EPFL

Thèse n° 9082

Resource Trade-Offs in Circuits and Systems: from Neurotechnology to Communications

Présentée le 30 août 2021

Faculté des sciences et techniques de l'ingénieur
Laboratoire de systèmes d'information et d'inférence
Programme doctoral en microsystèmes et microélectronique

pour l'obtention du grade de Docteur ès Sciences

par

Arda URAN

Acceptée sur proposition du jury

Prof. G. De Micheli, président du jury
Prof. V. Cevher, Prof. Y. Leblebici, directeurs de thèse
Prof. A. Emami, rapporteuse
Dr P. Muller, rapporteur
Prof. A. P. Burg, rapporteur

École
polytechnique
fédérale
de Lausanne

2021

The struggle to free myself of restraints
becomes my very shackles.
— Meshuggah, *Disenchantment*, Catch 33

To my family…

Acknowledgments

First and foremost, I would like to express my gratitude to my advisors Prof. Volkan Cevher and Prof. Yusuf Leblebici. I had the privilege of conducting diverse and exciting research projects under their patient guidance throughout my time at EPFL. They always supported my ideas and gave me the freedom to explore them, which influenced my professional and personal development as a researcher and as an engineer. I cannot thank them enough for giving me this unique opportunity.

I would like to extend my sincere thanks to Prof. Azita Emami, Dr. Paul Muller, Prof. Andreas Burg, and Prof. Giovanni De Micheli for serving in my thesis committee and reviewing the work presented here. I am grateful to Prof. Azita Emami also for hosting me at her laboratory in Caltech with great hospitality for three wonderful months. I also wish to thank Prof. Catherine Dehollain and Prof. Stéphanie Lacour for our fruitful collaboration, and Prof. Andreas Burg for generously sharing the test facilities of his laboratory.

I must mention Prof. Armin Tajalli as a *de facto* co-advisor of the second part of this thesis, and thank him for his close involvement. I have learned so much as a designer from his rigorous and systematic approach to high-speed circuits.

This thesis was made possible by the valuable contributions of many collaborators. I would like to begin with thanking Dr. Cosimo Aprile for his strong support during my first year and introducing me to the learning-based compression concepts. Special thanks to Dr. Kerim Türe for boosting the value of the neural recording chip project by adding the wireless capabilities, and also for helping me with my measurements. Many thanks to Dr. Florian Fallegger and Dr. Alix Trouillet for providing the electrodes and performing the surgery for *in vivo* validation.

It was a great pleasure working with Dr. Alain Vachoux since my first year at EPFL as his student, assistant, and colleague. I gratefully acknowledge his timely support for the design tools and the server infrastructure, which were essential for meeting the strict tapeout deadlines. I also thank Dr. Alexandre Schmid for his help with the laboratory organization.

I received invaluable assistance from our electronics technician Sylvain Hauser, who helped me get my PCBs running with countless reworks, and has always been ready to help when I needed. I had the chance to benefit from the experience of Peter Brühlmeier, who designed an excellent RF PCB. I very much appreciate the help of my colleagues Thomas Sanchez and Igor Krawczuk for setting up the machine learning experiments. I would also like to recognize the support of Gosia Baltaian and Patricia Vonlanthen for all administrative matters.

Acknowledgments

The journey towards this thesis has been inherently full of hurdles, and a global pandemic or the departure of an advisor certainly did not make it any easier. Therefore, my biggest thanks are to my wife, friends and family, from whom I found the strength to persevere.

The fun we had with my officemates and true friends, Dr. Cosimo Aprile and Dr. Jonathan Narinx, inside and outside ELD 335, will always remain as the fondest memories of my time at EPFL. I will miss Jonathan saving me from mountain tops on multiple occasions, and I will never forget Cosimo's greatest advice: "it is better to regret, than not to [sic]". Special thanks to Valentine, Rachele, and Victor for joining us occasionally.

I would also like to thank my colleagues at LSM: Dr. Kiarash Gharibdoust, Dr. Can Baltacı, Dr. Jury Sandrini, Dr. Elmira Shahrabi, Dr. Reza Navid Ranjandish, Dr. Cecilia Giovinazzo, Dr. Mustafa Kilic, Dr. Duygu Kostak, Dr. Bilal Demir, and Dr. Gain Kim; and my colleagues at LIONS: Thomas Sanchez, Igor Krawczuk, Dr. Baran Gözcü, Thomas Pethick, Paul Rolland, Fabian Latorre, Leello Dadi, Ahmet Alacaoğlu, Ali Kavis, Fatih Sahin, and Pedro Abranches, for the good times we shared throughout the years.

The work presented here was conducted in three cities: Lausanne, Sion, and Pasadena, where I had the chance to meet awesome people. I would like to thank Dr. Mish Ebrahim, Dr. Giovanni Resta, Dr. Bardiya Valizadeh, and Özge Kadıoğlu for being amazing chefs; Dr. Chris Ireland, Dr. Samantha Anderson, Dr. Kyriakos Stylianou, Dr. Alina Kampouri, Dr. Arunraj Chidambaram, and Dr. Andrzej Gladysiak for treating me as one of their officemates; Peanut, Lila, and Luna for being good dogs; Nancy McGuire for being such a lovely host and flatmate; Benyamin Haghi for our productive gym sessions, and counteracting lunch breaks at Ernie's; and finally, the SBB train crew for welcoming me on the InterRegio 90 every morning and evening, frequently at nights as well.

I do not always have luck, but when I do, I am the luckiest. The biggest luck I had during these years, and in my life in general, has been having my wife Dr. Pelin Kınık (π) always by my side. I could not have reached the end without her helping me to rise back to my feet every time I fell. I am also proud in the sense that our exchange of love notes in high school has evolved into exchanging love notes in our thesis acknowledgments.

I dedicate this thesis to my mother Sevda and grandmother Bingül, from whom I inherited the determination to endure this tough but rewarding journey. Sevgili annem ve anneannem, bu tezi tamamlamak için ihtiyacım olan azmi ve kararlılığı size borçluyum. Tüm özveriniz ve sınırsız desteğiniz için sonsuz teşekkürler. Sizi seviyorum.

Sion, June 22, 2021 Arda Uran

This work was supported by Hasler Foundation Program: Cyber Human Systems (project number 16066), and the Department of the Navy, Office of Naval Research (ONR) under the grant number N62909-17-1-2111.

Abstract

This thesis aims to explore and exploit trade-offs in integrated circuits and systems to overcome the fundamental bottlenecks faced by future data acquisition and communication systems. Specifically, we target the emerging implantable neurotechnology and the advancing high-speed communication domains as case studies.

The first part of the thesis addresses the extreme resource constraints of implantable neurotechnology. As the number of electrodes for neural recording reaches towards thousands, a single-chip solution is rendered infeasible in terms of both energy and area. To tackle this problem, we present an analog front-end architecture that applies various energy- and area-saving techniques to reduce the hardware resource demand of neural recording. The second challenge has been transferring the massive data throughput over a wireless link to allow neural recording in independent environments. Inspired by the previous ideas for compression and the emerging machine learning applications in neuroscience, we introduce an on-chip adaptive feature extraction framework to reduce the overall data rate to the feasible range.

The second part of the thesis focuses on the timing uncertainty problem in high-speed communication systems. The increasing data rates in wireline and wireless transceivers, and the emerging need for multitransceiver architectures necessitate the lowest energy and area for the frequency synthesis, which contradicts the current jitter-cost trade-off paradigm. To allow the use of area-efficient ring oscillators for jitter-sensitive applications under low-power constraints, we introduce a multiphase feedback phase-locked loop architecture, which suppresses the oscillator phase noise over a larger bandwidth.

Keywords:

Implantable neurotechnology, high-density neural recording, analog front-end (AFE), AC coupling, inverter-based low-noise amplifier (LNA), analog-to-digital converter (ADC), successive approximation register (SAR), monotonic switching, unit-length capacitor (ULC), on-chip compression, compressed Hadamard transform (CHT), wireless power and data transfer, machine learning, seizure detection, joint training, frequency synthesizer, phase-locked loop (PLL), loop bandwidth, Gardner's limit, multiphase feedback (MPF), voltage controlled ring oscillator, matrix phase detector (MPD).

Résumé

Cette thèse vise à explorer et exploiter des compromis dans les circuits et systèmes intégrés pour surmonter les goulots d'étranglement fondamentaux auxquels sont confrontés les futurs systèmes d'acquisition de données et de communication. Plus précisément, nous ciblons les applications émergentes dans les domaines de la neurotechnologie implantable et celui des communications à haut débit comme études de cas.

