Area- and Energy- Efficient Modular Circuit Architecture For 1,024-Channel Parallel Neural Recording Microsystem

by

Sung-Yun Park

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in the University of Michigan
2016

Doctoral Committee:

Professor Euisik Yoon, Chair
Associate Professor Joshua D. Berke
Professor Michael P. Flynn
Associate Professor David D. Wentzloff
Professor Emeritus Kensall D. Wise

DEDICATION

To my beloved family for all their support and patience during my years of study

ACKNOWLEDGMENTS

I would first like to thank my research advisor, Prof. Euisik Yoon, for all his support and guidance during my Ph.D. study at University of Michigan, Ann Arbor. This work would not have been possible without his insightful advice. I would also like to express my appreciation to my doctoral committee members, Prof. Ken Wise, Michael Flynn, David Wentzloff and Joshua Berke, for their participation, invaluable advice, and comments for this research work.

I would also like to thank my former and current co-workers in Yoon Lab: Seokjun, Sun-il, Jaehyuk, Khaled, Jihyun, Kyounghwan, Kyuseok, Xia, Fan, Patrick, Yu-Chih, Adam, Yu-Ju, Komal, Yu-Heng, Zhixiong and Buke for their helpful suggestions, discussions, and technical support.

Finally, my deepest gratitude goes to my parents, for all the sacrifices they made to make my life better. I would also like to express my sincere gratitude to my beloved family, Boram, John and Emily for all their love and understanding and support throughout the years.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGMENTS iii

LIST OF FIGURES ix

LIST OF TABLES xvi

ABSTRACT xvii

Chapter 1 Introduction 1

1.1 Motivation 1

1.2 Neural recording microsystem 5

1.2.1 Overview 5

1.2.2 Area- and energy- intensive circuit and architecture design 11

1.2.3 Summary 13

1.3 Thesis outline 15

1.3.1 Chapter 2: Characteristics of the neural signal 15

1.3.2 Chapter 3: 128-channel modular AFE architecture 15

1.3.3 Chapter 4: On-chip neural signal compressor 15

1.3.4 Chapter 5: PWM buck converter using analog-digital hybrid control 15

1.3.5 Chapter 6: Summary and future works 16

Chapter 2 Physical and electrical properties of cells in the brain 17

2.1 Structure of neurons[1] 17

2.2 Transmission of Information by Neurons[1] 19

2.3 LFP and AP in neurons 20

2.4 Electrical characteristics of LFP and AP 21

Chapter 3 128-channel modular Δ-ΔΣ analog front-end architecture 23

3.1 Introduction 23

3.2 Circuit architecture 26

3.2.1 Modular Δ-ΔΣ AFE architecture 26

3.3 Circuit implementation 27

3.3.1 LNA design 28

3.3.2 Programmable gain amplifier design 32

3.3.3 Δ-ΔΣ analog to digital converter 33

3.3.4 Digital filter and serializer 37

3.4 Experiment results 39

3.4.1 Measurement of LNA, PGA, and ADC 39

3.4.2 *In-vivo* measurement 42

3.4.3 Comparison metric 44

3.5 Summary and chapter conclusion 46

Chapter 4 On-Chip neural signal compressor with 128-channel analog front end 48

4.1 Introduction 48

4.2 Investigation of broadband neural signal 49

4.2.1 Revisiting to neural signal characteristics 49

4.2.2 Recent works for compression of LFP 55

4.2.3 Recent works for compression of AP 55

4.3 Circuit architecture 56

4.3.1 On-chip neural signal neural compressor with 128-channel Δ-ΔΣ AFE 56

4.4 Circuit implementation 60

4.4.1 Separation of broadband neural signals 61

4.4.2 DT-Δ-ΔΣ ADC 64

4.4.3 Reconfigurable spike waveform extractor 66

4.4.4 Digital controller and interface 69

4.5 Experiment results 70

4.5.1 LNA, HPF, LFP and ADC Measurement 71

4.5.2 *In-vivo* measurement 79

4.5.3 Summary and comparison 83

Chapter 5 PWM buck converter using analog-digital hybrid control 85

5.1 Introduction 85

5.1.1 General requirement 85

5.1.2 Switched capacitor and inductive dc-to-dc converters 87

5.1.3 PWM and PFM control 87

5.1.4 Motivation 90

5.2 Load-adaptive power transistor scaling 90

5.2.1 Analysis of power loss in a PWM buck converter 90

5.2.2 Switching power loss reduction in a PWM buck converter 93

5.3 Circuit architecture 96

5.4 Circuit implementation and operation 97

5.4.1 Power transistor design 97

5.4.2 Coarse digital controller 98

5.4.3 LCO in coarse digital controller 103

5.4.4 Mode arbiter 106

5.4.5 Fine analog controller 107

5.5 Experiment results 110

5.5.1 Transient response 111

5.5.2 Power consumption and PCE 113

5.5.1 Comparison 115

5.6 Summary and chapter conclusion 116

Chapter 6 Summary and future work 117

6.1 Summary 117

6.2 Future works 118

BIBLIOGRAPHY 120

LIST OF FIGURES

Figure 1-1 Architecture of neural recording microsystem (the optional blocks are grayed). 3

Figure 1-2 (a) Conceptual platform for a 1,024-channel parallel recording microsystem;(b) photograph of a fabricated single module with a 128-channel multi-shank probe, interposer, Si-cap, and 128-channel AFE chip; and (c) cross-sectional view of the interconnection between the probe and the AFE chip..... 4

Figure 1-3 Silicon microelectrode: (a) microphotograph and (b) conceptual schematic 6

Figure 1-4 10-channel active probe: (a) microphotograph and (b) conceptual drawing . 8

Figure 1-5 3-D microelectrode array: (a) microphotograph and (b) conceptual drawing 9

Figure 1-6 Utah Probe (a) photograph and (b) scaled drawing of an electrode side view (All values are in µm)..... 9

Figure 1-7 Microphotographs of IMEC's: (a) 455-channel monolithic active probe (54 parallel recordings) and (b) 966-channel monolithic active probe (384 parallel recordings)..... 11

Figure 1-8 0.013-mm^2-per-channel neural recording circuit: (a) block diagram and (b) microphotograph of the fabricated prototype..... 12

Figure 1-9 100-channel neural recording chip with sub-µW-per-channel power consumption: (a) block diagram and (b) microphotograph of the fabricated prototype..... 12

Figure 2-1 Structure of a neuron 18
Figure 2-2 An example of parallel recording of (a) broadband neural signals (b) filtered traces of (a) 21
Figure 3-1 Typical multi-channel neural recording AFE 23
Figure 3-2 Top level architecture of 128-channel neural recording Δ-ΔΣ AFE 27
Figure 3-3 On-chip signal compression in Δ-ΔΣ AFE and off-chip reconstruction 28
Figure 3-4 Schematic of single channel Δ-ΔΣ AFE 28
Figure 3-5 Schematic of Gm1 used for LNA 31
Figure 3-6 Schematic of PGA (a) and transconductor (G_{m2}) used for PGA (b) 33
Figure 3-7 Block diagram of CT Δ-ΔΣ ADC (a) and Modified quantizer inside the CT Δ-ΔΣ ADC (b) 35
Figure 3-8 SQNR of first and second order ΔΣ, and Δ-ΔΣ ADC with different OSR (a) SQNR of second order ΔΣ modulator and Δ-ΔΣ ADC with different bandwidth signals (b) 36
Figure 3-9 (a) Schematics of the first transconductor (G_{m3}) in CT Δ-ΔΣ ADC and (b) dynamic comparator used as 1b quantizer for Δ-ΔΣ ADC. The CMFB circuit for G_{m3} is not shown in (b). 37
Figure 3-10 Frequency response of the $sinc^2$ filter for the decimation of the output of the Δ-ΔΣ ADC 38
Figure 3-11 Die microphotograph of 128-channel AFE including test circuits. A single channel is enlarged in the right side 39
Figure 3-12 Measured gain and phase response of LNA 40
Figure 3-13 Measured input referred noise spectral density of LNA 41

Figure 3-14 Measured total harmonic distortion of LNA 41
Figure 3-15 Measured frequency response of PGA 42
Figure 3-16 32,768-point FFT of the output of Δ–ΔΣ ADC 42
Figure 3-17 0.6 second snapshot of In-vivo measurement for LFP and AP from a rodent 43
Figure 3-18 *In-vivo* data from multi-channel recordings: (a) broadband neural signal including LFP and spikes and (b) bandpass filtered spikes in the post-processing 44
Figure 3-19 Power spectra of the recorded and reconstructed neural signals from *in-vivo* measurement 44
Figure 3-20 E-A FoM with other state-of-the-art works 46
Figure 4-1 (a) Cross-correlation of the LFP from CA1 region of a rat's brain and (b) probe configuration used for the neural recordings 51
Figure 4-2 *CR* of spatial and temporal (ST) difference and temporal (T) of the LFP ... 52
Figure 4-3 *CR* of the LFP and AP through numerical calculation by applying (1) spatiotemporal correlation (LFP) and (2) temporal correlation (LFP) (3) spatiotemporal correlation (LPF+ AP) 53
Figure 4-4 Occurrence of AP from 32-channel, 2 second neural signal 54
Figure 4-5 Architecture of the on-chip neural signal compressor with a 128-channel Δ-ΔΣ AFE 58
Figure 4-6 Two ways to take spatial difference of the LFP (a) difference between adjacent channels and (b) difference between reference and others 60
Figure 4-7 Single channel signal acquisition blocks 61

Figure 4-8 Schematic of the SC-biquad filter. V_{CM} is the common mode voltage and φ_1 and φ_2 are non-overlapping clocks 63
Figure 4-9 The OTA used in the SC-biquad low-pass filter (a) and the CMFB circuit for the OTA 63
Figure 4-10 (a) Z-domain block diagram of the DT Δ-ΔΣ ADC (b) and the conceptual quantizer with reduced quantization level 64
Figure 4-11 Schematic of the DT Δ-ΔΣ ADC 65
Figure 4-12 Capacitor bank configuration for both of ADC and A-FIFO 67
Figure 4-13 Conceptual SAR/SS ADC with details of a dynamic comparator 68
Figure 4-14 2 modes of operation of the reconfigurable SAR/SS ADC: (a) SAR operation, (b) SS operation with A-FIFO 69
Figure 4-15 Die microphotograph of on-chip neural signal compressor with 128-channel AFE including test circuits 71
Figure 4-16 Measured gain and phase response of LNA 72
Figure 4-17 Measured common mode and differential gain of LNA 72
Figure 4-18 Measured input referred noise spectral density of LNA 74
Figure 4-19 Measured total harmonic distortion of LNA 74
Figure 4-20 Measured frequency response of PGA: gain adjustment 75
Figure 4-21 Measured frequency response of PGA: corner frequency adjustment 75
Figure 4-22 Measured frequency response of LPF: Gain adjustment 76
Figure 4-23 Measured frequency response of LPF: Corner frequency adjustment 76
Figure 4-24 Full-scale input vs. SNDR measurement of SAR ADC 77
Figure 4-25 DNL and INL measurement of Δ–ΔΣ ADC 77

Figure 4-26 32,768-point FFT of the output of DT Δ–ΔΣ ADC 78
Figure 4-27 Neural signal separation from the prerecorded broadband neural signal... 79
Figure 4-28 Two modes of operation for AP; normal (SAR) and compression (SS) mode, (a) both SAR and SS output of1 second AP and (b) Snapshot of 8 ms of (a) 79
Figure 4-29 Single channel in-vivo measurement of the LFP and AP 80
Figure 4-30 Measurement of the broadband neural recording in normal mode: (a) LFP (b) AP 81
Figure 4-31 (a) Measurement of the spatial difference of the local field potentials (LFP) in the compression mode and (b) Retrieved LFP from software (MATLAB) and (c) waveforms of AP..... 81
Figure 4-32 (a) Measurement of the compressed AP and (b) the waveforms of the excerpted AP 82
Figure 4-33 Digital interface power consumption (data handling and transmission) for LFP and cross-correlation between the channels 83
Figure 4-34 Digital interface power consumption (data handling and transmission) for AP and the average firing rate of the AP..... 83
Figure 5-1 (a) peak power consumption and (b) daily energy consumption of an implantable neural recording system..... 89
Figure 5-2 Four different power losses in a PWM buck converter; conduction (AC/DC) loss, overlap (IV) loss and switching (SW) loss, and the corresponding PCE of the converter..... 93

Figure 5-3 Four different power losses in a PWM buck converter: AC/DC loss, OV loss, and SW and the corresponding PCE of the converter while changing the size of the power transistor according to the loads..... 94

Figure 5-4 Duty cycles for a DCM Buck operation of two cases: M = 0.7 and M = 0.3. (5-2) is overlaid on the graph for M = 0.7 96

Figure 5-5 Architecture of the PWM buck converter with the hybrid controller 97

Figure 5-6 Block diagram of the coarse digital controller. The part of the power transistor array is also shown to aid the understanding..... 99

Figure 5-7 V/I converters inside IL Emulator (a) and Squared symbols are the generated duties (D) in two different VCR (M = 0.3 and 0.4) and the dotted lines are the necessary duties for each case (b) 101

Figure 5-8 (a) Operation of the proposed coarse digital control and (b) corresponding waveforms of V_P and V_N and I_L from the operation of (a)..... 103

Figure 5-9 (a) Single tap from the variable delay DPWM where the tap is assumed to be selected by the compensator. (b) Schematic of the single delay cell (D_F or D_R) 103

Figure 5-10 Functional block diagram describing the coarse digital control. The converter and the compensator actuates as Δ and Σ function, respectively 105

Figure 5-11 (a) A typical waveform of LCO. (b) 1000 cases of $|A_{LCO}|$ and T_{LCO} derived from the converter 106

Figure 5-12 Schematic of the mode arbiter 107

Figure 5-13 Conceptual operation of the mode arbiter (a) operation sequence (A to D) overlaid on the typical waveform of LCO (b) phase diagram of (a)..... 107

Figure 5-14 Block diagram for the analog control 109

Figure 5-15 (a) Operation of the proposed fine analog control. (b) Corresponding waveforms of V_P, V_N, and I_L from the operation of (a) 109

Figure 5-16 (a) A half circuit of the delay cell inside the last tap of the variable delay DPWM and (b) variation of delay (DF) by the additional I_E. 110

Figure 5-17 Chip microphotograph of the proposed PWM Buck converter 111

Figure 5-18 Step current response of the proposed PWM buck converter. I_{OUT} is changed from 50 µA to 4 mA. Measured at V_{OUT} = 1.0 V from V_{IN} = 3.3 V with f_{SW} = 0.96 MHz..... 112

Figure 5-19 Steady-state response of V_{OUT} and V_X. The converter is providing I_{OUT} = 900 µA and V_{OUT} = 1 V from VIN = 3.3 V 112

Figure 5-20 Power dissipation of analog and digital blocks..... 114

Figure 5-21 Measured power conversion efficiency at V_{OUT} = 1V from V_{IN} = 3.3V and f_{SW} = 0.96 MHz..... 114

LIST OF TABLES

Table 1-1 Performance summary of recent neural recording microsystems 14

Table 3-1 Performance comparison with the state-of-the-art AFEs 47

Table 4-1 Performance comparison with the state-of-the-art works 84

Table 5-1 Performance comparison with the state-of-the-art dc-to-dc converters ... 115

ABSTRACT

Area- and Energy- Efficient Modular Circuit Architecture for 1,024-Channel Parallel Neural Recording Microsystems

By

Sung-Yun Park

Chair: Euisik Yoon

In the past few decades, there has been significant progress in the development of neural recording systems with fast advance in electronics and micro-mechanical-systems (MEMS) technologies. The number of simultaneous recorded neurons has been doubled at every 7.4 years, following a trend similar to Moore's law. While various neural interface systems such as neural prosthetics can take the advantage of such rapid growth of the technologies, it is particularly beneficial to the comprehensive neuroscience research since high quality and parallel monitoring over a large number of neurons within small volume of brain can solely provide in-depth understanding brain's structure and activities.

This research focuses to develop system architectures and associated electronic circuits for a next generation neuroscience research tool, *i.e.* a massive-parallel neural recording

system capable of recording 1,024 channels simultaneously. Three interdependent prototypes have been developed to address major challenges in realization of the massive-parallel neural recording microsystems: minimization of energy and area consumption while preserving high quality in electrical recordings.

First, a modular 128-channel Δ-ΔΣ analog front-end (AFE) using the spectrum shaping technique has been designed and fabricated to propose an area-and energy efficient solution for neural recording AFEs. The fabricated AFE is also expandable to a 1,024-channel parallel recording system via hybrid assembly with the customized 3-dimension platform consisting of multi-shank probes, interposers, and silicon-encapsulation. The AFE achieved 4.84 fJ/C−s·mm^2 figure of merit that is the smallest the area-energy product among the state-of-the-art multichannel neural recording systems. It also features power and area consumption of 3.05 μW and 0.05 mm^2 per channel, respectively while exhibiting 63.3 dB signal-to-noise ratio with 3.02 μV_{rms} input referred noise.

Second, an on-chip mixed signal neural signal compressor was built to reduce the energy consumption in handling and transmission of the recorded data since this occupies a large portion of the total energy consumption as the number of parallel recording increases. The compressor reduces the data rates of two distinct groups of neural signals that are essential for neuroscience research: local field potentials (LFP) and action potentials (AP) without loss of informative signals. As a result, the power consumptions for the data handling and transmissions of the LFP and AP were reduced to about 1/5.35 and 1/10.54 of the uncompressed cases, respectively. In the total data handling and transmission, the measured power consumption per channel is 11.98 μW that is about 1/9 of 107.5 μW without the compression.

Third, a compact on-chip dc-to-dc converter with constant 1 MHz switching frequency has been developed to provide reliable power supplies and enhance energy delivery efficiency to the massive-parallel neural recording systems. The dc-to-dc converter has only predictable tones at the output and it exhibits > 80% power conversion efficiency at ultra-light loads, < 100 μW that is relevant power most of the multi-channel neural recording systems consume. The dc-to-dc converter occupies 0.375 mm^2 of area which is less than 1/20 of the area the first prototype consumes (8.64 mm^2).

CHAPTER 1
INTRODUCTION

1.1 Motivation

Since the first successful recording of *in-vitro* action potential by Adrian in the 1920s [1], there has been significant progress in the development of neural recording systems, especially in the last few decades. A journal article published in 2011 investigated 56 independent studies from 1950 to 2010 and reported that the number of simultaneous recorded neurons has roughly doubled every 7.4 years while following a trend similar to Moore's Law [2]. Indeed, current state-of-the-art neural recording systems have the capacity of approximately a few hundred high quality recording channels. [3][4][5][6].

This rapid growth is advantageous to most neural interface systems, such as neuroprosthetics, brain machine interface or brain computer interface (BMI or BCI), brain disorder diagnostic systems, personal entertainment, and so on. Moreover, it is particularly beneficial that comprehensive neuroscience research mandate the ability of high quality, parallel monitoring over a large scale of neurons within a small volume of the brain.

However, to realize a neural recording microsystem with a massive number of parallel recordings while maintaining high performance is not a trivial task. From an electrical

engineering point of view, in order to realize that neural recording microsystem, two major design issues must be addressed. First, the energy efficiency of the system should be maximized. As the number of simultaneous recording channels increases, power consumption in the system beyond some critical point (known as 80 mW/cm^2) leads to serious tissue damage through overheating, provided the energy consumption of the system is simply proportional to the channel count [7]. In addition, if considering fully implantable neural recording microsystems, risky and expensive surgical replacements may be required frequently due to the limited electrical energy source inside human or animal bodies. Many current neural recording systems have dealt with this energy issue, and their performance has continuously improved throughout the years. Several recent works already reached a few μW and even sub-μW per channel power consumption [8][9][10][11]. The other issue is the area the system occupies. Not only for the comprehensive neuroscience research, but also for almost all neural interface systems, the system should have a small form factor so that it can fit into the small volume of a targeted brain region. For instance, one electrode (one recording site) per one neuron is ideal for neuroscience research. Several outstanding studies have addressed this area-related issue [8][12].

However, most recent works have focused on only decreasing energy and area usage independently, even though the two issues are inter-dependent. In addition, in the view of a complete and self-sustained system, the other functional parts of the system should be considered and incorporated. Figure 1-1 describes a neural recording microsystem architecture consisting of some essential circuit blocks to be called a complete system (The stimulation, either electrical or optical and digital signal processing (DSP) blocks can be optionally included). The analog front-end (AFE) block is crucial to record and

precondition incoming neural signals. Needless to say, design optimization for this AFE block should be made. The design of the power management and digital interface blocks should also be carefully planned and constructed since they affect the total energy and area efficiency of the system. Otherwise, a standalone AFE block is subject to several practical limitations, such as heavy tethering and inefficient energy conversion due to the absence of the digital interface and power management blocks. Unfortunately, many recent works have focused on only the recording blocks, even though their final goals are to implement the complete neural recording systems.

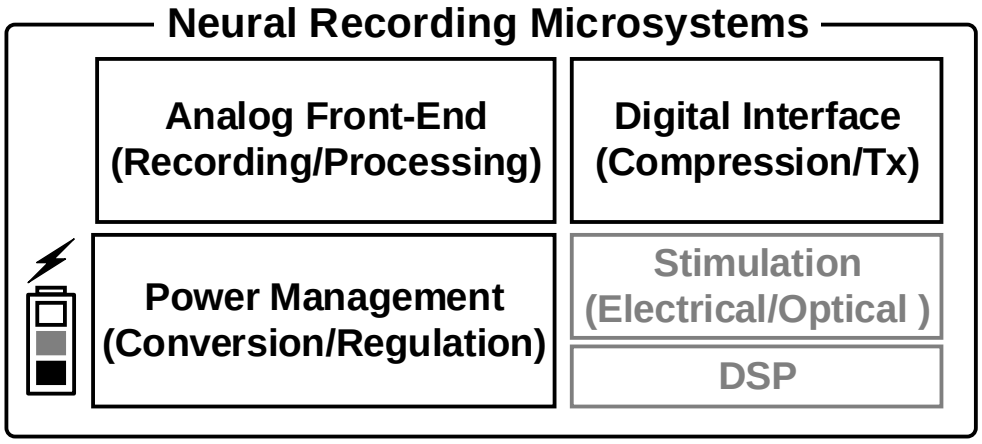


Figure 1-1 Architecture of neural recording microsystem (the optional blocks are grayed).

Our ultimate goal beyond this thesis is to build a neural recording microsystem facilitating 1,024-channel parallel recordings and the associated readout circuits [5]. **Error! Reference source not found.** (a) conceptually visualizes the proposed platform for the 1,024-channel parallel neural recording microsystem. To implement the proposed system, 8 modules consisting of a 128-channel neural probe and an application specific integrated circuit (ASIC) with 128 recording channels, an interposer, and a silicon-cap (Si-cap) will

be stacked as shown in Figure 1-2 (a). Each 128-channel probe has eight 6-mm long and 15-µm thick shanks, with 16 individual recording sites in each shank. For the connection between the 128-channel probe and the ASIC, an interposer is fabricated. The dimension of the interposer is 10.8 mm × 5 mm × 0.04 mm, and 135 through-silicon-vias (TSV) in the interposer make it possible for the ASIC to be flip-chip bonded as depicted in Figure 1-2 (c). In addition, a custom silicon cap (Si-cap) is fabricated to form a hermetic seal that prevents leakage and direct contact between the electrical components and tissues, except for the recording sites. This cap has an indium (In) sealing rim for solder bonding (appearing white in Figure 1-2 (b)). The photograph of a fabricated single module with a U.S. penny for scale is provided in Figure 1-2 (b).

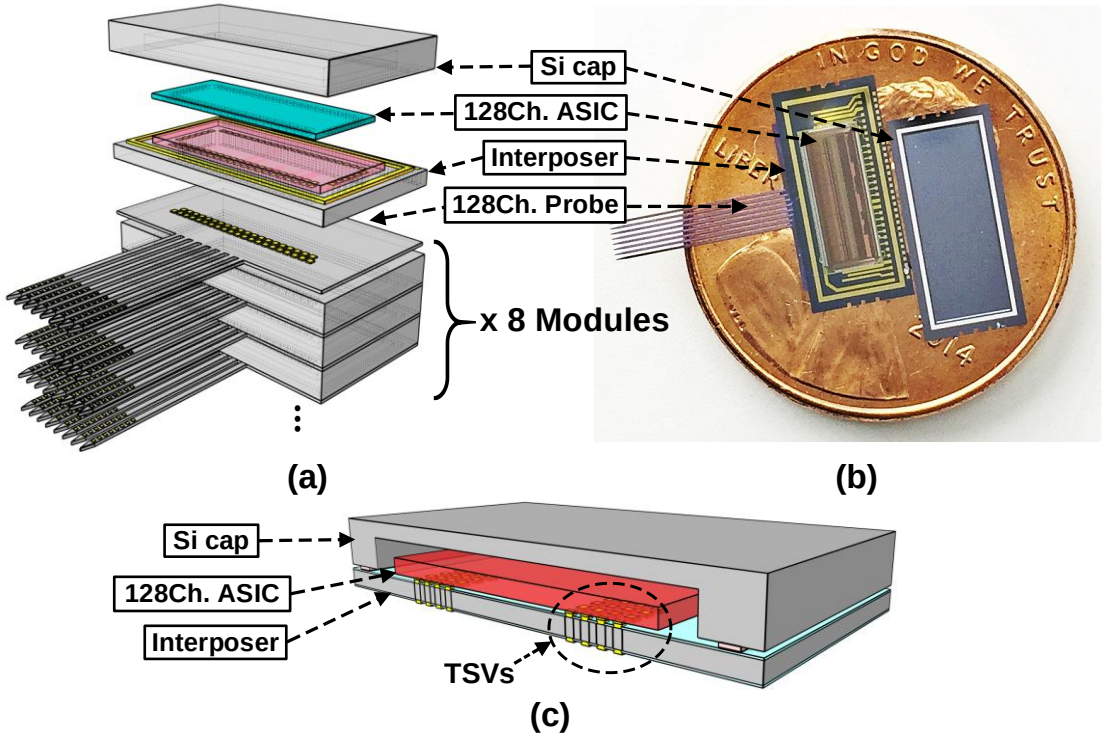


Figure 1-2 (a) Conceptual platform for a 1,024-channel parallel recording microsystem;(b) photograph of a fabricated single module with a 128-channel multi-shank probe, interposer,

Si-cap, and 128-channel AFE chip; and (c) cross-sectional view of the interconnection between the probe and the AFE chip

To reach that goal, the area and energy efficient architecture and circuit design not only for (1) the recording analog front-end (AFE), but also for (2) the digital interface, *i.e.* lossless neural signal compression scheme and (3) the power management were implemented. Those three inter-related projects are the main topic of this thesis. With the combination of the designs of the three essential functional blocks and the collaborative works related to the system assembly, a practical 1,024-channel parallel and high quality neural recording microsystem with optimum energy and area usage will be realized.

1.2 Neural recording microsystem

The ability to reliably and efficiently record neural signals from living matter has been of the most challenging goals in neuroscience and in the engineering of neural interfaces. This section will review the design principles and challenges involved with the advancement of various neural recording microsystems, including neural probes and active circuits. Even though this thesis is more closely related to the electrical circuit designs and architectures necessary for neural recording microsystems than for neural probes, the neural probes are nevertheless one of the most important building blocks in neural recording microsystems.

1.2.1 Overview

During the 1950s and 60s, early electrical recordings from nervous systems were achieved by attaching individual electrodes to a single wire, or to wire bundles. Even though this approach showed viable results throughout that period [13][14], it suffered many problems, such as poor fabrication precision, insertion damages, and less

reproducibility [15][16]. In terms of the miniaturization, reproducibility, and reliability of recordings, the pioneering work in neural recording microsystems might be Kensall D. Wise's 1969 introduction of the neural probe using microfabrication techniques (published in 1970) [17]. Figure 1-1 shows the microphotograph of the fabricated microelectrode and its schematic. This early work showed the substantially small size of the electrodes (2-μm diameter, 10–20-μm electrode distance), reduced tissue damage from insertion, and proved the feasibility of the micro-machined probe through the extracellular action potential (EAP) recordings. In addition, follow-up research in 1975 by the same author used a junction field-effect transistor (JFET) buffer in proximity to electrodes to minimize the inter-electrode parasitics. This can be regarded as a precursor of the system-level integration approach to neural recordings [18].

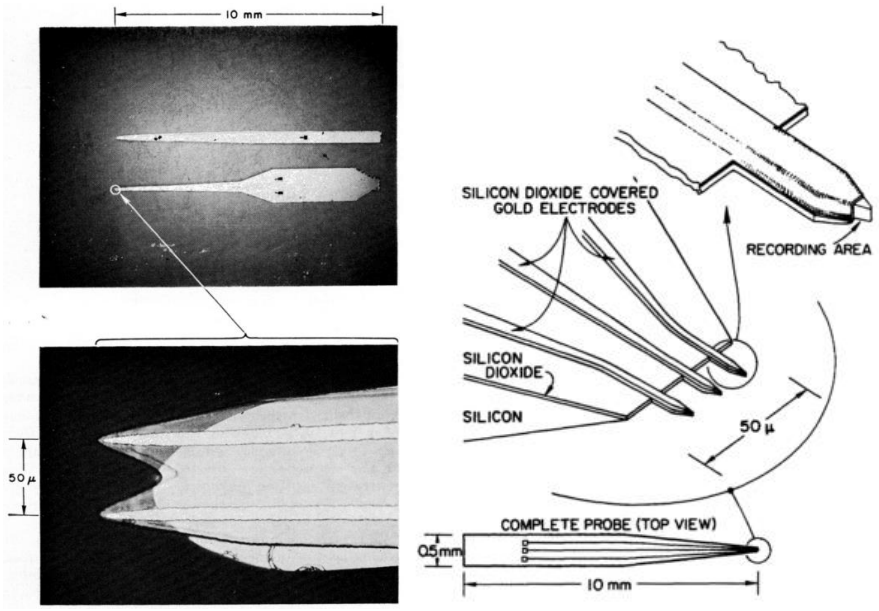
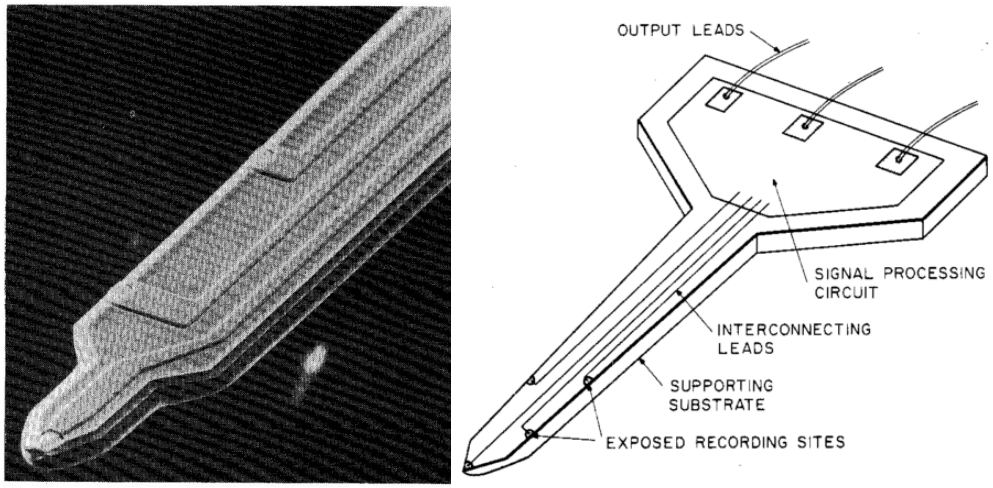


Figure 1-3 Silicon microelectrode: (a) microphotograph and (b) conceptual schematic

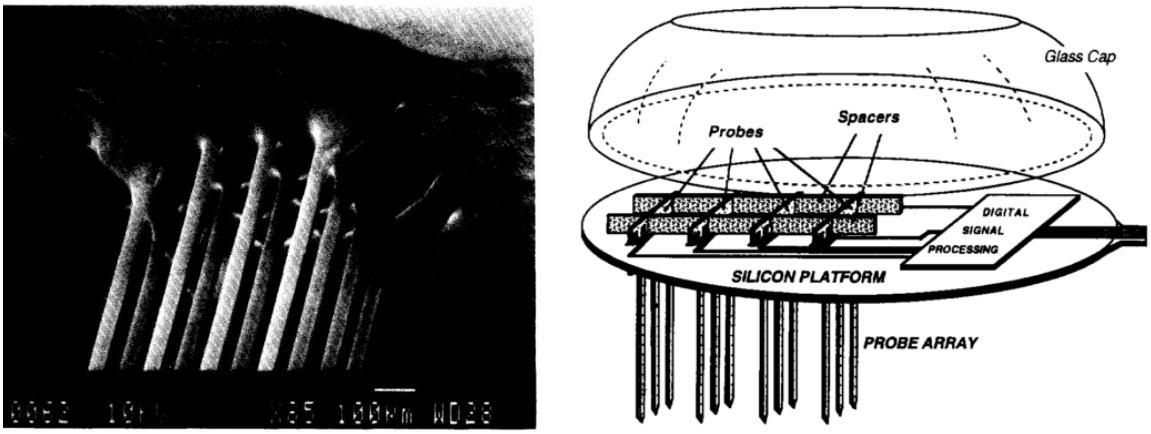
To extend and remedy possible issues with the initial effort, many researchers had suggested attractive solutions. As a result, the integration of the probe and the associated circuits in either hybrid or monolithic ways has opened up a new era for more reliable and reproducible neural recording microsystems. In 1985, Najafi and his colleagues reported a neural recording microsystem ('active probe'), which included a multielectrode array with 10 recording sites, an on-chip amplifier, and data multiplexing [19]. Figure 1-4 depicts the fabricated active probe and its conceptual drawing. The active circuitry was fabricated with a 6-µm, double-poly triply-implanted E/D NMOS process, and the probe was defined by the boron etch stop technique. This basic planar structure of the microfabricated probe with active circuitry is nowadays referred to as the "Michigan Probe." In 1990, the Michigan Probe was further improved upon when a 32-channel probe was developed with 8-channel parallel recording capability out of the 32 sites [20]. The active circuitry on this probe was fabricated with a 3-µm CMOS process. The next generation, 3-dimensional (3-D) neural recording microsystem was initiated in 1991 with the aid of the Michigan Probe concept (published in 1994) [21]. As shown in Figure 1-5, the array of the 2-dimensional Michigan Probes was perpendicularly assembled in the micromachined platform, with hybrid-assembled circuitry, to form the 3-D structure. In 2004, the monolithically integrated front-end amplifier and probe were fabricated with two-poly, one-metal CMOS process and reported [22]. The array of these probes was also assembled in the 3-D platform with an embedded data-compression application specific integrated circuit (ASIC) to become a 256-sites (32 parallel recording) fully implantable neural recording microsystem.



(a) (b)

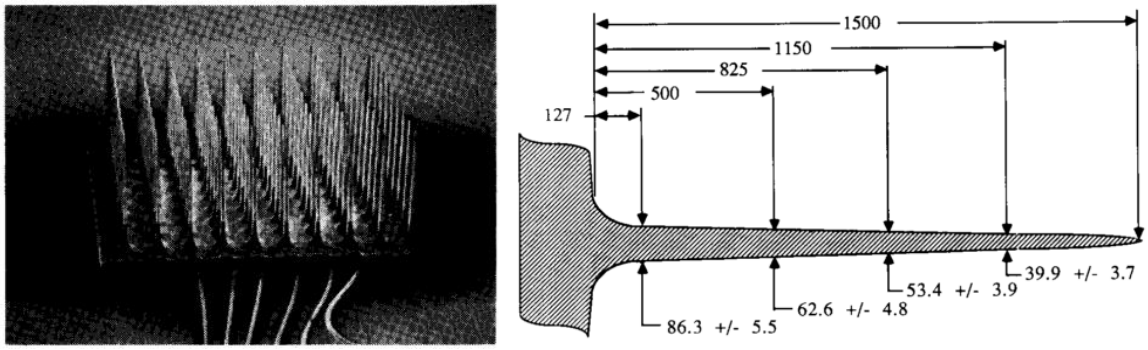
Figure 1-4 10-channel active probe: (a) microphotograph and (b) conceptual drawing

Another important milestone in the neural recording microsystem is based on the advent of the "Utah Probe" (or Utah Electrode Array (UEA)), developed by Normann at the University of Utah in 1988 (published in 1991) [23]. In contrast to the Michigan Probe, the Utah Probe is vertically processed to have about 100 pyramidal shape needles within a $4 \times 4\text{-mm}^2$ silicon substrate, as shown in Figure 1-6. Thanks to the mechanical strength of silicon and the needle shape, the Utah Probe is easy to insert into tissues. However it is unsuitable for chronic recording in human applications because of its stiffness [24]. In 2007, Harrison and his colleagues developed a neural recording microsystem based on the Utah Probe [25]. In their system, the active circuits--including the front-end amplifier, analog-to-digital converter (ADC), wireless data (433 MHz, FSK) and power transmission (13.56 MHz ISM band), on-chip spike detection, and data compression--were built on top of the Utah Probe to form a 100-channel neural recording microsystem.



(a) (b)

Figure 1-5 3-D microelectrode array: (a) microphotograph and (b) conceptual drawing

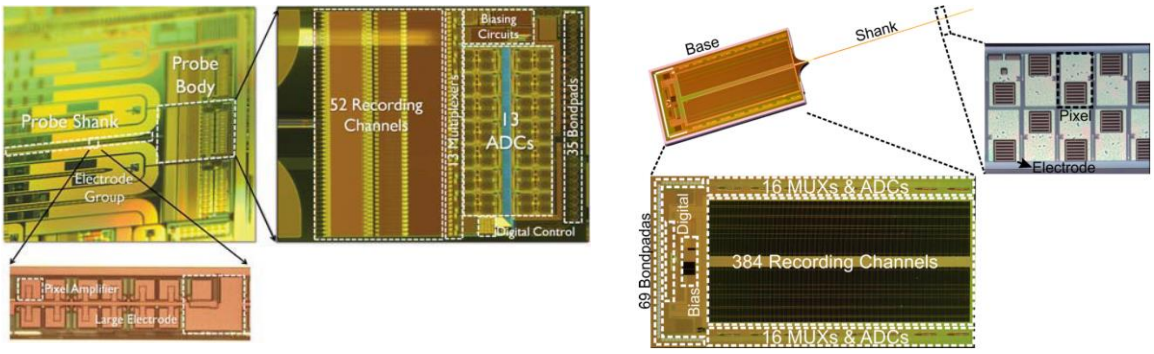


(a) (b)

Figure 1-6 Utah Probe (a) photograph and (b) scaled drawing of an electrode side view (All values are in μm)

Nowadays, state-of-the art neural recording systems have roughly over a hundred parallel recording capabilities, while exhibiting high-performance signal quality and system resource usage (power and area). As mentioned in the previous paragraph, a wireless 100-channel neural recording system combined with the Utah Probe was successfully demonstrated in the in-vivo environment. In 2009, the Δ-compressed, 256-channel neural recording system was proposed [3]. The active circuitry in the proposed microsystem was fabricated with a 0.35-μm CMOS process and hybrid-combined on top

of the Utah Probe. As part of the mobile neural acquisition system (Hermes series, Stanford University), a 96-parallel channel for the broadband recording (sub Hz ~ over 10 kHz) to fully cover both of local field potentials (LFP) and action potentials (AP, or spike) was also reported [26]. In the same year, the implantable 64-channel neural recording system using the Michigan Probe was demonstrated [27]. The platform consisted of two chips: one with 4 × 16-channel amplifiers and filters, and the other with a 64-channel signal processing unit and bidirectional power and data telemetry. In 2011, the University of Santa Cruz demonstrated a 128-channel simultaneous recording ASIC, facilitating on-the-fly spike-sorting hardware to investigate the field of neural prosthetics [4]. In 2012, the Institute of Microsystem Technology (IMTEK) in Europe announced a CMOS-based high-density silicon microprobe array with a 188-electrode shank and site selection circuits [28]. Since their approach uses the standard CMOS process, several years later they reported follow-up research, notably an on-probe CMOS amplifier to be built on the reported probe [29]. In addition, another European research group, Inter-university MicroElectronics Center (IMEC), reported a neuroscience research tool which has even higher recording channel counts (455) on a monolithically fabricated CMOS active probe with 55 parallel recordings and user programmability [30], as shown in Figure 1-7 (a). The IMEC is continuously pushing this monolithic approach and publicized the next version of the monolithic active probe, which has 384-parallel recording capability out of 966-channels in 2016 (Figure 1-7 (b)) [6].

