

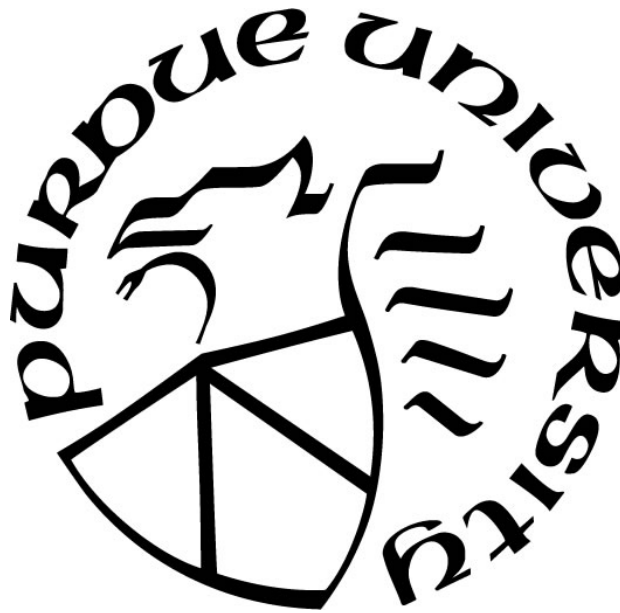
DIGITAL SIGNAL PROCESSING ARCHITECTURE DESIGN FOR CLOSED-LOOP ELECTRICAL NERVE STIMULATION SYSTEMS

by
Jui-Wei Tsai

A Dissertation

*Submitted to the Faculty of Purdue University
In Partial Fulfillment of the Requirements for the degree of*

Doctor of Philosophy



School of Electrical and Computer Engineering
West Lafayette, Indiana
December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Pedro P. Irazoqui, Chair

Department of Electrical and Computer Engineering

Dr. Kaushik Roy

Department of Electrical and Computer Engineering

Dr. Anand Raghunathan

Department of Electrical and Computer Engineering

Dr. Vijay Raghunathan

Department of Electrical and Computer Engineering

Dr. Matthew P. Ward

Department of Biomedical Engineering

Approved by:

Dr. Dimitrios Peroulis

Head of the Graduate Program

To my beloved wife, daughter and parents.

ACKNOWLEDGMENTS

The journey toward PhD degree is never easy, and I wouldn't have been able to complete it without tremendous help from many people.

First, I would like to thank my major advisor, Dr. Pedro Irazoqui, who appreciates my skills and lets me work in his lab throughout these years. Thank you for supporting my research and being patient with me. I'd also like to thank Dr. Kaushik Roy, Dr. Anand Raghunathan, Dr. Vijay Raghunathan, and Dr. Matthew Ward, for being interested in my work and willing to be my committee members. I'm also grateful for Dr. Michael Capano, the graduate coordinator of ECE Dept., who has been a patient mentor in my difficult time.

I would like to thank my lab colleagues, Dr. Hansraj Bhamra, Dr. Young-Joon Kim, Dr. Yu-Wen Huang, Dr. Oren Gall, Dr. John Lynch, Dr. Steven Lee, Dr. Rebecca Bercich, Dr. Henry Mei, Dr. Daniel Pederson, Dr. Quan Yuan, Dr. Kurt Qing, Dr. Muhammad Arafat, Dr. Jesse Somann, Dr. Jack Williams, Brandon Coventry, Henry Zhang, Grant (Zhi) Wang, Kyle Thackston, Chris Quinkert, Jay Shah, Vivek Ganesh, Gang Seo, Kaitlyn Neihouser, Ethan Biggs, Ryan Budde, and Gabriel Albors, the managing director of the Center for Implantable Devices. Thank you for many valuable discussions and helps in my experiments. I'm especially thankful for Dr. Hansraj Bhamra and Brandon Coventry who were always willing to talk with me whenever I felt hopeless about my PhD program.

I owe my thanks to Dr. Mark Johnson and Neal (Nien-Shiang) Chang, the manager of Taiwan Semiconductor Research Institute, Hsinchu City, Taiwan, who provided me with great technical assistances in ASIC design, especially its CAD tools. I also want to thank the teaching assistants of the ECE Senior Design course, Nathan Conrad, Shelly (Hengying) Shan, Steve Rausch and Sutton Hathorn, for giving me lot of valuable suggestions on the design of wireless devices.

I would like to thank the pastors of St. Thomas Aquinas, West Lafayette, IN, Fr. Patrick, Fr. Timothy, and Fr. Cassen. I want to give thanks to many brothers and sisters in Christ, Thomas Combiths, Susan Combiths, Maria (Eunjin) Cheon, En-Pei Han, Yu-Chen Lin, Jesyin Lai and

Tsung-Tai Yeh. Thank you for your kind-heartedness, generosity, and, most importantly, earnest praying for my PhD career. I also want to thank Master Kyu-Young Chai, the grand master of Chai Taekwondo, Lafayette, IN, and Coach Carlos Soto, the head coach of Impact Zone Training Center, West Lafayette, IN. You helped me build my physical and mental strength and showed your care about my life in the USA.

I'm also grateful for Dr. Ching-Chang Chien and Dr. Yu-Chuan Su, the professors of ESS Dept., NTHU, Taiwan, and Dr. Song-Nien Tang, the assistant professor of CS Dept., CYCU, Taiwan. I wouldn't have come to Purdue without you giving me the opportunity to work with you back in Taiwan. Thank you for always caring about my career at Purdue.

I would like to express my sincerest gratitude to my wife, Maria Teresa Kuo, who always stands with me in every despairing moment, my daughter, Katie Tsai, the little angel who is always a cheer for me, my parents, Mr. Jung-Chieh Tsai and Mrs. Yue-Chun Kuo, who bring me up and have always been my role model throughout these 35 years, my parents-in-laws, Mr. Jose (Yung-Chung) Kuo and Mrs. Bernadette (Hui-Fen) Hu, who believe in me and let Teresa stay with me in these years, my sister, Joanna (Jui-Yu) Tsai, and my brother-in-law, Jose (Chen-Wei) Kuo. It is your love that keeps encouraging me to stick with this program no matter what happened. I would also like to thank Mr. Andy (Te-Hsiu) Huang, my elementary school teacher, for your unceasingly care since my childhood. *"Don't forget why you started"*. It is what you told me in 2018 when I lost my assistantship, got stuck in my research, and even thought about going home, which still reminds and motivates me today. This thesis is especially dedicated to Mr. Jose (Yung-Chung) Kuo and Mr. Andy (Te-Hsiu) Huang, sadly, who left us on 2020. Hope you'll be proud to see me complete my PhD program, and hope you meet my beloved uncle Mr. Jung-Chin Tsai and grandparents in heaven.

Last but not the least, I would like to give this glory to the Lord. It's not by my strength, but by His grace, that I walked through each mountain and valley and finished this program. I would also like to thank whoever helped me and encouraged me in this journey yet I forgot to mention you in this acknowledgement. I wouldn't have made it without each of you, thank you.

"Trust in the LORD with all your heart, on your own intelligence do not rely; In all your ways be mindful of him, and he will make straight your paths." - Proverb 3:5-6

TABLE OF CONTENTS

LIST OF TABLES	9
LIST OF FIGURES	10
Abbreviations	14
ABSTRACT.....	16
1. INTRODUCTION	18
1.1 Background.....	18
1.2 Autonomous Nerve Control	20
1.3 Neural Response Telemetry	21
1.4 Electrically-Evoked Compound Action Potential	24
1.5 Real-Time Digital Signal Processing for Closed-Loop Neurostimulation.....	28
1.6 Motivation	31
1.7 Outline of Thesis	33
2. A DSP ARCHITECTURE FOR REAL-TIME EVOKED COMPOUND ACTION POTENTIAL RECOVERY IN NEURAL RESPONSE TELEMETRY SYSTEM.....	35
2.1 Introduction	35
2.2 Bidirectional-Filtered Coherent Averaging.....	40
2.3 Architecture Design.....	46
2.3.1 System Overview	46
2.3.2 Stimulation Controller.....	47
2.3.3 BFCA Core.....	49
2.3.4 Exponentially-Weighted Moving Average	54
2.3.5 Configurable Folded IIR filter Design	56
2.3.6 Output Buffer	60
2.4 Experiment Results.....	60
2.4.1 Hardware Implementations	60
2.4.2 <i>In-Vivo</i> Test Results	62
2.4.3 Efficacy Analysis	66
2.5 Conclusion of This Chapter.....	71

3. FREE: FIBER-RESPONSE EXTRACTION ENGINE ON A CUSTOM-MADE WEARABLE DEVICE FOR AUTONOMOUS NERVE ACTIVATION CONTROL.....	73
3.1 Introduction	73
3.2 System Overview.....	77
3.3 Architecture Design.....	80
3.3.1 BFCA Core in FREE.....	80
3.3.2 Peak Detector	81
3.3.3 Fiber Response Classifier.....	85
3.3.4 Output Buffer in FREE	87
3.4 PCB Prototype of Wearable Device	88
3.5 Experiment Results.....	92
3.5.1 Experiment Setup	92
3.5.2 Precision Comparison	94
3.5.3 In-Vivo Test Results.....	100
3.5.4 ASIC Implementation	102
3.6 Conclusion of This Chapter.....	104
4. CONCLUSION AND FUTURE WORK	105
4.1 Conclusion.....	105
4.2 Future Work.....	106
4.2.1 Half-Precision Floating-Point Computation.....	106
4.2.2 Data Compression of ECAP.....	108
4.2.3 Implantable Wireless Device	111
REFERENCES	116
VITA.....	126

LIST OF TABLES

Table 2.1 Comparison with other filtering techniques.....	71
Table 3.1 Peak Detection Algorithm	83
Table 3.2 List of the components used in the PMU PCB	91
Table 3.3 Performance comparison between hardware and software processing	103
Table 4.1 Comparison of the low-power COTS Bluetooth modules [151]	115

LIST OF FIGURES

Fig. 1.1 Block diagram of the autonomous nerve control (ANC) system and its applications. [35]	20
Fig. 1.2 A typical modern cochlear implant system that provides electrical stimuli to auditory nerve [42].	22
Fig. 1.3 Neural Response Telemetry (NRT) system in Nucleus CI24M cochlear implant [43].	23
Fig. 1.4 Cross-section of a nerve and group of nerve fibers (axons) [48].	24
Fig. 1.5 Classification of nerve fiber types [49].	25
Fig. 1.6 Classification of nerve fiber responses on ECAP waveforms plotted (A) against time axis and (B) as a function of conduction velocity. The ECAP responses are obtained from the left cervical vagus nerve of rat (<i>Conduction distance</i> = 8.0 ± 0.5 mm) [35].	25
Fig. 1.7 The three most commonly used methods for stimulus artifact reduction in ECAP recording: (a) alternating polarity, (b) subthreshold template subtraction, and (c) 2-pulse forward masking paradigm [50].	26
Fig. 1.8 Principle of the coherent averaging (CA) technique [52].	27
Fig. 1.9 High-level block diagram of a typical wireless wearable (or implantable) device for closed-loop neurostimulation.	29
Fig. 1.10 Illustration of a real-time digital signal processing (DSP) engine on a wireless device.	30
Fig. 2.1 Illustration of Conventional bidirectional neural response telemetry (NRT) systems. .	36
Fig. 2.2 (a) The NRT systems with a digital signal processor. (b) Details of stimulation and recording analog front-ends (AFE) and digital signal processor on the implant.	37
Fig. 2.3 Principle of the proposed bidirectional-filtered coherent averaging (BFCA) method combined with the alternating-polarity (AP) stimulation method for stimulus artifact rejection and distortion-free denoising of ECAP.	43
Fig. 2.4 Block diagram of the proposed DSP architecture for real-time ECAP recovery.	46
Fig. 2.5 Generation of the alternating-polarity stimulus pulse and time-locked windowing control in stimulation controller.	47
Fig. 2.6 Schematic of the BFCA core.	48
Fig. 2.7 Schematic of the DC calculator.	49
Fig. 2.8 (a) Time reversal via LIFO register and (b) its implementation with a two-port SRAM and two binary up/down counters.	51
Fig. 2.9 Timing diagram of the BFCA core under continuous neural data input.	53

Fig. 2.10 Schematics of the (a) exponentially-weighted moving-average (EWMA) calculator and (b) lead-one detector for estimation of weighting factor in EWMA.	55
Fig. 2.11 A conventional configurable 4-stage IIR filter.....	57
Fig. 2.12 (a) The proposed configurable, folded 4-stage IIR filter with shared multiply-add (MA) unit, (b) its timing diagram, and (c) shared MA operation of a biquad filter stage.....	59
Fig. 2.13 Schematic of the output buffer.	60
Fig. 2.14 Power consumption of the DSP architecture in 180-nm CMOS process.	61
Fig. 2.15 Circuit schematic of the stimulation and recording AFE.	62
Fig. 2.16 Setup of <i>in-vivo</i> electrical nerve stimulation for verification of proposed DSP architecture.....	63
Fig. 2.17 FPGA measurement results from stimulation trials with stimulus parameters $PW = 0.2$ ms, $PRF = 20$ Hz and $t_{train} = 1$ s: (a) windowed raw data and (b) computed ECAP responses of two stimulation trials with stimulus current amplitude of 0.2 mA and 0.4 mA, respectively.	65
Fig. 2.18 (a) FPGA measurement results of linear-phase filtered ECAP responses collected from stimulation trials with stimulus amplitude varying from 0 to 0.5 mA. (b) Amplitude growth function of nerve fiber responses designated by peaks on ECAP responses.....	66
Fig. 2.19 Root-mean-square (rms) value of stimulus artifacts in (a) raw data and (b) ECAP responses measured from FPGA. The rms value of noise floor in (b) is obtained from a 10-ms segment on each ECAP waveform containing no stimulus artifacts or nerve fiber responses.....	68
Fig. 2.20 Signal-to-noise ratio (SNR) of linear-phase filtered versus unfiltered ECAP responses measured from FPGA.	69
Fig. 2.21 Waveform distortion caused by forward filtering (ForFilt), coherent averaging (CA) and linear-phase filtering via BFCA: (a) demonstration and (b) a quantitative comparison using normalized mean-square error (NMSE) between filtered and original noise-free ECAP waveforms.....	70
Fig. 3.1 Top-level block diagram of the wireless wearable device in a closed-loop electrical nerve system and the proposed fiber-response extraction engine (FREE), a dedicated DSP engine for autonomous nerve activation control.	77
Fig. 3.2 (a) Block diagram of the proposed fiber-response extraction engine (FREE) and (b) the flowchart of its operation.	78
Fig. 3.3 (a) Architecture of the BFCA core in FREE and (b) the schematic of the maximal absolute value detector.....	80
Fig. 3.4 Illustration of the peak-detection principle.....	81
Fig. 3.5 (a) Architecture of the peak detector. Data path in the peak detector for (b) calculation of standard deviation of the noise floor during the training mode, updating of (c) percentage threshold and (d) noise threshold, and (e) peak detection. The multiplication operations are properly scheduled such that only one multiplication is performed for each clock cycle.....	84

Fig. 3.6 (a) Principle of fiber-response (FR) classification and (b) its architecture.	86
Fig. 3.7 Schematic of the output buffer in FREE.	87
Fig. 3.8 Schematic of the stimulation and recording analog front-end (AFE) in the wearable device.	88
Fig. 3.9 PCB prototype of the wireless signal processing (WSP) platform with major components annotated.....	89
Fig. 3.10 Block diagram of the power-management unit (PMU) for the power supply of the AFE and WSP platform.....	90
Fig. 3.11 The assembled PCB prototype of the wearable device.	91
Fig. 3.12 Illustration of experiment setup for performance comparison between hardware (HW) and software (SW) processing.	92
Fig. 3.13 Plots of the original neural signal recorded in a single stimulation trial that contains ECAP responses and the neural signal superimposed with 60-Hz noise and 2-Hz baseline drift as the input data for HW and SW comparison. The input data should be shifted by $+V_{cm}$ to meet the dynamic range of the ADC on AFE.....	93
Fig. 3.14 (a) An example of recovered ECAP via HW- and SW-processing versus the original ECAP waveform. (b) All HW- and SW-recovered ECAP responses from the entire input data set (66 trials).....	95
Fig. 3.15 Mean latency and amplitude of extracted (a) positive and (b) negative fiber responses. Both the positive and negative fiber responses extracted from HW have less amplitude variation than those from SW owing to the more effective removal of periodic noises on HW.	96
Fig. 3.16 Slope-activation data of extracted fiber responses from (a) HW and (b) SW and the predicted slope-activation relationship (i.e., rheobase currents I_{Rh} as a function of the percent activation level λ). The classification precision of fiber responses from HW and SW is reflected by the coefficient of determination R^2 which represents the goodness-of-fit of predicted I_{Rh}	99
Fig. 3.17 <i>In-vivo</i> test results (PW = stimulus pulse width = 0.2 ms; PRF = pulse repetition frequency = 20 Hz; t_{train} = stimulus train duration = 1 s) of HW- and SW- recovered ECAP waveforms and extracted fiber responses.	100
Fig. 3.18 (a) The amplitude growth function (AGF) of extracted fiber responses and (c) the mean latency of positive and negative fiber responses extracted from HW and SW in <i>in-vivo</i> tests.	101
Fig. 3.19 The ASIC implementation of proposed FREE in 180-nm CMOS technology: (a) die photo and breakdown of its (b) power and (c) area consumption.....	102
Fig. 4.1 Block diagram of a half-precision bit-serial floating-point adder [132].	107
Fig. 4.2 Block diagram of a half-precision floating-point multiplier, where (X_e, Y_e) and (X_f, Y_f) are exponents and fractions of two input signals, respectively, "load" and "reset" are control signals for each part, and Z_e and Z_f are the exponent and fraction of multiplier output [133].	107

Fig. 4.3 Illustration of DWT algorithm for data compression with 4 levels of decomposition [63].	108
Fig. 4.4 Data compression of an ECAP waveform using 2-level Haar wavelet DWT: (a) derivation of the threshold (THR_{WC}) based on the mean of absolute value of wavelet coefficients (μABS_{WC}), (b) original versus thresholded wavelet coefficients (WC_{THR}) of an ECAP waveform, and (c) original versus reconstructed ECAP waveform.	110
Fig. 4.5 Block diagram of the PMU with a combination of both WPT and a rechargeable battery as its power supply.	112
Fig. 4.6 (a) Schematic of the bq51003 (<i>Texas Instruments</i>) wireless power supply receiver [144]. (b) A receiving coil for the bq51003 (<i>Würth Elektronik Group</i>) with total size of 15 mm (diameter) \times 0.6 mm (height). (c) Top view of the bq500212AEVM-550 wireless power transmitter evaluation module (<i>Texas Instruments</i>) as the primary wireless power supply [145]. (d) Schematic of the bq21040 (<i>Texas Instruments</i>) lithium-ion battery charger [146].	113
Fig. 4.7 Top view of (a) the MIKROE-2471 BLE 3 Click (<i>MikroElektronika</i>) [148] and (b) the B204 USB dongle [150], both built with the NINA-B112 Bluetooth 4.2 module.	114

ABBREVIATIONS

ADC	Analog-to-Digital Converter
AFE	Analog Front End
AGF	Amplitude Growth Function
ANC	Autonomous Nerve Control
AP	Alternating Polarity
ASIC	Application-Specific Integrated Circuit
BFCA	Bidirectional-Filtered Coherent Averaging
BLE	Bluetooth Low-Energy
BS	Base Station
CA	Coherent Averaging
CMOS	Complementary Metal Oxide Semiconductor
CMRR	Common Mode Rejection Ratio
COTS	Commercial Off-the-Shelf
CP	Charge Pump
DAC	Digital-to-Analog Converter
DBS	Deep Brain Stimulation
DSP	Digital Signal Processing
DTFT	Discrete-Time Fourier Transform
DWT	Discrete Wavelet Transform
ECAP	Electrically-Evoked Compound Action Potential
ECG	Electroencephalogram
ECoG	Electrocorticogram
EEG	Electroencephalogram
ENG	Electroneurogram
ENS	Electrical Nerve Stimulation
EWMA	Exponentially-Weighted Moving Averaging
FDA	Food and Drug Administration
FIFO	First-In-First-Out
FIR	Finite Impulse Response

FPGA	Field-Programmable Gate Array
FREE	Fiber Response Extraction Engine
GUI	Graphic Users Interface
IC	Integrated Circuit
IIR	Infinite Impulse Response
LFP	Local Field Potentials
LIFO	Last-In-First-Out
LP	Linear Phase
MCU	Microcontroller Unit
NA	Neural Amplifier
NAP	Nerve Activation Profile
NMSE	Normalized Mean-Square Error
NRT	Neural Response Telemetry
PC	Personal Computer
PCB	Printed Circuit Board
PMU	Power Management Unit
RF	Radio Frequency
RLE	Run-Length Encoding
RX	Receiver
SAR	Stimulus Artifact Rejection
SoC	System-on-Chip
SQNR	Signal-to-Quantization-Noise Ratio
SNR	Signal-to-Noise Ratio
SPI	Serial Peripheral Interface
SRAM	Static Random-Access Memory
TX	Transmitter
UART	Universal Asynchronous Receiver-Transmitter
USB	Universal Serial Bus
VLSI	Very-Large Scale Integration
VNS	Vagus Nerve Stimulation
WPT	Wireless Power Transfer

ABSTRACT

Author: Tsai, Jui-Wei. PhD

Institution: Purdue University

Degree Received: December 2020

Title: Digital Signal Processing Architecture Design for Closed-Loop Electrical Nerve Stimulation Systems.

Committee Chair: Pedro Irazoqui

Electrical nerve stimulation (ENS) is an emerging therapy for many neurological disorders. Compared with conventional one-way stimulations, closed-loop ENS approaches increase the stimulation efficacy and minimize patient's discomfort by constantly adjusting the stimulation parameters according to the feedback biomarkers from patients. Wireless neurostimulation devices capable of both stimulation and telemetry of recorded physiological signals are welcome for closed-loop ENS systems to improve the quality and reduce the costs of treatments, and real-time digital signal processing (DSP) engines processing and extracting features from recorded signals can reduce the data transmission rate and the resulting power consumption of wireless devices. Electrically-evoked compound action potential (ECAP) is an objective measure of nerve activity and has been used as the feedback biomarker in closed-loop ENS systems including neural response telemetry (NRT) systems and a newly proposed autonomous nerve control (ANC) platform. It's desirable to design a DSP engine for real-time processing of ECAP in closed-loop ENS systems.

This thesis focuses on developing the DSP architecture for real-time processing of ECAP, including stimulus artifact rejection (SAR), denoising, and extraction of nerve fiber responses as biomedical features, and its VLSI implementation for optimal hardware costs. The first part presents the DSP architecture for real-time SAR and denoising of ECAP in NRT systems. A bidirectional-filtered coherent averaging (BFCA) method is proposed, which enables the configurable linear-phase filter to be realized hardware efficiently for distortion-free filtering of ECAPs and can be easily combined with the alternating-polarity (AP) stimulation method for SAR. Design techniques including folded-IIR filter and division-free averaging are incorporated to reduce the computation cost. The second part presents the fiber-response extraction engine (FREE), a dedicated DSP engine for nerve activation control in the ANC platform. FREE

employs the DSP architecture of the BFCA method combined with the AP stimulation, and the architecture of computationally efficient peak detection and classification algorithms for fiber response extraction from ECAP. FREE is mapped onto a custom-made and battery-powered wearable wireless device incorporating a low-power FPGA, a Bluetooth transceiver, a stimulation and recording analog front-end and a power-management unit. In comparison with previous software-based signal processing, FREE not only reduces the data rate of wireless devices but also improves the precision of fiber response classification in noisy environments, which contributes to the construction of high-accuracy nerve activation profile in the ANC platform. An application-specific integrated circuit (ASIC) version of FREE is implemented in 180-nm CMOS technology, with total chip area and core power consumption of 19.98 mm² and 1.95 mW, respectively.

1. INTRODUCTION

1.1 Background

Nervous system is a complex network spreading through the human body that carries message from brain to various parts of body for regulation of physiological functions, including breath, heart rate, sensation, speech, and even stomach movement during digestion [1]. These physiological functions may be modulated externally by stimulating particular branches of central or peripheral nervous systems, which is also known as neuromodulation [2-4]. Ever since the United States Food & Drug Administration (FDA) approved deep brain stimulation (DBS) as a valid treatment of tremor in 1997 [5], neuromodulation becomes an emerging therapeutic for various neuro- logical diseases. Neuromodulation not only provides another option for patients who're resistant to medication, but also possesses the capacity to target and dose a certain nerve and brain area more precisely, making it a popular treatment alternative to pharmaceutical approach. For instance, DBS utilizes implanted microelectrodes in the brain through which electrical stimulus is delivered to targeted brain area and has been employed in the treatments for Parkinson disease, chronic pain, and other neurological disorders including depression [6, 7]. Spinal cord stimulation (SCS) provides therapy for chronic and intractable pains by intervening in transmission of pain signals along the spinal cords with electrical pulses [8-10]. Applications of neuromodulation to other neurological or psychiatric disorders have been demonstrated and are still being investigated today [11, 12].

Electrical nerve stimulation (ENS) is one neuromodulation technique that involves stimulating nerves with electric current in order to modulate propagation of neural signals along nerve. Ever since the first patient-wearable ENS device was patented in the United States in 1974 [13], ENS has been widely used in clinical therapy for acute and chronic pains, and its application in the treatment for other neurological diseases has also received attention in these decades [14-16]. In human's nervous system, vagus nerve is the longest cranial nerve extending from the brain stem to the colon; it controls important sensory and motor functions, including the visceral sensation of lungs, heart, and digestive tract and muscles in the heart and digestive tract for the regulation of heart rate and food digestion, respectively [17, 18]. Vagus nerve stimulation (VNS) is one of

the most renowned ENS that was approved by the FDA for epilepsy treatment in 1997 and approved for major depressive disorder treatment in 2005 [19, 20]. Although the mechanism of VNS still requires more elucidation, studies have shown the efficacy of VNS and its mild side effects [21]. VNS has been approved for seizure reduction in Canada and more than 15 countries in Europe [22].

Most commercial neurostimulation systems today are in an open-loop manner, where devices with pre-programmed electrical stimulus is connected to targeted nerve or brain area via implanted microelectrodes, and stimulus parameters are tuned per week or month according to patient's subjective experience in treatment. For example, the VNS device produced by LivaNova and the responsive neurostimulation device produced by NeuroPace are two commercialized open-loop neurostimulators [23]. As the clinical experiences of open-loop neurostimulation accumulate, its problems become more evident, including low stimulation efficiency (either too much or too little dosing), slow reaction to patient's condition that easily causes patient's discomfort, and side effects associated with the therapy, all of which results from lacking objective measurement of how patients react to applied stimulus. A closed-loop neurostimulation system can improve stimulation efficiency and reduce discomforts and side effects on patients by recording physiological signals from patients and constantly adjusting stimulus strength in response to changes in recorded signals [24-26]. Efforts have been made in developing closed-loop neurostimulation systems and devices for various neurological diseases [27, 28]. In closed-loop VNS for epilepsy, stimulation is triggered at the onset of seizure, which can be detected in real-time based on heart rate change, electroencephalogram (EEG) and electrocardiogram (ECG) signals [29-32]. Closed-loop DBS comprises stimulation, sensing of biomarkers such as local field potentials (LFPs), action potentials, electrocorticogram (ECoG), and EEG, and detection of their features [33]. A review of closed-loop DBS systems and devices can be found in [34]. An FDA-approved closed-loop SCS system, the RestoreSensor system (Medtronic, Minneapolis, MN, USA), is also proposed, which automatically adjusts stimulus parameters according to patient's body position sensed by a 3-axis accelerometer [24]. In brief, a closed-loop neurostimulation system must be able to sense physiological signals effectively from patients, precisely locate the biomarkers in recorded signals, and detect change in biomarkers and adjust stimulus strength in response to the change in real-time.

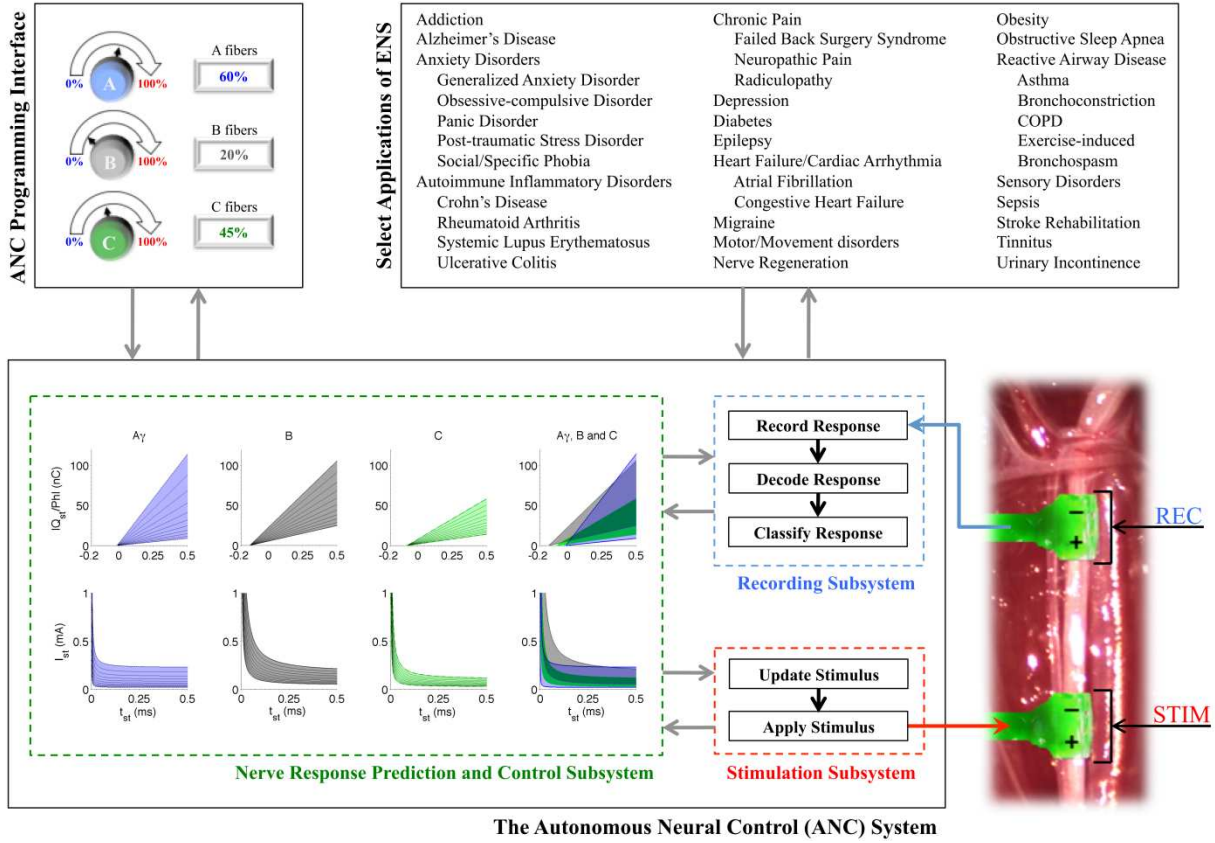


Fig. 1.1 Block diagram of the autonomous nerve control (ANC) system and its applications. [35]

1.2 Autonomous Nerve Control

A nerve comprises mainly bundle of cable-like nerve fibers (also called axon), each of which is a projection of nerve cell (neuron) that transmit electrical signals known as action potentials to different muscles, tissues and organs [36]. According to Gasser [37], nerve fibers can be classified into three types based on their physical features and signal conduction properties - group A (fast, myelinated), group B (slow, myelinated), and group C (slow, unmyelinated). It's believed that the electrical stimulation modulates the activity of nerve fibers and thus sensory and motor functions the nerve fibers map to, which is one explanation for the mechanism of VNS [38]. Based on this theory, if the activation of nerve fibers can be properly controlled, the efficacy of VNS can be greatly improved and the severity of side effects in open-loop VNS can also be minimized. To address this issue, Matthew et al. propose the autonomous nerve control

(ANC) system [35], a responsive closed-loop ENS system that automatically adjusts stimulus strength using measured nerve activation level.

Fig. 1.1 shows the block diagram of ANC system and its applications. In ANC, electrical stimulus is first applied onto a nerve via a stimulation electrode and the electrically-evoked compound action potential (ECAP) on the nerve in response to the stimulus is derived from the neural signals acquired from the recording electrode on the nerve adjacent to the stimulation electrode. ANC then identifies and classifies nerve fiber responses on ECAP waveform in real time. The amplitude responses of targeted nerve fiber together with stimulus parameters are clustered for construction of a patient-specific nerve activation profile (NAP), which predicts how nerve will respond to stimulus with any strength. In closed-loop stimulation, ANC constantly adjusts stimulus parameters according to the derived NAP to control the activation of targeted nerve fiber. ANC is first tested in VNS of rats to demonstrate its capacity to most efficiently control the activation of vagal A, B or C fibers [35] and can be applied to other ENS-based therapeutics for various neurological diseases, e.g. addiction, chronic pain, motor and sensory disorders.

ANC introduces great benefits to both patients and physicians. From patient's side, the period of treatment can be lowered to help patients save their time cost, and quality of treatment is also improved through the minimization of discomfort and side effects. For physicians, ANC provides an objective dosing standard based on the level of nerve activation, and save physicians from the time-consuming process of stimulus parameter tuning. ANC also enable physicians to selectively control the activation of fiber group (A, B or C fiber) and hence the physiological functions that fiber group maps to.

1.3 Neural Response Telemetry

It is estimated that around 466 million people worldwide suffer from some degree of hearing loss, 34 million of which are children [39]. Several causes of hearing loss includes genetics, aging, exposure to noise, infections, birth complications and traumas to ear. Hearing loss results in not only inconvenience to patients but also physical, psychological and social problems (e.g. headache, stress, low self-esteem, isolation from community, etc.). Ever since 1957, when

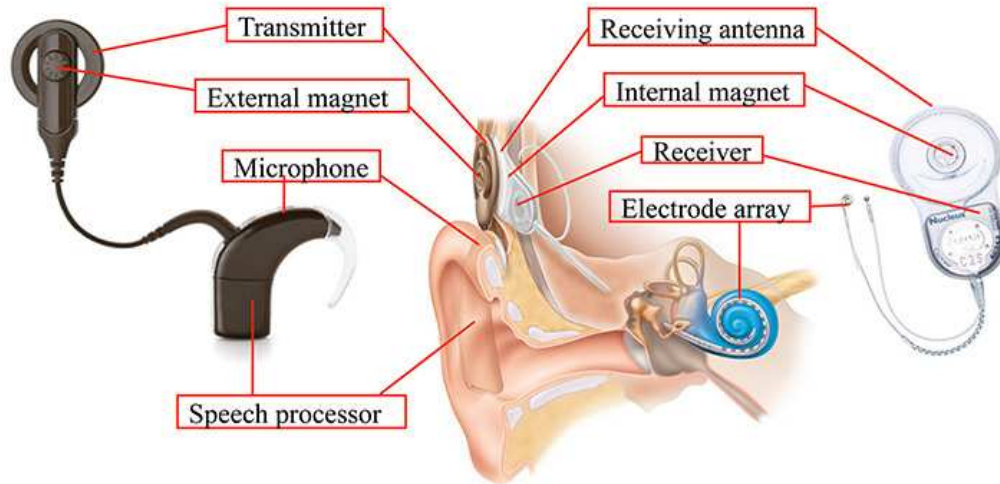


Fig. 1.2 A typical modern cochlear implant system that provides electrical stimuli to auditory nerve [42].

French physician Djourno and his colleagues regained the hearing of two totally deafened patients using electrical stimulation, cochlear implant has been a popular management for hearing loss that provides partial hearing to deafened patients. Today, cochlear implant is one of the most successful neural prosthesis with more than 120,000 people implanted worldwide [40, 41]. The goal of cochlear implant is to replace the normal acoustic hearing process with electrical signals that directly stimulate auditory nerve to restore functional hearing. Fig. 1.2 illustrates a typical modern cochlear implant system [42]. The sound is first sensed with a microphone, processed and encoded into digital signals by speech processor, and transmitted to the implant with the radio frequency (RF) transmitter. On the receiver that is placed under the skin behind the ear, digital signals are received with the antenna, decoded and converted into electric current. A stimulator on the receiver deliver the electric current to auditory nerve via the electrode array implanted in the cochlea, which's then interpreted as sound.

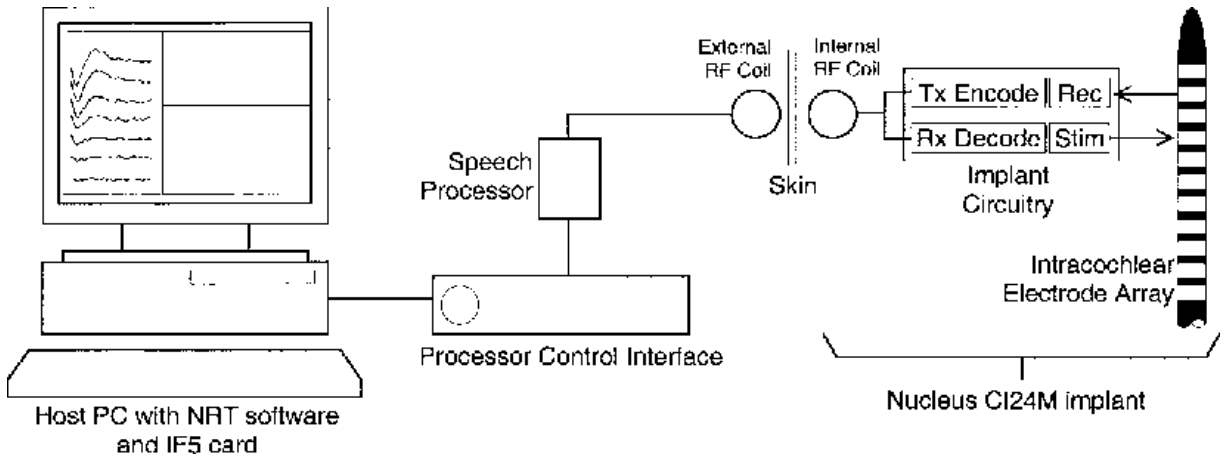


Fig. 1.3 Neural Response Telemetry (NRT) system in Nucleus CI24M cochlear implant [43].