La première partie de la thèse aborde les contraintes de ressources extrêmes de la neurotechnologie implantable. Comme le nombre d'électrodes pour l'enregistrement neuronal atteint des milliers, une solution à puce électronique unique est rendue irréalisable en termes d'énergie et de surface. Pour résoudre ce problème, nous présentons une architecture frontale analogique qui applique diverses techniques d'économie d'énergie et de surface pour réduire la demande en ressources matérielles de l'enregistrement neuronal. Le deuxième défi a été de transférer l'énorme débit de données sur une liaison sans fil pour permettre l'enregistrement neuronal dans des environnements indépendants. Inspiré par les idées précédentes pour la compression et les applications émergentes d'apprentissage automatique en neuroscience, nous introduisons un cadre d'extraction de caractéristiques sur puce pour réduire le débit de données global à la plage possible.

La deuxième partie de la thèse se concentre sur le problème de l'incertitude temporelle dans les systèmes de communication à haut débit. L'augmentation des débits de données dans les émetteurs-récepteurs filaires et sans fil, ainsi que le besoin émergent d'architectures multi-émetteurs-récepteurs nécessitent l'énergie et la surface les plus basses possibles pour la synthèse de fréquences, ce qui contredit le paradigme actuel de compromis gigue versus coût. Pour permettre l'utilisation d'oscillateurs en anneau de faible surface pour les applications sensibles à la gigue sous des contraintes de faible puissance, nous introduisons une architecture de boucle à verrouillage de phase à rétroaction multiphase et implémentons un synthétiseur de fréquence à large bande.

Mots clés : Neurotechnologie implantable, enregistrement neuronal haute densité, circuit frontal analogique, couplage AC, amplificateur à faible bruit, convertisseur analogique-numérique, registre d'approximation successive, commutation monotone , condensateur de longueur unitaire, compression sur puce, transformée de Hadamard compressée, puissance sans fil et transfert de données, apprentissage automatique, détection de crise, entraînement conjoint, synthétiseur de fréquence, boucle à verrouillage de phase, bande passante de boucle, Limite de Gardner, rétroaction multiphase, oscillateur en anneau commandé en tension, détecteur de phase à matrice.

Contents

List of Figures

List of Tables

List of Acronyms

ADC	Analog-to-digital converter
AFE	Analog front end
ANN	Artificial neural network
AP	Action potential
BER	Bit error rate
BGR	Bandgap reference
BJT	Bipolar junction transistor
BMI	Brain-machine interface
CC-LNA	Capacitively-coupled low-noise amplifier
CDAC	Capacitive digital-to-analog converter
CHT	Compressed Hadamard transform
CMFB	Common-mode feedback
CMOS	Complementary metal-oxide-semiconductor
CMRR	Common-mode rejection ratio
CP	Charge pump
CR	Compression ratio
CS	Compressed sensing
DAC	Digital-to-analog converter
DBS	Deep brain stimulation
DCT	Discrete cosine transform
DFT	Discrete Fourier transform
DNN	Deep neural network
DWT	Discrete wavelet transform
E-A FoM	Energy-area figure of merit
ECG	Electrocardiography
ECoG	Electrocorticography
EEG	Electroencephalography
ENOB	Effective number of bits
FAR	False alarm rate
FDIV	Feedback divider
FFT	Fast Fourier transform
FIR	Finite impulse response

FoM	Figure of merit
FSK	Frequency-shift keying
HFO	High-frequency oscillation
IC	Integrated circuit
iEEG	Intracranial electroencephalography
IR-UWB	Impulse radio ultra-wideband
IRN	Input-referred noise
JTF	Jitter transfer function
LA	Lock assist
LBCS	Learning-based compressive subsampling
LDO	Low-dropout
LF	Loop filter
LFP	Local field potential
LNA	Low-noise amplifier
LO	Local oscillator
LSB	Least significant bit
MCS	Multichannel compressed sensing
MEA	Microelectrode array
ML	Machine learning
MOM	Metal-oxide-metal
MOS	Metal-oxide-semiconductor
MPD	Matrix phase detector
MPD+CP	Matrix phase detector and charge pump
MPF	Multiphase feedback
MRI	Magnetic resonance imaging
MSB	Most significant bit
MUA	Multiunit spiking activity
NB	Narrowband
NEF	Noise efficiency factor
NMOS	N-channel metal-oxide-semiconductor
OOK	On-off keying
OTA	Operational transconductance amplifier
PBS	Phosphate-buffer saline
PCB	Printed circuit board
PD	Phase detector
PD+CP	Phase detector and charge pump
PEF	Power efficiency factor
PFD	Phase-frequency detector
PLL	Phase-locked loop
PMOS	P-channel metal-oxide-semiconductor
PRBS	Pseudorandom binary sequence
PSD	Power spectral density

PTAT	Proportional to absolute temperature
PVT	Process-voltage-temperature
PXF	Periodic transfer function
RIP	Restricted isometry property
RNN	Recursive neural network
RX	Receiver
SAR	Successive approximation register
sEEG	Stereotaxic electroencephalography
SFDR	Spurious-free dynamic range
SHS	Structured Hadamard sampling
SNR	Signal-to-noise ratio
SoC	System-on-chip
SOZ	Seizure onset zone
SST	Source-series terminated
SVM	Support vector machine
THD	Total harmonic distortion
TX	Transmitter
ULC	Unit-length capacitor
ULCDAC	Unit-length capacitor array digital-to-analog converter
UWB	Ultra-wideband
VCO	Voltage-controlled oscillator
WHO	World Health Organization
WPDT	Wireless power and data transfer
WPT	Wireless power transfer
XTAL	Crystal oscillator

Introduction

Integrated circuits (ICs) have benefited from the downscaling complementary metal-oxide-semiconductor (CMOS) technology for many years as Moore's Law predicted. The cost per transistor has decreased steadily as dies got smaller and wafers got bigger. The transistors have become faster, and the scaling supply voltage has reduced the energy consumption. As a result, more functionality could be integrated in a single chip with each new technology.

The progress in the semiconductor industry has been the epicenter for advancements in almost every other industry. In healthcare, for example, the advanced capabilities of wearable and implantable medical devices have enabled better treatments with higher patient comfort. In communications, the increasing speed of data transmission has allowed massive amounts of information to be instantly accessible.

Today the scaling trend has neared its end due to several technical and economical difficulties. Downscaling effort in the nanometer regime is disproportionately expensive, and small geometry effects adversely impact circuit performance, especially for analog circuits. The ambition for better technologies, on the other hand, has not seized to grow. Consequently, the new emphasis is on innovating new circuits and systems to sustain the progress. The goal of this thesis is to explore and exploit opportunities in integrated circuits and systems to solve the technical bottlenecks in two different application domains.

The first domain is implantable neurotechnology, which is extensively reviewed in Chapter 1. The primary research goal here is to remain extremely resource-frugal while extracting a large amount of information from the brain. Chapter 2 presents a resource-efficient analog front-end for maximizing the channel count within a single die [1]. Chapter 3 proposes a recording framework which leverages machine learning to reduce the data rate for wireless operation [2].

The second domain is high-speed communications, specifically frequency synthesis for transceivers. The main research goal here is to make resource-efficient ring oscillators feasible for jitter-sensitive applications. Chapter 4 introduces trade-offs in phase-locked loops, and describes a technique for maximizing the loop bandwidth to mitigate the oscillator noise. Chapter 5 presents the implementation details and the measured performance of the prototype [3]. We conclude the thesis with a summary of the main contributions and an outlook.

Implantable Neurotechnology Part I

1 Neural Recording Implants: A Review

The electrical nature of the nervous system was discovered in the late 18th century, when Luigi Galvani (1737-1798) observed the contraction of a frog's leg muscle when its exposed leg nerve was struck by an electrical spark [4]. The discovery was elaborated with Emil du Bois-Reymond's (1818-1896) capture of *the action potential* using a *galvanometer* [5], an instrument which converts electric current into mechanical motion by means of a magnetic coil. However, due to its large inertia, the galvanometer was not able to measure this rapid signal accurately. This may be considered as the beginning of the quest for a better neural interface.

The initial solutions for capturing the neural signals were also electromechanical. Julius Bernstein (1839-1917) improved the measurements by incorporating sampling action into the galvanometer [6]. The fundamental problem of inertia, was tackled by Willem Einthoven (1860-1927) with *the string galvanometer*, where the slow coil was replaced by a light-weight wire stretched between two poles of an electromagnet [7]. The string galvanometer became the foundation for electrocardiography, and played a role in Edgar Adrian's (1889-1977) discovery of the all-or-nothing principle of motor nerve action potentials [8].

