometry of the transformer determines its L, Q, k, and $f_{SR}$. The most efficient shape is a circle because as it has the shortest perimeter, thus least resistive loss given area. However, it is difficult to make a circle without violating DRC rules. Therefore, we resort to an octagon, more efficient than a square and relatively easy to implement. We would like to study the effect of the number of turns, the radius, width of the coils and turn-to-turn spacing have on L, Q, k, and $f_{SR}$. As shown in Table 6.2, we can maximize L by using a large number of turns, large radius and narrow spacing. At low frequencies, the width of the coils have negligible effect on L. To achieve a higher Q at the expense of reduced k, we should decrease the number of turns while increase the radius, width of the coil and turn-to-turn spacing. For a higher $f_{SR}$, we need to decrease the number of turns, radius and width while increasing the spacing. The final transformer design should be optimized leveraging the trade-offs provided by different geometries.

Conventional winding approaches, including parallel winding, interwound winding, overlay winding and concentric spiral winding usually are not symmetric [48]. We laid out a balun using a compact symmetrical bifilar transformer with inner common-ground tap (Fig. 6.6(a)). Here the turn ratio is 2:2, although an arbitrary winding ratio can be achieved
by modifying the coil lengths. Care is taken to ensure geometric symmetry in the four quads of the transformer layouts. The main coils are drawn in M9, and the crossovers are drawn in M8. The center tap is taken out on M7. The balun can also be configured as a transformer by floating the center tap (P4). Considering the k-Q trade-off provided by different geometries, we have chosen a 7-turn design with 5 \( \mu \text{m} \) ring width, 2 \( \mu \text{m} \) turn-to-turn spacing and 40 \( \mu \text{m} \) guard-ring spacing. The other dimensions are as labeled in Fig. 6.6(a). At 2.4 GHz, we simulated a L of 5.96 nH, a Q of 9.2 and a k of 0.9.

We imported the transformer layout into ADS Momentum and performed EM simulations. The n-port s-parameters from Momentum output can then be used in Spectre simulations. We first built a substrate model based on technology metal stack information. Then we used an inductor to calibrate the substrate model file by comparing the S-parameters of the simulated ADS output with that of the PDK. Key parameters such as Q, k, L can then be derived from Z-parameter simulations in Spectre:

\[
L = \frac{\text{imag}(Z_{11})}{2\pi f}
\]  
(6.5)

\[
Q = \frac{\text{imag}(Z_{11})}{\text{real}(Z_{11})}
\]  
(6.6)

\[
k = \frac{|M|}{\sqrt{L_p L_s}} = \frac{|\text{imag}(Z_{12})|}{\sqrt{\text{imag}(Z_{11}) \text{imag}(Z_{22})}}
\]  
(6.7)

To gain an intuitive understanding of the transformer and to facilitate transient-based simulations, we can build a circuit model for the custom transformer using the Z-parameter information. We used a relatively simple but sufficiently accurate model as shown in Fig. 6.6(b). This model accounts for the coupling among different inductor sections, resistive loss, parasitic capacitance and substrate loss. The effect of inter-turn fringing capacitance is usually small because the adjacent turns are almost equipotential and therefore is neglected in our model [144]. We can follow the procedure below for an approximate RLC transformer model:

1. Set up a testbench with “nport” device from “analogLib” linking to the “snp” file from
Figure 6.6: \( a) \) The layout of a symmetrical l anywhere planar transformer; \( b) \) A simplified transformer circuit model.
2. Obtain the Z-parameter from a 3-port "σp" simulation.

3. \( L_{1,2,3,4} \) are obtained from Eqn. 6.5 at low frequencies (i.e. 1 MHz).

4. \( R_{s1,s2,s3,s4} \) are obtained from \( 2\pi f \cdot L_{1,2,3,4} / Q_{1,2,3,4} \) at low frequencies.

5. \( k_{13,24,12,34} \) are obtained from Eqn. 6.7; curve-fitting to find \( k_{14,23} \) (you can start with \( k_{14,23} = k_{13} \).

6. From Eqn. 6.8, solve for \( C_p \).

\[
0 = 1 - \frac{R_s^2 C_p}{L_s} - (2\pi f_{SR})^2 L_s C_p
\]

(6.8)

7. Curve fitting to find the substrate loss resistance \( R_{sub} \) (you can start with a reasonable value, such as 3 kΩ).

6.3 RF front-end

Fig. 6.7 shows the schematic of the proposed RF front-end. Our RF front-end uses the following techniques to allow operation from a 300mV supply. First, we extensively utilize transformer coupling between stages to reduce headroom requirements. This strategy also allows inter-stage impedance transformation and single-ended to differential conversion. The entire front-end comprises single-stack nFETs with \( V_{ds} = V_{dd} \), maximizing utilization of the 300 mV supply. Secondly, we use forward-biased bulk-source junctions to lower threshold voltages where appropriate. Since our supply voltage is only 300 mV, this does not incur noticeable leakage currents. Thirdly, biasing key transistors in the RF signal path in moderate inversion optimize the trade-off between power consumption and \( f_r \).
Figure 6.7: Schematic of the RF front-end (folded LNA, transformer-feedback LC VCO, and quasi-passive mixer). Also shown is the microphotograph of an on-chip transformer, used throughout the front-end.
6.3.1 Folded LNA

As the first stage in the receiver chain, LNA needs to achieve low noise, high gain, good input matching, high reverse-isolation and moderate linearity. Common-source LNA (CSLNA) usually provides better noise performance using higher power consumption than the conventional common-gate LNA (CGLNA) when $f_T/f_o$ ($f_T$ normalized to the operating frequency) is relatively high. However, when $f_T/f_o$ decreases, its noise performance degrades linearly. On the other hand, the noise performance of CGLNA is independent of $f_T/f_o$. The gate noise of CSLNA is enhanced by its input $Q$, while that of CGLNA is insignificant even at low $f_T/f_o$. In addition, CSLNA usually requires off-chip matching components because its high-$Q$ input match makes the circuit more sensitive to process, voltage and temperature (PVT) variations. Lastly, CSLNA also exhibits inferior reverse isolation and stability due to the Miller effect from the feed-forward parasitic capacitor $C_{gd}$ [149]. Since our application is $f_T$-limited and gate noise is more pronounced in a deep sub-micron process, the RF performance advantage of CSLNA over CGLNA is minimal, if any. Therefore, we chose to employ CGLNA topology.

A cascode device is necessary to shield the input impedance from the effect of the load and enhance reverse isolation of the LNA. Since our application is narrow-band where capacitors are tuned out with inductors, we’re interested in the real part of the input impedance. This can be expressed as

$$R_{src} = \frac{R_{load} + r_o}{1 + g_m r_o} \tag{6.9}$$

Where $R_{load}$ represents the load resistance at the drain of the input device. Without a cascode device, the $R_p$ of the tuned output LC tank directly appears as $R_{load}$. In a deep sub-micron process, $r_o$ is limited even when the transistor is in saturation due to pronounced channel-length modulation and small $V_{ds}$. Typically, $r_o$ is on the order of 1 kΩ or less for > 1 mA bias current, while the $R_p$ of an on-chip 6 nH inductor with a Q of 10 is roughly 1 kΩs at 2.4 GHz. The input resistance increases by more than 100% as a result of $R_{load}$. However, when a cascode device is present, $R_{load}$ is the resistance looking into the source of the cascode device. Applying Eqn. 6.9 on the cascode device, the resistance is approximately
only 100 Ω, much less than an $R_p$ of 1 k Ω. Consequently, the input resistance and thus the $|S_{11}|$ is less affected by the load. However the necessity for a cascode device conflicts with the requirement that no transistor-stacking is allowed with a 300 mV supply voltage to achieve good RF performance. To reconcile these two requirements, a folded-cascode structure can be used where the signal is directed to the pFET cascode device (Fig. 6.8(a)). The current source can be realized by a tuned inductor to eliminate voltage overhead [70].

In this 65nm RF process, nFET transistors have better $g_m$ and $f_T$ compared to a larger pFET transistor with similar bias condition. We propose a transformer-coupled (XFMRC) topology to realize a nFET-only design and an intrinsic balun (Fig. 6.8(b)). Compared to their pFET counterpart, the nFET presents a lower input impedance at the cascode source, better shielding the amplifier input from the tuned load. In addition, a lower resistance also reduces the voltage level at the cascode input, improving the linearity. The reduced parasitic capacitance at the cascode input lowers the noise contribution from the cascode device at high frequencies. Coupled transformers also result in a higher effective inductance ($1+k^2$, $k \approx 0.8$ here) for a minimal increase in area. The ratio of the primary and secondary (N) can be $>1$ to achieve passive current gain and hence increase the overall voltage gain,
at the expense of degraded reverse isolation. In simulation, we can observe a 1.5 dB higher gain, 0.9 dB lower noise figure and 2 dB higher IIP3 for XFMRC-CGLNA compared to a conventional folded LNA. Lastly, XFMRC topology also provides free signal inversion, realizing an intrinsic balun. Compare with the on-chip baluns using CS and CG-LNA, XFMRC achieves superior matching between the differential signals.

Free signal inversion can be effectively used where in-phase and out-of-phase signals are simultaneous required. $G_m$-boosting is one example. It is typically used at the amplifier input to increase $g_m$ and reduce noise for a fixed bias current [149]. Here, since the transformer couples signal from the drain of the input transistor to the source of the cascode transistor, $g_m$-boosting technique can be applied to the cascode transistors. Leveraging on the intrinsic balun, the in-phase signal is capacitive-coupled to the gate of the cascode, while the out-of-phase signal to the source. Ideally, when the in-phase and out-of-phase signals have the same amplitude, the $g_m$ of the cascode is boosted by a factor of 2. For a given source impedance, the bias current of the cascode branch is reduced by a factor of 2. As a result, the same reverse isolation is achieved with lower power consumption.

By forward-biasing the bulk-source junction voltage of the three nFETs by 300 mV, we lower their threshold voltage by 10%. This allows us to bias the transistors in moderate inversion, increasing their $g_m/I_d \cdot f_T/f_o$ FOM by 100 to 200% compared to a weak-inversion operation. As we recall from the device characterization, a small length is desired to achieve high $f_T$ and $g_m/I_d \cdot f_T/f_o$ FOM. However, a length smaller than 100 nm should be avoided for its high $V_{TH}$ and low intrinsic gain $g_m/g_{ds}$. Weighing the trade-offs, we chose an optimal 100 nm for the length of the nFET transistors. We sized the input transistor 100 $\mu$m/100 nm to source sufficient current to provide a minimal NF and better than -10 dB [S11]. Thanks to the current saving in the cascode branches resulted from $g_m$-boosting, we can reduce the size of the cascode transistors to 50 $\mu$m/100 nm. Smaller cascode devices reduce parasitic capacitance and thus their noise contribution at high frequencies.
6.3.2 Transformer-feedback VCO

VCO usually dissipates a significant amount of power in order to achieve sufficiently high output swing and low phase noise. Ultra-low supply voltage conditions limit output signal swing and indirectly limits the phase noise performance. Existing solutions that use external inductors or MEMS resonators demonstrate low-power and lower phase noise [98][101][110][146]. However, they are not fully compatible with standard CMOS process.

Conventional low-voltage VCOs employ cross-coupled technique to realize negative-$g_{m}$ (Fig. 6.9(a)). Although there are no stacked transistors, the minimum supply voltage is limited by the voltage swing requirement, as load inductors only allow voltage swing above the top supply rail. On the other hand, a transformer-coupled VCO (TF-VCO) (Fig. 6.9(b)) enables the output signals to swing above the supply voltage and below the ground potential, thus increasing the output power and lowering the phase noise. Two mechanisms further reduce the phase noise. First, coupled transformers exhibit larger effective inductance per area and better quality factor compared to a single inductor. Secondly, transformer-feedback across the drain and source nodes enhances the cyclo-stationary noise property, similar to the Colpitts/Hartley design [62]. Transformer-feedback shows another advantage over capacitive-feedback by allowing a higher L/C ratio to achieve higher $R_{p}$ and lower power consumption. We simulated the output signal swing and phase noise of the standard cross-coupled VCO and the TF-VCO using low-$V_{TH}$ nFET transistors and simple transformer models (include $L$, $R_{s}$ and $k$). For the same output swing (1.3V differential) and current consumption (790 µA), TF-VCO requires a 200 mV lower voltage supply (300 mV instead of 500 mV) and achieves 3 dB better phase noise compared to standard cross-coupled VCO.

Fig. 6.9(b)b) shows a detailed schematic of the VCO. When the voltage at the source of the transistors swings below ground, the bulk-source pn-junction is forward-biased. To minimize the forward-bias leakage current into the substrate, the bulk of the cross-coupled transistors are tied to ground. We need to carefully choose the ratio of the primary and secondary coils of the transformer at the transistor drain and source ($L_{p}$ and $L_{s}$ respectively) to optimize the trade-off among output swing, power consumption and phase noise. For a fixed supply voltage and $L_{p}$, varying the ratio from 4:1 to 2:1 results in an increase in the
voltage swing and better far-off phase noise at the expense of higher current. We chose 2:1 ratio for a larger swing.

To cover process variations, we added varactors between the drain and source terminals of the cross-coupled pair ($C_p$ and $C_s$ respectively) for frequency tuning. The ratio of $C_p$ to $C_s$ also affect the VCO performance. As we increase the ratio from 1:1 to 1:4, the tuning range increases while the circuit consumes more current. There is a small improvement in the output swing and phase noise. We chose 1:1.8 to conserve current while achieving ≈ 8% tuning range over 300 mV change in the varactor voltage. We added a transistor at the drain of the cross-coupled pair to injection-lock the VCO to an external source for testing purposes.