Due to the difference in the structure of auditory nerve, the performance of cochlear implant can be unpredictable and vary in patients. This problem can be solved via the measurement of nerve function. In 1995, a bidirectional neural response telemetry (NRT) system was incorporated into the Nucleus CI24M cochlear implant to wirelessly monitor the ECAPs in response to the electrical stimuli on auditory nerve [44]. Fig. 1.3 shows the NRT system in Nucleus CI24M [43]. The stimulation parameters are first transmitted from speech processor to implant via RF link. On the implant side, digital signals are received and decoded (Rx Decode), and the electrical stimuli (Stim) corresponding to stimulation parameters are delivered onto auditory nerve via intra-cochlear electrode. Neural signals recorded (Rec) from electrode neighboring to the stimulation one are encoded digitally and transmitted (Tx Encode) back to speech processor and host PC. On the host PC, stimulus artifacts recorded along with neural signals are removed with dedicated algorithms. For example, the mask-probe paradigm proposed by Brown et al. [45] is adopted in Nucleus CI24M cochlear implant [43]. The ECAP response to the electrical stimulus is derived by coherently averaging all responses collected from a stimulation trial containing identical, repeated stimulus pulses. Features on the ECAP waveform are then identified, and stimulus strength is adjusted accordingly. Clinical studies have validated the capacity of NRT to wirelessly measure ECAP responses [43, 46, 47], and it can be equivalently applied to other ENS systems requiring wireless monitoring of ECAP responses on the nerve, including the newly proposed ANC platform.

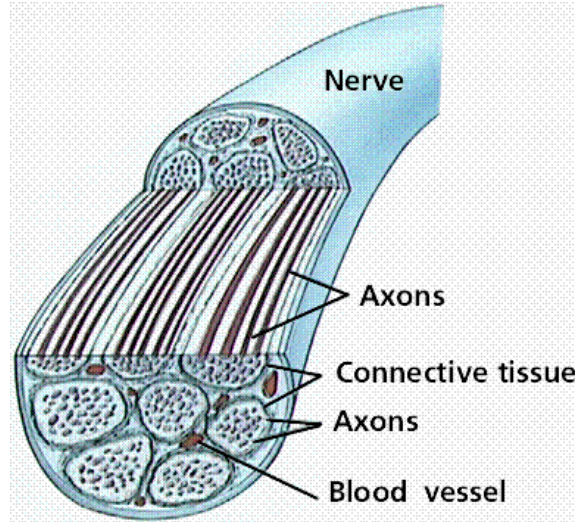


Fig. 1.4 Cross-section of a nerve and group of nerve fibers (axons) [48].

1.4 Electrically-Evoked Compound Action Potential

Fig. 1.4 illustrates the cross-section of a nerve and its nerve fibers (axons) on which action potentials propagate [48]. When electrical stimulus above stimulation threshold (the minimum required stimulation current to elicit an action potential) is applied to a nerve, groups of nerve fibers are activated simultaneously. The summation of all evoked action potentials from activated fibers is called compound action potential. Instead of the action potential of single axon, it is the electrically-evoked compound action potential (ECAP) that can be recorded externally as it propagates along the nerve. Several approaches to interfacing electrodes with nerves for ECAP recording have been referred in [49]. Cuff electrodes, which are designed to fit around nerve without invasion, possess the advantage of maintaining stable and long-term contact with nerve yet exerting little pressure. This makes cuff electrode suitable for implantation and the most popular electrode for nerve stimulation and ECAP recording.

Type of fiber	Diameter (micrometers)	Conduction velocity (m/sec)	General function
A- α	13–22	70–120	α -motoneurons, muscle spindle primary endings, Golgi tendon organs, touch
A- β	8–13	40–70	Touch, kinesthesia, muscle spindle secondary endings
A- γ	4–8	15–40	Touch, pressure, γ -motoneurons
A- δ	1–4	5–15	Pain, crude touch, pressure, temperature
B	1–3	3–14	Preganglionic autonomic
C	0.1–1	0.2–2	Pain, touch, pressure, temperature, postganglionic autonomic

Fig. 1.5 Classification of nerve fiber types [49].

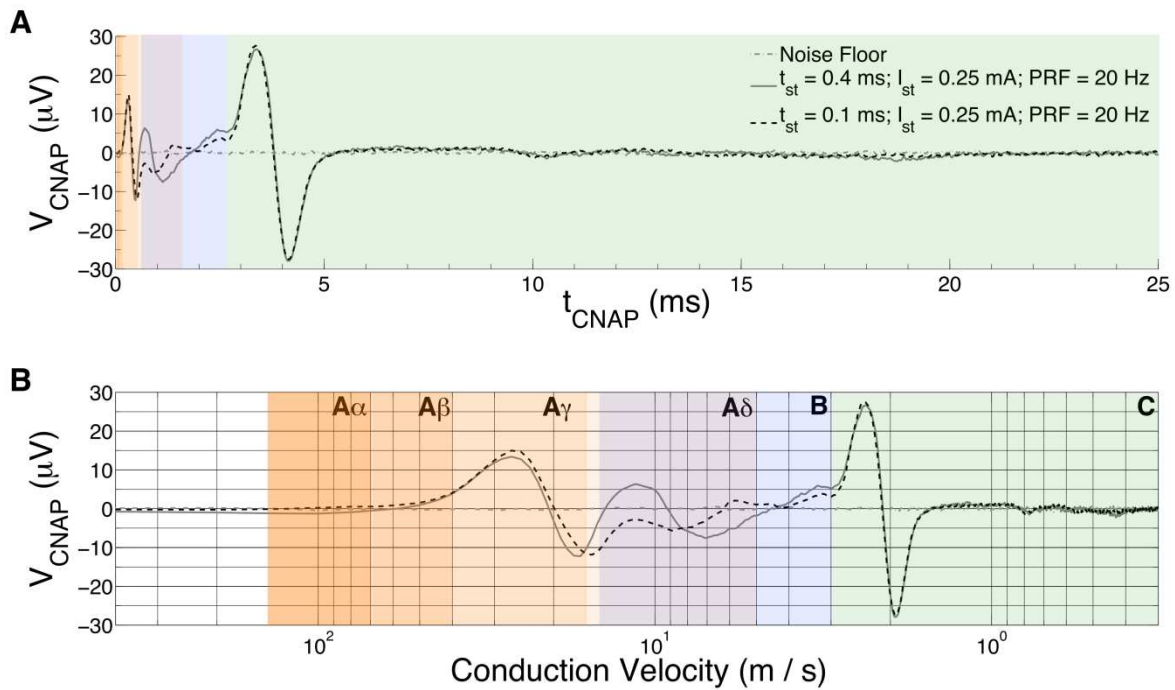


Fig. 1.6 Classification of nerve fiber responses on ECAP waveforms plotted (A) against time axis and (B) as a function of conduction velocity. The ECAP responses are obtained from the left cervical vagus nerve of rat (*Conduction distance* = 8.0 ± 0.5 mm) [35].

The action potentials of nerve fibers of the same group propagate along a nerve at a constant velocity called conduction velocity, which is proportional to the diameter of fibers [37]. Fig. 1.5 shows the classification of nerve fibers in letter systems established by Gasser [37] based on their conduction velocity [49]. When the ECAPs are recorded on a nerve at a fixed and known conduction distance (i.e., the distance between stimulation and recording electrode on a nerve), the responses of nerve fiber groups with specific conduction velocity form peaks with constant latency on ECAP waveform. As the stimulus strength is increased, the amplitude of peaks grow

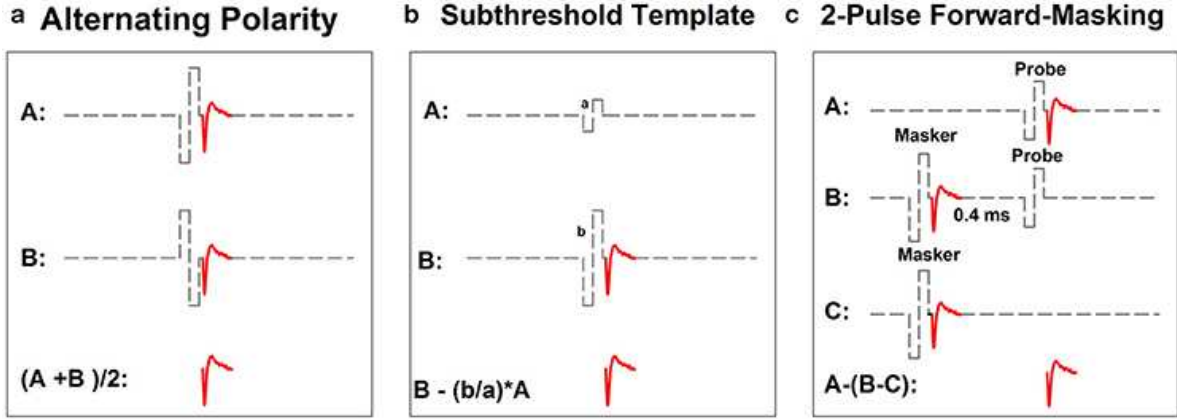


Fig. 1.7 The three most commonly used methods for stimulus artifact reduction in ECAP recording: (a) alternating polarity, (b) subthreshold template subtraction, and (c) 2-pulse forward masking paradigm [50].

accordingly owing to more fibers of that group activated. Fig. 1.6 shows the classification of nerve fiber responses on ECAP waveforms recorded at approximately 8-mm conduction distance [35], where the responses of A, B and C fibers, whose conduction velocities are listed in Fig. 1.5, peak separately within a fixed time range on ECAP waveforms. The latency and amplitude of these peaks, which indicate the type of activated nerve fiber and its activation level, respectively, are important biomedical features on ECAP waveforms. It should be kept in mind that a proper conduction distance must be chosen in order to separate the responses of different fiber groups while keep their response amplitudes.

Recording of ECAP on nerves is inevitably accompanied by stimulus artifact and ambient noise [50]. Such stimulus artifact usually contaminates the recorded ECAP signal and, at large enough amplitude, even saturates the recording amplifier, which hinders the amplifier from further recording. Stimulus artifact results mainly from the voltage gradients between the recording electrodes caused by current flowing through the tissues around nerves, and the electromagnetic coupling between stimulation and recording electrode [51], which can be reduced by increasing conduction distance (i.e., placing recording electrode further away from stimulation electrode). For implantable devices (e.g. cochlear implants) in which large enough conduction distance to completely eliminate stimulus artifact is impractical, additional techniques are required to suppress stimulus artifact. Fig. 1.7 illustrates the three most commonly used stimulus artifact

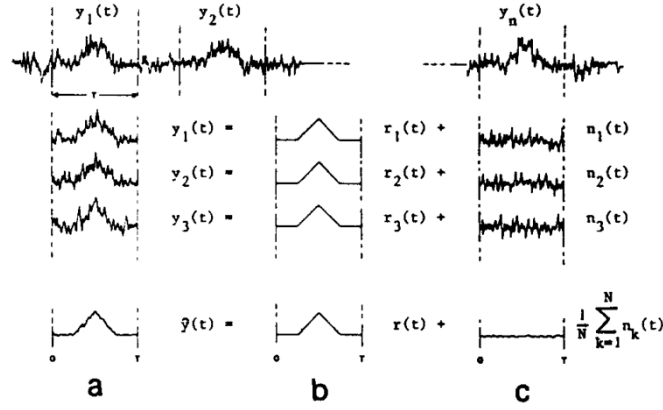


Fig. 1.8 Principle of the coherent averaging (CA) technique [52].

rejection (SAR) techniques in ECAP recording: alternating polarity, subthreshold template subtraction, and two-pulse forward masking paradigm [50]. The alternating polarity method in Fig. 1.7 (a) utilizes two stimulus pulses, a cathodal pulse and an anodal one, whose amplitude and shape are the same and polarity is opposite. On the ground that the polarity of ECAP response does not change with that of stimulus, stimulus artifact is removed by summing the cathodal and anodal responses whose stimulus artifacts have symmetric shape and opposite polarity. In the subthreshold template subtraction method, a pure stimulus artifact is evoked with a subthreshold stimulus pulse (i.e., stimulus below stimulation threshold), which serves as the template. The stimulus artifact is removed by subtracting the evoked response (ECAP plus artifact) with a scaled template. The two-pulse forward masking paradigm utilizes the refractory period of nerve where another stimulation leads to no ECAP [46] and aims to obtain a pure stimulus artifact within this period. As seen in Fig. 1.7 (c), either masker or probe pulse alone elicits both ECAP response and stimulus artifact, whereas the probe pulse within the refractory period that follows after the masker pulse elicits only a stimulus artifact. Artifact is then removed by summing the responses to the abovementioned stimuli (i.e., masker, probe, and masker plus probe).

Coherent averaging (CA), also called ensemble averaging, is a commonly used technique to recover evoked responses from recording noise and other signals that are not correlated to the evoked response and degrade the signal-to-noise ratio (SNR) [52-54]. The CA is based on the

principle that the response to the applied stimulus remains invariant throughout the entire stimulation, which is generally true for stimulation trials lasting for only seconds. Fig. 1.8 illustrates the principle of CA technique [52]. Assume that a series of N equidistant and identical stimuli are applied, and $y_i(t)$ is the output signal after the i -th stimulus that contains response $r_i(t)$ and noise $n_i(t)$, i.e., $y_i(t) = r_i(t) + n_i(t)$. The coherently averaged signal $\hat{y}(t)$ is the time-aligned averaging of all output signals from N stimuli, namely, $\hat{y}(t) = \frac{1}{N} \sum_{i=1}^N y_i(t) = r(t) + \frac{1}{N} \sum_{i=1}^N n_i(t)$, based on the invariance of response $r(t)$ (i.e., $r(t) = r_1(t) = \dots = r_N(t)$). The random noise plus uncorrelated signals $\frac{1}{N} \sum_{i=1}^N n_i(t)$ will then be averaged toward zero. CA is also equivalent to a low-pass finite-impulse-response (FIR) filter. Detailed descriptions and equations can be found in [52].

In short, ECAP is a direct and objective measurement of nerve activity and function and has been adopted as a biomarker in various diagnoses of neural diseases [49, 50] and closed-loop ENS systems such as previously mentioned cochlear implants and ANC platform, in combination with the stimulus-artifact-rejection techniques and CA. Its recording, processing and characterization still present challenges yet deserve more studies for improvement of neurological therapeutics.

1.5 Real-Time Digital Signal Processing for Closed-Loop Neurostimulation

As mentioned in Section 1.1, closed-loop neurostimulation requires continuous monitoring of physiological signals from patients and stimulation of nervous system with stimulus strength constantly adjusted in response to changes in recorded signals, which, until today, is achieved with medical equipments connected to patients via external cables in most clinical treatments. This is problematic, as these equipments and their setup are generally costly in time and money, and most importantly, the transcutaneous cable connection between nervous system and equipments results in not only patient's discomfort but also the risk of injury or inflammation on nerve, which degrades the quality of treatments. The progress in consumer electronics and semiconductor technologies has thus promoted the development of wireless wearable (or implantable) devices for various closed-loop neurostimulation systems in commercial off-the-shelf (COTS) components or application-specific integrated circuits (ASICs) [55-59].

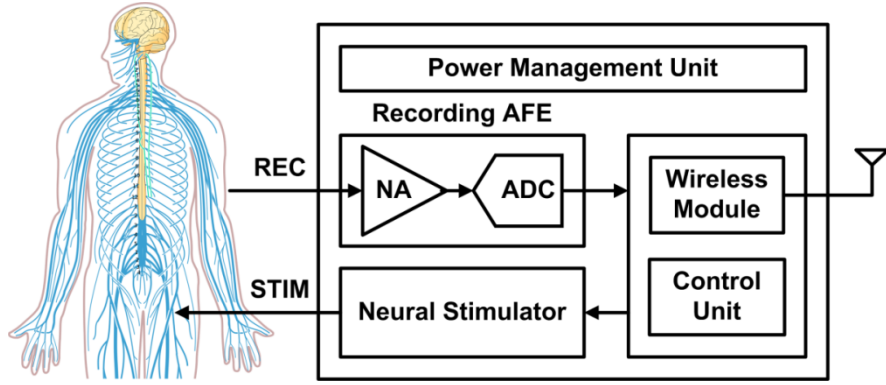


Fig. 1.9 High-level block diagram of a typical wireless wearable (or implantable) device for closed-loop neurostimulation.

Fig. 1.9 illustrates the high-level block diagram of a typical wireless, wearable (or implantable) closed-loop neurostimulation device. Neural signals from central or peripheral nervous system as well as other physiological signals (e.g. EEG, ECG) are recorded and digitized with neural amplifier (NA) and analog-to-digital converter (ADC) in recording analog front-end (AFE), respectively. The wireless module provides a bidirectional communication interface, by which digitized data from ADC are transmitted and user commands are received. Electrical stimuli are generated with the neural stimulator according to the stimulus parameters decoded from received user commands by the control unit, and delivered to targeted nerve or brain areas. The device is powered using battery or wireless power transfer (WPT), and the power supply of each building block is generated with the power management unit.

In order to extend the lifetime of battery and avoid excessive density of WPT that can heat up and damage tissues, power consumption is always the first consideration in designing wireless devices. The Federal Communications Commission restricts the maximum power density of electromagnetic field to 6 W/m^2 at 915 MHz and 10 W/m^2 at 2.4 and 5.8 GHz [60]. On the other hand, the resolution of neural or physiological signals recorded from wireless device and the corresponding data transmission rate must be high enough for users to distinguish the change in signals and adjust stimulus parameters accordingly. For instance, a 192-kbps sampling rate per channel ($8 \text{ bits} \times 24\text{-kHz}$ sampling frequency) is required for recording of action potential (also called "spike") from a neuron, and the resulting data rate of a 64-channel wireless neural

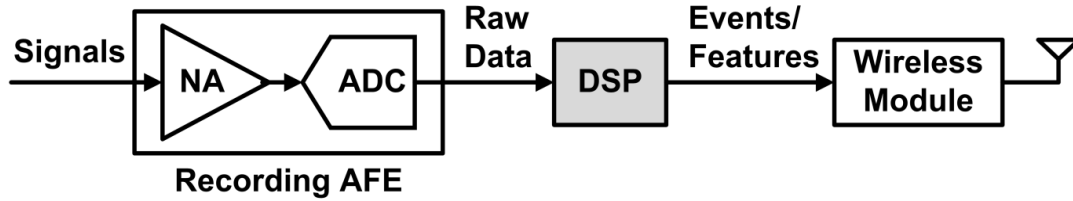


Fig. 1.10 Illustration of a real-time digital signal processing (DSP) engine on a wireless device.

recording device is as high as 11.71 Mbps [61]. Unfortunately, continuous transmission of recorded raw data at high rate is power-costly for the wireless module dominating the power consumption of wireless device. For example, the power consumption of a Bluetooth transceiver during transmission can reach up to 102.6 mW (57 mA at 1.8-V voltage supply) at 0.72-Mbps data rate [62]. Besides, data transmission at high rate results in high data error rate that also degrades the fidelity of recorded signals. As there's limited room for improvement in the power dissipation of wireless module, a better approach to saving power cost of wireless device in closed-loop systems is to reduce the data rate of device by sending only key information in recorded signals relevant to stimulation adjustment.

A real-time digital signal processing (DSP) engine capable of decoding recorded neural or physiological signals can effectively reduce the data transmission rate of wireless devices, and its role on a wireless device is illustrated in Fig. 1.10. Digitized data from ADC are processed by the DSP engine in real time, and only the detected events or extracted features on recorded signals are transmitted by the wireless module at full resolution, based on which the stimulation intensity is adjusted. Such DSP engine can be implemented in microcontroller, field-programmable gate array (FPGA) or ASIC, on which the DSP algorithms must be computationally efficient in order to minimize the implementation cost. Several examples of digital processor for neural signal processing are given as follows. A spike-sorting DSP chip in 90-nm complementary metal-oxide-semiconductor (CMOS) process is presented in [61] for detection and feature extraction of neuron spikes from 64 channels simultaneously, which has a power dissipation of only 130 μ W and reduces data rate from 11.71 Mbps to 1.02 Mbps. A neural signal processor is proposed for a 32-channel neural recording system which utilizes discrete wavelet transform (DWT) and run length encoding (RLE) for neural data compression [63]. This processor, implemented in 130-nm

CMOS process, consumes 800 μ W of power and reduces the maximum data rate of a 32-channel neural implant to 1 Mbps. A general-purpose wireless brain-machine-brain interface (BMBI) is reported in [64], which incorporates a microcontroller-based neural signal processor for digital filtering, feature extraction, spike detection, and compressed sensing. In [65], real-time algorithms for decoding of electroneurogram (ENG) are implemented onto an off-the-shelf DSP processor, which consist of denoising, spike detection, spike sorting by template matching, and classification. For optimal performance and lower area and power cost, very-large-scale integration (VLSI) architecture for real-time DSP and its implementation in either FPGA or ASIC are usually preferred.

1.6 Motivation

There has been significant progress in development of wireless wearable (or implantable) device and real-time DSP algorithm and architecture for various closed-loop neurostimulation systems. Surprisingly, today's closed-loop ENS systems that measure ECAP as feedback biomarker, such as the NRT system in cochlear implant and the newly proposed ANC platform [35], still rely on the offline processing of continuously recorded and transmitted neural data on software. For instance, the Nucleus CI24M cochlear implant incorporates a custom NRT software for post-processing of received neural data [43], including the SAR and CA techniques for recovery of ECAP responses described in previous sections and feature extraction from ECAP signals, and similar signal-processing steps are implemented on MATLAB software in ANC. The required data rate for transmission of neural data (e.g. 800-kbps input data rate in ANC) will be too high for wireless devices to work with these closed-loop ENS systems while satisfy the low power demand. It's favorable to have a DSP engine for ECAP processing that comprises SAR, denoising, and extraction of features such as fiber responses described in Section 1.4, to reduce the data rate of wireless device in these systems. A VLSI architecture of such DSP engine is especially desirable for performance and hardware cost optimization.

Although CA technique has been widely adopted in many ECAP-based closed-loop ENS platforms for noise removal, its efficacy is strongly dependent on the number of averaging (i.e., the number of stimuli). A large number of stimuli and hence long stimulus train duration is needed to boost the filtering capacity of CA, which also adds power consumption to wireless

device due to increase in the length of data recording and transmission time. Another deficiency of CA is its limited ability to remove periodic noises, such as electromagnetic interferences (e.g. 60-Hz power line) and baseline wanders (caused by patient's movement) which are prevalent in neural recording, especially when periodic noises are time-locked to stimulus pulses. Such noises, if not properly eliminated, will introduce inaccuracy onto biomedical features of ECAP (fiber responses), which adversely influences the tuning of stimulus parameters in closed-loop stimulation. Digital filters remove periodic noises more effectively and can be efficiently implemented in real-time DSP in finite impulse response (FIR) or infinite impulse response (IIR) structures [66]. Unfortunately, most digital filters, especially IIR filters which are more computationally efficient, have nonlinear phase response; This causes phase-frequency distortion of filtered signals (i.e., all frequency components of input signal shifted in time unequally) which results in the deformation of ECAP waveform and its features. It's possible to achieve zero-phase filtering and hence avoid distortion of ECAP waveform by applying a filter both forward and backward in time, which's also known as forward-backward filtering [67]. However, this technique requires a time-reversal operation on the entire input data stream (i.e., all neural data recorded during the stimulus train) and are still performed with offline software processing today. A distortion-free and computationally efficient filtering technique for more effective periodic noises removal and its VLSI architecture is essential to realize a real-time DSP engine for ECAP processing.

This thesis focuses on the design of a DSP engine dedicated to ECAP-based closed-loop ENS systems, including NRT and ANC systems, and its VLSI architecture. This real-time DSP engine performs SAR and filtering to recover ECAP from the stimulus artifact and noises and extracts fiber responses from recovered ECAP waveform. A computationally efficient filtering technique named bidirectional-filtered coherent averaging (BFCA) and its VLSI architecture is proposed for real-time denoising of ECAP, by which periodic noises are more effectively removed without introducing waveform distortion. With the DSP engine developed in this work, data transmission rate can be greatly reduced, which enables wireless devices to work with ECAP-based closed-loop ENS systems at reasonable power cost. The BFCA method removes periodic as well as random noises more efficaciously, which helps improve the precision of extracted biomedical features (e.g. fiber responses) and the performance of closed-loop stimulation.

1.7 Outline of Thesis

Chapter 2 presents a DSP architecture for real-time recovery of ECAP responses in NRT as well as other ECAP-based closed-loop ENS systems, which consists of SAR via alternating-polarity (AP) stimulation and denoising with the proposed BFCA method. The principle of the BFCA and its combination with AP technique is explained, and the VLSI architecture of BFCA algorithm AP stimulation-based SAR are described. Design techniques such as folded IIR filter and division-free averaging are presented for hardware efficient implementation. The stimulation and recording AFE circuitry interfacing with the DSP is also described. This DSP architecture is implemented on FPGA and verified in in-vivo ENS, and its efficacy is evaluated in terms of residual stimulus artifact, noise floor, and waveform distortion.

Chapter 3 extends the work of Chapter 2 and presents fiber-response extraction engine (FREE), the first real-time DSP engine designed for nerve activation control in closed-loop ENS using the ANC platform, to the best of our knowledge. Computationally efficient algorithms and VLSI architectures are presented for extraction of fiber responses from ECAP responses derived with the DSP architecture in Chapter 2. A custom-made wearable wireless device is built in printed circuit board (PCB) prototype that comprises a low-power FPGA onto which FREE is mapped, a Bluetooth transceiver, the stimulation and recording AFE circuitry described in Chapter 2 and a power-management circuitry, and can be powered with a single coin-cell battery. This wearable device is integrated into ANC system to verify the performance of FREE. Both offline and *in-vivo* experimental results show that compared with previous software-based processing in ANC, not only does FREE help reduces the required data transmission rate of wireless device, but the precision of extracted fiber responses is improved through the proposed BFCA. High-precision fiber responses obtained from FREE contributes to increase in the accuracy of NAP construction in ANC and hence closed-loop stimulation efficiency. FREE is also implemented in 180-nm CMOS technology, whose total chip area is 19.98 mm^2 and core power consumption is 1.95 mW at 1.8-V core voltage and 16-MHz system clock rate.

Finally, Chapter 4 draws a conclusion to this thesis and describes the future work for this research. The output data rate of FREE can be further reduced by employing DWT and RLE for compression of ECAP response, and its preliminary result is demonstrated. All the computations

in FREE can be implemented with the half-precision floating-point arithmetic that provides sufficient data precision whereas reduces the computation costs. Wireless powering technique can be further incorporated in order to make the wireless device implantable, and the required components in the power management unit are illustrated. Bluetooth transceivers with lower power consumption may be adopted to reduce the overall power cost of the wireless device.

2. A DSP ARCHITECTURE FOR REAL-TIME EVOKED COMPOUND ACTION POTENTIAL RECOVERY IN NEURAL RESPONSE TELEMETRY SYSTEM

This chapter presents the first digital signal processing (DSP) architecture for real-time recovery of electrically-evoked compound action potentials (ECAPs) from stimulus artifacts and periodic noises in bidirectional neural response telemetry (NRT) system. In this DSP architecture, a bidirectional-filtered coherent averaging (BFCA) method is proposed for configurable and distortion-free filtering of the ECAP waveforms, and the alternating-polarity (AP) stimulation method is utilized for rejecting stimulus artifacts overlapped with ECAPs, which can be easily incorporated into the proposed BFCA method. Design techniques including the configurable folded infinite-impulse-response (IIR) filter and division-free averaging are also presented for efficient hardware implementation. Synthesized in 180-nm CMOS process, the proposed DSP architecture consumes 0.97-mm^2 area and 2.38-mW power. The efficacy of the DSP architecture in recovering ECAPs from recorded neural data contaminated by overlapped stimulus artifacts and periodic noises is validated in *in-vivo* electrical nerve stimulations. Experiment results show that compared with the previous coherent averaging technique, the proposed DSP architecture improves the signal-to-noise ratio (SNR) of ECAP responses by 11 dB and achieves an 3.1% waveform distortion that is $17.1\times$ lower.

2.1 Introduction

Neural response telemetry (NRT) is an useful technique to wirelessly measure electrically-evoked compound nerve action potential (ECAP) for the study of the nervous system using implantable devices [46]. The measured ECAP reflecting the activity of the nerve being stimulated serves as an objective criterion for adjustment of stimulus parameters in closed-loop electrical nerve stimulation. Fig. 2.1 shows a conventional bidirectional NRT system for closed-loop stimulation [43, 68]. Bidirectional communication between the host personal computer (PC) and the radio-frequency (RF) transceiver on the implant is established with a base station (BS), and instructions from host PC are sent to implant and decoded to deliver user-defined stimulus train onto the nerve. Neural responses to the stimuli are recorded and digitized on the implant, and transmitted back to the host PC, on which data from the implant are processed to recover the

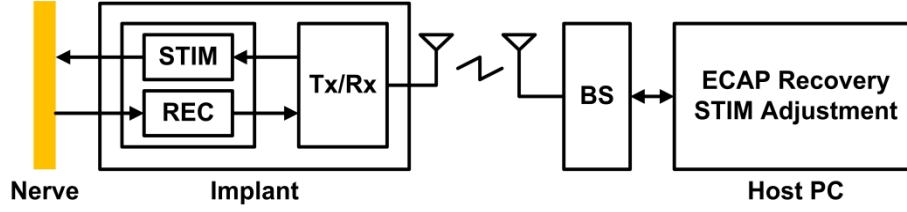


Fig. 2.1 Illustration of Conventional bidirectional neural response telemetry (NRT) systems.

ECAP responses and stimulus parameters are adjusted. NRT is first introduced into the electrical stimulation of the auditory nerve in cochlear implants decades ago, which has been widely applied in clinical treatment since then with over 200,000 patients implanted [40, 41]. A recently proposed response-driven electrical nerve stimulation platform, autonomous nerve control (ANC) [35], provides another promising application area for NRT. In ANC, ECAP responses to a pre-defined stimulus are decoded to identify targeted nerve fiber response, and the stimulation parameters are constantly updated according to a patient-specific nerve activation profile to control the activation level of nerve fiber. NRT system can be integrated into ANC for wireless measurement of ECAPs, offering close-loop electrical nerve stimulation on implantable devices with improved efficiency.

On the implant of a conventional NRT system, neural signals are continuously sampled and transmitted to host PC via RF transceiver during a stimulation trial that typically consists of a pulse train with specific stimulation rate (number of stimulus pulses per second) and duration [14]. As reported in [35], a 50-kHz sampling frequency and a 16-bit sampling precision (the analog-input precision of USB-6353, *National Instruments*) are required to resolve ECAPs with milli-volt amplitude after amplification, and the resulting data transmission rate will be 800 kbps on the implant. Whereas RF transceiver dominates the power consumption of the implant, continuous transmission of raw data wirelessly at such high data rate is not only power-costly for the implant, but also vulnerable to data loss in wireless transmission that degrades the fidelity of derived ECAP responses on host PC. A digital signal processing (DSP) hardware capable of recovering ECAP response from recorded raw neural data and extracting biomedical features of interest is thus welcome for minimizing data transmission rate in NRT systems. Fig. 2.2 (a) illustrates an NRT system with the above- mentioned DSP hardware, and details on the DSP unit

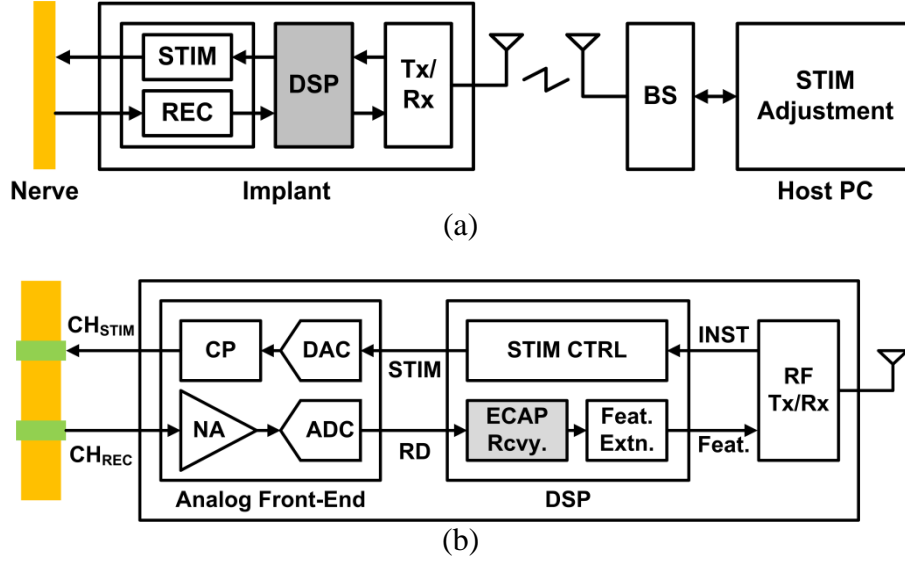


Fig. 2.2 (a) The NRT systems with a digital signal processor. (b) Details of stimulation and recording analog front-ends (AFE) and digital signal processor on the implant.

and the stimulation and recording analog front-end (AFE) on the implant are plotted in Fig. 2.2 (b). Based on the stimulation parameters in the decoded instruction received from host PC, a digitized stimulus waveform is generated from the stimulation controller on DSP unit and converted to a current stimulus on the stimulation electrode by the digital-to-analog converter (DAC) and the current pump (CP). Neural signals in response to the stimulus are picked up by the recording electrode in the neighborhood, conditioned by neural amplifier (NA) and digitized by analog-to-digital converter (ADC). On the DSP unit, digitized raw neural data (*RD*) are processed to obtain the ECAP response to the applied stimulus train, and biomedical features such as nerve fiber responses on the ECAP are extracted and transmitted back to host PC. Developing a hardware architecture for real-time ECAP recovery is the first step to the realization of the above-mentioned DSP unit, and the required signal-processing steps will be described later.

One challenge in ECAP recovery is the accompanying of electrical stimulus artifacts [51]. In implantable devices, a limited conduction distance (i.e., the distance between stimulation and recording electrodes) usually results in the overlapping of stimulus artifacts with ECAP responses, necessitating the utilization of stimulus artifact rejection (SAR) techniques on the

DSP unit. While many techniques for stimulus artifact rejection have been reported [69-73], those implemented in DSP hardware are first reviewed. In template subtraction method, a template of artifacts is derived from a series of pure stimulus artifacts recorded in sub-threshold stimulation, and recorded neural data is subtracted by the template signal to remove the stimulus artifacts. A hardware implementation of this method is reported in [74], where a low-cost infinite-impulse-response (IIR) temporal filter architecture is utilized for template generation. Another template generation method based on adaptive filtering and its hardware implementation is proposed in [75]. The main disadvantage of template subtraction is that in real-time stimulation, it's difficult to obtain an artifact template free of overlapped ECAP responses in the absence of accurate estimation of ECAP threshold. The forward masking method [45, 76, 77] aims to generate a pure stimulus artifact during the refractory period of nerve by utilizing a two-pulse stimulus, a high-amplitude masker pulse followed by a probe pulse. This pure stimulation artifact is then properly time-shifted and subtracted from recorded neural signals. Without prior knowledge of the refractory period of nerve being stimulated, however, this method fails if the probe pulse isn't completely within the refractory period, which induces artifacts plus ECAP responses. The alternating-polarity (AP) stimulation method [76], based on the fact that flipping stimulus pulse changes only the polarity of stimulus artifact instead of ECAP response, utilizes a cathodal pulse followed by an anodal one that has the same amplitude and opposite polarity. The resulting artifacts, which are identical in shape and opposite in polarity owing to the symmetry of stimulus pulses, cancel with each other by summing the cathodal and anodal responses within a period of biphasic stimulation. AP stimulation features low complexity and has been proved effective in removing artifacts overlapping with ECAP responses [35], and hence is preferred for implementation of real-time SAR on DSP hardware.

Another challenge in ECAP recovery is the presence of periodic noises such as electromagnetic interferences and baseline drifts which are commonly encountered on the implants. Although the coherent averaging of neural data is equivalent to a low-pass filter, as reported in [52], its performance is limited by the number of averaging cycles, and fails to effectively remove periodic noises, especially those time-locked to stimulus pulses. A programmable digital band-pass filter is necessary to periodic noise removal of ECAP whose frequency spectra varies with both nerve fiber distribution and conduction distance [78]. The major problem with digital

filtering, however, is its nonlinear phase response and the resulting frequency-dependent phase shift on filtered ECAP responses [66], which distorts both ECAP waveform and its biomedical features, especially the latency of nerve fiber responses reflecting distribution of nerve conduction velocity [37, 79]. The waveform deformation caused by nonlinear phase response of digital filters can be circumvented with zero-phase filtering (ex: `filtfilt` function in MATLAB). So far, this technique is only implemented on the software, which requires the entire raw data stream to be transmitted to host PC for offline processing. Wavelet filtering based on wavelet decomposition and reconstruction has been proved effective in removing low-frequency noise whereas maintaining the waveform shape and has been adopted for denoising of various biomedical signals [80, 81]. A hardware efficient very-large scale integration (VLSI) architecture of wavelet filtering is also presented for its real-time DSP implementation [82]. Nevertheless, unlike conventional digital filters, the programmability of passband in wavelet filtering is strictly limited due to the intrinsic property of discrete wavelet transform [83], making it unsuitable for filtering of ECAP. For real-time ECAP recovery, it is essential to develop a programmable and distortion-free filtering strategy and its computationally-efficient hardware implementation.