The all-or-nothing principle enabled the first shift from the slow mechanical domain to the faster inertialess electronic domain. Herbert Gasser (1888-1963) and Joseph Erlanger (1874-1965) could measure the action potential using a low-voltage *cathode ray oscillograph* thanks to its constant shape in response to periodic stimuli [9]. However, the intensity of the cathode ray oscillograph was not enough to capture weak single impulses in sensory neurons, thus the electronics could not achieve immediate success.

The invention of *the vacuum tube* changed the landscape for neural recording as with many other fields. Edgar Adrian used vacuum tube amplifiers to match weak neural signals to the large input range of another low-inertia instrument called *the capillary electrometer* [10]. The search for better amplifiers for neurophysiology led to the invention of *the differential amplifier* [11], which evolved to become an ubiquitous design element in electronics.

The rapid progress and miniaturization of electronics and radio technologies during the following decades created new possibilities for neuroscientific research. *The stimoceiver* [12] devised by José Delgado (1915-2011) is considered as the pioneer of wireless neural recording implants. This radio-controlled multichannel electronic headstage could both record and stimulate the electrodes implanted in the brain, allowing the investigation of neurological diseases and behavior on freely-behaving patients.

The invention of *the integrated circuit (IC)* in the late 1950s by Robert Noyce (1927-1990) [13] and Jack Kilby (1923-2005) [14] radically changed the field of electronics and led to custom ICs for implantable telemetry in the 1980s [15–17]. The rapidly increasing integration density following the Moore's Law [18] has fueled the race to the highest channel count and functionality. The number of electrodes that can be processed within a single chip reached a hundred in the 2000s [19] and it continues to increase. Today, we witness the emergence of thousand-electrode recording systems-on-chip [20].

In this chapter, we compile the contemporary applications, needs and challenges of wireless implantable multichannel neural recording ICs. Section 1.1 introduces the modern neural probes and their use cases. Section 1.2 surveys the circuit techniques for sensing neural potentials. The methods for wireless power and data transfer are discussed in Section 1.3. The current solutions for tackling the data rate bottleneck of high channel count are reviewed in Section 1.4. Finally, Section 1.5 summarizes the implications of this review for the design work which will be presented in Chapters 2 and 3.

1.1 Introduction

The human brain is a complex network of around 86 billion neurons [21]. The interaction between neurons is an electrochemical process [22]. A neuron receives excitatory and inhibitory inputs from other neurons via neurotransmitters shared through synapses, and the received neurotransmitters alter the membrane potential by activating certain ion channels. If the sum of the inputs rises the membrane potential above a certain threshold, a positive feedback rapidly activates other ion channels and generates an action potential (AP) which signals the target neurons[1]. A reverse process brings the membrane potential to its resting state.

1.1.1 Probing the Electrical Activity of the Brain

Probing the aforementioned electrical activity of neurons is the first stage of a neural interface. Intracellular APs can be probed using patch-clamped glass micropipettes [23]. However, this method destructs the cell membrane at least partially; therefore, it is only applicable for *in vitro* studies with isolated cells, or short-term *in vivo* studies [24,25]. Extracellular APs resulting from the movement of ions, and local field potentials (LFPs) created by the collective activity of groups of neurons, can be measured using neural probes [26]. Figure 1.1 shows various

[1] This process is also called as "spiking" or "firing", and the APs are also called as "spikes" in the literature.

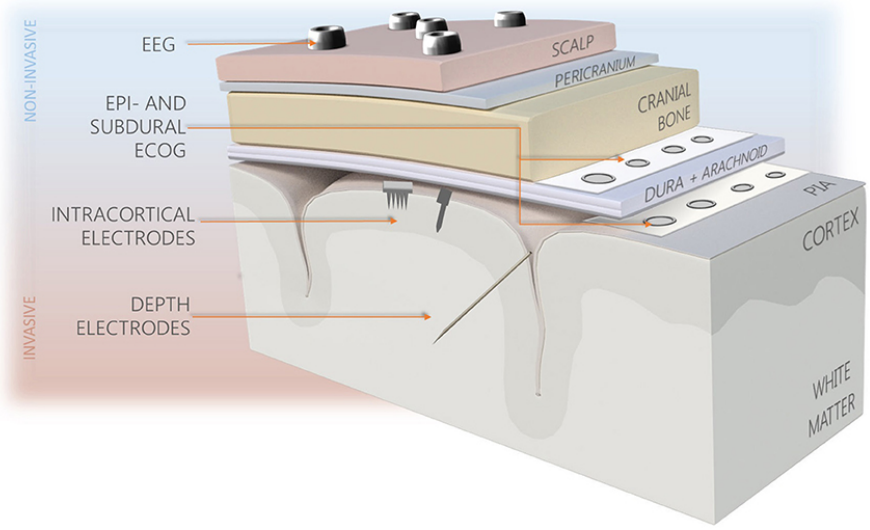


Figure 1.1 – Types of neural probes for sensing electrical activity. Spatial and temporal resolutions increase with the level of invasiveness. Reprinted from [27] under CC BY 4.0 license.

types of probes with different levels of invasiveness classified as depth electrodes, intracortical electrodes, electrocorticography (ECoG), and electroencephalography (EEG).

Depth electrodes are used to reach the subcortical regions of the brain, and they usually have a few recording sites at the tip of the lead [28]. Intracortical electrodes such as the Utah array and the Michigan probe have leveraged the silicon microfabrication technology to create dense microelectrode arrays (MEAs). The Michigan probe contains multiple recording sites along its shank [29]. The recording sites can be configured in a variety of arrangements depending on the application need [30]. Up to 1280 sites can be integrated in a single shank, and multi-shank configurations can expand the 1D probe into a 2D array [31]. The Utah MEA is a 2D arrangement of needle-like probes where the electrode site is at the tip [32].

Stiff penetrating MEAs offer the highest spatial and temporal resolution. However, they also damage the recording site and the recordings deteriorate over time due to foreign body reaction. Soft epidural or subdural ECoG grids placed on the surface of the brain have better mechanical compatibility with the surrounding tissue, hence better long-term performance [33]. ECoG grids can record the LFPs from superficial cortical neurons [26], and APs when the grid has high density and conformability [34].

Strong and synchronized cortical activity can be recorded through the scalp using EEG, without the costs and risks associated with implanting an invasive MEA. Although effective applications such as virtual keyboards and controlling wheelchairs have been demonstrated via EEG, the control of complex neuroprostheses have not yet been achieved [35]. This is mainly because the spatial resolution of EEG is limited to centimeters, and the signal attenuation is high due to the thick layer of tissues in between the electrode and the cortex [22]. As a result, advanced applications demand exclusively invasive recordings.

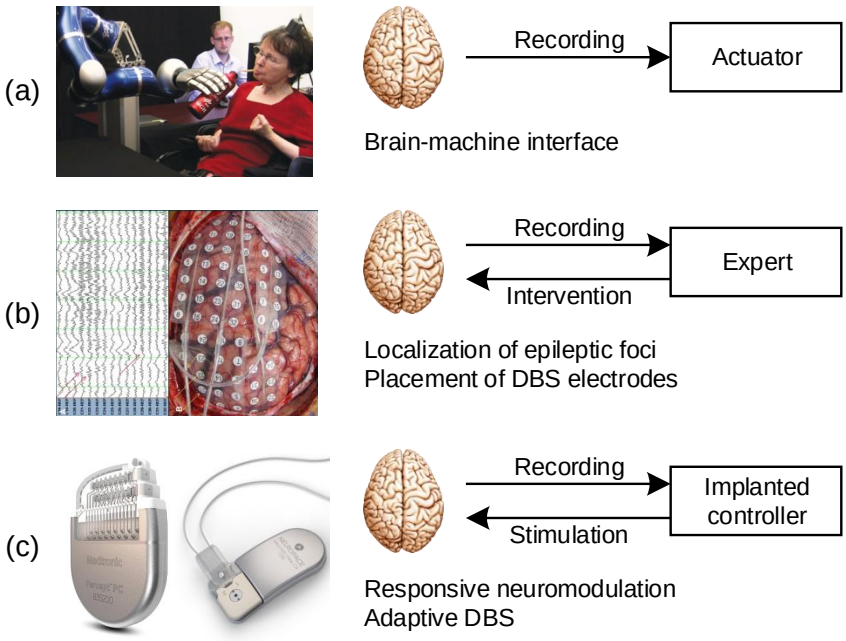


Figure 1.2 – Application examples for invasive neural recording. (a) Brain-machine interface. Image adapted from [36] with permission from Springer Nature. (b) Surgical aid. Image reprinted from [37] with permission from John Wiley and Sons. (c) Closed-loop stimulation. Image obtained from the product web pages of Percept PC, Medtronic Inc. and the RNS System, Neuropace Inc.