The VCO design procedure can be described as follows: First, we pick a load inductor with high Q capable of resonating out the parasitic capacitance and varactors at 2.4 GHz. Then we choose the transformer winding ratio $L_p/L_s$ to 2:1. Given the $R_p$ of the LC tank and $V_{gs}$ of 300 mV, we can size $W/L$ of the cross-coupled pair to achieve sufficient loop gain. Lastly, we add varactors at the drain and source of the cross-coupled pair for frequency
tuning. The transformer ratio, varactor ratio and transistor size are to be co-designed to achieve favorable trade-off among output swing, phase noise and power consumption.

When characterizing the receiver, the LO needs to be a known fixed frequency to properly down-convert the RF signal to IF. A frequency synthesizer is usually used to generate the LO. For simplicity, we chose to injection-lock the VCO to an external RF source. As shown in Fig. 6.9(b), the injection transistor is placed in parallel with the load LC tank. We want to maximize the transistor size for a large locking range. However, we need to also minimize the parasitic capacitance added to the tank, as a larger on-chip L/C ratio is desired for higher \( R_p \) and better Q. To trade-off these two dependence, we sized the transistor to 16\( \mu \)m/100nm, large enough to ensure >10\% locking range, small enough to minimize parasitic capacitance added to the tank. We can either injection-lock at the fundamental or its harmonics. The locking range is the largest when injecting at the fundamental, then the second harmonic, third, etc. Because we're targeting a low-IF architecture, where the LO is within a few MHz away from the desired RF signal, we chose to inject at the 2\(^{nd}\) harmonic to mitigate desensitizing the desired RF signal at the LNA input, at the expense of reduced locking range.

6.3.3 Quasi-passive mixer

As the last stage of the front-end path, the linearity of mixer dominates that of the receiver chain. It is also desired for the mixer to have gain to minimize noise contribution from the baseband. However, given the same input signal level and supply, higher gain usually results in poorer linearity. The trade-offs among gain, noise and linearity make RF down-conversion mixer the most challenging block in the ultra-low voltage receiver, where voltage headroom is scarce.

Active Gilbert-cell mixers are commonly used for its high gain and moderate linearity (Fig. 6.10(a)). However, its use in an ultra-low voltage application is limited for a few reasons. First, it requires two or more transistors stacked, resulting in a high supply voltage. Secondly, the LO drive needs to be large (> \( V_{gs} + V_{ds, sat} \)) to fully switch the transistor current from one polarity to another. This “hard switching” via a large LO signal
Figure 6.10: a) Active Gilbert-cell mixer; b) active switched-$g_{m}$ mixer; c) passive mixer; d) proposed quasi-passive mixer.
mathematically corresponds to multiplication by a square wave. Compared to a sinusoidal drive, this achieves higher conversion gain \(2/\pi\) instead of \(1/2\) and lower noise figure \([58]\). It either requires the VCO to have large output swing, or requires an additional buffer stage to increase swing. Either approach faces challenges to achieve high swing under a 300 mV supply. Lastly, operating the switch transistors in saturation requires significant headroom, reducing the headroom available for the load, resulting in lower conversion gain.

Alternatively, the LO and RF ports can be swapped to operate the differential pair transistors as switched-\(g_m\) cells to commutate the signal current to the output. As shown in Fig. 6.10(b), this topology lowers the minimum number of stacked transistors by one (the switch transistors). In addition, it also lowers the LO signal swing required to fully switch the transconductor cells by \(V_{dssat} (V_{gs})\). It theoretically has the same conversion gain as the Gilbert-cell mixers. While the noise from the switches appears at the output around zero-crossings of LO voltage, this noise appears at the output of switched-\(g_m\) mixers as common-mode noise and can be rejected. However, the flicker noise of the transconductor transistors in switched-\(g_m\) mixers is only partially mixed up and filtered out, whereas that in Gilbert-cell mixers is completely up-converted at the absence of mismatch. The way to understand it is to model \(1/f\) noise as a noise voltage source in series with the gate. This noise voltage is alternately multiplied by 0 and \(g_{m0}\) (intrinsic transconductance). Depending on the shape of the LO drive, part of this noise power "remains" in the baseband, while the other half is mixed up around harmonics. As a result, the switched-\(g_m\) mixer exhibits significantly more \(1/f\) noise due to the transconductor devices than does a Gilbert-cell mixer. Although resistive degeneration of the transconductor devices helps reduce this effect, it incurs a voltage drop that is undesirable given a tight headroom budget.

Passive mixer consumes negligible DC current, and usually exhibits superior noise and linearity when LO drive at the gate is large (Fig. 6.10(c)) [24]. LO swing should be larger than the \(V_{TH}\) to fully turn on and off the switches to ensure low switch on-resistance and avoid degrading the noise and linearity performance. In practice with a non-square LO drive, the requirement on the swing is even larger to reduce the time spent in the "middle voltage range", during which the switch transistors are neither completely turned on nor off. As we recall from the discussion on the inductively-coupled VCO, the output swing at
the gate of the cross-coupled pair is 800mV to 900mV, significantly larger than the $V_{TH}$ of the switch transistors ($< 300$ mV). While the LO swing is sufficient to drive the switch transistors, there are two other trade-offs that need to be carefully considered. First, the gain of the mixer is less than $2/\pi$, augmenting the noise contribution from the following baseband stages. Secondly, the input impedance is generally low (on the order of 200-500 $\Omega$s), severely loading down the equivalent $R_p$ of the LNA output LC tank ($> 1$ k $\Omega$s) and thus degrading the front-end RF gain.

In view of the trade-offs existing in the common active and passive mixer topologies, we propose a quasi-passive mixer topology, where it behaves like either a switched-$g_{m}$ active mixer or a passive mixer depending on the LO voltage. As illustrated in Fig. 6.10(d), Similar to a switched-$g_{m}$ mixer, the LO feeds into the source and the RF into the gate of the mixer. The difference lies in the supply of the mixer: it is tied to ground instead of the positive supply rail. When the voltage difference between the mixer gate bias voltage $V_{b_{mix}}$ and the LO input voltage $V_{LOin}$ exceeds the $V_{TH}$ of any two of the four mixer transistors, the mixer behaves like a switched-$g_{m}$ active mixer. Otherwise, it behaves like a passive mixer where all transistors are turned off and negligible current flow through them. The amount of time the transistors are on depend on $V_{b_{mix}}$ and the LO swing.

The conversion gain of the mixer can be approximated similarly to that of an active switched-$g_{m}$ mixer, with an additional term capturing the piecewise-linear behavior of the mixer. Specifically, the partially passive and partially active state of the mixer can be expressed in terms of conduction duty cycle. Taking finite switch-time into account, as in [58], a first-order approximation of the conversion gain $G_C$ of the quasi-passive mixer, loaded with resistors $R_L$ is expressed as,

$$G_{Conv} = \left( \frac{\sin(\pi D)}{\pi D} \right) \left( \frac{\sin(\pi \cdot f_{LO\tau_{sw}})}{\sin(\pi \cdot f_{LO\tau_{sw}})} \cdot g_{m0} \cdot R_L \right)$$  \hspace{1cm} (6.10)

The first term in Eqn. 6.10 takes the dynamic duty-cycle (D) into account. As D becomes 50%, the CG approaches that of a switched-$g_{m}$ mixer:

$$G_{Conv} = \frac{2}{\pi} \left( \frac{\sin(\pi \cdot f_{LO\tau_{sw}})}{\sin(\pi \cdot f_{LO\tau_{sw}})} \cdot g_{m0} \cdot R_L \right)$$  \hspace{1cm} (6.11)
For low LO frequency, the 2nd term approaches 1. The CG becomes the familiar $2/\pi$ multiplied by the small-signal gain $g_m R_L$, obtained from mixing the transconductors with a square wave. At high LO frequencies (in relation to $1/T_{sw}$), the 2nd term drops below 1, lowering the gain.

The input impedance of the mixer is a function of the LO voltage. It is important to understand how much we are loading the previous stage (here the VCO). Ignoring the capacitive component, Fig. 6.11 approximates the change in the conductance $g_{in}$ as a half sinusoid with its maxima occur at the peaks of LO+ and LO- waveforms. It is helpful to consider the mixer when in the active switched-$g_m$ state as a common-gate amplifier. Notice that the resistance looking up $M_a$ is approximately $1/g_{ma}$. The input resistance looking into the source of the switch transistors is that of a common gate amplifier. This resistance can be expressed as:

$$R_{src} = \frac{(1/g_{ma})||R_L + r_{ob}}{1+gmbr_{ob}}$$

(6.12)

Since all small-signal parameters vary with the LO input voltage, we can use pulse approximation to simplify the analysis. Assuming a duty cycle of D, the average $g_{in}$ looking into one pair of switch transistors $M_{a1,a2}$ or $M_{b1,b2}$ is the maximum input conductance ($g_{in,max}$) value multiplied by D. The sinusoidal deviation from the pulse approximation
is modified by $\sin(\pi D)/(\pi D)$. Lastly, the differential input conductance consists of two occurrences of the half sinusoids contributing from both pairs of the switch transistors. The input resistance can thus be expressed as:

$$R_{in} = \frac{R_{src,max}}{2D(\sin(\pi D)/\pi D)}$$ (6.13)

PSS and PSP analysis are used to simulate the input conductance of the quasi-passive mixer. We used realistic voltage swing from the VCO at the input in the simulations. Fig. 6.12 shows the calculated and simulated $R_{in}$ as a function of the duty-cycle. The two are in excellent agreement at large duty-cycles. At smaller duty-cycles, the sinusoidal approximation of the $g_{in}$ is less accurate as a result of the poorly conducting switches. Duty-cycle is a function of the mixer bias voltage and the input voltage swing. We can see that as the duty cycle increases, the input conductance drops, loading down the VCO more heavily.

6.3.4 VCO-mixer interface

We transformer-couple the LO to the mixer to reduce the number of transistors stacked in the mixer. As the input resistance of the mixer ($R_{in}$) is reflected at the VCO load through
the transformer, VCO should be co-designed with the mixer. The loading effect becomes more severe as $R_{in}$ reduces with an increasing duty-cycle. Fig. 6.13 shows the simulated loop gain and the differential output zero-to-peak voltage as a function of duty-cycle. The small-signal loop gain ($g_m R_p$) proportionally scales with the equivalent parallel resistance ($R_p$), which is directly loaded by the mixer $R_{in}$ multiplied by the $N^2$ ($N$ being the transformer winding ratio). In this case, the resistance reflected at the VCO load tank is 2× that of the mixer $R_{in}$. The loop gain drops by half (6 dB from an unloaded 14 dB) when loaded with the mixer $R_{in}$ of 350 Ω. Although the loading effect on the small-signal loop gain is drastic, the effect on the steady-state output swing is less pronounced. We observe a 28% decrease in the output swing when the duty-cycle changes increases from 20% to 40%.

The transformer that couples the LO to the mixer needs to be optimized to maximize the voltage swing into the mixer and minimize the loading effect from the mixer. Given the technology and transformer topology, we mainly focus on the winding ratio of the two secondary coil sections. We refer to the section in the VCO as $L_s$, the one in the mixer as $L_{sf}$. As we increase $L_{sf}/L_s$ from 1:1 to 2:1, the LO drive at the mixer input increases, while the loading effect of the mixer $R_{in}$ at the VCO load exacerbates. As a result, VCO phase noise degrades. The conversion gain of the mixer increases and linearity degrades. Thermal and flicker noise both increase as the time during which the mixer is active lengthens and
more instantaneous current flows in the mixer. We chose 1:1 ratio to reduce the loading effect at the expense of lower mixer conversion gain.

It is common to insert an intermediate buffer between the VCO and mixer to provide isolation of the resistive loading of the mixer input and increase the swing. As a result, the mixer conversion gain increases and phase noise improves. However, the RF buffers usually consume a significant amount of power (comparable to that of the VCO). As a result, we decide to bypass the buffer and directly transformer-couple the LO to the mixer output.

6.4 Baseband

6.4.1 Architecture

Active-RC or \( g_m C \) circuits are commonly used for the IF filters. Since it is difficult to implement a linear \( g_m \) element, active-RC filters offer better linearity performance compared to \( g_m C \) filters under ULV supplies [11]. 2-stage biquad filters are often used instead of a single-stage active-RC filter section to increase the overall loop gain and decrease in-band loss. However, the filter characteristics of biquad-based filters are more easily affected by parasitics and OTA non-idealities. Low-Q Tow-Thomas implementation was used to make the biquad more tolerant to parasitics and OTA non-idealities. [11] uses a 6th-order Butterworth filter implemented as a cascade of three biquads of ascending Q. The order was chosen to achieve maximal out-of-channel blocker attenuation. However, under 300mV supply, it is challenging to even achieve moderate DC gain (> 20 dB) and bandwidth, further lowering the pole Q of the filters. In addition, a 6th-order Butterworth filter requires 6 OTAs, consuming significant power (6mW).

Here we propose using passive frequency-translated filters instead of active-RC biquad filters to perform channel selection obviating the need for high-gain/bandwidth OTAs. Unlike active filters, the frequency-translated filters provide no gain; if not designed carefully they might even incur loss. Therefore, we interleave open-loop OTAs with the filters to provide gain (Fig. 6.14). The filters should be placed as early in the cascaded stages as possible to suppress out-of-channel blockers and intermodulation products. However, they can not be placed directly at the mixer output because the the resistance of the load resistors
(250 Ω) is so low that it incurs significant in-band loss and reduces the conversion gain. Therefore, we chose to place the first filter after a low-gain (20 dB) amplifier stage. Two stages of filters provide (15 dB) additional rejection of out-of-channel content compared to a single filter at the expense of narrower passband. The details is discussed later in this section. After ≈ 40 dB of gain from 2 stages of differential amplifiers, the third amplification stage performs differential-to-single-ended conversion (DTS), followed by a fourth stage of single-ended amplifier. A chain of inverters completes the limiting amplifier and provides a rail-to-rail output.