In this chapter, we present the first DSP architecture for real-time and distortion-free recovery of ECAPs from stimulus artifacts and periodic noises in bidirectional NRT systems. In this DSP architecture, a bidirectional-filtered coherent averaging (BFCA) method is proposed for configurable and distortion-free filtering of the ECAP waveforms, and the AP stimulation method is utilized to reject stimulus artifacts overlapped with ECAPs, which can be easily combined with the BFCA method. For hardware-efficient implementation, both the architectures of configurable folded IIR filter and exponentially-weighted moving averaging (EWMA) [84] are presented. Synthesized in 180-nm CMOS process, the proposed DSP architecture consumes 0.97-mm^2 area and 2.38-mW power. This DSP architecture is tested in *in-vivo* electrical nerve stimulations to verify its efficacy of removing overlapped stimulus artifacts and periodic noises. Compared with the previous coherent averaging technique, the proposed DSP architecture improves the signal-to-noise ratio (SNR) of ECAP responses by 11 dB and achieves an 3.1% waveform distortion that is $17.1\times$ lower.

This chapter is organized as follows. Section 2.2 describes the principle of bidirectional-filtered coherent averaging (BFCA) method for distortion-free artifact and noise removal on ECAP. Section 2.3 describes the proposed DSP architecture for real-time ECAP recovery. Section 2.4 presents results of FPGA and CMOS implementation of the DSP engine and its verification via *in-vivo* experiments, and Section 0 draws a conclusion of this work.

2.2 Bidirectional-Filtered Coherent Averaging

Coherent averaging is a useful method to extract evoked neural responses [52]. In this method, a stimulus train consisting of a series of identical and equidistant stimulus pulses is applied to the nerve. It's assumed that the nerve response to the same stimulus pulse in a stimulus train is invariant, which is generally valid for a stimulus train lasting for only a few seconds. An ECAP response to a stimulus train is obtained by systematically aligning and averaging of all evoked responses to a single stimulus pulse. During the averaging process, random noise components recorded with ECAPs are summed toward zero, contributing to a higher signal-to-noise ratio (SNR). Coherent averaging can be easily combined with AP stimulation method for SAR, in which an artifact-free ECAP response is attained by first aligning and summing the cathodal and anodal responses within an AP stimulus period and coherently averaging the summed waveform of all AP stimulus cycles [35].

A linear-phase programmable filter before coherent averaging is applicable to eliminate periodic noise interferences whereas avoid distorting ECAP waveforms in recorded raw neural data [85]. The simplest way to realize linear phase filters is to design finite-impulse-response (FIR) filters with symmetric or anti-symmetric impulse responses [86, 87]. Under the same frequency band and magnitude response specifications, however, FIR filters require much higher order than infinite-impulse-filter (IIR) filters and thus more computation costs. Several methods to derive an IIR filter with desired magnitude response and approximately linear phase in its pass band have been reported in [88]. Nevertheless, these methods require either an order increase in IIR filters or another filters to compensate the phase nonlinearity of IIR filters, both of which degrade the computation efficiency of original IIR filters, and their applications are limited to specific type of filter design due to a partially valid linear-phase relationship. Powell and Chau [89] proposed the linear-phase filter structure implemented with an IIR filter with transfer

function $H(z)$, cascaded by the same filter in time-reversed order, $H(z^{-1})$, which are realized using the local time-reversal of input data of filter $H(z)$. Powell and Chau's filter structure provides an exact linear-phase relationship over the entire frequency band while preserving the computation efficiency of an IIR filter, and can be applied to arbitrary IIR filters $H(z)$ designed with magnitude specification only. Another attractive feature of Powell and Chau's method is the utilization of block processing techniques, where a continuous input data stream is equally divided into finite sections, and each section is bidirectional-filtered with both $H(z)$ and $H(z^{-1})$. Without the requirement for additional data storage, block processing can be easily combined with the coherent averaging technique, where continuously recorded raw neural data are segmented into individual responses to a single stimulus. In this chapter, by integrating coherent averaging with Powell and Chau's linear-phase filter structure, we propose the bidirectional-filtered coherent averaging (BFCA) and its efficient hardware implementation for real-time, linear-phase filtering of ECAP responses.

Fig. 2.3 (a) illustrates the proposed BFCA method for distortion-free artifact and noise removal on ECAP, where TR and CA denote time-domain order reversal and coherent averaging, respectively. Recorded raw neural data from AFE, as plotted in Fig. 2.3 (b), consists of a series of ECAP responses evoked by an AP stimulus pulse train, as well as stimulus artifacts and periodic noise interferences. The raw data (RD) is continuously filtered by filter $H(z)$, whose outcome versus raw data before filtering are plotted in Fig. 2.3 (c). For each AP stimulus cycle, both the cathodal and anodal parts of continuously filtered raw data are sampled with a window time-locked to stimulus pulses, and the windowed data are reversed in time-domain order, as shown in Fig. 2.3 (d). By summing windowed cathodal and anodal responses, the stimulus artifacts, which are symmetric and aligned on time axis, are cancelled to restore the ECAP within an AP stimulus period. The ECAP response to applied stimulus train is computed by averaging ECAP of all AP stimulus cycles, referred as the mean ECAP (uCAP) response. The above-mentioned process is equivalent to the coherent averaging of continuously-filtered and time-reversed raw data, as illustrated in Fig. 2.3 (a). The uCAP response in reverse-time order is filtered with the same response $H(z)$, as shown by Fig. 2.3 (e), and converted back to continuous-time order with another TR operation. Fig. 2.3 (f) shows an ECAP response derived by applying BFCA on the raw data superimposed by periodic noises as seen in Fig. 2.3 (b), and

the original noise-free ECAP. It can be seen that applying BFCA effectively removes periodic noises on recorded raw data, and the resulting ECAP waveform is exactly the same as that of the original ECAP.

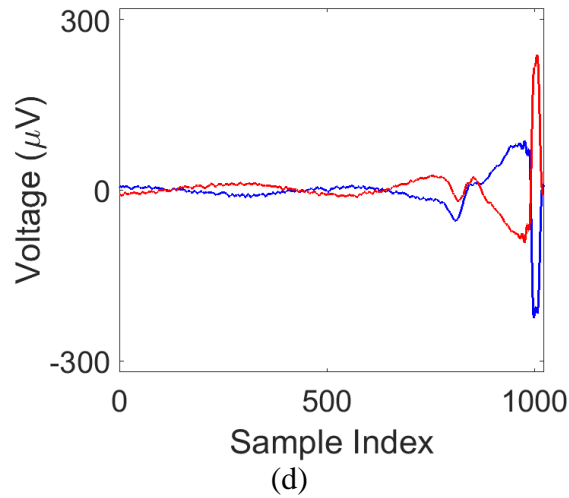
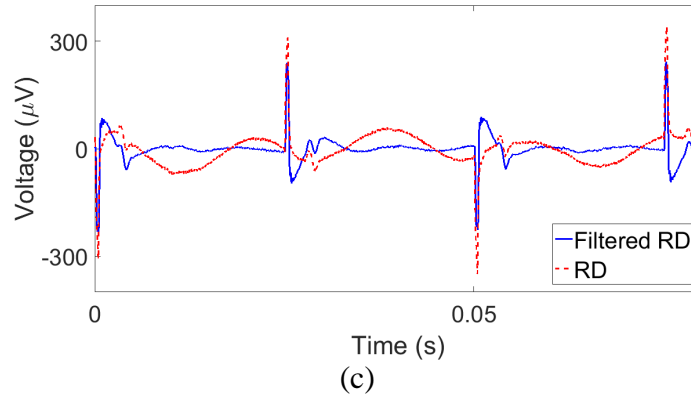
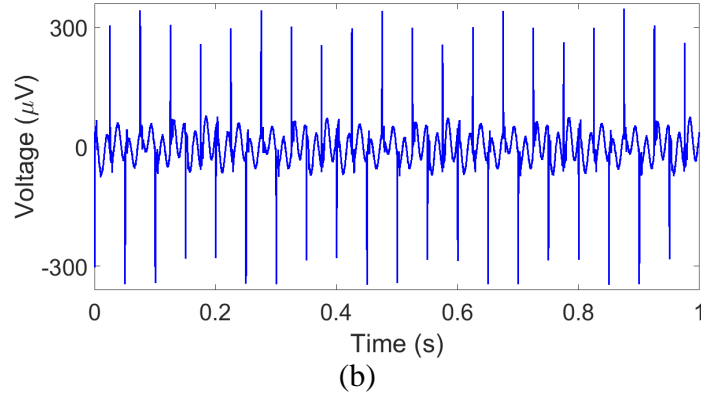
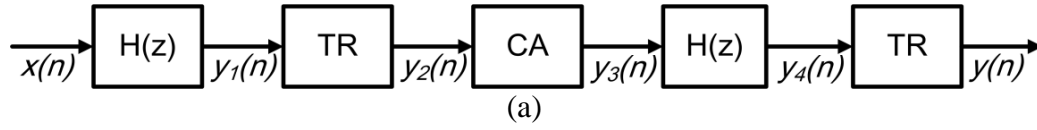


Fig. 2.3 Principle of the proposed bidirectional-filtered coherent averaging (BFCA) method combined with the alternating-polarity (AP) stimulation method for stimulus artifact rejection and distortion-free denoising of ECAP.

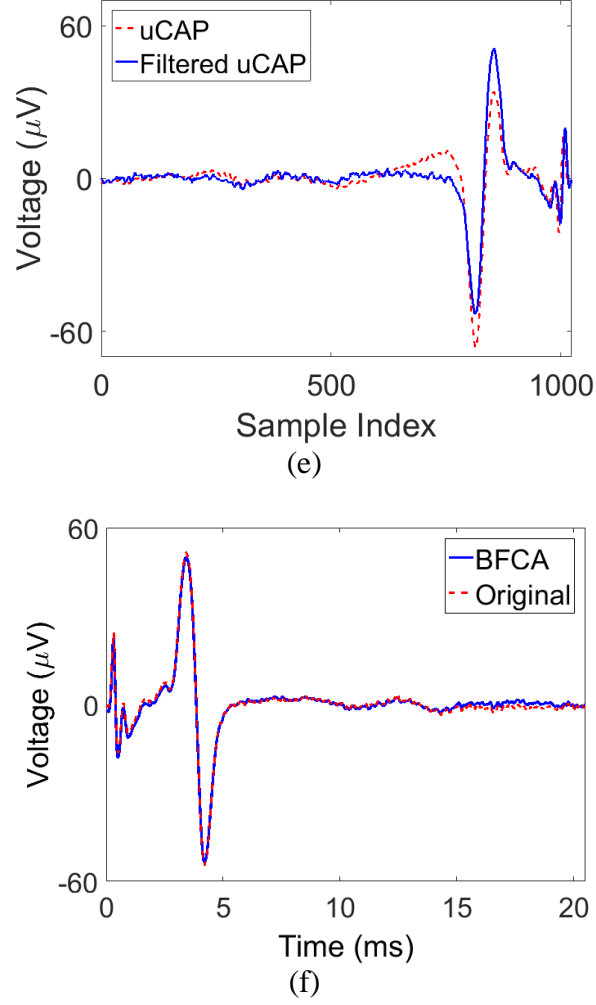


Fig. 2.3 continued.

A linear-phase relationship of BFCA shown in Fig. 2.3 (a) can be verified with its Discrete-Time Fourier Transform (DTFT). The DTFT of filtering process is expressed as

$$\begin{aligned} Y_1(e^{j\omega}) &= H(e^{j\omega})X(e^{j\omega}), \\ Y_4(e^{j\omega}) &= H(e^{j\omega})Y_3(e^{j\omega}). \end{aligned} \quad (2.1)$$

where $H(e^{j\omega})$ is the DTFT of filter response $H(z)$. The TR operation in Fig. 2.3 (a) is defined as

$$\begin{aligned} y_2(n) &= y_1(-n), \\ y(n) &= y_4(-n), \end{aligned} \quad (2.2)$$

and its DTFT is given by

$$\begin{aligned} Y_2(e^{j\omega}) &= Y_1(e^{-j\omega}), \\ Y(e^{j\omega}) &= Y_4(e^{-j\omega}). \end{aligned} \quad (2.3)$$

Assume that a stimulation train consists of N_{ST} AP stimulus pulses and, the time interval between cathodal and anodal stimulus pulses is T . The coherent averaging (CA) of filtered and time-reversed raw data in Fig. 2.3 (a) is defined as

$$\begin{aligned} y_3(n) &= \frac{1}{2N_{ST}} \sum_{k=0}^{2N_{ST}-1} y_2(n + kT) \\ &= y_2(n) * \frac{1}{2N_{ST}} \sum_{k=0}^{2N_{ST}-1} \delta(n + kT), \end{aligned} \quad (2.4)$$

where $*$ denotes convolution operation. The equivalent impulse response of CA is expressed as

$$h_{CA}(n) = \frac{1}{2N_{ST}} \sum_{k=0}^{2N_{ST}-1} \delta(n + kT), \quad (2.5)$$

and its DTFT, as derived in [52], is given by

$$H_{CA}(e^{j\omega}) = \frac{1}{2N_{ST}} \cdot \frac{\sin(\omega TN_{ST})}{\sin(\frac{1}{2}\omega T)}. \quad (2.6)$$

The DTFT of (2.4) can thus be written as

$$Y_3(e^{j\omega}) = H_{CA}(e^{j\omega}) Y_2(e^{j\omega}). \quad (2.7)$$

By summarizing (2.1)-(2.7), the transfer function of BFCA can be derived:

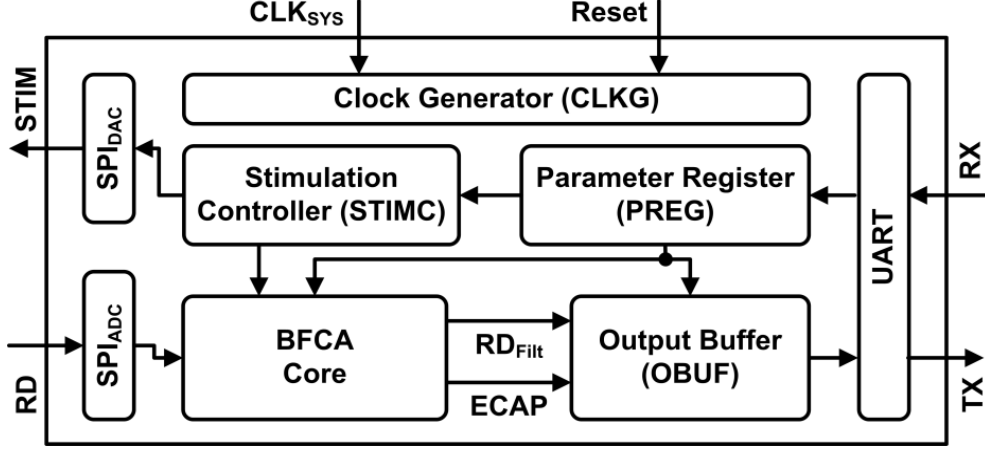


Fig. 2.4 Block diagram of the proposed DSP architecture for real-time ECAP recovery.

$$\begin{aligned}
 Y(e^{j\omega}) &= H_{CA}(e^{-j\omega})H(e^{-j\omega})H(e^{j\omega})X(e^{j\omega}) \\
 &= H_{CA}(e^{j\omega})|H(e^{j\omega})|^2 X(e^{j\omega}), \\
 H_{BFCA}(e^{j\omega}) &= \frac{Y(e^{j\omega})}{X(e^{j\omega})} = H_{CA}(e^{j\omega})|H(e^{j\omega})|^2,
 \end{aligned} \tag{2.8}$$

where the relationship $H_{CA}(e^{-j\omega}) = H_{CA}(e^{j\omega})$ can be verified from (2.6). It can be observed in (2.8) that the transfer function of BFCA, $H_{BFCA}(e^{j\omega})$, is real and positive. Therefore, no frequency-dependent phase shift will be introduced by BFCA, and the shape of ECAP waveforms characterizing the distribution of nerve conduction velocity can be preserved after applying BFCA to raw neural data. Moreover, the BFCA enables the filter $H(z)$ in Fig. 2.3 (a) to be realized with an infinite-impulse response (IIR) filter regardless of its nonlinear phase response, which saves more computation resources than linear-phase FIR filter under the same filter specification.

2.3 Architecture Design

2.3.1 System Overview

Fig. 2.4 shows the proposed DSP architecture for real-time ECAP recovery. The clock generator derives all clock signals from an external system clock. Two SPI masters SPI_{DAC} and SPI_{ADC} control the DAC and ADC in AFE for stimulus generation and neural data acquisition,

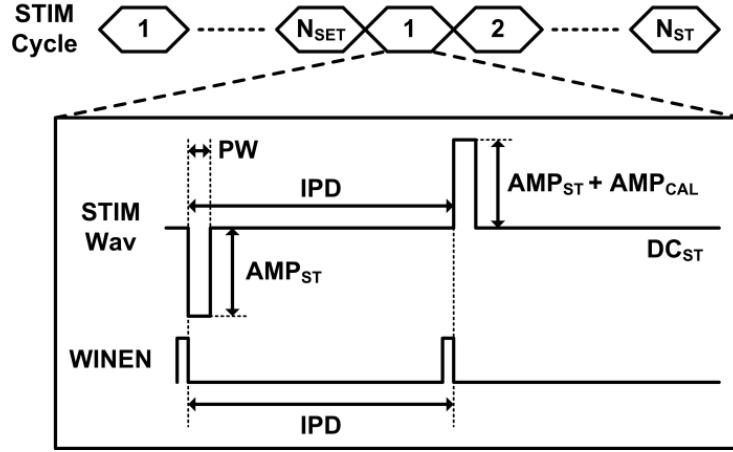


Fig. 2.5 Generation of the alternating-polarity stimulus pulse and time-locked windowing control in stimulation controller.

respectively. All system parameters, including the stimulation control and filter coefficients, are decoded from instructions serially loaded via the UART interface and stored into the parameter register. Based on received stimulation parameters, the stimulation controller derives the digital codes of the AP stimulus pulse train which are to be generated in AFE, and controls the time-locked windowing of raw data in BFCA. The raw data from the ADC are sampled at 50 kHz and digitized to 16 bits as required in [35]. The BFCA core, which is the hardware implementation of the proposed BFCA algorithm, compute the ECAP response to applied stimulation trial from digitized raw data. A configurable output buffer is also included, which selectively outputs either filtered raw data (RD_{Filt}) or ECAP responses according to the user's configuration, and its output data are then serially transmitted with the UART interface. Filtered raw data are transmitted in the beginning stimulation trials, by which users can check the balance between cathodal and anodal stimulus artifacts to ensure successful stimulus artifact cancellation. Once balanced stimulus artifacts are verified, only the artifact- and noise-free ECAP responses will be transmitted to users for further analysis.

2.3.2 Stimulation Controller

The stimulation controller is clocked at 1-MHz frequency to provide 1- μ s time resolution for stimulus pulse train. Fig. 2.5 shows the parameters for AP stimulus pulse generation. The number of AP pulses per train (N_{ST}) and the interphasic delay (IPD), i.e., time spacing between cathodal and anodal pulses, are determined by the pulse repetition frequency (PRF) and the

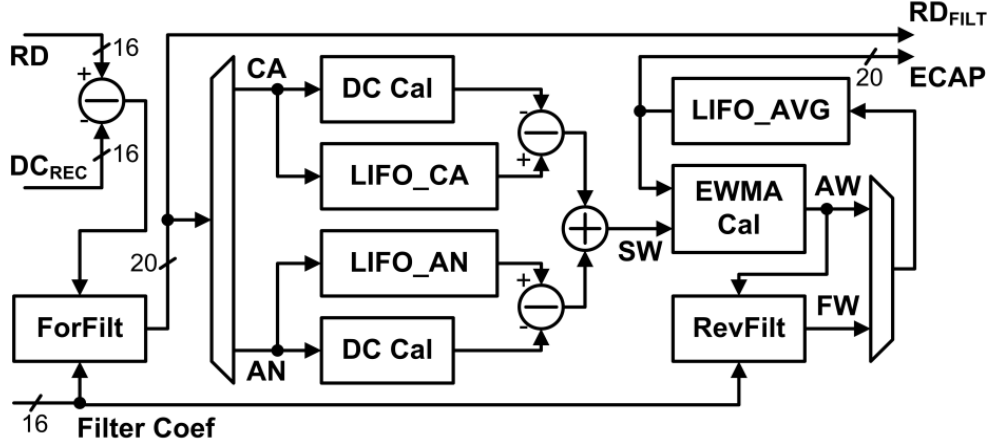


Fig. 2.6 Schematic of the BFCA core.

stimulus train duration (t_{train}) using the relationship $N_{ST} = (PRF \cdot t_{train})$ and $IPD = 1/(2 \cdot PRF)$, both of which are computed offline. Both the pulse width (PW) and IPD are represented in microseconds and stored into the parameter register as integers. The mid-code of the DAC in the stimulation AFE is assigned to the DC level of a stimulus pulse train (DC_{ST}), and the digital code of the stimulus pulse amplitude (AMP_{ST}) is determined offline by the desired current amplitude and the voltage-current relationship of the current pump in AFE. A signed parameter "amplitude calibration" (AMP_{CAL}) is added to anodal pulse amplitude in order to tune its resulting artifact amplitude for balance. Note that a series of N_{SET} settling cycle where no stimulus is applied (i.e., $AMP_{ST} = 0$) is appended before the AP stimulus train in order that the baseline current at the stimulation channel is first stabilized before the stimulus train starts.

The control signal for time-locked windowing of cathodal and anodal responses in digitized raw data is also generated in the stimulation controller. As seen in Fig. 2.5, a windowing-start signal $WINEN$ is launched at the rising edge of each stimulus pulse to start the windowing of recorded raw data. The cathodal and anodal stimulus artifacts of each AP stimulus pulse can thus be aligned on the time axis, as seen in Fig. 2.3 (d), and cancelled during the coherent averaging process.

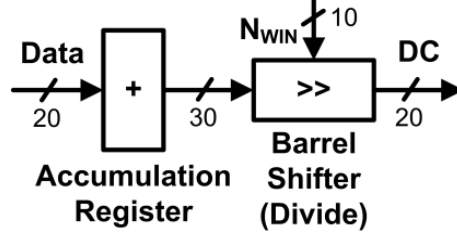


Fig. 2.7 Schematic of the DC calculator.

2.3.3 BFCA Core

Fig. 2.6 shows the schematic of the BFCA core. Digitized raw data from ADC in unsigned integers are converted to two's complement format by subtracting them by DC_{REC} , the mid-code of ADC (32768 for 16-bit ADC precision) corresponding to the common-mode voltage of recording AFE as will be described later. To avoid overflow in fixed-point computation [90], the word length of data paths in BFCA core are set to 20 bits. The signed raw data are filtered continuously with the forward filter (*ForFilt*), and its cathodal (CA) and anodal (AN) parts are windowed and stored into two last-in-first-out (LIFO) registers *LIFO_CA* and *LIFO_AN*, respectively. Note that the windowing of the filtered raw data is started after the settling cycles N_{SET} , when the outputs of the forward filter are settled. When windowed raw data are being stored into corresponding LIFO registers, its residual DC offset is calculated:

$$DC = \frac{1}{N_{win}} \sum_{n=0}^{N_{win}-1} x(n), \quad (2.9)$$

where $x(n)$ is the windowed raw data, and N_{win} is the windowing length which's set to a power of 2 such that division by N_{win} can be accomplished with right-shifting. Fig. 2.7 shows the schematic of the DC calculator in Fig. 2.6 for DC offset calculation of windowed cathodal and anodal raw data using (2.9), where the precision of the accumulation register is extended to 30 bits. The windowing length N_{win} is determined by the *IPD* of the AP stimulus. At 50-kHz sampling frequency, the value of N_{win} is programmable from 256 to 1024 to support a maximum *PRF* of 80 Hz, and the maximum windowing length is 20.48 ms, which is sufficient to cover the nerve fiber responses with slowest conduction velocity in ECAP response given a conduction distance less than 10 mm [35]. The cathodal and anodal responses stored in LIFO registers are

time-reversed, subtracted by their DC offsets, and summed to obtain an artifact-free ECAP waveform, denoted as the summed wave (SW). The uCAP is calculated by averaging SW s of all AP stimulation cycles using exponentially-weighted moving averaging (EWMA), whose principle will be described later. The updated averaging of SW from EWMA, denoted as the averaged wave (AW), is stored into the LIFO register $LIFO_AVG$. When the averaging process is completed at the end of a stimulus train, the AW is filtered in time-reversed order by the reverse filter ($RevFilt$). The outcome of the reverse filter, denoted as the filtered wave (FW), is stored back to the $LIFO_AVG$ and converted to forward-time ECAP as plotted in Fig. 2.3 (f). Both the forward filter and reverse filter are implemented in IIR filter structures with input and output word-length of 20 bits and internal word-length of 28 bits for overflow prevention, and have the same filter response by sharing the identical filter coefficients. Filter coefficients stored in the parameter registers are quantized to 16 bits and programmable online according to the frequency spectrum distribution of ECAP responses.

The impact of quantization noise on the BFCA algorithm is evaluated by comparing the performance of BFCA algorithms in the fixed-point precision specified in Fig. 2.6, versus that in floating-point precision. A set of offline-recorded raw neural data obtained from 40 stimulation trials is quantized to 16-bit precision and fed into both floating-point and fixed-point BFCA algorithms. The signal-to-quantization-noise ratio (SQNR) of ECAP responses obtained from the fixed-point BFCA algorithm is defined as

$$SQNR = 20 \log_{10} \left(\frac{S_{out,rms}}{N_{Q,rms}} \right), \quad (2.10)$$

with S_{out} and N_Q denoting the reference ECAP output in floating-point precision and quantization noise, respectively [90]. The averaged SQNR obtained from 40 raw neural data is 37.11 dB, which is sufficient for the discrimination of nerve fiber responses on ECAP waveforms.

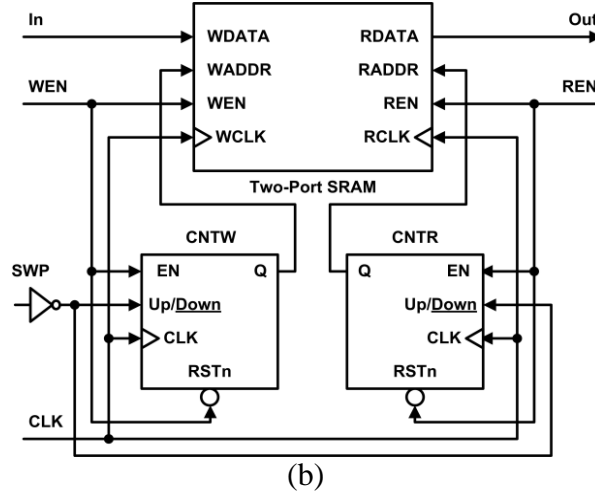
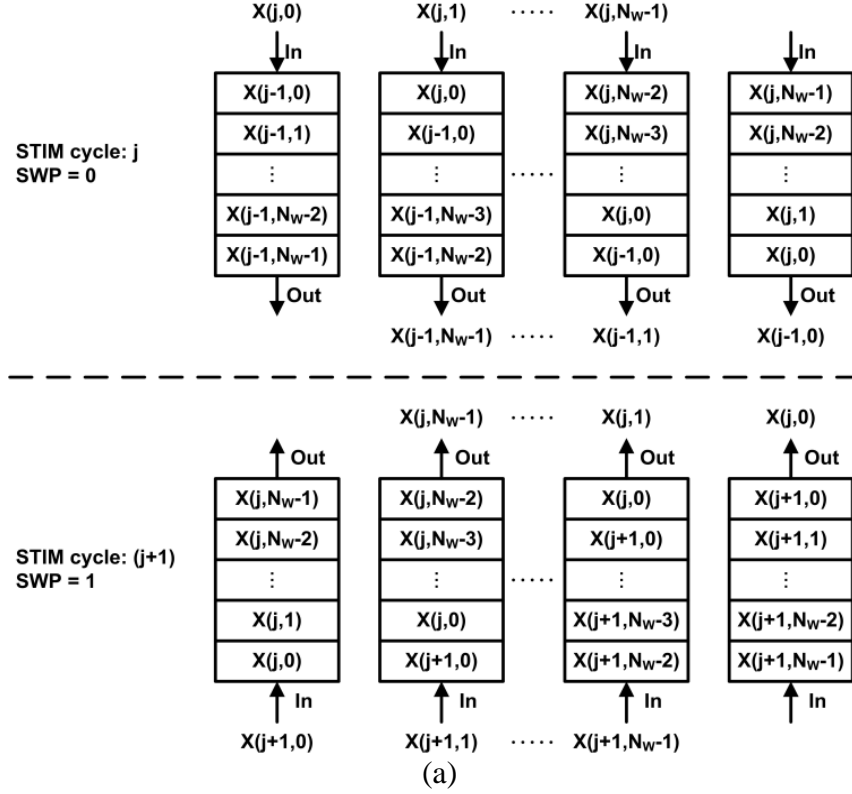


Fig. 2.8 (a) Time reversal via LIFO register and (b) its implementation with a two-port SRAM and two binary up/down counters.

Fig. 2.8 (a) illustrates the time reversal of recorded data using a LIFO register [89], where $X(j,n)$ denotes the n -th time index of input data in the j -th stimulation cycle, and N_w is the windowing length. The input data of the j -th stimulation cycle are written sequentially into LIFO from the top port, and the previously stored data in the $(j-1)$ -th cycle are read out from the bottom port in

reverse-time order. To reverse the order of data read-out, the port-swapping control (SWP) of the LIFO register is switched in the next cycle, flipping the write-in and read-out direction of the LIFO. In the $(j+1)$ -th stimulation cycle, data stored in the j -th cycle are read-out from the top port and thus in reverse-time order, and input data of the $(j+1)$ -th stimulation cycle are written into the LIFO via the bottom port. Such time-reversal operation is maintained by flipping SWP of the LIFO for each stimulation cycle. Fig. 2.8 (b) shows the implementation of a LIFO register using a two-port SRAM (one read port and one write port) with single clock signal CLK , and two bidirectional counters $CNTW$ and $CNTR$ for the write and read addresses generation, which are activated by the write-enable (WEN) and read-enable (REN) signals of the SRAM, respectively. The counting direction of counters is controlled by the SWP of the LIFO: when $SWP = 0$, activated counters are initialized to 0 and incremented to (N_W-1) ; when $SWP = 1$, activated counters are initialized to (N_W-1) and decremented to 0. Data access order on SRAM can be reversed via switching of SWP , and the time-reversal operation is thus achieved on the LIFO.

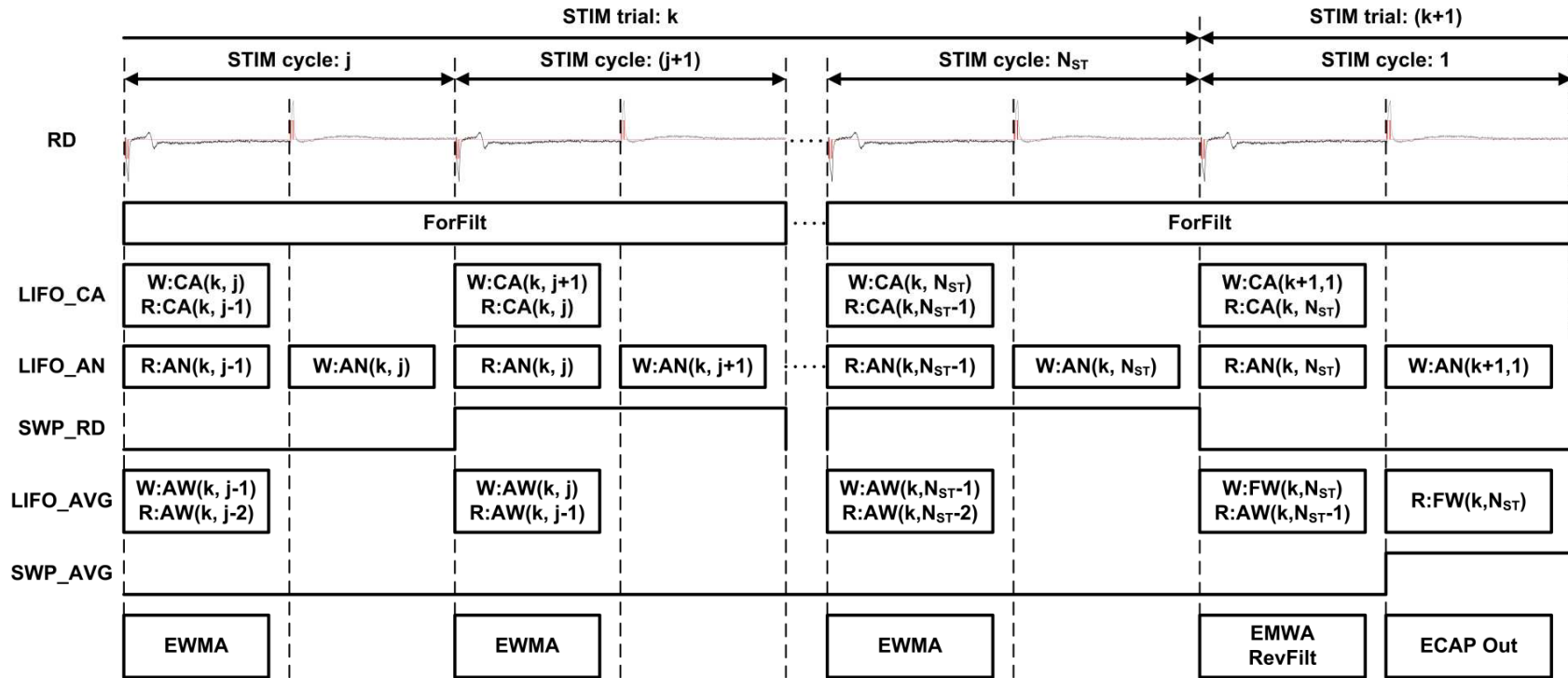


Fig. 2.9 Timing diagram of the BFCA core under continuous neural data input.

Fig. 2.9 shows the timing diagram of the BFCA core for continuous BFCA operation, where (k, j) denotes the j -th AP stimulus cycle in the k -th stimulation trial, and W and R represent data write-in and read-out of a LIFO register, respectively. The windowed cathodal and anodal responses from the *ForFilt* output in the j -th stimulus cycle, $CA(k, j)$ and $AN(k, j)$, are written into the *LIFO_CA* in the cathodal phase and the *LIFO_AN* in the anodal phase, respectively. In the $(j+1)$ -th AP stimulus cycle, when $CA(k, j+1)$ and $AN(k, j+1)$ are written into the *LIFO_CA* and the *LIFO_AN*, respectively, both the previously stored responses, $CA(k, j)$ and $AN(k, j)$, are read out time-reversely in the cathodal phase by flipping *SWP_RD*, the port-swapping control of both the *LIFO_CA* and *LIFO_AN*. Within the cathodal phase of each stimulus cycle, the summed wave $SW(k, j)$ is calculated from $CA(k, j)$ and $AN(k, j)$, and averaged with the $AW(k, j-1)$ read out from the *LIFO_AVG* using EWMA. Meanwhile, the averaged wave $AW(k, j)$ is written into the *LIFO_AVG* time-reversely. In the first cathodal stimulus cycle of the $(k+1)$ -th stimulation trial, the averaging of all SW s in the k -th stimulation trial is completed, which's represented as $AW(k, N_{ST})$. The $AW(k, N_{ST})$ is filtered time-reversely with the *RevFilt*, and the filtered wave $FW(k, N_{ST})$ is stored back into the *LIFO_AVG*. In the 1st anodal stimulus cycle, the ECAP response of the k -th stimulation trial is outputted in continuous-time order by reading out $FW(k, N_{ST})$ from the *LIFO_AVG* with its port-swapping control *SWP_AVG* flipped. While data stored in the k -th stimulation trial are read out from LIFO registers during the 1st AP stimulus cycle, the continuously filtered raw data of the $(k+1)$ -th stimulation trial, $CA(k+1, 1)$ and $AN(k+1, 1)$, are also windowed and written into the *LIFO_CA* and *LIFO_AN*, respectively. Therefore, no halting of raw data stream is required by the BFCA core, which enables the ECAP response of each stimulation trial to be calculated continuously.