1.1.2 Applications of Invasive Neural Recording

The applications of invasive multichannel neural recording can be classified intro three general categories as illustrated in Figure 1.2. The first category (a) uses the recorded signals to control an external device. The second category (b) aids the medical experts in their interventions by providing physiological markers. The third category (c) describes closed-loop systems which automatically act upon a detected brain state. Here we exemplify these classes within two application domains: brain-machine interfaces (BMIs) and neurological disorders.

Brain-Machine Interfaces (BMIs)

BMIs are systems that translate brain activity to interact with objects. Majority of the demonstrations target restoring lost functions or augmenting quality of life in disabled patients, such as cursor control, prosthesis control, and speech decoding.

Cursor control: One of the first showcases of invasive BMIs was controlling a computer cursor in 2D space. Hochberg *et al.* [38] implanted a tetraplegic subject with a 96-channel Utah MEA

in primary motor cortex, three years after spinal cord injury. The subject's movement intention caused spiking activity modulations in the arm area of the primary motor cortex, which were decoded to control a cursor and perform tasks such as opening an email or playing a game. In a similar experiment by Schalk *et al.* [39], the subjects were epilepsy patients implanted with 26 to 64-channel ECoG arrays covering large surfaces. Actual or imagined movements of different body parts caused variations in recorded spectral band powers, which were decoded to control a cursor. The time required to learn how to control this invasive interface was reported to be significantly shorter than it was for noninvasive EEG-based interfaces.

Prosthesis control: Recordings from the motor cortex also enable the control of prosthetic devices. In the 2006 study by Hochberg *et al.* [38], cursor intention was used to open and close a robotic arm. In the 2012 study by Hochberg *et al.* [36], two tetraplegic subjects performed reach and grasp tasks by imagining the movements. Each subject was implanted with a 96-channel Utah MEA in the hand area of the primary motor cortex, and the neural signals were decoded to control a robotic arm. Rastogi *et al.* [40] showed that grasping force can also be decoded from the neural signals. In a demonstration by Ajiboye *et al.* [41], a participant with spinal cord injury was able to control his own paralysed arm by decoding the intention from two 96-channel Utah MEAs in the hand area of the motor cortex, and stimulating the arm muscles through percutaneous electrodes.

Speech decoding: Cursor and prosthesis control BMIs can be coupled to a virtual keyboard to produce speech and text, but with speeds much below the natural rate of speech. Speech BMIs aim to achieve real-time synthesis by decoding brain signals. Anumanchipalli *et al.* [42], demonstrated speech synthesis in 5 patients implanted with high-density ECoG arrays, using a two-stage process where the first stage decodes neural signals into articulatory kinematic features, and the second stage maps the decoded kinematic features into acoustic features. Moses *et al.* [43] decoded question-and-answer dialogs from neural signals in 3 patients implanted with 128 or 256-channel ECoG arrays. Makin *et al.* [44] decoded spoken sentences directly from neural signals in 4 patients implanted with 128 or 256-channel ECoG grids.

Neurological Disorders

Epilepsy, Parkinson's disease, and dementia are three of the most common neurological disorders [45]. Research and clinical use of neural recording implants are critical for improving the treatment of these disorders.

Epilepsy: Epilepsy is a brain disorder characterized by an enduring predisposition to generate epileptic seizures, and by the neurobiologic, cognitive, psychological, and social consequences of this condition, as defined by the International League Against Epilepsy [46]. An epileptic seizure is a transient occurrence of symptoms due to abnormal excessive or synchronous neuronal activity in the brain. 50 million people worldwide were estimated to suffer from epilepsy by the World Health Organization (WHO) in 2019 [47].

Although the majority of epilepsy cases can be treated with anti-epileptic drug therapies, up to 37% are estimated to be drug-resistant (i.e. intractable) [48]. In this case, resection of the epileptogenic zone is a possible solution. A study by Mohan *et al.* [49] reported that among 284 patients who underwent epilepsy surgery, the percentages of patients who remained seizure-free at 5 and 10 years after the surgery were 47% and 38%, respectively. For the majority of patients who were not seizure-free, the seizure frequency reduced significantly.

Precise localization of the epileptic seizure onset zone (SOZ) is crucial to improve the success rate of epilepsy surgeries. An approximate presurgical assessment can be made with magnetic resonance imaging (MRI) and EEG recordings, but intracranial electroencephalography (iEEG) is required for higher accuracy. If the SOZ is expected to be in deep brain tissues, stereotaxic electroencephalography (sEEG) with depth electrodes is preferred. If the SOZ is expected to be closer to the brain surface, ECoG is preferred [50].

In cases where resective surgery contains the risk of severe adverse effects on neurocognitive functions, an alternative solution is implanted closed-loop neuromodulation [51]. This technique delivers electrical stimulation to the SOZ when abnormal patterns are detected in the continuously recorded ECoG activity. The recordings also allow the identification of patient-specific dynamics for personalizing the treatment [52].

Parkinson's disease: Parkinson's disease is a neurodegenerative movement disorder with cardinal motor symptoms being tremor, rigidity, bradykinesia/akinesia and postural instability, which are associated with dopamine deficit caused by the death of cells in the substantia nigra region of the brain [53]. A 2015 study by Feigin *et al.* [54] estimated 6.2 million people worldwide to suffer from Parkinson's disease.

There is no known cure for Parkinson's disease, nor a method to stop its progression at the time of this writing. The current treatments for symptomatic relief are mainly based on dopamine medications for early stages, and deep brain stimulation (DBS) for advanced stages [53]. DBS provides improvement in tremor for at least 10 years after implantation [55].

The success of DBS primarily depends on the precise placement of the stimulation electrodes. The localization is achieved by preoperative MRI for coarse mapping, and intraoperative microelectrode recording for searching activity patterns characteristic to the target region, such as the beta range activity [56]. Moreover, recordings obtained during the surgery has been found to predict the clinical improvements accurately at 6 months after the surgery [57].

Postoperative recording can improve the DBS therapy outcome by enabling personalized treatments. Adaptive DBS triggers stimulation based on the LFPs recorded on the stimulation electrodes. This method has been shown to be more effective and energy efficient than continuous stimulation [58]. Dyskinesia-related oscillations recorded from the motor cortex have been proposed as a potential feedback control signal to minimize the adverse effects [59]. A recent DBS system with integrated recording channels has demonstrated the correlation between patient-reported adverse events and recorded LFPs [28].

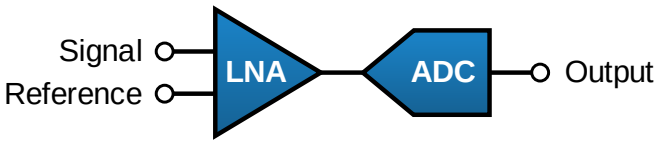


Figure 1.3 – A generic analog front-end.

Dementia: Dementia describes several diseases that affect memory, thinking, and behavior progressively and faster than the normal course of aging. According to the WHO [60], 47 million people worldwide were affected by dementia in 2015, and the number is predicted to reach 75 million in 2030 and 132 million by 2050.

60% to 70% of the dementia cases are caused by Alzheimer's disease [60]. It cannot be cured or stopped from progressing at the time of this writing, and the progress accelerates with age. Therefore, early diagnosis is crucial. Recent studies [61,62] have reported correlations between abnormalities in EEG markers and cognitive deficits in groups of Alzheimer's disease patients at different stages of the disease. Moreover, MEA recordings from the rat hippocampus have been useful in studying abnormalities related to Alzheimer's disease [63–65].

1.2 Analog Front-Ends for Sensing Neural Potentials

The vast majority of the demonstrations mentioned in Section 1.1.2 record the neural signals sensed through the electrodes using rack-mounted general-purpose neurophysiology instruments. Although this approach serves well for proof-of-concept research in the laboratory environment, translation to widespread use in independent environments calls for fully-implantable electronics.

A neural recording system can fit in a single IC and can be operated on a small battery or powered wirelessly when tailored to a specific application. As depicted in Figure 1.3, a generic recording channel consists of a low-noise amplifier (LNA) to boost weak neural signals, followed by an analog-to-digital converter (ADC) to produce a digital code for subsequent operations. This channel architecture is collectively referred to as an analog front end (AFE).