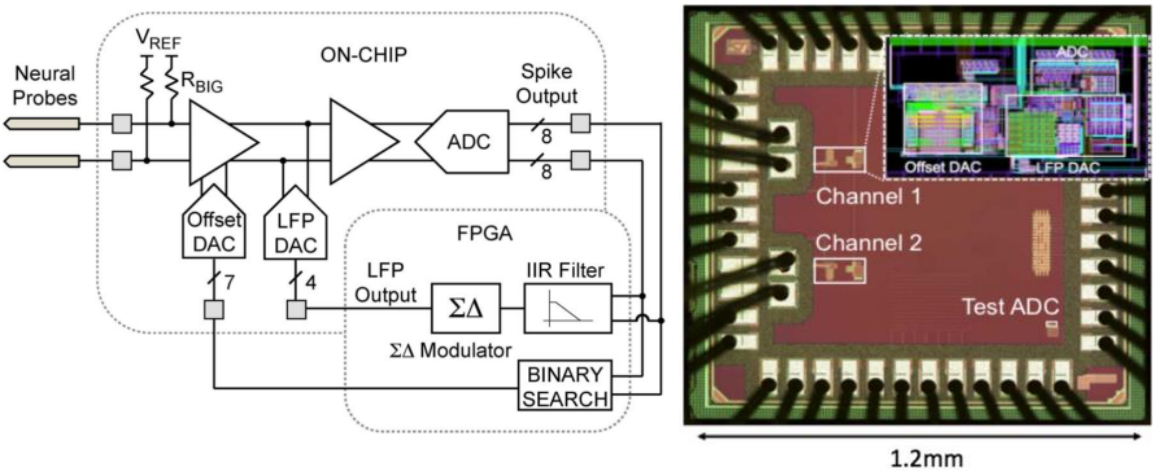


(a) (b)

Figure 1-7 Microphotographs of IMEC's: (a) 455-channel monolithic active probe (54 parallel recordings) and (b) 966-channel monolithic active probe (384 parallel recordings)

1.2.2 Area- and energy- intensive circuit and architecture design

As the number of the parallel recordings has steadily increased, the consumption of system resources such as energy (power) and area has increased as well. The high power consumption from the large amount of electrical recordings in the system inevitably leads to tissue damage, and a bulky system also limits chronic monitoring of brain activities and is often ill equipped to investigate small regions of the brain. This section will review the recent efforts related to minimizing energy and area consumption in the neural recording microsystem.



(a) (b)

Figure 1-8 0.013-mm^2-per-channel neural recording circuit: (a) block diagram and (b) microphotograph of the fabricated prototype

A single channel implementation in an extremely small area of 0.013 mm^2 using a mixed signal architecture and advanced technology, 65nm was reported in [8]. The fabricated ASIC can squeeze the area consumption extremely by using direct coupling to electrodes (DC-coupling) and can thereby eliminate the large on-chip capacitors that have been widely used with conventional AC-coupling in neural recordings. Figure 1-8 shows the block diagram and microphotograph of this work. Even though it consumes more than 10 times smaller area than others, the important functions depend on the external assistance (FPGA as shown in Figure 1-8 (a)). In addition, due to its front-end amplification using open loop, gain variation between channels might be expected. Similar work with this architecture was also published by the same author without any external assistance in 2014 [31]. However, the area consumption became almost doubled, 0.025 mm^2 per channel.

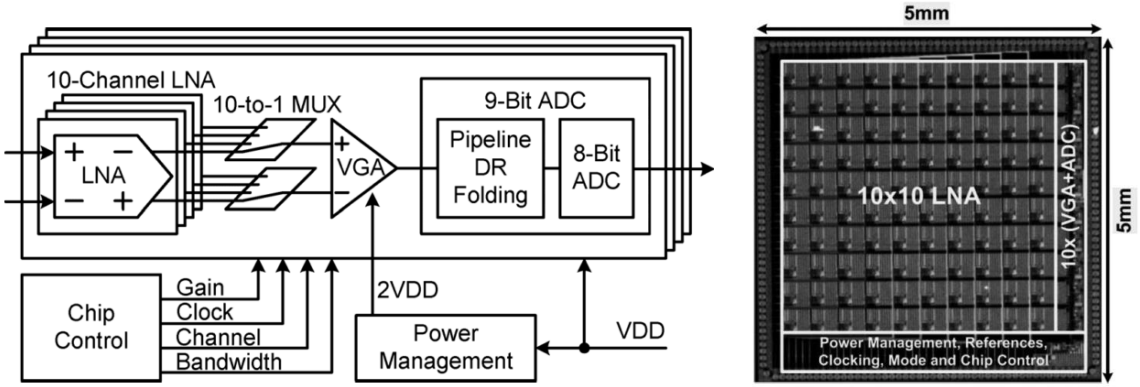


Figure 1-9 100-channel neural recording chip with sub-µW-per-channel power consumption: (a) block diagram and (b) microphotograph of the fabricated prototype

In 2011, a 16-channel neural recording interface that consumes only 1.13 μW per channel from a 0.5 V supply was presented [10]. The interface eliminates the power-hungry analog multiplexer and the associated buffer to reduce power consumption. Instead, the multi-channel successive approximation register (SAR) ADC was proposed to facilitate multiplexing among the channels. After two years, [11] reported a 100-channel broadband-recording (sub-Hz to 10 kHz) ASIC with only sub-μW (0.73 μW per channel) power consumption. Figure 1-9 shows the functional block diagram and the fabricated chip microphotograph of the reported work. This work might be able to achieve such low power consumption with the 0.45 V and 1.0 V dual supply with the consideration of the signal dynamic range (DR) throughout the acquisition channels. For the part of the ASIC where relatively high current is necessary for the low noise performance while exhibiting low DR, they used the 0.45V low supply voltage, while for the part where high DR is required they used the high supply 1.0 V. For facilitating different supplies, the interface also has an on-chip power management unit (a dc-to-dc converter).

1.2.3 Summary

This chapter dealt with the motivation for this thesis and the brief overview of the neural recording microsystems from 1970 to the present. At the end of the review, the studies highlighting two important circuit design parameters for neural recording microsystems area and energy consumption, were explained. Table 1-1 summarizes and compares the important circuit design specifications of the recent state-of-the-art neural recording microsystems.

Table 1-1 Performance summary of recent neural recording microsystems

	[6]	[5]	[11]	[30]	[12]	[8]	[4]	[26]	[10]	[27]	[3]	[25]	[22]
Year	2016	2015	2013	2013	2012	2011	2011	2011	2011	2009	2009	2007	2005
N. of channels	966	128	100	455	16	1	128	96	16	64	256	100	256
N. of recording	384	128	100	52	16	1	128	96	16	64	—	100	32
Power/Ch.[μW]	*49.06	3.05	0.73	*27.84	3.96	5.04	**23.44	**67.7	1.13	75	15	145	*94.5
Area/Ch. [mm^2]	*0.13	0.05	0.25	*0.19	0.07	0.013	**0.495	0.05	0.073	0.072	0.04	0.16	*0.625
IRN [μV_{rms}]	6.36	3.32	3.2	3.2	4.8	4.3	4.9	2.2	5.32	8	7	5.1	8.9
B.W.[kHz]	10	10.9	10	6	8.9	10	20	10	7.5	9.1	5	5	10
Sampling Rate [kHz]	30	25	20	30	6.25-50	20	40	31.25	30	62.5	10	15	20
ADC Resolution (Target/ENOB)	10/–	1/10.9	10/8.27	10/9.2	9/7	8/**7.18	6~9/–	10/9.72	8/7.32	8/–	***–	10/–	5/–
Ch Multiplexing (Analog/Digital)	Y(A)	N	Y(A)	Y(A)	Y(A)	N	Y(A)	N	Y(D)	Y(A)	N	Y(A)	Y(A)
Technology [μm]	0.13 SOI	0.18	0.18	0.18	0.25	0.065	0.35	0.13	0.13	0.5	0.35	0.5	3

*Include I/O power and interface circuits, otherwise power and area of AFE, **Estimated, ***No ADC, light gray: applicable to the Michigan Probe, dark gray: applicable to the Utah Probe

1.3 Thesis outline

1.3.1 Chapter 2: Characteristics of the neural signal

This chapter covers the biological structure of neurons in brain to elucidate basic characteristics of neural signals, and highlights the electrical characteristics of the neural signals for comprehensive neuroscience research. The challenges underlying in electrical recording of the neural signals will be briefly explained in the context of the neural signal characteristics.

1.3.2 Chapter 3: 128-channel modular AFE architecture

This chapter covers the architecture and circuit implementation of the implemented 128-channel Δ-ΔΣ analog front as a foundational work for the 1,024 neural recording microsystems. Then, the measurement results of the fabricated work and the comparison with the other state-of-the-art AFEs will be shown.

1.3.3 Chapter 4: On-chip neural signal compressor

This chapter introduces a lossless data compression scheme for neural recording microsystems. The strong motivation for the data compression in neural recording microsystems, and how the inherent natures of neural signals were exploited to realize the compression scheme will be explained. The implementation of the compressor with a 128-channel AFE and the measurement results will follow.

1.3.4 Chapter 5: PWM buck converter using analog-digital hybrid control

As an important supporting block in neural recording microsystems, this chapter deals with the power management circuit from its design and implementation to the measurement

results. The unique requirement of the dc-to-dc conversion for the neural microsystems will be addressed first, then the details of the design issue and measurement results will be delivered next.

1.3.5 Chapter 6: Summary and future works

This last chapter summarizes this thesis by pointing out the important results and suggests a possible future work to improve the works this thesis includes.

CHAPTER 2
PHYSICAL AND ELECTRICAL PROPERTIES OF CELLS IN THE BRAIN

The human brain is incredibly complex with an average of 21 billion neurons controlling sensation, movement, autonomic, mental processes, and so on. This vast network of cells constantly develops new interactions, interconnections, and adapts outputs based on the inputs to the brain. The functions of the human body require chemical and electrical interaction among neurons.

This chapter will briefly review the basic physical, chemical and electrical properties of the neurons.

2.1 Structure of neurons[1]

Neurons take information, process it, and generate the appropriate output based on the information. The organelles of a neuron include a nucleus, Golgi bodies, mitochondria, lysosomes and endoplasmic reticulum. The nucleus, Golgi apparatus, and endoplasmic reticulum are located to the soma or cell body of the neuron. Other organelles such as the mitochondria and endoplasmic reticulum are distributed throughout the neuron. A plasma membrane surrounds the cell while separating the extracellular environment from its contents.

[1]Chapter 2.1 and 2.2 are excerpted from a book, "Neuroscience Fundamentals for Rehabilitation" by L. Lundy-Ekman, 2002 [32].

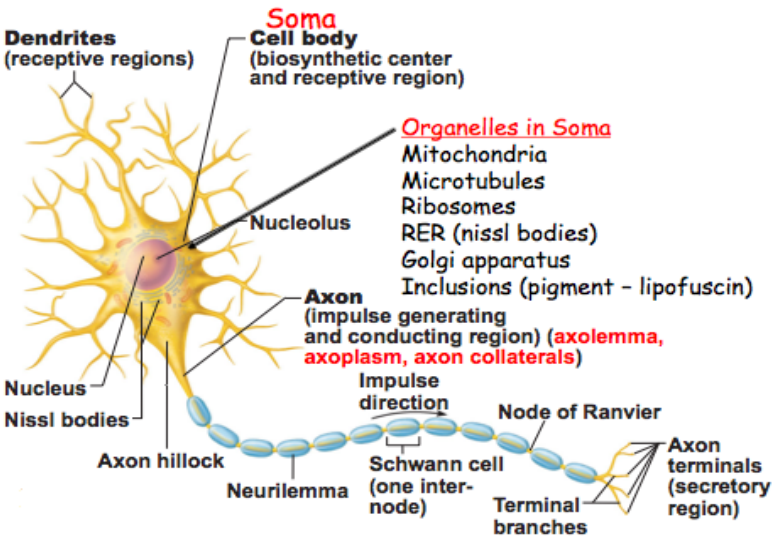


Figure 2-1 Structure of a neuron

A typical neuron has four parts: dendrites, axon, presynaptic terminals and soma. Dendrites, branchlike extensions that function as the main input of the cell, project from the soma. Figure 2-1 shows the structure of a motor neuron. Another process extending from the soma is the axon, reaching from the cell body to target cells while serving as the output unit of the cell. Most neurons have a single axon that arises from a specialized region of the cell, called the axon hillock. Axons vary in length. The shortest axons are less than 1 mm in length, whereas axons that transmit motor information from the spinal cord to the foot many be up to 1 m long. Axons end in presynaptic terminals, or fingerlike extensions that are the transmitting elements of the neuron. Neurons transmit information about their activity through the release of chemicals (called neurotransmitters) from the presynaptic terminals into the synaptic cleft. The synaptic cleft is the space between neurons, and servers as the site for inter-neuronal communication.

Most of the neurotransmitters are produced in the soma of the cell. Thus, the cell must have a mechanism for transporting neurotransmitters and other substances from the soma

to the presynaptic terminal. This process is called axoplasmic transport. When material travels from the soma along the axon toward the presynaptic terminal, the process is called anterograde transport. Some substances must be transported from the synapse back to the soma, which is known as retrograde transport.

2.2 Transmission of Information by Neurons[1]

Neurons acts by undergoing changes in electrical potential across the cell membrane. An electrical potential across a membrane exists when the distribution of ions creates a difference in electrical charge on each side of the cell membrane. Four types of membrane channels allow ions to flow the membrane: leak, modality-gated, ligand-gated and voltage-gated.

All channels serve as openings through the membrane. When the channels are open, ions including, $Na+$, K^+, Cl^- and Ca^{2+} diffuse through the openings. Leak channels allow diffusion of a small number of ions through the membrane at a slow continuous rate. The other channels are termed "gated" because they open in response to a stimulus and close when the stimulus is removed. Modality-gated channels, specific to sensory neurons, open in response to mechanical forces (*i.e.* stretch, touch, and pressure), temperature changes or chemicals. Ligand-gated channels open by a neurotransmitter binding to the surface of a channel receptor on a postsynaptic cell membrane. When open, the channels allow the flow of charged ions between the extracellular and intracellular environments of the cell, resulting in the generation of local potentials. Voltage-gated channels open by changes in electrical potential across the cell membrane. Changes in membrane potential produce a structural change of the channel that causes the channel to either open or close. Voltage-gated channels open almost instantaneously, and close as quickly. They are ultimately

responsible for the electrical signals that are the basis of information transfer in the nervous system. Voltage-gated channels are important in the release of neurotransmitters and the formation of action potentials.

A rapid change in electrical charge across the cell membrane transmits information along the length of an axon and elicits release of chemical transmitters to other neurons or to the electrically excitable membrane of a muscle. The difference in electrical charge, carried by ions, is referred to as the membrane's electrical potential. Three types of electrical potentials in neurons are essential for transmission of information: resting membrane potential, local potential, and action potential (AP).

2.3 LFP and AP in neurons

A local field potential (LFP, sometimes, called as micro-EEG) is an electrophysical signal recorded from the multiple nearby neurons by inserting metal, glass, or silicon probes into the relatively deep part of brain. The LFP can be obtained by measuring broadband (wide-band, ~40 kHz) neural signal and filtering the high frequency components of it. The unfiltered, broadband signal usually reflects the sum of action potentials from multiple cells within $50 - 350$ μm from the tip of the electrode and slow ionic events within $0.5 - 3$ mm from the tip of the electrodes. The LFP is a low pass filtered form of such broadband signals with about 300 Hz cutoff frequency [33].

Action potential (AP) is the reflection of the rising and falling electrical potential in membrane which is generated by special types of voltage-gated ion-channels embedded in a cell's plasma membrane. As explained in the section 2.2, if the channels shut down, the membrane potential stays in resting (called resting potential) and if the channel opens, the potential abruptly rises in either ways, positive or negative relative to the resting potential.

The action potential in neurons is known as spikes and the burst of those spikes is said as spike train. The emission of the neuronal AP is called a firing.

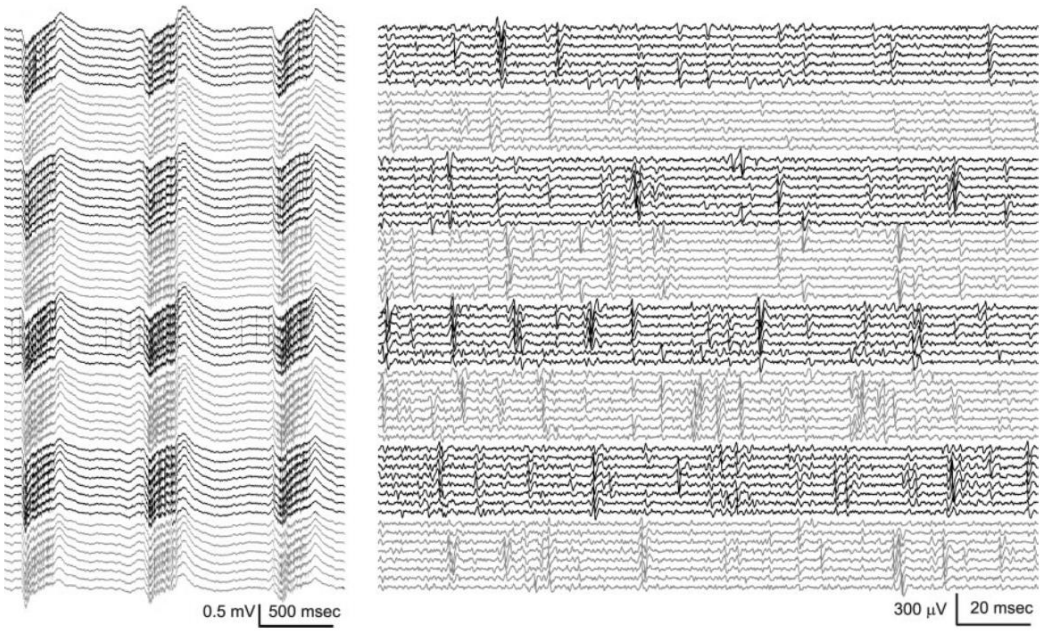


Figure 2-2 An example of parallel recording of (a) broadband neural signals (b) filtered traces of (a)

2.4 Electrical characteristics of LFP and AP

To understand brain activity, particularly for neuroscience research both local field potentials (LFP), average of field potential of nearby neurons surrounding monitoring sites and action potentials (or spikes) should be recorded simultaneously [33][34]. Figure 2-2 shows an example of (a) the broadband (AP on top of LFP) signal in layer V of the somatosensory cortex and (b) the filtered traces (0.5 − 5 kHz) of (a) [35]. The amplitudes of these signal ranges from the order of few μV to several mV and frequencies of them spans from DC to a few kHz. LFP, representing ensemble of the activity from the sets of

neurons surrounding the recording sites, can be found in the low frequency range (~1 − 300Hz). On the other hand, AP, representing single cell activity, is located in the higher frequency range (~300 – 10,000Hz). According to the nature of those signals, recording circuits have to be designed with sufficiently low input-referred noise (IRN) and high gain and dynamic range (DR) to measure both of the LFP and AP simultaneously. Furthermore, large amplitude fluctuation at near DC (~ 0 Hz) coming from electrode-tissue interface often exists in input signals and should be removed by high-pass filtering at very low cutoff frequencies [36]. This requires neural recording analog front-ends to facilitate the instrument level performance; over 60 dB signal DR, less than 5μV IRN, a wide range of bandwidth (larger than 10^5 ratio of minimum to maximum frequency) and larger than 40 dB flat-band gain. Those specifications may complicate the design of integrated electrical recording circuits, and to make even worse, constraints about resource usage such as power (energy) and silicon real-estate (area) make the designer even more perplexed.

CHAPTER 3
128-CHANNEL MODULAR Δ-ΔΣ ANALOG FRONT-END ARCHITECTURE

3.1 Introduction

For past few decades there has been significant advance in the development of neural recording systems with the fast advance in electronics and sensor technologies. Current state-of-the art neural recording systems are having roughly a hundred of recording channels while exhibiting high performance in terms of signal quality and resource usage.

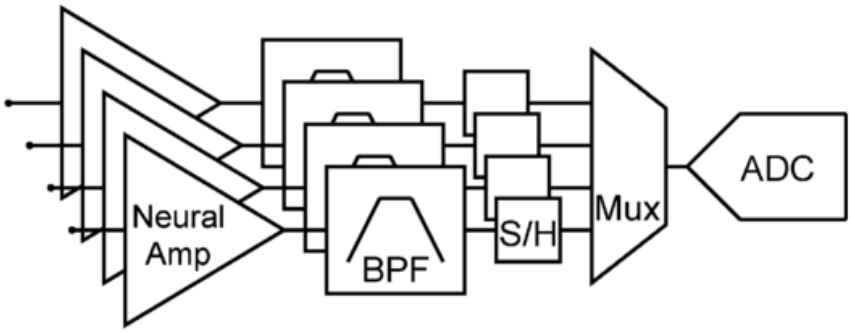


Figure 3-1 Typical multi-channel neural recording AFE

The implicit strategy of the conventional approaches to the effective usage of area and energy for implementing neural recording analog front-ends (AFE) is explained in the view of the topology selection: channel multiplexing and selection of ADCs. A neural recording AFE generally consists of a low noise amplifier (LNA), a programmable gain amplifier (PGA) with bandpass filtering, an analog multiplexer, and an analog-to-digital converter (ADC) in series. Figure 3-1 shows a typical multi-channel neural recording AFE [37]. Leaving the LNA behind this discussion due to its low noise performance (there is a strong relation between power and noise), the energy saving in the conventional neural signal

acquisition chains mostly comes from the selection of a specific types ADC, successive approximation register (SAR) ADC since there is no static power consumption and the switching power from the SAR-ADC is significantly reduced by using small size of capacitors and special control schemes [38][39][40]. As shown in Figure 3-1, the channel multiplexing is utilized for the area reduction since the SAR ADC consumes relatively large area compared with the other types of ADCs [41]. Many of recent papers took this approach even though their number of channel multiplexing is different [4][6][10][11][12][25].

However, the channel multiplexing has serious shortcomings. The multiplexing itself needs high speed buffers and switches, thus the additional power consumption from them occurs. Although there is optimizing strategy for the reduction of area and power consumption as reported in [41], it is obvious that the additional power consumption exists. In addition, the high quality recording might be deteriorated since it causes crosstalk between channels. The high quality recording requiring more than 60dB dynamic range (DR) can be deteriorated by the crosstalk since the modern CMOS switch only provide a few GΩ range off-resistance. More seriously, the channel multiplexing cannot provide true real time recording between since it depends on the time division multiplexing in analog domain. To maintain the true timing information between the recorded signal an additional compensation scheme is necessary [42].

To address those challenges, the architecture having an in-channel Δ-ΔΣ (Δ-modulated ΔΣ) ADC which is efficient for both the area and energy consumption while providing high quality recording with more than 60dB signal to noise ratio (SNR) is proposed, consequently the true parallel recording without channel multiplexing can be realized [5].

The Δ-ΔΣ ADC combined with the front-end (it is going to be called Δ-ΔΣ AFE.) signal conditioning blocks is able to compress the DR of neural signals without sacrificing signal integrity by taking a temporal difference of the oversampled neural signals (Δ–modulation) instead of recording raw signals. This is mainly due to the nature of neural signals where most of their energy are confined at low frequencies and they follow ~1/*f* curve in their signal spectrum [43][44]. For quantization, the ΔΣ-ADC is employed instead of the SAR-ADC since they are more compact than successive SAR ADC and both of Δ and ΔΣ modulator can take the advantage of oversampling [45]. The implementation of the ΔΣ ADC significantly reduces area, while its energy overhead is compensated by the DR compression scheme. The continuous time (CT) operation of the ΔΣ ADC also helps save energy due to its smaller bandwidth requirement than discrete time (DT) counterparts [46]. In addition, the Δ-ΔΣ ADC occupies only small area of 40 μm × 650 μm, thus the true parallel signal processing without any multiplexing scheme is possible. With this Δ–ΔΣ ADC, the Δ-ΔΣ AFE can achieve >10b resolution with only 3.05 μW power consumption within 40 μm × 1120 μm of area.

Overall, the Δ-ΔΣ AFE achieve the highest energy-area efficiency while maintaining the state-of-the-art low noise performance and facilitating high enough dynamic range to capture large swing of neural signals without distortion. This chapter is organized as follow. The Δ–ΔΣ AFE architecture is covered in the section 3.2. In the section 3.3, the main circuit blocks and the adopted (used) circuit design techniques for realizing the architecture are described in details. The section 3.4 presents the measurement results, obtained in two different kinds of experiments: electrical performance and *in-vivo* experiments with a rat. Finally, the section 3.5 states the conclusions of this paper.

3.2 Circuit architecture

3.2.1 Modular Δ-ΔΣ AFE architecture

Figure 3-2 Top level architecture of 128-channel neural recording Δ-ΔΣ AFEshows the top level architecture of the Δ-ΔΣ AFE including on-chip signal conditioning and off-chip signal restoration. The on-chip AFE consists of 128 neural signal acquisition channels, each including a low noise amplifier (LNA), a programmable gain amplifier (PGA), a Δ-ΔΣ analog to digital converter (ADC) and a digital decimation filter in series, and there are bias circuits and data serializer to support the data processing. There is also a programming shift register to change the gain and bandwidth setting of the LNA and PGA. All of the amplified, modulated neural signals from the 128-channel Δ-ΔΣ AFE is serialized onto a single bit and sent to the off-chip module for the signal restoration and storage. In addition, the number of channel can easily be expanded if the silicon real estate permits because each channel is self-contained (no multiplexing in analog domain). Since the incoming data are modulated, the signals should be processed to retrieve the original data. The signal spectra in Figure 3-3 conceptually illustrate the significant compression of the variance of the neural signals by the Δ-modulation. The compressed signal is then digitized by a ΔΣ-ADC with reduced resolution requirement and sent to the off-chip module. At the outside of the chip, the compressed neural signal is processed with Σ–modulation outside of the chip, and consequently the original signal is successfully retrieved.

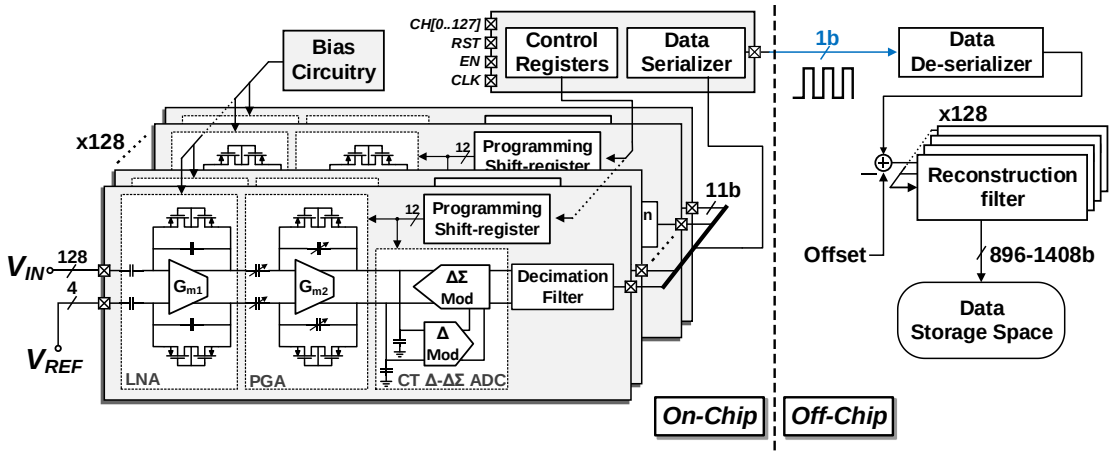


Figure 3-2 Top level architecture of 128-channel neural recording Δ-ΔΣ AFE

3.3 Circuit implementation

Figure 3-4 shows the schematic of the single channel AFE. The AFE operates fully-differentially to minimize any common mode (CM) variations and increase the DR of signals. Thanks to the CT-operation of the Δ-ΔΣ ADC, no sample and hold (S/H) circuit between the PGA and ADC is necessary and the anti-aliasing requirement is also relaxed [47]. All analog blocks are operated with 0.5 V supply except for the comparator that is connected to digital blocks operating with 1.0 V supply. The AFE consumes only 3.05 µW power while occupying small area of 0.05 mm^2 (45 µm × 1120 µm). The important parameters for *in-vivo* experiments such as low and high frequency corner (f_L and f_H) and total gain of the channel are externally programmable according to users' demand.

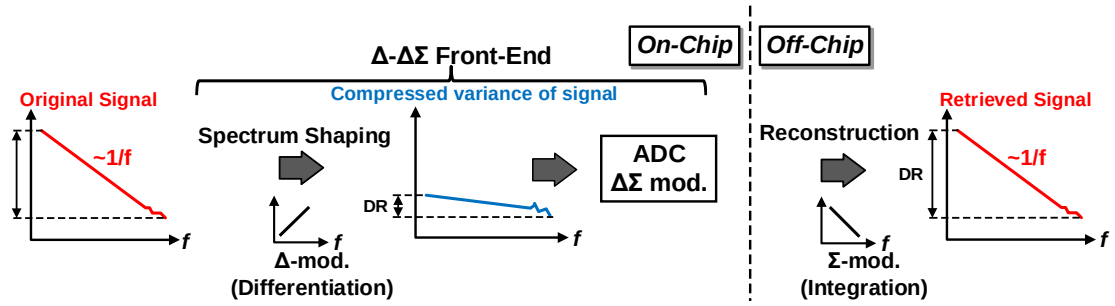


Figure 3-3 On-chip signal compression in Δ-ΔΣ AFE and off-chip reconstruction

3.3.1 LNA design

Since the LNA is located in the very first stage of the whole signal processing chain as shown in Figure 3-4, it must provide enough gain for the following stages to process the input signals while having low noise performance.

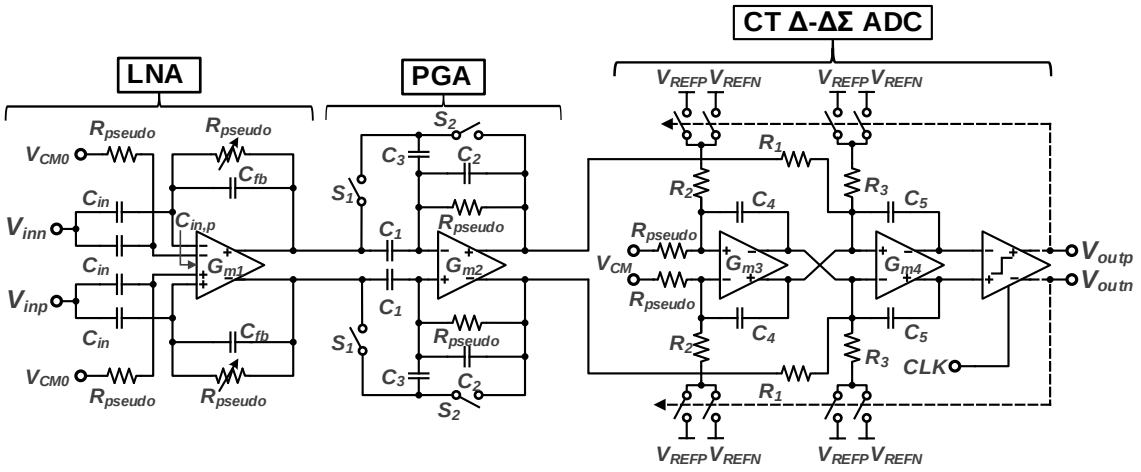


Figure 3-4 Schematic of single channel Δ-ΔΣ AFE

In addition, the LNA should reject large DC fluctuations (typically, 50~100 mV) coming from the electrode-tissue interface. In the LNA input neural signals are AC-coupled via the four capacitors, C_{in}, so that sub-Hz high pass corner frequency is formed with the combination of C_{in} and R_{pseudo} to suppress the large DC fluctuation. The design topology is

adopted from [48]. C_{in} must be made small enough to guarantee high input impedance at frequency of interests to minimize signal attenuation by the voltage dividing effect with electrodes, but large enough to avoid amplification of the intrinsic noise from LNA's operational transconductance block (G_{m1} in Figure 3-4). The amplification of the noise from G_{m1} is given by

$$\overline{v_{ni}^2} = \left(\frac{C_{in} + C_{fb} + C_{in,p}}{C_{in}} \right) \cdot \overline{v_{ni,OTA1}^2} \tag{2-1}$$

where $C_{in,p}$ is the parasitic input capacitance of G_{m1}, $\overline{v_{ni}^2}$ and $\overline{v_{ni,\mathrm{OTA1}}^2}$ are the input referred noise of the LNA and OTA, respectively. The gain of the LNA is generated by the closed loop feedback through C_{fb} and the ratio of C_{in} to C_{fb} makes overall gain. In this design, C_{in} and C_{fb} are chosen as 4.6 pF and 45 fF, respectively to limit the input referred noise by the capacitive amplification (1) to be less than 10% while maintaining reasonable input impedance, around 35 MΩ at 1 kHz (expecting the impedance of electrode as < 1 MΩ at 1 kHz) and generating gain of 40 dB. Even though the large capacitors are used for the closed loop gain and AC coupling, they do not play as a significant area overhead since they can be placed on top of the active circuitry thanks to the help from the selected technology, TSMC 0.18 μm process. For implementation of the large resistor, R_{pseudo}, the leakage current from a lateral bipolar junction transistor and a PMOS transistor is used [49]. The input referred noise from R_{pseudo} is

$$\overline{v_{ni,R}^2} = \left(\frac{V_{n,R}}{1 + sR_{pseudo}C_{fb}} \right) \cdot \left(\frac{1}{A^2} \right) \tag{2-2}$$

where $V_{n,R}$ is the thermal noise of R_{pseudo} and A is 40dB closed-loop gain of the LNA, respectively. According to (2), the input referred noise from R_{pseudo} is largely attenuated by

both of *A* and *R*-*C* network (-20 dB/dec) after sub-Hz frequency, $1/(2\pi R_{pseudo}C_{fb})$, therefore the noise contribution from R_{pseudo} is negligible compared to the thermal noise or flicker noise from G_{m1}. Figure 3-5 shows the transistor level schematic of G_{m1}. The thermal noise density of G_{m1} is given by (3)

$$\overline{v_{ni,th}^2} \approx \frac{8kT}{3} \cdot \left(\frac{1}{g_{m1} + g_{m3}} \right) \cdot \Delta f \tag{2-3}$$

where k is Boltzmann constant, T is absolute temperature, and g_{m1} and g_{m3} are the transconductance of M_1 and M_3, respectively. To achieve low noise performance of the LNA all transistors in G_{m1} are designed to operate in the subthreshold region where the transconductance efficiency is maximized. The OTA also has two complementary inputs (M_1–M_4) as depicted in Fig. 3-5. The complementary inputs increase theoretically the transconductance of the first stage by a factor of 2, so that reduce the input thermal noise by factor of $\sqrt{2}$ by (3). The input referred flicker noise of G_{m1} is also given by (4)

$$\overline{v_{ni,1/f}^2} \approx \frac{1}{C_{ox}} \cdot \left(\frac{K_n g_{m1}^2}{(WL)_1} + \frac{K_p g_{m3}^2}{(WL)_3} \right) \cdot \left(\frac{1}{g_{m1} + g_{m3}} \right)^2 \cdot \frac{1}{\Delta f} \tag{2-4}$$

where K_n and K_p are flicker noise coefficients of the standard 1.8 V NMOS and PMOS, respectively. The large gate area of the input transistors (800 μm/0.35 μm, 840 μm/0.25 μm for M_1, M_2, and M_3, M_4, respectively) used to reduce the flicker noise.

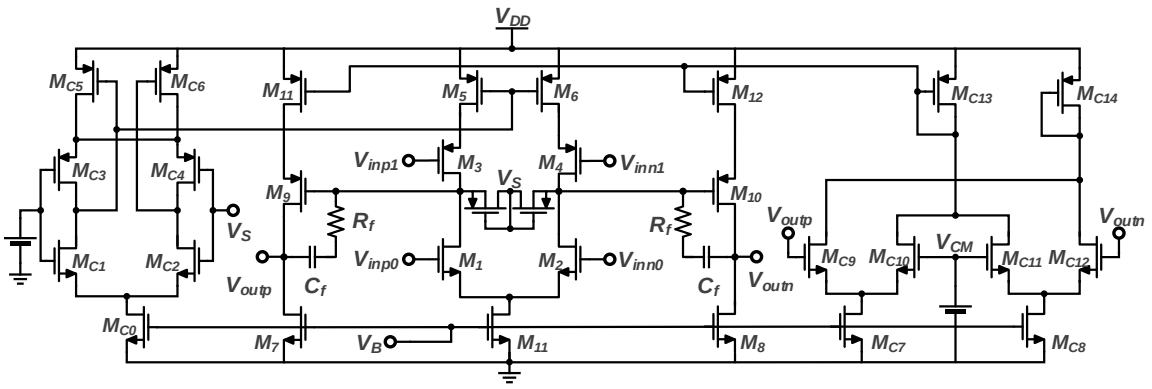


Figure 3-5 Schematic of Gm1 used for LNA

Due to the squeezed 0.5V supply voltage, a single DC voltage cannot effectively provides the proper bias both for NMOS (M_1 and M_2) and PMOS (M_3 and M_4) transistors. The input DC-bias for M_1 and M_2 comes from the output CM voltage, V_{CM} which is half of the supply voltage and that for M_3 and M_4 is generated using two diode connected PMOSs (not shown in Figure 3-5). The squeezed supply also degrades the common mode rejection ratio (CMRR) and power supply rejection ratio (PSRR) of G_{m1}. Thus, the dual tail currents (M_0, M_5 and M_6) are used to reduce the common mode gain.

Since G_{m1} has two amplifications stage, it need a frequency compensation network to guarantee enough phase margin. The Miller compensation using a 738 fF capacitor, C_f is used to move the dominant and second dominant pole and a 744 kΩ resistor, R_f is used for nulling of the transmission zero created from the Miller capacitor, C_f. The each stage of the transconductor also needs the common mode feedback (CMFB). Two schematics of the two different CMFB circuits are provided in Figure 3-5 as well. The transistors used for the CMFB are indicated as the lower case of 'c' in Figure 3-5. The LNA generates 3.05 μV_{rms} input referred noise through 0.5 Hz to 10.7 kHz while consuming 1.6 µA static current, however, its power consumption remains sub-µW levels thanks to the 0.5V supply.

The figure of merits for LNA such as NEF and NEF^2V_{DD} are comparable or better compared with other state-of-the-art works.

3.3.2 Programmable gain amplifier design

Between the LNA and Δ-ΔΣ ADC a PGA is inserted to provide more gain, an additional bandpass filtering and to drive the following Δ-ΔΣ ADC as shown in Figure 3-4. The transistor level schematic of the PGA and its transconductor is depicted in Figure 3-6. The voltage gain of the PGA can be adjusted by varying its feedback factor using the two switches S_1 and S_2 as shown in Figure 3-6. Conventionally, a common approach for the gain adjustment is to use a switch to connect or disconnect a feedback path. However, this scheme may bring about signal distortion at very low frequency due to the reactance created by the off-state resistance of the control switches. To avoid that distortion, the "flip-over-capacitor" scheme is adopted from [50]. As shown in Fig. 5, there are two control switches, their complementary, and the associated with each flip-over-capacitor, C_1 and C_2. Hence, functions either as a part of the input capacitor or as a part of the feedback capacitor according to the states of the two control switches. Thanks to this operation, the off-state resistance of the corresponding switch is always excluded from the feedback loop and behaves only as a negligible load to the LNA and the PGA, posing no distortion to the frequency response of the PGA. By flipping each to either input or output node, four gain settings: 0, 3, 9, 15dB can be achieved. The capacitance values of C_0, C_1, C_2 and C_3 are 850 fF, 385 fF, 165 fF and 85fF, respectively. The DC biasing points of the PGA are regulated by a fully balanced pseudo-resistor with a fixed resistance (R_{pseudo} in Figure 3-5 (a)). The R_{pseudo} is large enough to ensure that the resultant high-pass corner frequency is

lower than the lowest high-pass corner frequency of the LNA. To guarantee enough driving capability for the following Δ-ΔΣ ADC, the transconductor for the PGA has a buffer. For the buffers, the low threshold voltage transistors are employed. The bias currents for the OTA are mostly concentrated on the buffer. To provide enough phase margin and minimize the change of high frequency corner for the different gain setting, the compensation capacitors are also selectable according to gain setting as shown in Figure 3-6 (b). The power consumption of the PGA is 0.42 μW and it is able to provide about 800 mV_{pp} output swing.