6.4.2 Analog IF amplifier

Fig. 6.14 shows the schematic of the analog IF amplifiers. Due to severe headroom limitation, we used two single-ended self-biased inverters instead of OTAs as IF amplifiers. While eliminating the current source is necessary for our ULV application, this topology is more susceptible to supply noise and requires a clean supply during operation. Here the DC bias of the nFET and pFET is separated while the IF signal is amplified by both devices,
Compared to a conventional common-source stage, this current-reused topology doubles the transconductance and gain, reduces the input-referred noise by half for the same bias current. We forward-bias the bulk-source junction of the two FETs by 300 mV to reduce their $V_{TH}$. Because the $V_{TH}$ of the pFET is higher than that of nFET, the two FETs are biased to allow a higher $V_{GS}$ at the pFET to achieve the same level of inversion as the nFET. Given $V_{gs}$, the pFET and nFET are sized to 8μm/100nm and 20μm/100nm to source enough current (50μA/branch) to provide sufficient bandwidth ($\approx$ 20 MHz). The diode connection through the feedback resistor sets the input and output DC levels. The high-pass filter corner is set to $\approx$ 300 kHz by $C_{AC}$ (10 pF) and $R_{fb}$ (1 MΩ) after the Miller-gain reduction. Because a precise low-pass bandwidth is not critical as long as it is significantly higher than the IF frequency, it is set by the intrinsic bandwidth of the amplifier. All four stages of analog IF amplifiers use the same transistors and DC-biasing. We chose smaller biasing resistors of the DTS converter (200 kΩ) because there is no reduction from Miller gain.

Input-referred offsets resulting from transistor mismatches shift the DC operating point and reduce amplifier gain. In addition, the DC offsets result in a non-50% duty-cycle waveform at the output of the inverter stages subsequent to the analog IF amplifiers, when large, can even latch up the inverters. Consequently, adequate differential gain is required to provide common-mode rejection and attenuate the input-referred offset voltage of the IF amplifier chain to negligible levels. We AC-couple after each differential stages to reduce the propagation of DC offsets. To further reduce the effect of DC offsets, we DC-couple the output of the DTS converter to the input of the subsequent amplifier. Because the input DC voltage of the fourth amplifier stage is defined through resistive-feedback, the output DC voltage in the DTS stage is more accurately defined at the presence of mismatch. In simulation, we can see that even with 10 mV offsets in the input transistors, we can still faithfully amplify and square up a 100 μV signal though the chain of IF amplifiers and inverters.
6.4.3 Frequency-translated IF filter

Passive frequency-translated bandpass filters (FTBPFs) perform channel selection without the need for OTAs. These integrated FTBPFs are evolved from the original N-path filtering concept [129][36]. The principle is often used in the RF front-ends where high-Q filtering is required to prevent compression of the LNA by out-of-band blockers [85][121][6]. In a mixer-first architecture, the voltage-sampling mixers employ the same principle to perform frequency translation as well as blocker filtering at the mixer input [24].

The FTBPFs offer several advantages that allows effective integration of high-Q filters. First, their center frequency is precisely controlled by the clock. Secondly, because they are consisted of minimum channel-length switches and MOS capacitors, they occupy a small area, consume no power (excluding the clock generation) and scale well with technology. Moreover, 1/f noise and second-order nonlinearity are no longer an issue.

Fig. 6.15(a) shows the implemented differential 8-path FTBPF consisted of 8 pairs of switches and 4 capacitors. The conventional 8-phase FTBPF uses 8 grounded baseband capacitors at the end of the 8 switching paths. Here use 4 floating capacitors to save the total capacitance area by a factor of four. The MOS switches are clocked by 8 periodic clocks, each with a duty-cycle of 1/8 and phase-shifted by 1/8 of the period. One side of the switches are connected to the baseband capacitors, and the other side is connected to a large output impedance seen at the output of the IF amplifiers. Every 1/8 of the clock period, only 1 floating capacitor is connected to the IF amplifier while the rest are floating. We forward-bias the bulk-source junction of the switches to 300 mV to reduce its on resistance. Ideally, the input impedance of the FTBPF seen by the IF amplifier is that of the baseband impedance but frequency shifted to IF. A low-Q baseband impedance is converted to a high-Q bandpass one with the center precisely controlled by the IF clock [85].

The high-Q bandpass characteristic exhibits a low impedance to the out-of-channel interferers located outside of the passband of the FTBPF, while maintaining a large impedance for the signal contents near \( f_{IF} \). In addition to the desired high-Q BPF centered around the IF frequency, scaled copies at odd harmonics of the IF frequency will also appear at the
Figure 6.15: Schematic of the IF amplifier and filter chain.
output spectrum. This is usually not a concern to when the filters are used in the RF front-end because these components usually experience enough attenuation along the receiver chain to not compress any node. However, when used here where the IF center frequency is much smaller than the low-pass bandwidth of the IF amplifiers, the components near the closest odd harmonics experience little to no filtering. Therefore, it is desired to choose a relatively large IF frequency so that most of its odd harmonics fall outside the passband of the amplifiers. In addition, we should limit the rail-to-rail voltage of the IF clock to reduce the magnitude of the copies at the odd harmonics.

The discussion above on the 8-phase filter can be generalized to a M-phase filter. We want to maximize M for two reasons. First, when the source impedance has a shunt capacitive part (per our case here), increasing M increases the effective shunt impedance and minimizes the in-band loss of the desired signal caused by the capacitive part [5]. However, the bandwidth is independent of M for a given total capacitor size (M\(\cdot C_{BB}\)) [85]:

\[
BW = \frac{1}{M(R_s + R_{SW})C_{BB}}
\]

(6.14)

Where \(R_s\) is the output impedance of the source (here the IF amplifier), \(R_{SW}\) is the on-resistance of the switches, and \(C_{BB}\) is the baseband load. Note that since we have a floating capacitor, \(C_{BB} = C_B \cdot 2\). Second, M-phase filter folds any components at \((M-1)^{st}\) and \((M+1)^{st}\) harmonics of the IF are folded to the desired frequency components around IF with normalized folding gain proportional to 1/(M-1) and 1/(M+1), respectively [37]. Consequently, a large M pushes the closest folding components to higher frequencies and lowers the folding gain. Our choice of \(M = 8\) is limited by the highest speed (\(M \cdot f_{IF}\)) the 8-phase clock generation logic can operate under 300 mV supply (\(\approx 50\) MHz). In addition, designing a FTBPF with > 8 phases has marginal advantage on reducing the in-band loss and folding gain while adding complexity.

A Johnson-counter generate 8 phases from a single clock source that runs at 8\(\times\) the IF frequency. As shown in Fig. 6.15(b), a Johnson counter is consisted of 4 flip-flops connected in a loop. Because there are 16 combinations for all possible output phases, we inserted additional combinational logic (an AND and NOR gates) to lock the output phases to the
desired combination. Because the addition of the combinational logic introduces phase delays in the 8 edges, we used another set of flip-flops to synchronize or "re-time" the clock edges. Eight sets of two clock edges with 225 degree phase lag are then combined by AND gates to form 8 phases with 1/8 duty-cycle.

We forward-bias the bulk-source junction of the switches to lower their turn-on voltage. Because the switches still require a voltage higher than 300 mV to fully turn on, we used an on-chip level shifter to convert the switch gate to 600 mV. The level shifter is consisted of two latches. We sized the cross-coupled transistors small to avoid undesired latch-up. Unlike the FTBPF design under nominal supply voltages, the rail-to-rail voltage on the IF filter clock is limited. As a result, we have to size the switch transistors wide to reduce their on-resistance. Here we chose 1μm/100nm to achieve an on-resistance of 1 kΩ with 600 mV clock drive. The larger parasitic capacitance at the gate node incur larger power consumption in the 8-phase generation circuitry. To the first order, the switching power scales proportional with the increase in the capacitance (P = fC • V^2_D). Fortunately, the logic switches at only a few MHz IF frequency that it consumes negligible power (2.2 μA at 600 mV) even with the increased capacitance. We chose 20 pF baseband floating capacitor (equivalent to 40 pF C_{BB}).

6.5 Experimental results

The chip is fabricated in a 65 nm CMOS process. Fig. 6.16 shows the measured performance of the LNA. We used a vector network analyzer and a noise figure analyzer to measure the s-parameter and gain/NF respectively. External tapped-capacitor transformer is used to achieve |S_{11}| of -20 dB at 2.42 GHz (Fig. 6.16(a)). Because the input Q of a common-gate amplifier is relatively, the input-matching is rather wide-band. In fact, we can observe better than -10 dB of |S_{11}| from 2 to 2.6 GHz. The LNA achieves a 4.7 dB noise figure (NF) with 20.2 dB of gain at 2.46 GHz (Fig. 6.16(b)). We notice that the frequency of maximum gain deviates from the minimum NF (2.6 GHz vs. 2.46 GHz). The output LC tank frequency is measured to be higher than simulation because we overestimated the parasitic capacitance associated with the tank. While the noise figure is dominated by the input stage, whereas the gain is determined by the output LC tank.
Figure 6.16: Measured performance of the LNA: a) $\text{s11}$; b) gain and noise figure; c) IP3.

To account for a range of input frequencies and PVT variations, we incorporated tuning capabilities in the oscillator. Fig. 6.17(a) shows the relatively linear tuning curve of the VCO, demonstrating a tuning range of 280 MHz over 300 mV of varactor voltage change. To characterize the receiver, we used an external RF source at 2× the LO frequency with an amplitude of 350 mV to injection-lock the VCO. Fig. 6.17(b) shows the phase noise performance of the transformer coupled VCO. We achieve a phase noise of 112.9 dBc/Hz and -140.9 dBc/Hz at 1 MHz frequency offset when free-run and locked respectively.

Fig. 6.18 shows the measured performance at the mixer output. We used a spectrum analyzer to characterize the performance at a few MHz of IF frequency. We measured 20.6 dB of gain and 5.1 dB NF at 6 MHz IF frequency (Fig. 6.18(a)). Because the mixer is active for part of the period, we still observe significant flicker noise at 3 MHz or below. The 3 dB high-pass frequency corner occurs below 1 MHz, determined by the high-pass corner formed by the AC-coupling R-C network between the LNA and the mixer gate. We achieve an IP3 of 21 dBm at the mixer output (Fig. 6.18(b)). The linearity is compromised as a result of the ultra-low supply voltage. Alternatively, we can trade-off front-end LNA gain and noise figure with linearity.

To accurately measure the gain at the last stage of the IF amplifier before the signal is squared up from a chain of inverters, we used a small RF input power (-100 dBm) to avoid
Figure 6.17: Measured VCO performance: a) tuning curve; b) phase noise (free-run and injection-locked).

Figure 6.18: Measured performance at the mixer output: a) gain and noise figure; b) IIIP3.
saturating the amplifier output, critical under a 300 mV supply. We measure a conversion gain of 83 dB and a total DSB NF of 6.1 dB at 6 MHz IF (Fig. 6.19). Despite the high gain, the total NF is not degraded significantly compared to when measured at the mixer output. The 3 dB bandwidth occurs at 8 MHz as we target low-IF applications. Fig. 6.20 shows the normalized response of the full receiver with the filters turned on. We can achieve a steep channel select filtering of 12 dB and 30 dB at 1 MHz offset with one and two filters on, respectively. It is interesting to note that although the two filters are identical in design, the cascaded response is not linear. This can be explained by the fact that the FTBPF is a linear-time-variant system, and the high-order up/down-conversions modify the effective impedance observed at the IF fundamental frequency [85]. As the voltage swing at the output of the last IF amplifier stage is larger than the previous stage, the filter switch resistance decreases as a result of larger $V_{gs}$. The input resistance at a large frequency offset can be expressed as:

$$ R_{in}|\Delta f > f_o = \frac{R_{sw} + R_s}{8D} - R_s \quad (6.15) $$

As the duty cycle approaches to 1/8, the resistance approaches to the average switch resistance $R_{sw}$. The total output resistance at a large frequency offset from the IF decreases with $R_{sw}$. Consequently, the filter roll-off is steeper at the last stage of the IF amplifier.

The performance of on-chip transformers is critical to our topology. We use a symmetrical interleaved planar bifilar topology for the transformers to achieve a 6nH self-inductance with a coupling coefficient k of 0.824 and Q of 12.6. Similar to [48], the transformer structure has an octagonal shape with two parallel coils that are symmetrically interleaved side-by-side on the low-loss top metal (M9) to maximize k and Q. We performed EM simulation in ADS to obtain s-parameters of the transformer and built a simplified RLC circuit model for simulation. To verify the accuracy of the EM-simulation and the model, we also fabricated stand-alone transformer test structure and open/short/through de-embedding test structures. We probed the transformer to obtain the 2-port s-parameter and followed standard open/short/through de-embedding procedures to remove the parasitic impedance associated with the pads [60][23].
Figure 6.19: Measured gain and NF performance at the IF output.

Figure 6.20: Measured filter response normalized to the IF frequency.
Fig. 6.21: Test structures used for de-embedding: a) open; b) short.

Fig. 6.22 shows that the simulated parameters \( L, Q, k \) closely match those obtained from the measured s-parameters after de-embedding. The deviation in \( Q \) is possibly due to a pessimistic estimate of substrate loss. Another possibility is the choice for the short test structure. We connected two signal pads to ground only in the outer boundary of test structure. This structure is not as accurate as the one with ground connection on the sides and a through connection between the signal pads.