Given a total power and area budget, the number of channels that can be integrated on a single system-on-chip (SoC) depends on the energy and area efficiency of the single channel. Therefore, optimizing the AFE is crucial in the design of an implantable multichannel neural recording system. Depending on the application needs and resource restrictions, the AFE architecture can be adapted in many ways, as reflected by the number of publications in the past decades. This section intends to provide an overview of circuit specifications, challenges, and current solutions from both LNA and ADC perspectives.

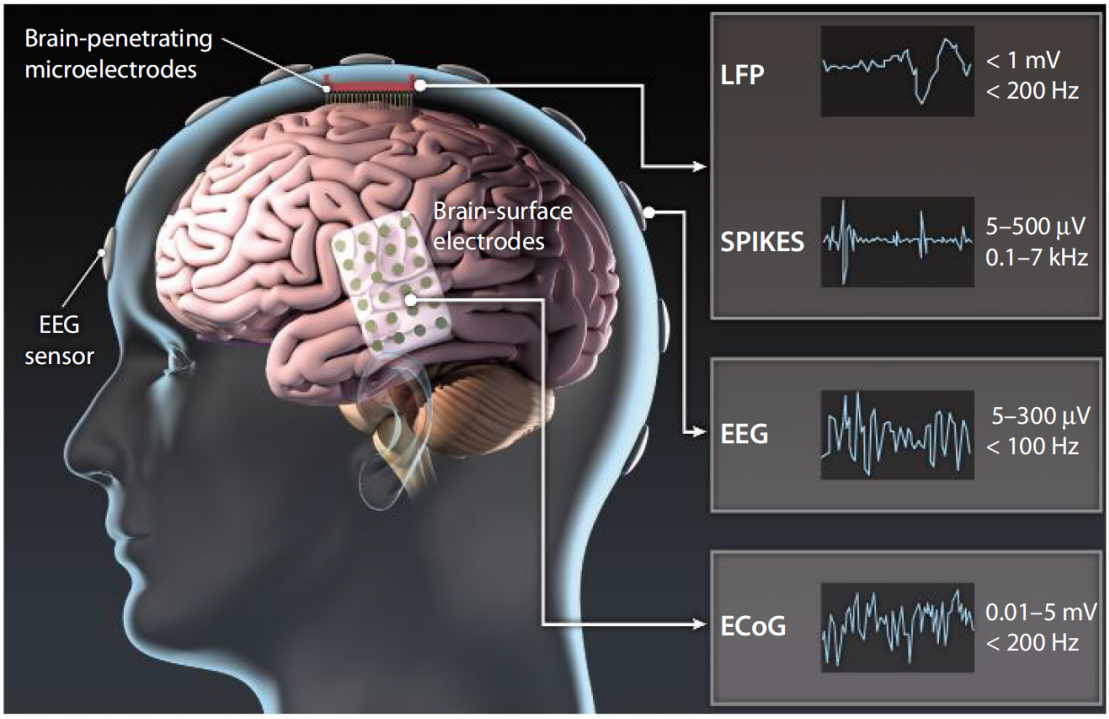


Figure 1.4 – Bandwidth and amplitude of different types of neural signals. Reprinted from [68] with permission from the American Association for the Advancement of Science (AAAS).

1.2.1 Specifications

Modern neurophysiology instruments are designed to investigate the detailed mechanisms of neural activity with up to 16-bit resolution and 30 kHz sampling rate [66, 67], which are well above the requirements of many practical applications. However, the implant power and area restrictions demand an optimized recording front-end in order to maximize resource usage efficiency. Here we focus on the factors that influence the choice of channel count, bandwidth, and resolution for an implantable system.

Bandwidth

Figure 1.4 illustrates the expected bandwidth and the amplitude of neural signals at different levels of invasiveness. The highest frequency neural signal of interest is the AP with 7 kHz bandwidth. ECoG and LFP signals have lower bandwidths around 200-500 Hz.

Optimizing the LNA bandwidth and the ADC sampling rate for the target signal bandwidth saves energy. The current drawn by most LNA topologies is proportional to the bandwidth, and the power requirements of Nyquist-rate ADCs scale linearly with the sampling rate. Choosing the sampling rate as 3 times higher than the amplifier bandwidth would be sufficient to digitize the signal reliably without aliasing [69]. To distinguish the shape of each AP produced by a group of neurons, a sampling rate between 10-40 kS/s is sufficient [70].

Resolution

Assuming the maximum peak of spiking activity at 1 mV and the minimum noise and interference from neighboring neurons at 10 μV [71, 72], signal-to-noise ratio (SNR) at the recording site is less than 100 [70, 72], which corresponds to 6 to 7 bits of dynamic range. Recording with higher fidelity does not necessarily mean better performance from the application perspective. For example, an analysis by Even-Chen *et al.* [69] finds that decoding movement intention does not benefit from higher resolution above 7 bits.

Avoiding overdesign is essential for implantable AFEs as increasing resolution generally requires a disproportionate increase in power and area. As a simple approximation for LNAs, the bias current of a differential pair has to be scaled up 4 times in order to decrease the input-referred thermal noise by half. On the ADC side, empirical analysis shows that power increases by 2 times per additional bit for low to moderate resolution (below ~10-bit), and by 4 times for higher resolutions [73]. The area trend shows an increase of 2.2 times per bit [74], but the relationship is weaker due to non-scaling overheads of peripheral circuits [75].

A resolution-optimized design keeps the noise contribution of the LNA and the ADC below the electrode site noise to avoid corrupting the signal. Olsson and Wise [76] measured the site noise as 9.6 μV, and the site-referred noise of the amplifier as 8.9 μV, which led to a total noise of 13.1 μV. Assuming 1 mV peak amplitude, a 5-bit digitizer was chosen since the site and amplifier noise would have dominated the quantization noise at higher resolutions.

Channel count

The reliability of recordings depends not only on sufficient fidelity, but also on the alignment of the target signal source with the electrode [77]. To increase the chances of alignment, high channel count is desired. Furthermore, some of the electrodes in an electrode array may be inactive due to site damage, connectivity issues, or degradation [78]. High channel count can provide the redundancy to compensate for the lost electrodes. Moreover, the increased spatial density and coverage can compensate for the limitations in time-series information such as low sampling rate or low SNR [79].

High spatial density is especially important when high-frequency activity is of interest. As the electrode spacing decreases, the increase in correlation is lower for high frequencies (>30 Hz) than for low frequencies (<30 Hz), according to a study on a 4 mm-spaced 256-channel ECoG grid implanted on human speech cortex [80]. Therefore, dense electrode arrays can capture more high-frequency content.

Marblestone *et al.* [70] discussed the theoretical limits of electrode density for discriminating spiking activity of individual neurons. Considering the exponential decay of the signal with increasing distance to neuron, the interference from neighboring neurons at the recording site, and the current spike sorting algorithms, 40 μm electrode spacing would be required to detect spikes from every neuron, which corresponds to ~10 neurons per electrode, and ~7.5

million electrodes for covering an entire mouse brain. The state-of-the-art MEAs can integrate up to a thousand electrodes with down to 20 µm spacing [20, 31, 81].

1.2.2 Amplification

There has already been a few comprehensive reviews on LNAs. Ng *et al.* [82] compiled the amplifier topologies used for neural recording. Bagheri *et al.* [83] reviewed the techniques for noise and offset cancellation. Here we explore both aspects quantitatively by comparing published solutions from the resource efficiency point of view.

Noise

Since the amplifier is the first active stage, its noise contribution irreversibly corrupts the actual neural potential and propagates through the signal chain. Therefore, low-noise design techniques are essential to ensure that the circuit noise is kept below the background noise level of the neurons, which is in the order of 10 μV_{rms} as discussed previously.

Noise efficiency factor (NEF) is a widely accepted metric that allows comparing different topologies based on the noise-current-bandwidth trade-off. It normalizes the input-referred circuit noise of an amplifier against the thermal noise of an ideal bipolar junction transistor (BJT) [17]:

$$NEF = V_{rms,in}\sqrt{\frac{2I_{tot}}{4kTU_T\pi B}}, \tag{1.1}$$

where $V_{rms,in}$ is the root-mean-square input-referred noise, I_{tot} is the total current, U_T is the thermal voltage, and B is the bandwidth of the amplifier. Considering only the thermal noise contribution, NEF of an ideal differential amplifier is limited to $n\sqrt{2}$ [84] where n is the weak inversion slope factor with $1 < n < 1.7$ [85].