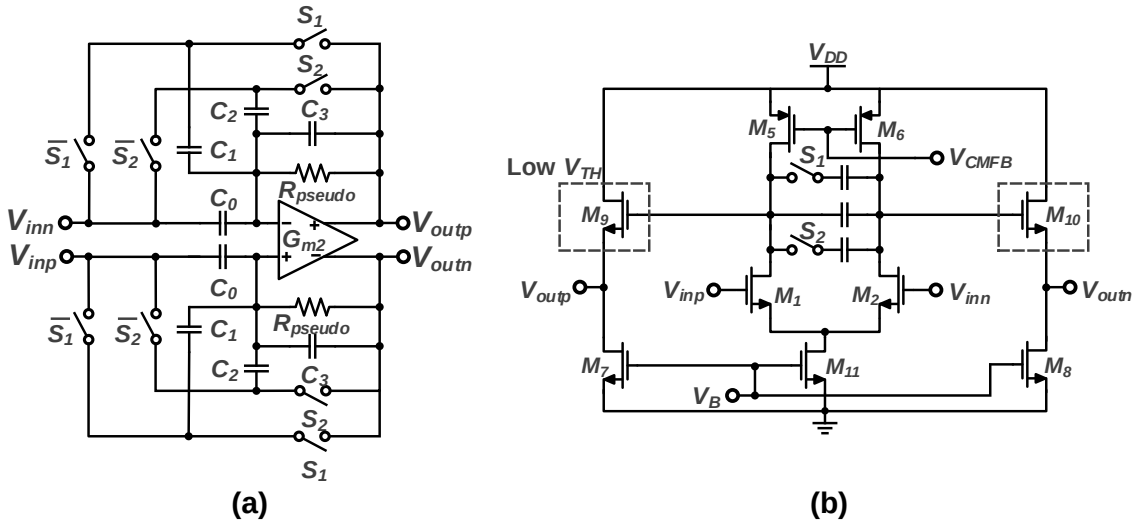


Figure 3-6 Schematic of PGA (a) and transconductor (G_{m2}) used for PGA (b)

3.3.3 Δ-ΔΣ analog to digital converter

Figure 3-7 (a) shows the block diagram of the proposed CT Δ-ΔΣ ADC where a Δ-modulator is cascaded with a ΔΣ ADC. The structure of the Δ-modulator is a first order and that of the ΔΣ modulator is a first order, single loop, feedback architecture with a single-bit quantizer. The feedback DACs are realized with non-return-zero (NRZ) signals. The feedback coefficients, k_1 and k_2 are determined as 1 and 1.5 by applying the inverse

invariant transformation (IIT) to the discrete second order ΔΣ ADC with the NRZ feedback waveform [51]. Even though the loop filter looks like second order, as shown in Figure 3-7 (a), the overall noise transfer function (NTF) of the Δ-ΔΣ ADC is same as the first order ΔΣ modulator considering off-chip integration (Σ modulation) to restore the original input signals. Thus the overall quantization noise power has similar form of the first order ΔΣ modulator given by (5)

$$\overline{e_n^2} = \frac{\Delta^2 \pi^2}{36} \cdot \frac{1}{OSR^3} \tag{2-5}$$

where Δ and OSR is the quantization step and oversampling ratio (OSR), respectively [52]. However, compared to the first order ΔΣ modulator the quantization step is much smaller since it is nothing but the prediction error from the Δ-modulator. In the proposed CT Δ-ΔΣ ADC, the quantization step is bounded (here, δ in Figure 3-7 (b)) within the maximum derivative of the input signal times sampling interval ($T_S = 1/f_S$). As shown in Figure 3-7 (b) quantization noise becomes smaller, therefore lower noise floor coming from the quantization process can be achieved. Figure 3-8 (a) shows the numerical simulation for the signal to quantization noise ratio (SQNR) of a first, second order ΔΣ modulator, and a Δ-ΔΣ ADC by changing OSR from 4 to 64. For the Δ-ΔΣ ADC and ΔΣ ADC, $\delta = 0.025$ and $\Delta = 1$ is used, respectively. As indicated the slope of the SQNR improvement is same as the first order ΔΣ modulator (+9dB by doubling OSR), however, Δ-ΔΣ ADC has about 30dB offset compared to the first order ΔΣ modulator thanks to the reduced quantization noise. The SQNR improvement is even more dramatic once the energy of input signal is bounded within smaller bandwidth. Another numerical simulation (fixed OSR = 32) with the four input signals where their bandwidths are different but the total energies are same is shown in Figure 3-8 (b). To aid understanding, the conceptual probability density

functions (PDF) are also depicted in Figure 3-8 (b). While the SQNR of the second order ΔΣ modulator remains same, that of the Δ-ΔΣ ADC becomes higher as the bandwidth of signal decreases. This characteristic is particularly beneficial to the neural signals including LFP whose energy is mostly located low frequency region, < 600 Hz and AP which have higher bandwidth with small energy. The first order Δ-ΔΣ ADC with 32 over-sampling ratio (OSR), 800 kHz can achieve over 10b resolution in this architecture with the aid of the additional 30dB improvement from the Δ-modulator.

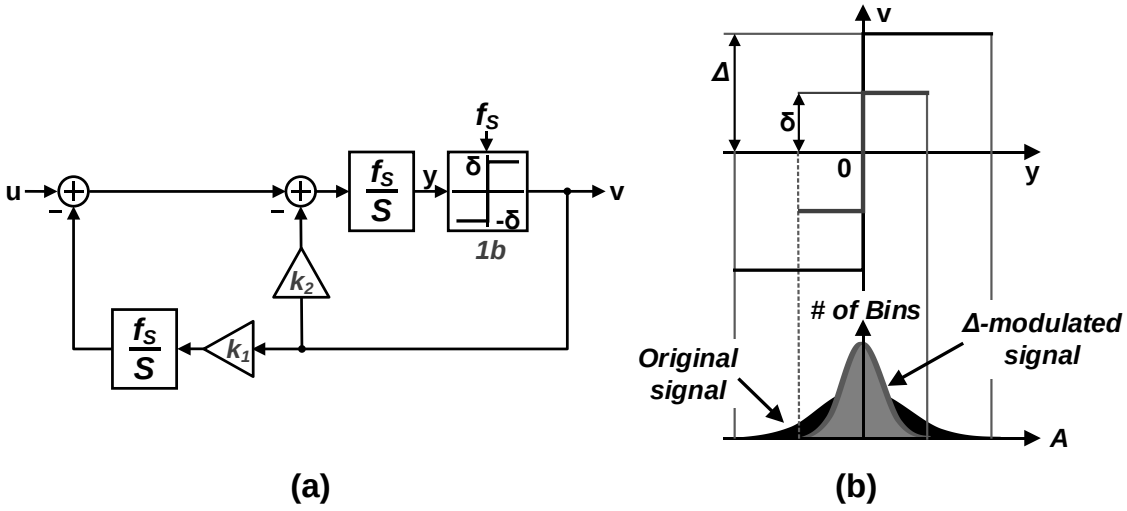


Figure 3-7 Block diagram of CT Δ-ΔΣ ADC (a) and Modified quantizer inside the CT Δ-ΔΣ ADC (b)

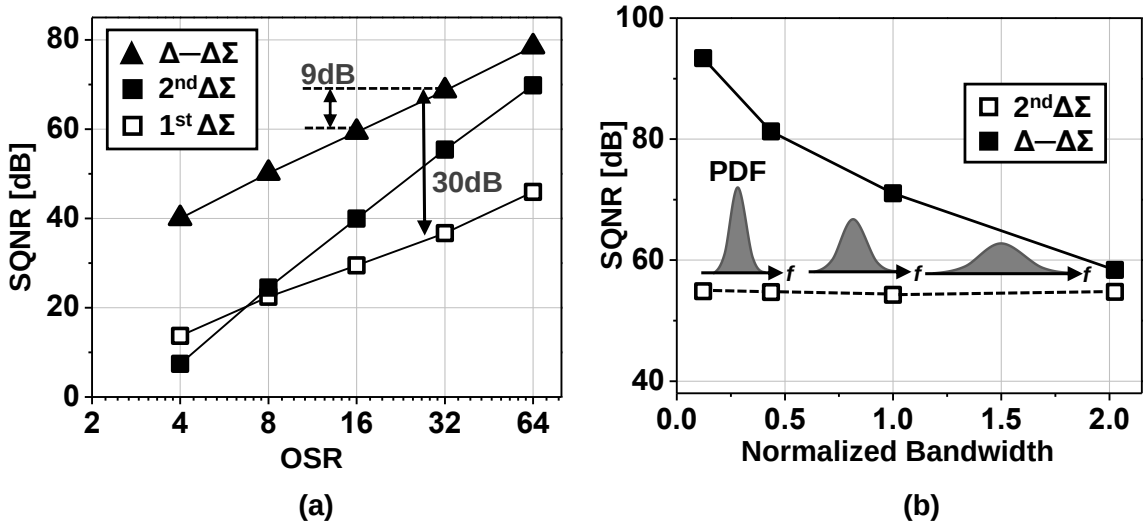


Figure 3-8 SQNR of first and second order ΔΣ, and Δ-ΔΣ ADC with different OSR (a) SQNR of second order ΔΣ modulator and Δ-ΔΣ ADC with different bandwidth signals (b)

As shown in Figure 3-4, the proposed Δ-ΔΣ ADC consists of two transconductance cells, g_{m2} and g_{m3}, two DACs, and a 1b quantizer (dynamic comparator). The DAC in the Δ-ΔΣ ADC is implemented with NMOS switches, passives, and references (V_{CM} ± 20 mV, V_{CM} = 250 mV). The passive values are R_1 = 625 kΩ, R_2 = 1.3 MΩ, R_3 = 2MΩ, and C_4 = C_5 = 2 pF. Figure 3-9 (a) shows the schematics of the first transconductor, g_{m3} and dynamic comparator in the proposed CT Δ-ΔΣ ADC. The second transconductor, g_{m4} is the simple single stage OTA, therefore it is not shown in this paper. The gain, bandwidth, and DR requirement g_{m3} is more stringent than those of g_{m4} since the first integrator using g_{m3} should generate the delayed replica of the input signals with the precision of over 60dB SNR. It consists of 3 stages with Miller compensation using passives, C_f and R_f. The OTA, g_{m3} has 79.8 dB dc gain and 1.050 MHz unity gain bandwidth with 57º phase margin while consuming 2.2 µA current. The last stage is a buffer to drive the second integrator. M_0~M_{10} are the standard 1.8 V transistors and M_{11} and M_{12} are the low V_{TH} transistor for the proper output common mode, 0.25 V (V_{DD}/2). Figure 3-9 (b) shows the schematic of the dynamic

comparator used for the 1b quantization. For this comparator 1V supply (V_{DDH}) is used since 0.5V supply makes the decision time of the comparator slow, consequently, it deteriorates the overall performance without any ELD compensation circuit [53].

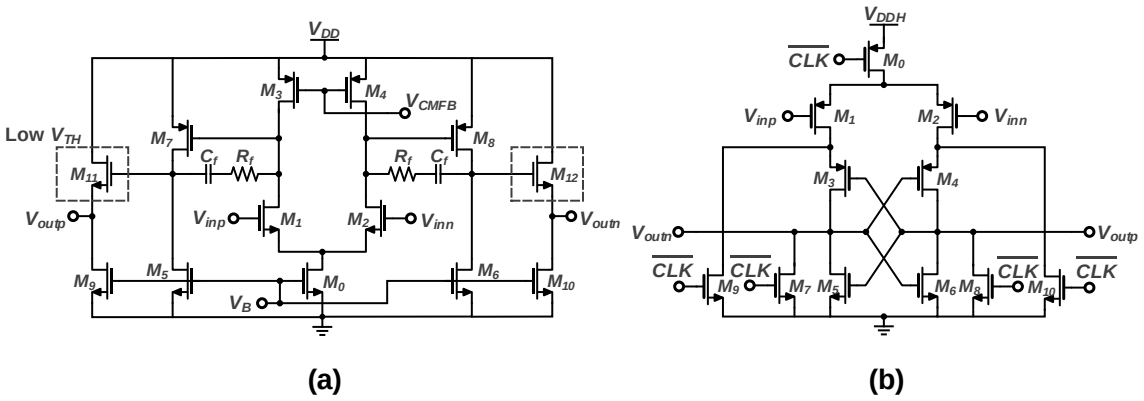


Figure 3-9 (a) Schematics of the first transconductor (G_{m3}) in CT Δ-ΔΣ ADC and (b) dynamic comparator used as 1b quantizer for Δ-ΔΣ ADC. The CMFB circuit for G_{m3} is not shown in (b).

3.3.4 Digital filter and serializer

To minimize the effort for the assembly of the whole system, it is better that all 128-channel data is serialized onto 1b. However, the 1b serialization makes the total output data rate of all 128-channel with f_S = 800 kHz become over 100 Mb/s, which incurs high power consumption for the data transmission. The better trade-off is to implement on-chip decimation filters for the each individual channels to reduce the data rate even though they require additional area overhead.

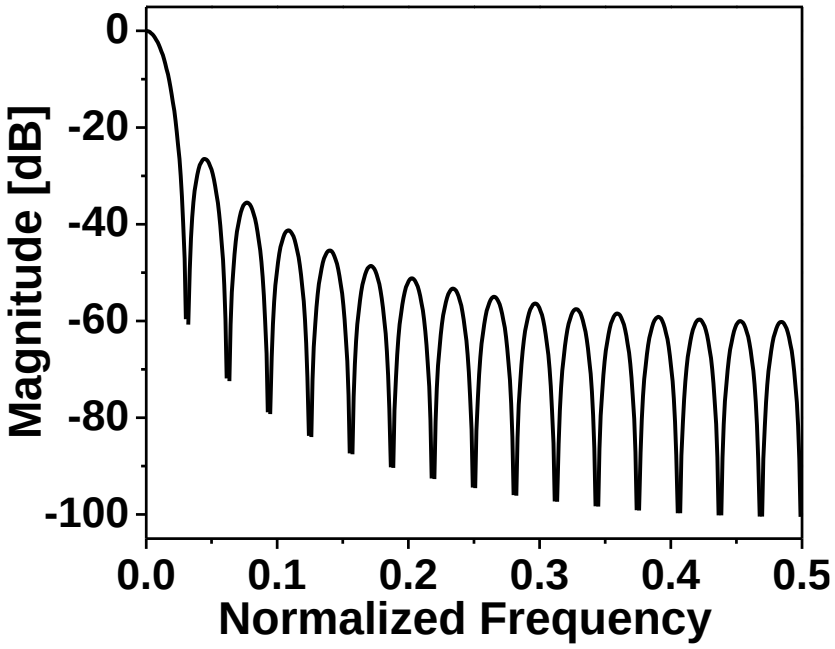


Figure 3-10 Frequency response of the $sinc^2$ filter for the decimation of the output of the Δ-ΔΣ ADC

Since overall noise transfer function considering the off-chip signal restoration is the first order, $sinc^2$ filter is able to provide the enough attenuation of the out-band noise. The equation (6) shows the transfer function of the 2nd order *sinc* filter.

$$H(e^{j2\pi f}) = \left(\frac{sinc(Nf)}{sinc(f)} \right)^2 \tag{2-6}$$

where f and N are the normalized frequency and number of AP, 32 by the OSR of the ADC. Figure 3-10 shows the frequency response of the $sinc^2$ filter. Even though the $sinc^2$ filter introduce a droop into the in-band frequency response, the overall SNR degradation is less 0.5 dB which is a minute portion of the overall over 60 dB SNR requirement. The $sinc^2$ filter is implemented using auto-place and routing (APR) and it occupies only the area of 0.00288 mm^2 (40 μm × 72 μm).

3.4 Experiment results

The 128-channel Δ-ΔΣ AFE was fabricated in 0.18 μm 1P6M CMOS process from TSMC. Figure 3-11 shows a microphotograph of the fabricated chip. The 128-channel AFE occupies 5950 μm × 1500 μm including digital interface for the data serializer and filters, and programming registers for the test patterns. There are four same bias circuits and each is shared by 32 channels. The enlarge figure shows a single-channel AFE which occupies 45 μm × 1120 μm with 5 μm channel separation to reduce crosstalk between channels (2.5 μm for left and right).

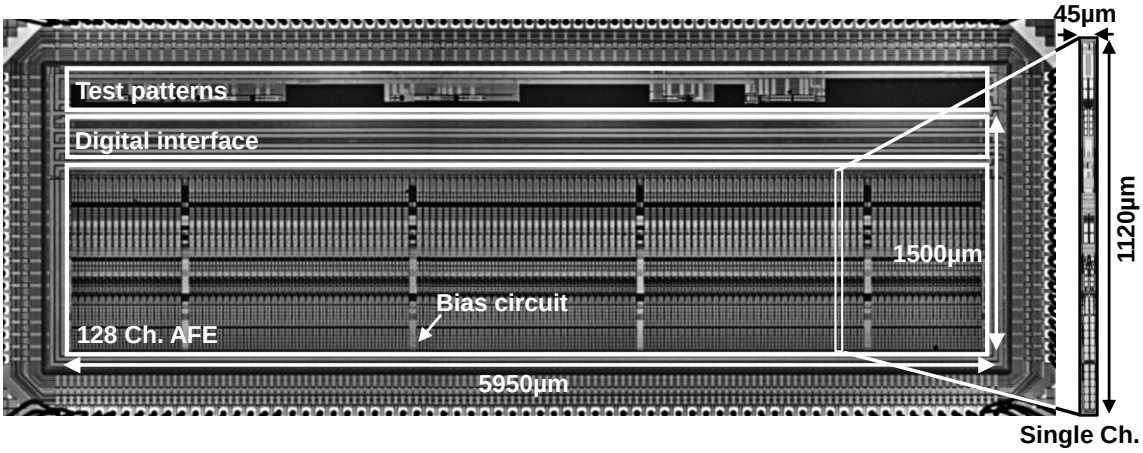


Figure 3-11 Die microphotograph of 128-channel AFE including test circuits. A single channel is enlarged in the right side

3.4.1 Measurement of LNA, PGA, and ADC

The measured frequency response of the LNA is shown in Figure 3-12. It realizes a mid-band gain of 38.5 dB from 0.4 Hz low corner frequency (f_L) 10.9 kHz high frequency corner (f_H). The IRN of the LNA, measured using the Agilent dynamic signal analyzer 35670A, with the maximum bandwidth is shown in Figure 3-13. The integrated IRN from 0.5 Hz to 12.7 kHz is about 3.32 μV_{rms} which is smaller than 10 μV_{rms}, the general noise

requirements of the neural amplifier [36]. Figure 3-14 shows the measurement of total harmonic distortion (THD) with considering up to 20 harmonics. The input sine wave was generated with the SR560 (Standford amplifier) with 10× gain and 30 mHz to 300 KHz bandpass filter and the output is measured at the output of the LNA with Agilent 35670A. The measured THD is 0.516%.

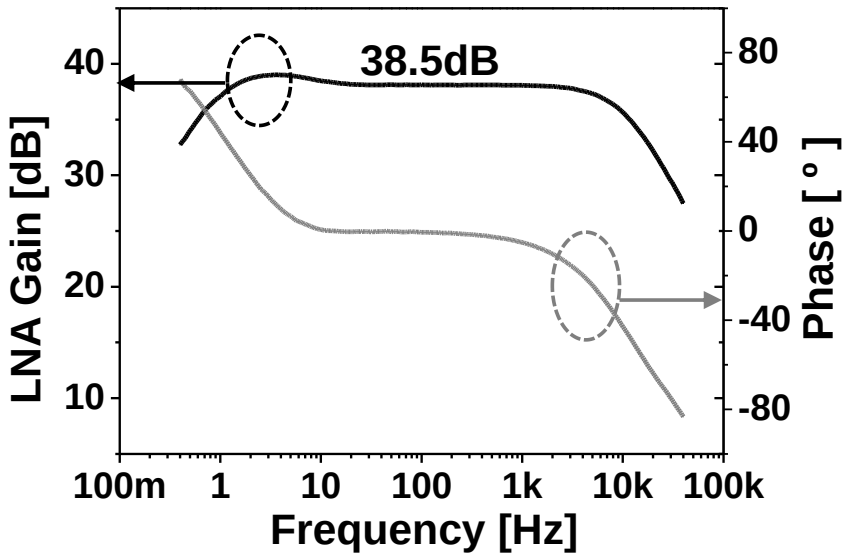


Figure 3-12 Measured gain and phase response of LNA

The measured gain of PGA is shown in Figure 3-15. The voltage gain of the overall system can be adjusted by changing the gain of the PGA, where 4 level voltage gains of -1.05 dB, 2.47 dB, 8.22 dB, and 14.36 dB can be selected, as depicted in Figure 3-15. Thanks to the additional selectable capacitors in the PGA, f_H does not changed by the gain settings. Figure 3-16 depicts 32,768-points FFT analysis of the Δ–ΔΣ ADC. The second largest spur (3^{rd} harmonics) is measured at -86dB. The spurious free dynamic range (SFDR) is about 74dB and the calculated signal to noise and distortion ratio is 67.4dB, equivalent to 10.9b effective number of bit (ENOB). The output band noise is shaped with +20dB/dec line

which is the typical frequency characteristic of the ΔΣ ADC. The power consumption of the ADC is measured to be about 1.68μW from 0.5 V for the analog block and 1.0 V for the digital block (a comparator) and its FoM is 35.2fJ/C-s.

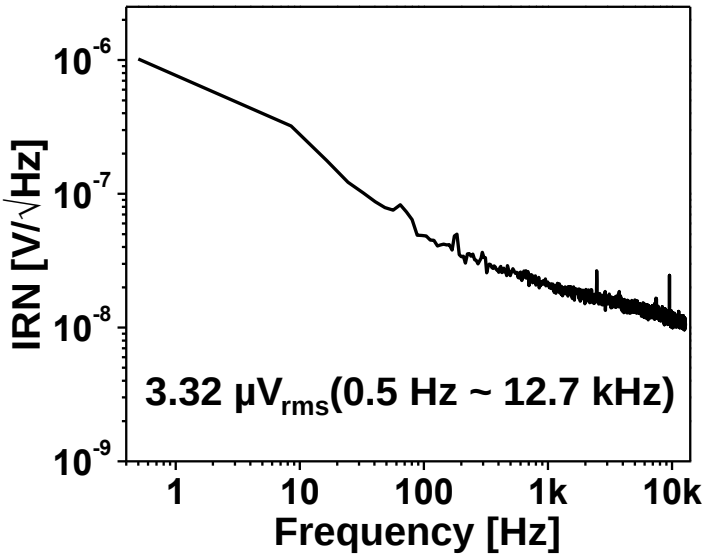


Figure 3-13 Measured input referred noise spectral density of LNA

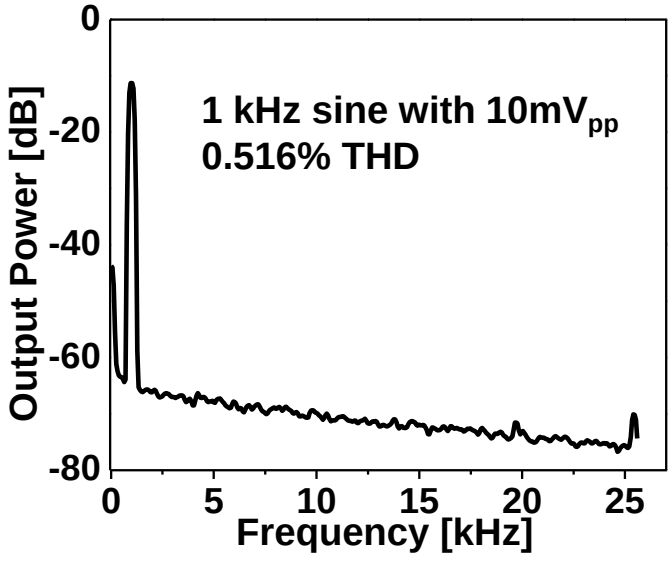


Figure 3-14 Measured total harmonic distortion of LNA

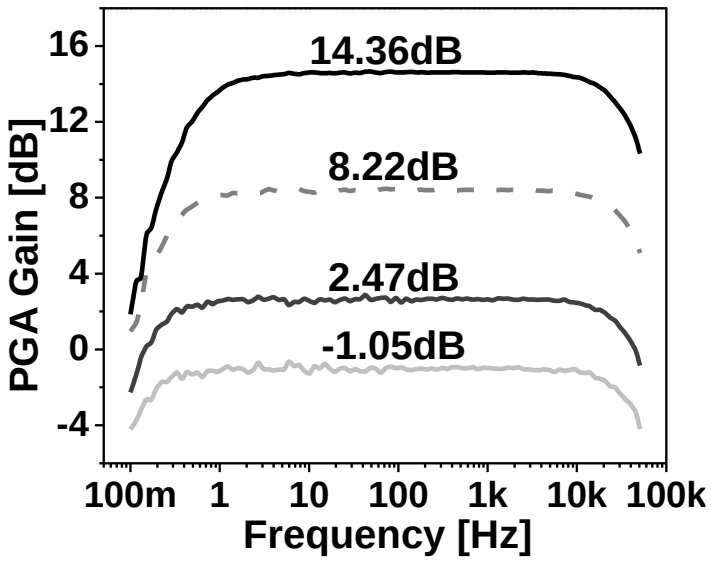


Figure 3-15 Measured frequency response of PGA

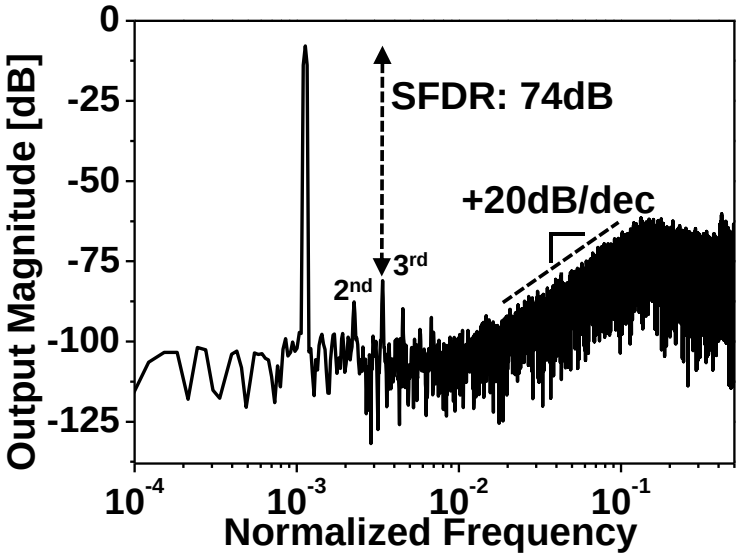


Figure 3-16 32,768-point FFT of the output of Δ–ΔΣ ADC

3.4.2 *In-vivo* measurement

For the *in-vivo* signal measurement, a silicon multi-shank probe was implanted in the neo-cortex of a rodent and externally interfaced with the fabricated chip. Figure 3-18 (a) shows about 0.6 s snapshot of the recorded LFP and AP (spike) and Figure 3-17 (b) shows

the only AP after band-pass filtering of the left plot with 300 Hz and 7.5 kHz low and high frequency corner. For clear visualization, the part of the recorded LFP is enlarged. As shown in the left waveform, the small spikes which have amplitude around 200 μV are superimposed on the relatively large fluctuation of LFP which has around 1.2 mV. Figure 3-18 depicts ~1s neural signals from the 128-channels to check the feasibility of the multi-channel recording. As same as Figure 3-17, the waveform of both of the LFP and spikes are plotted in Figure 3-18 (a) and the band-pass filtered spikes are depicted in Figure 3-18 (b). While the amplitude of LFP span around a few mV (~2-3 mV), the spikes does around a few hundred μV (~200 μV). Figure 3-19 shows the magnitude spectra of the reconstructed and compressed neural signals from the measurement. As expected, the recorded raw signal exhibits a ~1/*f* slope since exhibiting the energy of the signal are mostly located in LFP regime. The full variance of 75.3dB is modulated on the chip through the built-in Δ-modulator into the 48.2dB, then the signal is fully recovered by the Σ-modulator in the off-chip as shown in Figure 3-19.

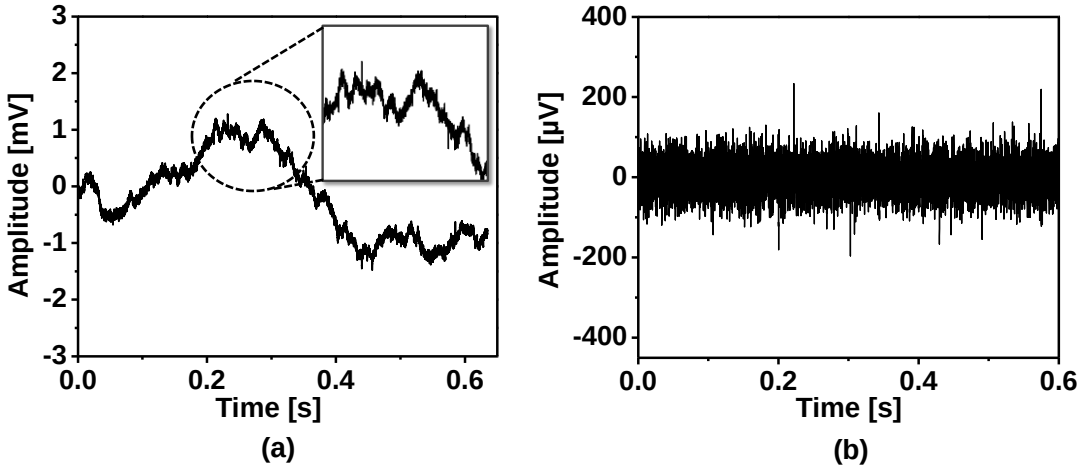


Figure 3-17 0.6 second snapshot of In-vivo measurement for LFP and AP from a rodent

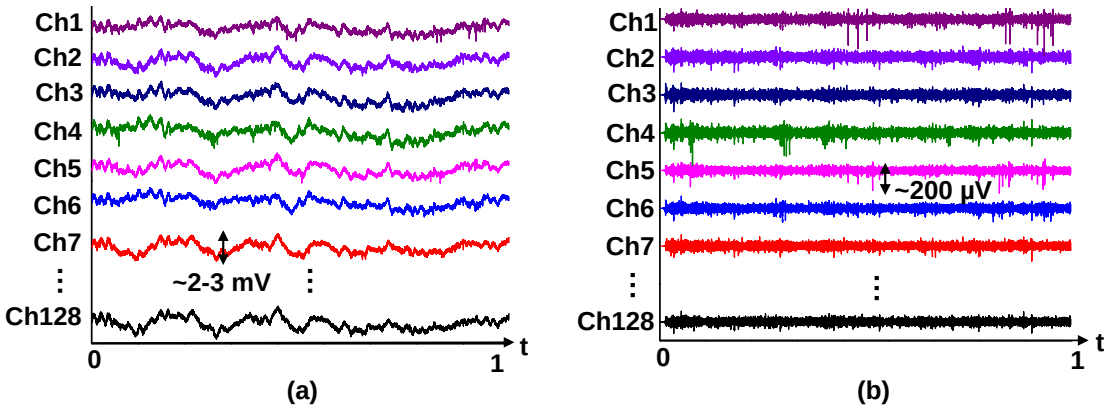


Figure 3-18 *In-vivo* data from multi-channel recordings: (a) broadband neural signal including LFP and spikes and (b) bandpass filtered spikes in the post-processing

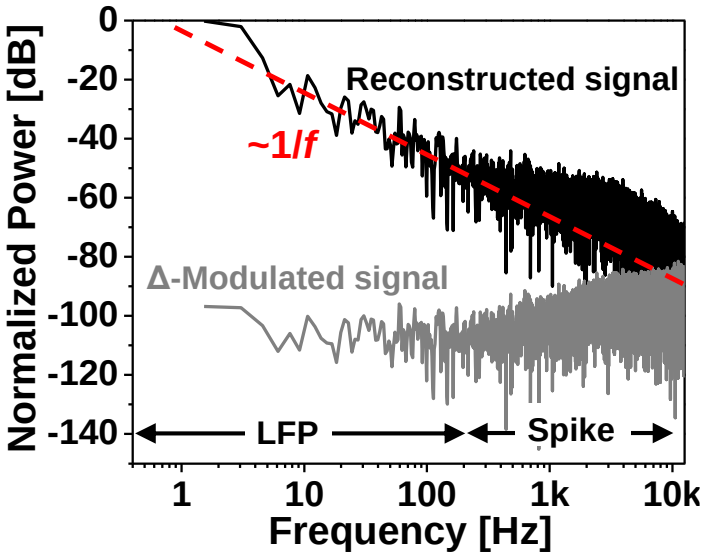


Figure 3-19 Power spectra of the recorded and reconstructed neural signals from *in-vivo* measurement

3.4.3 Comparison metric

A performance comparison with state-of-the-art work is summarized in Table I. For a fair comparison, we define a Ch. FoM (3-7)

$$Ch.FoM = \frac{P_{Ch}}{2^{ENOB,Ch} \cdot f_S} \tag{3-7}$$

where P_{Ch} is the total power consumption of a single channel AFE and *ENOB,Ch* is the measured ENOB from the output of the AFE, not the individual ADC. Even though the Ch. FoM has same analytic form as Walden's FoM [54] used for judging the performance of ADCs, it can tell us the performance of an AFE by simply replacing the power and ENOB with those of an AFE. Definitely, larger ENOB with smaller power consumption indicates better performance. The gain of an AFE does not affect the Ch. FoM since it deteriorates the SNDR of the final output once the gain increases to the point where signal is saturated and starts being distorted. From the calculation, we achieve 96.82fJ/C-s Ch. FoM with 10.3b ENOB and 3.05 μW power consumption of a single AFE. We also evaluate another new FoM, called Energy-Area FoM (E-A FoM). The E-A FoM indicates how efficiently both of the area and energy are used for the realization of a single AFE. With the two defined FoMs, our work and the other state-of-the-art works are displayed together in Figure 3-20. The bottom left is the desired direction to pursue as indicated with the arrow. To our knowledge, our work achieved the best E-A FoM of 4.84fJ/C-s·mm^2.

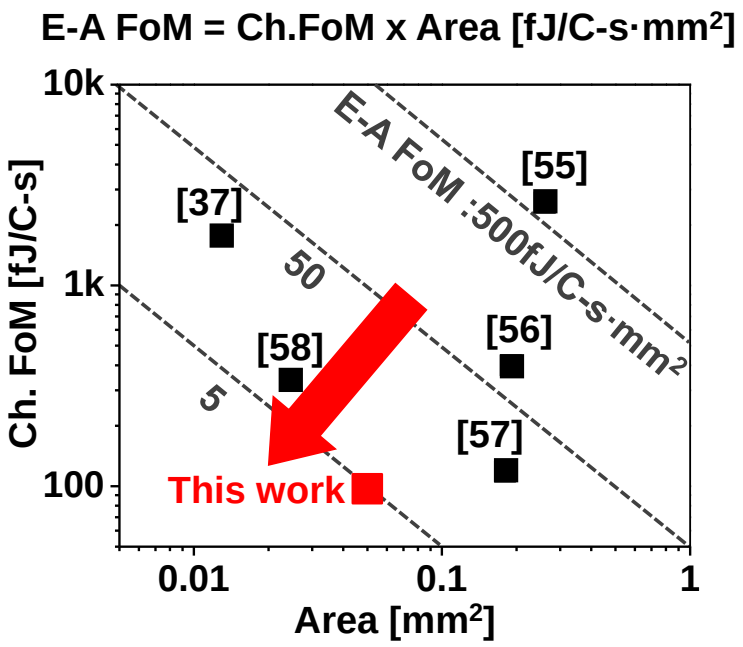


Figure 3-20 E-A FoM with other state-of-the-art works

3.5 Summary and chapter conclusion

We reports an energy and area-efficient 128-channel modular architecture incorporating Δ-modulated ΔΣ signal acquisition for 1,024-channel parallel brain activity monitoring platform. The frontend circuit employed the spectrum equalizing and Continuous Time (CT)-ΔΣ quantization to use the inherent spectral characteristics of brain signals in energy and area efficient manner. With those techniques, the dynamic range of the acquired signals has been compressed by 27dB and the frontend achieved the smallest energy-area product, 4.84 fJ/conv·mm^2, which is the most critical figure of merit for the platform of the 1,024-channel parallel recording. The fabricated circuits consumes 0.05 mm^2 area and 3.05μW per channel from a 0.5V supply and exhibits 63.8dB SNDR, 3.02 NEF, and 4.56NEF^{2}VDD.

Table 2-1 Performance comparison with the state-of-the-art AFEs

		[55]	[37]	[56]	[57]	[58]	**This work**
Power/Ch. [µW]		68	5.04	7.02	0.73	1.84	**3.05**
Input-referred noise [μV_{rms}]		2.2	4.9/4.3	3.2	3.2	7.5	**3.32**
Bandwidth [kHz]		10	10	6	10	*10	**10.9**
NEF		4.5	5.99	3.08	1.57	3.6	**3.02**
NEF^2V_{DD}		*24.3	17.96	17.13	1.12	12.9	**4.56**
CMRR		–	75	60	73	–	**> 60dB**
ADC ENOB		9.72	7.15	9.2	8.27	8.2	**10.9**
ADC sampling freq. [KHz]		31.25	20	*30	*20	20	**25**
Ch. Multiplexing		No	No	Yes	Yes	Yes	**No**
FoM [fJ/C-s]	ADC only	42	219.7	84	22	34	**35.2**
	**Ch. FoM	*2580	*397.87	*1774.3	*118.24	*312.85	**96.82**
E-A FoM [fJ/C-s·mm^2]		*670.8	*23.07	*75.60	*21.28	*8.07	**4.84**
Recording channels		96	1	54	100	64	**128**
Core supply voltage [V]		1.2	0.5	1.8	0.45	1.0	**0.5**
Area/Ch. [mm^2]		0.26	0.013	0.19	0.18*	0.0258	**0.05**
Technology		0.13µm	65nm	0.18µm	0.18µm	65nm	**0.18µm**

*Estimated, **$P/f_S \cdot 2^{ENOB}$, P and ENOB are from a single channel

CHAPTER 4
ON-CHIP NEURAL SIGNAL COMPRESSOR WITH 128-CHANNEL ANALOG FRONT END

4.1 Introduction

Over the past few years, the main focus of the design of neural recording microsystems has been increasing the number of parallel recording sites in more area-and energy efficient ways [5][6][55][56][58][59][60]. Current state-of-the-art neural recording chips can be realized with sub-μW (0.73 μW) of power and 0.013 mm^2 of area per channel consumption [37]. As the numbers of the parallel recording sites and the implementation efficiency have steadily increased, the data rate of the recorded signals has constantly increased as well. The increased data transmission rate, *i.e.* larger data bandwidth, inevitably leads to higher power consumption for the data handling and transmission in the digital interface circuits, and it frequently overwhelms the power consumed in the recording circuits. For instance, 10b data stream with 30 kHz sampling rate from only 128 channels forms about 40 Mb/s data rate when it is serialized into a single bit. Assuming that 30 pF loading capacitance exists in an I/O buffer and 3.3 V standards I/O transistors are used for the buffer, about 12 mW of power is dissipated in only that buffer. When this amount of power is compared with the current state-of-the-art AFE's power consumption (usually < 10 μW per channel), it is more than 20 times larger. If the ultimate goal is to build a neural recording microsystem with 128, 256, 1,024 or more simultaneous recording channels in the future, power consumption in the digital interface will become an ever-greater part of the total power consumption. Otherwise, the data stream should be parallelized into bulky wire

bundles while enduring the high possibility of tissue infections from tethering and failures of the experiment with freely behaving animals. For the fully implantable systems to which the data stream should be sent wirelessly, the situation become even worse since the wireless data transmission results in higher power consumption than the wireline.