We measured the SNR at the last stage of the analog IF amplifier. Based on the SNR requirement of Zigbee standard (7 dB SNR at the demodulator for a 2 MHz channel bandwidth), this receiver achieves -91 dBm sensitivity. Fig. 6.23(a) shows the selectivity of the receiver, defined as the maximum input power at different frequency offsets required to achieve 7 dB SNR at the analog IF output. We chose an IF frequency of 3.3 MHz in this experiment. We can see that the selectivity profile is shaped by the frequency translated filter, which contains gain peaks at the IF frequency and its odd harmonics (after a scaling factor). Therefore, our sensitivity is best at the IF frequency of ±3.3 MHz, and followed by its odd harmonics of ±9.9 MHz, ±16.5 MHz, etc. Because we do not have I/Q paths to reject image, we see approximately equal gain from the positive and negative sidebands. We also set up for BER testing using a FPCA to generate a pseudo-random number sequence (PRNS). The limited IF output from the chip is demodulated on the FPGA and compared with the transmitted data sequence. Fig. 6.23(b) shows the BER obtained for a range of
Figure 6.22: The simulated and measured parameters of the transformers: a) $L$; b) $Q$; c) $k$. 
Table 6.3: Performance Comparison

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply voltage [V]</td>
<td>0.3</td>
<td>0.6</td>
<td>0.4</td>
<td>0.75</td>
<td>0.5</td>
</tr>
<tr>
<td>Power [mW]</td>
<td>1.6</td>
<td>32.5</td>
<td>0.33</td>
<td>11.3</td>
<td>26.2</td>
</tr>
<tr>
<td>RF $f_{IN}$ [GHz]</td>
<td>2.46</td>
<td>2.4</td>
<td>2.34</td>
<td>1.95</td>
<td>5.6</td>
</tr>
<tr>
<td>Voltage gain [dB]</td>
<td>83</td>
<td>67</td>
<td>N/A</td>
<td>31.5</td>
<td>27.1</td>
</tr>
<tr>
<td>NF [dB]</td>
<td>6.1</td>
<td>16</td>
<td>7</td>
<td>3.5</td>
<td>8.6</td>
</tr>
<tr>
<td>IP3 [dBm]</td>
<td>-21.5</td>
<td>-10.5</td>
<td>-7.5</td>
<td>-10.5</td>
<td>-17.9</td>
</tr>
<tr>
<td>PN@1 MHz [dBc/Hz]</td>
<td>-140</td>
<td>-115</td>
<td>-106</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Technology</td>
<td>65nm CMOS</td>
<td>90nm CMOS</td>
<td>130nm CMOS</td>
<td>90nm CMOS</td>
<td>180nm CMOS</td>
</tr>
<tr>
<td>Area [mm²]</td>
<td>2.496</td>
<td>2.9</td>
<td>4.4</td>
<td>4.3</td>
<td>N/A</td>
</tr>
</tbody>
</table>

RF input power. When input with a BFSK signal at 200kbps modulated with a pseudo-random number sequence, we achieved a sensitivity of -91.5 dBm for a BER of 10e-3 (this is consistent with theory given the 3 dB advantage of QPSK over coherently demodulated BFSK). We also performed over-the-air link test, where the FPGA modulates the baseband data onto a 2.4 GHz carrier in a commercial transmitter chip, and the custom IC receives the data over the air and performs demodulation in the FPGA. Fig. 6.23(c) shows a waveform showing the transmitted and correctly received bit sequence.

A performance comparison with the relevant recently published state-of-the-arts is shown in Table 6.3. To the best of our knowledge, this radio allows a lower supply voltage and achieves higher total gain than any GHz-range radio reported to date. Our ultra-low supply voltage results in limited headroom and degraded linearity performance compared to the other receivers operating off a higher supply voltage. As we’ll discuss in the next section, the linearity significantly improves as we operate off a 400 mV or higher supply.

6.6 Discussion

6.6.1 Supply dependence

As we target a variety of energy-harvesting applications where the voltage at the harvester output varies, the design should be functional over a range of supply voltages. We characterize the performance of the receiver at 300 mV, 400 mV and 500 mV supply voltages
Figure 6.23: Receiver system test: a) far-out selectivity; b) BER at 200 kbps; c) wireless transmitted and received bit sequence.
Table 6.4: Measured receiver performance over supply variation

<table>
<thead>
<tr>
<th></th>
<th>300 mV</th>
<th>400 mV</th>
<th>500 mV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current (mA)</td>
<td>5.37</td>
<td>7.61</td>
<td>10.9</td>
</tr>
<tr>
<td>Gain (dB)</td>
<td>82.8</td>
<td>82.8</td>
<td>80.8</td>
</tr>
<tr>
<td>NF(DSB) (dB)</td>
<td>6.1</td>
<td>5.5</td>
<td>6.2</td>
</tr>
<tr>
<td>IIP3 (dBm)</td>
<td>-20.6</td>
<td>-11</td>
<td>-14</td>
</tr>
<tr>
<td>LO PN at 1 MHz(Locked) (dBc/Hz)</td>
<td>-140.9</td>
<td>-142.2</td>
<td>-143.3</td>
</tr>
<tr>
<td>LO PN at 1 MHz(Free-run) (dBc/Hz)</td>
<td>-112.9</td>
<td>-110.3</td>
<td>-107.9</td>
</tr>
</tbody>
</table>

Table 6.5: Measured performance breakdown under 0.3/0.4/0.5 V supplies

<table>
<thead>
<tr>
<th></th>
<th>LNA</th>
<th>Mixer</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>at the output of</td>
<td>20.2/20.2/20.2</td>
<td>20/21/21.3</td>
<td>82.8/82.8/80.8</td>
</tr>
<tr>
<td>Gain (dB)</td>
<td>4.7/4.9/5.2</td>
<td>5.3/5.2/5.2</td>
<td>6.1/5.5/6.2</td>
</tr>
<tr>
<td>NF (dB)</td>
<td>-20.8/-15.9/-16.4</td>
<td>-21.5/-11/-14</td>
<td></td>
</tr>
</tbody>
</table>

(Table 6.4). To further understand the effect of supply change on each stage, we also measured performance at the output of the LNA, mixer and IF amplifiers separately (Table 6.5).

The LNA current is fixed when the supply voltage is increased. Our proposed transformer-coupled topology maximizes voltage headroom where \( V_{ds} \approx V_{dd} \). Due to the enhanced channel-length modulation in deep sub-micron process, an 100 mV increase in \( V_{dd} \) (\( V_{ds} \)) brings about a 30% increase in the bias current given the same \( V_{gs} \). This effect is exacerbated as the \( V_{TH} \) is reduced from the body-effect when the bulk is tied to the supply. Therefore, we reduced \( V_{gs} \) so that the LNA carries the same current across different supply voltages. While the LNA current is fixed, the current in the VCO increases with the supply as the gate of the cross-coupled transistors is biased at the supply voltage. In addition, as the larger LO swing increases the instantaneous current in the mixer, indirectly sourcing more current from the VCO. As a result, the total current increases with the supply voltage.

The voltage gain stays approximately the same. The 2 dB decrease in the total voltage gain at 500 mV supply is due to the reduced gain in the IF amplifiers. As the inversion coefficient increases with the increased \( V_{gs} \), the intrinsic gain \( g_m/g_{ds} \) is decreased. Nonetheless, the RF gain from the LNA and mixer stage stays approximately the same. A higher LO drive increases the mixer current and \( g_m \) during its active state, in turn slightly increasing
the mixer conversion gain.

The noise figure of the LNA degrades slightly as the supply increases. To fix the current in the LNA, we need to decrease $V_{gs}$ and move the bias of the transistors toward weak inversion. The noise figure of the LNA degrades as the inversion coefficient decreases. The noise figure of the RF front-end stays approximately the same, while that of the IF amplifiers improves at 400 mV. As the current in the IF amplifiers increase, the input-referred noise decreases. The linearity improves significantly at 400 mV supply. As expected, a larger $V_{ds}$ leads to more superior linearity when the LNA is biased in the moderate inversion. In addition, a larger LO drive increases the $V_{gs}$ and $V_{ds}$ of the mixer switch transistors, reducing their on-resistance and improve the mixer linearity. However, the linearity degrades at 500 mV, as LNA enters weak inversion region of operation.

6.6.2 Effect of mixer gate bias voltage

As discussed earlier, the input impedance of the mixer ($R_{in}$) is reflected at the VCO tank through the transformer, loading down the VCO. The bias voltage of the mixer gate ($V_{bmix}$) changes $R_{in}$, in turn affecting the performance of the VCO. Fig. 6.24(a) shows the operating current of the oscillator during free-run and injection-locked. As $V_{bmix}$ increases, the total parasitic capacitance dominated by $C_{gs}$ at the mixer source node increases, adding phase lag to the oscillator feedback loop and reducing the free-running frequency. We decreased the capacitance of the varactors to maintain a fixed free-running frequency. As a result, the average current consumption is lowered at high $V_{bmix}$. Fig. 6.24(b) shows the VCO output power including the buffer loss (simulated 6.3 dB). As the $V_{bmix}$ increases, the loaded $R_{p}$ of the VCO tank reduces. Along with the reduced average current, the output power is lowered by 4 dB as the $V_{bmix}$ is increased from 140 to 260 mV.

Fig. 6.24(c)6.24(d) shows the phase noise measured at 10kHz and 1MHz. We can see that the closed-in phase noise does not change significantly, whereas the far-out phase noise degrades by 5 dB as $V_{bmix}$ increases from 140 to 260 mV. The loading of the mixer reduces the $R_{p}$ and Q of the VCO tank, as well as the output signal swing. According to Leeson's model on phase noise [66],
Figure 6.24: Effect of mixer bias voltage on the measured VCO: a) average current; b) output power (after buffer); c) phase noise at 10 kHz frequency offset; d) phase noise at 1 MHz frequency offset.
\[ L(\Delta \omega) = 10 \cdot \log \left\{ \frac{2FkT}{P_{\text{sig}}} \cdot \left[ 1 + \left( \frac{\Delta \omega}{2Q\Delta \omega} \right)^2 \right] \cdot \left( 1 + \frac{\Delta \omega_{v/f}}{\Delta \omega} \right) \right\} \]  
(6.16)

In general, the phase noise degrades with a smaller signal power and Q. However, the phase noise mechanisms is more complicated at low frequency offsets. In addition to flicker noise, there is also noise generated through AM-to-PM conversion. When the output swing modulates the nonlinear parasitic capacitance, the envelope of the signal varies. While amplitude noise does not affect phase noise to the first order, the nonlinearity of the capacitance converts a portion of the amplitude noise to phase noise at low frequency offsets. Therefore, we do not see significant degradation in the closed-in phase noise at high \( V_{\text{bmix}} \) as the output swing reduces, and less amplitude noise is converted to phase noise.

As we recall from an earlier discussion, \( V_{\text{bmix}} \) and the LO drive determine the duty-cycle, or period of time that the mixer is active. We now analyze its effect on the receiver performance. Fig. 6.25(a) shows the measured gain and NF at the mixer output. We see that as the duty-cycle increases with \( V_{\text{bmix}} \), the gain peaks while the NF falls at mid-way (\( \approx 180 \) mV) of the bias voltage range. According to Eqn. 6.10, the conversion gain (\( G_c \)) is an increasing function of the small-signal transconductance (\( g_m \)) and a decreasing function of the duty-cycle (\( D \)). When \( V_{\text{bmix}} \) is low, the gain is limited by the low \( g_m \) resulted from small swing \( V_{gs} \). However, when \( V_{\text{bmix}} \) is high, the gain is limited by the large duty-cycle. The NF increases with \( V_{\text{bmix}} \) as the transistors are turned-on for a longer period of time, therefore contributing more noise.

Fig. 6.25(b) shows that the measured linearity at the mixer output peaks at approximately 220 mV. The maximum \( V_{gs} \) when the transistors are on is \( V_{\text{bmix}} + V_{\text{LO,zero}} \). Given the same LO drive, the higher \( V_{\text{bmix}} \) increases the transistor's inversion coefficient, improving the linearity. However, the loading effect at the VCO output exacerbates as \( V_{\text{bmix}} \) increases, decreasing the LO drive and degrading the linearity. At small \( V_{\text{bmix}} \), \( V_{gs} \) is small during the small fraction of period that the transistor is on. As a result, the conversion is low while the linearity improves.
6.6.3 Comparison of mixer topology

To further demonstrate the compelling trade-offs our proposed quasi-passive mixer offers, we compare measured performance with the active switched-$g_m$ mixer. We tie the node between the load resistors to 300 mV to transform the mixer into an active switched-$g_m$ mixer. Fig. 6.26(a) shows the measured conversion gain of the RF front-end with the quasi-passive and the active mixer. We see a 6 dB higher gain above the high-pass corner in the active mixer when compared to the quasi-passive mixer. Given the same $V_{bmix}$ and LO drive, the maximum current and hence conversion transconductance $g_m$ in the quasi-passive mixer is are lower when compared with the active mixer as the drain-source voltage ($V_{ds}$) is reduced. The effect of $V_{ds}$ on current is even more pronounced in a deep sub-micron process. As a result, the gain of this mixer is lowered.

Fig. 6.26(b) shows the measured noise figure of the RF front-end with the quasi-passive and the active mixer. The noise of the quasi-passive mixer is $\approx 2$ dB lower than its switched-$g_m$ counterpart as a result of two mechanisms. First, the 1/f noise is reduced as the current in the mixers reduces. Although 1/f noise still partially remains in the baseband at the mixer output, the magnitude is lower. Secondly, the mixer does not contribute thermal noise during its passive state. If the duty-cycle is low, the total amount of noise eliminated during its passive state could be significant. Consequently, although the gain is lower than
Figure 6.26: Measured RF front-end performance comparison with the active mixer and quasi-passive mixer: a) gain; b) noise figure; c) IIP3; d) VCO phase noise.

A conventional switched-$g_m$, the noise is non-intuitively lower.