2.3.4 Exponentially-Weighted Moving Average

Let $CA(j, n)$ and $AN(j, n)$ denote the windowed cathodal and anodal responses in the j -th AP stimulus cycle, respectively. The summed wave (SW) of the j -th stimulation cycle is calculated by

$$SW(j, n) = (CA(j, n) - DC_{CA}) + (AN(j, n) - DC_{AN}), \quad (2.11)$$

$$AW(j, n) = \begin{cases} SW(j, n), & j = 1 \\ (1 - K_{EWA}) \cdot AW(j - 1, n) + K_{EWA} \cdot SW(j, n), & j > 1 \end{cases} \quad (2.13)$$

where $AW(j, n)$, the averaged wave as described earlier, is the EWMA of SW s from the first j AP stimulus cycles of a stimulus train. The uCAP response is thus the EMWA of SW s from all AP stimulus cycles, i.e., $uCAP(n) = AW(N_{ST}, n)$. The weighting coefficient of EWMA in (2.13), K_{EWA} , is adjusted to the number of AP stimulus cycles (N_{ST}) and calculated by

$$\begin{aligned} N_{RS} &= \lfloor \log_2 N_{ST} \rfloor, \\ K_{EWA} &= 2^{-N_{RS}}. \end{aligned} \quad (2.14)$$

It's worth mentioning that the number of right-shifting (N_{RS}) in (2.14) is equivalent to the number of bits after leading-one in the binary expression of N_{ST} and thus can be easily calculated. For example, an N_{ST} equal to 20, whose binary expression is "10100", will give $N_{RS} = 4$. Since the value of K_{EWA} is a power of 2, the multiplication by K_{EWA} in (2.13) is done by right shifting. The computation of EWMA requires only right-shifting and addition and hence can be implemented on the hardware efficiently. Fig. 2.10 (a) illustrates the schematic of the EWMA calculator. Both the SW calculated with (2.11) and the AW of the previous AP stimulus cycle which is read out from *LIFO_AVG* register are right-shifted by N_{RS} derived from the lead-one detector, and summed to form an updated AW . Fig. 2.10 (b) shows the combinational logic implementation of the lead-one detector. The binary expression of N_{ST} in 16-bit precisions is bit-reversed (*Rev*), inverted, and incremented by 1 (*Inc*). The number of bits after leading-one of N_{ST} (*POS*) is derived via exclusive-OR (XOR) of *Rev* and *Inc*, which is expressed in thermal code, and the number of right-shifting (N_{RS}) can be obtained via the bit-summing of *POS*.

2.3.5 Configurable Folded IIR filter Design

Both the forward filter and reverse filters in the BFCA core adopt an eight-order IIR filter, which is implemented with 4 cascaded bi-quadratic (biquad) IIR filters [91]. The difference equation of a direct-form II [66] biquad IIR filter stage is written as

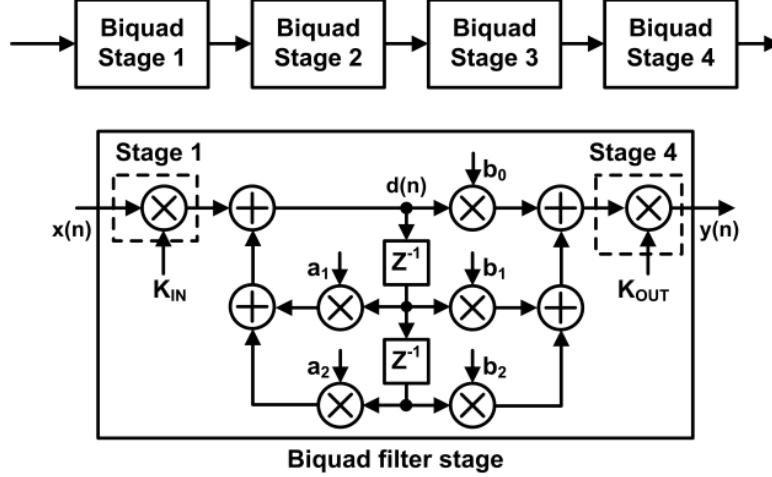


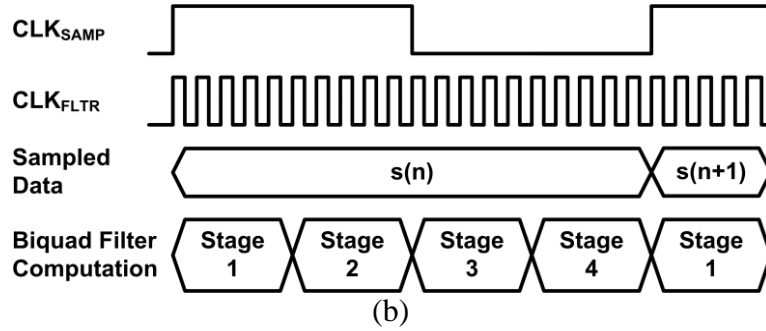
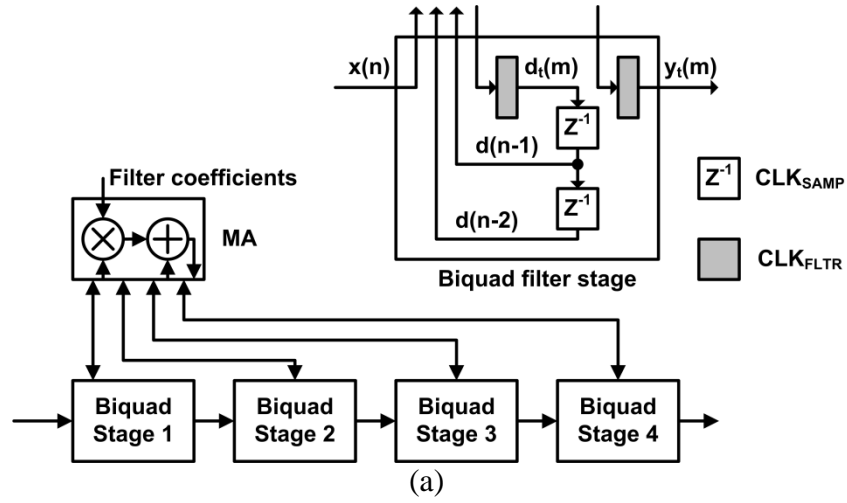
Fig. 2.11 A conventional configurable 4-stage IIR filter

$$\begin{cases} d(n) = x(n) - \sum_{k=1}^2 a_k d(n-k) \\ y(n) = \sum_{k=0}^2 b_k d(n-k) \end{cases}, \quad (2.15)$$

where $x(n)$ and $y(n)$ are the input and output signals, respectively, and a_k and b_k are the filter coefficients. Fig. 2.11 shows the conventional implementation of an eight-order IIR filter with 4 cascaded biquad filter stages, each of which requires 5 multiplications and 4 additions. The input and output of the eight-order IIR filter are multiplied by the input-scaling constant (K_{IN}) and output-scaling constant (K_{OUT}), respectively, to save the internal word-length of biquad filter stage while avoid overflow. The eight-order IIR filter in Fig. 2.11 requires 22 multiplications and 16 additions, which is excessive and not feasible for its hardware implementation on the FPGA.

To further save the implementation cost, folding technique can be utilized for minimizing the number of arithmetic units in IIR filters [92], and several folded IIR filter architectures have been proposed [93, 94]. In our design, a folded direct-form II IIR filter architecture modified from [94] is presented, as shown in Fig. 2.12 (a). A single multiply-add (MA) unit controlled by a faster clock CLK_{FILT} is shared by 4 biquad filter stages, and one MA operation is performed per CLK_{FILT} cycle. For each biquad filter stage, the $d(n)$ and $y(n)$ defined in (2.15) are computed in multiple CLK_{FILT} cycles, and the temporary outcomes of $d(n)$ and $y(n)$ from the MA unit are accumulated in two registers d_i and y_i , respectively, which are also clocked by the CLK_{FILT} . Delay elements of each biquad filter stage store the computed $d(n)$ of the previous sampling

clock (CLK_{SAMP}) cycles, namely $d(n-1)$ and $d(n-2)$. Fig. 2.12 (b) shows the timing diagram of the folded eight-order IIR filter architecture. The input sample $s(n)$ and delay elements of each biquad filter stage are first updated at the rising edge of the sampling clock CLK_{SAMP} , and filtered data are computed by 4 cascaded biquad filter stages at the rate of CLK_{FILT} . The MA operation of a biquad filter stage for computing $d(n)$ and $y(n)$ using (2.15) is shown in Fig. 2.12 (c), where $d_i(m)$ and $y_i(m)$ denote the value of accumulation registers d_i and y_i in the m -th CLK_{FILT} cycle, respectively. In the first CLK_{FILT} cycle of each biquad filter stage, the input of a biquad filter stage $x(n)$ is stored into the d_i register. Note that in the first biquad filter stage, a MA operation is required in the first CLK_{FILT} cycle to multiply input sample $s(n)$ by K_{IN} . The $d(n)$ and $y(n)$ of a biquad filter stage are calculated following the order specified in Fig. 2.12 (c), and the resulting $d(n)$ and $y(n)$ are latched in accumulation registers d_i and y_i , respectively, throughout the rest of CLK_{FILT} cycles. When the computation of $y(n)$ is completed, the next biquad filter stage will be started, where the received $y(n)$ from previous stage is directly stored into the d_i register and the calculation of $d(n)$ and $y(n)$ are followed. In the forth biquad filter stage, an extra MA operation is required to multiply K_{OUT} with calculated $y(n)$ stored in the y_i register. The rate of the CLK_{FILT} is 32 times faster than that of CLK_{SAMP} so that computation of all biquad filter stages are finished within a CLK_{SAMP} cycle. When this folded IIR filter is disabled, the input sample $s(n)$ is directly passed to the filter output without any MA operation.



CLK _{FLTR} Cycle	Shared MA Operation
1	$d_t(1) = s(n) * K_{IN}$ (Stage 1) $d_t(1) = x(n)$
2	$d_t(2) = d_t(1) - d(n-1) * a_2$
3	$d_t(3) = d_t(2) - d(n-2) * a_3$ $d(n) = d_t(3)$
4	$y_t(4) = d(n) * b_1$
5	$y_t(5) = y_t(4) + d(n-1) * b_2$
6	$y_t(6) = y_t(5) + d(n-2) * b_3$ $y(n) = y_t(6)$
7 (Stage 4)	$y(n) = y_t(6) * K_{OUT}$

(c)

Fig. 2.12 (a) The proposed configurable, folded 4-stage IIR filter with shared multiply-add (MA unit), (b) its timing diagram, and (c) shared MA operation of a biquad filter stage.

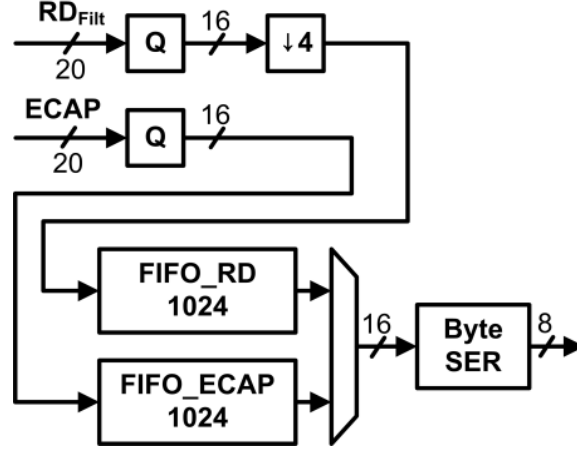


Fig. 2.13 Schematic of the output buffer.

2.3.6 Output Buffer

Fig. 2.13 shows the schematic of the output buffer. Both the RD_{Filt} and computed ECAP responses are quantized to 16 bits by preserving their most significant parts (*i.e.*, the most significant 16 bits). In the first and last AP stimulation cycles, the windowed cathodal and anodal parts of the RD_{Filt} are down-sampled by 4 and stored into a first-in-first-out (FIFO) register $FIFO_RD$ with word-depth of 1024, which provides to users the information on both the settling of stimulus train and the balance of stimulus artifacts in recorded neural data. The computed ECAP responses from the BFCA core is directly stored into the 1024-word FIFO register $FIFO_ECAP$. The windowed RD_{Filt} and ECAP responses are selectively read out at the end of each stimulation trial according to user's configuration and serialized into byte streams for serial data transmission using the UART interface. The FIFO registers can be easily implemented with two-port SRAM [95].

2.4 Experiment Results

2.4.1 Hardware Implementations

The proposed DSP architecture in Fig. 2.4 is mapped to the Microsemi IGLOO2 FPGA (M2GL025) on the FUTUREM2GL-EVB evaluation board. This FPGA was programmed using Verilog and Microsemi Libero-SoC development software. The mapped architecture requires 6936 (25.04%) logic elements (LEs), 3035 (10.96%) D flip-flops (DFFs), 8 (25.81%) large

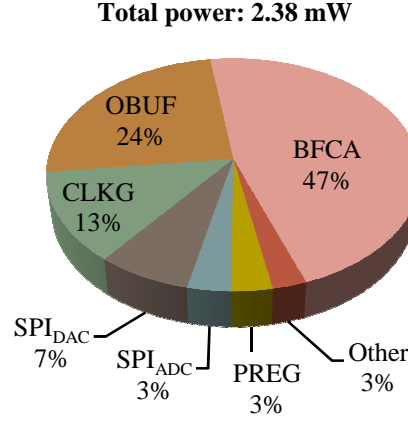


Fig. 2.14 Power consumption of the DSP architecture in 180-nm CMOS process.

SRAMs (LSRAMs), each with size of 18×1024 bits, and 8 (23.53%) MACC units, each of which contains an 18×18 bits multiplier. Synthesized in 180-nm CMOS process, the DSP architecture occupies 0.97-mm^2 silicon area which's dominated by on-chip SRAMs in the BFCA core and the output buffer, and the total power consumption is 2.38 mW at 16-MHz system clock rate and 1.62-V core voltage, whose detail is plotted in 0. Low-power design techniques such as clock gating can be applied to further reduce power consumption.

Fig. 2.15 shows the circuit schematic of the stimulation and recording AFE interfacing with the DSP architecture. The DAC8832, a 16-bit precision, micro-power, SPI-compatible serial interface digital-to-analog converter from *Texas Instruments*, is used to convert the digitized stimulus waveform into a voltage output. With the on-chip matched bipolar offset resistors, the DAC8832 can be configured to provide a bipolar voltage output for the AP stimulus pulse generation by connecting an external operational amplifier to its dedicated pins [96], as pointed out in Fig. 2.15. The DAC8832 can also be reset to a mid-scale code which corresponds to 0-V voltage output in bipolar mode. The OPA191 from *Texas Instruments* is chosen as the external operational amplifier for DAC8832 owing to its high precision ($\pm 5\text{-}\mu\text{V}$ offset voltage and $\pm 5\text{-pA}$ input bias current), wide gain bandwidth (2.5 MHz), low quiescent current (140 μA) and wide supply range [97]. A Howland current pump is employed and implemented with a LT6375 voltage-difference amplifier from *Analog Devices* [55]. A DC-blocking capacitor is used in the current pump to avoid direct current injections into nerves, and a resistor trimmer ranging from 0

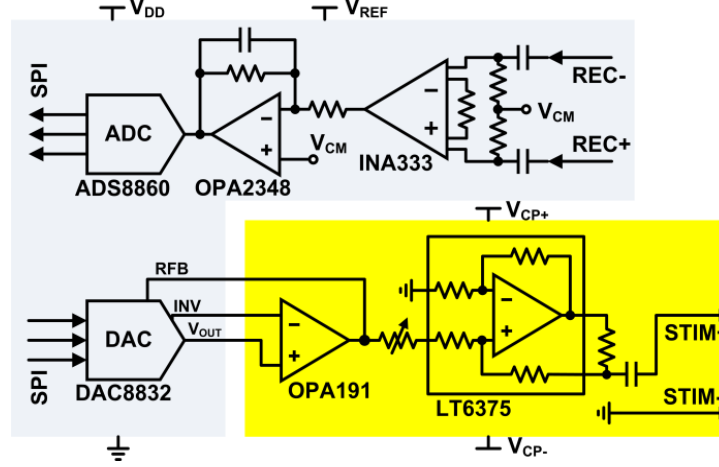


Fig. 2.15 Circuit schematic of the stimulation and recording AFE.

to 1k Ohm (SMUA102PET, *Ohmite*) is added before the LT6375 to balance the resistor network, which helps the current pump achieve high common-mode rejection ratio (CMRR) and high output impedance [98]. The current pump is properly designed to deliver up to 1.5-mA stimulus current, and based on the analysis in [55], the charge-balance error of this current pump is less than 0.3%, which's acceptable for generating symmetric stimulus pulses in the AP stimulation. Neural signals are differentially recorded with a capacitively-coupled precision instrumentation amplifier (INA333, *Texas Instruments*) and conditioned by an active filter constructed with a micro-power, low-noise operational amplifier (OPA2348, *Texas Instruments*). The recording front end has a total gain of 500, and a bandwidth from 1.6 Hz to 20 kHz, much wider than that of ECAP responses to prevent the active filter from distorting the ECAP waveforms in recorded neural signals. Recorded neural signals are digitized with the ADS8860, a 16-bit precision, micro-power, SPI-compatible serial interface analog-to-digital converter from *Texas Instruments*. The supply voltage of both the LT6375 and OPA191 (V_{CP}) is ± 10 V, providing sufficient headroom for the output voltage swing of the current pump. The analog supply voltage (V_{REF}) and common-mode voltage (V_{CM}) of two amplifiers are 3.0 V and 1.5 V, respectively, and the digital supply voltage (V_{DD}) of the ADC and DAC is 3.3 V.

2.4.2 In-Vivo Test Results

Fig. 2.16 illustrates the setup of *in-vivo* electrical nerve stimulations for verification of the proposed DSP architecture. Following the surgery procedure described in [35], the stimulation

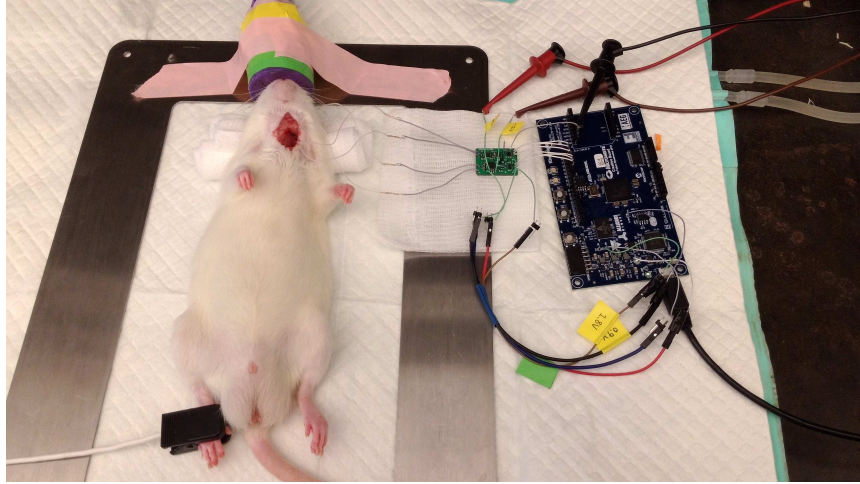


Fig. 2.16 Setup of *in-vivo* electrical nerve stimulation for verification of proposed DSP architecture.

and recording electrodes depicted in Fig. 2.2 (b), both of which are made of silicone cuff electrodes, are attached to the left cervical vagus nerve of a male rat. These two electrodes are connected to the differential stimulation and recording channels of the AFE, whose supply voltages are generated with external power supplies. The programmed FPGA evaluation board is powered at 5 V and controlled by the host PC via an USB-UART bridge on the board. Output data from the DSP architecture are plotted with MATLAB R2016a software. A band-pass elliptic filter with 0.2-3 kHz pass-band, 20-dB stop-band attenuation and 0.1-dB passband ripples is adopted in the proposed BFCA method, whose filter coefficients are derived with the Filter Designer in MATLAB.

A conduction distance of 8 mm is measured after the implant of electrodes, and a series of stimulation trials with varying amplitude are applied to the nerve, whose parameters are listed as follows: $PW = 0.2$ ms, $PRF = 20$ Hz and $t_{train} = 1$ s. The stimulus current amplitude is limited to 0.5 mA to avoid amplifier saturation during the recording of stimulus artifacts. Fig. 2.17 plots the windowed raw data of two stimulation trials with stimulus current amplitude of 0.2 mA and 0.4 mA, respectively, where cathodal and anodal stimulus artifacts are symmetric before and after filtering. The corresponding linear-phase (LP) filtered ECAP responses are plotted against unfiltered ECAP responses in Fig. 2.17 (b). Clearly, linear-phase filtering of ECAP responses using the proposed BFCA method can effectively reduce periodic noise interferences and

preserve the waveform of ECAP responses, especially the amplitude and latency of peaks on ECAP waveforms representing the activation level of certain nerve fiber groups [37].

To prove the validity of measured ECAP responses, a series of stimulation trials are applied to the nerve by varying stimulus amplitude from 0 mA to 0.5 mA with 0.05-mA increment. Fig. 2.18 (a) plots the linear-phase filtered ECAP responses computed by the FPGA against stimulus amplitude, where twenty ECAP responses are collected per stimulus amplitude. It can be seen that consistent ECAP waveforms are measured under the same stimulus amplitude, and that the responses of activated nerve fiber groups, distinguished by positive and negative peaks with constant latency and amplitude proportional to the applied stimulus strength, are also visible on measured ECAP waveforms. Fig. 2.18 (b) plots the amplitude growth function (i.e. peak-to-peak amplitude versus stimulus strength) of fiber responses marked in Fig. 2.18 (a). The 3 peak groups on ECAP waveforms are classified into $A\gamma$, $A\delta$ and C fiber groups, respectively, based on their conduction velocity defined in [35]. Note that the response amplitude of C fiber drops slightly after 0.3-mA stimulus amplitude, which may result from its low positive peak amplitude at its maximal activation level.

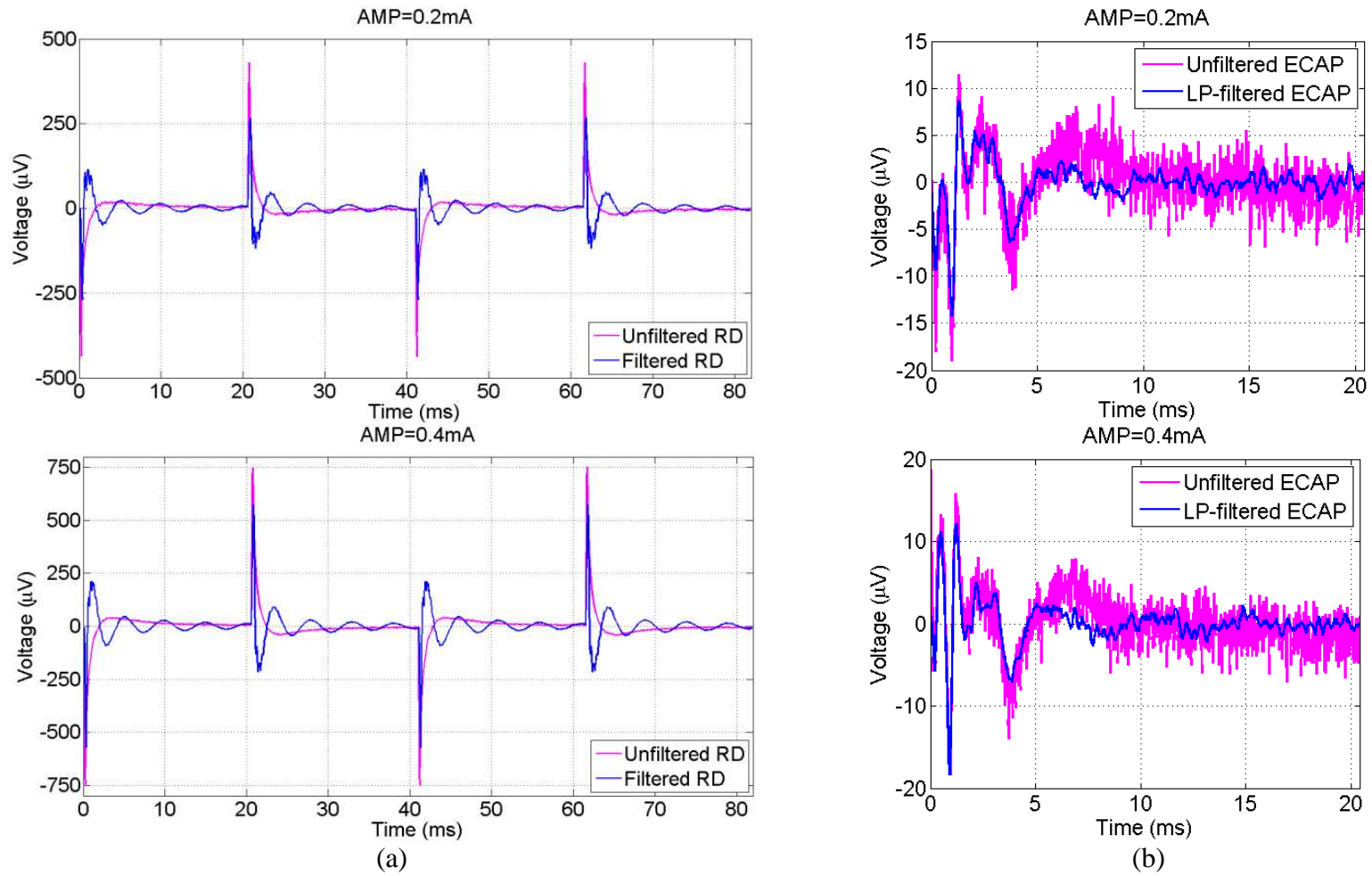


Fig. 2.17 FPGA measurement results from stimulation trials with stimulus parameters $PW = 0.2$ ms, $PRF = 20$ Hz and $t_{train} = 1$ s: (a) windowed raw data and (b) computed ECAP responses of two stimulation trials with stimulus current amplitude of 0.2 mA and 0.4 mA, respectively.

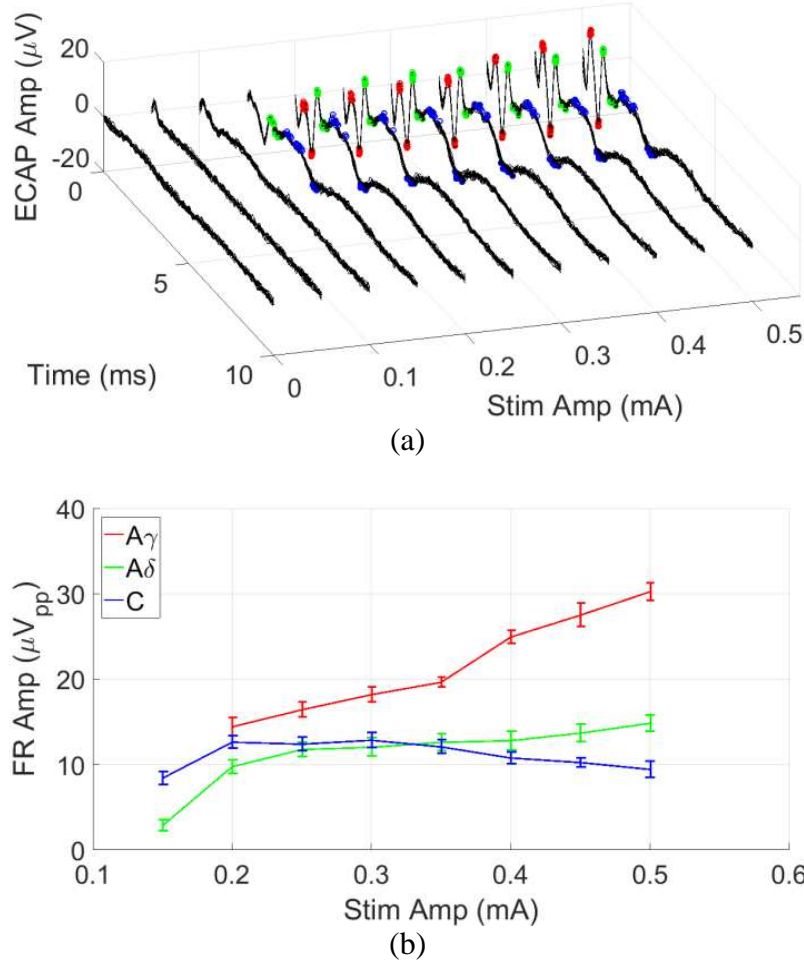
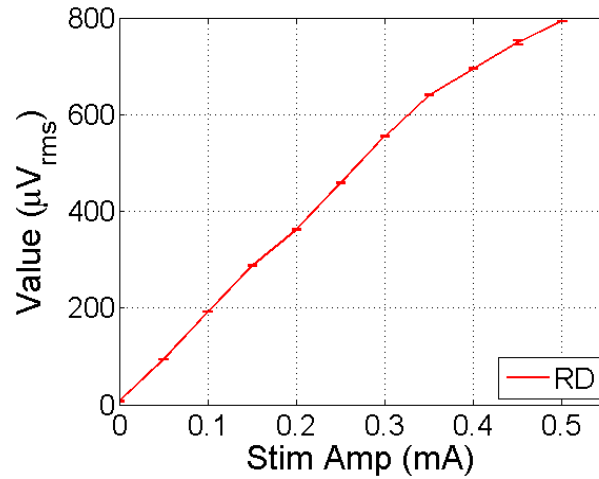


Fig. 2.18 (a) FPGA measurement results of linear-phase filtered ECAP responses collected from stimulation trials with stimulus amplitude varying from 0 to 0.5 mA. (b) Amplitude growth function of nerve fiber responses designated by peaks on ECAP responses.

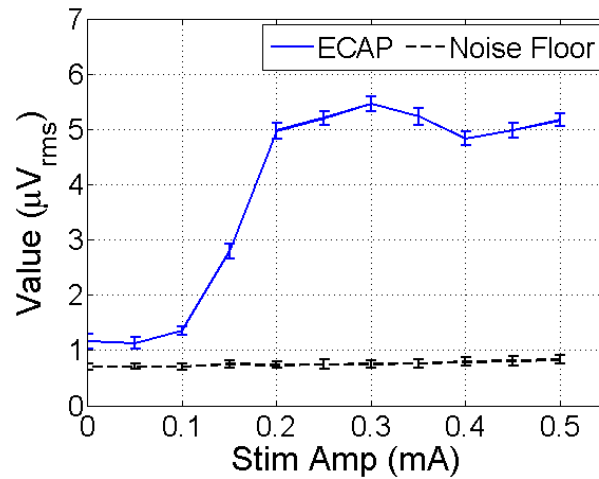
2.4.3 Efficacy Analysis

The efficacy of the AP stimulation method for stimulus artifact rejection in the DSP architecture is demonstrated by comparing the amplitude of stimulus artifacts in raw data and ECAP responses from a total of 220 stimulation trials in Fig. 2.18. In our analysis, the stimulus artifacts are defined as the waveforms within a 10-ms window starting from the onset of stimulus pulses, as seen in Fig. 2.17 (a), and the root-mean-square (rms) values of stimulus artifacts are calculated for both raw data and ECAP responses. Fig. 2.19 (a) and (b) plot the mean rms values of stimulus artifacts in raw data and ECAP responses from the FPGA, respectively, under different

stimulus amplitudes. Note that the mean rms value of the noise floor and its standard deviation in Fig. 2.19 (b) is obtained from a 10-ms segment on ECAP waveforms containing no stimulus artifacts and nerve fiber responses. While the stimulus artifacts in raw data grow proportionally with the applied stimulus amplitude as plotted in Fig. 2.19 (a), the rms values of the stimulus artifact on recovered ECAP responses are approximately 1.6 times higher than that of the noise floor for stimulus amplitude below 0.15 mA. As the stimulus amplitude is increased above 0.15 mA, where nerve fiber responses are visible on ECAP waveforms as seen in Fig. 2.18 (a), the rms values of stimulus artifact on ECAP responses are of the same order of the amplitude of nerve fiber responses plotted in Fig. 2.18 (b). This implies that the stimulus artifacts overlapped with ECAPs as seen in Fig. 2.17 (a) are successfully removed. The rms value of stimulus artifact on measured ECAP responses is reduced by a factor of 115 on average.



(a)



(b)

Fig. 2.19 Root-mean-square (rms) value of stimulus artifacts in (a) raw data and (b) ECAP responses measured from FPGA. The rms value of noise floor in (b) is obtained from a 10-ms segment on each ECAP waveform containing no stimulus artifacts or nerve fiber responses.

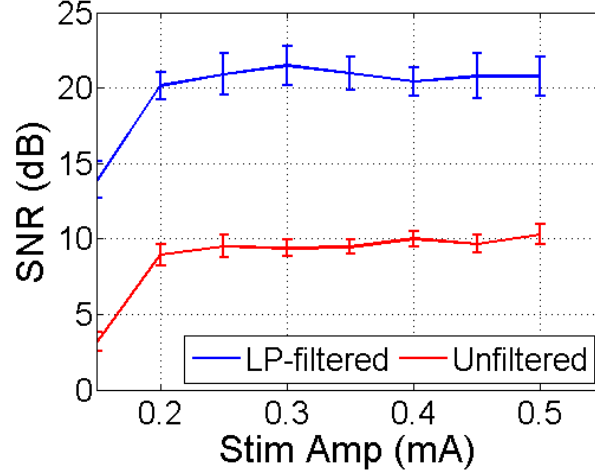


Fig. 2.20 Signal-to-noise ratio (SNR) of linear-phase filtered versus unfiltered ECAP responses measured from FPGA.

The performance of linear-phase filtering using the proposed BFCA method is quantified with the signal-to-noise ratio (SNR) improvement and waveform distortion in ECAP responses. In SNR analysis, only the ECAP waveforms from the FPGA containing recognizable nerve fiber responses are considered. The SNR of ECAP waveforms is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{FR_{rms}}{NF_{rms}} \right), \quad (2.16)$$

where FR_{rms} and NF_{rms} are the rms voltage of nerve fiber responses and noise floor, respectively. The value of FR_{rms} and NF_{rms} are calculated from the first 5-ms interval of ECAP waveforms containing nerve fiber responses and a 10-ms segment on ECAP waveforms containing background noises only, respectively. Fig. 2.20 plots the mean SNR of unfiltered and linear-phase filtered ECAP responses from the FPGA under different stimulus amplitude. For stimulus amplitude above 0.2 mA, the SNR of linear-phase filtered ECAPs is 20.8 dB on average, whereas that of unfiltered ECAPs is 9.6 dB; this facilitates the discrimination of nerve fiber responses on ECAP waveforms. In waveform distortion analysis, ECAP responses are recovered from a set of offline-recorded neural data in [35] superimposed by 60-Hz sine waves with -6-dB SNR and random phase as periodic noises; the SNR of sine waves can be calculated with (2.16), where the fiber responses are defined as the first 6-ms interval of the noise-free ECAP waveform

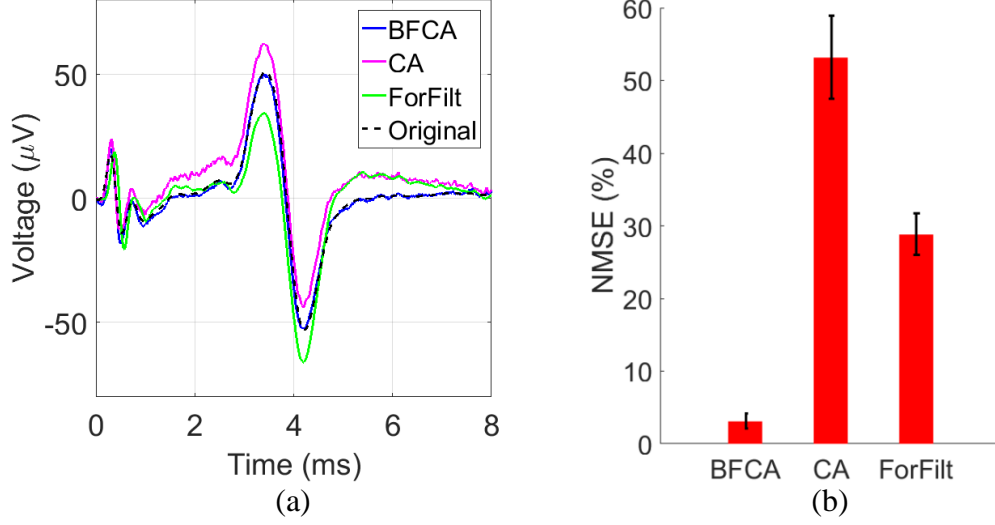


Fig. 2.21 Waveform distortion caused by forward filtering (ForFilt), coherent averaging (CA) and linear-phase filtering via BFCA: (a) demonstration and (b) a quantitative comparison using normalized mean-square error (NMSE) between filtered and original noise-free ECAP waveforms.

in Fig. 2.21 (a). An example of the original noise-free ECAP waveform versus the ECAP waveforms after the forward filtering (i.e., continuous filtering of raw data before coherent averaging only), coherent averaging, and linear-phase filtering via the BFCA is given in Fig. 2.21 (a). The level of waveform distortion after filtering is quantified with the normalized mean-square error (NMSE) between the original and filtered ECAP waveforms, which is defined as

$$\text{NMSE} = \frac{\sum_n (x_n - y_n)^2}{\sum_n (y_n - \bar{y})^2}, \quad (2.17)$$

where x_n and y_n denote the samples of filtered and original ECAP waveform, respectively. The \bar{y} in (2.17) denotes the DC offset of the original ECAP waveform which is calculated by

$$\bar{y} = \frac{1}{N} \sum_n y_n, \quad (2.18)$$

where N is the number of samples in ECAP waveforms and is equal to the windowing length N_{win} . Fig. 2.21 (b) plots the mean NMSE of ECAP waveforms obtained with the forward filtering, coherent averaging and BFCA in 66 trials. The NMSE of coherent averaging is up to 53.1%,

Table 2.1 Comparison with other filtering techniques

	[80]	[35]	This work
Signal	Neuron Spike	ECAP	ECAP
Filtering Technique	Wavelet Filtering	Coherent Averaging	BFCA
Frequency Selectivity	No	No	Yes
SNR (dB)	9*	9.6	20.8
Waveform Distortion (%)	4*	53.1	3.1

* The best cases of reported SNR and waveform distortion are listed here.

which results from its deficiency in removing periodic noises. The forward filtering removes periodic noises better, but its NMSE is still 28.8% due to the nonlinear phase response of IIR filters. The NMSE of linear-phase filtering via the proposed BFCA method is only 3.1%. Table 2.1 shows a comparison of this work with other filtering techniques. Compared with the coherent averaging method used in [35], the proposed BFCA method improves SNR by 11 dB and achieves an 3.1% waveform distortion that is $17.1\times$ lower. Beside, with its IIR filters that can have arbitrary frequency response, the BFCA method provides frequency selectivity which is useful in characterizing the high- and low- frequency components of ECAP responses [99]. To our best knowledge, this is the first DSP architecture for programmable and distortion-free filtering of ECAP responses in real-time.