4.2 Investigation of broadband neural signal

4.2.1 Revisiting to neural signal characteristics

As mentioned in chapter 3, to understand in-depth brain activity, both LFP and AP should be recorded simultaneously [34]. The sum of the two different signals constitutes wide bandwidth from sub Hz (typically, 0.1 ~ 0.5 Hz) to 10 kHz (typically, 5 ~ 7.5 kHz) and wide dynamic range from a few tens of μV to mV (100 μV ~ 4 mV). The neuroscientists who explore extensive brain activities and try to understand the system of neuronal connections require high-quality recordings with more than 10b resolution and < 10 μV_{rms} input referred noise (IRN), from as many channels as possible in a small volume of brain. Thus system requirements become difficult to meet when incorporating many high-quality recording channels with given energy and area constraints. Our recent work for the recording analog front-end (AFE) strategically used the inherent neural signal characteristic to simultaneously reduce the area and energy consumption in the implementation while preserving high performance [5]. Here, we will take a look again at the neural signal characteristic used in chapter 2, then investigate other inherent features of neural signals to further explore the possibility of signal compression.

Broadband neural signals, that is, LFP and AP on top of the LFP (LFP + AP), have inherent ~1/*f* characteristic in their spectra [43][44]. The ~1/*f* characteristic (high temporal

correlation) means that the broadband neural signals have higher energy in their low frequency than in their high frequency, which entails that the LFP have similar characteristic to the broadband neural signals because most of the energy of the neural signals comes from the LFP. In other words, once the LFP are separated from the broadband signals by a proper filtering, they have even higher temporal correlation since all of the high-frequency contents of the broadband neural signals are filtered out. This high temporal correlation can contribute to the data compression by taking the temporal difference of the signals. In addition to this known ~1/*f* characteristic, we can surmise another inherent characteristic of the LFP when we observe the characteristics of the LFP from multi-channel neural recordings.

Slow varying LFP are aggregated sum of the signals from local neurons. Since they are so-called "average values" from multiple recording sites, the cross-correlation between the local LFP from the neighboring recording sites (spatial cross-correlation) might be high. In particular, since the distance between recording sites becomes smaller to record the activities of larger numbers of neurons in a given volume, the spatial cross-correlation between the LFP from neighboring channels becomes higher. As most of the lossless data-compression schemes use the statistical redundancy to represent data, this high spatial cross-correlation in the LFP also suggests that there is possibility of reducing the data rate of LFP using this characteristic. Figure 4-1 (a) shows the spatial cross-correlation of the neighboring LFP from 16 recording sites from two different probe shanks (for shank 1 and 3, each has 8 recording sites). The signals were recorded from the CA1 region of a rat's brain in Buzsaki's laboratory (New York University, NY). The total signal duration used for this calculation is 2 seconds, and they are low pass filtered with 625 Hz corner

frequency by the commercial software (MATLAB, v.2013). The distance between the recording sites increases linearly from 25 µm to 50 µm (25 µm between sites 1 and 2, and 50 µm between sites 7 and 8) as shown in Figure 5-1(b). Figure 4-1(a) demonstrates high spatial cross-correlation (mostly, > 0.9).

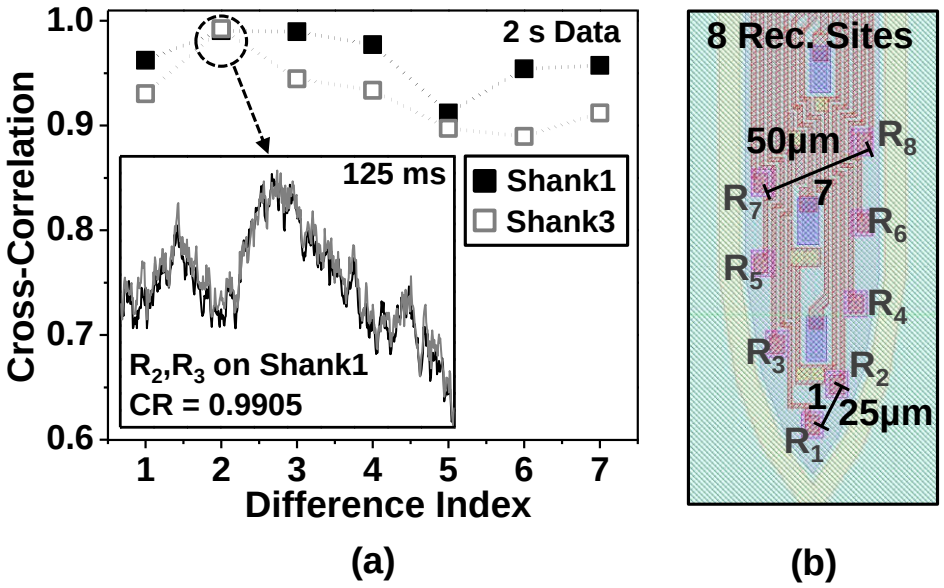


Figure 4-1 (a) Cross-correlation of the LFP from CA1 region of a rat's brain and (b) probe configuration used for the neural recordings

To confirm the previous arguments, we performed another numerical calculation to obtain the compression ratio (*CR*) of the LFP. The temporal difference (temporal correlation, equivalently ~1/*f*) and spatial/temporal difference (spatiotemporal correlation) are applied for the LFP. For the calculation of the *CR*, the ideal entropy encoder, *i.e.* Huffman encoder (a built-in function in MATLAB, v.2013) is used and the data is quantized into 10b. The *CR* is given by

$$CR = \frac{N_o}{N_e} \tag{4-1}$$

where N_o and N_e are the numbers of bits in the original data sets and encoded data sets, respectively. Figure 4-2 shows the *CR* of the LFP from 32 recording sites where eachof eight sites are positioned on a single shank and there are four shanks. The 15, 16, 23, 24 indices are omitted because they are too noisy to extract the useful information. In this case, the length of the LFP is 2 seconds. The signals are filtered and quantized with the same conditions used for Figure 4-1. The average *CR*s when only temporal and both spatial/temporal correlation are applied are about 4.3 and 5.8, respectively. As expected, the temporal correlation characteristic provides moderate compression; however, higher compression is achieved by employing the spatiotemporal correlation.

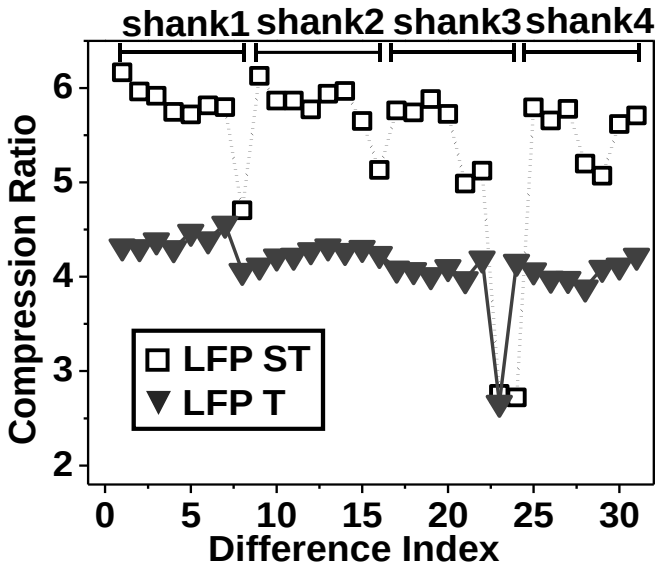


Figure 4-2 *CR* of spatial and temporal (ST) difference and temporal (T) of the LFP

To further study the neural signal characteristics, another numerical calculation was performed: *CR*s for (1) the LFP using spatiotemporal correlation, (2) the LPF using temporal correlation, and (3) the broadband neural signals (again, LFP + AP) using spatiotemporal correlation. The last calculation is intended to prove the assumption that the

compression based on the temporal correlation might be more effective if the signals has only low frequency components. As shown in Figure 4-3, while the compression using the spatiotemporal correlation of only the LFP exhibits the best performance, the *CR* using the spatiotemporal correlation of the broadband signals is three times lower than the best case. This degradation can be explained with the fact that the AP are single cell activities (unit activities) and their firing rates are sparse and rather uncorrelated to the each other relative to the LFP. At times, the propagating spikes can be detected in several recording sites simultaneously; however, their relative amplitudes and shapes are somewhat different, since the amplitudes and phases of the propagating spikes are attenuated and lagged with a faster pace than LFP while diverging. Based on the above numerical calculation and argument, we can conclude that the LFP and AP should be separated for better compression because of their different characteristics.

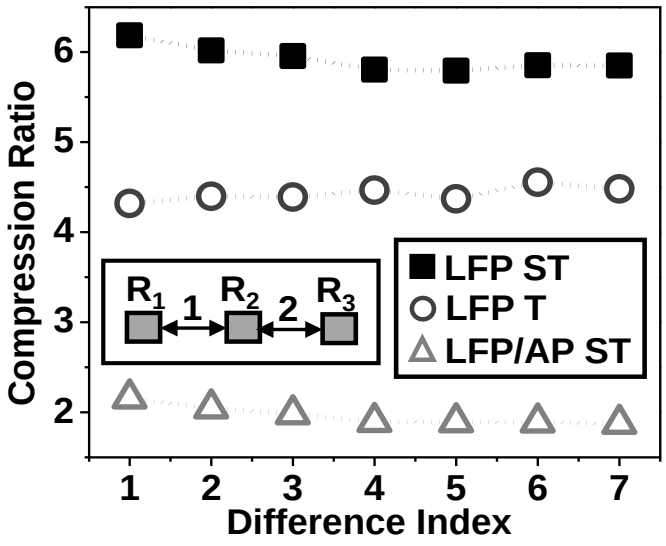


Figure 4-3 *CR* of the LFP and AP through numerical calculation by applying (1) spatiotemporal correlation (LFP) and (2) temporal correlation (LFP) (3) spatiotemporal correlation (LPF+ AP)

Since AP should be separated from LFP for higher *CR* of LFP, another compression strategy for AP must be developed. Generally speaking, the AP is very sparse in nature. The principal mechanism of information transmission in extracellular neural recordings is the change in firing rate of AP in the individual neurons. The AP is the short waveform of a few milliseconds of duration whose rate of occurrence ranges from a few tens to hundreds per second. Figure 4-4 shows the spike counts from the 32 channels, 2-second data. From the broadband neural signals, the AP are extracted using a software filter (MATLAB, v.2013) with 750 Hz high-pass cutoff frequency and simple dual amplitude thresholding [61]. Indeed, the average firing rate of the AP is only 61, which is very sparse compared to the whole 40,000 points of the data set. Equivalently, we can say the duty of the AP is only about 3%. A simple hand calculation can provide the estimated bandwidth reduction. For instance, assuming a case of 3-ms duration for each AP, the calculated *CR* is 10.93(2000 ms / (13 ms × 61) ≈ 10.96).

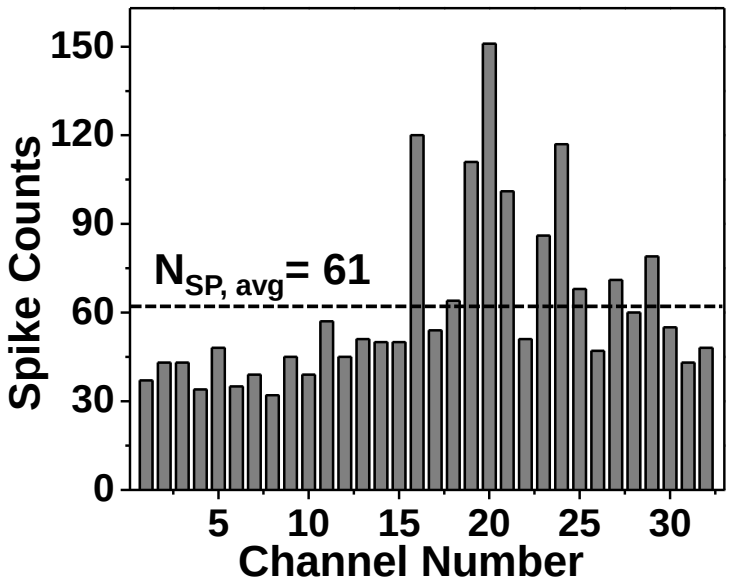


Figure 4-4 Occurrence of AP from 32-channel, 2 second neural signal

4.2.2 Recent works for compression of LFP

Intensive research has been performed on the lossy or lossless bio-signal compression schemes [62][63][64]. In some cases lossy compression is preferred if the distortion caused by the compression does not affect the relevance of the restored signals. For instance, lossy compression is chosen based on the sparseness of the energy coefficients of the signals [65][66]. In [67] the compression was achieved with the discrete packet wavelet transform (DPWT). [68] performed modeling with antireflection all-pole filters while assuming the electromyography (EMG) signals are wide-sense-stationary (WSS). For the case of the lossless compression, the electroencephalography (EEG) signals are compressed using spatiotemporal correlations [69]. However, there has been relatively little research into LFP. The validity of our approach to compressing LFP is rooted in observing that the basic principles of compression are similar; either spatial or temporal correlation, or both, in terms of time domain or frequency domain.

4.2.3 Recent works for compression of AP

Many of the current recording systems dedicated to the measurement of AP exploit this sparseness of the AP to reduce the data transmission rate. In those systems, filtering LFP out of the broadband recording is performed first, and then the detection and separation of the AP from the background signal (mostly noise) is achieved through the analog, digital, or mixed-signal AP detectors. After the AP are detached from the background signals, only the timing information of the AP such as time stamp, inter-spike intervals, or both is sent to the off-chip to minimize the data transmission rate. However, this approach inherently accompanies loss of information such as overall shapes, the derivatives or $2^{nd}/3^{rd}$

derivatives, or refractory periods of the AP, which are useful clues for follow-up studies. Therefore, to preserve AP's waveform to prevent loss of such important information, AP compression schemes often incorporate dedicated memory to store preambles of AP (data from a starting to detection point of the AP) or provide fast and slow paths for AP; the fast path is used for the detection of the AP, and the slow path with a long delay time (a few hundred μs) is used for the propagation of the replica of the AP. In [70] and [71] the analog linear delay lines were implemented using the 12-tap all-pass filters. In their implementation, the entire AP waveform can remain intact even though the AP detector senses the AP later than their occurrence, thanks to the slow path having about 640-μs linear delay through the all-pass filters. Nonetheless, the delayed waveform in their implantations was distorted due to the equiripple nature of the all-pass filter throughout its frequency response, deteriorating the signal integrity. As mentioned previously, another way to avoid such distortion of the AP waveforms is to store a certain amount of waveforms in memory before the occurrence of the AP. For a satisfactory study of the AP, the necessary length of the preambles of AP is generally known to be between 500 and 800 μs, equivalent to about 16 samples considering 20- to 31.25-kHz sampling rates for the AP. In [72], the dedicated 16 bytes (8-bit resolution and 16 data storage places) per channel memory was implemented with the on-chip static random access memory (SRAM) and showed the feasibility of the work in in-vivo environments. Recently, analog memories using on-chip capacitors are reintroduced to keep the full waveform of the AP [73].

4.3 Circuit architecture

4.3.1 On-chip neural signal neural compressor with 128-channel Δ-ΔΣ AFE

The main purpose of the proposed on-chip neural signal compressor is to reduce the power consumption from the data handling and transmission (let's call it as digital interface hereafter) for multi-channel neural recording microsystems while minimizing the additional system resources to implement it. Therefore, the on-chip compressor was implemented with mixed-signal design by using the existing functional blocks in the 128-channel neural AFE. In this work, we proposed a spatiotemporal compressor for LFP by using the Δ-ΔΣ analog-to-digital converter (ADC) [5] and applying channel-to-channel difference and entropy encoder, and a waveform extractor for AP by employing the reconfigurable analog first-in first-out (A-FIFO) scheme which reuses the on-chip capacitor bank dedicated in the in-channel ADC for the reduction of the data rate of AP.

Figure 4-5 shows the proposed architecture of the 128-channel modular neural recording AFE with the on-chip neural signal compressor. The 16 channels forms 1 group and there are 8 groups in the system. The off-chip signal restoration is also drawn to aid understanding. The AFE consists of 128 signal acquisition channels including the low noise amplifiers (LNA), programmable gain amplifiers (PGA) with the low and high pass filters (LPF/HPF), and AP waveform extractors consisting reconfigurable SAR/SS ADC including A-FIFO, Δ-ΔΣ ADCs, and digital decimation filters. There are bias circuits and two data serializers to support the data processing in the channels. Additionally, a spike detector exists for each of the spike waveform extractor. There is also a programming shift register to change the gain and bandwidth settings of the LNA, PGA, and threshold level of the spike detection. Even though Figure 4-5 conceptually visualizes the PGA/LFP/HPF block, this block actually consists of more complicated components. The details of this block will be explained in the next paragraph. All of the amplified, processed LFP and AP

from the 128-channel AFE is serialized onto two single bit wires (each for the LFP and the AP) and sent to the off-chip module for the signal restoration and storage. Since the incoming LFP are Δ-modulated in both time and spatial domains, they should be processed to retrieve the original data by applying Σ-modulation and addition between channels after decoding. However, any off-chip signal processing for the recorded AP is unnecessary since the on-chip processing for them is nothing but the waveform extraction.

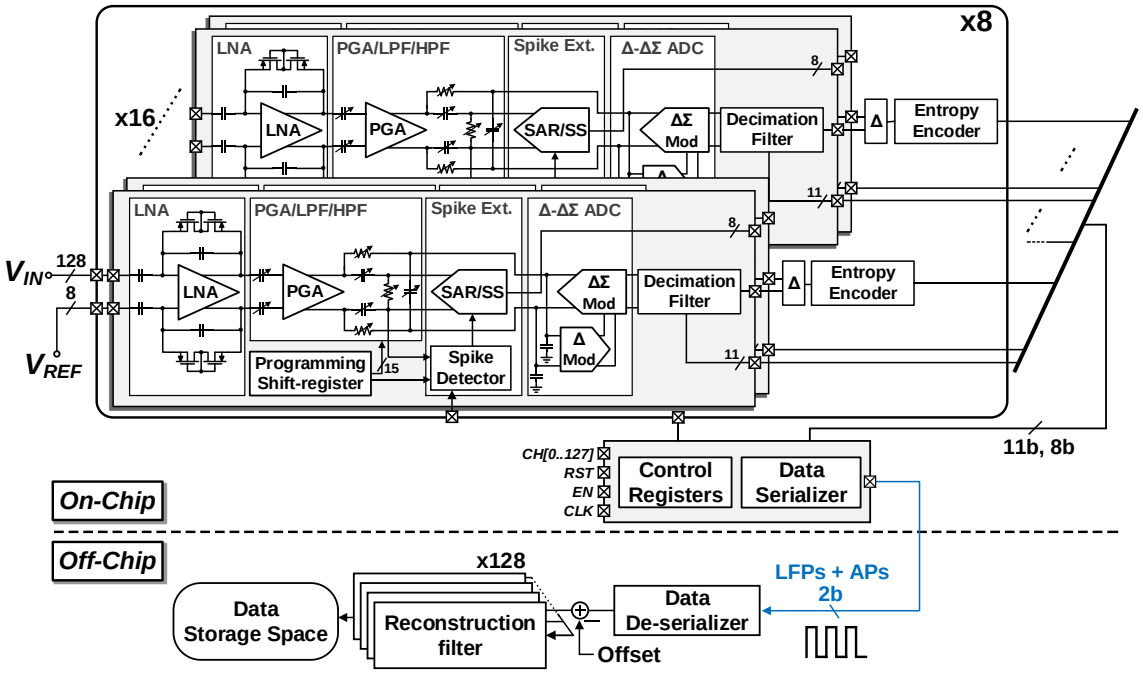


Figure 4-5 Architecture of the on-chip neural signal compressor with a 128-channel Δ-ΔΣ AFE

As shown in Figure 4-5, there are two signal pathways for LFP and AP in the proposed architecture. The incoming broadband signals (LFP + AP) are amplified and pre-conditioned (bandpass filtered with 0. 4Hz – 9.5 kHz) in the LNA, then the LPF and AP are amplified and filtered again through the anti-aliasing filter (AAF) plus switched capacitor (SC) biquad LPF and continuous time (CT) PGA, respectively. To meet various

users' demands, the gains and bandwidths for two signal chains are fully programmable within the prefixed ranges. The conditioned LFP are temporally compressed via Δ-modulation, and then digitized by a ΔΣ-ADC and sent to an on-chip decimation filter. After processed through the decimation filter, the LFP in each channel are subtracted (spatial compression) each other according to the external setting which is determined by the information based on the physical configuration of the electrical probes to maximize the spatial correlation between channels. The spatial difference is taken at every 16 channels. There are two ways to make that spatial difference as shown in Figure 4-6. One out of the 16 channels becomes the reference channel and the differences between the adjacent channels from the reference or the difference between the reference and the rest of the channels are calculated as shown in Figure 4-6 (a) and (b), respectively. The calculation results are given as

$$\Delta_n = x_{\mathrm{mod}(n+ref,N)} - x_{\mathrm{mod}(n+ref+1,N)} \tag{4-2}$$

$$\Delta_n = x_{ref} - x_{\mathrm{mod}(n+ref+1,N)} \tag{4-3}$$

where N = 16, n = 0, 1 …15 integers, and *ref* is the number of selected channel by the external programming. (4-2) and (4-3) are the results from Figure 4-6 (a) and (b), respectively.

For the preconditioned AP, the proposed architecture supports two different modes of operations; normal and compression mode. In the normal mode, raw AP data can be sent to the off-chip for the case that users need to process the recorded AP in their own software. In this mode, the in-channel SAR ADC quantizes the preconditioned raw AP. The compression mode support to transmit the detected AP waveforms (2.5 ~ 4 ms length for each AP) while providing the programmability for the length of the AP waveform to be

sent. In this compression mode, the preambles of the detected AP are stored in the on-chip capacitors which were used as the binary computation for the SAR ADC (DAC in the SAR ADC) in the normal mode. For the quantization in this mode, the in-channel SAR ADC is reconfigured into the single slope (SS) ADC since the capacitors are used as the storage place (analog first-in-first-out (A-FIFO)) for the AP preambles. Therefore, the proposed reconfigurable A-FIFO scheme does not require any dedicated space and the quantizer does not need to operate all the time in the compression mode.

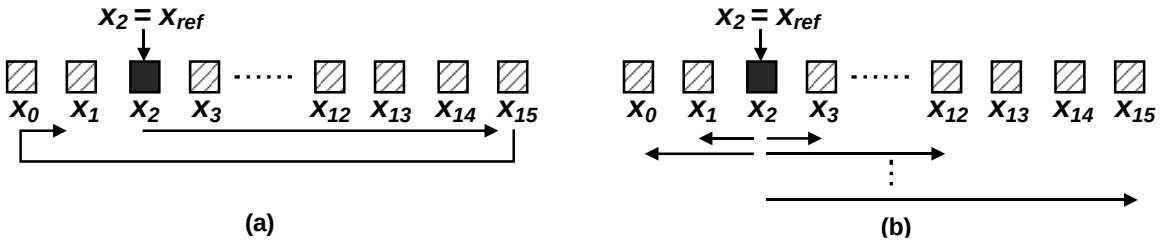


Figure 4-6 Two ways to take spatial difference of the LFP (a) difference between adjacent channels and (b) difference between reference and others

4.4 Circuit implementation

Circuit blocks for a single channel signal acquisition is shown in Figure 4-7. Since the design requirement of the LNA is same as the previous design in the chapter 3, the same block is used. The PGA block with HPF function is also designed with the same topology we previously implemented in the chapter 3 by simply adding cut-off frequency modulation with the pseudo-resistors. The details of those circuits are found in [5]. For implementing the SC-filter, as shown in Figure 4-7 an anti-aliasing filter (AAF) should be located before the SC-filter to prevent any noise folding. Since the design requirement of that AAF is not tight, a simple capacitive feedback closed-loop AAF was implemented.

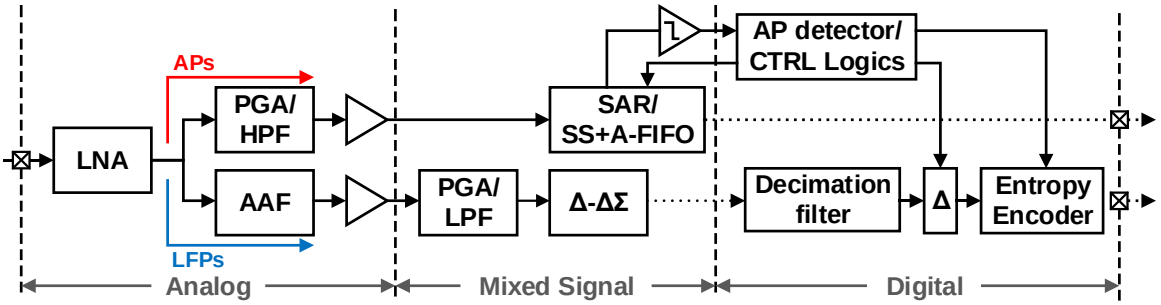


Figure 4-7 Single channel signal acquisition blocks

4.4.1 Separation of broadband neural signals

To effectively compress the incoming neural signals by applying different compression strategies to LFP and AP, the LFP and AP are separated after the initial amplification. As briefly mentioned in the section 4-3, the LFP are separated using the SC-biquad filter (the PGA/LFP block in Figure 4-7). The generally accepted corner frequency to isolate the LFP from the broadband neural signals is known as 150 ~ 300 Hz. To realize such a low frequency CT-active filter requires bulky passive components. Since the part of our main goal is to minimize the system resource, the filter is implemented using a SC-filter to reduce area consumption. Even though the transconductance (OTA) used in the SC-biquad filter definitely dissipates larger power due to their higher bandwidth requirement than CT-filters, the increased power consumption affect little to the overall power consumption as the cut-off frequency for this purpose is low enough. Figure 4-8 shows the schematic of the SC-biquad filter. The frequency of the non-overlapping clocks, φ_1 and φ_2 was set as 64 kHz considering 2 kHz sampling frequency of LPF (32 oversampling ratio (OSR)). The SC-biquad filter has the programmable gain (0, 6, 9.5 and 12 dB) and corner frequency (150, 200, 250 and 300 Hz) for users by adjusting the ratio of C_1 to C_3 and the feedback capacitors (C_{F1} and C_{F2}) of OTAs. The gain of the SC-biquad low-pass filter is determined as

$$A = \frac{C_1}{C_3} \tag{4-4}$$

where A is the gain of the low-pass filter and C_1 varies from 40.8 to 163.2 fF and C_3 is 40.8 fF. The corner frequency of the low-pass filter is also given by

$$\omega_{-3dB} = \sqrt{\frac{C_2 \cdot C_3 \cdot f_S^2}{C_{F1} \cdot C_{F2}}} \tag{4-5}$$

where ω_{-3dB} is the -3dB low-pass corner frequency and f_S is the sampling frequency of the clocks (φ_1 and φ_2).

$$Q = \sqrt{\frac{C_{F1}}{C_{F2}} \cdot \frac{C_2 \cdot C_3}{C_4^2}} \tag{4-6}$$

where Q is the quality factor of the filter and C_4 is set as 56.4 fF. According to (18) and (19), the corner frequency can be adjusted by keeping Q (constant ratio of C_{F1} to C_{F2}) as a constant (here, Q is set as about 0.7). Figure 4-9 shows the schematic of the OTA and its common mode feedback (CMFB) used for the SC-biquad filter. V_{CMFB} and V_B are the desirable common mode (CM) voltage and bias voltage, respectively. To enhance the gain and bandwidth of the OTA, the local positive feedback is used [74]. This class-A amplifier gain and gain-bandwidth product (GBW) are given by

$$A = g_{m1} \cdot R_{out} \cdot \frac{B}{1-\alpha} \tag{4-7}$$

$$GBW = \frac{g_{m1}}{2\pi C_L} \cdot \frac{B}{1-\alpha} \tag{4-8}$$

where $\alpha = (W/L)_{5,6}/(W/L)_{3,4}$, $B = (W/L)_{3,4}/(W/L)_{7,8}$, and g_{m1} and R_{out} is the OTA of M_1 and the output resistance of the OTA, respectively.

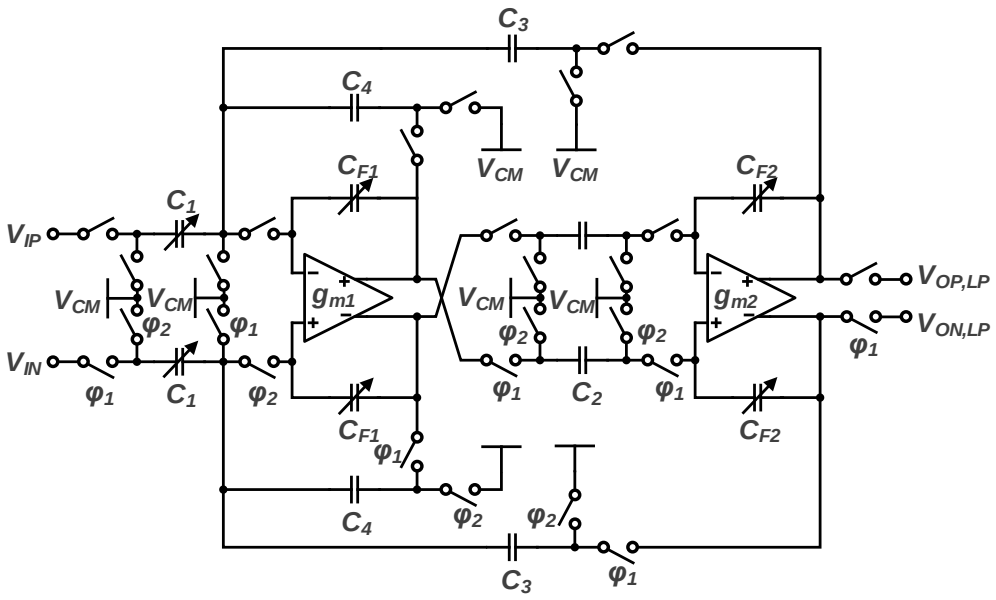


Figure 4-8 Schematic of the SC-biquad filter. V_{CM} is the common mode voltage and φ_1 and φ_2 are non-overlapping clocks

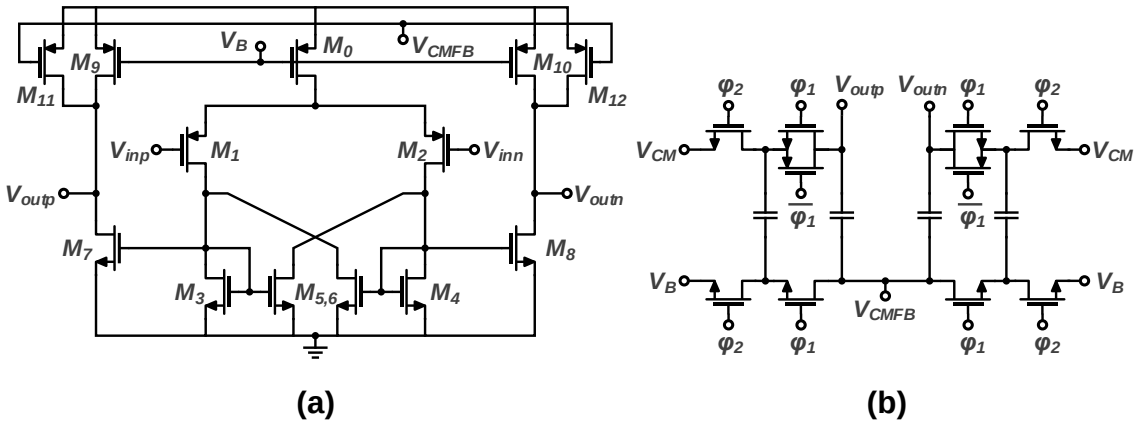


Figure 4-9 The OTA used in the SC-biquad low-pass filter (a) and the CMFB circuit for the OTA

The overall power consumption the SC-biquad filter is 0.444 μW and the AAF only consumes 0.045 μW.

4.4.2 DT-Δ-ΔΣ ADC

For the quantization of the recorded LFP, a discrete time (DT) Δ-ΔΣ ADC is employed. Figure 4-10 (a) shows the block diagram of the DT Δ-ΔΣ ADC where K_Q is the scaling constant to reduce the quantization noise. Figure 4-10 (b) depicts the quantization noise reduction in the Δ-ΔΣ ADC conceptually. Its circuit architecture is same as the previous CT Δ-ΔΣ ADC and the design parameters such bandwidth requirement, clock frequency, and scaling coefficients are different.

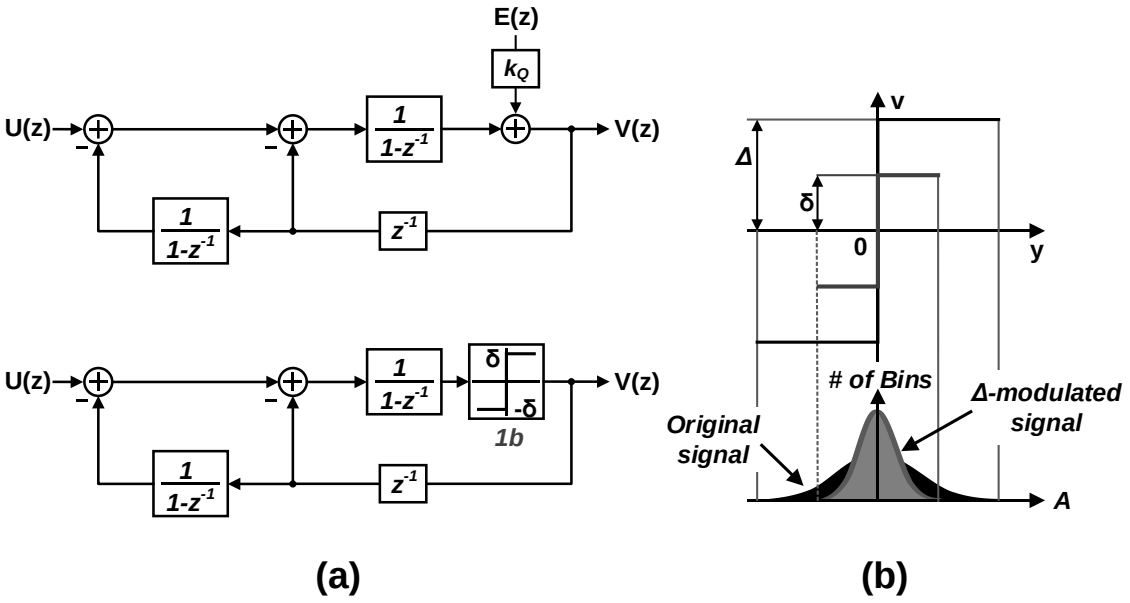


Figure 4-10 (a) Z-domain block diagram of the DT Δ-ΔΣ ADC (b) and the conceptual quantizer with reduced quantization level

Generally speaking, CT Δ-ΔΣ ADCs are more efficient in terms of energy consumption than DT counterparts, however its component sizes become larger when the bandwidth of the input signal become smaller [75]. Since we are dealing with very low frequency signals (the sampling frequency is 2 kHz.), the increased power consumption from the high bandwidth requirement (The rule of thumb is that the bandwidth of the OTA should be

more than 5 times of the sampling frequency) in the OTAs used in DT Δ-ΔΣ ADCs does not play a significant role in the overall power consumption of the system. In addition, the channel-to-channel variation can be improved by using the DT implementation since the matching property of the on-chip metal-insulator-metal (MIM) capacitors (±1%) is better than the resistors (±30%) [76]. Figure 4-11 shows the schematic of the DT-Δ-ΔΣ ADC. The capacitor C_1, C_2, C_3, C_4 and C_F are 200, 100, 250, 50, and 1 pF, respectively. V_{REFP}, V_{REFN}, and V_{CM} are the reference and common mode voltages and φ_1 and φ_2 are 64 kHz non-overlapping clocks.

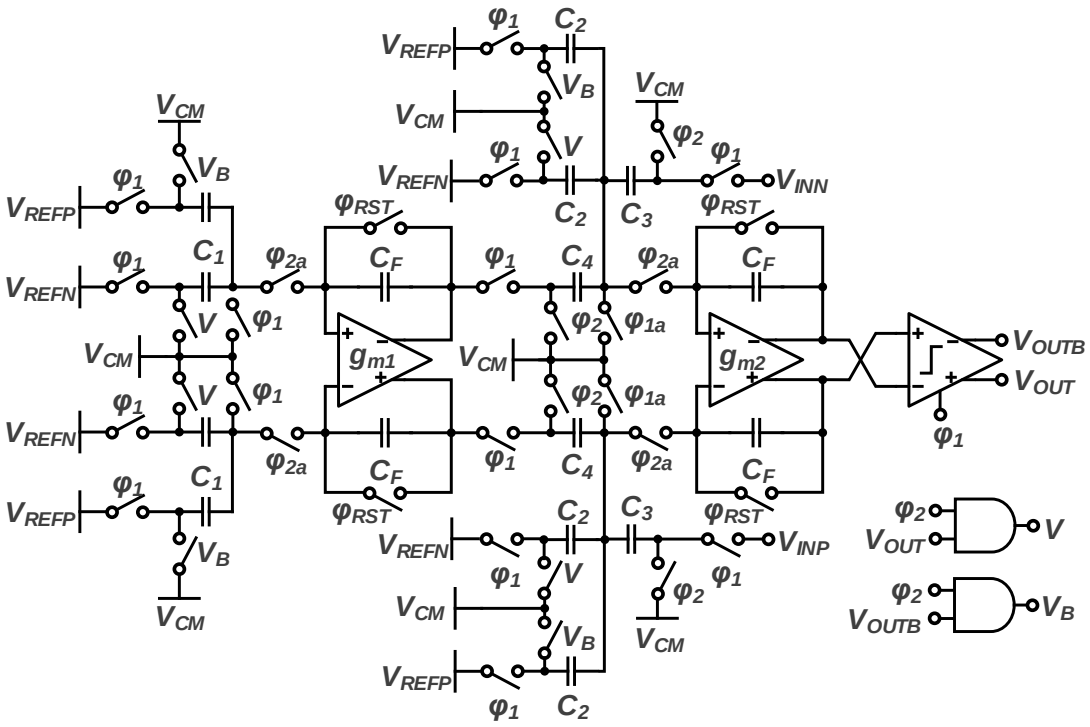


Figure 4-11 Schematic of the DT Δ-ΔΣ ADC

The OTAs used in the DT Δ-ΔΣ ADC denoted as g_{m1} and g_{m2} in Figure 4-11 have the same circuit topology as the one in Figure 4-9 while having different design parameters such as gain, bandwidth, noise, and etc. For the proposed architecture the Δ-ΔΣ ADC can be used

only for the LFP whose bandwidth is ten times smaller than the AP, the high performance can be achieved without high frequency clock. Therefore, for the proposed architecture, discrete time Δ-ΔΣ ADC might not play as a significant energy overhead while occupying relatively small area. The overall power consumption of the DT Δ-ΔΣ is about 0.837 μW and its SNDR and ENOB are 61.87 dB and 9.98 bit at 100 Hz input frequency.