While NEF captures the current efficiency of a circuit, it does not reflect its power efficiency due to the lack of supply voltage (V_{DD}) in the equation. Power efficiency factor (PEF) [86] accommodates this need and allows comparison between amplifiers in terms of noise-power-bandwidth:

$$PEF = NEF^2 V_{DD} = \frac{V_{rms,in}^2 2P_{tot}}{4kTU_T\pi B}. \tag{1.2}$$

There are a number of candidate operational transconductance amplifier (OTA) topologies for LNA implementation. We compare the most commonly used topologies in the literature based on their NEFs and PEFs in Figure 1.5 [83, 84, 86–100, 100–124, 124–133, 133–140, 140–152]. Telescopic cascode, inverter-based and current-reuse topologies have demonstrated higher efficiency compared to the other known topologies. Nevertheless, NEF and PEF should be evaluated together with the individual advantages and drawbacks of each topology.

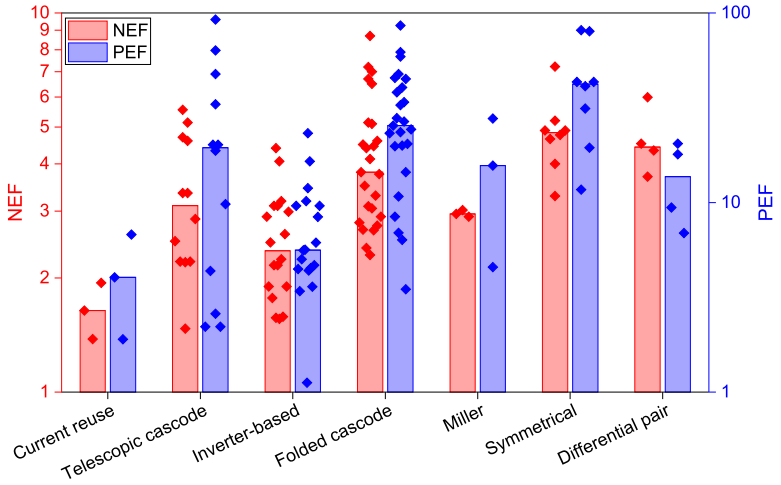


Figure 1.5 – Comparison of published NEF and PEF values for various amplifier topologies. The bars represent median values.

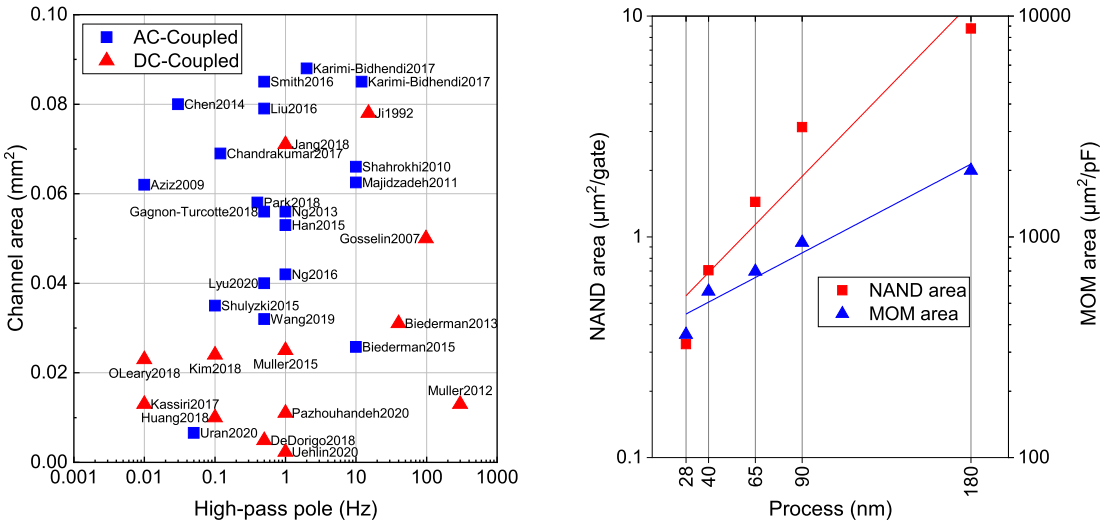
Telescopic cascode amplifier can achieve a higher gain per current thanks to its cascode transistors, which also do not contribute to noise. Consequently, this architecture has achieved the NEF values closest to the $n\sqrt{2}$ limit among the architectures with a single differential pair. The drawback of the cascode transistors is the large voltage headroom requirement which does not permit aggressive supply scaling.

The inverter-based and current reuse topologies have multiple differential pairs in one current branch, thus have lower theoretical NEF limits [153] as observed in Figure 1.5. The inverter-based amplifier has a stack of P-channel metal-oxide-semiconductor (PMOS) and N-channel metal-oxide-semiconductor (NMOS) differential pairs that provides additional gain per unit current. However, the lack of cascode transistors decreases its robustness against supply noise. The current reuse topology aims to amplify multiple channels at once with one stack of differential pairs. Therefore, it requires a large voltage headroom and avoiding crosstalk is a challenge [92].

Flicker noise dominates thermal noise in the neural frequency band between sub-1 Hz and 100 Hz. Chopper stabilization [138, 154, 155] is a common technique for removing flicker noise, at the expense of reduced input impedance and switching artifacts at the output. These adverse effects must be controlled by additional circuits such as impedance boosting and ripple reduction [83].

Electrode offset

The electrode-tissue interface generates a potential in the range of hundreds of mV, and can vary in the range of tens of mV from channel to channel and over time [71, 138, 156]. Most analog circuits cannot accommodate such variability. The most common and straightforward



(a) (b)

Figure 1.6 – (a) Area and high-pass pole location trade-off. (b) Rate of progress for NAND gate and metal-oxide-metal (MOM) capacitor area efficiencies.

work-around is to use capacitive coupling [129]. However, this method can be very area-consuming when the resulting high-pass pole needs to be kept as low as 1 Hz and below. In this case, the coupling capacitor and the feedback resistor need to be in the order of pF and GΩ, respectively. Other methods to handle the electrode offset rely on active cancellation by means of analog [17, 134] or mixed-signal [86, 132] feedback paths.

Figure 1.6a illustrates the trade-off between the circuit area and the high-pass cut-off based on published front-ends [1, 16, 86, 100, 102, 108, 121, 124, 131, 132, 140, 144, 152, 155, 157–163]. Mixed-signal methods have an obvious area advantage, especially in smaller process nodes since the digital footprint shrinks faster than the capacitance density as observed in Figure 1.6b.

1.2.3 Digitization

Figure 1.7 shows the sampling rate and resolution choices in the neural recording literature for the three most common ADC architectures [1, 19, 31, 83, 86, 90, 91, 94, 95, 97–102, 104–107, 109–116, 120, 125–127, 130–133, 135, 136, 141, 143, 146, 147, 149, 151, 152, 157, 158, 161, 162, 164–183]. The reported effective numbers of bits (ENOB) have been between 6 and 12, with the majority being around 8 to 10. The sampling rates cover the kHz range as this specification depends on the signal band of interest.

Successive approximation register (SAR) ADC is the most commonly used architecture for low-to-medium resolutions thanks to its high versatility and scalability. A charge-redistribution SAR ADC is based on discrete-time charge manipulation on the sampled signal, therefore it can be built without any static power consumption. For high resolution, $\Delta^2\Sigma$ topology has

been used thanks to its oversampling nature. It also eliminates the need for an LNA as it samples directly the electrodes [110].

Voltage-controlled oscillator (VCO)-based ADCs have recently gained popularity thanks to their amenability to process and voltage scaling. This topology operates in frequency domain rather than in voltage domain and mostly uses digital elements, thus it can work with less headroom and have a compact footprint. The voltage-to-frequency nonlinearity is the main limitation for VCO ADCs; therefore, they need additional nonlinearity correction stages [170], or they need to be placed in a feedback loop [86].

A true comparison of ADC performances is difficult due to a multitude of specifications driving the design choices. Nevertheless, two simple figure of merits (FoMs) have been adopted in the literature that capture power, sampling rate, and resolution as a single metric: Walden FoM (FoM_W) and Schreier FoM (FoM_S). Definitions of these FoMs were covered in detail by Murmann [184]. FoM_W assumes 2 times increase in power per additional bit, which holds true for low to moderate resolution ADCs up to about 8 bits. Higher resolution ADCs align better with FoM_S which assumes 4 times increase in power per additional bit.

Figure 1.8a and 1.8b plot the energy efficiencies of standalone ADCs and complete AFEs in the literature. The energy per sample raises by an order of magnitude when the overhead of the LNA is added to the ADC. The increasing energy cost of high resolution is clear in both figures.