2.5 Conclusion of This Chapter

This chapter presented the first DSP architecture for real-time recovery of ECAP responses from stimulus artifacts and periodic noises in bidirectional NRT systems. A BFCA method was proposed for configurable and distortion-free filtering of ECAPs, and the AP stimulation method that can be combined with the BFCA is utilized for rejecting overlapped stimulus artifacts. Design techniques including the configurable folded IIR filter and EWMA were also presented for hardware- efficient implementation of the DSP architecture. Synthesized in 180-nm CMOS

process, the total area and power consumption of this DSP architecture are 0.97 mm^2 and 2.38 mW , respectively. The proposed DSP architecture was tested in *in-vivo* electrical nerve stimulations to verify its efficacy of recovering ECAPs from overlapped stimulus artifacts and periodic noises, and experiment results showed that compared with the previous coherent averaging technique, the proposed DSP architecture improves the SNR of ECAP responses by 11 dB and achieves an 3.1% waveform distortion that is $17.1\times$ lower. This is the first step to realizing the real-time DSP engine in Fig. 2.2 (b). The principle and VLSI architecture of feature extraction from ECAP waveforms, the complete DSP engine, and its integration with wearable wireless devices will be discussed in the next chapter.

3. FREE: FIBER-RESPONSE EXTRACTION ENGINE ON A CUSTOM-MADE WEARABLE DEVICE FOR AUTONOMOUS NERVE ACTIVATION CONTROL

This chapter continues the work of Chapter 2 and presents FREE (fiber-response extraction engine), the first digital signal processing (DSP) engine dedicated to nerve activation control in closed-loop electrical nerve stimulation (ENS) systems. FREE adopts a newly proposed bidirectional-filtered coherent-averaging (BFCA) method combined with the alternating-polarity (AP) stimulation for stimulus artifact rejection and distortion-free filtering of electrically-evoked compound nerve action potentials (ECAPs) in real-time, and its hardware architecture are illustrated. The algorithms and VLSI implementation of real-time fiber-response extraction, including peak detection on ECAP and fiber-response classification are also explained. A custom-made wearable device powered by a single coin battery is realized in a printed circuit board prototype that integrates the FREE, a low-power wireless transceiver, a stimulation and recording analog front-end, and a power management unit. FREE reduces the data transmission rate of wearable devices to 16.4 kbps for ECAP output and 192 bps for fiber-response output, which are $49\times$ and $4167\times$ lower than that of software processing, respectively. Experimental results show that compared with the previous software-processing techniques, FREE improves the precision of fiber response classification in terms of amplitude precision by up to $3.1\times$ in noisy environments, which boosts the accuracy of nerve activation profiles by up to 62.9%. An application-specific integrated circuit version of FREE implemented in 180-nm CMOS process consumes 1.95-mW core power at 1.8-V supply.

3.1 Introduction

Ever since the U.S. Food and Drug Administration (FDA) approved deep-brain stimulation as a valid treatment for tremor in late 90's [5], electrical neuromodulation becomes an emerging therapeutic for many neurological diseases [4, 100-102]. The main advantage of electrical neuromodulation is its capacity to target and dose a certain nerve and brain area more precisely, which also makes it a popular treatment for nerve disease alternative to pharmaceutical approaches. Nowadays, most commercial neuromodulation systems are configured in open-loop manner [23], where pre-programmed electrical stimuli are delivered to nerve and adjusted after

weeks or months based on patient's subjective experiences. As the drawbacks of open-loop systems gradually appear, including poor efficiency (either too much or too little dosing) and slow reaction to patient's clinical symptoms, which easily causes patient's discomfort and other side effects, efforts have been made in building closed-loop neuromodulation systems which automatically adjust stimulus strength in real-time based on patient's physiological responses and significantly improve the drawbacks of open-loop systems [24, 27, 103].

Electrical nerve stimulation (ENS) is one neuromodulation techniques that has been widely adopted in clinical therapies for pain, epilepsy and depression [14, 15, 104, 105]. The evoked compound action potential (ECAP) is the sum of action potentials from nerve fibers in response to the electrical stimulus that can be recorded on the nerve [49]. It reflects the activity of the nerve being stimulated and is an objective measure of patient's nerve physiology and stimulation efficiency in closed-loop neural stimulations. The most renowned closed-loop ENS system employing ECAP as the feedback physiological signal is the neural response telemetry in cochlear implants [41]. A newly proposed ENS platform, autonomous nerve control (ANC) [35], also utilizes ECAP for stimulation parameter adjustment in closed-loop system: it decodes ECAPs for construction of the patient-specific nerve activation profile (NAP) that describes the relationship between the stimulus strength and the activation level of nerve fibers. By precisely controlling patient's nerve activation based on the derived NAP, ANC helps mitigate patient response variability and maximizes the efficacy of closed-loop ENS.

Traditional closed-loop neuromodulation platforms require wire connections between patient's nervous system and external instruments for neural stimulation and recording. Such tethered cables not only degrade the signal quality and restrict patient's movement, but also introduce the risk of infection and injury due to tension on cables attached to the nervous system. To solve these problems, wearable devices and application-specific integrated circuits (ASICs) capable of simultaneous stimulation and recording have been developed for various closed-loop neuromodulation systems [55, 57, 106, 107], on which neural signals are continuously recorded and wirelessly transmitted to a host personal computer (PC) for post processing. Whereas the radio-frequency (RF) transceiver for data telemetry dominates the power consumption of wearable devices, data transmission rate becomes a major limiting factor for their realization. For example,

a data rate of 1.96 Mbps is required to transmit raw data from 128 channels (1-kS/s sampling rate and 15-bit resolution) [107], and an 800-kbps data rate will be required for raw data transmission in the ANC system [35]. Not only is continuous wireless transmission of raw data at such high rate power-costly, but it's also subject to data loss during transmission that impairs the fidelity of recorded neural signals. Real-time digital signal processing (DSP) techniques have been employed to reduce output data rate of wearable devices or ASICs for closed-loop neuromodulation systems. For instance, [108] reports a fully-integrated neuromodulation system-on-chip (SoC) that operates 64 acquisition channels with digital compression by sending spike events and firing rate, and [109] reports a 128-channel bidirectional closed-loop neural interface system with field-programmable gate array (FPGA) based real-time spike sorting. A general-purpose brain-machine-brain interface (BMBI) is also reported in [64], which incorporates a microcontroller-based digital neural signal analyzer for time and frequency domain feature extraction and compressed sensing of neural signals. Beside electrical neuromodulation, real-time DSP, including spike detection and data compression, has also been applied for combined optogenetics and multi-channel neural recording [110].

Regardless of the progress in ENS and real-time DSP, today's closed-loop ENS systems, especially the newly-proposed ANC platform, still rely on offline processing of continuously transmitted neural signals to remove stimulus artifact [111] and noises from ECAP and extract biomedical features such as nerve fiber responses [37] that serve as references for stimulus strength adjustment. A real-time DSP for artifact and noise rejection and feature extraction of ECAP is hence desirable for reducing the output data rate of wearable devices applied to ANC as well as other closed-loop ENS systems. Furthermore, neural signals recorded with wireless devices are inevitably accompanied with periodic noises including power-line interference and baseline wander, which introduces errors onto extracted biomedical features and degrades the stimulation efficiency. For instance, the amplitude error of nerve fiber responses on the ECAP caused by noises can easily decrease the accuracy of NAP in ANC. The coherent-averaging technique [52], which's been used for noise removal of the ECAP in ENS systems, including ANC, is inefficient in removing periodic noises, especially those time-locked to the stimulus pulse train. It's therefore essential to have an efficient DSP strategy for real-time periodic noise

removal which is also capable of preserving the waveform morphology of the ECAP and the precision of extracted biomedical features.

In this chapter, we present a fiber-response extraction engine (FREE) for real-time artifact and noise rejection and feature extraction of the ECAP. FREE employs the newly proposed bidirectional-filtered coherent-averaging (BFCA) method presented in Chapter 2 for distortion-free filtering of the ECAP, which can be easily combined with the alternating-polarity (AP) stimulation method for stimulus artifact rejection. Nerve fiber responses on recovered ECAP waveforms are identified according to the user-defined response latencies. Resource-optimized architecture design of FREE, including the BFCA and real-time fiber response extraction is also presented. FREE is implemented on a custom-made and coin battery-powered wearable printed circuit board (PCB) integrating a low-power FPGA, a Bluetooth transceiver, a stimulation and recording analog front-end (AFE), and a power management unit (PMU), and tested both offline and *in-vivo*. Compared with the previous software-based ECAP processing in [35], FREE not only reduces the maximum data rate of wearable devices to 16.4 kbps that is at least $49\times$ lower, but also improves the precision of fiber response classification in terms of amplitude precision by up to $3.1\times$ in noisy environments, which boosts the accuracy of NAP construction by up to 62.9%. An ASIC implementation of FREE is demonstrated whose total chip area and core power consumption of 19.98 mm^2 and 1.95 mW, respectively. To our best knowledge, FREE is the first DSP engine designed for ANC platform to facilitate nerve activation control on wearable devices, and can be applied to other closed-loop ENS systems utilizing the ECAP as their feedback biomarker.

The rest of this chapter is organized as follows. Section 3.2 presents an overview of FREE on wireless wearable devices. Section 3.3 presents the architecture design of each building module in FREE. Section 3.4 presents the custom-made wireless wearable device in PCB prototype integrating FREE. Section 3.5 presents offline and *in-vivo* tests for performance comparison between FREE and previous software-based signal processing [35] in terms of amplitude and latency variation of fiber responses and the resulting NAP in noisy environments, and the ASIC implementation of FREE. Finally Section 3.6 draws a conclusion to this work.

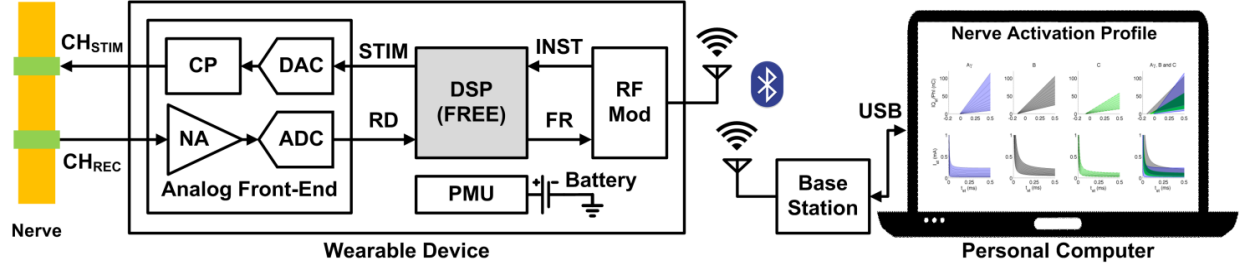


Fig. 3.1 Top-level block diagram of the wireless wearable device in a closed-loop electrical nerve system and the proposed fiber-response extraction engine (FREE), a dedicated DSP engine for autonomous nerve activation control.

3.2 System Overview

Fig. 3.1 presents the top-level block diagram of the wireless wearable device in a closed-loop ENS system and the proposed FREE, a DSP engine dedicated to autonomous nerve activation control. This device comprises the AFE for stimulation and recording of neural signals, the proposed FREE for real-time artifact and noise rejection and feature extraction of ECAPs in digital domain, a RF module in charge of wirelessly receiving instructions from and transmitting biomedical features to the host PC, and the PMU connected to batteries for power supply of abovementioned modules. Digitized stimulus waveforms are first generated from FREE based on decoded instructions, and converted to current stimuli on stimulation electrode by a digital-to-analog converter (DAC) and a current pump (CP) in the AFE for electrical nerve stimulation. The responsive neural signals recorded on the neighboring electrode are conditioned with a neural amplifier (NA) and digitized by an analog-to-digital converter (ADC). On FREE, digitized raw neural data (RD) are processed to recover the ECAP response to the applied stimulus, and nerve fiber responses on the ECAP are extracted as its biomedical features, which are then fed to the RF module and wirelessly sent back to the PC. A base station is connected to the host PC via the USB port for wireless communication between the wearable device and the PC, and Bluetooth standard is adopted for low-power and short-distance wireless communication required by battery-powered wearable devices nowadays. On the host PC, the NAP of fiber groups of interest, which describes the extent of nerve fiber response to given stimulus strength, is derived based on the fiber responses from the wearable device, and stimulus parameters are

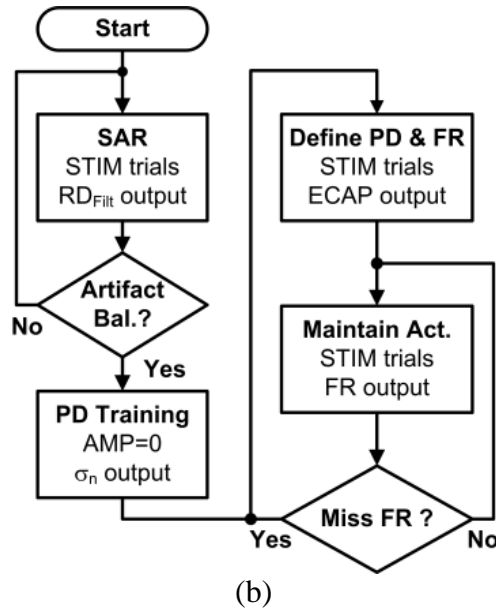
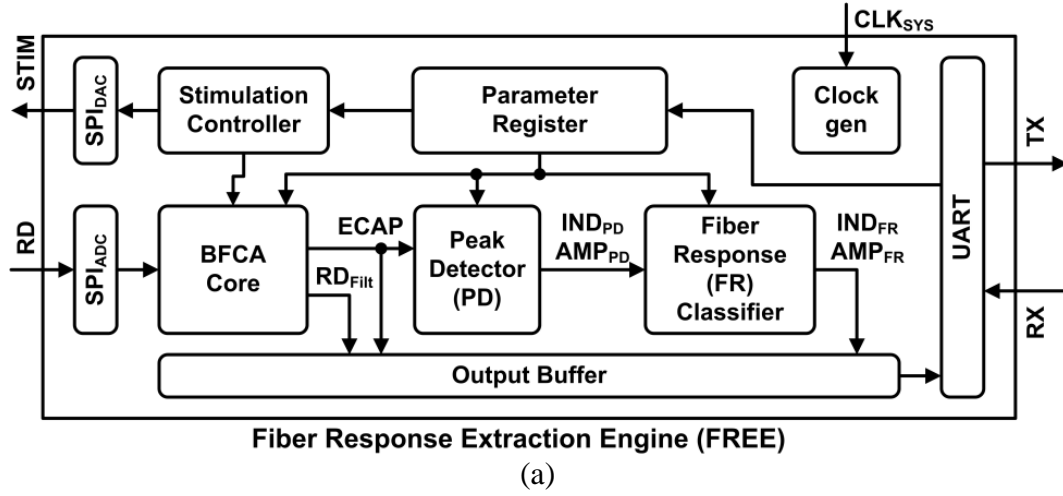


Fig. 3.2 (a) Block diagram of the proposed fiber-response extraction engine (FREE) and (b) the flowchart of its operation.

adjusted according to the predicted NAP and delivered to the wearable device to maintain desired activation level of the fiber group.

Fig. 3.2 (a) shows the overall block diagram of the proposed FREE. Two SPI masters SPI_{DAC} and SPI_{ADC} interface with the DAC and ADC in the AFE for stimulus generation and neural data acquisition, respectively. The stimulation controller generates AP stimulus pulse trains with 1-ns time resolution which is transmitted to the AFE with the SPI_{DAC} . Details on the generation of AP

stimuli and its parameters have been described in Chapter 2. The quantized raw data (RD) from the ADC are sampled at 50 kHz, as is adopted in [35], and both the SPI_{DAC} and SPI_{ADC} are configured to transmit and receive 16-bit digitized data, respectively, same as the precision of the data acquisition board used in [35]. In the BFCA core, the ECAP response to the applied stimulus pulse train is recovered from digitized raw data via stimulus artifact rejection and distortion-free filtering, as described in Chapter 2. Fiber responses, which peak on ECAP waveforms, are detected with the peak detector preset with ripple and noise floor thresholds, and targeted fiber responses with specific conduction velocity, whose peaks are located within a fixed time range on the ECAP waveform [37], are then identified and classified with the fiber-response classifier. The output buffer selectively outputs filtered raw data (RD_{Filt}), recovered ECAP response, or index and amplitude of extracted fiber responses, according to user's configuration, and data are then serially outputted with the UART interface. All system parameters, including the stimulus parameters, filter coefficients, thresholds for peak detection and time indices of targeted fiber responses, are serially loaded from the UART interface and stored into the parameter register.

Fig. 3.2 (b) illustrates the flow chart of the operation of FREE for nerve activation control using the ANC platform. At the beginning stimulation trials RD_{Filt} are outputted first such that users can adjust the stimulus parameters to balance the stimulus artifacts in RD_{Filt} for stimulus artifact rejection (SAR). After stimulus artifact rejection, a stimulation trial with zero stimulus amplitude is first applied for peak detector (PD) training, in which noise floor on recorded neural data and its standard deviation (σ_n) are calculated by the BFCA core and peak detector, respectively, and the computed σ_n is sent to users. A series of stimulation trials with nonzero amplitude are then applied, and the users, based on the outputted ECAP responses, can define the parameters for the peak detector and fiber-response classifier for fiber-response extraction. During the stimulation trials for constructing NAPs and maintaining the activation level of targeted fiber group, only extracted fiber responses will be outputted to minimize the amount of data transmission and the resulting power consumption on the wearable device. ECAP responses can always be retransmitted whenever there's any missing or invalid fiber response (For example, fiber responses fall out of predefined time range or deviate from predicted NAP, both of which can be identified with their index and amplitude sent from FREE).

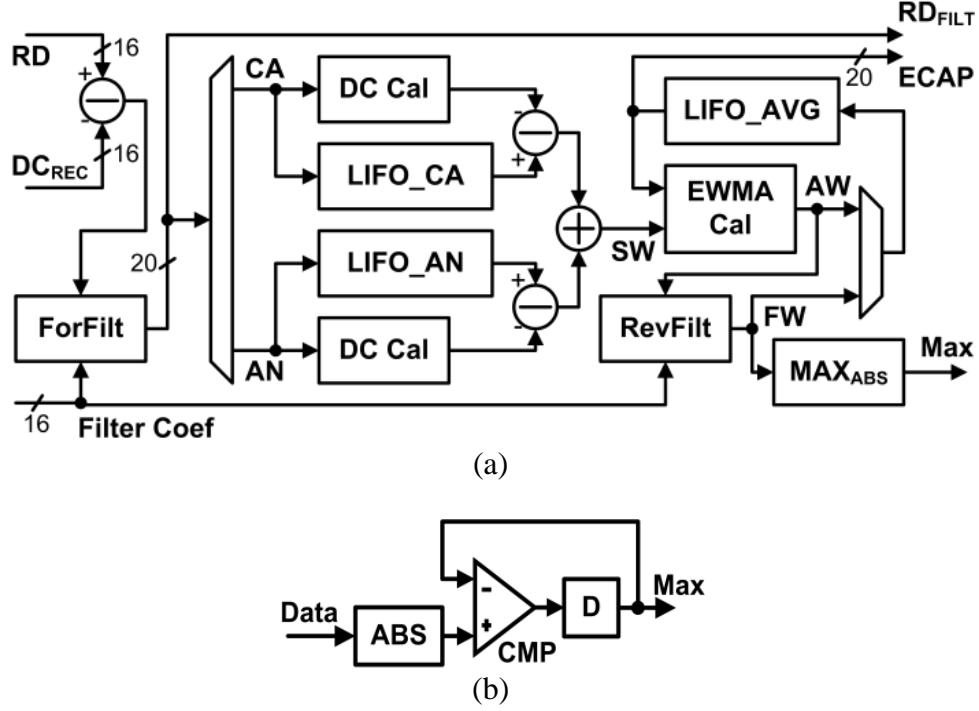


Fig. 3.3 (a) Architecture of the BFCA core in FREE and (b) the schematic of the maximal absolute value detector.

3.3 Architecture Design

3.3.1 BFCA Core in FREE

FREE adopts the bidirectional-filtered coherent-averaging (BFCA) algorithm combined with the AP stimulation method, as presented in Chapter 2, for SAR and distortion-free filtering of ECAP in real-time. Fig. 3.3 (a) shows the architecture of the BFCA core in FREE, where the hardware implementation of the division-free exponentially-weighted moving average (EWMA) calculator, the resource-sharing biquad IIR filter used for both forward and reverse filters, and the last-in-first-out (LIFO) register are the same as those described in Chapter 2. In BFCA core the time-locked windowing of raw-data is controlled by the stimulation controller to ensure the synchronization between the stimulus pulse and windowing and hence the alignment of cathodal and anodal stimulus artifacts [35]. In FREE the windowing length (N_{win}) is also programmable from 256 to 1024, allowing recovered ECAP responses with duration from 5.12 ms to 20.48 ms to be stored at 50-kHz sampling frequency.

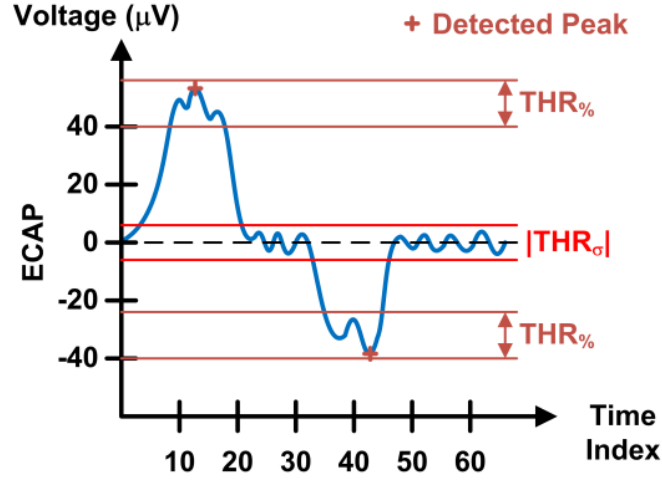


Fig. 3.4 Illustration of the peak-detection principle.

It should be noticed that a maximal absolute value detector (MAX_{ABS}), whose schematic is shown in Fig. 3.3 (b), is integrated into the BFCA core in FREE for the detection of maximal absolute value of an ECAP response that will be used in the peak detector. When the output of the reverse filter, which's the recovered ECAP response in time-reversal order as explained in Chapter 2, is written into the $LIFO_AVG$ register, it's also fed into the MAX_{ABS} . The absolute value of the first sample is directly stored into the register. In the following clock cycles the previously stored data in the MAX_{ABS} is compared with the absolute value of incoming samples, and replaced when the absolute value of incoming sample is larger. At the end of reverse filter output, the value stored in MAX_{ABS} will be the maximal absolute value of recovered ECAP response.

3.3.2 Peak Detector

As stated in [35], an ECAP response contains separate peaks on time axis that are contributed by activated fiber groups with different conduction velocity. A peak detector is hence required in FREE to identify possible fiber responses and should be able to distinguish fiber responses from background noises. Fig. 3.4 illustrates the principle of peak detection in FREE based on amplitude thresholding [112]. The amplitude of noise threshold (THR_{σ}) is estimated based on the standard deviation of noise floor on ECAP responses (σ_n) multiplied with an empirical constant

C , namely, $THR_{\sigma} = C \cdot \sigma_n$. Using the same formula in [113], the standard deviation of noise floor σ_n is calculated by

$$\sigma_n = \sqrt{\left(\sum_{i=0}^{N_{win}-1} x_i^2 - \left(\sum_{i=0}^{N_{win}-1} x_i\right)^2 \cdot \frac{1}{N_{win}}\right) \cdot \frac{1}{N_{win}}}, \quad (3.1)$$

where x_i is the i -th sample of the ECAP response from the BFCA core and N_{win} is the windowing length described earlier. The N_{win} is a power of 2 such that the division by N_{win} can be implemented with a right-shifting operation. It should be mentioned that the optimal value of C that best estimates the amplitude of activated fiber responses in ECAPs is still being studied today, which requires not only offline statistic analysis on pre-recorded ECAP responses but also relies on subjective decisions from physiologists based on their clinical experiences [114-116]. Nevertheless, amplitude thresholding and its hardware implementation are still useful for future study of real-time data compression [110, 117] of ECAP waveforms and hence will be described here. To better locate the local maxima and minima on ECAP waveforms at the presence of the rippling resulted from residual random noises after filtering and the quantization noise in fixed-point arithmetic, a percentage threshold ($THR_{\%}$) is also preset such that only the peaks with amplitude variation greater than $THR_{\%}$ will be detected. This percentage threshold ($THR_{\%}$) is given by $THR_{\%} = Max \cdot \%$, where Max is the maximal absolute value of ECAP response calculated by the BFCA core, and $\%$ is the user-defined percentage of amplitude variation ranging from 0 to 1.

Table 3.1 Peak Detection Algorithm

Input: $x(i)$, $i = 0, 1, \dots, N_{win} - 1$: i -th sample of the recovered ECAP,
 $THR_{\%}$: percentage threshold, THR_{σ} : noise threshold;
initial $state = INI$, $i_{pos} = 0$, $i_{neg} = 0$, $Ind_{pos} = \emptyset$, $Amp_{pos} = \emptyset$,
 $Ind_{neg} = \emptyset$, $Amp_{neg} = \emptyset$;
for $i = 0, \dots, N_{win} - 1$
 switch ($state$)
 case INI : **if** $(x(i_{pos}) \geq x(i) + THR_{\%})$ $state = POS$;
 else if $(x(i) \geq x(i_{neg}) + THR_{\%})$ $state = NEG$;
 end if
 if $(x(i) \geq x(i_{pos}))$ $i_{pos} = i$;
 else if $(x(i) \leq x(i_{neg}))$ $i_{neg} = i$;
 end if
 break;
 case POS : **if** $(x(i) \geq x(i_{pos}))$ $i_{pos} = i$;
 else if $(x(i_{pos}) \geq x(i) + THR_{\%})$
 if $(x^2(i_{pos}) \geq THR_{\sigma}^2)$
 $Ind_{pos} = Ind_{pos} \cup \{i_{pos}\}$;
 $Amp_{pos} = Amp_{pos} \cup \{x(i_{pos})\}$;
 end if
 $i_{neg} = i$; $state = NEG$;
 end if
 break;
 case NEG : **if** $(x(i) \leq x(i_{neg}))$ $i_{neg} = i$;
 else if $(x(i) \geq x(i_{neg}) + THR_{\%})$
 if $(x^2(i_{neg}) \geq THR_{\sigma}^2)$
 $Ind_{neg} = Ind_{neg} \cup \{i_{neg}\}$;
 $Amp_{neg} = Amp_{neg} \cup \{x(i_{neg})\}$;
 end if
 $i_{pos} = i$; $state = POS$;
 end if
 break;
 end switch
end for
Output: Ind_{pos} , Amp_{pos} , Ind_{neg} , Amp_{neg}

Table 3.1 demonstrates the details of the peak detection algorithm in FREE, where i_{pos} and i_{neg} denote the index of temporarily positive and negative peaks, respectively. Note that $THR_{\sigma}^2 = C^2 \cdot \sigma_n^2$ is computed instead to avoid the square root calculation of σ_n in (3.1), and the square

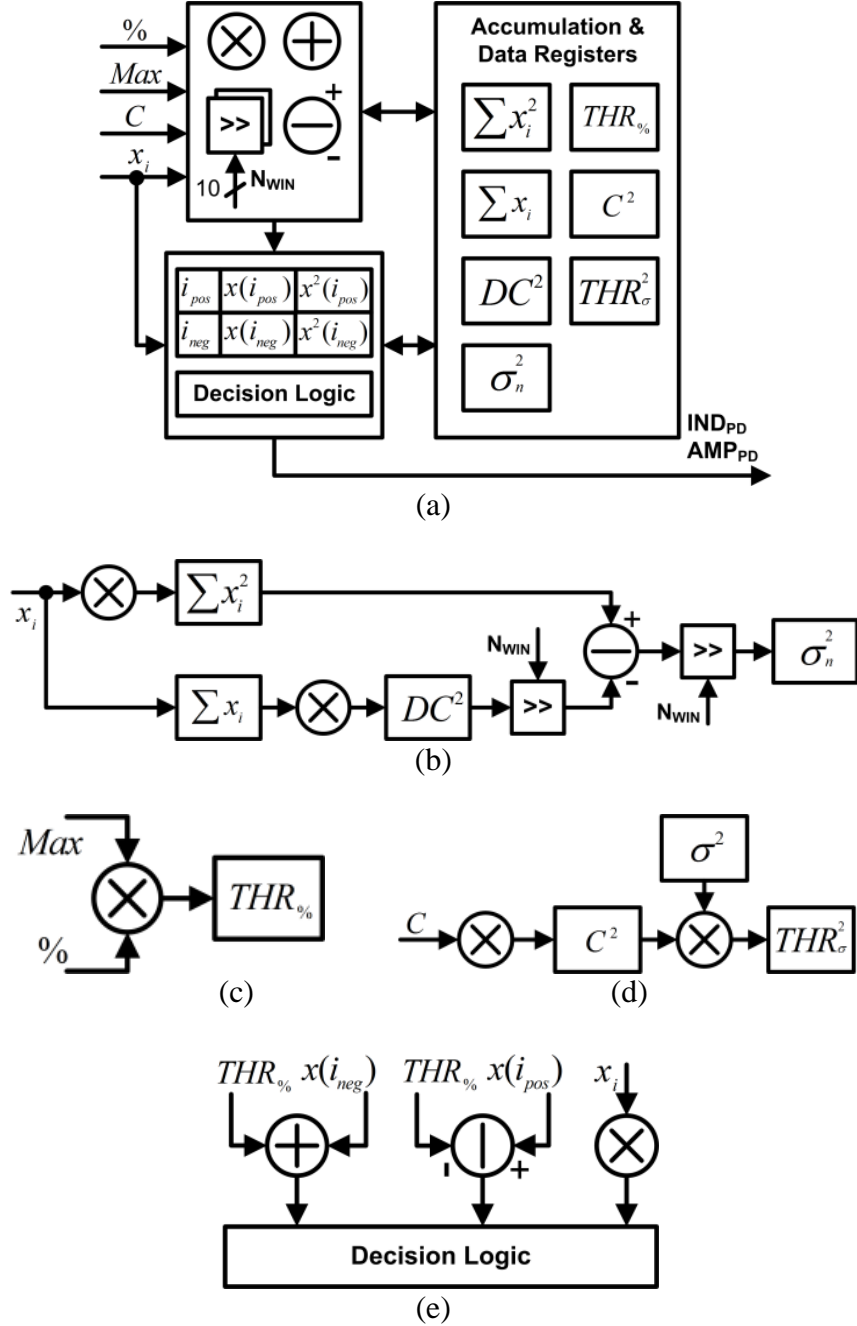


Fig. 3.5 (a) Architecture of the peak detector. Data path in the peak detector for (b) calculation of standard deviation of the noise floor during the training mode, updating of (c) percentage threshold and (d) noise threshold, and (e) peak detection. The multiplication operations are properly scheduled such that only one multiplication is performed for each clock cycle.

value of the input sample is compared with THR_{σ}^2 for amplitude thresholding, as described in [113]. The architecture of the peak detector, which's the hardware implementation of Table 3.1,

is shown in Fig. 3.5 (a). The accumulation and data registers store the value of two thresholds $THR_{\%}$ and THR_{σ}^2 , and the temporary results during their derivation. The index, amplitude and amplitude square of temporary positive and negative peaks are also stored in a register file. The decision logic is the realization of the case statement in Table 3.1, which can be implemented as a finite-state machine. As illustrated in Fig. 3.2 (b), σ_n^2 is computed during the training phase of the peak detector. In other stimulation trails $THR_{\%}$ and THR_{σ}^2 are updated right after FREE receives an instruction containing updated C and $\%$. Fig. 3.5 (b), (c), (d) and (e) illustrate the data-path arrangement in the peak detector for calculating σ_n^2 in the training phase, updating of $THR_{\%}$ and THR_{σ}^2 , and peak detection. By storing the temporary results into data registers, the multiplication operations in the computation of σ_n^2 and THR_{σ}^2 can be properly scheduled such that only one multiplication is performed for each clock cycle. As a result, only one multiplier is required in the peak detector, which saves the cost of its hardware implementation. On detecting a positive or negative peak, its index and amplitude are outputted to the fiber-response classifier, together with the flag signals indicating a detection and the polarity of detected peak. For fixed-point implementation, $\%$ is quantized into 16-bit precision, and C is encoded as a 8-bit unsigned integer with 5-bit fractional length, assuming C no greater than 6 as stated in [113].

3.3.3 Fiber Response Classifier

Given a fixed conduction distance on the nerve, i.e., the distance between the stimulation and recording electrodes in Fig. 3.1, fiber responses with specific conduction velocity peak within a fixed time range on an ECAP waveform, as shown in [35]. Real-time classification of fiber responses is hence possible by properly defining the time index of targeted fiber groups and locating the peaks on an ECAP waveform with index in the proximity of that defined by users. Fig. 3.6 (a) illustrates the principle of fiber-response classification. Two time indexes IND_{UP} and IND_{UN} are first defined by users, which mark the approximate time range of positive and negative peaks of the targeted fiber group (A fiber in Fig. 3.6 (a) as an example), respectively, according to the conduction distance and the ECAP responses collected in the stimulation trials indicated in Fig. 3.2 (b). The fiber response of that fiber group, represented by its positive and negative peaks, is then classified by finding the positive peak nearest to IND_{UP} and negative peak nearest to IND_{UN} , respectively.

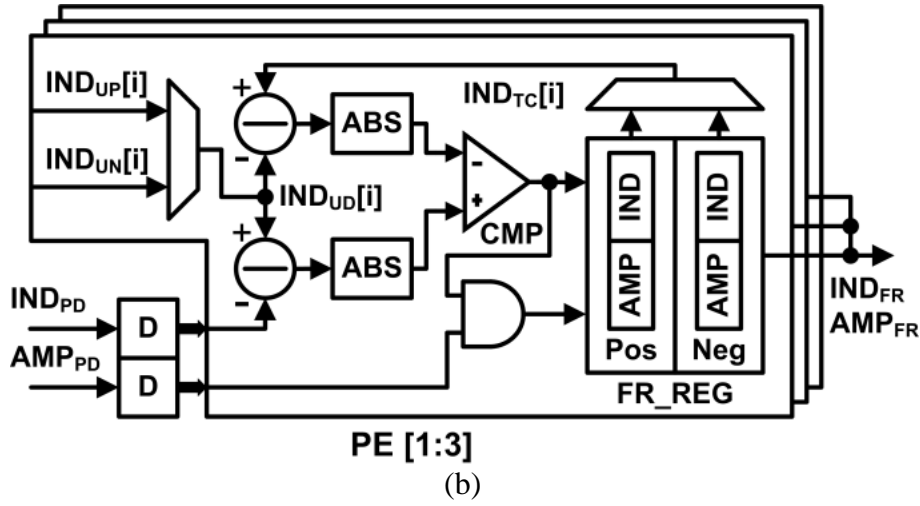
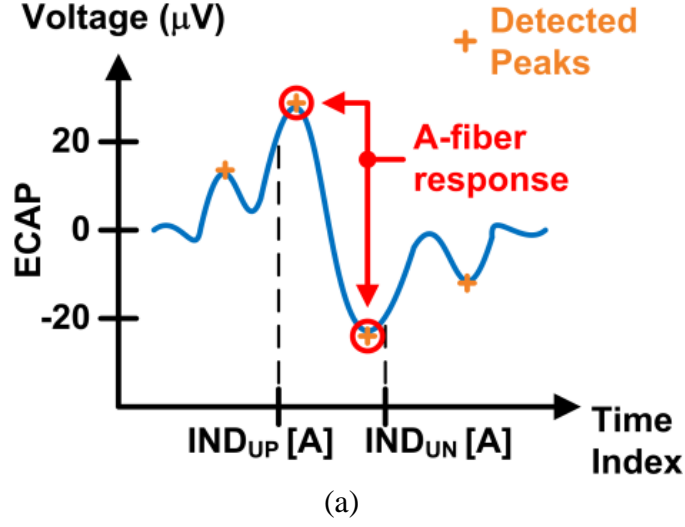


Fig. 3.6 (a) Principle of fiber-response (FR) classification and (b) its architecture.