4.4.3 Reconfigurable spike waveform extractor

In this proposed architecture, there are two different mode of operation as explained in the section 4.3.1. In the normal mode, the recorded AP are sent to the off-chip storage place without any compression. In this mode, a low power SAR ADC digitizes the incoming AP. On the other hand, in the compression mode the waveforms of AP are sampled and stored in 16 × A-FIFO$_U$ (unit A-FIFO) temporally, and not converted into digital formats until the spike detector issues the conversion onset signals. The A-FIFO is implemented by reusing the capacitor banks (DAC) used in the SAR ADC in the normal mode, and the sampled data on the capacitor (analog memory) is quantized by comparing with ramp signal that is externally provided. In this mode, it means that the quantization of the stored AP is done with a SS ADC consisting of the ramp signal and the comparator used in the SAR ADC. Figure 4-12 shows the capacitor configuration and related transistors used for both DAC in the 8b-SAR ADC and the A-FIFO. The unit capacitance (C_U) for the SAR ADC is 17.5 fF which is a half value of the minimum capacitor provided by the given fabrication process (35 fF from TSMC 0.18 μm). C_U is realized by simply stacking the two minimum capacitors (not shown here). Figure 5-12 highlights two circuits, 2C_U

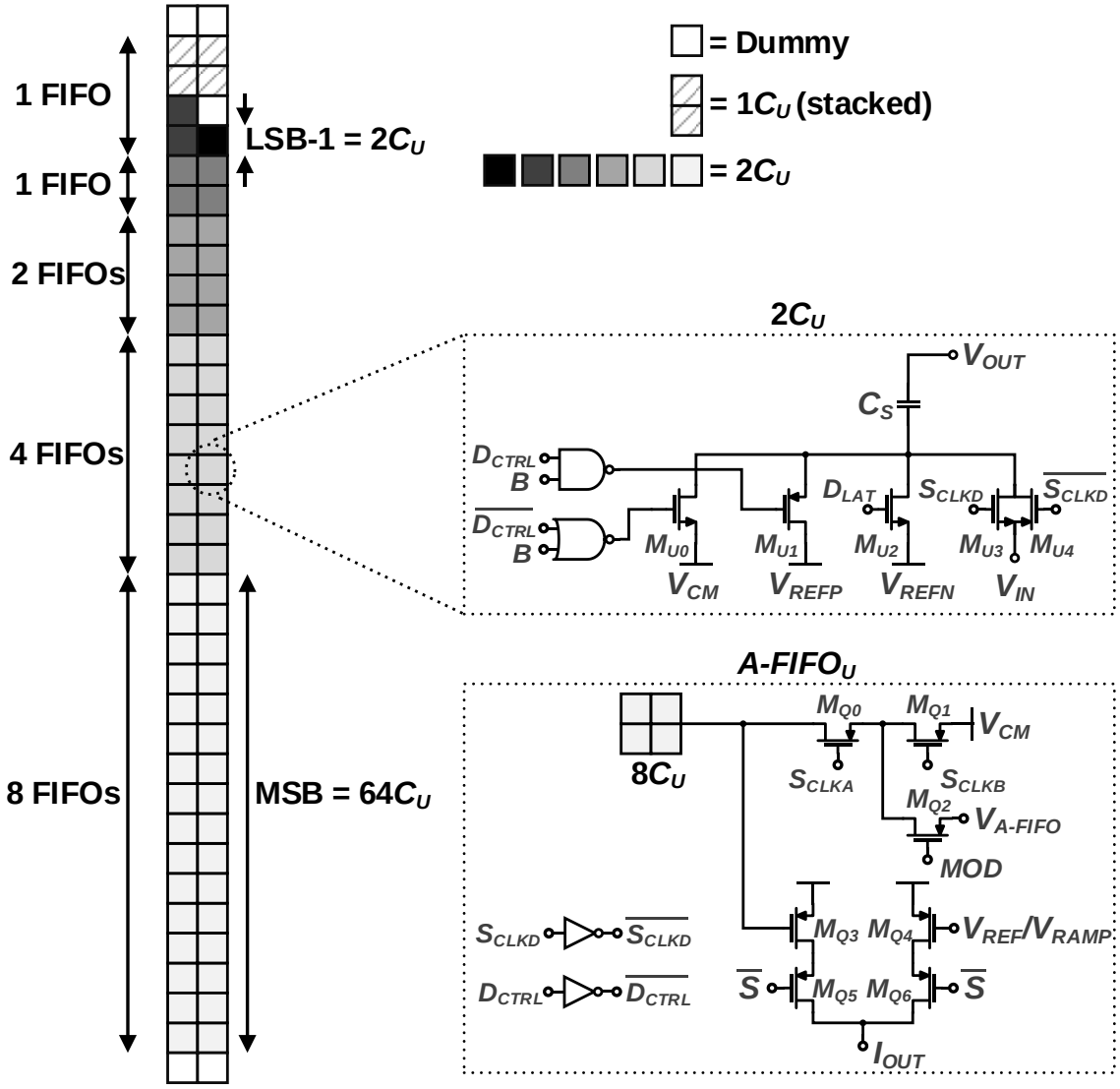


Figure 4-12 Capacitor bank configuration for both of ADC and A-FIFO

and A-FIFO$_U$ which are important to realize the reconfigurable ADC. The $2C_U$ has 3 single transistor switches ($M_{U0\text{-}2}$), 1 transmission gate ($M_{U3,\ 4}$) and logics. Since the $2C_U$ necessitates only a single 4×4 μm^2 MIM capacitor, the DAC has the $2C_U$ as a unit. An A-FIFO$_U$ consists of 8 C_U (140 fF) and 16 A-FIFO$_U$ becomes the A-FIFO as shown in Figure 5-12. The node voltage, $V_{A\text{-}FIFO}$ is connected to the same nodes in other A-FIFO$_U$. The A-FIFO$_U$ consists of 3 switches ($M_{Q0\text{-}2}$) and a differential pair ($M_{Q3\text{-}6}$). The differential pair converts the analog value of the connected A-FIFO$_U$ by comparing V_{REF} or V_{RAMP}

depending on the mode of operation, and delivers the comparison result to the cross-coupled pair in the dynamic comparator as a current format (I_{OUT}). The rest part of the comparator except for the differential pair is shown in Figure 4-13 with the conceptual capacitor banks. Overall, the dynamic comparator has 16 difference-differential (DD) input stage and converts each A-FIFO$_U$ sequentially. Figure 4-14 depicts the control signals of the three switches in Figure 4-12 according to the two different modes of operation.

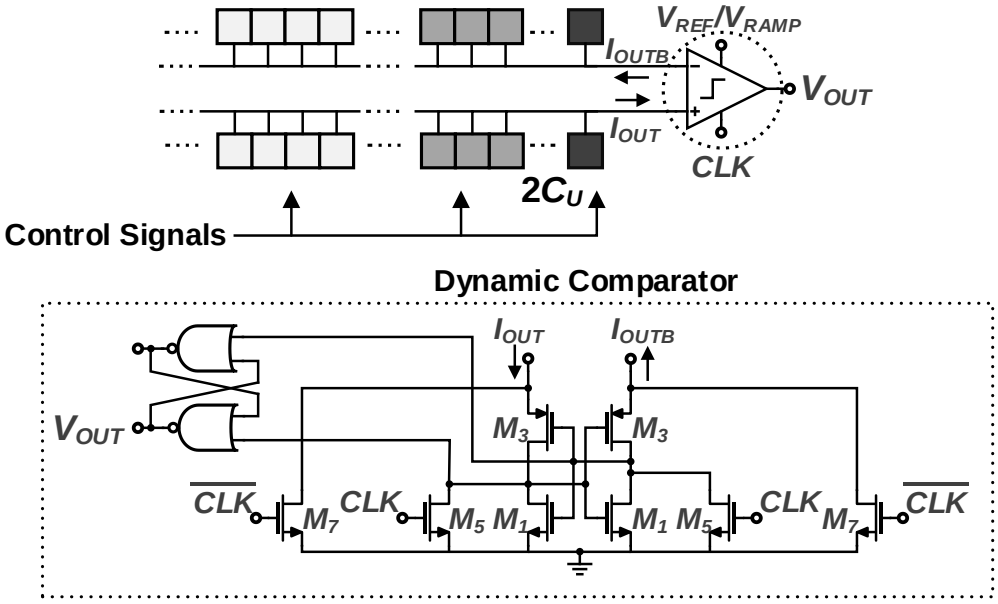


Figure 4-13 Conceptual SAR/SS ADC with details of a dynamic comparator

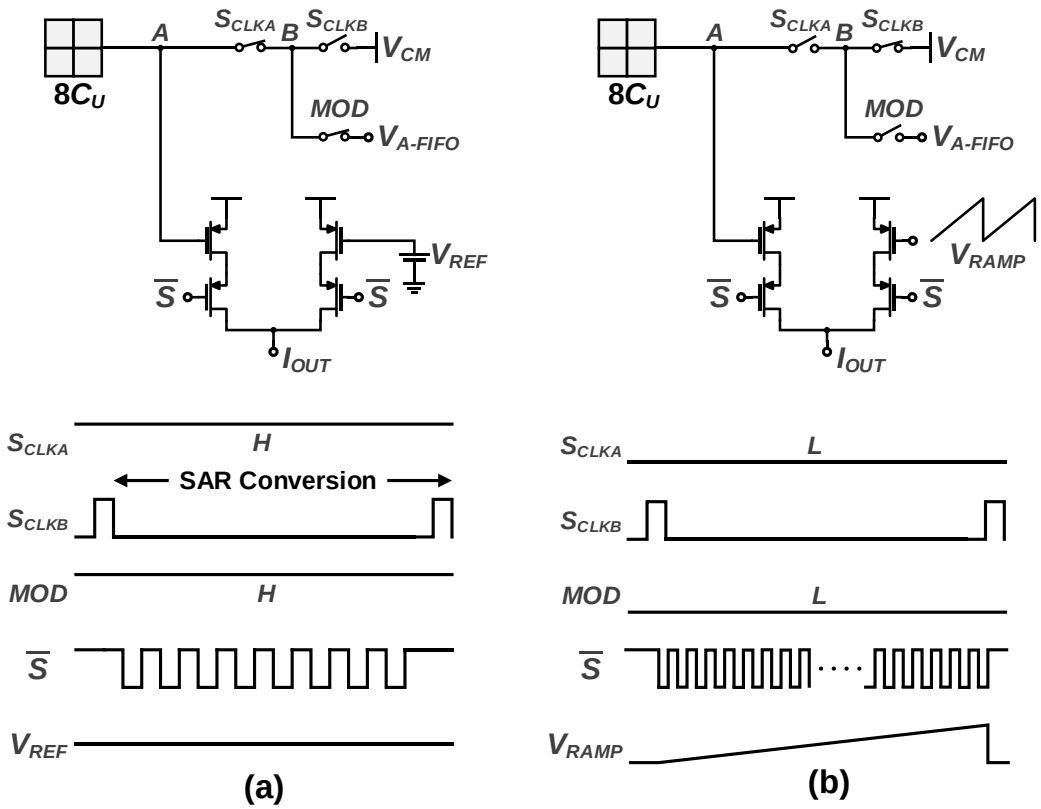


Figure 4-14 2 modes of operation of the reconfigurable SAR/SS ADC: (a) SAR operation, (b) SS operation with A-FIFO

4.4.4 Digital controller and interface

Figure 4-15 shows the digital controller and interface circuits. All data latches and control signals except for the ramp signal for the SAR, SS and Δ-ΔΣ ADC are generated internally. The spike detection is performed by a simple dual thresholding where the threshold voltages are provided externally. The equation for the threshold voltages are gives as

$$\sigma_n = median\left[\frac{|x|}{0.6745}\right] \tag{4-9}$$

where x is the recorded raw AP and the threshold voltage is determined as $4\sigma_n$ [61]. The channel-to-channel difference and the entropy-encoding of the recorded LFP are taken in the LFP processing block. The digital interface collects the two different signal paths (LFP and AP) and then serializes into two external data wires. For all digital implementations the auto place and routing (APR) and Verilog HDL are used.

4.5 Experiment results

The on-chip neural signal compressor with 128-channel AFE was fabricated in TMSC 0.18 µm 1P6M CMOS process. Figure 4-15 shows a microphotograph of the fabricated chip. The overall die area is 15.08 mm^2. Even though it is hard to clearly differentiate between the on-chip neural signal processor and the 128-channel AFE due to the mixed-signal nature of the compression mechanism, we can roughly draw the boundary between the two as shown in Figure 4-15. The neural signal compressor includes all controllers for the SAR and SS ADCs, 128 frame counters for the spike extraction and 128 entropy (Huffman) encoders for the AP, and 128 decimation filters, 120 spatial Δ generators and interrogator for LFP. In addition, there are a digital interface for data-serializing of the LFP and AP, and programming registers. The neural signal compressor and digital interface occupy 6080 µm × 930 µm. The 128-channel AFE consists of 128-LNA, high-pass PGA, low-pass PGA, AAF, buffers, SAR ADC and Δ-ΔΣ ADC. They takes the area of 6080 µm × 1250 µm including eight bias circuits, each is shared by 16 channels.

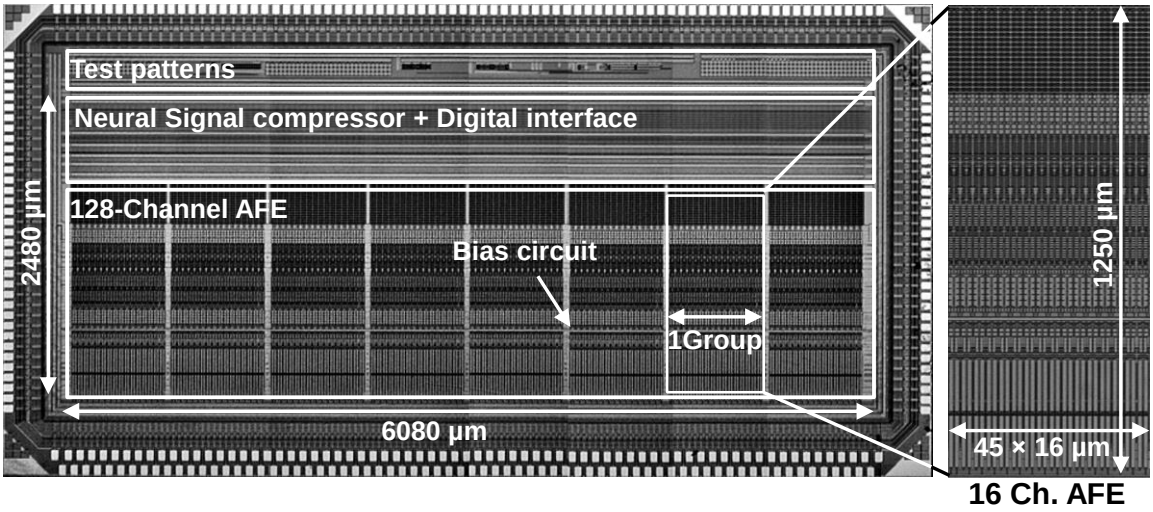


Figure 4-15 Die microphotograph of on-chip neural signal compressor with 128-channel AFE including test circuits

4.5.1 LNA, HPF, LFP and ADC Measurement

The frequency response of the LNA is shown in Figure 4-16. It has a mid-band gain of 37.8 dB from 0.4 Hz low corner frequency (f_L) 9 kHz high frequency corner (f_H). To measure the common mode rejection ratio (CMRR) of the LNA, the common and differential gains are measured and shown in Figure 4-17. The CMRR is about 60 dB. The IRN of the LNA measured using the Agilent 35670A is also shown in Figure 4-18. The integrated IRN from 0.4 Hz to 10.3 kHz is about 5.18 μV_{rms} which satisfies the neural recording requirement for the IRN (< 10 μV_{rms}) [36]. For the comparison, the simulated IRN is drawn with the measured one. Figure 4-19 shows the total harmonic distortion (THD) of the LNA with 1 kHz sine wave input. The 3rd order harmonics was found when the input voltage is 3.2 mV_{pp}. The THD is about -64.1 dB that is equivalent to about 0.062%. The power consumption of the LNA is 1.045 µW from a 0.5 V supply. Based on the foregoing

measurements, the important performance metrics for the LNA such as NEF and NEF^2VDD are calculated as 2.56 and 3.28, respectively.

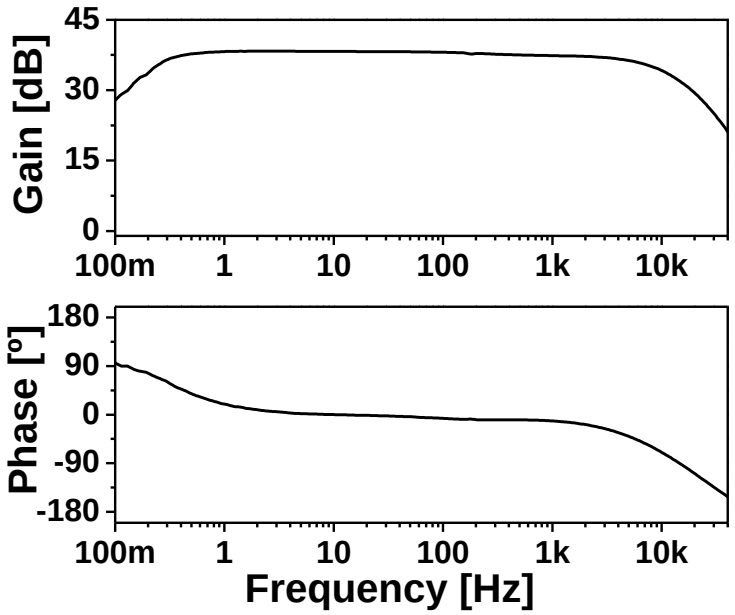


Figure 4-16 Measured gain and phase response of LNA

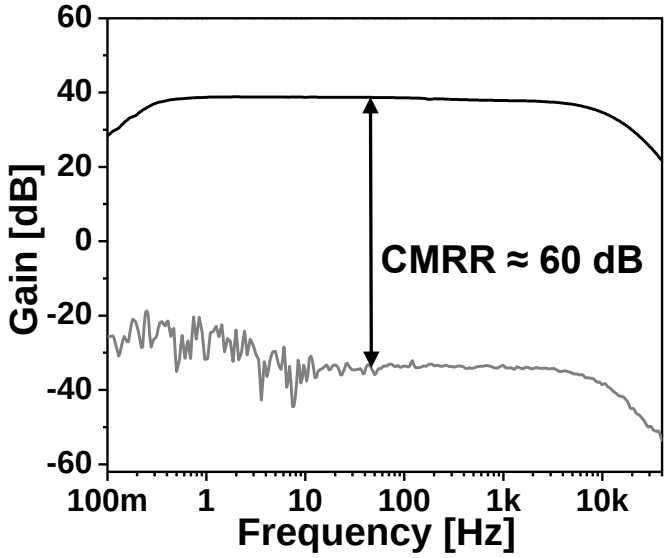


Figure 4-17 Measured common mode and differential gain of LNA

The measured gain response of the PGA for AP is shown in Figure 4-20 and 4-21. The voltage gain of the overall system can be adjusted by changing the internal setting of the PGA, where 4 level voltage gains of 7.6 dB, 9.9 dB, 17.1 dB, and 23.1 dB can be selected, as depicted in Figure 4-20. The corner frequency adjustment of the PGA is shown in Figure 4-21. The low frequency corner (f_H) can be set from 85 to 590 Hz by the external setting through the scan chain. Another PGA for LFP is also measured and gain and corner frequency changes are shown in Figure 4-22 and 4-23, respectively. According to Figure 4-22 and 4-23, the gain and the low frequency corner (f_L) are adjusted from -1.41 to 9.67 dB and from 166 to 298 Hz (4 and 8 different gain and frequency settings), respectively. Figure 4-24 depicts 1,024-points FFT analysis of the SAR ADC. The spurious free dynamic range (SFDR) is about 60.5 dB and the calculated signal to noise and distortion ratio is 48.76 dB, equivalent to 7.81b effective number of bit (ENOB). The power consumption of the ADC is measured to be about 0.11 μW from 0.5 V for the analog block and 1.0 V for the digital block (a comparator) and its figure of merit (FoM) is 19.61fJ/C-s. Figure 4-25 shows the differential nonlinearity (DNL) and integral nonlinearity (INL) of the SAR ADC. As described in Figure 4-25, there is no missing code in the implemented SAR ADC. The maximum and minimum DNL are 0.629 and -0.4182, respectively. Figure 4-26 depicts 32,765-points FFT analysis of the Δ-ΔΣ ADC. From Figure 4-26 the spurious free dynamic range (SFDR) can be read about 74.8 dB and the calculated signal to noise and distortion ratio is 61.87 dB, equivalent to 9.98b effective number of bit (ENOB). The output band noise is shaped with +20dB/dec line which is the typical frequency characteristic of the ΔΣ ADC. The power consumption of the ADC is measured to be about 0.837 μW from 0.5 V

for the analog block and 1.0 V for the digital block (a comparator) and its FoM is 414.15 fJ/C-s.

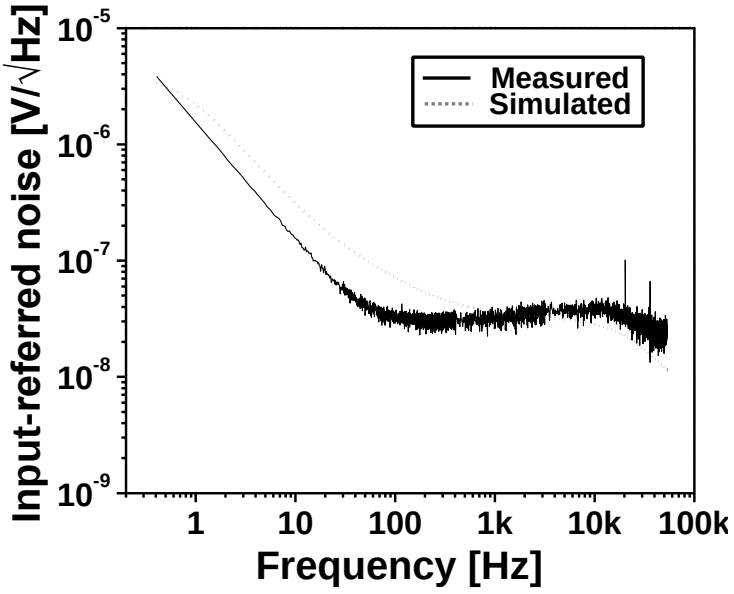


Figure 4-18 Measured input referred noise spectral density of LNA

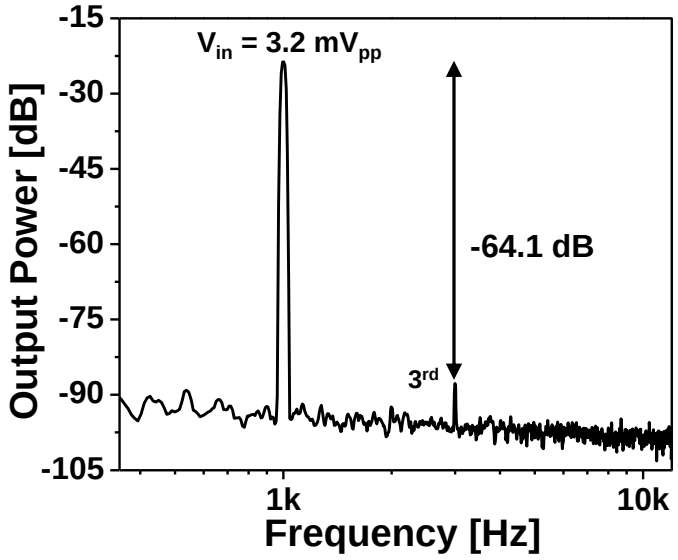


Figure 4-19 Measured total harmonic distortion of LNA

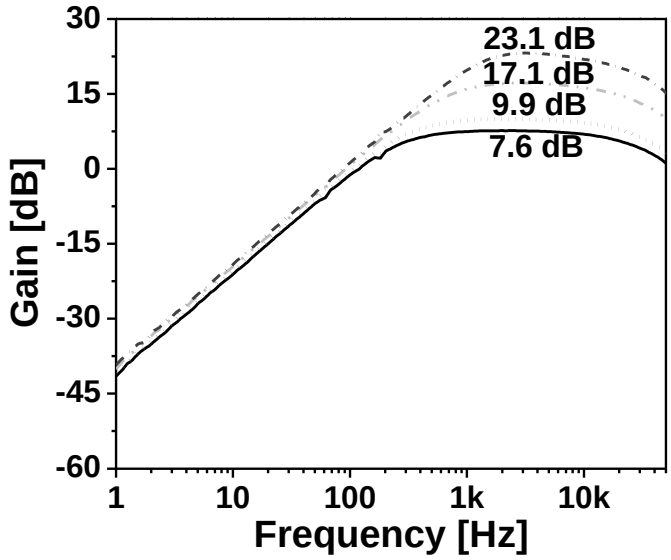


Figure 4-20 Measured frequency response of PGA: gain adjustment

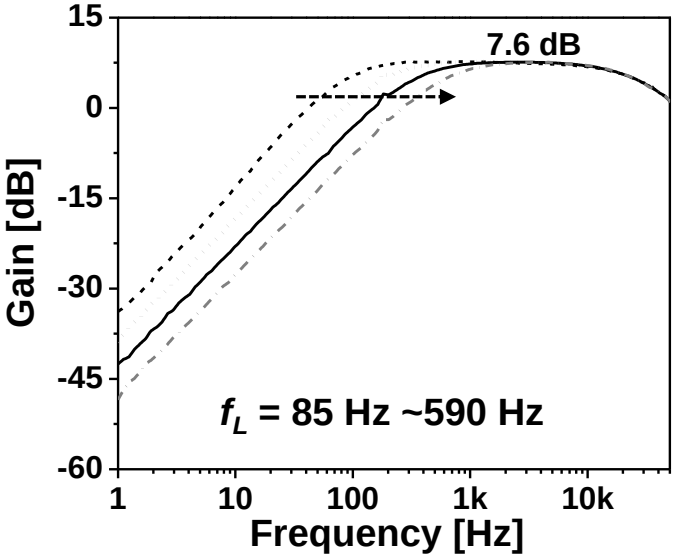


Figure 4-21 Measured frequency response of PGA: corner frequency adjustment

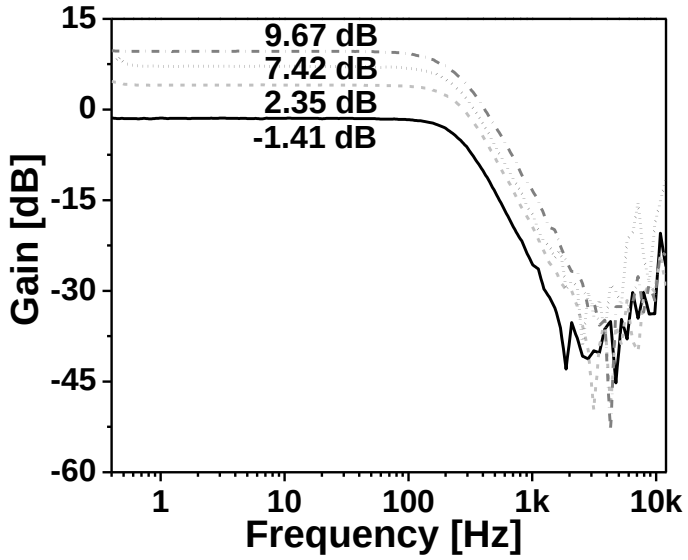


Figure 4-22 Measured frequency response of LPF: Gain adjustment

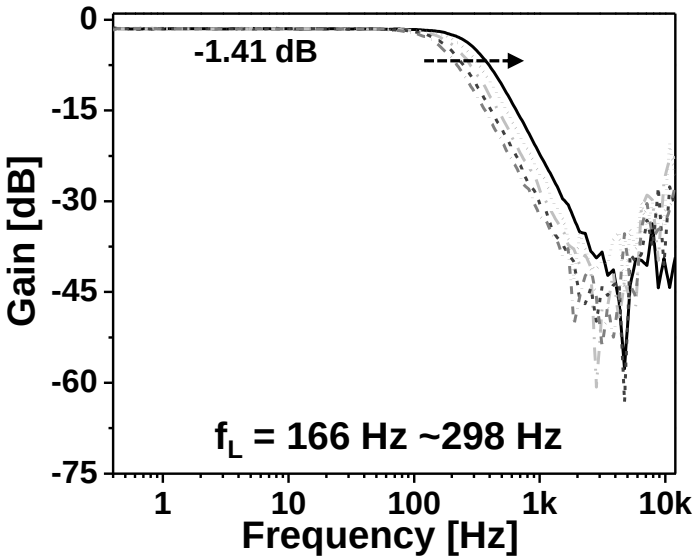


Figure 4-23 Measured frequency response of LPF: Corner frequency adjustment

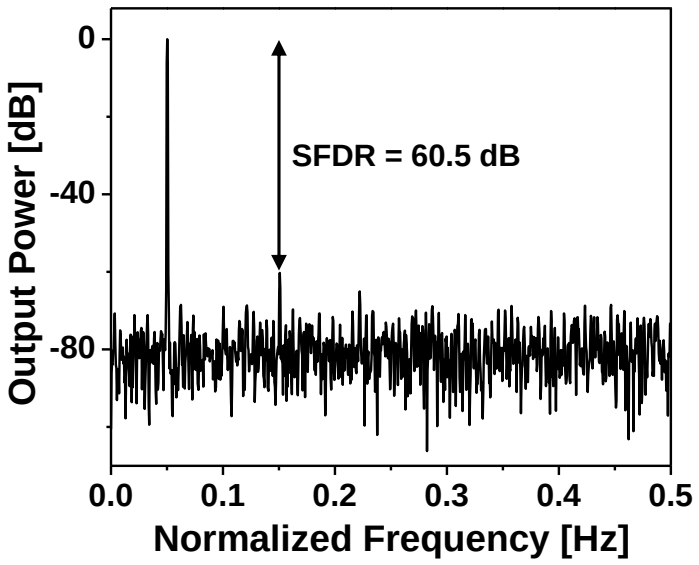


Figure 4-24 Full-scale input vs. SNDR measurement of SAR ADC

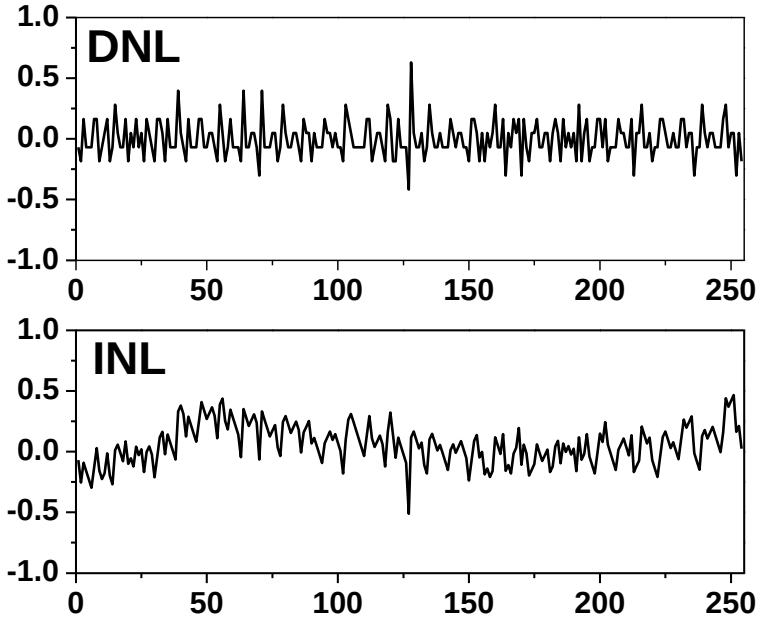


Figure 4-25 DNL and INL measurement of Δ–ΔΣ ADC

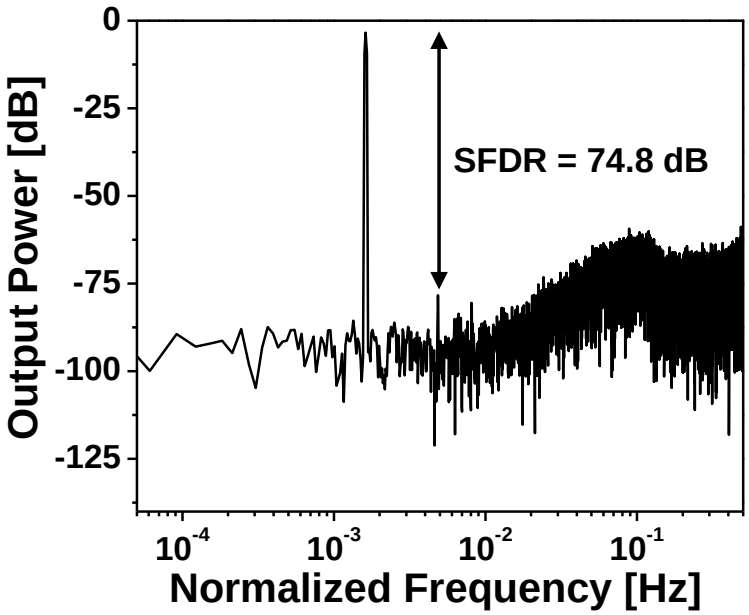


Figure 4-26 32,768-point FFT of the output of DT Δ–ΔΣ ADC

To confirm the functionality of the signal acquisition, the 1 second pre-recoded broadband neural signals are applied to the arbitrary selected channel. Figure 4-27 shows the input and separated LPFs and AP. For this measurement, the gain of the two PGAs are set with the lowest setting (-1.41 and 7.6 for LFP and AP, respectively) and f_H and f_L are 520 and 210 Hz, respectively. The overall fluctuation in the broadband input signals are well reflected on the extracted LFP. Figure 4-28 shows the two different modes of operation: normal and compression mode for AP. As expected, the only necessary waveforms of AP are extracted when the compression mode is on. For the clear visualization of the recorded signals, one of the AP are expanded and plotted on the right of Figure 4-28.

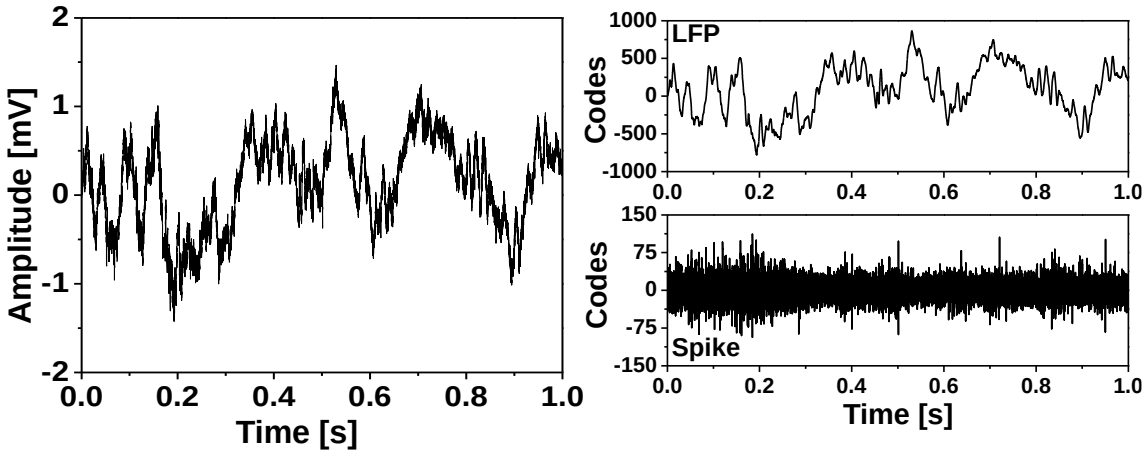


Figure 4-27 Neural signal separation from the prerecorded broadband neural signal

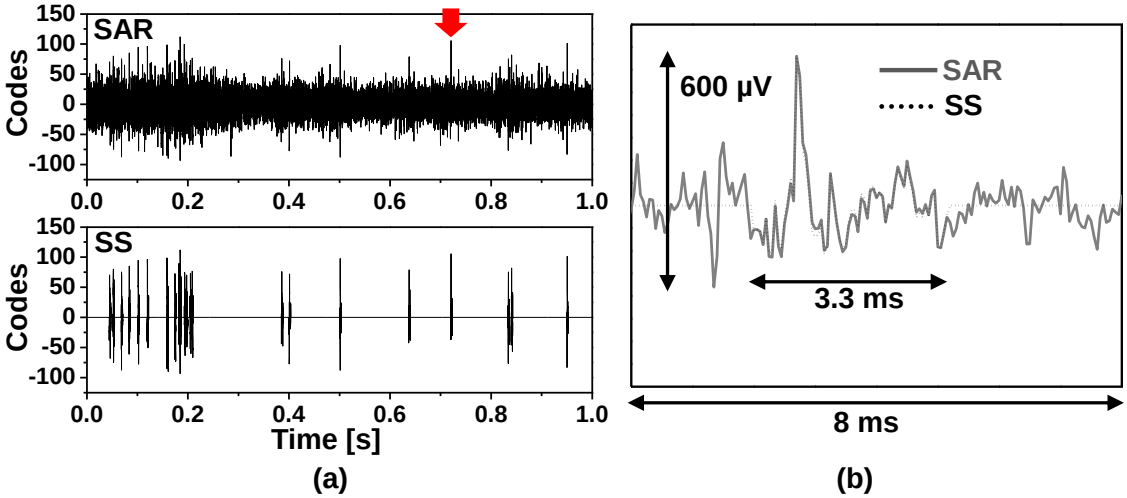


Figure 4-28 Two modes of operation for AP; normal (SAR) and compression (SS) mode, (a) both SAR and SS output of1 second AP and (b) Snapshot of 8 ms of (a)

4.5.2 *In-vivo* measurement

For the *in-vivo* signal measurement, a 32-channel tetrode was implanted in a rat and externally interfaced with the fabricated chip. Figure 4-29 shows the single channel measurement of LFP and AP. The overall gains and bandwidths for the LFP and AP measurement are 47.6 dB with 170 Hz high pass cutoff frequency and 47.9 dB with 455

Hz high pass cutoff frequency, respectively. Figure 4-30 (a) shows about 5 second snapshot of the recorded LFP and AP with the lowest gains for the PGAs and f_L = 248 Hz and f_H = 450 Hz corner frequency settings. For the compression mode, Figure 4-31 depicts the compressed and restored waveforms of the LFP ((a) and (b)). As shown in Figure 4-31, the amplitudes of the compressed LFP are much smaller than the restored ones. Figure 4-32 (a) depicts the separated AP when the LFP in Figure 4-31 was measured. For the AP, only the active parts have the codes, otherwise all data remain constant ('0' in those cases). Short (43 ms) snapshot of the AP is enlarged in Figure 4-32 (b) for better visualization.