Recently, energy-area figure of merit (E-A FoM) (Area$\times FoM_W$) has been proposed as a new metric for comparing the total cost of neural recording AFEs [116]. Accordingly, Figure 1.9 adds the area dimension to Figure 1.8b. Although the popular SAR ADC exhibits superior energy efficiency up to 10-bit resolution, its energy-area efficiency lags behind the emerging VCO and $\Delta^2\Sigma$-based topologies.

1.3 Wireless Telemetry

Wireless communication is critical for the safety of long-term neural recording implants since it eliminates burr holes and through-skull wires which are prone to physical damage or infections [185]. In this section, we discuss the techniques and considerations for wireless power and data transfer (WPDT) with examples from the literature.

1.3.1 Powering the Implant

The power consumption of implanted electronics is primarily constrained by thermal dissipation, because excessive temperature elevation damages the electrical properties of brain cell membranes and ion channels [70]. The widely accepted limit for safe operation is 1 to 2°C [186–189]. Marblestone *et al.* [70] estimated a maximum of 40 mW/cm^2 power density for staying below 2°C.

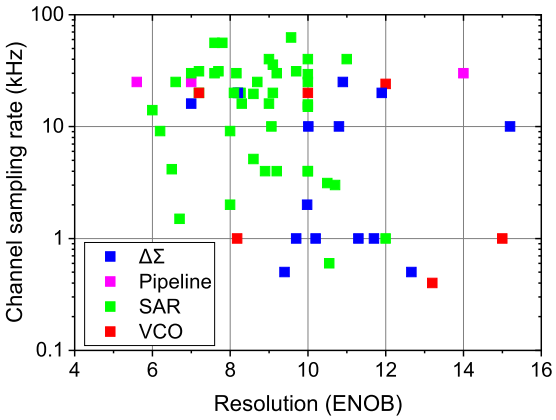


Figure 1.7 – Sampling rate and resolution choices in the literature.

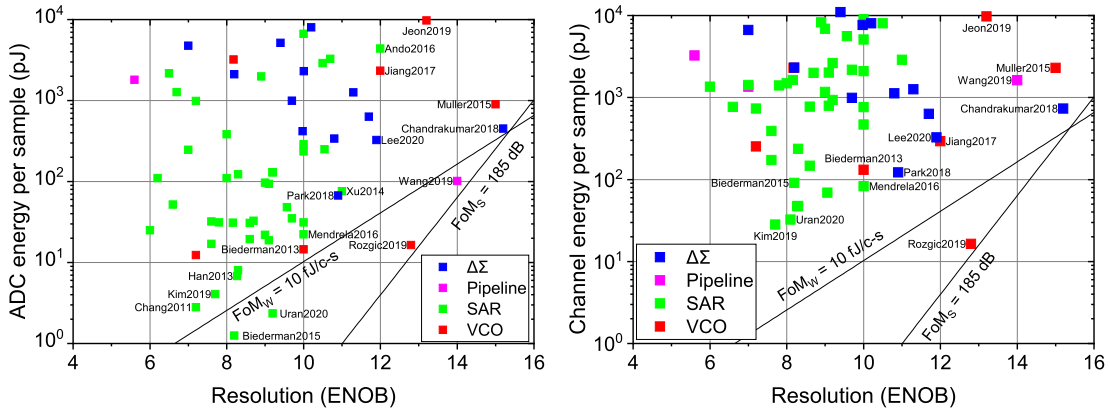


Figure 1.8 – Energy efficiencies of published (a) ADCs and (b) AFEs.

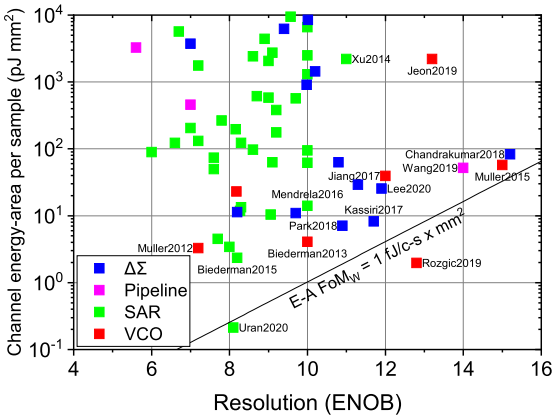


Figure 1.9 – Energy-area efficiencies of published AFEs.

Another constraining factor for power consumption is the capacity of the power source. The majority of the implantable interfaces rely on an implanted battery or wireless power transfer (WPT). The small capacity of biocompatible batteries implies that an implant's size grows with its power need, which in turn limits the choice of where the implant is placed. For example, deep brain stimulators are placed in the chest area with the wires passing through the neck because the device is too large to fit inside the skull [190]. Moreover, batteries have to be replaced via surgery approximately every 2 to 8 years [191]. Rechargeable batteries prolong the battery lifetime a few more years, but the patient is responsible to charge it regularly [192].

WPT has been used to recharge implanted batteries, or even to eliminate them completely by directly powering the implant. The two most common WPT strategies are inductive coupling for superficial implants, and ultrasound for deep implants. The former is more efficient for cortical implants as the depth is 5-10 mm under the scalp. Similar to the case of batteries, the power need influences the size of the implant with inductive coupling. High power over long distance enforces low field frequency for electromagnetic safety, which in turn requires a large coil diameter. On-chip coils become feasible below 1 mW and 1 cm [115, 132].

Apart from WPT, energy harvesting from body resources has also been investigated. A glucose biofuel cell reported by Zebda *et al.* [193] was able to generate 193.5 $\mu W/cm^2$ average power. An enzyme-based biofuel cell by Bollella *et al.* [194] generated 10 µW for powering a temperature sensor with memory and wireless data connectivity. Although such power levels do not match the multichannel neural recording requirements yet, the combined progress in biofuel cells and low-power circuit techniques is promising towards fully-autonomous implants.

1.3.2 Communication

The choice of the communication standard is driven by a multidimensional trade off between power, range, data rate, and complexity. We can categorize the current options for medical applications in three categories: passive, narrowband (NB), and ultra-wideband (UWB) [195].

Passive communication uses a pair of reader and tag, which are generally inductively coupled. The reader sends data to the tag by modulating the magnetic field, and the tag talks back by load modulation or backscattering. This strategy has been popular since the first examples of integrated WPDT [196], mainly because load modulation requires minimal overhead on the implant side, thus it can achieve below 10 pJ/b efficiency [126, 127]. As with inductive power links, the range of passive communications is in the centimeters range.

NB transmitters rely on modulating a carrier frequency using frequency-shift keying (FSK) [168, 172] or on-off keying (OOK) [93]. The main advantage of NB over passive techniques is extending communication distance to the meters range, at the expense of higher power consumption. To illustrate, Bonfanti *et al.* [105] reported 3.5 mA increase in current for extending the range from 5 cm to 10 m with an FSK transmitter.

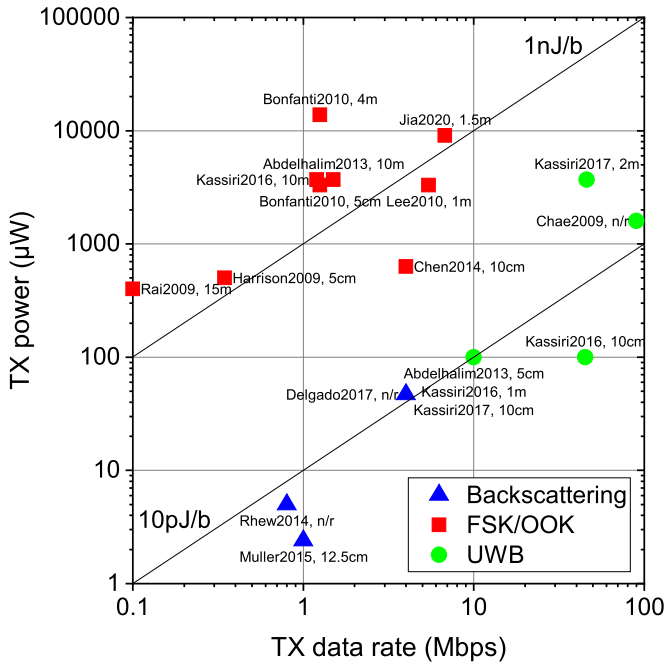


Figure 1.10 – Comparison of data rate and transmitter power for various strategies.