Fig. 3.6 (b) presents the hardware architecture of the fiber-response classifier in FREE which supports the classification of 3 targeted fiber groups. When a peak is detected by the peak detector, its index (IND_{PD}) and amplitude (AMP_{PD}) are first latched in registers and fed into 3 processing elements (PE), each in charge of the classification of a fiber group. Based on the polarity of detected peak, either the IND_{UP} or IND_{UN} of a targeted fiber group is assigned to the user-defined peak index (IND_{UD}) in each PE. The PE then calculates and compares the time-index distance to the user-defined peak between detected and temporarily-classified peak, namely, $|IND_{PD} - IND_{UD}[i]|$ and $|IND_{TC}[i] - IND_{UD}[i]|$, where $IND_{TC}[i]$ denotes the index of temporarily-classified positive or negative peak in the i -th fiber group, depending on the polarity

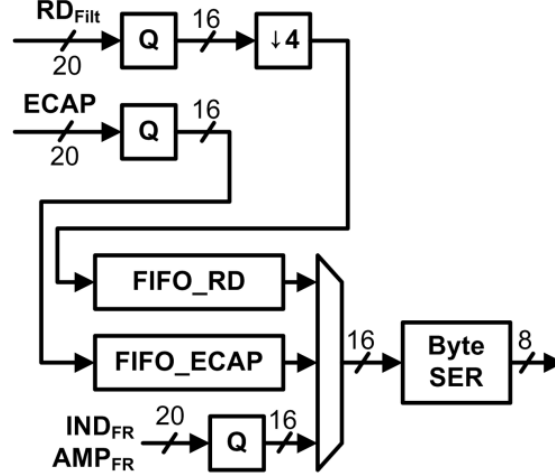


Fig. 3.7 Schematic of the output buffer in FREE.

of the detected peak, and $IND_{UD}[i]$ is the user-defined peak index in that group. Each PE has a fiber-response register (FR_REG) storing the index and amplitude of temporarily-classified positive and negative peaks. At the beginning of classification, the index and amplitude values of temporarily-classified positive and negative peaks in the FR_REG are initialized at 0. Upon receiving a detected peak with time-index distance to IND_{UD} smaller than that of temporarily-classified peak, i.e., $|IND_{PD} - IND_{UD}[i]| < |IND_{TC}[i] - IND_{UD}[i]|$, the index and amplitude values of temporarily-classified peak previously stored in the FR_REG are replaced with the IND_{PD} and AMP_{PD} , respectively. At the end of an ECAP waveform, the positive and negative peaks stored in the FR_REG will be the peaks closest to the IND_{UP} and IND_{UN} on time axis, respectively. The classified positive and negative responses stored in the FR_REG of 3 PEs are then serially read out on receiving user's instructions.

3.3.4 Output Buffer in FREE

Fig. 3.7 shows the schematic of the output buffer in FREE, which is the same as that in Chapter 2 except for additional an output of index and amplitude values of classified fiber responses. The RD_{Filt} , recovered ECAP responses and the amplitude value of fiber responses are quantized to 16 bits by preserving their most significant parts (i.e., the most significant 16 bits). The RD_{Filt} in FREE are also down-sampled by 4, and only the RD_{Filt} in the first and last AP stimulation cycles are stored into the first-in-first-out (FIFO) register $FIFO_RD$. Both the FIFO registers $FIFO_RD$ and $FIFO_ECAP$ have a total size of 16.4 kbits ($16 \text{ bit} \times 1024 \text{ word-depth}$), and the total size of

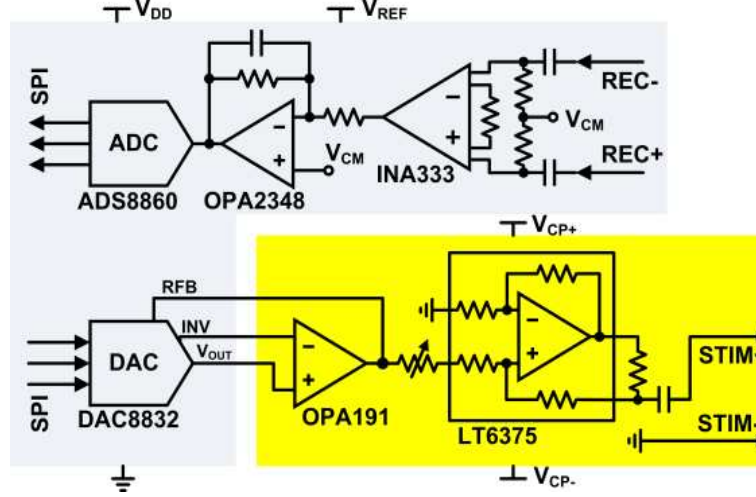


Fig. 3.8 Schematic of the stimulation and recording analog front-end (AFE) in the wearable device.

fiber response outputs is 192 bits ($16 \text{ bit} \times (2 \text{ peak indexes} + 2 \text{ peak amplitudes}) \times 3 \text{ fiber groups}$). The RD_{Filt} , ECAP and fiber responses can also be selectively read out at the end of each stimulation trial according to user's instructions and are serialized into byte streams for data transmission using the UART interface.

3.4 PCB Prototype of Wearable Device

In order to demonstrate the efficacy of the proposed FREE in reducing the data transmission rate of wearable wireless devices in closed-loop ENS systems, a printed circuit board (PCB) prototype of the wearable device in Fig. 3.1 is implemented.

Fig. 3.8 shows the circuit schematic of the stimulation and recording AFE, whose topology is the same as that described in Chapter 2. The ADC and DAC are implemented with the ADS8860 (*Texas Instruments*) and DAC8832 (*Texas Instruments*), respectively, both featuring 16-bit precision, micro-power and SPI-compatible serial interface. The DAC8832 combined with the external operational amplifier OPA191 (*Texas Instruments*) can be configured in bipolar output operation for AP stimulus pulse generation, whose pin connection is plotted in Fig. 3.8 [96]. A Howland current pump is employed and implemented with the LT6375 (*Analog Devices*) voltage-difference amplifier. The current pump adopts a DC-blocking capacitor to avoid direct

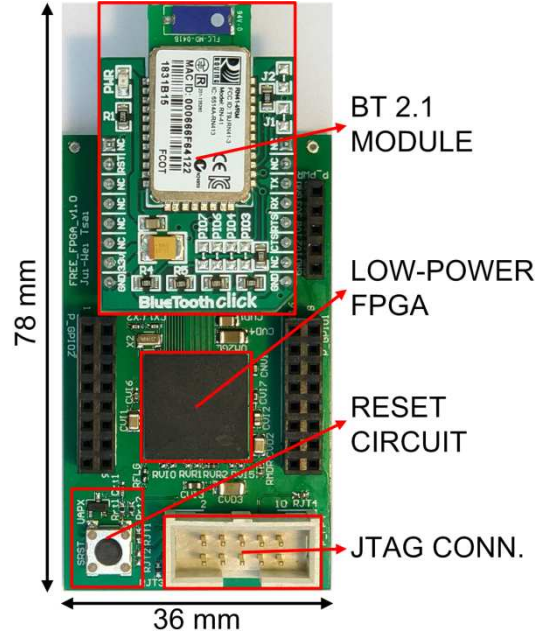


Fig. 3.9 PCB prototype of the wireless signal processing (WSP) platform with major components annotated.

current injection into nerves, and a resistor trimmer (SMUA102PET, *Ohmite*) to balance the resistor network, which helps improve its common-mode rejection ratio (CMRR) and output impedance [118]. Up to 1.5-mA stimulus current can be provided by this current pump. Neural signals are differentially recorded with a capacitively-coupled precision instrumentation amplifier (INA333, *Texas Instruments*) and amplified with an active filter implemented with the OPA2348 (*Texas Instruments*). The total gain of the recording AFE is 500, and its bandwidth is 1.6-20k Hz. The dual supply voltage of the LT6375 and OPA191 ($\pm V_{CP}$) is set to ± 10 V. The analog supply voltage (V_{REF}) and common-mode voltage (V_{CM}) of two amplifiers are 3.0 V and 1.5 V, respectively, and the digital supply voltage (V_{DD}) of ADC and DAC is 3.3 V.

Fig. 3.9 demonstrates the PCB prototype of the wireless signal-processing (WSP) platform. This WSP platform is built with a low-power FPGA (M2GL025-VFG256, *Microsemi Cor.*) as the main processor onto which FREE is mapped. On the programmed FPGA, FREE occupies 8275 (29.88%) logic elements (LEs), 3680 (13.29%) D flip-flops (DFFs), 8 (25.81%) large SRAMs (LSRAMs), each with size of 18×1024 bits, and 12 (35.29%) MACC units, each of which contains an 18×18 bits multiplier. The MIKROE-958 Bluetooth Click (*Mikro-Elektronika*) built

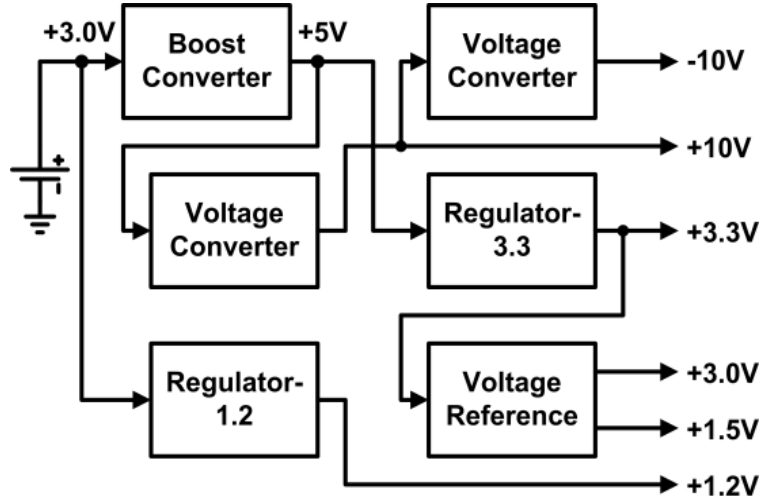


Fig. 3.10 Block diagram of the power-management unit (PMU) for the power supply of the AFE and WSP platform.

based on the RN-41 (*Microchip Tech.*) low-power, class-1 Bluetooth radio module is chosen as the example RF module in Fig. 3.1 and mounted on the WSP board. The RN-41 features an on-chip antenna, compatibility with Bluetooth 2.1 standard, the UART interface, and easy-to-configure property for instant USB cable replacement. Thanks to the proposed FREE, data transmission and reception on the RN-41 take place only at the end of each stimulation trial. Such advantage enables the RN-41 to be configured in sniff mode, where the radio wakes up at a specific interval set to 250 ms in our case, and sleeps in very low power mode (with current drain around 2 mA) for the rest of the time [119]. Compared with normal continuous mode with an average current consumption of 30 mA, RN-41 in sniff mode only drains 8-mA average current [120]. Other major components of the WSP board, including a JTAG connector for FPGA programming and the reset circuitry for both power-on and manual reset, are also annotated in Fig. 3.9. The core power of the FPGA is 1.2 V, and 3.3-V power supply is used for the I/O power of the FPGA and other active components, including RN-41 and APX803S-31SA-7 (*Diodes Incorporated*) in the reset circuitry. The total size of the WSP board is 78 mm × 36 mm.

Table 3.2 List of the components used in the PMU PCB

Component	Company	Part number
Boost Converter	Texas Instruments	TPS61032PWPR
Voltage Converter	Texas Instruments	TL7660CDGKT
Regulator-3.3	Texas Instruments	TPS78233DDCR
Regulator-1.2	Texas Instruments	TPS78001DDCT
Voltage Reference	Texas Instruments	REF2030AIDDCT

Fig. 3.10 illustrates the block diagram of the power-management unit (PMU) for power supply generation of the stimulation and recording AFE and the WSP platform from a single 3-V, 620-mAh CR2450 coin battery. First, a TPS61032PWPR boost converter from *Texas Instruments* is chosen to convert the voltage from a single coin battery to 5 V, owing to its 20- μ A quiescent current, wide input voltage range (1.8-5.5 V) [121], and 93% conversion efficiency at 25-mA

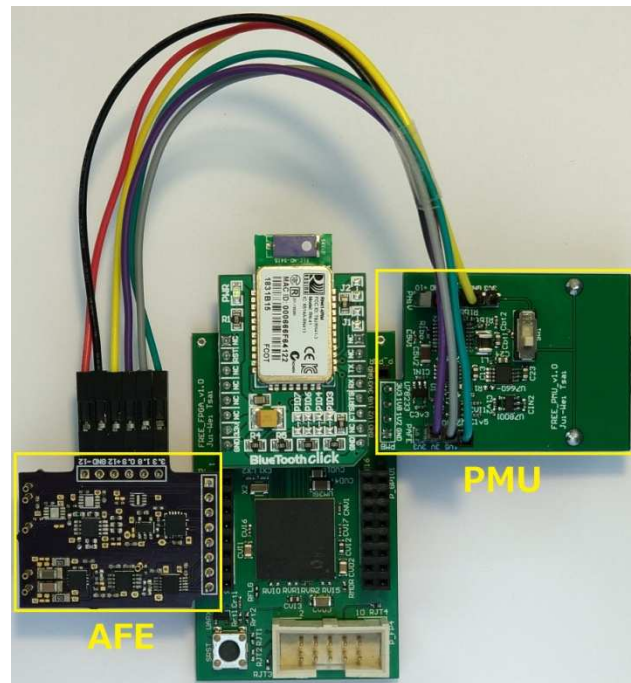


Fig. 3.11 The assembled PCB prototype of the wearable device.

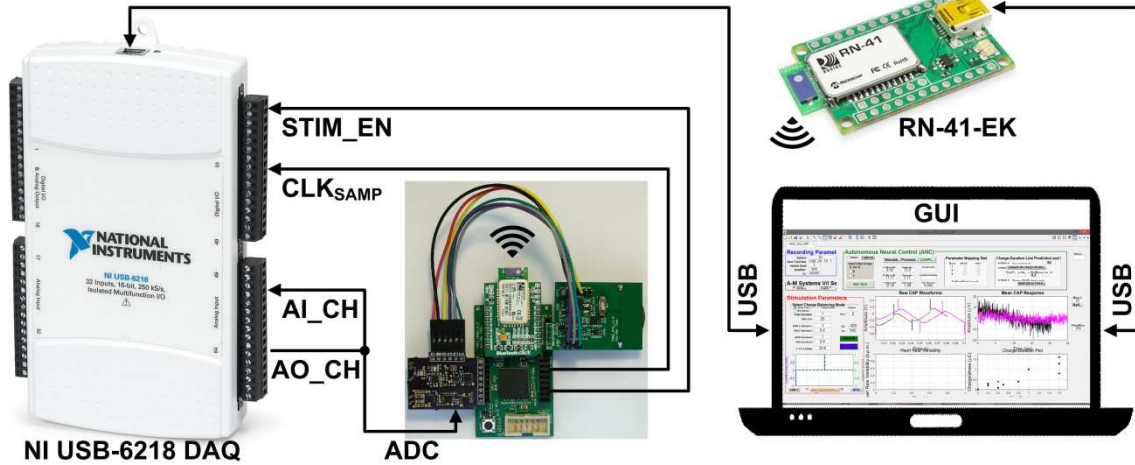


Fig. 3.12 Illustration of experiment setup for performance comparison between hardware (HW) and software (SW) processing.

output current (simulated with WEBENCH® Power Designer) [122]. Two voltage converters (TL7660, *Texas Instruments*), one configured as a positive-voltage doubler and the other as a negative-voltage converter as described in [123], are adopted to deliver ± 10 -V supply voltage. The 3.3-V and 1.2-V supply voltage are generated by down-regulating the 5-V boost converter output and 3-V coin battery output with low drop-out voltage regulators, and the 3-V and 1.5-V voltage in the recording AFE are generated with a voltage reference. Table 3.2 lists all components used in the implementation of the PMU. The assembled PCB prototype of the wearable device in Fig. 3.1 is shown in Fig. 3.11, in which the AFE and PMU boards are marked, and the coin battery is on the bottom side of the PMU board. The measured power consumption of the wearable device is 234 mW (3-V battery output \times 78-mA current consumption).

3.5 Experiment Results

3.5.1 Experiment Setup

The efficacy of the BFCA core in SAR and distortion-free noise removal has been presented in Chapter 2. In this chapter, we further demonstrate the efficacy of FREE in not only reducing the data transmission rate of a wearable device but also improving the accuracy in the prediction of the NAP in the ANC platform compared with previous software-based processing in [35]. Fig. 3.12 illustrates the experiment setup for performance comparison between hardware (HW) and

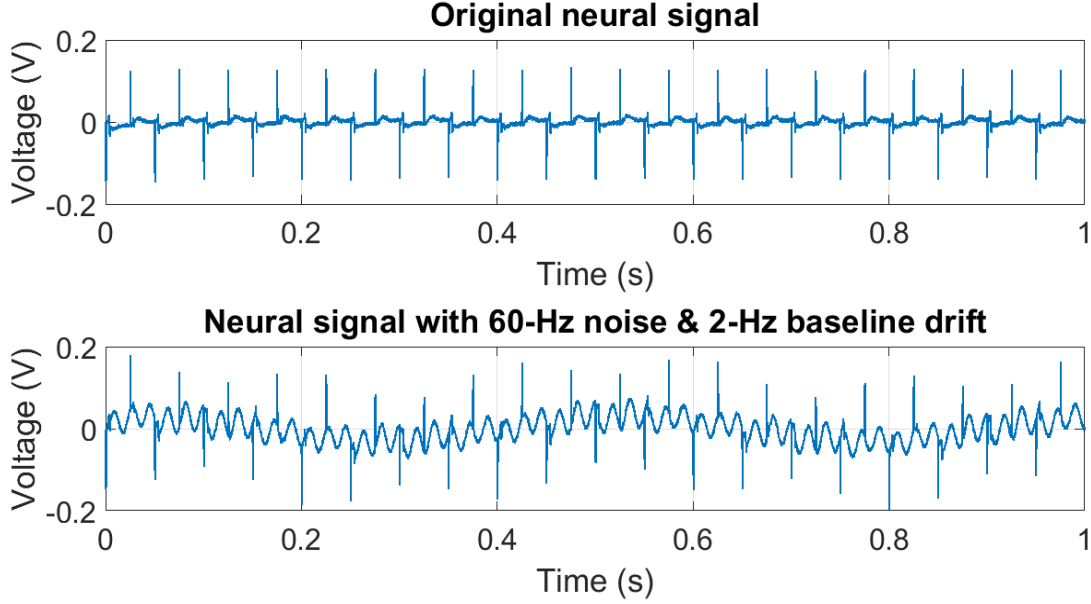


Fig. 3.13 Plots of the original neural signal recorded in a single stimulation trial that contains ECAP responses and the neural signal superimposed with 60-Hz noise and 2-Hz baseline drift as the input data for HW and SW comparison. The input data should be shifted by $+V_{cm}$ to meet the dynamic range of the ADC on AFE..

software (SW) processing, in which SAR, noise removal, and feature extraction of ECAPs are performed using FREE on a FPGA and MATLAB R2016a (*MathWorks*), respectively. A RN-41-EK (*Microchip Tech.*) Bluetooth evaluation kit serves as the base station in Fig. 3.1, which's controlled by a MATLAB-based graphic users interface (GUI) on the PC via its USB port. The RN-41-EK is configured in master mode which establishes connection to the RN-41 module on the wearable device configured in slave mode [119]. A data acquisition board (USB-6218, *National Instruments*) interfaced with the PC via an USB port is used to generate test signals which are fed into the ADC input and processed with the FPGA on the wearable device. The test signals generated from the analog output channel of the USB-6218 are also sent to its analog input channel for SW processing. Two control signals from FREE, $STIM_EN$ (the flag signal pulled high during the stimulation train) and CLK_{SAMP} (the 50-kHz sampling clock), are fed to the digital input channels of the USB-6218; Both the signal output on the AO_CH and acquisition on the AI_CH are enabled by the $STIM_EN$ signal and synchronized to the sampling of the ADC with the CLK_{SAMP} signal.

Fig. 3.13 shows the input test signals for performance comparison, where the pre-recorded neural signals presented in [35] containing valid fiber responses are superimposed with periodic noises, including power-line interference and baseline wander simulated by 60-Hz and 2-Hz sinusoidal waves [124-126], respectively, with -6-dB signal-to-noise ratio (SNR) and random phase shift for both waves. The SNR is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{FR_{rms}}{A_{rms}} \right), \quad (3.2)$$

where FR_{rms} is the root-mean-square (rms) amplitude of fiber responses waveform on the "original ECAP", i.e., the ECAP derived from original noise-free neural signals, and A_{rms} is the rms amplitude of the sine waves. In the computation of A_{rms} the fiber responses are defined as the first 6-ms interval of ECAP waveform as seen in Fig. 3.14 (a). A band-pass elliptic filter with 0.1-4 kHz pass-band, 20-dB stop-band attenuation and 0.1-dB passband ripple is adopted for both the forward and reverse filters in the BFCA core.

3.5.2 Precision Comparison

An example of the recovered ECAP via HW- and SW-processing versus the original ECAP waveform is plotted in Fig. 3.14 (a). It can be seen the ECAP obtained from HW-processing is bettered regained than that from SW-processing in terms of waveform distortion, defined as the mean squared Euclidean distance between the original ECAP and the recovered ECAP waveform in (2.17), which proves that real-time BFCA algorithm on FREE rejects periodic noises better than traditional coherent averaging used in the SW processing [35]. Fig. 3.14 (b) plots all HW- and SW-recovered ECAP waveforms and extracted fiber responses from pre-recorded neural signals in 66 stimulation trials, each with unique stimulus parameters.

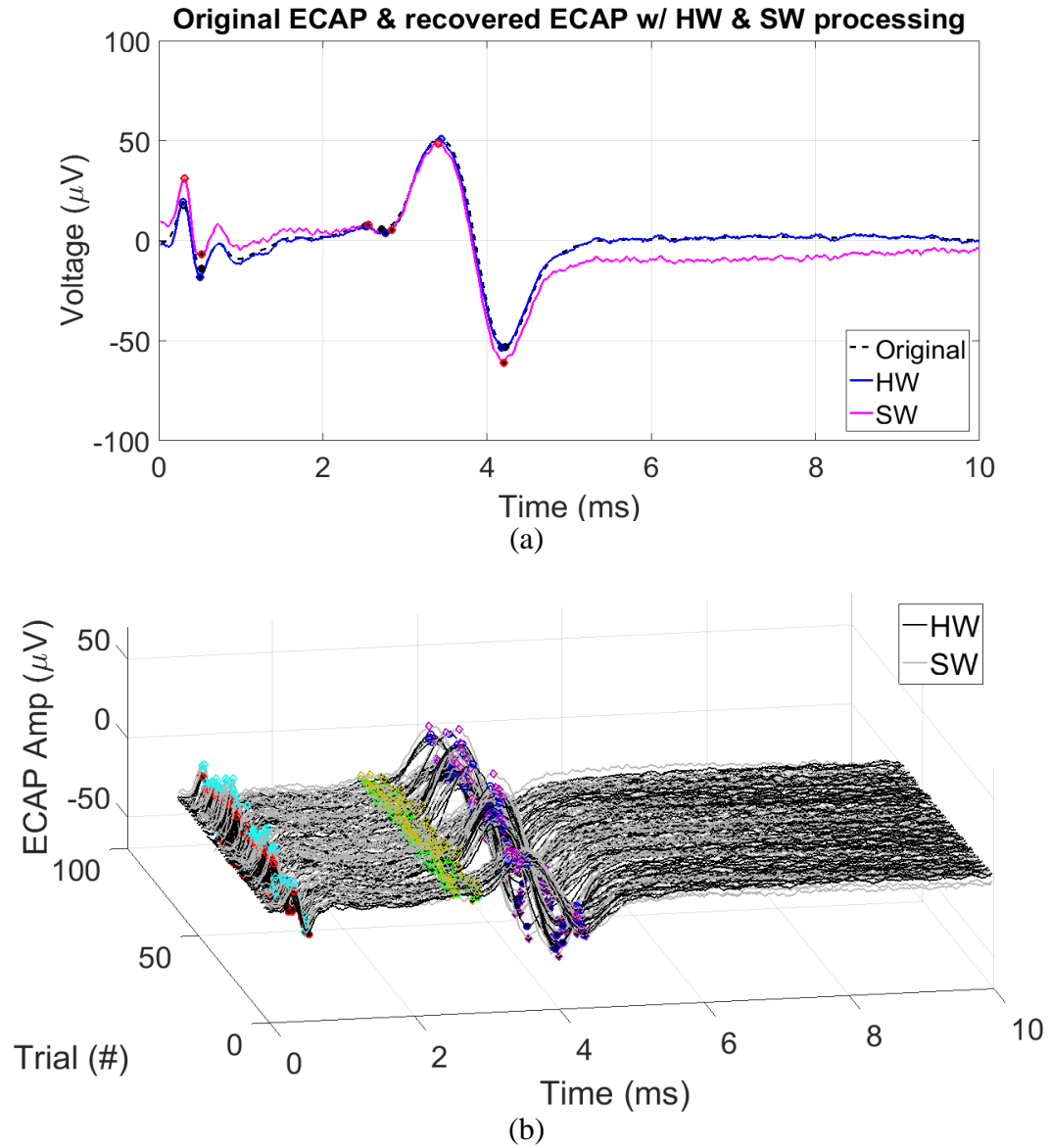


Fig. 3.14 (a) An example of recovered ECAP via HW- and SW-processing versus the original ECAP waveform. (b) All HW- and SW-recovered ECAP responses from the entire input data set (66 trials).

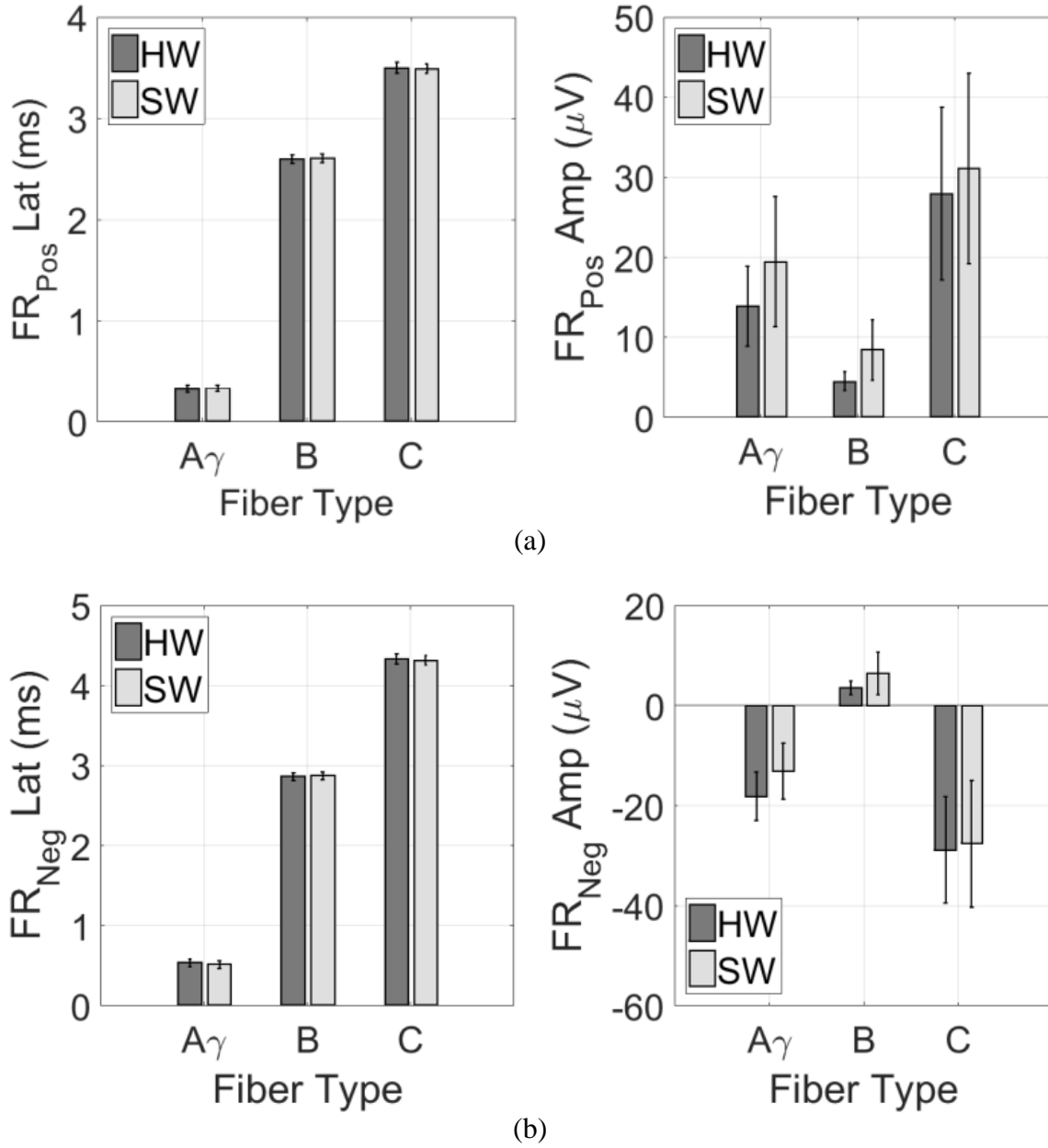


Fig. 3.15 Mean latency and amplitude of extracted (a) positive and (b) negative fiber responses. Both the positive and negative fiber responses extracted from HW have less amplitude variation than those from SW owing to the more effective removal of periodic noises on HW.

The precision of fiber-response classification on HW and SW is first compared in terms of fluctuation in the latency and amplitude of extracted fiber responses [99]. Fig. 3.15 (a) and (b) show the mean latency and amplitude of positive and negative responses, respectively, of 3 fiber groups (A γ , B, and C fibers given an 8-mm conduction distance reported in [35]) extracted from

both HW and SW and their variations. The responses of 3 fiber groups on both HW- and SW-recovered ECAP waveforms are recognizable and hence there's little difference between HW and SW in the latency variation of positive and negative fiber responses. Nevertheless, the HW has higher classification precision indicated by its positive and negative responses having a lower amplitude variation than that from SW, especially for A γ and B fibers with low SNR. Such amplitude variation results mainly from the amplitude error of fiber responses caused by periodic noises on ECAP waveforms. Note that positive local minima within the time range of B fiber are extracted as its negative responses on both HW and SW due to a low yet identifiable response amplitude of B fiber.

To compare the overall classification precision, the slope-activation relationships of both HW and SW, which predicts the threshold current (i.e., rheobase current, I_{Rh}) versus fiber activation level and is the key to constructing the NAP in the ANC platform [35], are derived based on extracted fiber responses and the corresponding stimulus parameters. As reported in [35], the stimulus-response data are first sorted, and the amplitude of the positive fiber response (i.e., positive peak) is converted into an activation level by normalizing it with the largest fiber response amplitude representing the maximal activation. Fiber responses with similar activation level (i.e., with difference less than a preset error tolerance), along with the associated stimulus parameters are then clustered, and clusters with at least 2 stimulus-response pairs are selected. The rheobase current, i.e., the slope of the charge-duration line of each chosen cluster corresponding to an activation level, is computed using the least-square linear regression. The slope-activation relationship of a fiber group can be predicted by applying linear regression on the abovementioned slope-activation data and is modeled by the equation $\widehat{I_{Rh}} = A \cdot r^\lambda$, where $\lambda = \frac{AMP_{FR}}{AMP_{max}} \cdot 100$ is the activation level in percentages, A is the rheobase current for 0% activation level, and r is a constant reflecting the rate of growth of rheobase current with respect to the activation level. Fig. 3.16 (a) and (b) illustrate the slope-activation data derived using fiber responses from HW and SW, respectively, together with the corresponding equations modeling the slope-activation relationship and coefficient of determination (R^2) representing the goodness of fit. The amplitude error of fiber responses caused by periodic noises introduces inaccuracy into the process of clustering and computation of rheobase current for each cluster,

which causes the slope-activation data to deviate from the predicted model. With the real-time removal of periodic noises on ECAPs, slope-activation data derived from the HW-extracted fiber responses fit better to the predicted model as indicated by R^2 in Fig. 3.16, suggesting that higher overall classification precision is achieved on HW. This helps more accurately constructing a NAP which estimates the stimulus parameters to maintain desired nerve activation level.

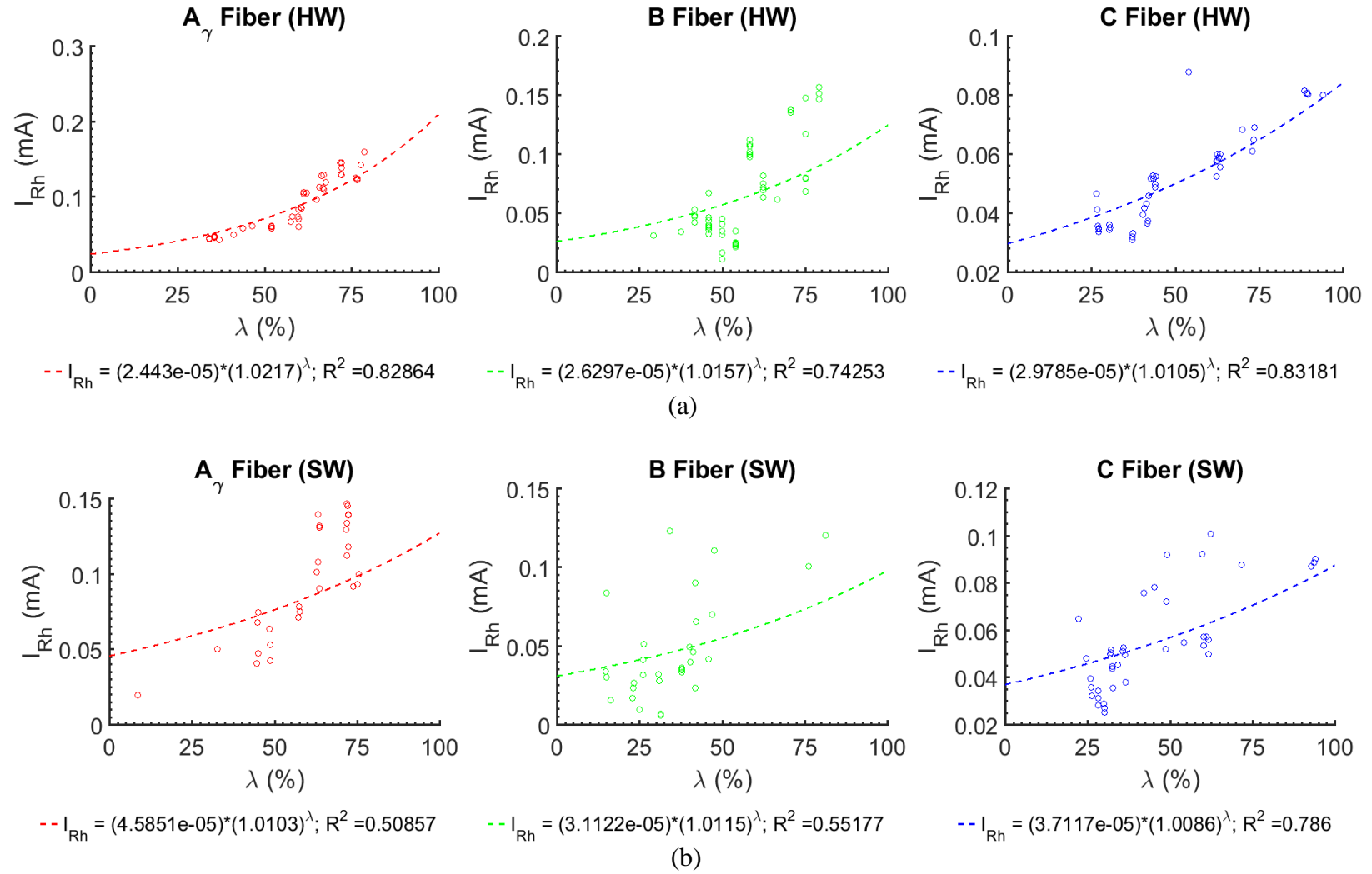


Fig. 3.16 Slope-activation data of extracted fiber responses from (a) HW and (b) SW and the predicted slope-activation relationship (i.e., rheobase currents I_{Rh} as a function of the percent activation level λ). The classification precision of fiber responses from HW and SW is reflected by the coefficient of determination R^2 which represents the goodness-of-fit of predicted I_{Rh} .

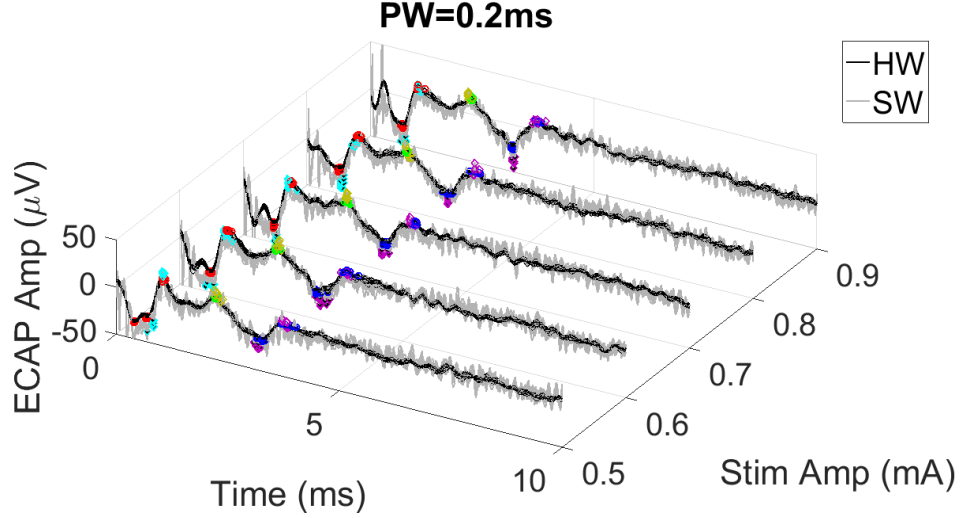


Fig. 3.17 *In-vivo* test results (PW = stimulus pulse width = 0.2 ms; PRF = pulse repetition frequency = 20 Hz; t_{train} = stimulus train duration = 1 s) of HW- and SW- recovered ECAP waveforms and extracted fiber responses.

3.5.3 In-Vivo Test Results

The performance comparison between HW- and SW- processing is further accomplished on a male Long-Evan rat *in-vivo*. Two custom-made silicone cuff electrodes in differential configuration serve as the stimulation and recording electrodes depicted in Fig. 3.1. These two electrodes are attached to the cervical vagus nerve of the rat following the surgical procedure in [35], and connected to the stimulator output and the differential amplifier input in Fig. 3.8. An 8-mm conduction distance is measured after the implant of electrodes. For data acquisition in SW processing, the output of the OPA2348 is connected to the analog input channel of the USB-6218.