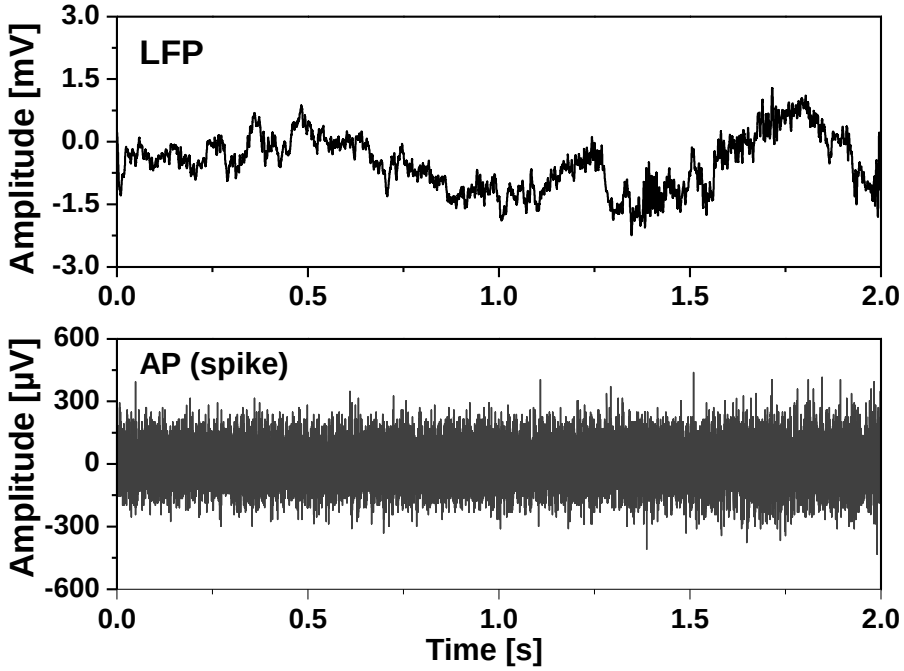


Figure 4-29 Single channel in-vivo measurement of the LFP and AP

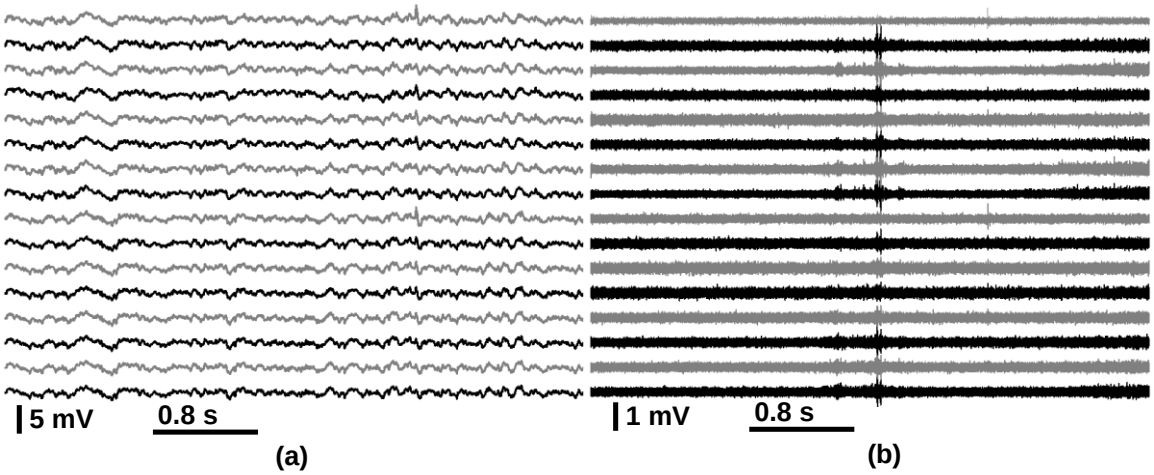


Figure 4-30 Measurement of the broadband neural recording in normal mode: (a) LFP (b) AP

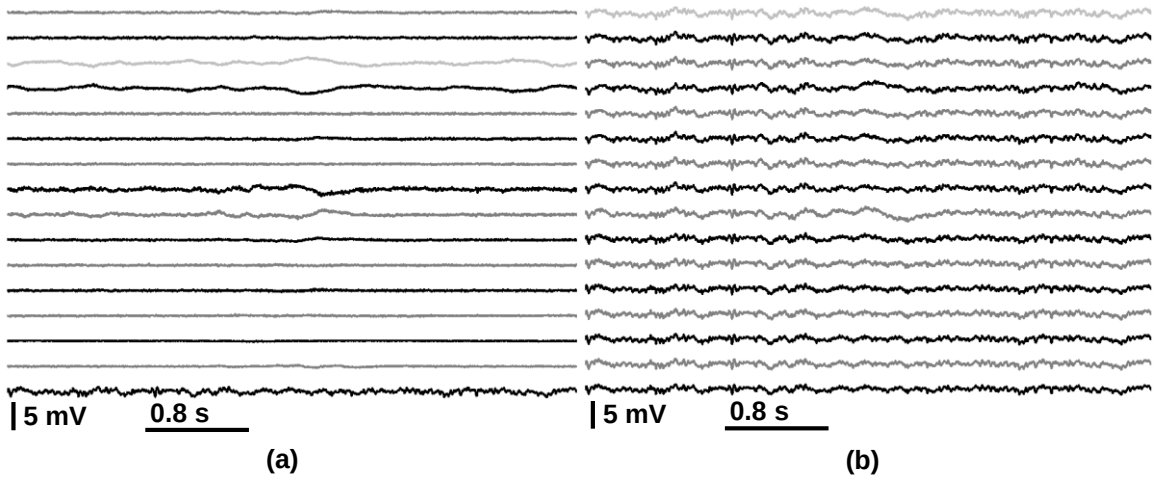


Figure 4-31 (a) Measurement of the spatial difference of the local field potentials (LFP) in the compression mode and (b) Retrieved LFP from software (MATLAB) and (c) waveforms of AP

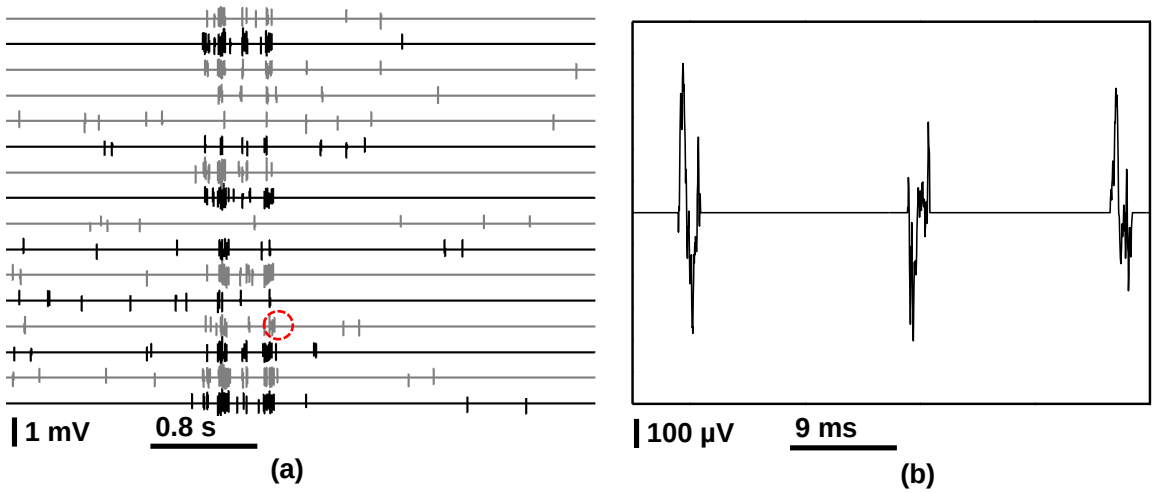


Figure 4-32 (a) Measurement of the compressed AP and (b) the waveforms of the excerpted AP

The last, but most important experiment is the power measurement for the two different modes of operation. Figure 4-32 shows the 1 minute average power measurement and the calculated cross-correlation based on the measured LFP. As shown in this figure, the overall average power follows well with the cross-correlation between channels: it means the compression based on the spatiotemporal correlation is effective. The average power consumption per channel for the LFP is measured as 3.59 µW. Figure 4-33 shows another average power measurement for the AP recording. As anticipated, the average power consumption is proportional to the AP firing rate. The average power consumption per channel for the AP is measured as 8.39 µW.

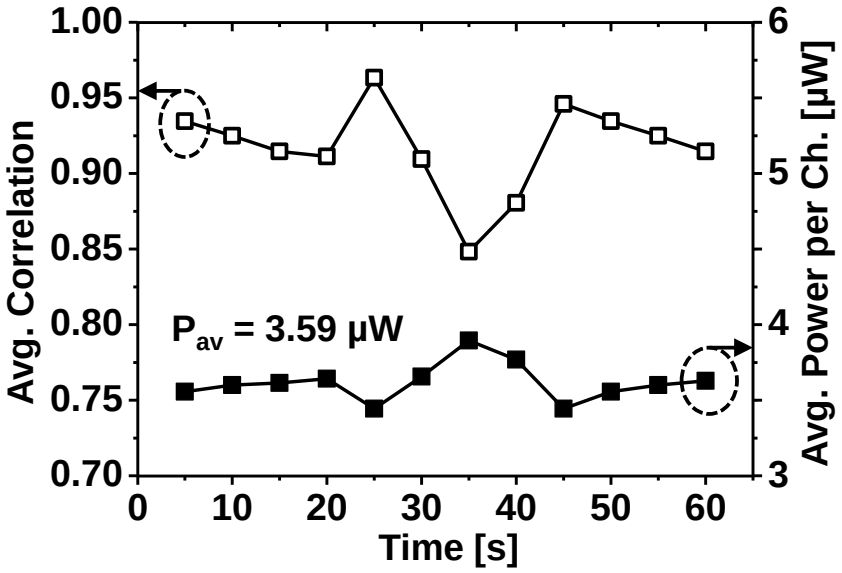


Figure 4-33 Digital interface power consumption (data handling and transmission) for LFP and cross-correlation between the channels

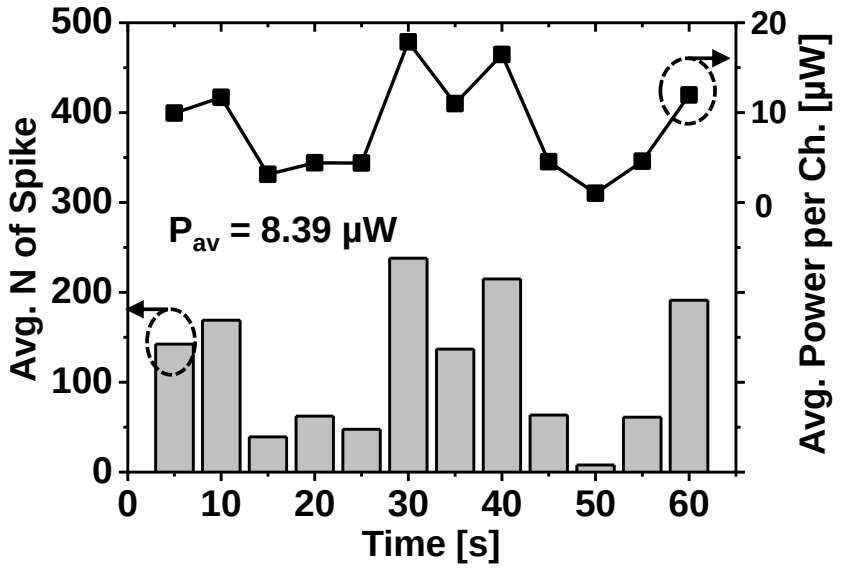


Figure 4-34 Digital interface power consumption (data handling and transmission) for AP and the average firing rate of the AP

4.5.3 Summary and comparison

The overall performance of the fabricated neural signal compressor is summarized in Table 4-1 by comparing other state-of-the-art works. The measured power consumption of the AFE and the on-chip neural signal compressor in the compression mode are 3.37 µW and 11.98 µW per channel, respectively. This work achieved the smallest power consumption per channel while simultaneously recording the broadband neural signals. At the same time, the proposed work accomplished high quality recording in terms of IRN and resolutions. Considering the power consumption of the digital block; 107.5 µW per channel in the normal mode, the power reduction in the digital block to transmit the data might be significant, estimated as about 1/9.

Table 3-1 Performance comparison with the state-of-the-art works

	[72]	[77]	[56]	[6]	**This work**
Analog power/Ch. [µW]	42.5	7.7	27.84	49.06	**3.37**
Digital power/Ch. [µW]	95.6	43			**11.98**
Analog area/Ch. [mm^2]	*0.144	–	0.19	0.12	**0.059**
Digital area/Ch. [mm^2]	*0.116	–			**0.039**
Data comp. for LFP/AP	–/Yes	No/Yes	–/–	–/–	**Yes/Yes**
Lossless LFP/AP	–/Yes	–/Yes	–/–	–/–	**Yes**
f_S for LFP/AP [KHz]	–/30	1/20	2.5/30	2.5/30	**2/25**
Bits/Sample for LFP/AP	–	–	10	10	**11/8**
ENOB for LFP/AP [bit]	–/7	> 9	9.2	–	**9.98/7.81**
FoM for LFP/AP [fJ/C-s]	–/**1927.1	–/–	*397.9	–	**414.15/19.61**
IRN [μV_{rms}]	5.9	3.3/3.1	3.2	6.36	**5.18**
Bandwidth [kHz]	9	10	*6	10	**9.2**
NEF	4.9	3.8	3.08	–	**2.61**
NEF^2VDD	*6.73	–	17.13	–	**3.41**
CMRR [dB]	–	> 70	60	> 60	**60**
# of chan. for LFP/AP	–/16	6/24	52 (455)	384(966)	**128/128**
Supply voltage [V]	1.2	–	1.8	1.8	**0.5**
Technology	0.18 µm	0.18 µm	0.18 µm	0.13 µm SOI	**0.18 µm**

* Estimated, ** Calculated based on 7.4 µW of power consumption in ADC

CHAPTER 5
PWM BUCK CONVERTER USING ANALOG-DIGITAL HYBRID CONTROL

5.1 Introduction

In the previous chapters, the architecture and circuit design of the 128-channel Δ-ΔΣ AFE and the on-chip neural signal compressor to efficiently use the system resources; area and energy in the analog blocks and energy in the digital block were discussed. When it comes to the energy in the system, the overall discussion up to this point was about how to consume the energy wisely. In addition to that, another important consideration is that how efficiently deliver the necessary energy to the system from primary energy sources. Despite of the high energy efficiency of the 128-channel Δ-ΔΣ AFE and the digital interface, the most part of the energy might be wasted if the energy delivery to the system is inefficient. Particularly, if considering the future fully implantable neural recording microsystems, the energy delivery and management must be more critical issue. This chapter deals with a dc-to-dc converter that is one of the most important components in the power management unit.

5.1.1 General requirement

The efficient energy usage is one of the most critical issues for the implantable neural recording systems since they are commonly located in the energy-limited environments such as human or animal bodies. Due to such restricted environmental conditions the implantable systems have no choice but depending on only several types of energy sources such as batteries, inductive wireless energy transfer, or a combination of them

[78][79][80][81][82][83]. Unfortunately, like the scenarios of all battery-powered devices, the implantable biomedical systems must be also replaced after a certain time period even though the replacement requires expensive and risky surgical procedures [84][85].

The high energy efficiency in the implantable biomedical systems can be achieved in two ways. First, the battery's lifetime is maximized if the total power consumption of the individual functional blocks in the system is minimized. In recent years many low power techniques have been developed to improve the energy efficiency of the implantable biomedical systems, and consequently the state-of-the art systems only consumes less than few mW [4][12][37][57][86]. Second, the dedicated power management in the systems also contributes to the efficient energy usage as the proper distribution of the primary energy source to the each individual block not only guarantees the high performance but also minimizes unnecessary power consumption [87][88].

However, most of the recent low power techniques for the implantable biomedical systems must be compromised if the power management is not properly engaged in the system. In essence many low power techniques for the implantable biomedical systems are enjoying the aggressive power supply scaling as the dynamic voltage scaling (DVS) technique is applied to the advanced digital circuits for high energy efficiency. [4][12][37][57][10][86][89]. While the scaled supply makes it possible for the systems to realize economic power consumption, the squeezed voltage headroom of the circuits incurs poor power supply rejection ration (PSRR) of the systems. To remedy this problem, linear regulators (linear dc-to-dc converters) are widely employed in the conventional power management of the implantable biomedical systems for the stable regulation of power

supplies. However the power management fails to secure high power conversion efficiency (PCE) if the linear regulators are employed since they dissipate away the large drop-out voltage resulted from the battery voltage (3 ~ 5V) and the required supply (0.5 ~ 1.5V) [90][91]. Therefore, it is necessary to investigate other types of power converters for implantable biomedical systems which operate with high voltage conversion ratio (VCR, ratio of the output to input voltage) while providing high power conversion efficiency (PCE).

5.1.2 Switched capacitor and inductive dc-to-dc converters

In broad sense, power converter can be divided into linear converter (regulator) and switching converter and the latter is further classified into switched capacitor (SC) and inductive converter. Generally, the SC converter can be fully integrated with other circuits on a same die and show relatively high PCE (60-90%) [87][88][92][93][94]. However, it provides only the discrete VCR and requires complex compensation techniques to generate a wide range of VCRs [95]. In the battery operating system where usually requires a large number of VCRs, that disadvantage is difficult to overcome. On the other hand, the inductive switching converter can fully cover all possible VCRs in a given topology (for example, buck, boost, fly-back and so on) and exhibit high PCE in general [96][97][98][99][100][101][102][103][104][105]. Even though the inductive power converter need several external components (typically 2−3 external components for basic topologies), it is tolerable for the implantable biomedical systems considering the typical form factor of the systems [106][107][108]. Therefore the inductive power converter can be a good candidate for implantable neural recording systems instead of the linear regulator.

5.1.3 PWM and PFM control

In spite of some attractive characteristics of inductive DC-to-DC converters, care must be taken in several aspects when employing this type of power converter to the implantable biomedical systems. In terms of the operating frequency of those converters the control mechanism of the converters can be categorized into the two different extremes: pulse frequency modulation (PFM) and pulse width modulation (PWM). The PFM converters can provide high PCE (~90%) for wide range of loads thanks to their adjustable switching frequency with load conditions [99][100][101]. However, the PFM control has serious shortcomings when combined with power noise-sensitive loads. For instance, the output ripples are larger by the lower switching frequency, which magnifies the impact of switching noise on loads. Moreover, since the frequency of the output spurs is a function of load currents, it varies from a particular load condition to another [109][110]. Thus, predicting their precise location in order to mitigate their impact on the load side becomes extremely difficult. On the other hand, the PWM converters can provide stable output with relatively fast transient while exhibiting high PCE over medium to high loads. In addition, their output spurs are predictable thanks to their fixed switching nature, therefore the PWM converter are able to provide clean supplies if including proper filters at their output. However, their PCE typically suffers at light load condition, in particular less than a few mW that the most of the implantable biomedical systems consume [106]. For instance, in operations where a few of low power blocks in an implantable recording system are the only active parts, low PCE is inevitable with the PWM converters. This case actually happens as shown in Figure 5-1 (a) where the percentages of power consumptions of the common functional blocks in an implantable biomedical systems are described [106]. The calculation in Figure 5-1 is based on the assumption that the 100 channel AFE consumes

10 μW per channel, the electrical stimulation block (E-Stim) dissipates 1−4 mW and the data transmitter and receiver (Tx and Rx) uses 1 mW. Since the analog front-end (AFE) always operates while the other blocks such as data communications and electrical/optical stimulations are active based on the systems request (or users' request), the operation of the power conversion (dc-to-dc conversion) is mostly dedicated to the low power regions. Therefore high energy efficiency is not achieved if the conventional PWM converter is directly engaged in the systems. Moreover, considering the battery life time, the energy consumption should be considered instead of the peak power consumption. Based on the estimated values from Figure 5-1 (a), the daily energy consumption can calculated as shown in Figure 5-1 (b). Even though the AFE only consumes a small fraction of the total power consumption, around 0.97 %, it takes about 87 % of the total daily energy consumption. Hence, if the PCE on low power region is low, the battery life time becomes shorter. In conclusion, it is obvious that high energy efficiency cannot be achieved if the conventional PWM converters are employed for implantable recording systems.

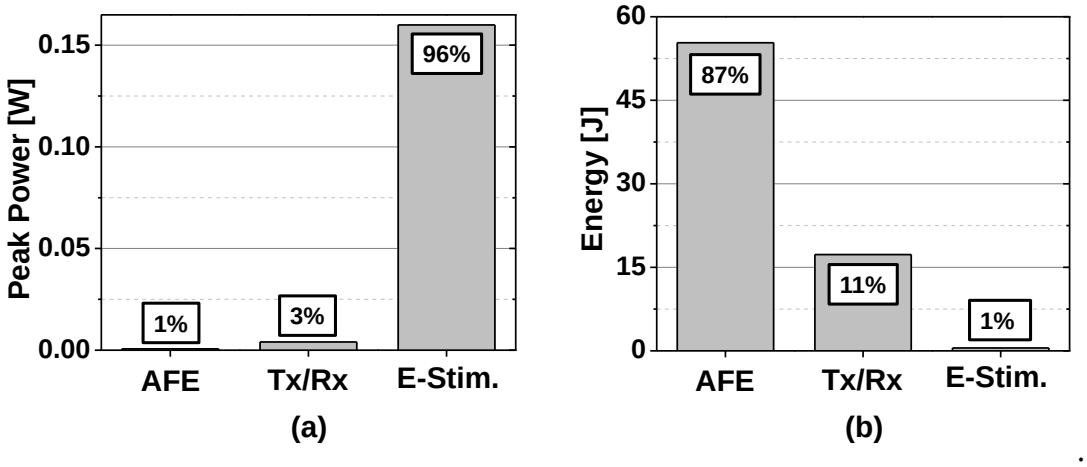


Figure 5-1 (a) peak power consumption and (b) daily energy consumption of an implantable neural recording system

5.1.4 Motivation

A dc-to-dc converter must provide high PCE at light loads to extend life time of an implantable neural recording microsystem while securing high supply integrity. Here, a PWM buck converter (inductive) with the load-adaptive power transistor scaling scheme and hybrid (analog/digital) coarse/fine control scheme is presented to achieve high PCE even at ultra-light load condition, even less than 100μA. Since it operates with fixed frequency its output spurs are predictable and can be completely removable. Thus, it is possible to guarantee high integrity of power supply as well. The rest of this chapter is organized as follows. The section 5.2 focuses on the modifications made to traditional buck converter architecture and summarizes the changes necessary for the proposed buck converter. The section 5-3 describes the architecture and circuit design of this PWM buck converter and its operation in details. Measurement results are presented in the section 5-4, followed by comparison with the other state-of-the-art works in the section 5-5.

5.2 Load-adaptive power transistor scaling

In this chapter, the principal mechanism to improve PCE of this PWM buck will be explained. For the explanation, the analysis of power loss in a conventional buck converter is presented first, and then a remedy to reduce power loss is described.

5.2.1 Analysis of power loss in a PWM buck converter

This section explores how different types of power loss related to loading conditions of a constant switching inductive dc-to-dc converter, particularly in a PWM buck converter and how to minimize the dominant power loss to enhance the PCE. For this discussion, all design parameters such as switching frequency, values of passives, and so on are assumed

to be selected well to reduce the total power loss of the converter. As the converter's load current varies over its full range, the relative contribution of the different types of power loss also varies. Without loss of generality the total power loss can be decomposed into four different categories: AC/DC conduction loss coming from Ohmic loss of power transistors in the converter and overlap (IV) loss resulted from imperfect switching timings of the NMOS and PMOS transistors, and switching (SW) loss dissipated for charging of the gate capacitances of the power transistors. There are other types of power losses by eddy current and core saturation in the inductor, however they are mostly negligible compared to the four listed above [105]. The four different losses and the total PCE with varying load current from 30 μA to 5 mA are illustrated in Figure 5-2 where the converter steps a 3.3V supply down to generate 1 V output while operating at 1 MHz fixed switching frequency with a 6.8 μH inductor and 1.2 μF capacitor. All parasitic resistance of the passives is assumed to be 1 Ω. As shown in Figure 5-2, while IV and AC/DC conduction loss are monotonically increased from 30μA to 5mA, the SW loss is independent of the change of the load currents and remains constant (~80μW) due to the fixed size of the power transistors. Specifically, the SW loss dominates all of the other three losses at less than 1 mA loads, which make this PWM converter exhibit poor PCE at the light loads. Thus, in order to improve the PCE in such light loads, the SW loss should be reduced without increasing other losses. The SW loss of a power converter is given by

$$P_{SW,loss} = C_G \cdot V_{SW}^2 \cdot f_{SW} \tag{5-1}$$

where C_G is the total gate capacitance of power transistors, V_{SW} is voltage swing of the gate driving signal, and f_{SW} is the switching frequency of the converter. From (5-1) the SW loss can be smaller by reducing each of f_{SW}, V_{SW} or C_G, or combination of them. The typical

way to reduce the SW loss is changing f_{SW} according to loads, which is what the PFM control does. Since the PFM control has an inevitable disadvantage for the sensitive loads as explained in the section 5.1.3, this method is not further discussed in this section. Nonetheless, one exceptional technique of this PFM technique has worthy of note. The auto selectable frequency pulse width modulation (ASFM) scheme, where the operating frequencies can be chosen among fixed sets of frequencies, was proposed to achieve the improved PCE while generating predictable spurs at the output [103]. However, their selectable frequencies are bounded to several discrete numbers, for instance f_{SW}/N where $N = 2i$ and $i = 0, 1, 2, 3$; thus, the more complicated frequency sets and control circuits are necessary when trying to achieve near-optimal PCE or to cover a broader range of loads. The effective ways to reduce C_G was also proposed in [111] where the gate width of the power transistors is changed to reduce the effective C_G. Thanks to the reduction of the effective C_G in their scheme, their PCE is improved, however the presented scheme only showed the proof-of-concept without any automatic control for changing of C_G. For the reduction of V_{SW}, [112] proposed a technique to reduce the gate charge which is called the gate charge modulation (GCM) where the amplitudes of the gate driving potential are altered according to load conditions. Another variation to reduce V_{SW} is to exploit the resonance between the additional off-chip inductor and C_G in the gate of a power transistor [113]. Despite of various techniques which have their own advantages, most of works cannot provide high PCE for the light loads which most of the implantable neural recording systems demand.

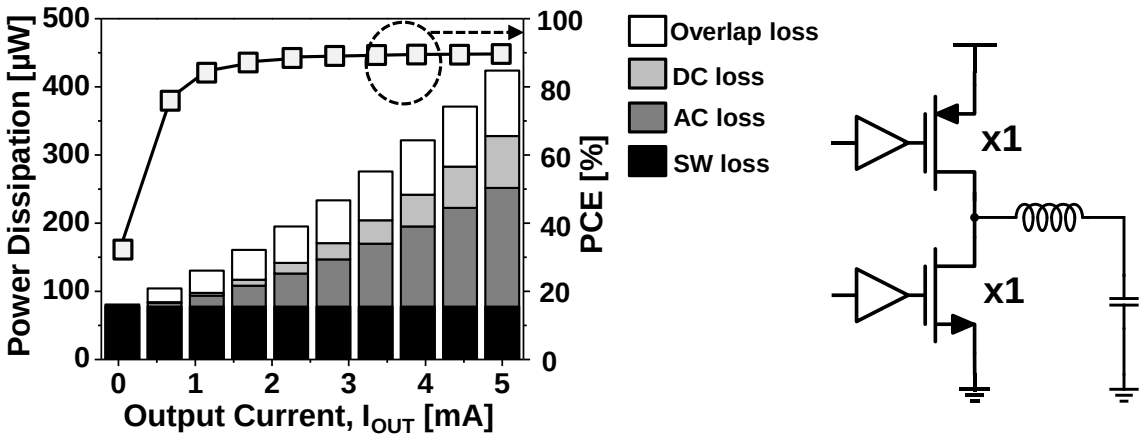


Figure 5-2 Four different power losses in a PWM buck converter; conduction (AC/DC) loss, overlap (IV) loss and switching (SW) loss, and the corresponding PCE of the converter

5.2.2 Switching power loss reduction in a PWM buck converter

Even though [111] did not propose any automatic control schemes to set the size of power transistors, their analytical solution about the optimal size of power transistors to minimize power loss of a converter and experiment results to support their theory is noteworthy. According to their results, the optimal width (W_{OPT}) of the power transistor in a buck converter to minimize the total power loss is given by

$$W_{OPT} \propto \frac{P_{OUT}}{\sqrt{f_{SW}}} = \frac{V_{OUT} \cdot I_{OUT}}{\sqrt{f_{SW}}} \tag{5-2}$$

where P_{OUT} and f_{SW} are the output power and switching frequency of the given power converter, respectively [111]. As implicated in (5-2), W_{OPT} is proportional to the load current when the operating frequency and the output voltage (V_{OUT}) are constant. As a verification of the work in [111] another numerical calculation is performed using (5-2) and is illustrated in Figure 5-3 where the width of the power transistors is divided by 16 and allocated in 16 unit size of them, and other conditions except for the width are same as

in Figure 5-2. Even though the AC and DC conduction loss and overlap loss are also affected by the variable size of the power transistors, their incensements are minute due to the small load currents. On the other hand, as the effective gate capacitance becomes smaller the SW loss is significantly decreased, and the overall PCE is correspondingly improved from less than 40% to over than 80% even at the extreme light load, 35μA that is the favorable power range for implantable neural recording microsystems. Consequently if we can realize the load-adaptive scaling of the power transistors scheme satisfying (2) while consuming small additional power (much less than the SW loss at light loads) and exhibiting low complexity, the overall PCE will be enhanced.

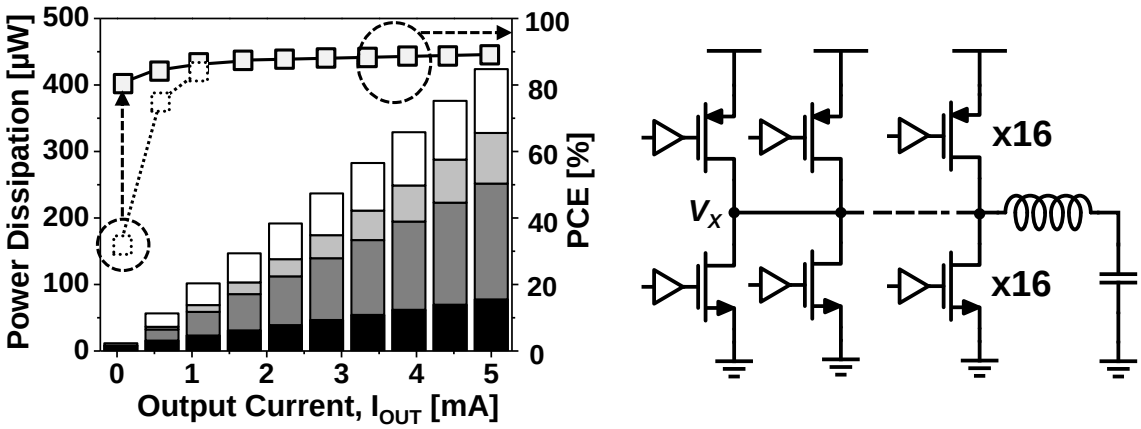


Figure 5-3 Four different power losses in a PWM buck converter: AC/DC loss, OV loss, and SW and the corresponding PCE of the converter while changing the size of the power transistor according to the loads

Unfortunately, most of the recent works must have the dedicated controllers and sensors to realize the load-adaptive power transistor scaling [103][104][112][113][114]. In [103] the parasitic resistance of the inductor and active RC low pass filters are used for the estimation of the current loading condition and accommodated the finites state machine

(FSM) to calculate the necessary size of the power transistor. In [114], the dedicated ADC for this purpose was implemented on chip. However, in this architecture, the load-adaptive scaling of power transistor is realized by reusing the existing blocks without any dedicated sensors and estimators. The implemented PWM buck converter operates in discontinuous conduction mode (DCM) throughout the required loads from 35µA to 4mA with the given design parameters [115][116]. In DCM mode of operation, the duty cycle is calculated with (5-3)

$$M = \frac{2}{1+\sqrt{1+\frac{4K}{D^2}}} \tag{5-3}$$

where M, K and D are VCR, critical value ($K = 2Lf_S / R$, R is load), and duty cycle, respectively. Using (5-3) the required duty cycles for both of two different VCR (0.3 and 0.7, equivalent to V_{OUT} = 1 V from V_{IN} = 2.5 and 3.3 V) are shown in Figure 5-4. Since the required duty is proportional to the output current, I_{OUT} in the DCM mode, it can be used for the selection of the number of the power transistors (N, in Figure 4-4). Even though this is not an optimal control following (5-2) precisely (The ideal case for M = 0.7 is indicated in the dashed line in Figure 5-4), the additional power loss from the deviation is not significant. For instance, the additional power dissipation at the 35 µA by the deviation from the ideal is calculated as about 1.2 µW which is only 5 % of total power dissipation at this load condition. In addition, another advantage of this control comes for the realization of it. The selection of the number of transistor is discrete, thus the digital controller might be favorable. In this case, no digital-to-analog conversion (DAC) which is usually implemented with a power consuming ramp generator in the conventional digitally controlled converters is necessary because the selection of the power transistors

itself performs the DAC function [117]. The details of the realization will be dealt in the section 5.4.

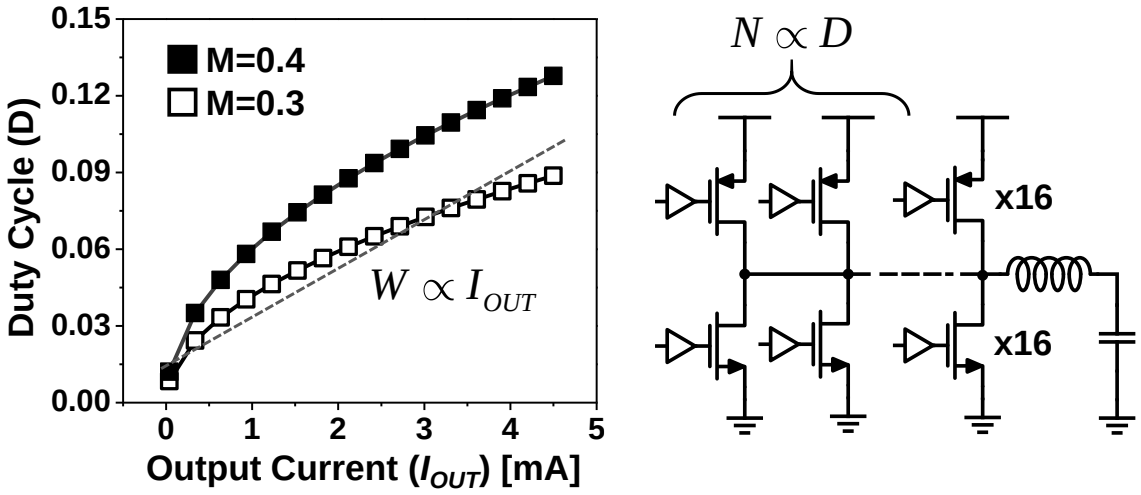


Figure 5-4 Duty cycles for a DCM Buck operation of two cases: M = 0.7 and M = 0.3. (5-2) is overlaid on the graph for M = 0.7

5.3 Circuit architecture

Figure 5-5 shows the architecture of the proposed PWM buck converter to improve the light load efficiency where V_{IN} and V_{OUT} are the input and output voltages and V_{REF}, V_{REFH}, and V_{REFL} are the reference voltages. The buck converter steps 2.5-3.3V input voltage (V_{IN}) down to 1V output (V_{OUT}). The converter incorporated the analog-digital hybrid controller and the array of power transistors for the load-adaptive scaling of the width of power transistors. The power transistors of the converter have an array of the 16 unit-sized power transistors. The controller has two different modes: digital (coarse) and analog (fine) controls. In the digital control mode, the overall operation is similar with the conventional digital controller [117]. It uses a variable-delay digital pulse width modulation (variable delay DPWM), a digital compensator, an inductor current emulator, an ADC. In analog

control mode the overall operation is performed within the error bounds from the coarse digital control while reusing the functional blocks of the digital control. The bounded error from the coarse digital control is amplified and delivered to the variable delay DPWM for the final tuning. Since the converter has two control modes, there should be a functional block to make a decision for which mode should be used. The mode arbiter not only governs the mode of controls but also extract the current control information of the converter based on the limit cycle oscillations (LCO) from the oversimplified digital control. On the whole, the LCO are the phenomena which should be avoided for the stable output control. However, the LCOs are exploited for the control of the converter in the proposed architecture.

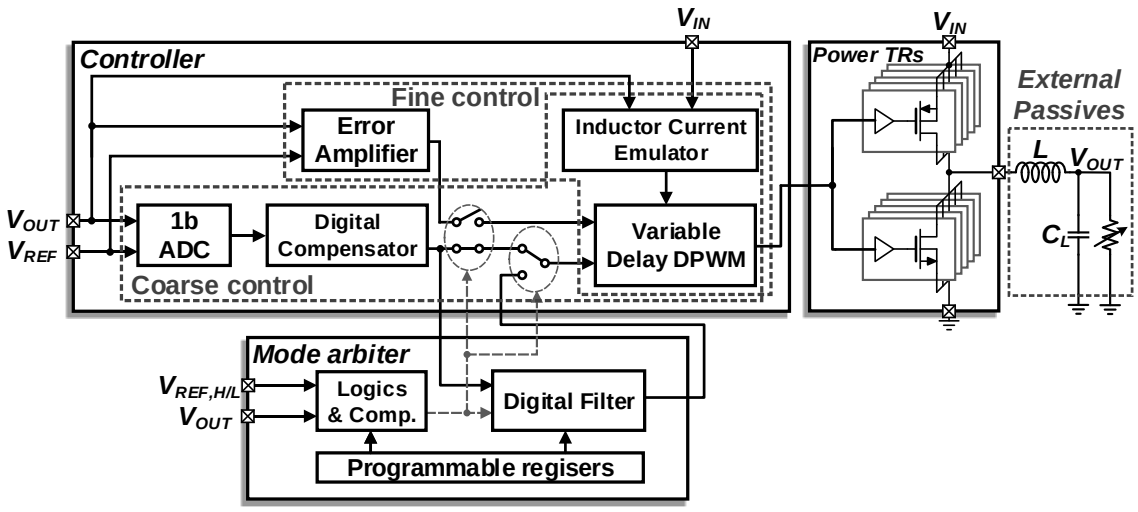


Figure 5-5 Architecture of the PWM buck converter with the hybrid controller

5.4 Circuit implementation and operation

5.4.1 Power transistor design

To achieve the high PCE at light loads, the sizing of the unit power transistor is one of the most important issues in the proposed converter. For the design of the unit-sized power

transistors, two major design criteria were considered. First, the sum of the widths of all unit power transistors (the maximum size of the power transistors) was designed by considering for the Ohmic loss (AC/DC conduction loss) to be less than half of the total power dissipation at maximum current (4 mA), and consequently the SW loss and IV loss can occupy the other half of the total allowable power loss. The reason why we set this standard is that the portion of both conduction losses becomes dominant as the output current increases and researches more than half of the total power loss as shown in Figure 5-3. Second, once the SW loss is confined into the certain value to guarantee the target PCE (>80%) for the minimum load current, the other losses can be ignored because the other loss will be much smaller than the SW loss due to the tiny load current. However, another constant loss comes from the analog blocks should be included in this procedure. To guarantee over 80% efficiency at the lightest load condition (35 μW), the SW loss are set to be less than 5 μW considering 6.5 μW constant power consumption from the analog blocks. When a single PMOS and NMOS power transistor are selected, C_{TOT} (total capacitance at node V_X in Figure 5-3) should be less than 600 fF at f_{SW} = 1 MHz and V_{SW} = 3.3 V. The gate capacitance (C_{GS}) and overlap capacitance (C_{GD}) of a single PMOS and NMOS are 90 fF and 90 fF, and 0.95 fF and 1.2 fF, respectively. Thus, C_{TOT} equals to be 550 fF. For maximum current, the power dissipation from the Ohmic loss (AC/DC conduction loss) becomes about 170 μW by the 1.15 and 0.9 Ω on-resistance of the 120 μm and 75μm widths of PMOS and NMOS transistors.

5.4.2 Coarse digital controller

Figure 5-6 shows the block diagram of the coarse digital control. The controller operates only when the output of the converter stays in transient. The error voltage is

quantized by the 1b-analog to digital converter (ADC) and integrated through the compensator to control the number of delay cells in a loop. The ADC and compensator are implemented minimally using a comparator and a 16b-counter to reduce the power dissipation. Once the compensator calculates the necessary duty (D), the variable delay DWPM consisting of the array of the delay line and unit-sized gate drivers turns on or off the unit-size power PMOS and NMOS transistors. The PMOS and NMOS driving signals are denoted as V_P and V_N in Figure 5-6. The variable delay DPWM can realize D using the loop delays designated by the counter, and adaptively perform the power gating by choosing the necessary number of cells for a specific load without any additional control for the selection of the power transistors. The forward and reverse delay cells, D_F and D_R in the DPWM are controlled by the inductor current emulator (I_L Emulator) where the driving current I_F and I_R are generated for the delay cells. In nominal 3.3 V input and 1.0 V output, the D_F and D_R are 4.5 ns and 16.5 ns, respectively.