Fig. 6.26(c) shows the measured linearity of the RF front-end with the quasi-passive and the active mixer. As the conversion gain is lowered, IIP3 is also improved with the quasi-passive mixer by $\approx 5$ dB. Although a higher supply allows for a larger headroom, potentially improving the linearity, the advantage at 300 mV is minimal as the headroom is reduced by the voltage drop across the load resistors. This voltage drop is larger in the active mixer when compared to the quasi-passive mixer as the instantaneous current is larger.

Fig. 6.26(d) shows the measured phase noise while the VCO is loaded with the quasi-passive and the active mixer. We do not observe any noticeable differences between the
the two scenarios. This can be explained by the similar input resistance of the two mixers. When $V_{gs} < V_{TH}$, the active mixer has significantly larger sub-threshold current than the quasi-passive mixer as a result of larger $V_{ds}$. As a result, the input resistance of an active mixer is lower. However, when $V_{gs} > V_{TH}$, there is current flowing in all four transistors of the quasi-passive mixer, whereas there is only significant current flowing in two transistors of the active mixer. Consequently, the input resistance of the active mixer is slightly higher. On average, the input resistance looking into the source of the two types of mixers are similar.

It is also interesting to compare the performance of proposed quasi-passive mixer with the conventional passive mixer. The main difference between the two are the location of the RF and LO inputs. When used in a receiver, the source of the quasi-passive mixer interfaces with the VCO, whereas that of the passive mixer interfaces with the LNA. Table 6.6 shows the simulated performance of the RF front-end with 3 mixer topologies: the quasi-passive, passive with the same mixer bias and LO drive and passive mixer with higher mixer bias and LO drive. The second topology implies using the voltage at the VCO transformer’s secondary coil when driving the gate of the passive mixer, whereas the third topology implies driving the mixer gate with the voltage at the primary coil. We can see that for the same mixer bias and LO drive, passive mixer provides a higher impedance, lower gain and similar NF and IIP3. For a higher mixer bias and LO drive, the performance of the passive mixer improves as the switches can be fully turned on and off. It has a similar gain, better NF and IIP3 at the expense of lower input impedance. Whereas passive mixer seems to be more superior when driven hard, the low input impedance directly loads down the LNA output tank, heavily degrading the RF front-end gain and thus increasing the noise contribution from the IF stages. A low-noise IF transimpedance amplifier (TIA) is required in the case for insufficient RF front-end gain.

6.7 Conclusion

Ultra-low supply voltages present significant challenges for RF receiver designs. In this paper, we presented a 2.4 GHz wireless receiver operating from a 300 mV supply voltage, the lowest reported to date. The chip was fabricated in a standard 65nm CMOS process.
Table 6.6: Simulated mixer performance

<table>
<thead>
<tr>
<th></th>
<th>Quasi-passive</th>
<th>Passive (same drive)</th>
<th>Passive (large drive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIN</td>
<td>305 $\Omega$</td>
<td>498 $\Omega$</td>
<td>150 $\Omega$</td>
</tr>
<tr>
<td>Gain</td>
<td>0.2 dB</td>
<td>-2.2 dB</td>
<td>0.6 dB</td>
</tr>
<tr>
<td>NF(DSB)</td>
<td>12.1 dB</td>
<td>12 dB</td>
<td>5.7 dB</td>
</tr>
<tr>
<td>IIP3</td>
<td>-0.8 dBm</td>
<td>0.23 dBm</td>
<td>7.25 dBm</td>
</tr>
</tbody>
</table>

We used a transformer-coupled technique throughout the RF front-end to avoid stacking transistors and thus maximize the voltage headroom. We forward-bias bulk-source junction of the transistors in the RF and IF amplifiers to reduce the $V_{TH}$, making it possible to bias key transistors in the moderate inversion to optimize the trade-off between power consumption and $f_T$. In the baseband, frequency-translated filters were exploited to perform channel-selection without the need for OTAs. Despite the ultra-low supply voltage, the receiver still achieves -91.5 dBm sensitivity for a BER of $10^{-3}$ when input with a BFSK signal modulated by a PRNS. This work demonstrates the feasibility of operating GHz-range receivers under a supply voltage as low as 300 mV, creating new opportunities for directly interfacing integrated circuits with a variety of energy harvesters.
Chapter 7

WIRELESS SENSING SYSTEM DESIGN AND INTEGRATION

Sensing systems made from off-the-shelf (OTS) components usually exhibit high power consumption (> 20mW). For instance, OTS amplifier and ADC consume ~100 µW; MSP430 consumes ~500 µW (in low-power mode); transmitter (Nordic) consumes 20~30 mW. In continuous transmission mode, a typical coin-cell battery (Zinc Air Size 5, 33 mAh) lasts approximately 2 hours. To extend battery life, we have developed low-power custom wireless ICs used in BASN applications. The first SoC (Neurec) was an integrated biosignal transmission interface chip which wirelessly streams a digitized bio-signal waveform over 15 m. It consumed 500µW under 1V supply [100]. Later, we developed the “Bumblebee”, a 7.6mm×8.7mm PCB that contains the “Neurec” IC, all necessary passive components, and a 0.17 gram battery allowing over 3 days of continuous recording. It can reliably transmit bio-physiological data (such as EMG and ECG) over 10 m to a USB-compatible 433 MHz receiver [87].

Though the “Bumblebee” show great potential, it still faces many challenges, among which the most critical issue is node lifetime. In many applications, such as long-term monitoring of chronic illnesses, limited battery lifetimes severely undermine the deployment of BANs, since the required node operating lifetime is effectively indefinite. A large battery increases the form factor of the node, making the node unwearable or uncomfortable, while a small battery requires frequent changing and reduces wearer compliance. UHF RF-powering provides a battery-free solution for wireless sensing tags. In an RFID system, a reader sends power and control data to the tags. The tags modulate the reflection coefficient and backscatter the sensor data to the reader. We developed a fully-passive 900 MHz RFID tag IC with integrated low-noise amplifier and full EPC Class 1 Generation 2 protocol compatibility [141]. The tags achieved -12 dBm sensitivity. This translates into a 3 m range with an OTS RFID reader, enabling previously impossible recording scenarios like in-flight recording from...
small insects.

One problem with RFID-like systems is that they require infrastructure (base station or reader) close by (within a few meters). There is an emerging need for active miniaturized systems that stream bio-physiological data in the far field, which would enable continuous monitoring of patient health. To extend the range in battery-less operations, harvesting ambient energy offers an appealing alternative to RFID systems. A variety of energy harvesting mechanisms exist, so the correct choice is highly application dependent. We are particularly interested in thermal energy harvesting, as body area sensors allow access to a reasonably consistent temperature gradient.

In a recent work, we demonstrated the feasibility of powering a wireless biosignal acquisition chip solely from a thermoelectric harvester and/or RF power [148]. The chip contains integrated supply regulation, analog front-end, DSP and transmitter (TX). A closed-loop power management monitors the harvested energy and duty-cycles the TX accordingly. An on-chip boost converter boosts the TEG voltage from 30 mV to 1.35 V at 38% efficiency.

In this chapter, I’ll present details to the design and integration of the “Bumblebee” and the battery-less BASN SoC.
Part I

THE “BUMBLEBEE”

7.1 The circuit design of the wireless tag IC

As shown in Fig. 7.1, the neural interface comprises an analog front-end with gain variable from 42 to 78dB, an 8b ADC, and a 100kb/s 2-FSK transmitter. The system operates in the Medical Implant Communications Service (MICS, 402 to 405MHz) and 433MHz ISM bands. A new transmitter architecture allows high power efficiency for low output power MICS-band communications. Also presented is a new complementary fully-differential low-noise analog front end with a Noise Efficiency Factor (NEF) of 2.48.

A programmable amplifier, shown in Fig. 7.2(a), amplifies microvolt-level signals over a 25mHz to 11.5kHz bandwidth. AC coupling at the inputs of both the LNA and VGA prevent offset amplification. A fully differential closed-loop architecture is used to ensure sufficient linearity and supply rejection. The LNA inputs are AC-coupled using 20pF capacitors and high-resistance (>100GΩs) MOS-bipolar pseudo-resistors. Separation of the signal and bias paths allows the input to simultaneously drive the n- and pFETs of the input stage. This complementary-input strategy doubles the effective transconductance for a given bias current while the output noise remains constant, thus reducing the input-referred noise voltage by a factor of two [45]. Thick-oxide MOS transistors with large gate areas are used at the input to reduce gate leakage while minimizing 1/f noise. The variable gain amplifier (VGA) consists of a complementary rail-to-rail folded-cascode core with programmable capacitive feedback. Six-level variable gain is set by selecting the feedback capacitors; the six variable high pass corners are set by programming the feedback transconductor bias current. Alternately, high impedance pseudoresistor feedback can be selected to obtain a sub-Hz high pass corner. The VGA is sampled by an 8b SAR ADC, designed to operate at sample rates from 10 to 100 kS/s.

The system streams one channel of continuous spike data sampled at 8b, 9.1kS/s with
interleaved synchronization packet headers. We target a 100kb/s datarate and the maximum allowable MICS-band EIRP of -16dBm (25μW) to maximize communication range for untethered animal in-vivo experimentation. Low power transmitters traditionally suffer from poor global efficiency since the frequency generation and modulation overhead begin to compete with the power amplifier dissipation. These effects are addressed by employing a new transmit architecture that operates entirely at the on-chip crystal reference frequency (44.545MHz) and drives a 9× frequency multiplying power amplifier, as shown in Fig. 7.2(b). The baseband FSK data directly modulates the reference oscillator using capacitor pulling. A 4pF on-chip capacitor creates a 16.4kHz Δfref, resulting in approximately 148kHz FM frequency deviation at the 400MHz carrier.

The reference clock drives a 9-stage DLL. Each non-inverting delay stage in the DLL consists of a current-starved inverter controlled by two feedback loops for frequency and duty-cycle control, respectively. The operation of the edge-combiner (EC) is dependent on the overlap of rising and falling edges for a period of T/18, where T is the period of the reference input at 44.545MHz. Edge-combiner performance is critically dependent on equally-spaced edges from the delay cells of the DLL. Balancing delay-stage loads, symmetric
Figure 7.2: a) Low-noise analog front-end schematic; b) Transmitter edge combiner schematic details and the output match.
layout, and dual-edge locking are essential for maximizing output power and minimizing reference spurs. This topology allows operation of the entire DLL at 44.545MHz, allowing quartz stability without the need for an RF carrier frequency PLL/DLL. This technique can be generalized to other multiplication factors, determined by the number of stages in the DLL and switching legs in the edge-combiner. Both DLL loops must demonstrate sufficient bandwidth to track the FSK baseband signal.

The edge-combiner behaves like a high-efficiency non-linear power amplifier, and produces pulses of current based on overlap of separate delay cells in the DLL. This current is absorbed by a tapped-capacitor LC matching network, which transforms the TX source impedance to match a 500 antenna and attenuates out-of-band spurs. The measured $|S_{22}|$ is shown in Fig. 7.2(b), showing better than -10dB return loss over a 30MHz bandwidth. This tank produces a sinusoidal voltage waveform of amplitude, $2/\pi I_{tail}R_p$, where $R_p$ is the loaded tank impedance and $I_{tail}$ is the edge-combiner DC current.

The proposed system was implemented in a 2.51mm$^2$ die, shown in Fig. 7.4(a), using a 0.13µm CMOS process. All biasing, logic, and passives are integrated on-chip except for two quartz resonators and 7 passive components, used for antenna impedance matching, DLL loop filtering, and system clock generation. The overall transfer function for the analog front end and the measured LNA input-referred voltage noise spectrum are shown in Fig. 7.3(a). The measured gain is adjustable between 42dB and 78dB. The integrated noise from 0.1Hz to 25.6kHz is 1.9µV$_{rms}$. The power dissipation of the entire analog front-end, including ADC and biasing, is 75µW.

The measured transmitter FSK deviation of 145 kHz is shown in Fig. 7.3(b). The quartz-based transmitter provides 2ppm frequency stability over the relevant 25 to 45°C MICS temperature specification, eliminating the need for frequency tracking algorithm as in [14]. The transmitter consumes 400µW with a -16dBm output power at a 100kb/s datarate with an edge combiner PA efficiency of 16%. The data rate for the transmitter is limited to approximately 200kb/s by the 1s settling time of the DLL. The in-band measured spurs are less than -30dBc, conforming to the MICS standard. Reference spurs, offset from the carrier by multiples of fref, result from unequal delay in the delay cells due to random process mismatch and systematic layout related load imbalances. The measured reference
Figure 7.3: a) Measured gain and noise response of LNA and VGA; b) Transmitter spectrum with FSK symbols overlaid.
spurs at 44.5MHz offset from the carrier are -22dBc, falling below -40dBc with a narrow-band antenna.

System functionality was verified by applying artificial neural spikes of 160µV_{pp} amplitude at the input, and reconstructing the signal from digital data taken from a commercial FSK receiver board located at a distance of 15m from the transmitter, using a 1/4-wave 500 antenna pair. Fig. 7.4(b) shows the reconstructed signal overlaid on the original spike and also summarizes the performance of the neural streaming system.

7.2 The design of the WBAN node

The Bumblebee is an ultra lightweight, high-performance, low power wireless sensor ideal for electromyography (EMG) and other sensing applications[87]. It records a fully differential signal, amplifies and digitizes the signal, and wirelessly transmits the data to a paired receiver.

7.2.1 The PCB

Fig. 7.5 shows the functional block diagram. The entire system is powered from a single 0.13 gram coin-cell battery, which is subsequently regulated to 1V. Other than a regulator IC, all functionality is performed by a single custom-designed IC. There is no non-volatile memory within this chip, so an external programmer is required to load appropriate settings for the chip (mode of operation, gain values, transmit power) upon powerup. The entire device is completely self-contained and compatible with a variety of recording paradigms (EMG, spike-based, audio).