A series of stimulation trials with varying amplitude are applied to the nerve with the following stimulus parameters: PW = stimulus pulse width = 0.2 ms, PRF = pulse repetition frequency = 20 Hz and t_{train} = stimulus train duration = 1 s. The stimulus amplitude ranges from 0.5 mA to 0.9 mA with 0.1-mA increment to obtain observable fiber responses on ECAP waveforms without amplifier saturation. Fig. 3.17 plots ECAP waveforms and extracted responses of 3 fiber groups (A δ , B, and C fibers) from both HW and SW against stimulus amplitude, in which ten ECAP waveforms are collected per stimulus amplitude. The peak-to-peak amplitude of fiber responses

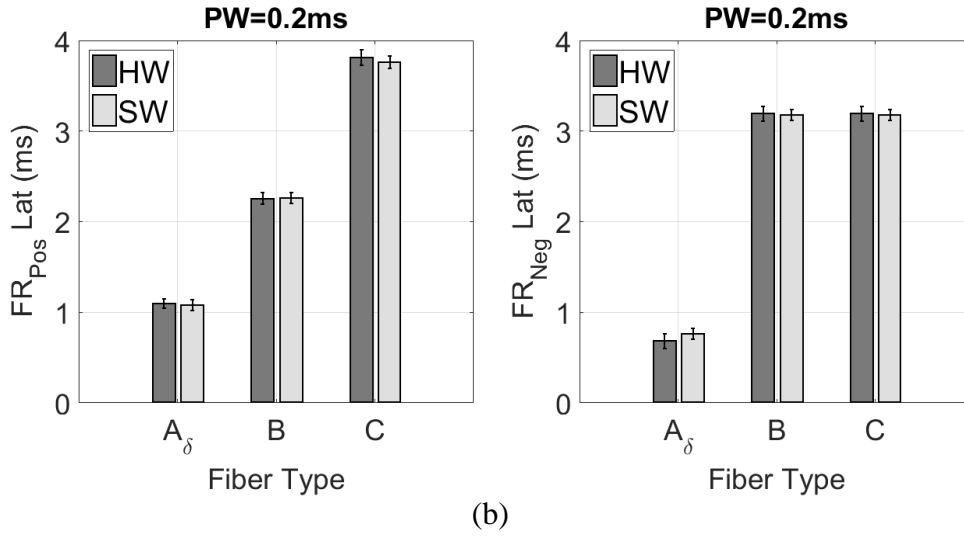
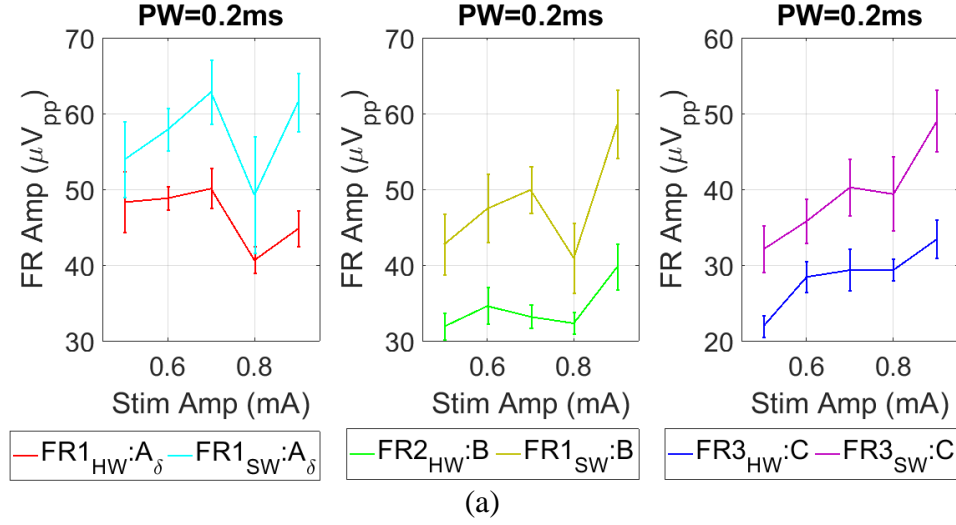


Fig. 3.18 (a) The amplitude growth function (AGF) of extracted fiber responses and (c) the mean latency of positive and negative fiber responses extracted from HW and SW in in-vivo tests.

versus stimulus amplitude, also known as the amplitude growth function (AGF), and the mean latency of positive and negative responses, along with their variations, are plotted in Fig. 3.18 (a) and (b), respectively. Both HW- and SW- recovered ECAP waveforms contain peaks contributed by both fiber responses and noises (both periodic and random) in the proximity of user-defined fiber response index (i.e., IND_{UP} and IND_{UN}), and hence the mean latency and variation of positive and negative responses from HW and SW are nearly identical. At each stimulus amplitude, however, the peak-to-peak amplitude variation of 3 fiber responses from HW is lower

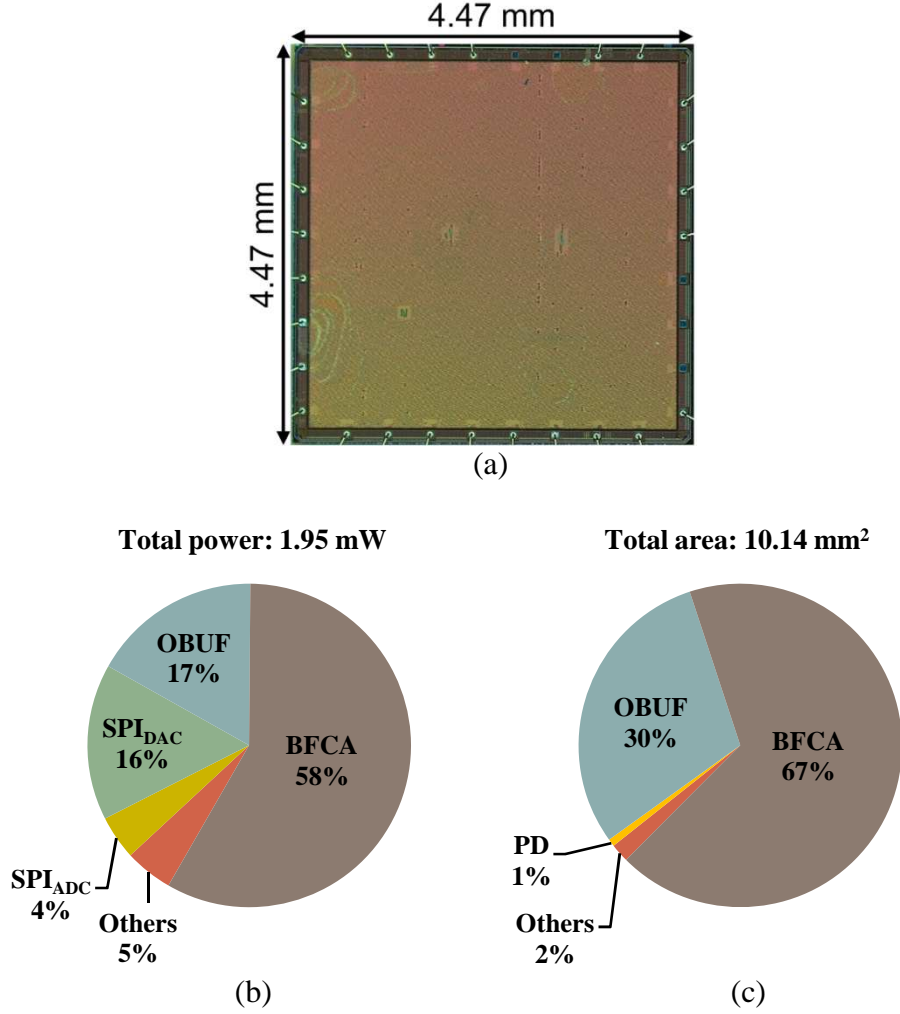


Fig. 3.19 The ASIC implementation of proposed FREE in 180-nm CMOS technology: (a) die photo and breakdown of its (b) power and (c) area consumption.

than that from SW, which's necessary for constructing NAP with higher accuracy, as verified in Fig. 3.15 and Fig. 3.16.

3.5.4 ASIC Implementation

The proposed FREE is further implemented in 180-nm CMOS technology. Fig. 3.19 (a) shows the micrograph of the fabricated chip, whose total core and chip area are 10.14 mm² and 19.98 mm², respectively. The core and I/O voltages of the chip are 1.8V and 3.3V, respectively, and the system clock frequency is 16 MHz. To the best of our knowledge, this is the first digital signal processor dedicated to the newly-proposed ANC platform. This chip is tested with the same

Table 3.3 Performance comparison between hardware and software processing

	Software [35]	This work
Filtering Technique	Coherent Averaging (CA)	BFCA
Processing Unit	Intel Core i5-3230M	ASIC
Freq. [MHz]	2,600	16
Power [mW]	28,440	1.95 (at 1.8V)
Latency [ms]	11.5	20.5
Max Data Rate [kbps]	800	16.4
Amplitude Variation (B Fiber) [μ V]	3.74	1.22
Coeff. of Determination (A Fiber)	50.9%	82.9%

setup as that in Fig. 3.12, except that the FPGA is replaced with the chip under test, and an 1.8-V low drop-out voltage regulator (TPS78218DDCT, *Texas Instruments*) down-regulating the 3-V coin battery output to 1.8V is used as the core voltage supply of the chip. At 1.8-V core voltage and 16-MHz clock rate, the measured active and stand-by power are 1.95 mW and 0.3 mW, respectively. Fig. 3.19 (b) and (c) show the breakdown of power and area consumption, respectively, which's estimated with the layout result and post-layout simulation in Cadence Innovus Implementation System. The power consumption of FREE is dominated by memory banks in both the BFCA core and the output buffer and the fixed-width multiplier of IIR filters in the BFCA core, which can be further reduced by employing the low-power memory architectures [127, 128] and power-efficient fixed-width multipliers [129, 130].

Table 3.3 shows the comparison with the SW processing, the MATLAB-based ECAP processing in [35] comprising SAR, denoising and fiber response extraction, on an Intel® Core™ i5-3230M Processor. The SW processing requires wearable neuromodulation devices to continuously transmit recorded neural data at 800-kbps (16-bit precision \times 50-kHz sampling frequency) data transmission rate. In contrast, with real-time ECAP processing via FREE, the maximum data rate becomes 16.4 kbps for both RD_{Filt} and ECAP output (on the ground that t_{train} is greater than 1 s and both RD_{Filt} and ECAP output has 16.4 kbits), at least 49 \times lower than that of SW, and 192 bps

for fiber response output which is $4167\times$ lower. The computation latency, i.e., the required time for signal processing in ANC, is estimated over 66 stimulation trials. Whereas the latency of FREE is of the same order of that of SW, the computation power of FREE is much less than that of SW, on which the average power of the CPU is 28.44 W. Moreover, FREE improves the precision of fiber-response classification and the accuracy of NAP construction in noisy environments. For example, in B fiber classification, the amplitude variation of the positive response from FREE is 1.22 μV , which is $3.1\times$ lower than that from the SW; in constructing the NAP of A γ fiber, the coefficient of determination of the slope-activation profile derived from the fiber responses from FREE is 82.9%, which is increased by 62.9% compared with the SW.

3.6 Conclusion of This Chapter

In this chapter, a fiber-response extraction engine (FREE), the first DSP engine dedicated to nerve activation control using the newly-proposed ANC platform, was presented. FREE employs the DSP architecture presented in Chapter 2 for stimulus artifact rejection and distortion-free filtering of ECAP waveforms. Computationally-efficient algorithms for fiber response extraction on ECAPs and their VLSI architectures were also explained. FREE was implemented on a custom-made and coin battery-powered wearable PCB integrating a low-power FPGA, a Bluetooth transceiver, a stimulation and recording AFE and a power management unit. FREE reduces the maximum data rate of wearable devices to 16.4 kbps, which is at least $49\times$ lower than that of software processing. Experimental results also show that compared with the previous software-processing, FREE improves the precision of fiber response classification by $3.1\times$ in noisy environments, which increases the accuracy of nerve activation profiles by up to 62.9%. An ASIC implementation of FREE was also presented whose total chip area and core power consumption of 19.98 mm^2 and 1.95 mW, respectively. FREE facilitates nerve activation control on wearable devices by reducing the data rate and power costs and improving the precision of NAP in noisy environments, and can be applied to other closed-loop ENS systems utilizing the ECAP as their feedback biomarkers.

4. CONCLUSION AND FUTURE WORK

4.1 Conclusion

Electrical neurostimulation is an emerging therapeutic for various neurological diseases and possesses the advantage over pharmaceutical approach to dose the targeted area of brain or nerve more precisely. Closed-loop neurostimulation approaches increase the stimulation efficacy and minimize side-effects and patient's discomfort by constantly tailoring the stimulation parameters according to feedback physiological signals from patients. To improve the quality and reduce the costs of treatments, wireless neurostimulation devices capable of both stimulation and telemetry of recorded physiological signals can be introduced into closed-loop neurostimulation in replacement of laboratory instruments. In view of the data transmission rate and the resulting power consumption of wireless devices, a real-time DSP processor processing and extracting features from recorded signals is desired; its VLSI architecture and implementation in FPGA and ASICs are especially attractive for optimal computation power and cost. ECAP is an objective measure of the nerve activity and condition and has been adopted as the feedback biomarker in closed-loop ENS systems including NRT in cochlear implants and a newly proposed ANC platform. This thesis focuses on the development of a DSP engine and its VLSI architecture for real-time processing of ECAP, including SAR, denoising, and extraction of nerve fiber responses as biomedical features. When integrated with the wireless device applied in closed-loop ENS systems, not only does such DSP engine reduce required data rate of the wireless device, but it also improves the precision of extracted biomedical features by removing artifacts and noises on ECAPs, which facilitates the tailoring of stimulation parameters and boosts the efficacy of closed-loop ENS systems.

Chapter 2 describes a DSP architecture for recovery of ECAP responses in NRT and other ECAP-based closed-loop ENS systems. A newly proposed BFCA technique enables the configurable linear-phase filter to be realized hardware efficiently for distortion-free filtering of ECAPs, and this technique can be easily combined with AP stimulation method for SAR. This DSP architecture also incorporates folded-IIR filter and division-free averaging to reduce the computation cost. The DSP architecture is mapped onto a low-power FPGA, and it's proved in

in-vivo tests to reject stimulus artifacts overlapped with ECAP responses and remove periodic noises more effectively without distorting ECAP waveforms.

A fiber-response extraction engine (FREE) is presented in Chapter 3 for nerve activation control in closed-loop ENS using the ANC platform. FREE utilizes the DSP architecture proposed in Chapter 2 for SAR and denoising of ECAPs, and the DSP architecture of computationally efficient peak detection and classification algorithms for fiber response extraction from ECAPs in real time. FREE is mapped onto a custom-made and battery-powered wearable wireless device incorporating a low-power FPGA, a Bluetooth transceiver, a stimulation and recording AFE circuitry and a power-management circuitry. In comparison with previous software-based signal processing, FREE demonstrates its capacity to not only reduce the data rate of wearable devices but also improve the precision of extracted biomedical features (fiber responses) reflected by their amplitude and latency variation. It is also verified that fiber responses extracted from FREE, which possesses higher precision than those from software processing, helps boost the accuracy of NAP construction in ANC. An ASIC version of FREE is implemented in 180-nm CMOS technology, whose total chip area is 19.98 mm^2 and core power consumption is 1.95 mW at 1.8-V core voltage and 16-MHz system clock rate.

4.2 Future Work

4.2.1 Half-Precision Floating-Point Computation

All the computations in FREE, including the BFCA method combined with the AP stimulation based SAR in Chapter 2 and the peak detection and classification in Chapter 3, are accomplished with fixed-point arithmetic in which the data width is 20 bits. Since the arithmetic operations in those algorithms consist mainly of addition, subtraction, multiplication, and logical shift, they can be implemented in half-precision floating-point (HPFP) arithmetic to further save the required data width in FREE yet still achieve desired data precision. The data width of a HPFP number is 16 bits (1-bit sign, 5-bit exponent and 11-bit fraction) and hence this format is also referred as "binary16" in IEEE 754-2008 standard [131].

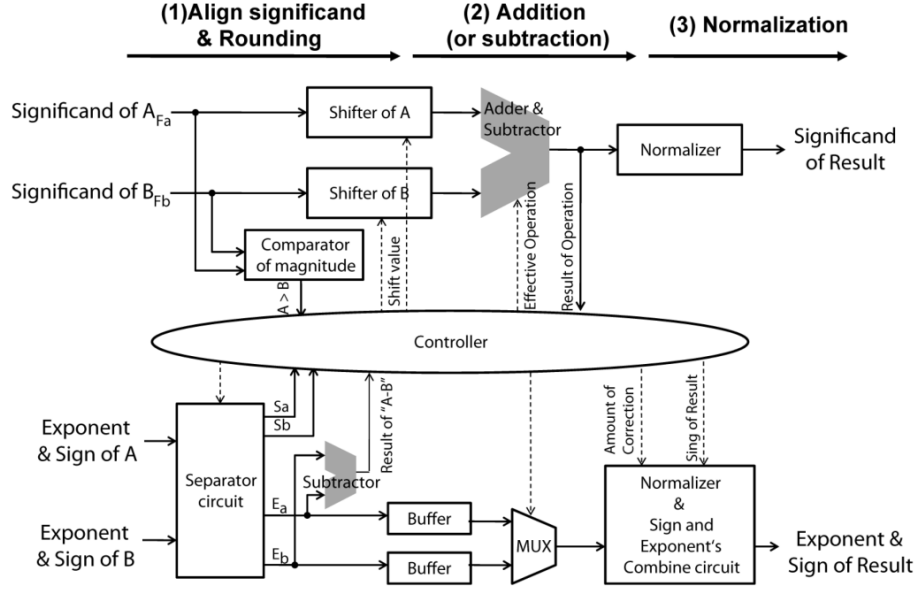


Fig. 4.1 Block diagram of a half-precision bit-serial floating-point adder [132].

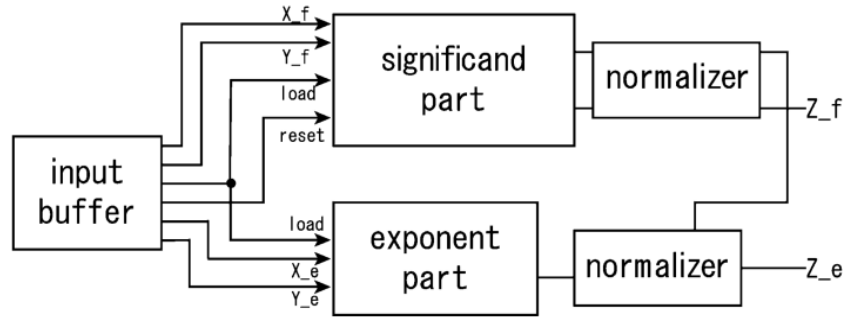


Fig. 4.2 Block diagram of a half-precision floating-point multiplier, where (X_e, Y_e) and (X_f, Y_f) are exponents and fractions of two input signals, respectively, "load" and "reset" are control signals for each part, and Z_e and Z_f are the exponent and fraction of multiplier output [133].

As stated in Section 2.3.3, the dynamic range of 16-bit ADC output in 2's complement format is -32768 to 32767 and can be expressed in HPFP format, whose the maximum representable value is 65504 . To avoid overflow in computations, data in HPFP format can be scaled by factors equal to power of 2, which corresponds to subtracting the exponent of a HPFP number. Addition, subtraction, and multiplication can be implemented in hardware according to the floating-point arithmetic defined in IEEE 754 standard [134]. Fig. 4.1 illustrates the block diagram of a HPFP adder [132], and the block diagram of a HPFP multiplier is also shown in Fig. 4.2 [133]. The left shifting in (2.13) and right shifting in (2.9) can be achieved by adding and subtracting the

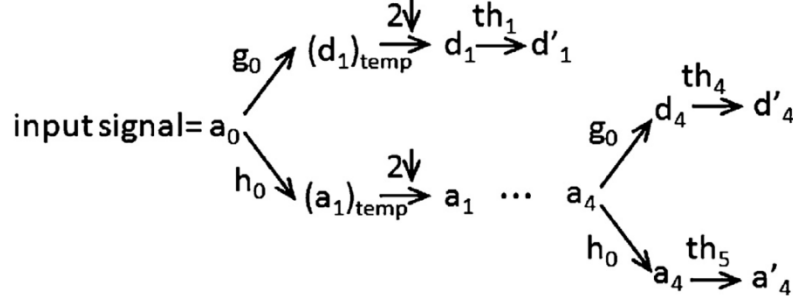


Fig. 4.3 Illustration of DWT algorithm for data compression with 4 levels of decomposition [63]

exponent of a HPFP number, respectively. The filter coefficients in the BFCA method can be first derived in single-precision floating-point format and then converted into HPFP format. Implementing the computations in FREE in HPFP arithmetic can effectively reduce required data width and the resulting hardware costs, including area and power consumptions.

4.2.2 Data Compression of ECAP

In close-loop ENS, the morphology of recorded ECAPs is crucial for neurologists to identify valid nerve fiber activation and determine the approximate latency of fiber responses. As mentioned in Section 3.5.4, the required data transmission rate for an ECAP output in FREE is 16.4 kbps. This can be further reduced by applying real-time data compression technique to the ECAP outputs from the BFCA core. The discrete wavelet transform (DWT) in combination with run-length encoding (RLE) is one popular data compression technique which has been utilized for data reduction in various biomedical systems such as neural recording [63] and bladder pressure monitoring [135]. This technique has the advantage of preserving the temporal information and the shape of detected events on time window, and hence is applicable to ECAP waveforms on which the latency and amplitude of fiber responses on ECAPs must be maintained after compression.

DWT decomposes signals into different frequency bands with multiple stages of low and high pass filters. Fig. 4.3 illustrates DWT algorithm for data compression with 4 levels of decomposition [63], where h_0 and g_0 are the filter coefficients of low and high pass filters, respectively, and a_j and d_j represent approximation and detail coefficients at j -th level, respectively. At each level, a_j is filtered by h_0 and g_0 to generate temporary outputs $(a_{j+1})_{\text{temp}}$ and $(d_{j+1})_{\text{temp}}$ which are

down-sampled by 2 to obtain a_{j+1} and d_{j+1} . After decomposition a pre-defined threshold th_j is applied to detail coefficients d_{j+1} , whose resulting value is denoted by prime ('), namely, d'_{j+1} . At the last 4th level, thresholds are applied to both approximate and detail coefficients to generate a'_4 and d'_4 , respectively. The thresholding operation will keep the significant high-energy coefficients contributed by events and zero the insignificant low-energy coefficients resulted from noises, which is equivalent to the wavelet-based filtering [80, 81].

A preliminary result of data compression of the ECAP waveform in Fig. 3.14 (a) is demonstrated with a 2-level DWT, where the Haar wavelet is selected owing to its computation simplicity (which requires only addition and subtraction) [117, 136]. Fig. 4.4 (a) shows the derivation of quantization threshold of the wavelet coefficients of an ECAP waveform. First, the wavelet coefficients (denoted as "WC" in Fig. 4.4) of noises are computed by applying the DWT to the noise waveform obtained using the BFCA method, and then the mean of absolute value of wavelet coefficients (μABS_{WC}) in each decomposition level is computed [82]. The quantization threshold for the wavelet coefficients of an ECAP waveform (THR_{WC}) is derived by multiplying the μABS_{WC} with an empirical scaling constant, which is set to 5 in this demonstration. The wavelet coefficients of an ECAP waveform, obtained from the DWT, are compared against the THR_{WC} , below which the wavelet coefficients are quantized to zero. Fig. 4.4 (b) shows the wavelet coefficients of an ECAP waveform before and after quantization. The quantized wavelet coefficients are encoded with RLE, where sequence of zeros is replaced with a word representing zero followed by a zero-count word, and non-zero words are unchanged. As an example in [63], a 40-word data sequence

BD000A0000000A000000CB0A0000000000D00000D

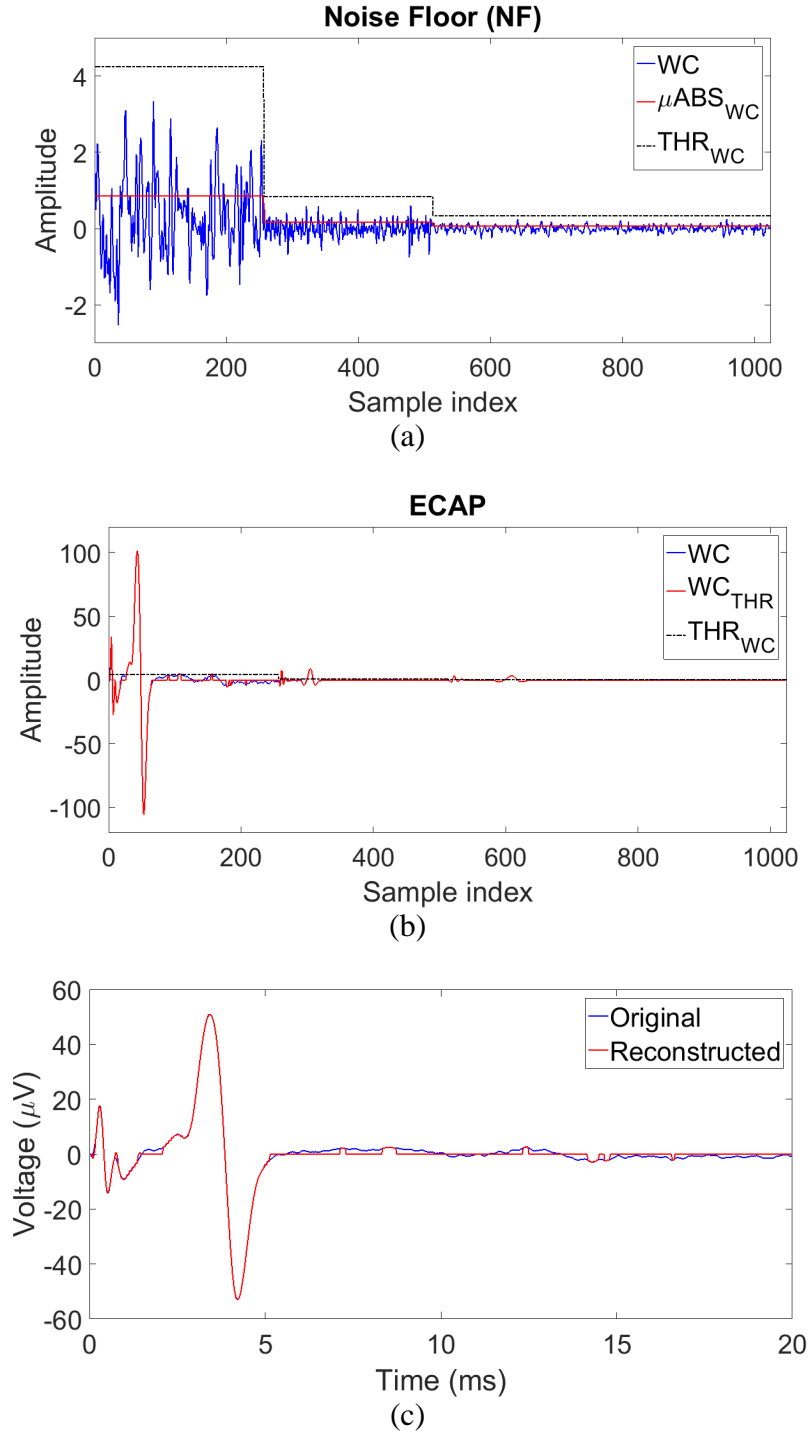


Fig. 4.4 Data compression of an ECAP waveform using 2-level Haar wavelet DWT: (a) derivation of the threshold (THR_{WC}) based on the mean of absolute value of wavelet coefficients (μABS_{WC}), (b) original versus thresholded wavelet coefficients (WC_{THR}) of an ECAP waveform, and (c) original versus reconstructed ECAP waveform.

will be reduced to the sequence "BDX3AX7AX6CB0AX10DX5D" after RLE, where X stands for zero. Given an ECAP waveform with windowing length of 1024, the total number of non-zero elements in the quantized wavelet coefficients in Fig. 4.4 (b) is 175, and the length after RLE is 217 in this example. Assume the precision of each wavelet coefficient is 16 bits. The total size of encoded wavelet coefficients is 3472 bits, which is $4.72\times$ lower than that of ECAP waveform. Fig. 4.4 (c) shows a comparison between original ECAP waveform and reconstructed ECAP waveform after data compression. It can be seen that the shape of nerve fiber responses and its time-axis location are preserved on the reconstructed ECAP waveform.

What needs further study in the data compression of ECAP waveforms is the optimal value of THR_{WC} representing the amplitude threshold of activated fiber responses, and the level of decomposition in DWT that achieves the maximal compression of an ECAP whereas maintains the latency and shape of fiber responses. Moreover, to generate an optimal compression, it's desirable to find a wavelet basis whose shape resembles the signal to be compressed so that the original can be reconstructed with the fewest nonzero wavelet coefficients. So far, studies have shown the Symlets 4 is the optimal wavelet basis for compression of neural spikes, whose shape best matches that of spikes and requires moderate computations [137]. The wavelet basis that best matches ECAP waveforms thus deserves investigation. For implementation of real-time DSP, several approaches to designing the VLSI architecture of DWT have been presented, including the pyramid algorithm [138, 139] and lifting scheme [140-142]. Implementing RLE in VLSI architecture is also feasible and an implementation example can be found in [63].

4.2.3 Implantable Wireless Device

To make the wearable wireless device presented in Section 3.4 chronically implantable, it must be combined with the WPT strategies to eliminate the need for constant battery replacement [143]. Fig. 4.5 illustrates the block diagram of the PMU which utilizes a combination of both WPT and a rechargeable battery as its power supply. The powering coil receives the power from electromagnetic fields, and the wireless receiver converts the received electromagnetic power into a regulated DC voltage output. The DC voltage output of the wireless receiver is fed to a battery charger that can charge a 3.7-V lithium-ion rechargeable coin-cell battery, power the system, or both. The powering coil, wireless receiver, and the battery charger are all available in

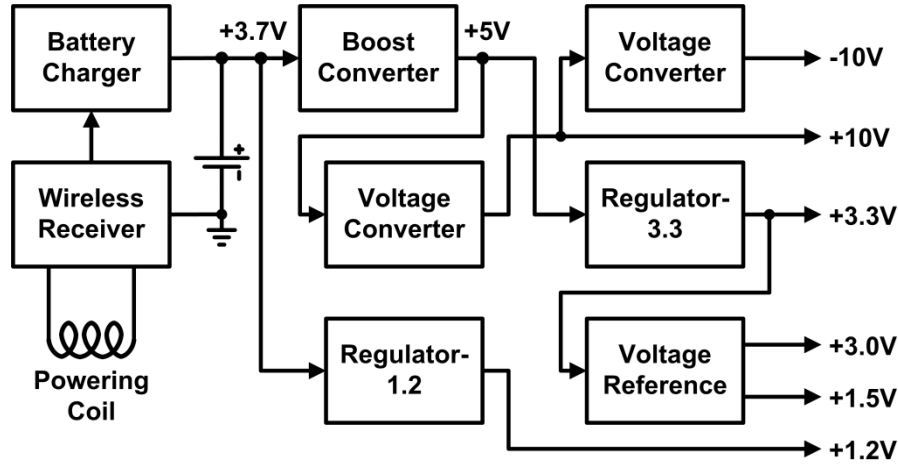
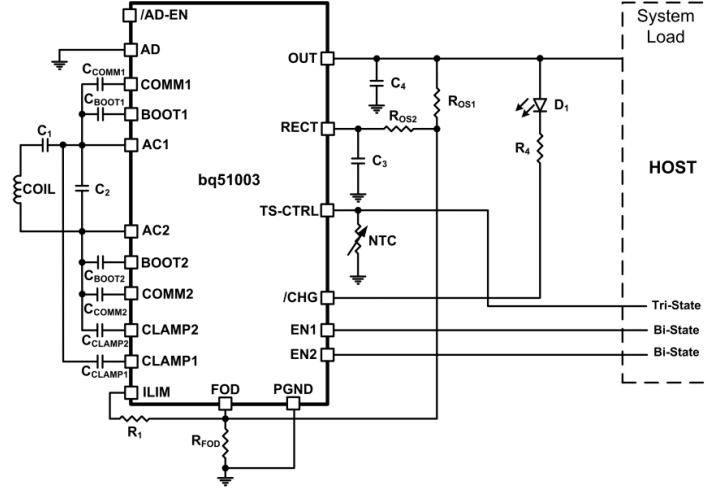
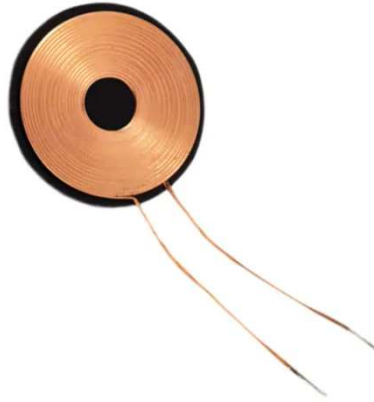


Fig. 4.5 Block diagram of the PMU with a combination of both WPT and a rechargeable battery as its power supply.

COTS components. Fig. 4.6 (a) shows the schematic of the bq51003 (*Texas Instruments*) wireless power supply receiver as an example [144], which utilizes near-field inductive coupling for WPT. A receiver coil for the bq51003 wireless receiver is shown in Fig. 4.6 (b) (*Würth Elektronik Group*), whose total size is 15 mm (diameter) \times 0.6 mm (height). Fig. 4.6 (c) shows the bq500212AEVM-550 wireless power transmitter evaluation module (*Texas Instruments*), which uses a 5-V USB port as the power supply and is compatible with the bq51003 wireless receiver [145]. Combined with the receiver coil in Fig. 4.6 (b), the bq51003 wireless receiver provides 5-V regulated output voltage at 500-mA loading current. The schematic of an example lithium-ion battery charger bq21040 (*Texas Instruments*) is shown in Fig. 4.6 (d). This battery charger has an input voltage supply range from 3.5 V to 28 V, and provides up to 800-mA charging current at 4.2-V regulated voltage output [146] which can charge the 3.7-V lithium-ion battery and supply the system loading. On the other hand, a newly proposed WPT technique, named cavity resonator based WPT [147], employs a WPT chamber with circulating electromagnetic fields as the primary power transmitter and a biaxial receiver coil system to enable the wireless powering of devices implanted in free-moving animals. This technique has been adopted in the design of the Bionode, a closed-loop neuromodulation device [55], and is also applicable to the powering of the wireless device presented in Section 3.4.



(a)



(b)



(c)

Fig. 4.6 (a) Schematic of the bq51003 (*Texas Instruments*) wireless power supply receiver [144]. (b) A receiving coil for the bq51003 (*Würth Elektronik Group*) with total size of 15 mm (diameter) \times 0.6 mm (height). (c) Top view of the bq500212AEVM-550 wireless power transmitter evaluation module (*Texas Instruments*) as the primary wireless power supply [145]. (d) Schematic of the bq21040 (*Texas Instruments*) lithium-ion battery charger [146].

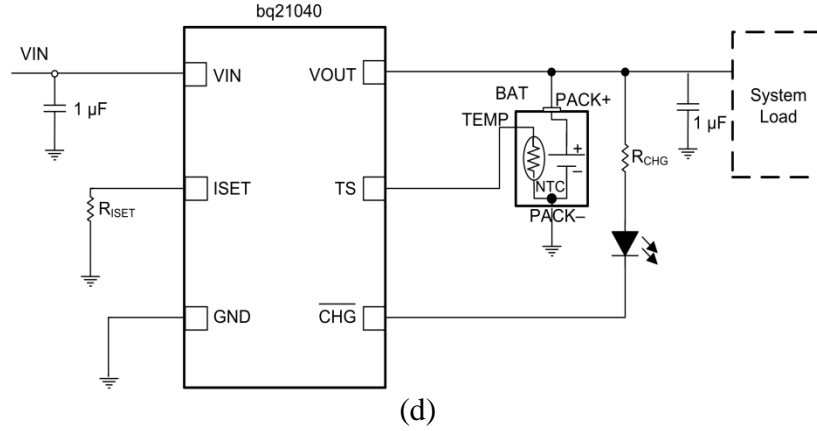
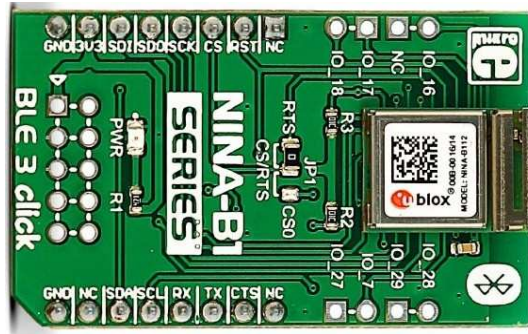
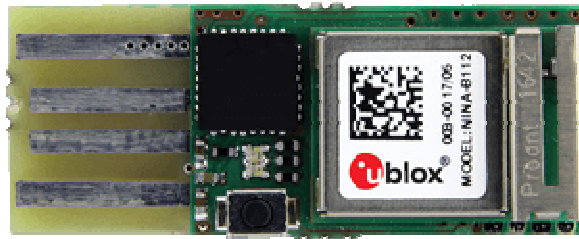


Fig. 4.6 continued.

The power consumption of the RF module in Fig. 3.1 can be further reduced to facilitate the realization of an implantable device by using other low-power Bluetooth transceivers. Fig. 4.7 (a) shows the MIKROE-2471 BLE 3 Click (*MikroElektronika*) [148] as an example, which is built



(a)



(b)

Fig. 4.7 Top view of (a) the MIKROE-2471 BLE 3 Click (*MikroElektronika*) [148] and (b) the B204 USB dongle [150], both built with the NINA-B112 Bluetooth 4.2 module.