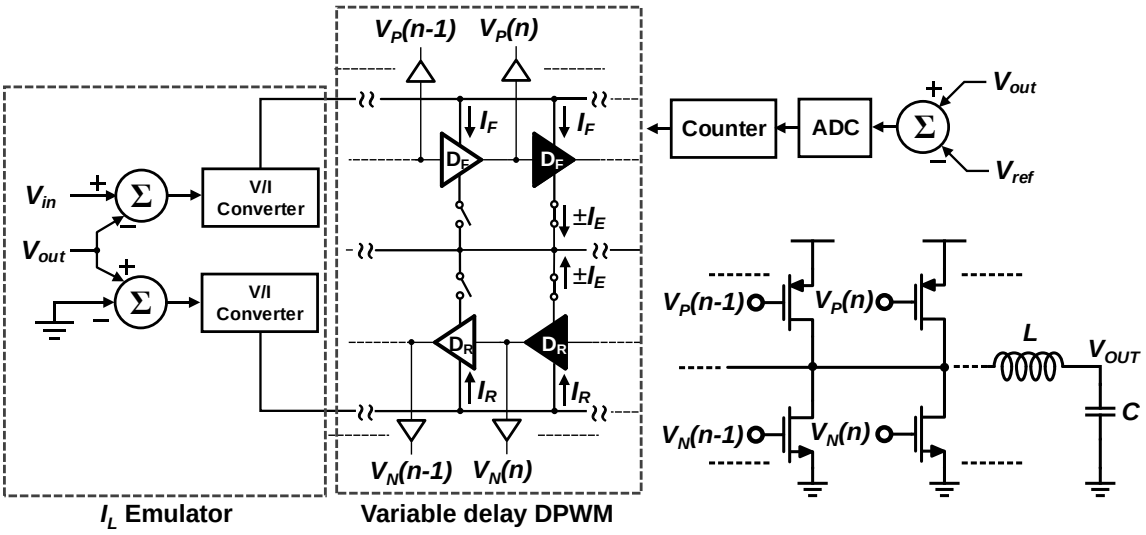


Figure 5-6 Block diagram of the coarse digital controller. The part of the power transistor array is also shown to aid the understanding

The I_L Emulator performs two functions. First, it provides different delays for the different input supply. As shown in Figure 5-7 (a) the I_L Emulator consists of two voltage to current (V/I) converters that generate the driving current I_F and I_R using $V_{IN} - V_{OUT}$ and V_{OUT}. Since the delay cells are inversely proportional to the input current I_F and I_R, the D_F and D_R are inversely proportional to $V_{IN}–V_{OUT}$ and V_{OUT}, respectively. When the necessary duty (D) is increased by the input supply (V_{IN}) decrease, D_F also increased accordingly, therefore the coarse digital can cope with the input supply variation. Figure 5-7 (b) shows the generated duty cycles for two different VCR (M = 0.3 and 0.4) with the necessary duties calculated using (3) (drawn with dot lines). While there are errors compared to the ideal cases, the generated delay is following the general tendency within the give load currents. The errors do not make any troubles in the overall control because it will be compensated in the fine analog control. The other function the I_L Emulator provides is the indirect zero current detection (ZCD). The ratio of the D_F to D_R is critical to prevent reverse I_L which occurs in the DCM mode operation. Under the DCM mode of the operation with synchronous design, the ZCD functions for the detection and elimination of the reverse inductor current is mandatory for the low power operation. In the proposed architecture the necessary condition for the ZDC is automatically satisfied by generating D_F and D_R which meet (5-4). For the perfect ZCD function in a buck converter, (4) should be met where D_1 and D_2 are the on time of PMOS and NMOS, respectively and L is the inductance [118].

$$\frac{V_{OUT} - V_{IN}}{L} D_1 = \frac{V_{OUT}}{L} D_2 \tag{5-4}$$

As shown in Figure 5-7 the driving currents, I_F are I_R for the forward and reverse delays are formed through the two transconductance cells having V_{IN}, V_{OUT}, and GND their inputs in a way that I_F and I_R are proportional to $V_{IN} - V_{OUT}$ and V_{OUT}. Since the delays are inversely

proportional to the driving current, D_F and D_R are inversely proportional to $V_{IN} - V_{OUT}$ and V_{OUT}. Consequently, (4) are satisfied thanks to the fact that the total delays in forward and reverse are D_1 and D_2 which are $n \times D_F$ and $n \times D_R$, and consequently the ZCD function is achieved. In nominal condition where V_{IN} = 3.3 V and V_{OUT} = 1 V the ratio of D_1 to D_2 is 2.3 ideally and becomes about 2.2308 considering the nominal values of D_F and D_R.

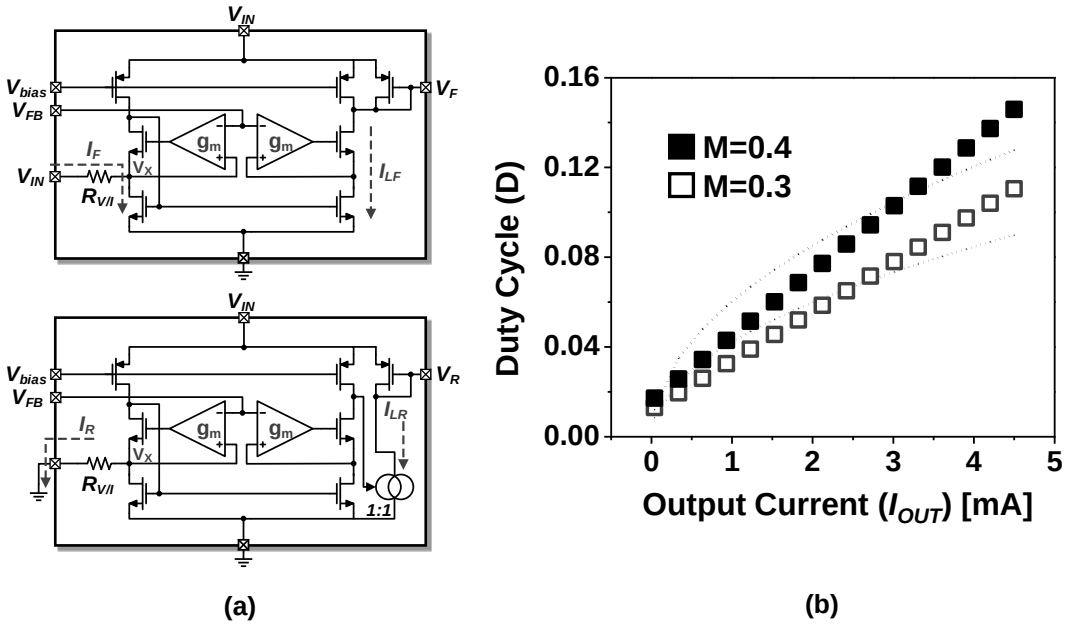


Figure 5-7 V/I converters inside IL Emulator (a) and Squared symbols are the generated duties (D) in two different VCR (M = 0.3 and 0.4) and the dotted lines are the necessary duties for each case (b)

The operation of the coarse digital control is conceptually shown in Figure 5-8 (a). If the 'n' is assumed to be a calculated value from the compensator, the clock (*CLK*) propagates toward the n^{th} tap of the forward delay (from left to right) while turning on the unit-sized PMOS power transistors sequentially. Once the clock reaches to the n^{th} tap, it turns off all the PMOS power transistors at once, then the *CLK* propagates backward to the start point. At the beginning of the backward traveling, the 'n' numbers of NMOS power

transistors are turned-on at once and they are turned off sequentially while the clock is travelling in the D_R delay line. Figure 5-8 (b) depicts the corresponding waveforms of the clock, PMOS and NMOS driving signals (V_{PMOS}, V_{NMOS}), and the inductor current (I_L). As the rising edge of the *CLK* coming each V_{PMOS} go to the logic 'H' to drive the PMOS transistors sequentially, thus I_L increased as the PMOS transistors becomes turned on. Once *CLK* reaches to the designated tap (*n*), the all of the V_P go back to the logic L and all V_N are changed into the logic 'H' state after the fixed dead time (the dead time is not shown.) and go back to the logic 'L' state as *CLK* travels back. Consequently, I_L decreases as *CLK* propagates backward. Figure 4-9 (a) shows the schematics of the single tap of the variable delay DPWM and the unit delay cell where the delay tap is assumed to be selected by the compensator. The variable delay DPWM consists of 16 D_F, D_R and D_T and logic gates corresponding to the 16b compensator. The variable delay DPWM is implemented using logic gates, and thus does not dissipate the static power. The delay cell has the fully-differential CMOS thyristor structure adopted from [119]. Since the delay cell provides fast state flipping thanks to the positive feedback from M_1 and M_2 when asserted by *CLK*, it is robust to the supply variation. At 2.5 V supply the delay varies less than 5 % for the nominal value at 3.3 V. The transistor M_3 provide the reference (I_F or I_R) current and the additional current can be added and subtracted at node V_C to change the delay (also denoted as $\pm I_E$ in Figure 5-7(b)). At 3.3V nominal supply voltage D_F, D_R, and D_T are 4.5 ns, 16.5 ns and 2 ns, respectively. The single delay cell consumes about 0.5 µW at 3.3 V with 1 MHz clock.

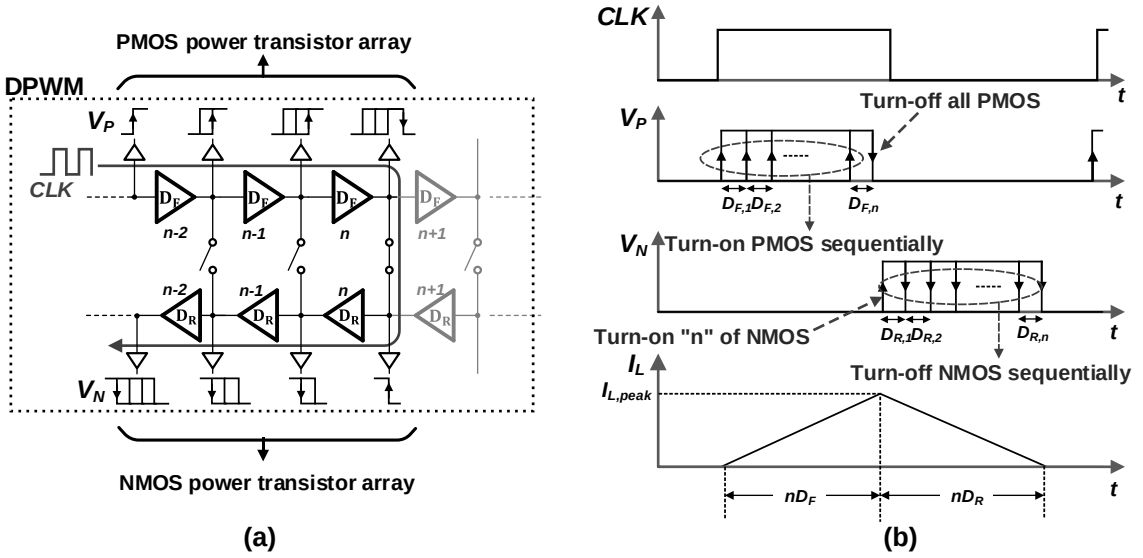


Figure 5-8 (a) Operation of the proposed coarse digital control and (b) corresponding waveforms of V_P and V_N and I_L from the operation of (a)

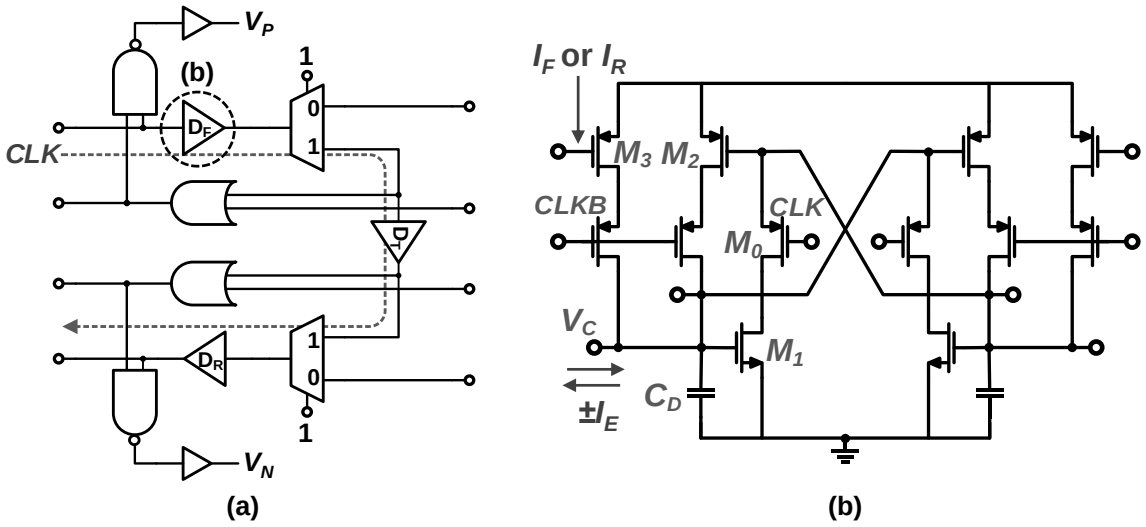


Figure 5-9 (a) Single tap from the variable delay DPWM where the tap is assumed to be selected by the compensator. (b) Schematic of the single delay cell (D_F or D_R)

5.4.3 LCO in coarse digital controller

A specific property of the digital controller is that the small oscillations of the output voltage around V_{REF} could occur in steady state. Those oscillations are caused by nonlinear

quantization effects in the ADC and the variable delay DPWM. Generally, when the DPWM resolution is low, compared to that of the ADC, for some operating conditions quantized DPWM output cannot results in steady-state error. Then, the voltage loop compensator changes the duty ratio control signal between two or more adjacent discrete duty ratio values and the oscillation known as limit cycle oscillations (LCO) occur [120]. In the proposed scheme, particularly, the LCOs come from the coarse control nature, i.e., the 1b-ADC which has no steady state conditions. Although undesirable in steady-state, the LCOs contain useful information about the controlled systems. For example, [121] extracted the amplitudes and frequencies of the LCOs and used them to tune the PID coefficients in the controller. In the proposed architecture, the average duty (D_{avg}) is extracted from the oscillation at V_{OUT} and D_{avg} is provided for the fine analog control. When observing the digital control, it can be understood as 1st order 1bit incremental $\Delta\Sigma$ modulator as shown in Figure 5-10 where the Δ functions is performed in the power converter itself and the Σ function is accomplished in the a digital integrator (compensator). The filter (F (n)) takes the output of the integrator and average of it (D_{avg}), thus the output of F (n) is able to provide the estimated duty necessary for the control of the converter once enough time for the calculation of D_{avg} is allowed. The resolution of D_{avg} is given by (5-5) where N and M are the resolution and data points to be used for the average calculation, respectively [122].

$$N = \log_2 M \text{ (bit)} \tag{5-5}$$

To calculate D_{avg} with enough resolution (enough data points, M) the frequencies of LCOs should be known in advance. In addition to that, the amplitude information of the LCOs should be provided for the system to capture them. Generally, the analytical equation

for the LCO is hard to get and needs some assumptions if the describing function for the estimation is used [120][123]. Instead of the analytical description, in this work the extensive numerical simulations were performed to measure the characteristics of the LCOs with the fixed power converter design parameters like V_{IN}, V_{OUT}, f_{SW}, L, and C. Figure 5-11 (a) shows a typical waveform of LCOs and the extracted absolute amplitudes ($|A_{LCO}|$) and periods (T_{LCO}) while the load current varies from 30μA to 5mA. As shown in Figure 5-11 (b), the amplitude and frequency of LCOs are bounded within 3–24 mV and 25–125 KHz, respectively with the given load specifications. From the known T_{LCO}, the M for the calculation of D_{avg} can be set from 64–200 points which guarantees the resolution of D_{avg} to be at least 4.8bit using (4). The A_{LCO} is also utilized to distinguish the onset of the LCOs from the transient responses of V_{OUT} in the mode arbiter.

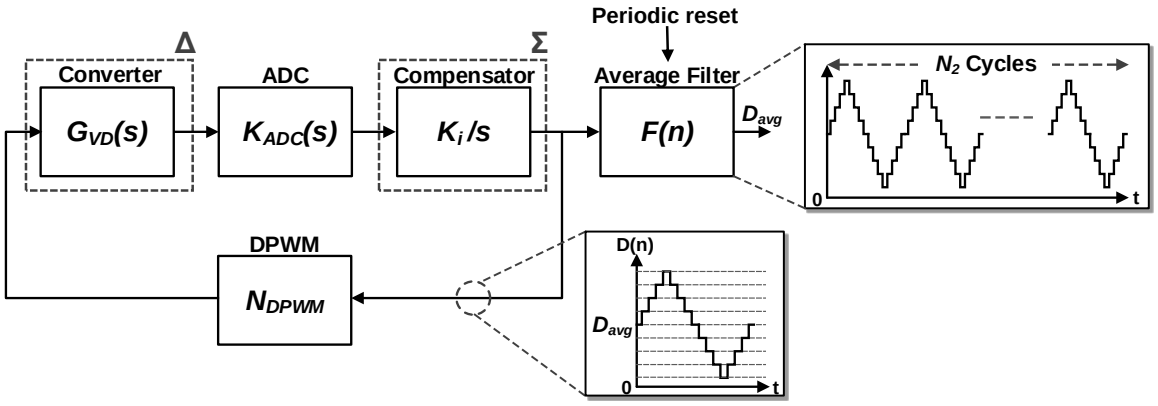


Figure 5-10 Functional block diagram describing the coarse digital control. The converter and the compensator actuates as Δ and Σ function, respectively

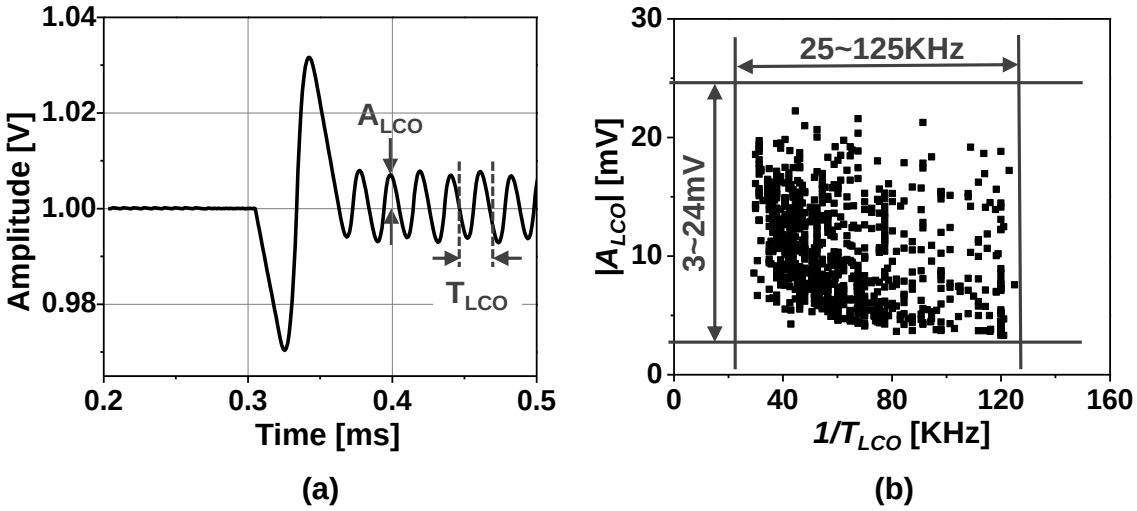


Figure 5-11 (a) A typical waveform of LCO. (b) 1000 cases of $|A_{LCO}|$ and T_{LCO} derived from the converter

5.4.4 Mode arbiter

Figure 5-12 shows the schematic of the mode arbiter which governs the transition between the analog-digital control modes and provides D_{avg}. Based on knowledge of the extracted $|A_{LCO}|$ and T_{LCO} the time window and voltage window to watch and calculate D_{avg} from LCOs are set to be up to 200 μs and 100 mV, respectively. When the output converges into the designated 100 mV voltage window (V_{REFH}–V_{REFL}), N_1-bit counter starts counting the number of zero-crossings of the output. Until the number of zero-crossings reaches a threshold the average filter is inactive by the stop command from the N_1-bit counter and waits for the stabilization of LCOs. After zero crossings of N_1 times, the average filter is initiated and takes the output of the digital compensator, n, and generates D_{avg} (N) during N_2–N_1 counts. Figure 5-13 conceptually illustrates the operation of the mode arbiter. When the output stays outside of the voltage window, from A to B, the two counters are frozen by the two comparators. After the output converges into the window, the N_1- and N_2-bit counters start counting up their values. The average filter starts calculating from C and

stops at D. At D, the duty cycle is fixed into '*N*' from the average filter and then the analog fine control is initiated. For flexibility of the operation, N_1 and N_2 are programmable to be 2–4 and 6–12, respectively.

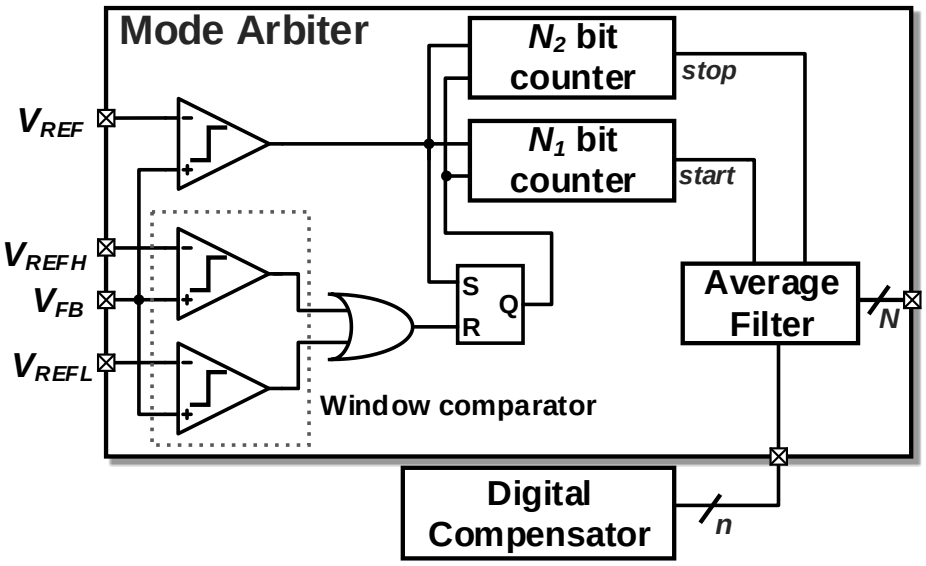


Figure 5-12 Schematic of the mode arbiter

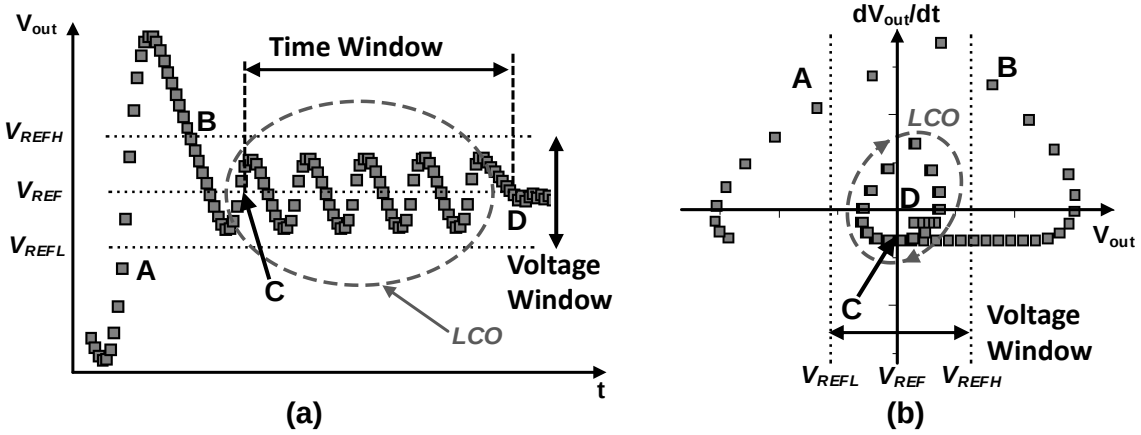


Figure 5-13 Conceptual operation of the mode arbiter (a) operation sequence (A to D) overlaid on the typical waveform of LCO (b) phase diagram of (a)

5.4.5 Fine analog controller

Once the mode arbiter estimates D_{avg} (N) the operation of the ADC and variable delay DPWM are frozen, and the analog control is initiated to begin the fine adjustment by the mode arbiter. Figure 5-14 shows the block diagram for the analog control. In this control the finite V_E is amplified and converted into the error current (I_E) through the error amplifier (transconductance cell, g_m), I_E is added/subtracted from I_F and I_R generated by the I_L emulator to decide an appropriate delay of the last tap. Thus, the forward and reverse delays in the last tap are inversely proportional to $I_F \pm I_E$ and $I_R \pm I_E$. The dynamic range of delays in the last tap is designed as $\pm 1.5 D_F$ and $\pm 1.5 D_R$ of the digital transient mode to fully cover both the error from the coarse digital control and the rounding error of the digital average filter. Figure 5-15 describes the operation of the fine analog control. *CLK* propagates back and forth through '*N*' tap in the same way as the coarse digital control in the transient mode. Figure 5-15 (b) also illustrates the waveforms of *CLK*, V_P, V_N, and I_L. The different operation from the coarse digital control is that the delay of the last (N^{th}) tap is adjusted not only by the I_L Emulator, but also by V_E in the fine analog control, and consequently I_L is changed to $I_L \pm \Delta I_L$ to compensate the error voltage. Figure 5-16 shows the half circuit of a delay cell where D_F or D_R is modulated by V_E. The output of the error amplifier is expressed as g_m times V_E (dependent current source). After the mode arbiter finishes the calculation of D_{avg}, it generates V_{LOCK} to connect the error amplifier to the last tap. Then, the delay is adjusted by I_E to compensate V_E as shown in Figure 5-16 (b).

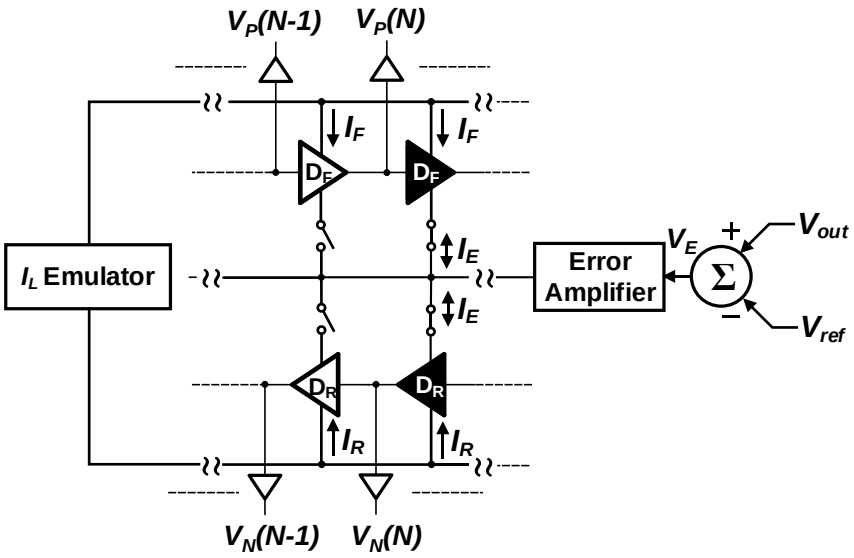


Figure 5-14 Block diagram for the analog control

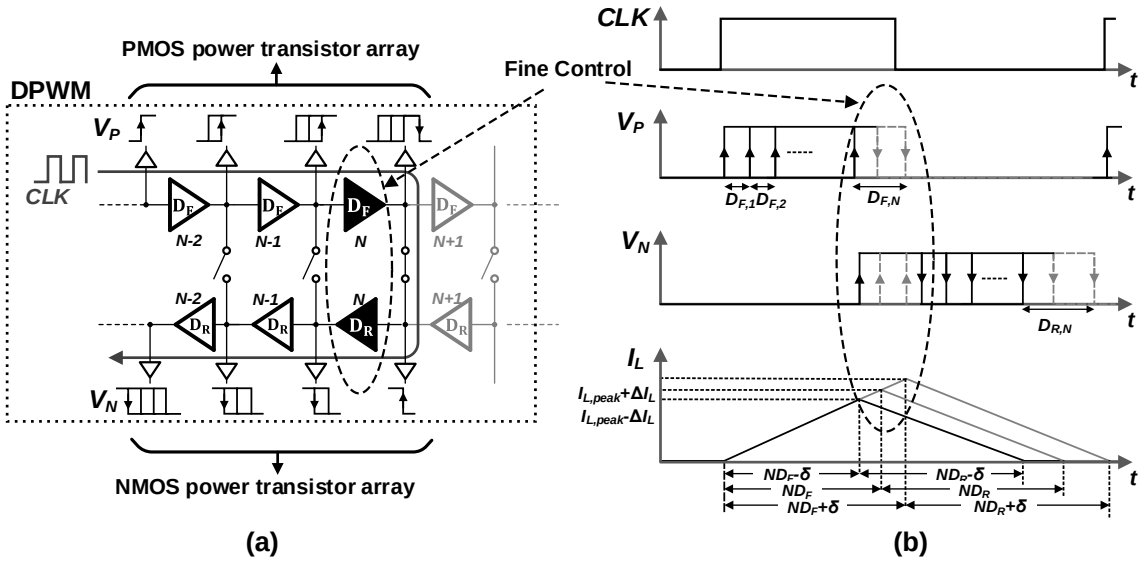


Figure 5-15 (a) Operation of the proposed fine analog control. (b) Corresponding waveforms of V_P, V_N, and I_L from the operation of (a)

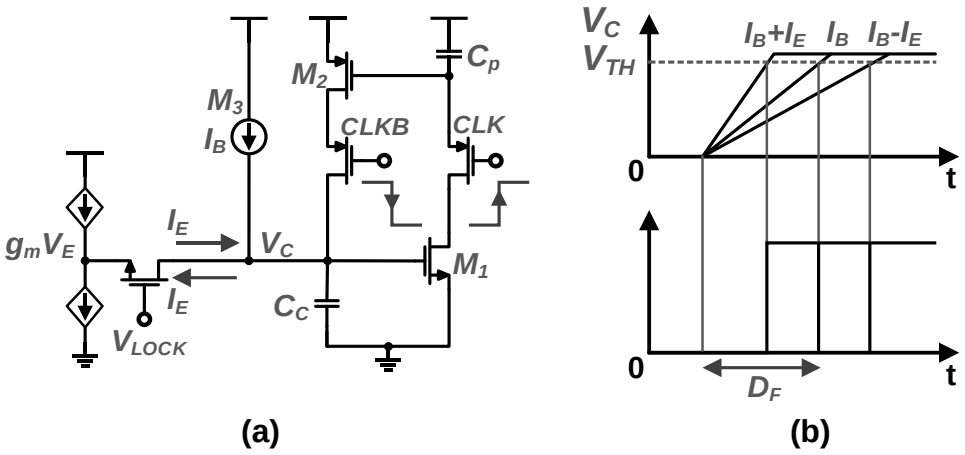


Figure 5-16 (a) A half circuit of the delay cell inside the last tap of the variable delay DPWM and (b) variation of delay (DF) by the additional I_E.

5.5 Experiment results

The proposed PWM converter was fabricated in 0.18 µm standard CMOS process with 1P6M. The chip is implemented within 1250 µm × 300 µm, the area of 0.375 mm^2 using only 3.3V I/O transistors. The chip microphotograph is depicted in Figure 5-17. Some important blocks such as the I_L emulator, power transistors, variable delay DPWM, and mode arbiter are highlighted. Majority of the occupied silicon area is accounted for the implementation of the array of the power transistors and their gate drivers.

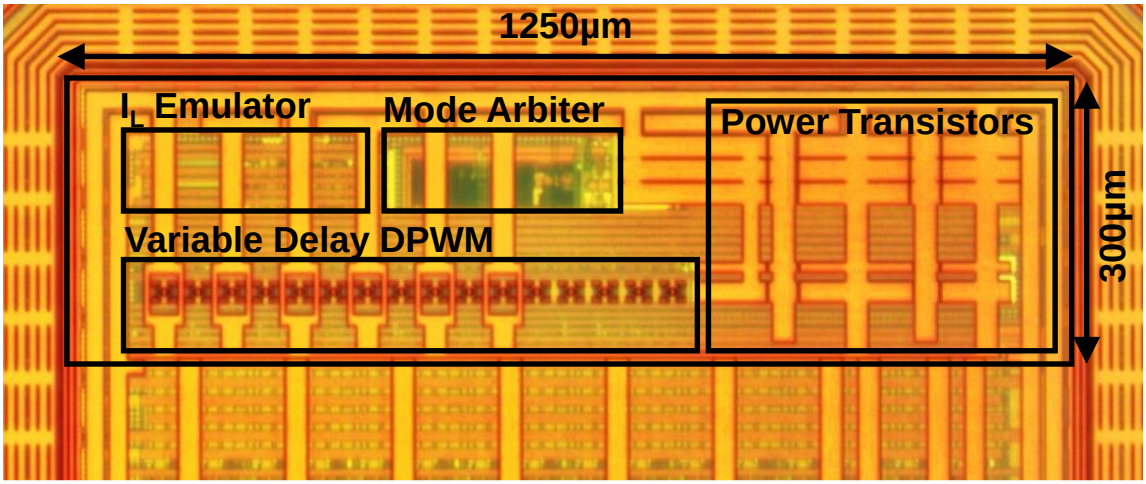


Figure 5-17 Chip microphotograph of the proposed PWM Buck converter

5.5.1 Transient response

The transient response of V_{OUT} by the abrupt output current change from 50 µA to 4 mA is shown in Figure 5-18. The converter operates at about 1 MHz frequency and the desired V_{OUT} is 1 V. Once the output is out of the voltage window by the current change, the mode arbiter unlocks the analog control, which is indicated with the mode arbiter start signal. As shown in the enlarged figure of the Figure 5-18 the LCOs are initially observed and are used by the mode arbiter to calculate D_{avg} as explained in the section IV. Finally, the LCO is suppressed and there's no oscillation in the output. In the maximum current transient the drop of V_{OUT} is about 240 mV. Figure 5-19 depicts the waveforms of the inductor voltage (V_X), *CLK*, and V_{OUT} when the converter is in steady state (governed by the fine analog control). In Fig. 18 the converter is providing 900 µA output current from the 3.3*V* input supply. No steady state oscillation (LCO) is observed at V_{OUT}. From V_X, the ZCD function can be confirmed. The on-times of PMOS and NMOS transistors are 39 ns and 77 ns, respectively so the ratio of two values is about 1.97. Even though there is some deviation from the ideal value, 2.3, there is no significantly power loss observed by the reverse I_L thanks to the proposed indirect ZCD function.

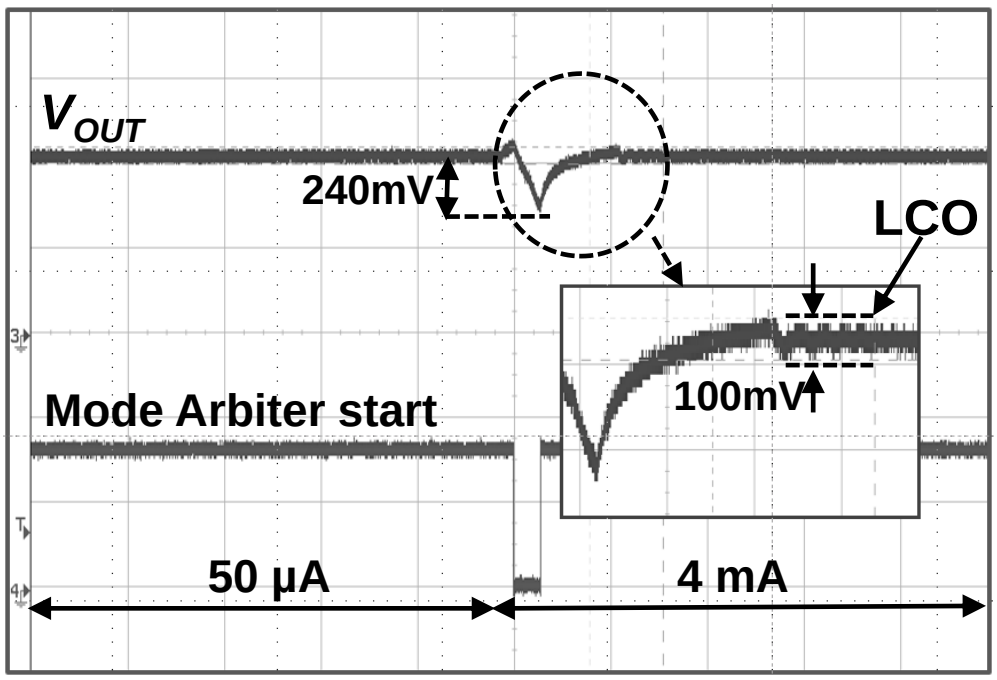


Figure 5-18 Step current response of the proposed PWM buck converter. I_{OUT} is changed from 50 µA to 4 mA. Measured at V_{OUT} = 1.0 V from V_{IN} = 3.3 V with f_{SW} = 0.96 MHz

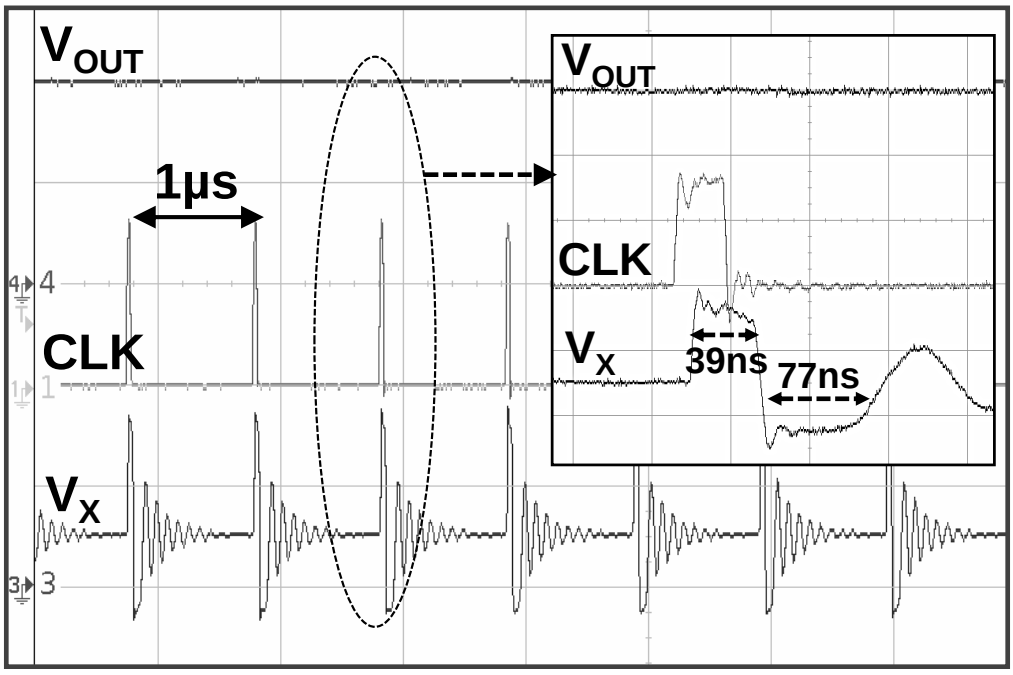


Figure 5-19 Steady-state response of V_{OUT} and V_X. The converter is providing I_{OUT} = 900 µA and V_{OUT} = 1 V from VIN = 3.3 V

5.5.2 Power consumption and PCE

Figure 5-20 shows the separately measured power dissipation of the analog and digital blocks in the proposed buck converter while sweeping the output current from 35 µA to 4.1 mA. The digital block only consumes about 6µW at the lightest load, I_{OUT} = 35 µA. For the comparison, 80 µW horizontal line is drawn which is the calculated SW loss where the size of the power transistors are not scaled and it size is fixed as the maximum value (W_P = 1.6 mm and W_N = 0.95 mm). Without scaling of the power transistor width the calculated PCE is about 34 % (See Figure 5-2), therefore we can estimate that the improvement of the PCE at the lightest load is 37 %. The analog blocks including all transconductors in the system consume only 7 µW which is constant across the all possible output load conditions. The overall PCE is also measured from the 3.3V nominal supply while sweeping the full range of I_{OUT} as shown in Figure 5-21. The PCE is 71 % at 35 µA load current, however, it recovers and reaches to over 80% from 45 µA load current. The 86.3% maximum PCE is measured at 1.4 mA load current.