7.2.2 Receiver

We developed a low-cost companion receiver for the Bumblebee. Since the receiver sits at a PC or basestation, it is not power-constrained and can utilize off-the-shelf electronics. This receiver reads in the wireless data transmitted from the Bumblebee, performs clock and data recovery, and reconstructs the analog signal as well as sends the digital sample data over a standard USB connection to a computer terminal for further processing. Fig. 7.6
Figure 7.4: a) The microphotograph; b) Reconstructed measured data and system performance summary.
contains the block diagram of this receiver.

This receiver demodulates the signal, decodes the packet and outputs the raw data both via USB to a computer system as well as a reconstructed analog signal. Demodulation occurs within the Melexis EVB7122 receiver. The resulting bitstream is processed by a TI MSP430F2012 microcontroller which runs a custom clock and data recovery (CDR) algorithm. The programming of the microcontroller was done in C and is entirely interrupt driven to achieve accurate timing and avoid any operating system overhead.

7.2.3 The system

The Bumblebee functionality comes from a single chip, so the final product can be extremely small since minimal support circuitry is required. The Bumblebee uses a thin 4-layer PCB measuring only 7.6 x 8.7 mm². Low weight is critical to allow in vivo experimentation on small animals (mice, rats) and insects (locusts, moths). The fully-populated PCB weighs 0.18g without battery, 0.3g with a 337 battery, and 0.34g with a size 5 zinc-air battery. Fig. 7.7 shows the front and back of the Bumblebee with the unencapsulated chip.
mounted with chip-on-board (COB) connections. The deployed board used epoxy encapsulation to protect the bondwires.

The Seiko S-11L10 Super-Low Output Low Dropout CMOS Voltage Regulator was chosen in conjunction with a 1μF bypass capacitor both because of its small size and low quiescent current, consuming only 9μA during operation. Two different batteries were deemed suitable for the Bumblebee: the ZincAir Size 5 and the Silver Oxide 337. Each has a tradeoff between power density and total weight. ZincAir batteries have significantly higher energy density when compared to a standard Silver Oxide coin-cell, however commercially available sizes in this technology are larger than some Silver Oxide batteries.

For applications where the absolute minimum weight is needed, the Silver Oxide size 337 is ideal. It weighs only 0.13g and provides approximately 8.3mAh at 1.55V. When a slightly higher weight can be tolerated the ZincAir size 5 battery is the best choice. Weighing 0.17g, this battery provides 35mAh at 1.25V; a more than 3x increase in stored energy over the 337 battery for only 0.04g of additional weight.
Figure 7.7: Front and back of the Bumblebee board on a quarter. A thin 30 gauge monopole wire antenna a few cm long allows >10m range.
Part II

A BATTERYLESS ENERGY HARVESTING WIRELESS BODY-AREA NETWORK (WBAN) SOC

Energy harvesting from ambient energy sources, such as thermal gradients or mechanical vibrations, potentially provides indefinite lifetime. Examples such as [41][67] have shown that commercial thermopiles can be applied to BANs. To eliminate battery changing, nodes can operate solely from energy harvesting instead of using a battery, although this introduces new challenges. The full system must consume less energy than the amount harvested, high power components such as the transmitter must be heavily duty-cycled, and the node must cope with time varying harvested energy profiles [16].

To ensure sustained operation of the node using harvested energy, on-node processing to reduce the amount of data transmitted, power management, and ultra-low power (ULP) circuits are key. Recent advances in ultra-low power chip design techniques, with many targeting wireless sensor networks, have enabled a push toward long lifetime BAN devices performing complex applications. For example, [56] presents an ECG acquisition and processing SoC with a 3-channel analog front-end (AFE) and flexible, generic DSP components. Clock-gating and duty-cycling are used to reduce power, but without voltage reduction techniques (VDD=1.2V), the minimum power is 31.1µW. The SoC does not include a transmitter or voltage regulators. An SoC for EEG seizure detection integrates a COTS radio and does not include voltage regulation or power management [128]. The system in [20] integrates several chips with solar cells and a battery to accomplish near-perpetual operation for measuring intraocular pressure. The system consumes 3.3fW/bit at 400mV when taking one measurement every hour. These works show that energy harvesting could provide a viable power supply for ULP BAN circuits. However, integration of a complete wireless, flexible, easily deployable BAN node on an SoC that supports closed-loop power management and energy harvesting has yet to be demonstrated.
We utilize recent advances in energy harvesting, low voltage boost circuits, dynamic power management, subthreshold processing, bio-signal front-ends, and low power RF transmitters to realize an integrated reconfigurable wireless BAN SoC for ECG, EMG, and EEG applications with autonomous power management for completely battery-free operation [148]. This SoC can run indefinitely from energy harvested from body heat while worn, and potentially decreases cost by having high integration and targeting a wide range of bio-electric sensing applications.

7.3 System overview

Conventional wireless sensors use batteries (Fig. 7.8), limiting node lifetime and reducing user compliance due to the requirement for charging or replacing batteries. These sensors often require high data transmission rates, which quickly drain battery energy. More sophisticated approaches that include on-node processing and duty-cycling of the power-hungry radio are available [13], but the lifetime of such COTS nodes is usually limited to a few days, and custom BAN chips have not yet integrated radios, processing, and power management with energy harvesting. In contrast, we propose a wireless BAN chip powered by energy harvested from human body heat using a thermoelectric generator (TEG). This, in conjunction with ULP circuits, intelligent duty cycling of power-hungry blocks (e.g., the transmitter), and a programmable power management system allows for indefinite operation of the chip.

To demonstrate, we present a chip targeting ExG applications.

To achieve flexible data acquisition and processing while operating the node solely from harvested energy, we propose a system architecture, illustrated in Fig. 7.9, which comprises four subsystems. First, the energy harvesting/supply regulation section boosts a harvested supply input as low as 30mV up to a regulated 1.35V using an off chip storage capacitor. It provides five regulated voltage supplies to the rest of the chip, and generates a bandgap reference. Second, the four-channel AFE subsystem provides bio-signal acquisition with programmable gain and sampling rate, amplifying bio-signals as low as a few μVs while consuming <4μW/channel. A variable gain amplifier (VGA) maximizes the signal at the input to an 8-bit successive-approximation (SAR) analog to digital converter (ADC), reducing the ADC resolution requirement. Third, the acquired data is sent to a subthreshold digital
processing subsystem that also performs mode control and power management (including power/clock-gating of blocks and dynamic voltage scaling (DVS)) based on the available energy on the storage capacitor. The digital section includes a custom digital power management (DPM) processor, general purpose microprocessor (MCU), programmable FIR, 1.5kB instruction SRAM/ROM, 4kB data memory FIFO, and dedicated accelerators for ECG heart rate (R-R) extraction, atrial fibrillation (AFib) detection, and EEG band energy calculation. The DPM is responsible for power management, node control, data flow management, and overseeing all processing on-node. Finally, a sub-mW 400/433 MHz MICS/ISM band frequency-multiplying transmitter (TX) performs BFSK transmission up to 200 kbps. The TX has low instantaneous power consumption to avoid the need of large filtering capacitors on the supplies and is intelligently duty-cycled to achieve low average power consumption.
Figure 7.9: System block diagram for the proposed chip comprising the energy harvesting/supply regulation, AFE, subthreshold DSP, and TX subsystems [148].

7.4 Energy harvesting/supply regulation subsystem

The energy harvesting subsystem is designed to harvest energy from RF, thermoelectric, or solar power sources and to provide regulated voltages to the rest of the chip.

7.4.1 Harvesting Energy from a Thermoelectric Generator (TEG)

At the heart of thermal energy harvesting is the thermoelectric generator (TEG), which generates a voltage proportional to the difference of the temperatures applied to the two sides, or the hot and cold plates. In our application, one side is placed on the warm skin, while the other side is exposed to the cooler ambient air. Thermoelectric generators (TEGs) are constructed of thermopiles in series. This arrangement places a bound on the maximum voltage achievable from a given temperature difference for a given size. The Seebeck coefficient of a conventional thermocouple (bismuth telluride) is $\pm 0.2 \text{mV}/^\circ\text{C}$. Assuming a temperature gradient of $1^\circ\text{C}$, a $11cm^2$ TEG will generate much less than 1V. In addition,
the surrounding air presents a large thermal resistance that dramatically reduces the effective temperature gradient across the thermopiles, further limiting the voltage available at the TEG output [67].

To quantify how much power can be harvested, we placed a COTS TEG of 44cm$^2$ (Laird) on different parts of the human body. Fig. 7.10 shows that this TEG can harvest approximately 60μW at room temperature and 200μW at 6°C. However, the voltage available at the TEG output is only tens of mV, requiring a high conversion ratio boost converter to generate a usable supply voltage. On-body thermal harvesting is difficult because the large thermal resistance of the human body and the ambient air reduce the effective temperature gradient across the thermopiles. Even when the ambient temperature is well below the deep body temperature (i.e., 37°C), the real thermal gradient on the thermopiles is around 0.3°C [67].

7.4.2 Boost Converter and RF Kick-start

This work employs the boost converter architecture proposed in [17] to utilize the low voltages available from a body-worn TEG. Fig. 7.11(a) shows the diagram of the energy harvesting subsystem. Due to its increased efficiency at high conversion ratios and minimum usable input voltage (30mV in this work), the converter is well suited for harvesting the
TEG energy input. By modulating the two power switches in the boost converter, energy is first stored in the inductor and then transferred onto the storage capacitor. The precise timing of turning on/off the switches is critical for high conversion efficiency. The switching frequency is adjusted in a feedback loop to synchronize the PMOS turn-off with the inductor current zero-crossing, preventing current flowing backward through the PMOS switch. The measured efficiency is 38% on our chip when converting from 30mV to 1.35V. For low power levels at high supply voltages, the dynamic power dissipation caused by the switches and control logic in the designed converter becomes a significant fraction of the overall power consumption, limiting the ability to achieve high efficiency.

While the boost converter supports a 30mV power source, the internal oscillator and control logic need 600mV for startup. This requires a one-time pre-charge of the \( V_{\text{BOOST}} \) node. Previously reported start-up mechanisms use batteries and mechanical switches [17][103], requiring bulky off-chip components. Instead, we use wireless RF power for the kick-start. Incident RF power around -10dBm is provided wirelessly for 1-2 seconds and rectified through an RF rectifier front-end consisting of a 6-stage charge pump (Fig. 7.11(a)). We chose 6 stages to optimize for sensitivity instead of efficiency. In the case of kick-starting the boost converter, we charge a (storage) capacitor instead of powering a relatively constant voltage supply. Fewer stages exhibit higher efficiency, shorten the charging time, but require a larger RF input. As an alternative, self-synchronous rectifiers could be more efficient if the forward conduction loss is less than threshold voltage of the diode-connected devices [94]. However, they require fully-differential RF inputs. A shunt regulator clamps \( V_{\text{BOOST}} \) to 1.35V to prevent over-voltage during the RF powering. A low-voltage bandgap-based power-on-reset (POR) resets the chip after \( V_{\text{BOOST}} \) reaches 1V. Hysteresis in the POR trigger levels allows POR resetting only when \( V_{\text{BOOST}} \) drops below a critical voltage where the chip fails to function correctly. \( V_{\text{KILL}} \) is determined by the minimum \( V_{\text{BOOST}} \) voltage required to generate correct reference voltages and sustain conversion. Fig. 7.11(b) shows a measurement of the RF kick-start. After the voltage at the TEG output settles, a short RF burst wirelessly charges the storage capacitor at \( V_{\text{BOOST}} \). Shortly after the voltage reaches 600mV, the boost converter turns on and charges the capacitor to a regulated voltage. The node can then continue to run indefinitely from the TEG, unless \( V_{\text{BOOST}} \) drops below \( V_{\text{KILL}} \).
Figure 7.11: (a) Schematic of the energy harvesting section; (b) Measured startup sequence for the chip. A RF pulse kick-starts $V_{\text{BOOST}}$, allowing the boost converter to turn on.
due to a prolonged period of consumption exceeding the harvested energy.

7.4.3 Chip Resuscitation

In the case that $V_{\text{BOOST}}$ decreases below $V_{\text{KILL}}$, the chip will blackout and will automatically shut-off. The chip can be ‘revived’ with the same startup sequence of RF powering and harvesting from the TEG. A 90 instruction (132 byte total) subthreshold ROM can re-boot the chip and execute a default AFib detection algorithm. We chose to store an AFib detection algorithm as an indicative example of our chip performing long term illness monitoring, though objectively it is possible to miss the rare event when the chip is in shut-off mode. The chip is able to recognize its program state through a latch structure that stores this information.

7.4.4 Supply Regulation

The low voltage TEG input is boosted and subsequently regulated. Fig. 7.12(a) details the supply regulation circuits. All biases are generated on-chip. On-chip supply regulation is provided by four sub-$\mu$W linear regulators: 1.2V (AFE), 0.5V (DSP), 1.0V (TX local oscillator (LO)), and 0.5V (TX power amplifier (PA)). A programmable switched-capacitor DC-DC converter provides an output from 0.25V to 1V in 50mV steps (Fig. 7.12(b)). A 3-bit resistor DAC (RDAC) generates a reference for the desired output level based on a control word from the DPM. The arrangement of the capacitors in the array varies according to the desired output range, in a manner similar to [102]. An external capacitor (Cext) ensures that the voltage ripple due to the switching operation is minimized. The subthreshold DSP accelerators can either be connected to the 0.5V supply or variable supply (switched-capacitor DC-DC converter) through PMOS headers, enabling DVS for additional power savings.