Table 4.1 Comparison of the low-power COTS Bluetooth modules [151]

Transceiver	NINA-B112	BL652	ZL70103	CC2640R2F	BMD-350
Energy Efficiency (nJ/b)	9-19	9-19	15-19	11-23	12-25
I_{DC} at 0 dBm (mA)	5.3	5.3	5.3	6.1	7.1
VDD (V)	1.7-3.6	1.7-3.6	2.8-3.5	1.8-3.6	1.7-3.6
Duplex	N/R	Full	Half	Full	Full
Physical size (mm ³)	14×10×4	14×10×2	6×5×2	7×7×2	8.7×6.4×1.5
Antenna	Internal	Internal	External	External	Internal
Max DR (Mbps)	1	1	0.8	1	2

with the NINA-B112 (*u-blox*) Bluetooth Low-Energy (BLE) module and can also be mounted on the WSP board in Fig. 3.9. The NINA-B112 BLE module featuring Bluetooth 5.0 standard has a maximum data rate of 1 Mbps, a module size of 14 mm × 10 mm (including the on-chip antenna) a supply voltage range of 1.7-3.6 V, and a current consumption of 5.3 mA at 0-dBm transmitter power [149]. It also supports serial communication via the UART interface. Fig. 4.7 (b) shows the B204 USB dongle (*u-blox*) [150]; it also uses the NINA-B112 BLE module, provides access to UART over USB, and thus can serve as the base station in Fig. 3.1. A comparison of the low-power COTS Bluetooth modules can be found in [151] and is given in Table 4.1. Furthermore, the Bluetooth transceiver on the WSP board can be programmed in standby mode by default, where the current consumption is only a few microamperes (e.g., 2.2 μ A for NINA-B112), and waken up by the dedicated control pins on FREE for data transmission.

All the boards in the PCB prototype shown by Fig. 3.11 can be further miniaturized through PCB layout, and integrated in a small package for implants using the rigid-flex PCB technology [110] or the PCB assembly technique presented in [55].

REFERENCES

- [1] J. K. Mai, and G. Paxinos, *The Human Nervous System*: Elsevier Science, 2011.
- [2] M. Hallett, "Transcranial magnetic stimulation and the human brain," *Nature*, vol. 406, no. 6792, pp. 147, 2000.
- [3] M. A. Nitsche, L. G. Cohen, E. M. Wassermann, A. Priori, N. Lang, A. Antal, W. Paulus, F. Hummel, P. S. Boggio, and F. Fregni, "Transcranial direct current stimulation: state of the art 2008," *Brain stimulation*, vol. 1, no. 3, pp. 206-223, 2008.
- [4] E. S. Krames, P. H. Peckham, A. Rezai, and F. Aboelsaad, "What is neuromodulation?," *Neuromodulation*, pp. 3-8: Elsevier, 2009.
- [5] FDA, "Medtronic Activa tremor control system P960009," [Online]. Available: http://www.accessdata.fda.gov/cdrh_docs/pdf/p960009.pdf.
- [6] M. L. Kringelbach, N. Jenkinson, S. L. F. Owen, and T. Z. Aziz, "Translational principles of deep brain stimulation," *Nature Reviews Neuroscience*, vol. 8, no. 8, pp. 623-635, 2007/08/01, 2007.
- [7] J. S. Perlmuter, and J. W. Mink, "Deep Brain Stimulation," *Annual Review of Neuroscience*, vol. 29, no. 1, pp. 229-257, 2006/07/21, 2006.
- [8] T. Cameron, "Safety and efficacy of spinal cord stimulation for the treatment of chronic pain: a 20-year literature review," *Journal of Neurosurgery: Spine*, vol. 100, no. 3, pp. 254-267, 2004.
- [9] R. B. North, D. H. Kidd, M. Zahurak, C. S. James, and D. M. Long, "Spinal Cord Stimulation for Chronic, Intractable Pain: Experience over Two Decades," *Neurosurgery*, vol. 32, no. 3, pp. 384-395, 1993.
- [10] P. L. Gildenberg, "History of Electrical Neuromodulation for Chronic Pain," *Pain Medicine*, vol. 7, no. suppl_1, pp. S7-S13, 2006.
- [11] M. Hariz, P. Blomstedt, and L. Zrinzo, "Future of brain stimulation: New targets, new indications, new technology," *Movement Disorders*, vol. 28, no. 13, pp. 1784-1792, 2013/11/01, 2013.
- [12] P. A. Spagnolo, and D. Goldman, "Neuromodulation interventions for addictive disorders: challenges, promise, and roadmap for future research," *Brain*, vol. 140, no. 5, pp. 1183-1203, 2016.
- [13] D. Maurer, "Transcutaneous stimulator and stimulation method," US Patent 3817254, 1974.
- [14] F. Rattay, *Electrical Nerve Stimulation: Theory, Experiments and Applications*: Springer Science & Business Media, 2013.
- [15] K. E. Nnoaham, and J. Kumbang, "Transcutaneous electrical nerve stimulation (TENS) for chronic pain," *Cochrane Database of Systematic Reviews*, no. 3, 2008.
- [16] D. M. Walsh, T. E. Howe, M. I. Johnson, F. Moran, and K. A. Sluka, "Transcutaneous electrical nerve stimulation for acute pain," *Cochrane Database of Systematic Reviews*, no. 2, 2009.
- [17] E. Agostoni, J. E. Chinnock, M. D. B. Daly, and J. G. Murray, "Functional and histological studies of the vagus nerve and its branches to the heart, lungs and abdominal viscera in the cat," *The Journal of physiology*, vol. 135, no. 1, pp. 182-205, 1957.
- [18] H.-R. Berthoud, and W. L. Neuhuber, "Functional and chemical anatomy of the afferent vagal system," *Autonomic Neuroscience*, vol. 85, no. 1, pp. 1-17, 2000/12/20/, 2000.

- [19] C. A. Edwards, A. Kouzani, K. H. Lee, and E. K. Ross, "Neurostimulation Devices for the Treatment of Neurologic Disorders," *Mayo Clinic Proceedings*, vol. 92, no. 9, pp. 1427-1444, 2017/09/01/, 2017.
- [20] F. R. Carreno, and A. Frazer, "Vagal Nerve Stimulation for Treatment-Resistant Depression," *Neurotherapeutics*, vol. 14, no. 3, pp. 716-727, 2017/07/01, 2017.
- [21] D. A. Groves, and V. J. Brown, "Vagal nerve stimulation: a review of its applications and potential mechanisms that mediate its clinical effects," *Neuroscience & Biobehavioral Reviews*, vol. 29, no. 3, pp. 493-500, 2005/05/01/, 2005.
- [22] S. C. Schachter, and C. B. Saper, "Vagus Nerve Stimulation," *Epilepsia*, vol. 39, no. 7, pp. 677-686, 1998/07/01, 1998.
- [23] E. B. Dalkilic, "Neurostimulation devices used in treatment of epilepsy," *Current treatment options in neurology*, vol. 19, no. 2, pp. 7, 2017.
- [24] F. T. Sun, and M. J. Morrell, "Closed-loop neurostimulation: the clinical experience," *Neurotherapeutics*, vol. 11, no. 3, pp. 553-563, 2014.
- [25] A. O. Hebb, J. J. Zhang, M. H. Mahoor, C. Tsiokos, C. Matlack, H. J. Chizeck, and N. Pouratian, "Creating the Feedback Loop: Closed-Loop Neurostimulation," *Neurosurgery Clinics*, vol. 25, no. 1, pp. 187-204, 2014.
- [26] B. Rosin, M. Slovik, R. Mitelman, M. Rivlin-Etzion, Suzanne N. Haber, Z. Israel, E. Vaadia, and H. Bergman, "Closed-Loop Deep Brain Stimulation Is Superior in Ameliorating Parkinsonism," *Neuron*, vol. 72, no. 2, pp. 370-384, 2011/10/20/, 2011.
- [27] M.-C. Lo, and A. S. Widge, "Closed-loop neuromodulation systems: next-generation treatments for psychiatric illness," *International Review of Psychiatry*, vol. 29, no. 2, pp. 191-204, 2017.
- [28] V. Nagaraj, S. T. Lee, E. Krook-Magnuson, I. Soltesz, P. Benquet, P. P. Irazoqui, and T. I. Netoff, "Future of Seizure Prediction and Intervention: Closing the Loop," *Journal of Clinical Neurophysiology*, vol. 32, no. 3, 2015.
- [29] S. Ramgopal, S. Thome-Souza, M. Jackson, N. E. Kadish, I. Sánchez Fernández, J. Klehm, W. Bosl, C. Reinsberger, S. Schachter, and T. Loddenkemper, "Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy," *Epilepsy & Behavior*, vol. 37, pp. 291-307, 2014/08/01/, 2014.
- [30] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic Seizure Detection in EEGs Using Time-Frequency Analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703-710, 2009.
- [31] M. Zijlmans, D. Flanagan, and J. Gotman, "Heart Rate Changes and ECG Abnormalities During Epileptic Seizures: Prevalence and Definition of an Objective Clinical Sign," *Epilepsia*, vol. 43, no. 8, pp. 847-854, 2002/08/01, 2002.
- [32] A. L. I. Shoeb, T. Pang, J. Guttag, and S. Schachter, "NON-INVASIVE COMPUTERIZED SYSTEM FOR AUTOMATICALLY INITIATING VAGUS NERVE STIMULATION FOLLOWING PATIENT-SPECIFIC DETECTION OF SEIZURES OR EPILEPTIFORM DISCHARGES," *International Journal of Neural Systems*, vol. 19, no. 03, pp. 157-172, 2009/06/01, 2009.
- [33] W. Bouthour, P. Mégevand, J. Donoghue, C. Lüscher, N. Birbaumer, and P. Krack, "Biomarkers for closed-loop deep brain stimulation in Parkinson disease and beyond," *Nature Reviews Neurology*, vol. 15, no. 6, pp. 343-352, 2019/06/01, 2019.

- [34] M. Parastarfeizabadi, and A. Z. Kouzani, "Advances in closed-loop deep brain stimulation devices," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, pp. 79, 2017/08/11, 2017.
- [35] M. P. Ward, K. Y. Qing, K. J. Otto, R. M. Worth, S. W. M. John, and P. P. Irazoqui, "A Flexible Platform for Biofeedback-Driven Control and Personalization of Electrical Nerve Stimulation Therapy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 475-484, May 2015.
- [36] L. Squire, D. Berg, F. E. Bloom, S. Du Lac, A. Ghosh, and N. C. Spitzer, *Fundamental Neuroscience (4th Edition)*: Academic Press, 2013.
- [37] H. S. Gasser, "The Classification of Nerve Fibers," *The Ohio Journal of Science*, vol. 41, pp. 145-149, 1941.
- [38] S. E. Krahel, "Vagus nerve stimulation for epilepsy: A review of the peripheral mechanisms," *Surgical neurology international*, vol. 3, no. Suppl 1, pp. S47-S52, 2012.
- [39] W. H. Organization, "Deafness and hearing loss," [Online]. Available: <https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss>, Mar. 2020.
- [40] F. G. Zeng, S. Rebscher, W. Harrison, X. Sun, and H. Feng, "Cochlear Implants: System Design, Integration, and Evaluation," *IEEE Reviews in Biomedical Engineering*, vol. 1, pp. 115-142, Nov. 2008.
- [41] F. A. Spelman, "The past, present, and future of cochlear prostheses," *IEEE Engineering in Medicine and Biology Magazine*, vol. 18, no. 3, pp. 27-33, May-Jun. 1999.
- [42] Y. Brand, P. Senn, N. Dillier, M. Kompis, and J. H. J. Allum, "Cochlear implantation in children and adults in Switzerland," *Swiss medical weekly*, vol. 144, no. w13909, 2014.
- [43] N. Dillier, W. Lai, B. Almqvist, C. Frohne, J. Muller-Deile, M. Stecker, and E. von Wallenberg, "Measurement of the electrically evoked compound action potential via a neural response telemetry system," *Annals of Otology Rhinology and Laryngology*, vol. 111, no. 5, pp. 407-414, May 2002.
- [44] P. M. Carter, A. R. Fisher, T. M. Nygard, B. A. Swanson, R. K. Shepherd, M. Tykocinski, and M. Brown, "Monitoring the electrically evoked compound action potential by means of a new telemetry system," *Annals of Otology, Rhinology and Laryngology*, vol. 104, no. 9 (Suppl. 166), pp. 48-51, 1995.
- [45] C. J. Brown, P. J. Abbas, and B. Gantz, "Electrically evoked whole-nerve action potentials: Data from human cochlear implant users," *Journal of the Acoustical Society of America*, vol. 88, no. 3, pp. 1385-1391, 1990.
- [46] C. J. Brown, P. J. Abbas, and B. J. Gantz, "Preliminary experience with neural response telemetry in the nucleus CI24M cochlear implant," *Otology & Neurotology*, vol. 19, no. 3, pp. 320-327, May 1998.
- [47] D. Cafarelli Dees, N. Dillier, W. K. Lai, E. von Wallenberg, B. van Dijk, F. Akdas, M. Aksit, C. Batman, A. Beynon, S. Burdo, J. M. Chanal, L. Collet, M. Conway, C. Coudert, L. Craddock, H. Cullington, N. Deggouj, B. Frayssé, S. Grabel, J. Kiefer, J. G. Kiss, T. Lenarz, A. Mair, S. Maune, J. Müller -Deile, J. P. Piron, S. Razza, C. Tasche, H. Thai-Van, F. Toth, E. Truy, A. Uziel, and G. F. Smoorenburg, "Normative Findings of Electrically Evoked Compound Action Potential Measurements Using the Neural Response Telemetry of the Nucleus CI24M Cochlear Implant System," *Audiology and Neurotology*, vol. 10, no. 2, pp. 105-116, 2005.

- [48] W. K. Purves, D. Sadava, G. H. Orians, and H. C. Heller, *Life: The Science of Biology*, 4th ed.: Sinauer Associates Inc, 1994.
- [49] J. L. Parker, N. H. Shariati, and D. M. Karantonis, "Electrically evoked compound action potential recording in peripheral nerves," *Bioelectronics in Medicine*, vol. 1, no. 1, pp. 71-83, 2018.
- [50] S. He, H. F. B. Teagle, and C. A. Buchman, "The Electrically Evoked Compound Action Potential: From Laboratory to Clinic," *Frontiers in Neuroscience*, vol. 11, pp. 339, 2017.
- [51] K. C. McGill, K. L. Cummins, L. J. Dorfman, B. B. Berlizot, K. Luetkemeyer, D. G. Nishimura, and B. Widrow, "On the nature and elimination of stimulus artifact in nerve signals evoked and recorded using surface electrodes," *IEEE Transactions on Biomedical Engineering*, vol. BME-29, no. 2, pp. 129-137, Feb. 1982.
- [52] O. Rempelman, and H. H. Ros, "Coherent averaging technique: A tutorial review Part 1: Noise reduction and the equivalent filter," *Journal of Biomedical Engineering*, vol. 8, no. 1, pp. 24-29, Jan, Jan. 1986.
- [53] D. T. T. Plachta, N. Espinosa, M. Gierthmuehlen, O. Cota, T. C. Herrera, and T. Stieglitz, "Detection of baroreceptor activity in rat vagal nerve recording using a multi-channel cuff-electrode and real-time coherent averaging." pp. 3416-3419.
- [54] W. W. Surwillo, "Recovery of the cortical evoked potential from auditory stimulation in children and adults," *Developmental Psychobiology*, vol. 14, no. 1, pp. 1-12, 1981/01/01, 1981.
- [55] D. J. Pederson, C. J. Quinkert, M. A. Arafat, J. P. Somann, J. D. Williams, R. A. Bercich, Z. Wang, G. O. Albors, J. G. R. Jefferys, and P. P. Irazoqui, "The bionode: a closed-loop neuromodulation implant," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 18, no. 1, pp. 1-20, 2019.
- [56] Y. Lo, Y. Kuan, S. Culaclii, B. Kim, P. Wang, C. Chang, J. A. Massachi, M. Zhu, K. Chen, P. Gad, V. R. Edgerton, and W. Liu, "A Fully Integrated Wireless SoC for Motor Function Recovery After Spinal Cord Injury," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 3, pp. 497-509, 2017.
- [57] H.-G. Rhew, J. Jeong, J. A. Fredenburg, S. Dodani, P. G. Patil, and M. P. Flynn, "A fully self-contained logarithmic closed-loop deep brain stimulation SoC with wireless telemetry and wireless power management," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2213-2227, 2014.
- [58] A. Bagheri, S. R. I. Gabran, M. T. Salam, J. L. P. Velazquez, R. R. Mansour, M. M. A. Salama, and R. Genov, "Massively-Parallel Neuromonitoring and Neurostimulation Rodent Headset With Nanotextured Flexible Microelectrodes," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, no. 5, pp. 601-609, 2013.
- [59] D. Loi, C. Carboni, G. Angius, G. N. Angotzi, M. Barbaro, L. Raffo, S. Raspopovic, and X. Navarro, "Peripheral Neural Activity Recording and Stimulation System," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, no. 4, pp. 368-379, 2011.
- [60] R. A. Bercich, D. R. Duffy, and P. P. Irazoqui, "Far-Field RF Powering of Implantable Devices: Safety Considerations," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2107-2112, 2013.
- [61] V. Karkare, S. Gibson, and D. Marković, "A 130- μ W, 64-Channel Neural Spike-Sorting DSP Chip," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 5, pp. 1214-1222, 2011.
- [62] J. Lee, Y. Su, and C. Shen, "A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi." pp. 46-51.

- [63] Y. Yang, A. M. Kamboh, and A. J. Mason, "A configurable realtime DWT-based neural data compression and communication VLSI system for wireless implants," *Journal of Neuroscience Methods*, vol. 227, pp. 140-150, 2014/04/30/, 2014.
- [64] X. Liu, M. Zhang, B. Subei, A. G. Richardson, T. H. Lucas, and J. V. d. Spiegel, "The PennBMBI: Design of a General Purpose Wireless Brain-Machine-Brain Interface System," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 2, pp. 248-258, 2015.
- [65] D. Pani, G. Barabino, L. Citi, P. Meloni, S. Raspopovic, S. Micera, and L. Raffo, "Real-Time Neural Signals Decoding onto Off-the-Shelf DSP Processors for Neuroprosthetic Applications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 9, pp. 993-1002, 2016.
- [66] A. V. Oppenheim, and R. W. Schaffer, *Discrete-time Signal Processing*, 3rd ed.: Prentice Hall, 2010.
- [67] J. O. Smith, *Introduction to Digital Filters with Audio Applications*: W3K Publishing, 2007.
- [68] W. Liu, K. Vichienchom, M. Clements, S. DeMarco, C. Hughes, E. McGucken, M. Humayun, E. de Juan, J. Weiland, and R. Greenberg, "A neuro-stimulus chip with telemetry unit for retinal prosthetic device," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 10, pp. 1487-1497, Oct. 2000.
- [69] L. F. Heffer, and J. B. Fallon, "A novel stimulus artifact removal technique for high-rate electrical stimulation," *Journal of Neuroscience Methods*, vol. 170, no. 2, pp. 277-284, May 2008.
- [70] A. E. Hines, P. E. Crago, G. J. Chapman, and C. Billian, "Stimulus artifact removal in EMG from muscles adjacent to stimulated muscles," *Journal of Neuroscience Methods*, vol. 64, no. 1, pp. 55-62, Jan. 1996.
- [71] H. Liang, and Z. Lin, "Stimulus artifact cancellation in the serosal recordings of gastric myoelectric activity using wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 7, pp. 681-688, Jul. 2002.
- [72] D. T. O'Keeffe, G. M. Lyons, A. E. Donnelly, and C. A. Byrne, "Stimulus artifact removal using a software-based two-stage peak detection algorithm," *Journal of Neuroscience Methods*, vol. 109, no. 2, pp. 137-145, Aug. 2001.
- [73] D. A. Wagenaar, and S. M. Potter, "Real-time multichannel stimulus artifact suppression by local curve fitting," *Journal of Neuroscience Methods*, vol. 120, no. 2, pp. 113-120, Oct. 2002.
- [74] K. Limnusun, H. Lu, H. J. Chiel, and P. Mohseni, "Real-time stimulus artifact rejection via template subtraction," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 3, pp. 391-400, Jun. 2014.
- [75] A. Mendrela, J. Cho, J. Fredenburg, V. Nagaraj, T. Netoff, M. Flynn, and E. Yoon, "A bidirectional neural interface circuit with active stimulation artifact cancellation and cross-channel common-mode noise suppression," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 4, pp. 955-965, Apr. 2016.
- [76] L. H. M. Mens, "Advances in Cochlear Implant Telemetry: Evoked Neural Responses, Electrical Field Imaging, and Technical Integrity," *Trends in Amplification*, vol. 11, no. 3, pp. 143-159, Sep. 2007.

- [77] R. Charlet de Sauvage, Y. Cazals, J. E. Erre, and J. A. Aran, "Acoustically derived auditory nerve action potential evoked by electrical stimulation: An estimation of the waveform of single unit contribution," *Journal of the Acoustical Society of America*, vol. 73, no. 2, pp. 616-627, 1983.
- [78] B. H. Brown, "Frequency analysis used for interpretation of human nerve action potentials obtained from surface electrodes," *Medical & Biological Engineering*, vol. 6, no. 5, pp. 493-502, 1968.
- [79] P. J. Maccabee, N. F. Hassan, R. Q. Cracco, and J. A. Schiff, "Short latency somatosensory and spinal evoked potentials: power spectra and comparison between high pass analog and digital filter," *Electroencephalography and Clinical Neurophysiology*, vol. 65, no. 3, pp. 177-187, 1986.
- [80] A. B. Wiltschko, G. J. Gage, and J. D. Berke, "Wavelet filtering before spike detection preserves waveform shape and enhances single-unit discrimination," *Journal of Neuroscience Methods*, vol. 173, no. 1, pp. 34-40, 2008/08/15/, 2008.
- [81] P. B. Patil, and M. S. Chavan, "A wavelet based method for denoising of biomedical signal." pp. 278-283.
- [82] S.-W. Chen, and Y.-H. Chen, "Hardware Design and Implementation of a Wavelet De-Noising Procedure for Medical Signal Preprocessing," *Sensors*, vol. 15, no. 10, pp. 26396-26414, 2015.
- [83] P. J. Van Fleet, *Discrete Wavelet Transformations: An Elementary Approach with Applications*, 2nd ed.: John Wiley & Sons, 2011.
- [84] J. S. Hunter, "The exponentially weighted moving average," *Journal of Quality Technology*, vol. 18, no. 4, pp. 203-210, Oct. 1986.
- [85] J. C. Principe, and J. R. Smith, "Design and Implementation of Linear Phase FIR Filters for Biological Signal Processing," *IEEE Transactions on Biomedical Engineering*, vol. BME-33, no. 6, pp. 550-559, 1986.
- [86] J. McClellan, T. Parks, and L. Rabiner, "A computer program for designing optimum FIR linear phase digital filters," *IEEE Transactions on Audio and Electroacoustics*, vol. 21, no. 6, pp. 506-526, Dec. 1973.
- [87] J. McClellan, and T. Parks, "A unified approach to the design of optimum FIR linear-phase digital filters," *IEEE Transactions on Circuit Theory*, vol. 20, no. 6, pp. 697-701, Nov. 1973.
- [88] M. A. Al-Alaoui, "Linear phase low-pass IIR digital differentiators," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 697-706, Feb. 2007.
- [89] S. R. Powell, and P. M. Chau, "A technique for realizing linear phase IIR filters," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2425-2435, Nov. 1991.
- [90] B. Widrow, and I. Kollar, *Quantization Noise: Round-Off Error in Digital Computation, Signal Processing, Control, and Communications.*, New York, NY, USA: Cambridge Univ. Press, 2008.
- [91] B. P. McGovern, R. F. Woods, and C. McAllister, "Optimised multiply/accumulate architecture for very high throughput rate digital filters," *Electronics Letters* vol. 31, no. 14, pp. 1135-1136, 1995.
- [92] K. K. Parhi, *VLSI Digital Signal Processing Systems: Design and Implementation*: John Wiley & Sons, 2007.

- [93] M. A. Basiri M, and N. M. Sk, "Configurable folded IIR filter design," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 12, pp. 1144-1148, Dec. 2015.
- [94] A. Hemdani, M. W. Naouar, I. S. Belkhodja, and E. Monmasson, "Design of a digital IIR filter for active filtering applications," in *Proc. IEEE Int. Mediterranean Electrotechnical Conf.*, Mar. 2012, pp. 1107-1112.
- [95] H. Shousheng, and M. Torkelson, "Designing pipeline FFT processor for OFDM (de)modulation," in 1998 URSI International Symposium on Signals, Systems, and Electronics. Conference Proceedings (Cat. No.98EX167), Pisa, Italy, 1998, pp. 257-262.
- [96] Texas_Instruments, "16-Bit, Ultra-Low Power, Voltage Output Digital-to-Analog Converter," DAC8832 datasheet, Feb. 2006, [Revised Sep. 2007].
- [97] Texas_Instruments, "36-V, Low-Power, Precision, CMOS, Rail-to-Rail Input/Output, Low Offset Voltage, Low Input Bias Current Op Amp," OPAx191 datasheet, Dec. 2015, [Revised Oct. 2019].
- [98] Texas_Instruments, "A comprehensive study of the howland current pump," AN-1515 A, AN-1515 A, Jan. 2008, [Revised Apr. 2013].
- [99] A. J. Loutit, T. Maddess, S. J. Redmond, J. W. Morley, G. J. Stuart, and J. R. Potas, "Characterisation and functional mapping of surface potentials in the rat dorsal column nuclei," *The Journal of physiology*, vol. 595, no. 13, pp. 4507-4524, 2017.
- [100] J.-S. Brittain, and H. Cagnan, "Recent trends in the use of electrical neuromodulation in Parkinson's disease," *Current behavioral neuroscience reports*, vol. 5, no. 2, pp. 170-178, 2018.
- [101] K. Kumar, and S. Rizvi, "Historical and present state of neuromodulation in chronic pain," *Current pain and headache reports*, vol. 18, no. 1, pp. 387, 2014.
- [102] R. S. Fisher, and A. L. Velasco, "Electrical brain stimulation for epilepsy," *Nature Reviews Neurology*, vol. 10, no. 5, pp. 261, 2014.
- [103] P. Afshar, A. Khambhati, S. Stanslaski, D. Carlson, R. Jensen, S. Dani, M. Lazarewicz, J. Giftakis, P. Stypulkowski, and T. Denison, "A translational platform for prototyping closed-loop neuromodulation systems," *Frontiers in neural circuits*, vol. 6, pp. 117, 2013.
- [104] D. M. Labiner, and G. L. Ahern, "Vagus nerve stimulation therapy in depression and epilepsy: therapeutic parameter settings," *Acta neurologica scandinavica*, vol. 115, no. 1, pp. 23-33, 2007.
- [105] K. A. Sluka, and D. Walsh, "Transcutaneous electrical nerve stimulation: basic science mechanisms and clinical effectiveness," *The Journal of pain*, vol. 4, no. 3, pp. 109-121, 2003.
- [106] S. Stanslaski, P. Afshar, P. Cong, J. Giftakis, P. Stypulkowski, D. Carlson, D. Linde, D. Ullestad, A.-T. Avestruz, and T. Denison, "Design and validation of a fully implantable, chronic, closed-loop neuromodulation device with concurrent sensing and stimulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 4, pp. 410-421, 2012.
- [107] A. Zhou, S. R. Santacruz, B. C. Johnson, G. Alexandrov, A. Moin, F. L. Burghardt, J. M. Rabaey, J. M. Carmena, and R. Muller, "A wireless and artefact-free 128-channel neuromodulation device for closed-loop stimulation and recording in non-human primates," *Nature biomedical engineering*, vol. 3, no. 1, pp. 15-26, 2019.

- [108] 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 Journal of Solid-State Circuits*, vol. 50, no. 4, pp. 1038-1047, 2015.
- [109] J. Park, G. Kim, and S. Jung, "A 128-Channel FPGA-Based Real-Time Spike-Sorting Bidirectional Closed-Loop Neural Interface System," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2227-2238, 2017.
- [110] G. Gagnon-Turcotte, Y. LeChasseur, C. Bories, Y. Messaddeq, Y. D. Koninck, and B. Gosselin, "A Wireless Headstage for Combined Optogenetics and Multichannel Electrophysiological Recording," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 1, pp. 1-14, 2017.
- [111] A. Bahmer, O. Peter, and U. Baumann, "Recording and analysis of electrically evoked compound action potentials (ECAPs) with MED-EL cochlear implants and different artifact reduction strategies in Matlab," *Journal of Neuroscience Methods*, vol. 191, no. 1, pp. 66-74, 2010.
- [112] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Network-Computation in Neural Systems*, vol. 9, no. 4, pp. R53-R78, Nov, 1998.
- [113] G. P. Seu, G. N. Angotzi, F. Boi, L. Raffo, L. Berdondini, and P. Meloni, "Exploiting all programmable SoCs in neural signal analysis: A closed-loop control for large-scale CMOS multielectrode arrays," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 4, pp. 839-850, 2018.
- [114] J. A. Undurraga, R. P. Carlyon, J. Wouters, and A. Wieringen, "Evaluating the noise in electrically evoked compound action potential measurements in cochlear implants," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 7, pp. 1912-1923, 2012.
- [115] E. K. Glassman, and M. L. Hughes, "Determining electrically evoked compound action potential thresholds: A comparison of computer versus human analysis methods," *Ear and hearing*, vol. 34, no. 1, pp. 96, 2013.
- [116] A. Botros, B. van Dijk, and M. Killian, "AutoNRT™: An automated system that measures ECAP thresholds with the Nucleus® Freedom™ cochlear implant via machine intelligence," *Artificial intelligence in medicine*, vol. 40, no. 1, pp. 15-28, 2007.
- [117] M. A. Shaeri, and A. M. Sodagar, "A Method for Compression of Intra-Cortically-Recorded Neural Signals Dedicated to Implantable Brain-Machine Interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 485-497, 2015.
- [118] Texas_Instruments, "A comprehensive study of the howland current pump," AN-1515 A, Jan. 2008, [Revised Apr. 2013].
- [119] Microchip_Technology, "Bluetooth Data Module Command Reference & Advanced Information User's Guide ", RN-BT-DATA-UG, Mar. 2013.
- [120] Microchip_Technology, "RN41/RN41N Class 1 Bluetooth Module," RN-41-DS, Nov. 2013.
- [121] Texas_Instruments, "96% Efficient Synchronous Boost Converter With 4A Switch," TPS6103x datasheet, Sep. 2002, [Revised Mar. 2015].
- [122] "WEBENCH® Power Designer," [Online]. Available: <https://webench.ti.com/power-designer/>.
- [123] Texas_Instruments, "CMOS Voltage Converter," TL7660 datasheet, Jun. 2006.

- [124] J. A. Van Alste, and T. S. Schilder, "Removal of Base-Line Wander and Power-Line Interference from the ECG by an Efficient FIR Filter with a Reduced Number of Taps," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 12, pp. 1052-1060, Dec. 1985.
- [125] J. A. van Alsté, W. Van Eck, and O. E. Herrmann, "ECG baseline wander reduction using linear phase filters," *Computers and Biomedical Research*, vol. 19, no. 5, pp. 417-427, Oct. 1986.
- [126] S. Hargittai, "Efficient and fast ECG baseline wander reduction without distortion of important clinical information," in 2008 Computers in Cardiology, Bologna, 2008, pp. 841-844.
- [127] K. Kim, H. Mahmoodi, and K. Roy, "A Low-Power SRAM Using Bit-Line Charge-Recycling," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 446-459, 2008.
- [128] B.-D. Yang, "A Low-Power SRAM Using Bit-Line Charge-Recycling for Read and Write Operations," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 10, pp. 2173-2183, 2010.
- [129] Y.-H. Chen, C.-Y. Li, and T.-Y. Chang, "Area-Effective and Power-Efficient Fixed-Width Booth Multipliers Using Generalized Probabilistic Estimation Bias," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 3, pp. 277-288, 2011.
- [130] L. D. Van, and C.-C. Yang, "Generalized low-error area-efficient fixed-width multipliers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 8, pp. 1608-1619, 2005.
- [131] "IEEE Standard for Floating-Point Arithmetic," *IEEE Std 754-2008*, pp. 1-70, 2008.
- [132] H. Park, Y. Yamanashi, K. Taketomi, N. Yoshikawa, M. Tanaka, K. Obata, Y. Ito, A. Fujimaki, N. Takagi, K. Takagi, and S. Nagasawa, "Design and Implementation and On-Chip High-Speed Test of SFQ Half-Precision Floating-Point Adders," *IEEE Transactions on Applied Superconductivity*, vol. 19, no. 3, pp. 634-639, 2009.
- [133] H. Hara, K. Obata, H. Park, Y. Yamanashi, K. Taketomi, N. Yoshikawa, M. Tanaka, A. Fujimaki, N. Takagi, K. Takagi, and S. Nagasawa, "Design, Implementation and On-Chip High-Speed Test of SFQ Half-Precision Floating-Point Multiplier," *IEEE Transactions on Applied Superconductivity*, vol. 19, no. 3, pp. 657-660, 2009.
- [134] J.-M. Muller, N. Brunie, F. de Dinechin, C.-P. Jeannerod, M. Joldes, V. Lefèvre, G. Melquiond, N. Revol, and S. Torres, *Handbook of Floating-Point Arithmetic*: Springer, 2018.
- [135] R. Karam, S. J. A. Majerus, D. J. Bourbeau, M. S. Damaser, and S. Bhunia, "Tunable and Lightweight On-Chip Event Detection for Implantable Bladder Pressure Monitoring Devices," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 6, pp. 1303-1312, 2017.
- [136] C. K. Chui, *An Introduction to Wavelets*: Elsevier, 2016.
- [137] K. G. Oweiss, "A systems approach for data compression and latency reduction in cortically controlled brain machine interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 7, pp. 1364-1377, 2006.
- [138] M. Vishwanath, R. M. Owens, and M. J. Irwin, "VLSI architectures for the discrete wavelet transform," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 42, no. 5, pp. 305-316, 1995.

- [139] A. Grzeszczak, M. K. Mandal, and S. Panchanathan, "VLSI implementation of discrete wavelet transform," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 4, no. 4, pp. 421-433, 1996.
- [140] H. Chao-Tsung, T. Po-Chih, and C. Liang-Gee, "Flipping structure: an efficient VLSI architecture for lifting-based discrete wavelet transform," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1080-1089, 2004.
- [141] K. G. Oweiss, A. Mason, Y. Suhail, A. M. Kamboh, and K. E. Thomson, "A Scalable Wavelet Transform VLSI Architecture for Real-Time Signal Processing in High-Density Intra-Cortical Implants," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 6, pp. 1266-1278, 2007.
- [142] A. M. Kamboh, M. Raetz, K. G. Oweiss, and A. Mason, "Area-Power Efficient VLSI Implementation of Multichannel DWT for Data Compression in Implantable Neuroprosthetics," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 1, no. 2, pp. 128-135, 2007.
- [143] K. Agarwal, R. Jegadeesan, Y. Guo, and N. V. Thakor, "Wireless Power Transfer Strategies for Implantable Bioelectronics," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 136-161, 2017.
- [144] Texas_Instruments, "Highly Integrated Wireless Receiver Qi (WPC v1.2) Compliant Power Supply," bq51003 datasheet, Dec. 2013, [Revised Jul. 2018].
- [145] Texas_Instruments, "bq500212A bqTESLA Wireless Power TX EVM," Jul. 2013, [Revised May 2016].
- [146] Texas_Instruments, "0.8-A, Single-Input, Single Cell Li-Ion and Li-Pol Battery Charger," bq21040 datasheet, Apr. 2016, [Revised Jan. 2019].
- [147] H. Mei, K. A. Thackston, R. A. Bercich, J. G. R. Jefferys, and P. P. Irazoqui, "Cavity Resonator Wireless Power Transfer System for Freely Moving Animal Experiments," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 775-785, 2017.
- [148] "BLE 3 Click," [Online]. Available: <https://www.mikroe.com/ble-3-click>.
- [149] "NINA-B1 series datasheet," [Online]. Available: <https://www.u-blox.com/en/docs/UBX-15019243>.
- [150] "Blueprint B204 application note," [Online]. Available: <https://www.u-blox.com/en/docs/UBX-17060841>.
- [151] T. Zhan, S. Z. Fatmi, S. Guraya, and H. Kassiri, "A Resource-Optimized VLSI Implementation of a Patient-Specific Seizure Detection Algorithm on a Custom-Made 2.2 cm² Wireless Device for Ambulatory Epilepsy Diagnostics," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1175-1185, 2019.

VITA

Jui-Wei (Johnson) Tsai received the B.S. degree in Engineering and System Science (ESS) with honor in 2007 and M.S. degrees in Electrical Engineering in 2009, all from National Tsing Hua University (NTHU), Hsinchu City, Taiwan. He is currently working toward Ph.D. degree in Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA.

From 2010 to 2013, he was a research assistant of NTHU, Hsinchu City, Taiwan, working on the application of micro-electromechanical system (MEMS) transducers and their integration with electronics systems. In 2013, he joined the Center for Implantable Devices, Purdue University, West Lafayette, IN, USA, as a research assistant. His research interests include algorithms and VLSI architectures for signal processing, ASIC design, applications of MEMS transducers in consumer electronics, and development of wireless biomedical devices.