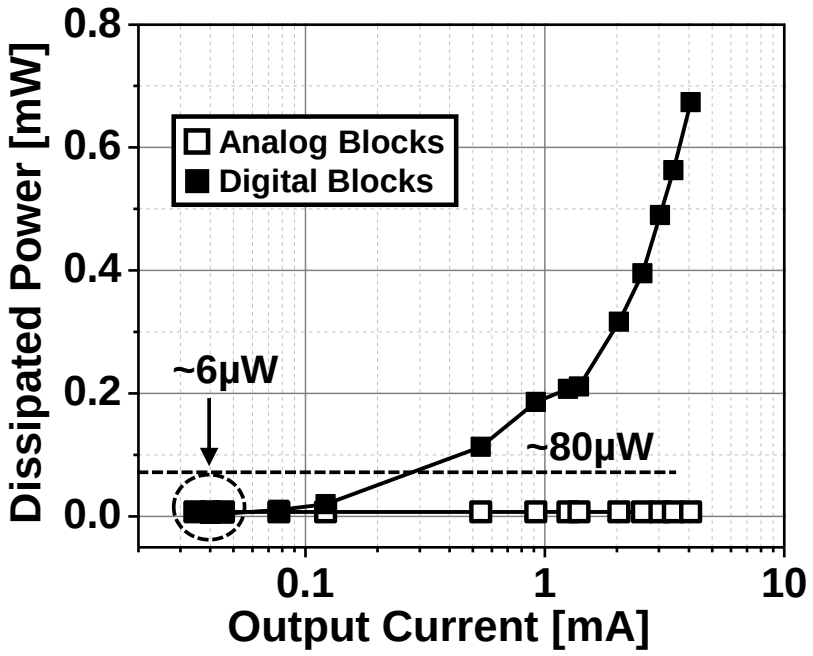


Figure 5-20 Power dissipation of analog and digital blocks

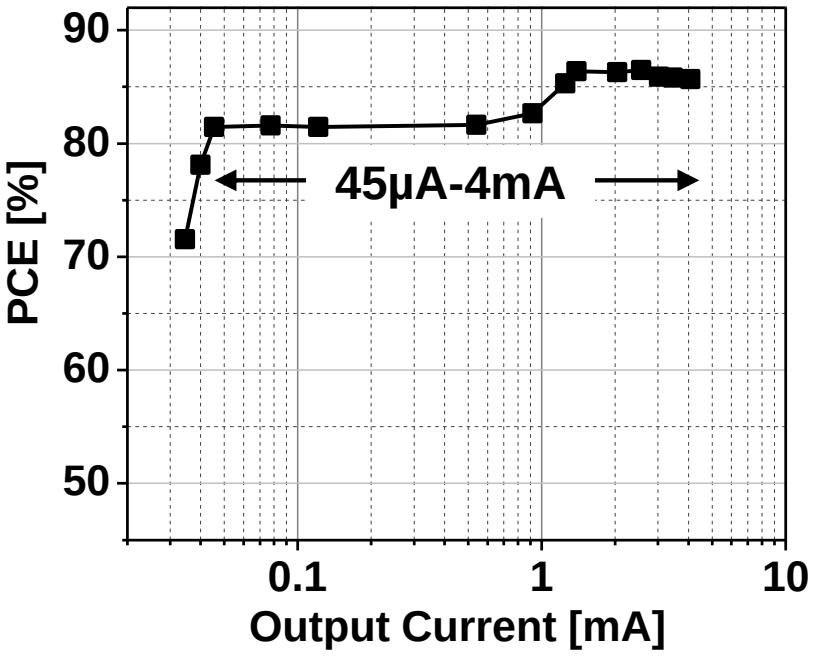


Figure 5-21 Measured power conversion efficiency at V_{OUT} = 1V from V_{IN} = 3.3V and f_{SW} = 0.96 MHz

5.5.1 Comparison

The comparisons with the other state-of-the-art power converters designed for the improvement of the light load efficiency with different control techniques are shown in Table 5-1. [42-43] used the PWM control while [30] adopted the mode hopping between PWM and PFM. As mentioned in the section II, [36] utilized a set of frequencies according to the different loading. Compared to the other works, the proposed converter is able to facilitate more than 100 times smaller loads driven by the PWM control while maintaining over 80 % PCE. In addition to that, the switching frequency and the values of the external inductor and capacitor stay similar range with other works.

Table 4-1 Performance comparison with the state-of-the-art dc-to-dc converters

	This work	[99]	[103]	****[111]	[112]
Technology	**0.18µm CMOS**	45nm CMOS	0.35µm CMOS	1.5µm CMOS	0.5µm CMOS
Input voltage	**2.5–3.3V**	2.8–4.2V	1.8 – 3V	5V	3.6V
Output voltage	**1V**	0.6–1.2V	0.9V	2.5V	1/1.5/1.8V
Output current range	**0.035–4mA**	0.020-10mA (PFM) 10-100mA (PWM)	5–500mA	20–80mA	10–400mA
LC filter	**L=6.8µH, C=1.2µF**	L=10µH C=2µF	L=2.2µH, C=2.2µF	L=0.1µH, C=0.03µF	L=3.3µH, C=4.7µF
Operating method	**PWM**	PFM/PWM	[1]ASFPWM	PWM ([2]WS)	PWM ([3]GCM)
f_{SW} @ PWM	**0.96MHz**	N/A	0.25/0.5/1/2 MHz	10MHz	3MHz
PWM driven load range with > 80% PCE	**45µA–4mA**	*10mA–100mA	40 –500mA	Max. PCE < 80%	20–400mA
Peak Efficiency	**86.5%**	87.4%(PFM) 87.2%(PWM)	*~90%	*~40%	*~89%

[1]ASFPWM: Auto Selectable Frequency Pulse Width Modulation, [2]WS: Width Scaling
[3]GCM: Gate Charge Modulation, *Estimated

5.6 Summary and chapter conclusion

In this paper, we have presented a 1 MHz PWM buck converter providing the load-adaptive scalable sizing of power transistors with the analog-digital hybrid controller for the application of the implantable biomedical systems. To minimize power consumption of the power converter, especially at ultra-light loads, the digital controller dynamically changes the size of the power transistor without any dedicated blocks for the purpose. Moreover, the digital controller is minimally implemented and the arising error from its coarse nature is effectively compensated by the analog control. Thus the proposed buck converter is able to achieve high PCE of larger than 80% at less than 100 μA loads. Furthermore, thanks to its constant switching frequency the buck converter does not generate any unpredictable output spurs. The buck converter showed 71% efficiency at minimum 35 μA load and 86.3% peak efficiency at 1.4 mA. From 45 μA to 4.1 mA loads, it achieved over 80% efficiency generating 1V output from a 3.3V single supply. The converter occupies 0.375 mm^2 of active area using 0.18 μm standard CMOS process and only requires one external 1.2 μF capacitor and one 6.8 μH inductor. Therefore, the proposed converter is suitable for implantable neural recording microsystems consuming small current while requiring high energy efficiency and supply integrity.

CHAPTER 6
SUMMARY AND FUTURE WORK

6.1 Summary

The goal of this research is to provide the architecture and associated electronics for the 1,024-channel neural recording microsystems as shown in Figure 1-2. In order to achieve this goal, it is important to address specific circuit and architecture design challenges; minimizing area and energy consumption while not compromising circuit performance such as noise, resolution, robustness, and robustness. The three interdependent projects; modular 128-channel Δ-ΔΣ analog front-end (AFE) circuit, on-chip neural signal compressor, and a dc-to-dc converter for power management have been carefully designed to accomplish this goal.

The modular 128-channel Δ-ΔΣ analog front-end (AFE) was designed using the spectrum shaping (or equalizing) technique and fabricated in TSMC 0.18 μm CMOS process. The fabricated AFE is able to be expandable to a 1,024-channel parallel recording system via hybrid assembly with the customized 3-dimension platform. The AFE achieved 4.84 fJ/C−s·mm^2 figure of merit that is the smallest the area-energy product up to date.

Then, the on-chip mixed signal neural signal compressor with a 128-channel AFE was built in the same CMOS process to minimize the energy consumption overhead in digital blocks, especially, in an I/O block. The compressor reduces the power consumption in the digital blocks transmitting LFP and AP by 1/5.35 and 1/10.54 of the uncompressed cases,

respectively. The measured power per channel consumption in digital blocks is 11.98 μW that is about 1/9 of 107.5 μW that is the case without using the compressor.

For the sake of the overall system, not only for the energy usage efficiency, but the efficient energy delivery to the system is also considered. The on-chip dc-to-dc converter (PWM buck) was developed to facilitate reliable power supplies and enhance energy delivery efficiency to the 1,024-channel parallel neural recording system. The dc-to-dc converter has predictable and removable tones at the output and it exhibits > 80% power conversion efficiency at ultra-light loads, < 100 μW that is relevant power most of the multi-channel neural recording systems consume.

6.2 Future works

Although several contributions have been made in this research to realize the 1,024-channel parallel recording microsystem, there are still some areas of improvements in the design. For further improvements and future work the followings are suggested.

- Hybrid assembly to realize the proposed 1,024-channel neural recording system is one of the most urgent thing.
- An event-driven (or asynchronous) digital interface circuit is necessary. Even though the fabricated compressor effectively reduces data rate, the power consumption reduction of the digital interface was not as significant as that of the data rate since the current digital interface constantly operates (synchronous operation). The power consumption in the digital interface is reduced significantly if the interface can stay in sleep mode where there is no spike.
- Integration of the designed dc-to-dc converter with the AFE is necessary. The power converter was designed and fabricated independently and its performance was

measured separately. To realize the complete system, the power converter should be on a single die with the analog front-end (AFE).

- The proper interface and protocol with the outside storage place should be developed. Currently, interface DUT with the external data storage place (a PC) is built with NI instruments which has only 25 Mbps bandwidth. That bandwidth is not enough to handle 1,024-channel simultaneous recording. In addition, the communication protocol developed and applied to the AFE is not the standardized one. For future expandability and applicability, the AFE should follow standard communication protocol.
- To further reduce the data rate and save the energy consumption in the system, digital signal processing (DSP) units for the feature extraction and possibly, classification are necessary.

BIBLIOGRAPHY

[1] S. A. Hodgkin, "Edgar Douglas Adrian, Baron Adrian of Cambridge," *Trans. Am. Neurol. Assoc.*, vol. 103, no. August 1977, pp. 303–304, 1978.

[2] I. H. Stevenson and K. P. Kording, "How Advances in Neural Recording Affect Data Analysis.," *Nat. Neurosci.*, vol. 14, no. 2, pp. 139–42, Feb. 2011.

[3] J. N. Y. Aziz, K. Abdelhalim, R. Shulyzki, R. Genov, B. L. Bardakjian, M. Derchansky, D. Serletis, and P. L. Carlen, "256-Channel Neural Recording and Delta Compression Microsystem With 3D Electrodes," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 995–1005, Mar. 2009.

[4] M. Chae, W. Liu, Z. Yang, T. Chen, J. Kim, M. Sivaprakasam, and M. Yuce, "A 128-Channel 6mW Wireless Neural Recording IC with On-the-Fly Spike Sorting and UWB Tansmitter," in *IEEE International Solid State Circuits Conference*, 2011, pp. 146–148.

[5] S.-Y. Park, J. Cho, K. Na, and E. Yoon, "Toward 1024-Channel Parallel Neural Recording : Modular Δ-ΔΣ Analog Front-End Architecture with 4.84fJ/C-s·mm2 Energy-Area Product," in *IEEE Symposium on VLSI Circuits*, 2015, pp. 112–113.

[6] C. M. Lopez, S. Mitra, J. Putzeys, B. Raducanu, M. Ballini, A. Andrei, S. Severi, M. Welkenhuysen, C. Van Hoof, S. Musa, and R. F. Yazicioglu, "A 966-electrode neural probe with 384 configurable channels in 0.13 μm SOI CMOS," in *IEEE International Solid-State Circuits Conference*, 2016, pp. 392–393.

[7] T. M. Seese, H. Harasaki, G. M. Saidel, and C. R. Davies, "Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of muscle tissue to chronic heating.," *Lab. Invest.*, vol. 78, no. 12, pp. 1553–1562, 1998.

[8] R. Muller, S. Gambini, and J. M. Rabaey, "A 0.013mm 5 μW DC-Coupled Neural Signal Acquisition IC with 0.5V Supply," in *IEEE International Solid State Circuits Conference*, 2011, pp. 302–304.

[9] S.-J. Kim, L. Liu, L. Yao, W. L. Goh, and M. Je, "A 0.5-V Sub- μW / Channel Neural Recording IC with Delta-Modulation-Based Spike Detection," in *IEEE Asian Solid-State Circuits Conference*, 2011, pp. 219–222.

[10] W. Liew, X. Zou, and Y. Lian, "A 0.5-V 1.13-μW/Channel Neural Recording Interface with Digital Multiplexing Scheme," in *European Solid State Circuits Conference*, 2011, pp. 219–222.

[11] D. Han, Y. Zheng, R. Rajkumar, G. Dawe, and M. Je, "A 0.45V 100-Channel Neural-Recording IC with Sub-μW/Channel Consumption in 0.18 μm CMOS," in *IEEE International Solid State Circuits Conference*, 2013, pp. 290–292.

[12] K. M. Al-ashmouny, S. Chang, and E. Yoon, "A 4 μW/Ch Analog Front-End

Module With Moderate Inversion and Power-Scalable Sampling Operation for 3-D Neural Microsystems," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 5, pp. 403–413, 2012.

[13] D. H. Hubel, "Tungsten microelectrode for recording from single units," *Science (80-.).*, vol. 125, no. 3247, pp. 549–550, 1957.

[14] S. Url, T. J. Archive, and T. Archive, "Long-Term Recording from Single Neurons in Brain of Unrestrained Mammals Felix Strumwasser Long-Term Recording from Single Neurons in Brain of," vol. 127, no. 3296, pp. 469–470, 2008.

[15] D. Banks, "Neurotechnology," *IEEE Eng. Sci. Eductation*, vol. 7, no. June, pp. 135–44, 1998.

[16] K. C. Cheung, "Implantable microscale neural interfaces," *Biomed. Microdevices*, vol. 9, no. 6, pp. 923–938, 2007.

[17] K. D. Wise, J. B. Angell, and A. Starr, "An Integrated-Circuit Approach to Extracellular Microelectrodes," *IEEE Trans. Biomed. Eng.*, vol. BME-17, no. 3, pp. 238–247, 1970.

[18] K. D. Wise and J. B. Angell, "A low capacitance multielectrode probe for use in extracellular neurophysiology," *IEEE Trans. Biomed. Eng.*, vol. BME-22, no. 3, pp. 212–220, 1975.

[19] K. Najafi, K. D. Wise, and T. Mochizuki, "A High-Yield IC-Cfompatible Multichannel Array Reco rding," *Electron Devices IEEE Trans.*, vol. 32, no. 7, pp. 1206–1211, 1985.

[20] J. Ji, K. Najafi, and K. D. Wise, "A Scaled Electronically-configurable Multichannel Recording Array Functional Design for On-chip Signal praceeging," *Sensors And Actuators*, vol. 23, pp. 589–591, 1990.

[21] A. C. Hoogerwerf and K. D. Wise, "A Three-Dimensional Microelectrode Array for Chronic Neural Recording," *IEEE Trans. Biomed. Eng.*, vol. 41, no. 12, pp. 1136–1146, 1994.

[22] R. H. Olsson and K. D. Wise, "A Three-Dimensional Neural Recording Microsystem With Implantable Data Compression Circuitry," *IEEE J. Solid State Circuits*, vol. 40, no. 12, pp. 2796–2804, 2005.

[23] P. K. Campbell, K. E. Jones, R. J. Huber, K. W. Horch, and R. A. Normann, "A silicon-based, three-dimensional neural interface: Manufacturing processes for an intracortical electrode array," *IEEE Trans. Biomed. Eng.*, vol. 38, no. 8, pp. 758–768, 1991.

[24] M. HajjHassan, V. Chodavarapu, and S. Musallam, "NeuroMEMS: Neural probe microtechnologies," *Sensors*, vol. 8, no. 10, pp. 6704–6726, 2008.

[25] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A Low-Power Integrated Circuit for a Wireless 100-Electrode Neural Recording System," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, 2007.

[26] R. M. Walker, H. Gao, P. Nuyujukian, K. Makinwa, K. V Shenoy, T. Meng, and B.

Murmann, “A 96-Channel Full Data Rate Direct Neural Interface in 0.13μm CMOS,” in *IEEE Symposium on VLSI Circuits*, 2011, pp. 144–145.

[27] A. Sodagar, G. Perlin, and Y. Yao, “An Implantable 64-Channel Wireless Microsystem for Single-Unit Neural Recording,” *IEEE J. Solid-State Circuits*, vol. 44, no. 9, pp. 2591–2604, 2009.

[28] K. Seidl, M. Schwaerzle, I. Ulbert, H. P. Neves, O. Paul, and P. Ruther, “CMOS-Based High-Density Silicon Microprobe Arrays for Electronic Depth Control in Intracortical Neural Recording–Characterization and Application,” *J. Microelectromechanical Syst.*, vol. 21, no. 6, pp. 1426–1435, Dec. 2012.

[29] M. Kuhl and Y. Manoli, “A 0.01 mm2 fully-differential 2-stage amplifier with reference-free CMFB using an architecture- switching-scheme for bandwidth variation,” in *IEEE European Solid-State Circuits Conference*, 2015, pp. 287–290.

[30] C. M. Lopez, A. Andrei, S. Mitra, M. Welkenhuysen, W. Eberle, C. Bartic, R. Puers, R. F. Yazicioglu, and G. Gielen, “An Implantable 455-Active-Electrode 52-Channel CMOS Neural Probe,” in *IEEE International Solid-State Ciircuits Conference*, 2013, pp. 288–290.

[31] R. Muller, H.-P. Le, W. Li, P. Ledochowitsch, S. Gambini, T. Bjorninen, A. Koralek, J. M. Carmena, M. M. Maharbiz, E. Alon, and J. M. Rabaey, “A Miniaturized 64-Channel 225 μW Wireless Electrocorticolgraphic Neural Sensor,” in *IEEE International Solid State Circuits Conference*, 2014, pp. 412–414.

[32] L. Lundy-Ekman, “Neuroscience Fundamentals for Rehabilitaton,” in *Philadelphia : W.B. Saunders*, 2002.

[33] G. Buzsáki, C. a Anastassiou, and C. Koch, “The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes.,” *Nat. Rev. Neurosci.*, vol. 13, no. 6, pp. 407–20, Jun. 2012.

[34] G. Buzsáki, “Large-Scale Recording of Neuronal Ensembles,” *Nat. Neurosci.*, vol. 7, no. 5, pp. 446–51, May 2004.

[35] J. Csicsvari, D. a Henze, B. Jamieson, K. D. Harris, A. Sirota, P. Barthó, K. D. Wise, and G. Buzsáki, “Massively Parallel Recording of Unit and Local Field Potentials with Silicon-Based Electrodes.,” *J. Neurophysiol.*, vol. 90, no. 2, pp. 1314–23, Aug. 2003.

[36] T. Jochum, T. Denison, and P. Wolf, “Integrated Circuit Amplifiers for Multi-Electrode Intracortical Recording.,” *J. Neural Eng.*, vol. 6, no. 1, p. 012001, Feb. 2009.

[37] R. Muller, S. Gambini, and J. M. Rabaey, “A 0.013 mm2, 5 μW, DC-Coupled Neural Signal Acquisition IC with 0.5 V Supply,” *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 1–12, 2012.

[38] S. Chang, K. Al-ashmouny, and E. Yoon, “A 0.5V 20fJ/Conversion-Step Rain-to-rail SAR ADC with Programmable Time-Delayed Control Units for Low-Power Biomedical Application,” in *IEEE European Solid State Circuit Conference*, 2011, pp. 8–11.

[39] S. Lee, S. Park, H. Park, and J. Sim, “A 21 fJ/Conversion-Step 100kS/s 10-bit ADC

With a Low-Noise Time-Domain Comparator for Low-Power Sensor Interface," *IEEE J. Solid-State Circuits*, vol. 46, no. 3, pp. 651–659, 2011.

[40] D. Zhang, A. Bhide, and A. Alvandpour, "A 53-nW 9.1-ENOB 1-kS/s SAR ADC in 0.13um CMOS for Medical Implant Devices," *IEEE J. Solid-State Circuits*, vol. 47, no. 7, pp. 1585–1593, 2012.

[41] M. S. Chae, W. Liu, and M. Sivaprakasam, "Design Optimization for Integrated Neural Recording Systems," *IEEE J. Solid State Circuits*, vol. 43, no. 9, pp. 1931–1939, 2008.

[42] V. Majidzadeh, A. Schmid, and Y. Leblebici, "A 16-Channel, 359 μW, Parallel Neural Recording System using Walsh-Hadamard Coding," in *IEEE Custom Integrated Circuits Conference*, 2013, vol. 2, no. 4, pp. 1–4.

[43] S. Venkatraman, C. Patten, and J. M. Carmena, "Exploiting the 1/f structure of neural signals for the design of integrated neural amplifiers.," in *IEEE Engineering in Medicine and Biology Society Conference*, 2009, vol. 2009, pp. 2050–3.

[44] C. Bédard, H. Kröger, and a. Destexhe, "Does the 1/f Frequency Scaling of Brain Signals Reflect Self-Organized Critical States?," *Phys. Rev. Lett.*, vol. 97, no. 11, p. 118102, Sep. 2006.

[45] S. Balagopal and V. Saxena, "A Low-Power Single-Bit Continuous-Time ΔΣ Converter with 92.5 dB Dynamic Range for Biomedical Applications," *J. Low Power Electron. Appl.*, vol. 2, no. 4, pp. 197–209, Jul. 2012.

[46] S. Yan and E. Sanchez-Sinencio, "A Continuous-Time ΔΣ Modulator With 88-dB Dynamic Range and 1.1-MHz Signal Bandwidth," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 75–86, 2004.

[47] S. Pavan and P. Sankar, "Power Reduction in Continuous-Time Delta-Sigma Modulators Using the Assisted Opamp Technique," *IEEE J. Solid-State Circuits*, vol. 45, no. 7, pp. 1365–1379, Jul. 2010.

[48] F. Zhang, J. Holleman, and B. P. Otis, "Design of Ultra-Low Power Biopotential Amplifiers for Biosignal Acquisition Applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 4, pp. 344–355, 2012.

[49] R. R. Harrison and C. Charles, "A Low-Power Low-Noise CMOS Amplifier for Neural Recording Applications," *IEEE J. Solid State Circuits*, vol. 38, no. 6, pp. 958–965, 2003.

[50] X. Zou, X. Xu, L. Yao, and Y. Lian, "A 1-V 450-nW Fully Integrated Programmable Biomedical Sensor Interface Chip," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1067–1077, Apr. 2009.

[51] J. A. Cherry and W. M. Snelgrove, "Excess Loop Delay in Continuous-Time Delta–Sigma Modulators," *IEEE Trans. Circuits Syst. Analog Digit. Signal Process.*, vol. 46, no. 4, pp. 376–389, 1999.

[52] B. E. Boser and B. A. Wooley, "The Design of Sigma-Delta Modulation Analog-to-Digital Converters," *IEEE J. Solid-State Circuits*, vol. 23, no. 6, pp. 1298–1308, 1988.

[53] M. Ortmanns and F. Gerfers, “Compensation of Finite Gain-Bandwidth Induced Errors in Continuous-Time Sigma–Delta Modulators,” *IEEE Trans. Circuits Syst. I*, vol. 51, no. 6, pp. 1088–1099, 2004.

[54] R. H. Walden, “Analog-to-Digital Converter Survey and Analysis,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, 1999.

[55] H. Gao, R. M. Walker, P. Nuyujukian, K. A. A. Makinwa, K. V Shenoy, B. Murmann, and T. H. Meng, “HermesE : A 96-Channel Full Data Rate Direct Neural Interface in 0.13 μm CMOS,” *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 1043–1055, 2012.

[56] C. M. Lopez, A. Andrei, S. Mitra, M. Welkenhuysen, W. Eberle, C. Bartic, R. Puers, R. F. Yazicioglu, and G. G. E. Gielen, “An Implantable 455-Active-Electrode 52-Channel CMOS Neural Probe,” *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 248–261, 2014.

[57] D. Han, Y. Zheng, R. Rajkumar, G. S. Dawe, and M. Je, “A 0.45 V 100-Channel Neural-Recording IC With Sub- μW/Channel Consumption in 0.18 μm CMOS.,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 6, pp. 735–46, 2013.

[58] W. Biederman, D. J. Yeager, N. Narevsky, J. Leverett, R. Neely, J. M. Carmena, E. Alon, and J. M. Rabaey, “A 4.78 mm2 Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels With Digital Compression and Simultaneous Dual Stimulation,” *IEEE J. Solid-State Circuits*, vol. 50, no. 4, pp. 1038–1047, 2015.

[59] K. A. Ng and Y. P. Xu, “A multi-channel neural-recording amplifier system with 90dB CMRR employing CMOS-inverter-based OTAs with CMFB through supply rails in 65nm CMOS,” in *IEEE International Solid-State Circuits Conference*, 2015, pp. 1–3.

[60] R. Muller, H. P. Le, W. Li, P. Ledochowitsch, S. Gambini, T. Bjorninen, A. Koralek, J. M. Carmena, M. M. Maharbiz, E. Alon, and J. M. Rabaey, “A Minimally Invasive 64-Channel Wireless μECoG Implant,” *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 344–359, 2014.

[61] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, “Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering,” *Neural Comput.*, vol. 16, no. 8, pp. 1661–87, Aug. 2004.

[62] S. M. Jalaleddine, C. G. Hutchens, R. D. Strattan, and W. a Coberly, “ECG data compression techniques--a unified approach.,” *Biomed. Eng. IEEE Trans.*, vol. 37, no. 4, pp. 329–43, 1990.

[63] G. Antoniol and P. Tonella, “EEG data compression techniques,” *IEEE Trans. Biomed. Eng.*, vol. 44, no. 2, pp. 105–114, 1997.

[64] D. Staudenmann, I. Kingma, A. Daffertshofer, D. F. Stegeman, and J. H. Van Dieen, “Improving EMG-based muscle force estimation by using a high-density EMG grid and principal component analysis,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 4, pp. 712–719, 2006.

[65] E. Carotti, J. C. De Martin, R. Merletti, and D. Farina, “Compression of surface EMG signals with algebraic code excited linear prediction,” *Med. Eng. Phys.*, vol.

29, no. 2, pp. 253–258, 2007.

[66] J. A. Norris, K. B. Englehart, and D. F. Lovely, “Myoelectric signal compression using zero-trees of wavelet coefficients,” *Med. Eng. Phys.*, vol. 25, no. 9, pp. 739–746, 2003.

[67] L. Brechet, M. F. Lucas, C. Doncarli, and D. Farina, “Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2186–2192, 2007.

[68] E. S. G. Carotti, J. C. De Martin, R. Merletti, and D. Farina, “Compression of Multidimensional Biomedical Signals With Spatial and Temporal Codebook-Excited Linear Prediction,” *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, pp. 2604–2610, 2009.

[69] Y. Wongsawatt, S. Oraintarat, and T. Tanakat, “Lossless Multi-channel EEG Compression,” *IEEE Int. Symp. circuits Syst.*, vol. 6, pp. 2–5, 2006.

[70] B. Gosselin and M. Sawan, “Adaptive detection of action potentials using ultra low-power CMOS circuits,” in *IEEE Biomedical Circuits and Systems Conference*, 2008, pp. 209–212.

[71] B. Gosselin and M. Sawan, “An Ultra Low-Power CMOS Automatic Action Potential Detector.,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 4, pp. 346–53, Aug. 2009.

[72] B. Gosselin, A. E. Ayoub, J. Roy, M. Sawan, F. Lepore, and A. Chaudhuri, “A Mixed-Signal Multichip Neural Recording Interface With Bandwidth Reduction,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 3, no. 3, pp. 129–141, 2009.

[73] S. Mitra, J. Putzeys, F. Battaglia, C. M. Lopez, M. Welkenhuysen, C. Pennartz, C. van Hoof, and R. F. Yazicioglu, “24-Channel Dual-Band Wireless Neural Recorder With Activity-Dependent Power Consumption,” in *IEEE International Solid-State Circuits Conference*, 2013, pp. 292–293.

[74] K. U. L. B. Ohri and M. J. Callahan, “Integrated PCM Codec,” no. 1, pp. 38–46, 1979.

[75] Z. Li, T. S. Fiez, and A. A. D. Abstract, “A 14 Bit Continuous-Time Delta-Sigma A/D Modulator With 2.5 MHz Signal Bandwidth,” vol. 42, no. 9, pp. 1873–1883, 2007.

[76] M. J. M. Pelgrom, a. C. J. Duinmaijer, and a. P. G. Welbers, “Matching properties of MOS transistors,” *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[77] S. Mitra, J. Putzeys, F. Battaglia, C. M. Lopez, M. Welkenhuysen, C. Pennartz, C. Van Hoof, and R. F. Yazicioglu, “24-Channel Dual-Band Wireless Neural Recorder With Activity-Dependent Power Consumption,” *Analog Integr. Circuits Signal Process.*, vol. 83, pp. 317–329, 2015.

[78] T. Cameron, G. E. Loeb, R. a. Peck, J. H. Schulman, P. Strojnik, and P. R. Troyk, “Micromodular Implants to Provide Electrical Stimulation of Paralyzed Muscles and Limbs,” *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 781–790, 1997.

[79] K. D. Wise, D. J. Anderson, J. F. Hetke, D. R. Kipke, and K. Najafi, "Wireless Implantable Microsystems: High-Density Electronic Interfaces to the Nervous System," *Proc. IEEE*, vol. 92, no. 1, pp. 76–97, Jan. 2004.

[80] R. Sarpeshkar, C. Salthouse, J. J. Sit, M. W. Baker, S. M. Zhak, T. K. T. Lu, L. Turicchia, and S. Balster, "An Ultra-Low-Power Programmable Analog Bionic Ear Processor," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 711–727, 2005.

[81] J. Georgiou and C. Toumazou, "A 126-μW Cochlear Chip for a Totally Implantable System," *IEEE J. Solid-State Circuits*, vol. 40, no. 2, pp. 430–443, 2005.

[82] L. S. Y. Wong, S. Hossain, A. Ta, J. Edvinsson, D. H. Rivas, and H. Nääs, "A Very Low-Power CMOS Mixed-Signal IC for Implantable Pacemaker Applications," *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2446–2456, 2004.

[83] S. O'Driscoll, T. Meng, K. Shenoy, and C. Kemere, "Neurons to Silicon: Implantable Prosthesis Processor," *IEEE Int. Solid -State Circuits Conf.*, pp. 552–553, 2006.

[84] B. Gosselin, "Recent advances in neural recording microsystems.," *Sensors (Basel).*, vol. 11, no. 5, pp. 4572–97, Jan. 2011.

[85] R. R. Harrison, "The Design of Integrated Circuits to Observe Brain Activity," *Proc. IEEE*, vol. 96, no. 7, pp. 1203–1216, Jul. 2008.

[86] T. Denison, K. Consoer, W. Santa, G. Molnar, and K. Mieser, "A 2 μW, 95nV/√Hz, chopper-stabilized instrumentation amplifier for chronic measurement of bio-potentials," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 1–6, May 2007.

[87] M. Alioto, E. Consoli, and J. M. Rabaey, "' EChO ' Reconfigurable Power Management Unit for Energy Reduction in Sleep-Active Transitions," *IEEE J. Solid-State Circuits*, vol. 48, no. 8, pp. 1921–1932, 2013.

[88] O. Al-Terkawi Hasib, M. Sawan, and Y. Savaria, "A Low-Power Asynchronous Step-Down DC-DC Converter for Implantable Devices," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 3, pp. 292–301, 2011.

[89] M. Yip, J. L. Bohorquez, and A. P. Chandrakasan, "A 0.6V 2.9 μW Mixed-Signal Front-End for ECG Monitoring," in *IEEE Symposium on VLSI Circuits*, 2012, pp. 66–67.

[90] A. C. Paglinawan, W. Y. Chung, Y. H. Wang, S. C. Cheng, W. H. Lee, A. Albason, C. C. Chuang, and S. L. Lou, "A System Design with Mode-Switching Power Management for Closed-Loop Implantable Glucose Biosensor," *J. Med. Biol. Eng.*, vol. 31, no. 5, pp. 321–330, 2011.

[91] E. Matei, E. Lee, J. Gord, P. Nercessian, P. Hess, H. Stover, T. Li, and J. Wolfe, "A Biomedical Implantable FES Battery-Powered Micro-Stimulator," in *IEEE Custom Integrated Circuits Conference*, 2008, no. Cicc, pp. 317–324.

[92] S. Bang, A. Wang, B. Giridhar, and D. Blaauw, "A Fully Integrated Successive-Approximation Switched-Capacitor DC-DC Converter with 31mV Output Voltage Resolution," in *IEEE International Solid-State Circuits Conference*, 2012, no. June 2009, pp. 6–7.

[93] T. Starner and W. Electronics, “Capacitive Power-Management Circuit for Micropower Thermoelectric Generators with a 2.1uW Controller,” in *IEEE Solid-State Ciircuits Conference*, 2011, pp. 300–302.

[94] M. Wieckowski, G. K. Chen, M. Seok, D. Blaauw, D. Sylvester, and A. Arbor, “A Hybrid DC-DC Converter for Sub-Microwatt Sub-1V Implantable Applications,” in *IEEE Symposium on VLSI Circuits*, 2009, vol. 91, pp. 166–167.

[95] T. Van Breussegem and M. Steyaert, “A fully integrated gearbox capacitive DC/DC-converter in 90nm CMOS: Optimization, control and measurements,” in *IEEE Workshop on Control and Modeling for Power Electronics*, 2010, pp. 1–5.

[96] M. Wens and M. Steyaert, “Basic DC-DC Converter Theory,” in *Design and Implementation of Fully-Integrated Inductive DC-DC Converters in Standard CMOS*, Dordrecht: Springer Netherlands, 2011.

[97] J. Wibben and R. Harjani, “A High-Efficiency DC–DC Converter Using 2 nH Integrated Inductors,” *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 844–854, Apr. 2008.

[98] P. Hazucha, G. Schrom, J. Hahn, B. A. Bloechel, P. Hack, G. E. Dermer, S. Narendra, and D. Gardner, “A 233-MHz 80%– 87% Efficient Four-Phase DC–DC Converter Utilizing Air-Core Inductors on Package,” *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 838–845, 2005.

[99] S. Bandyopadhyay, Y. K. Ramadass, and A. P. Chandrakasan, “20μA to 100mA DC-DC Converter with 2.8 to 4.2V Battery Supply for Portable Applications in 45nm CMOS,” in *IEEE Solid-State Ciircuits Conference*, 2011, pp. 2010–2012.

[100] T. Huang, C. Hsieh, Y. Yang, Y. Lee, Y. Kang, K. Chen, C. Huang, Y. Lin, and M. Lee, “A Battery-Free 217 nW Static Control Power Buck Converter for Wireless RF Energy Harvesting With -Calibrated Dynamic On / Off Time and Adaptive Phase Lead Control,” *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 852–862, 2012.

[101] Y. K. Ramadass and A. P. Chandrakasan, “Minimum Energy Tracking Loop With Embedded DC–DC Converter Enabling Ultra-Low-Voltage Operation Down to 250 mV in 65 nm CMOS,” *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 256–265, Jan. 2008.

[102] J. Xiao, A. V Peterchev, J. Zhang, and S. R. Sanders, “A 4μ-A Quiescent-Current Dual-Mode Digitally Controlled Buck Converter IC for Cellular Phone Applications,” *IEEE J. Solid-State Circuits*, vol. 39, no. 12, pp. 2342–2348, 2004.

[103] T. Y. Man, P. K. T. Mok, and M. Chan, “An Auto-Selectable-Frequency Pulse-Width Modulator for Buck Converters with Improved Light- Load Efficiency,” in *IEEE International Solid State Circuits Conference*, 2008, no. 1, pp. 24–26.

[104] S. S. Kudva and R. Harjani, “Fully-Integrated On-Chip DC-DC Converter With a 450X Output Range,” *IEEE J. Solid-State Circuits*, vol. 46, no. 8, pp. 1940–1951, Aug. 2011.

[105] S. Kim and G. Rincon-Mora, “Achieving High Efficiency Under Micro-Watt Loads with Switching Buck DCDC Converters,” *IEEE J. Low Power Electron.*, no. 404, 2009.

[106] S. Chang, K. Alashmouny, M. Mccormick, Y. Chen, and E. Yoon, "BioBolt : A Minimally-Invasive Neural Interface for Wireless Epidural Recording by Intra-Skin Communication," in *IEEE Symposium on VLSI Circuits*, 2012, vol. 2.

[107] H.-G. Rhew, J. Jeong, J. a. Fredenburg, S. Dodani, P. Patil, and M. P. Flynn, "A wirelessly powered log-based closed-loop deep brain stimulation SoC with two-way wireless telemetry for treatment of neurological disorders," in *IEEE Symposium on VLSI Circuits*, 2012, pp. 70–71.

[108] J. Lee, D. Kipke, and M. Flynn, "A 64 Channel programmable closed-loop deep brain stimulator with 8 channel neural amplifier and logarithmic ADC," *IEEE J. Solid State Circuits*, vol. 45, no. 9, pp. 76–77, Jun. 2008.

[109] C. Tao and A. Fayed, "Output Spectrum Analysis of Buck Converters in DCM with PFM Control," in *IEEE International Symposium on Circuits and Systems*, 2012, pp. 2267–2270.

[110] C. Tao and A. A. Fayed, "A Low-Noise PFM-Controlled Buck Converter for Low-Power Applications," *IEEE Trans. Circuits Syst. I*, vol. 59, no. 12, pp. 3071–3080, 2012.

[111] S. Musunuri and P. L. Chapman, "Improvement of Light-Load Efficiency Using Width-Switching Scheme for CMOS Transistors," *IEEE Power Electron. Lett.*, vol. 3, no. 3, pp. 105–110, 2005.

[112] M. D. Mulligan, B. Broach, and T. H. Lee, "A 3MHz Low-Voltage Buck Converter with Improved Light Load Efficiency," in *IEEE Solid-State Ciircuits Conference*, 2007, pp. 2–4.

[113] Y. Chen, F. C. Lee, L. Amoroso, and H. P. Wu, "A Resonant MOSFET Gate Driver with Efficient Energy Recovery," *IEEE Trans. Power Electron.*, vol. 19, no. 2, pp. 470–477, 2004.

[114] D. Ma, W. H. Ki, and C. Y. Tsui, "An Integrated One-Cycle Control Buck Converter with Adaptive Output and Dual Loops for Output Error Correction," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 140–149, 2004.

[115] S.-Y. Park, J. Cho, K. Lee, and E. Yoon, "PWM Buck Converter with >80% PCE in 45uA-to-4mA Loads Using Analog-Digital Hybrid Control for Biomedical Systems," in *IEEE International Solid-State Circuits Conference*, 2015, pp. 218–220.

[116] S. Park, J. Cho, K. Lee, and E. Yoon, "A PWM Buck Converter With Load-Adaptive Power Transistor Scaling Scheme Using Analog-Digital Hybrid Control for High Energy Efficiency in Implantable Biomedical Systems," *Trans. Biomed. Circuits Syst.*, vol. 9, no. 6, pp. 885–895, 2015.

[117] K. Wang, N. Rahman, Z. Lukić, and A. Prodić, "All-Digital DPWM/DPFM Controller for Low-Power DC-DC Converters," in *IEEE Applied Power Electronics Conference and Exposition*, 2006, pp. 719–723.

[118] C. Lu, H. Hsieh, and L. Lu, "A 0.6 V Low-Power Wide-Range Delay-Locked Loop in 0.18 m CMOS," *IEEE Microw. Wirel. Components Lett.*, vol. 19, no. 10, pp. 662–664, 2009.

[119] G. Kim, M. Kim, B. Chang, and W. Kim, “A Low-Voltage, Low-Power CMOS Delay Element,” *IEEE J. Solid-State Circuits*, vol. 31, no. 7, pp. 966–971, Jul. 1996.

[120] A. V. Peterchev and S. R. Sanders, “Quantization Resolution and Limit Cycling in Digitally Controlled PWM Converters,” in *IEEE Power Electronics Specialists Conference*, 2001, vol. 2, no. 1, pp. 465–471.

[121] H. Li, A. Feizmohammadi, and A. Prodic, “Limit-Cycle Based Auto-Tuning System for Digitally Controlled Low-Power SMPS,” in *Twenty-First Annual IEEE Applied Power Electronics Conference and Exposition*, 2006, pp. 1143–1147.

[122] J. Markus, J. Silva, and G. C. Temes, “Theory and Applications of Incremental ΔΣ Converters,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 51, no. 4, pp. 678–690, 2004.

[123] Z. Zhao and A. Prodic, “Limit-Cycle Oscillations Based Auto-Tuning System for Digitally Controlled DC –DC Power Supplies,” *IEEE Trans. Power Electron.*, vol. 22, no. 6, pp. 2211–2222, 2007.