7.5 Closed-loop power management

To prevent node blackout, the SoC must be aware of the available energy on the storage capacitance and adjust its mode of operation accordingly. If $V_{\text{BOOST}}$ is decreasing, the
Figure 7.12: a) Top level diagram of the supply regulation subsystem; b) Diagram of the switched-capacitor variable DC-DC converter.
SoC must switch modes and consume less power and energy. When harvested energy is abundant again, the chip should recover itself to a mode of full operation. The always-on DPM (digital power manager) is a custom-ISA (Instruction Set Architecture) chip controller that implements the closed-loop power management scheme. The DPM is also responsible for node control, data flow management, and overseeing all processing on the node.

The DPM monitors $V_{BOOST}$ through the ADC, where it is sampled after first being scaled to the ADCs full input range (Fig. 7.13). To do so, $V_{BOOST}$ is halved by a resistive divider, buffered to reduce the output impedance, and compared with a reference voltage. The difference is amplified (gain of 4) through a difference amplifier. Both amplifiers in Fig. 7.13 are implemented as two-stage differential-input, single-ended-output OTAs. The DPM issues an instruction that selects an input channel of a 5-input mux (the other 4 channels belong to the 4 channel AFE as explained later) that enables the ADC to digitize the scaled $V_{BOOST}$ ($V_{CAP_{DIG}7:0}$) and send it to the DPM.

Based on the $V_{CAP_{DIG}}$ value, the DPM selects the node operating mode in a spotlight fashion based on programmable digital threshold values. The DPM compares $V_{CAP_{DIG}}$ to two 8-bit threshold values (green threshold, yellow threshold) to set the DPM operating mode (green, yellow, or red). The DPM is capable of jumping from the current mode to any mode, regardless of its previous state (i.e. green to red mode). Each operating mode limits the subset of blocks that can be powered-on (Fig. 7.14(a)), capping the maximum power consumption for the mode. Normal operation is green mode (e.g. $V_{BOOST} > 1.3$V),
which allows all blocks to be on. In yellow mode (e.g. $1.1\text{V} < V_{BOOST} < 1.3\text{V}$), the DPM duty cycles the transmitter based on available energy. In red (e.g. $V_{BOOST} < 1.1\text{V}$), the transmitter, accelerator blocks, and the AFE are clock- and power-gated. Even if the node program requires one of those blocks to be turned on and functional, the stoplight power management will override this request. The DPM transitions immediately from mode to mode when the 8-bit digital $V_{CAP_{DIG}}$ value updates without requiring an additional instruction. By sampling $V_{BOOST}$ and reacting immediately, the DPM closes the loop for a sustainable power management scheme.

This power management scheme provides a flexible platform. The programmer can set the mode threshold values and the timing of instructions to sample $V_{BOOST}$ for possible mode transition. Thus, the programmer can bypass power management restrictions if transmitting or processing the data is deemed more important than keeping the node alive (i.e. the node has detected a rare EMG muscle movement event and must transmit the real-time data immediately using the power-hungry wireless transmitter, regardless of the capacitor energy). When the stoplight scheme is used in parallel with DPM branch instructions, the programmer can also control how the node reacts during a mode change. Fig. 7.14(b) displays an example scenario that uses the DPM stoplight flexibility by sending EMG data when an event is detected during green mode until the transition to yellow (at $V_{BOOST} = 0.96\text{V}$ in this example). In yellow, a subroutine samples and stores the raw input data and shuts off other blocks. When $V_{BOOST}$ rises above the green threshold, another subroutine post-processes the buffered raw data at a higher frequency using DVS before recuperating the chip back to green mode. Fig. 7.14(c) shows measured waveforms of the closed-loop DPM stoplight scheme cycling through the three modes as the boost converter input is swept from 250mV to 20mV and back.

### 7.6 Flexible biosignal datapath

#### 7.6.1 DPM as a Flexible Signal Path Controller

In addition to the DPMs power management responsibility, the DPM manages the data signal path. The DPM executes instructions from a 1.5kB instruction memory and provides
Figure 7.14: a) An example that makes use of the ‘stoplight’ scheme. Here, the green threshold is set to $V_{CAP_{DIG}} = 205$. A subroutine in the yellow mode ensures timely processing of stored data when the chip recuperates from red mode. b) Measured closed-loop power management ‘stoplight’ scheme. $V_{BOOST}$ is sampled at 256Hz. The node automatically updates its ‘stoplight’ status according to the $V_{BOOST}$ value.
a lower energy alternative to using generic MCUs for controlling the node (Fig. 7.15(a)). Fig. 7.15(c) shows the DPMs custom instruction set architecture (ISA), which is designed to facilitate node management. The DPM controls the data memory, input channels of the AFE and ADC, sampling rate, transmission rate, clock frequency creation and distribution, bus management for flexible and timing-defined data flow, time delays, and clock-gating and DVS voltage of the digital blocks, as summarized in Fig. 7.15(b).

7.6.2 Analog Front End (AFE)

The chip has four independently selectable bio-signal input channels, each with a fully-differential chopper-stabilized low-noise amplifier (LNA) and variable-gain amplifier (VGA) [147]. We chose 20 kHz as the chopper clock frequency to be significantly higher than the flicker-noise corner of the OTA for effective flicker-noise reduction. Input chopper switches are placed before the input capacitors. Compared to a topology where the switches come after the input capacitors, our arrangement reduces the amplification of any OTA offsets that might saturate its output. Any mismatch in the input capacitors results in common-mode to differential-mode gain. Since this amplifier is effectively AC-coupled, an off-chip capacitor and resistor are used to block any DC offset voltage at the electrode interface. One drawback of the topology is that the input impedance is relatively low. Periodic steady-state simulations reveal that the input impedance is a few MOs, sufficient for our ECG, EMG, and EEG applications. A programmable Gm-C filter reduces the switching ripple to below the noise floor. Along with the VGA, our amplifiers provide 7-step digitally-programmable gain (40-78dB) from DC to 320 Hz at 3μW/channel. A 5-input mux allows the sub-μW 8-bit SAR ADC to sample any of the four channels as well as the V\textsubscript{BOOST} node for monitoring stored energy.

7.6.3 Subthreshold Digital Signal Processing Subsystem

Fig. 7.16(a) shows the subthreshold DSP subsystem. To achieve ultra-low power consumption, we implement ASIC versions of the heart rate extractor (R-R), atrial fibrillation detector (AFib), and energy band extractor/envelope detector (ENV DET). An 8-bit
Figure 7.15: a) Summary of DPM energy saving when compared to our general purpose MCU; b) Diagram of the $V_{BOOST}$ monitoring path; c) Summary of the DPM ISA.
Figure 7.16: a) Block diagram for the proposed subthreshold data processing subsystem; b) Measured energy-delay curves for MCU, RR+AFib accelerator, and 30-tap, 1-channel FIR.
RISC ISA general purpose processor (GPP) MCU executes generic computations, and a re-programmable FIR performs digital filtering. A digital packetizer streams serial data to the transmitter. Two memory arrays store the program (Instruction Memory, IMEM) and bio-signal data (Data Memory, DMEM). A DMA achieves easy FIFO control and low memory latency for the DMEM. Two 8-bit switch-box busses, controlled by the DPM, configure the connections of all the processing accelerators, MCU, DMA, and packetizer. Each input/output bus port has a 4-bit address. Having two busses eases data steering and simplifies the control instructions. To support the stoplight scheme described in Section IV, each processing element has a clock-gate and two PMOS headers [29], one connected to 0.5 V, the other to the variable voltage for DVS. The clock generator block (CLK GEN) distributes a programmable clock signal (frequency and phase) to each of the processing units. The chip can process data flexibly with the MCU, use the highly efficient hardware accelerators, or cascade accelerators with MCU processing. It also can stream data on the transmitter, store and burst data, or do event based transmission. The main components work as follows:

- **MCU:** The 8-bit GPP MCU is a subthreshold RISC based on the PIC series (Microchip). The MCU is designed to run arbitrary programs and functions down to 0.26V, 1.2 kHz. Fig. 7.16(b) shows the energy-delay (E-D) curve for the MCU. The MCU consumes 0.7nW to 1.4μW measured power (0.26-0.55V) and 1.5pJ/op at the default 0.5V, 200kHz setting. The MCU shares the IMEM with the DPM. Fig. 7.17 summarizes the organization of memory, MCU, and DPM, and their capabilities. A multiplexer steers each instruction to either the MCU or DPM (INSTSTEER) based on a special code word. When the MCU is executing instructions, the DPM automatically goes into a low power sleep mode. When the DPM is executing instructions, the MCU is either turned off or clock gated to save state. In this way, we retain the energy efficiency of the DPM as a chip controller and the generic flexibility of the MCU without requiring extra instruction memory space. The MCUs instructions are programmed at the same time as the DPM during the chips pre-deployment.

- **Instruction and Data Memories:** Because the chip operates in subthreshold in an
Figure 7.17: Integration of DPM and GPP MCU. DPM and MCU share the same memory, and instructions are steered to the appropriate processor. The MCU is power-gated when idle. Both efficient control of the node and generic processing are achieved.

N-strong technology, the SRAMs use an 8T bitcell and the zero leakage read-buffer from [127]. To eliminate half-select instability during a write, both reads and writes apply to full rows of memory. The DMEM is split into four 1kB banks that can be individually power-gated by NMOS footers being overdriven to 1.2V when active to ensure low levels of ground bounce. Measured results show reliable operation down to 0.3V at 200 kHz with IMEM read energy of 12.1pJ per read at 0.5V and leakage energy per cycle at 200 kHz of 6.6pJ.

- DMA: The DMA is an efficient subthreshold accelerator to interface between the DMEM and the rest of the SoC. It is programmed by one instruction of the DPM and effectively treats the DMEM as a FIFO to support efficient streaming. A clock multiplexor synchronizes the DMA clock rate to the component it interfaces to by choosing between several clock rates. A memory controller uses separate DMEM banks during green and yellow modes for easier data management. To solve the half-select stability issue during writes, we use a row buffer and only write a row when all words are ready. When the difference between the write pointer address and read pointer address is greater than or equal to 4 bytes, a DMA flag is raised, which signifies
to the DPM that there is a full packet of data. This simple and efficient mechanism of interrupting for transmission limits overflows.

- **Programmable FIR Filter**: A four-channel, programmable, max-30 tap, and synthesizable filter was designed to enable operation in the subthreshold regime down to 300mV (measured). The FIR filter is capable of filtering all four-channels of the AFE. The programmable options include coefficient selection, number of taps, and number of filters. When power is critical and data fidelity can be compromised, a half-taps mode allows for a 15-tap filter. The direct-form implementation of an FIR requires as many adders and multipliers as there are taps, costing area and leakage. Due to the small sampling rate for ExG signals, each result can instead be computed serially over multiple faster clock cycles using only one multiplier and one adder. This architecture results in a 30x reduction in area per channel and a measured 1.1pJ per tap at 350mV. The architecture saves valuable chip area and reduces leakage current. For further power reduction, each individual channel can be clock-gated. A measured energy-delay curve is given in Fig. 7.16(b).

- **Envelope Detector**: For EEG signals, knowing the signal power within a specific frequency band is useful for determining neural activity in the $\alpha$, $\beta$, low-$\gamma$ and $\gamma$ frequency bands [147]. The ENV DET circuit computes the average signal power within a specified frequency band. This block receives data directly from the FIR filter and has four input channels corresponding to the channel outputs of the filter. Eqn. 7.1 computes the average signal power $p_x$ of input signal $x$,

\[
p_x = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2
\]

(7.1)

where $N$ is the summing window size. To reduce the computation complexity, $N$ is set to powers of two ($2^7 - 2^7$), which allows for division to be implemented as a simple right shifting of the data. Further, $x$ values are rounded to the nearest power of four and the square results come from a lookup table. The rounding reduces the number
of bits required during data transformation as the lower two bits are always 0. The ENV DET consumes 3.5mW (measured) at 0.5V and 200 kHz.

- R-R Extraction: The heart-rate extractor accelerator is a simple version of the popular Pan-Tomkins algorithm [96]. This R-R algorithm calculates the heart rate by means of time windowing and thresholding, after an initial 4 second time frame where the R-R accelerator gains a baseline DC value for the heart waveform. The time stamp given to two consecutive peaks is the difference in the number of samples between them. For this reason, we can achieve a desired accuracy by changing the sampling rate and using the R-R accelerator in a DVS fashion to accommodate the faster or slower processing rate needed. Once an R-R time has been calculated, a pulse is output, which signifies to the AFib accelerator a new R-R sample is ready.

- AFib Accelerator: The atrial fibrillation detector is an ASIC accelerator that detects the arrhythmia using an implementation of the clinically validated algorithm described in [63]. It receives its inputs from the R-R accelerator and outputs an AFib_flag signal to the DPM signifying the detection of atrial fibrillation. The algorithm uses only 12 R-R intervals. Many variables in the algorithm, such as the margin of error, are programmable. The algorithm uses a pattern recognition scheme that quantifies the entropy in these 12 R-R intervals. If the entropy is more than a programmable threshold, then an AFib event is reported.

7.6.4 Low Power RF Transmitter

To allow operation from harvested power, the peak current consumption of the chip must be minimized. We utilize a frequency-multiplying transmitter architecture to reduce the synthesizer power by operating the LO at 1/9 the carrier frequency [97]. We use equally spaced edges generated from the cascaded ring oscillators to drive the edge-combiner (EC) embedded PA to perform the frequency multiplication.

The use of frequency multiplication allows harmonic injection-locking from the crystal oscillator. Instead of using a PLL, injection-locking a low-frequency ring oscillator to an
on-chip crystal reference eliminates the longer settling times, therefore allowing aggressive duty-cycling of the transmitter to further save power. Directly injection-locking the multi-phase ring oscillator using a single-phase reference introduces significant mismatch. We used cascaded multi-phase injection-locking to correct the phase and amplitude mismatches.

On-chip BFSK modulation is accomplished by pulling the quartz reference clock. By modulating the load capacitor, we can pull the crystal frequency by 200 ppm. After 9× multiplication, the resulting frequency deviation is approximately 100 kHz, achieving >100kbps datarate. The TX circuit consumes 160μW when transmitting at its maximum data rate of 200kbps [97]. In bio-signal raw data mode, the transmitter operates at a 100% duty-cycle, while the R-R extraction mode, explained in Section VI, reduces the duty cycle and average transmitter power consumption to 0.013% and 190 nW respectively. The packetizer contains a programmable packet header and CRC to allow compatibility with commercial receivers.

7.7 System measurement

An ECG experiment was performed on a healthy human subject. In our in-vivo experiments, we used self-adhesive surface electrodes (Kendal Meditrace 535) to acquire ECG signals on the order of a few mVs. One electrode is attached to the chest, while the second (reference) electrode is attached to the abdomen. The two electrode leads are twisted before interfacing with the SoC to minimize interference. First, the chip was set to ECG raw data mode (consuming 397μW from the 1.35V $V_{\text{BOOST}}$ node) (Fig. 7.18). Our chip was paired to an unmodified TI CC1101 receiver and a wireless link was successfully established in the 433 MHz ISM band. The reconstructed ECG (dashed) closely matched the actual ECG. Also shown is the transmitted data, received data and clock waveforms. The zoomed-in section shows one 44b packet of data, including 9b header, 32b data, and 3b CRC.

Next, the chip ran an R-R interval extraction algorithm on the MCU and transmitted measured heart-rate every 5s operating from a 30mV supply voltage (Fig. 7.19(a)). Every 5s, $V_{\text{BOOST}}$ is sampled to check for sufficient available energy, in which case the crystal oscillator is enabled for 20ms before the TX transmission, which takes 650s including turn-on time and transmission of a 24-bit packet. The heart-rate extractor algorithm measures the R-R interval with a time resolution of $(1/128)$s (Fig. 7.19(b)). The extremely low duty-
Figure 7.18: Measured waveforms showing correct data reconstruction from wirelessly streamed data.

cycle of TX and crystal oscillator dramatically reduces the power consumption of the TX and crystal oscillator to negligible amounts.

In AFib detection mode, the R-R and AFib accelerators enable the TX and transmit the last 8 beats of raw ECG (buffered in the DMEM) only when a rare AFib event occurs. Measurement results for the AFib demo are presented in Fig. 7.19(b). A pre-recorded set of AFib data from MIT-BIH database is used for this demo (www.mit.edu). Detection occurs 12 R-R intervals after the inception of an AFib event. A pattern recognition algorithm determines if an AFib has occurred [63]. The total chip power in both the R-R and AFib modes is 19µW, and the chip is powered exclusively from a 30mV harvested input. Fig. 7.20(a) presents a current breakdown of the R-R extraction demo. The current is nearly evenly distributed among different components, and selective transmission significantly reduces the average power consumption of the transmitter. Fig. 7.20(b) shows the micrograph of the 2.5mmx3.3mm batteryless BAN SoC (130nm CMOS), and Fig. 7.21(a) gives a performance summary.

Fig. 7.21(b) shows a performance comparison table with recent BAN SoCs. This work is
Figure 7.19: a) Measured system response in the R-R extraction mode, where the R-R interval information is wireless transmitted every 5s. Also shown is the measured accuracy of the results. b) Measured system AFib demo experiment using R-R extractor and AFib accelerator. Normal and AFib ECG waveforms are from MIT-BIH database. The last 8 beats of raw ECG are stored in DMEM and streamed over TX if AFib is detected.
Figure 7.20: a) Current breakdown for R-R extraction mode; b) chip die photo.
### Energy Harvesting

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{in}$</td>
<td>30 mV</td>
</tr>
<tr>
<td>$V_{reg}$/V_{in}</td>
<td>&gt;1.25 V</td>
</tr>
<tr>
<td>Kick-Start</td>
<td>RF @ 10 dBm</td>
</tr>
<tr>
<td>$V_{clam}$</td>
<td>1.35 V</td>
</tr>
<tr>
<td>Efficiency</td>
<td>38%</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>2.9 V</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>&lt;4 V</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>0.3-0.5 V</td>
</tr>
</tbody>
</table>

### System Power

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>14 µA (R-R) (0.013% TX duty cycle)</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>0-320 Hz</td>
</tr>
<tr>
<td>CMRR</td>
<td>&gt;70 dB</td>
</tr>
<tr>
<td>Supply</td>
<td>1.35 V</td>
</tr>
</tbody>
</table>

**RR FOM**: Power (nW) / frequency (MHz) / # of taps / input bit length / coefficient bit length

### Table: This Work vs. Other Works

<table>
<thead>
<tr>
<th>Parameter</th>
<th>This Work</th>
<th>[8]</th>
<th>[24]</th>
<th>[9]</th>
<th>[24]</th>
<th>[7]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensors</td>
<td>ECG, EMG, EEG</td>
<td>ECG</td>
<td>Neural ECG, EMG, EEG</td>
<td>EEG, TIV</td>
<td>EEG, TIV, TIV Pressure</td>
<td></td>
</tr>
<tr>
<td>Supply Voltage</td>
<td>300 mV, –10 dBm</td>
<td>1.2 V</td>
<td>1 V</td>
<td>1 V</td>
<td>1.2 V</td>
<td>0.4 W, 0.5 V</td>
</tr>
<tr>
<td>E Harvesting</td>
<td>Thermal, RF</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Solar</td>
</tr>
<tr>
<td>AFE</td>
<td>4-channel</td>
<td>5-channel</td>
<td>3-channel</td>
<td>1-channel</td>
<td>18-channel</td>
<td>4-channel</td>
</tr>
<tr>
<td>Power Mgmt.</td>
<td>DPM, Clock gating, Power gating</td>
<td>Clock gating</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Power gating</td>
</tr>
<tr>
<td>GPP MCU</td>
<td>1.5 pJ/Instr. @ 200 kHz (5-GHz RISC ISA)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>25.8 pJ/Instr. @ 73 MHz (32-bit CORTEX-M3)</td>
</tr>
<tr>
<td>Accelerators</td>
<td>Programmable</td>
<td>FIR, AFE, MCU, §V DET, DMA, Packetizer</td>
<td>FIR, Encryption, DMA</td>
<td>ASOC DSP</td>
<td>FIR, Packetizer, Compression</td>
<td>x</td>
</tr>
<tr>
<td>Memory</td>
<td>5.5 kB (0.3%–0.7%)</td>
<td>43 kB (1.2%)</td>
<td>x</td>
<td>x</td>
<td>20 kB (1.2%)</td>
<td>5 kB (0.4%)</td>
</tr>
<tr>
<td>DVS</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Digital Power</td>
<td>2.1 µW</td>
<td>–12 µW</td>
<td>N/A</td>
<td>2.1 µW</td>
<td>500 µW</td>
<td>2.1 µW (MCU)</td>
</tr>
<tr>
<td>TX (at T=0°C)</td>
<td>200 kHz</td>
<td>x</td>
<td>100 kV</td>
<td>x</td>
<td>1 Mbps (on-body link)</td>
<td>x</td>
</tr>
<tr>
<td>TX Power (100% on)</td>
<td>160 µW</td>
<td>x</td>
<td>400 µW</td>
<td>x</td>
<td>2.5 mW</td>
<td>x</td>
</tr>
<tr>
<td>TX Power (on)</td>
<td>–18.5 dBm</td>
<td>x</td>
<td>–18 dBm</td>
<td>x</td>
<td>–6 dBm</td>
<td>x</td>
</tr>
<tr>
<td>TX Band</td>
<td>400/43 MHz</td>
<td>x</td>
<td>400/43 MHz</td>
<td>x</td>
<td>20.40 MHz</td>
<td>x</td>
</tr>
<tr>
<td>Total Chip Power</td>
<td>19 µW</td>
<td>31.1 µW</td>
<td>500 µW</td>
<td>77.1 µW</td>
<td>2.4 mW</td>
<td>7.7 µW</td>
</tr>
<tr>
<td>Note on Total Power</td>
<td>1-channel AFE, 8-bit ADC, DSP (R-Reduction), and TX duty-cycled at 0.05%</td>
<td>AFE, 12-bit ADC, and streaming TX with 100% duty cycle</td>
<td>1-channel AFE, 8-bit ADC, and streaming TX with 100% duty cycle</td>
<td>1-channel AFE, 12-bit ADC, and DSP (EFG feature extraction)</td>
<td>4-channel AFE, 10-bit ADC, DSP (data compression, FIR), SRAM, TX at 5% duty cycle</td>
<td>4-channel AFE, 10-bit ADC, DSP (data compression, FIR), SRAM, TX at 5% duty cycle</td>
</tr>
</tbody>
</table>

Figure 7.21: a) Measured performance summary; b) Performance comparison with state-of-the-art BAN nodes.
the first wireless bio-signature processing chip enabling battery-free operation. The chip can be powered from an input as small as 30mV, enabling thermal energy harvesting. For on-chip power management we augment power and clock-gating by using the DPM, a custom chip controller, to intelligently handle energy consumption based on the available energy. The closed-loop power management 'stoplight scheme enables potentially indefinite operation while the node is worn. Our 5.5kB of on-chip memory remains operational down to a subthreshold 0.3V and is compatible with the flexible subthreshold datapath. In the heartbeat extraction mode where the transmitter is duty-cycled, the entire chip, including VDD regulation, only consumes 19μW. To the best of the authors' knowledge, this system has lower power, lower minimum input supply voltage, and more complete system integration than other reported wireless BAN SoCs to date.
Chapter 8

CONTRIBUTIONS

Recent advances in ultra-low power chip design techniques, many originally targeting wireless sensor networks, will enable a new generation of body-worn devices for health monitoring. Off-the-shelf chips often consume too much power and are usually too bulky for body-worn applications. In this thesis, I’ve presented ULP analog and RF circuit design techniques as well as system-level innovations that enable the next-generation wireless sensors. While integrating more functionalities on-chip, we focus on reducing the power consumption to extend the battery life of the BSNs.

Recent development in IC design allows for battery-free operation scenarios using energy harvesting. Limited conversion efficiency of boost converter motivates reducing supply voltages to operate circuits directly off the harvested voltage. Reduced supply voltage is also a projected trend as a result of technology scaling. Although most digital functionalities scale well with reduced supply voltage, analog/RF designs become increasingly challenging. We used a transformer-coupled technique to maximize the voltage headroom to allow 2.4 GHz operation from a 300 mV supply.

Fig. 8.1 shows the timeline of the projects I’ve worked on during my Ph.D. program:

1. “Neurec V1” and “Neurec V2”: A 500μW fully integrated neural interface that wirelessly streams a digitized neural waveform over 15m. Later, we made a self-contained miniaturized wireless sensing platform housing “Neurec V2” that continuously streams biosignal waveforms for as long as 3 days on a 33 mAh coin-cell battery. This work was originally presented in ISSCC 2009, then ISSCC DAC 2010. The design techniques used in the low-noise bioamplifier on the chip are reorganized into a journal paper published in TBioCAS 2012.

2. “SoCWISP”: A fully-passive 900MHz RFID tag IC with addressability, full EPC Class
Figure 8.1: Research summary showing the die photos of the chips that I taped out during my Ph.D. program.
1. Generation 2 (Gen 2) protocol compatibility, a $1.25 \mu V_{rms}$ integrated noise chopper-stabilized micropower sensor interface amplifier, and an 8b ADC. This work was originally presented in ISSCC 2012, then reorganized into a journal published at JSSC 2010.

3. “ECoG”: A 6.4$\mu W$ ECoG/EEG processing IC with $0.4 \mu V_{rms}$ noise floor intended for emerging BCI applications. This chip conditions the signal and simultaneously extracts energy in four fully-programmable frequency bands. This work was originally presented in CICC 2010, then expanded into a journal published at TCAS I 2011.

4. “ILTX”: A tag architecture that employs an active transmitter to avoid the “self-jamming” problem present in RFID systems, reducing the complexity of reader design. This scheme allows remote placement of the receiver and extends data transmission range. This work was presented at RFIC 2011.

5. “FMOS V1”, “FMOS V2” and “FMOS V3”: A method for integrating circuitry into the lid wafer to form a sub-0.1mm$^3$, sub-mW, 1.5 GHz temperature-compensated chip-scale oscillator. The first version of the oscillator design was presented at FCS 2011. The second version focuses on low-jitter performance, and was presented at FCS 2012. A third version improving the phase noise and output power of the second version was taped out summer 2012.

6. “BASN”: We utilize the state-of-the-art in low power RF transmitters, low voltage boost circuits, subthreshold processing, biosignal front-ends, dynamic power management, and energy harvesting to realize an integrated reconfigurable wireless body-area network (BAN) SoC capable of autonomous power management for battery-free operation. This work was originally presented at ISSCC 2012, then expanded into a journal published at JSSCC 2012.

7. “LVRX V1” and “LVRX V2”: A 2.4 GHz receiver that operates from a supply voltage of only 300 mV, the lowest report to date. It allows direct powering from various energy harvesting sources. This work will be presented at ISSCC 2013.
I have contributed in advancing the state-of-the-art ULP circuit design techniques, including low-noise bioamplifiers for ExG (e.g. ECG, EMG, EEG, etc.) and neural applications, analog signal processors targeted for but not limited to ECoG applications, transmitter and receiver for short-range wireless. In addition, I have also helped realizing the next-generation wireless sensing system ICs that aim at prolonging battery life, even eliminating the battery leveraging emerging energy-harvesting solutions. Lastly, a few of the works presented in this thesis (such as the “Bumblebee”, “SoCWISP”) have not only been published as literatures, but also deployed at various labs assisting scientists to perform experiments that are difficult if not impossible to be done otherwise. Beyond academia, the “FMOS” work is even making splashes in the industry while getting commercialized.
BIBLIOGRAPHY