Passive and NB transmitters are effective for data rates up to the Mbps range. However, the data rate requirement for high-resolution multichannel systems has reached the order of tens of Mbps. A relatively recent technique that achieves up to 90 Mbps [95] is UWB, which relies on modulating a carrier frequency with time-domain pulses. The data rate and the range of UWB depend on the frequency band and the topology used as studied by Kassiri *et al.* [110, 172]. For short range up to 10 cm, the 3.1-10.6 GHz band achieved 45 Mbps while consuming 0.1 mW. The range increased up to 1 m by operating in the sub-1 GHz band, at reduced data rate of 10 Mbps. The delay- and VCO-based topologies also produced different results. The former achieved 10 Mbps at 10 cm consuming 0.1 mW, and the latter achieved 46 Mbps at 2 m consuming 3.7 mW.

Figure 1.10 summarizes the solution space proposed in the literature [93, 95, 97, 104, 105, 110, 114, 126, 127, 133, 164, 168, 172, 197]. Based on this graph, we can conclude that backscattering has been favored for its low power consumption when data rate is in the order of Mbps and the transmission range is limited to a few centimeters. For higher data rates, UWB has been a solution with approximately the same energy efficiency (pJ/b), and with range up to a few meters. FSK/OOK-based NB transmitters can increase the range up to around 10 meters at the expense higher energy per bit.

The power overhead of wireless communication largely depends on the laws of electromagnetics and the transmission medium. Therefore, the circuit-level innovations progress slowly. The only shortcut to save power is to reduce the data rate and to use duty cycling, where the data is shipped in bursts and then the radio is shut down until the next cycle.

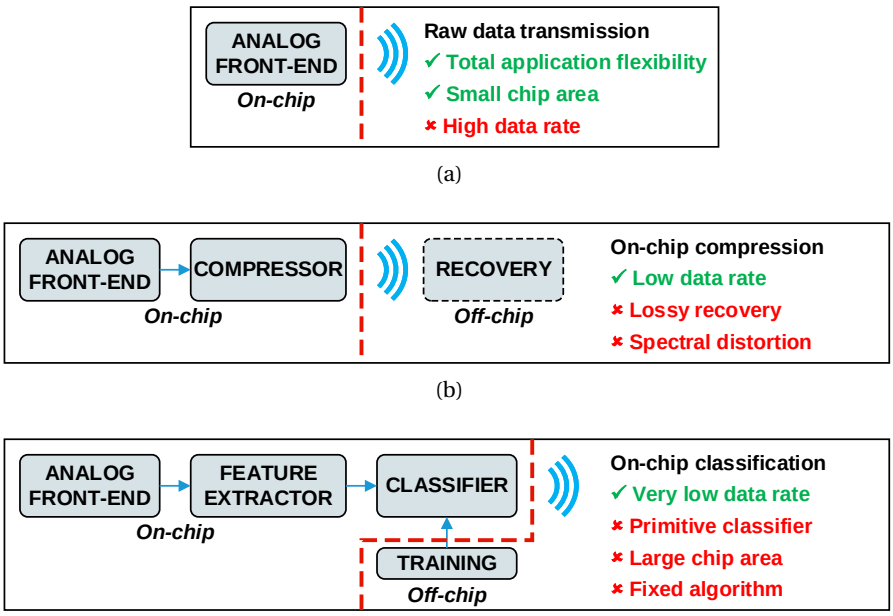


Figure 1.11 – Advantages and disadvantages of (a) raw data transmission, (b) on-chip compression, and (c) on-chip classification.

1.4 Data Rate Reduction Techniques

The straightforward method to create a neural recording application is to send the recorded data off-chip and process at the receiver side, as done by the majority of demonstrations discussed in Section 1.1.2. Nevertheless, the strive for high channel count, high resolution and high sampling rate inevitably leads to a data deluge. As an example, a 10-bit, 20 kS/s AFE produces a throughput of 200 kb/s per channel. For a thousand channels, this means 200 Mb/s overall throughput that needs to be transferred via the wireless link. This data rate is currently not feasible due to the physical and technological limitations of implant, as reviewed in Section 1.3. The established medical transceivers support up to 800 kb/s [198], whereas the recently proposed UWB transmitters report from 10 to 90 Mb/s.

To work around the data rate bottleneck, two approaches have been dominantly used. The first approach reduces the data throughput by compressing the raw data stream. The second approach aims to eliminate the need for high-speed data transmission altogether by performing the user-end classification task on the chip. Figure 1.11 summarizes the advantages and disadvantages of each approach and the following sections discuss them in detail.

1.4.1 On-Chip Compression

On-chip compression aims to reduce the transmitted data rate by discarding parts of the raw data while retaining the information content. Here we focus on three compression approaches

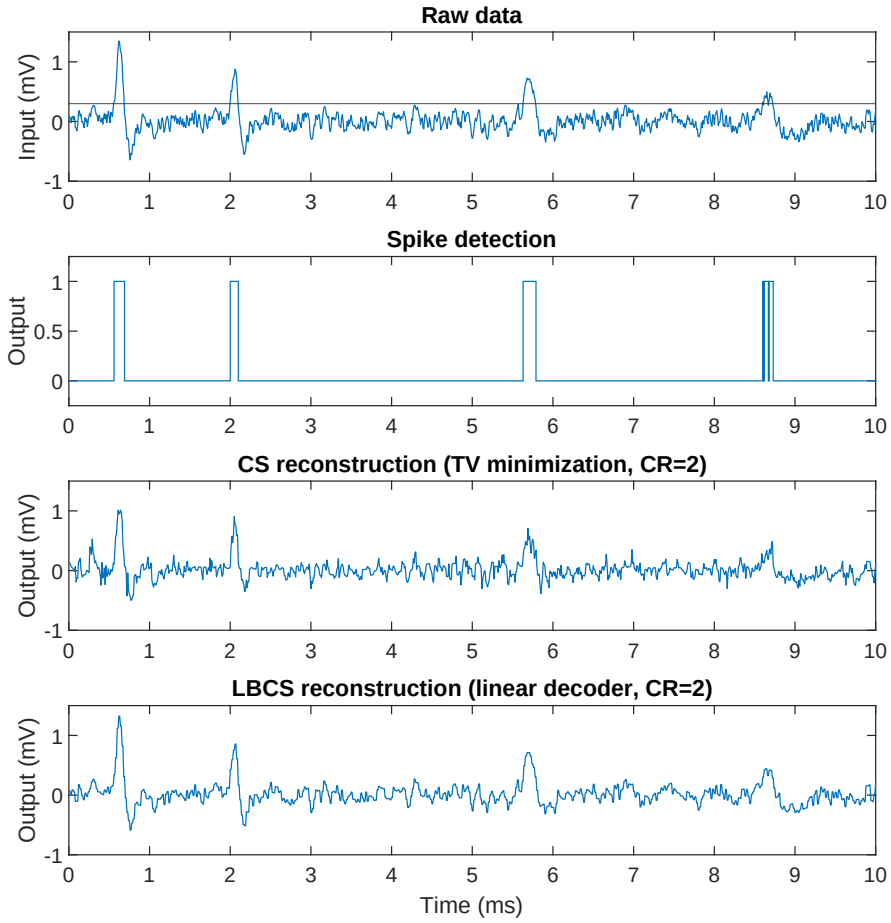


Figure 1.12 – Three on-chip compression approaches: spike detection, CS, and LBCS.

for neural signals as illustrated in Figure 1.12: spike detection, compressed sensing (CS), and learning-based compressive subsampling (LBCS).

Spike Detection

Spiking rate, amplitude and timing serve as neural features for many BMIs [36, 38, 41]. In this case, transmitting only spike information would be sufficient for such applications. Spike detection compares the neural signal stream to a threshold determined based on a statistical metric, thus it can be easily implemented on-chip. Harrison [199] transmitted only the presence and the timing of detected spikes from a 100-channel, 10-bit, 30 kS/s raw data stream. Assuming a firing rate around 10 Hz, transmitting only the 7-bit address of the channel reduced the data rate from 30 Mb/s to an average of 7 kb/s. Olsson and Wise [76] also included the amplitude as a coarse indicator of the spike shape.

Multichannel MEAs record multiple neurons per electrode. Therefore, transmitting only the timing and amplitude data leads to information loss. Spike sorting aims to detect individual spikes and associate them with individual neurons by clustering the spike shapes; however, implementing it on-chip requires a large area [200]. An alternative strategy has been transmitting a time window around the detected spike for off-chip spike sorting [106, 109, 135, 143].

The loss of the global signal shape limits the use of spike detection to only a few cases. Majority of neural recording applications require the full waveform for further feature engineering such as extracting spectral powers [28, 42, 51] or for visual diagnosis [28]. In addition, the bandwidth saving thanks to spike detection is highly dependent on the firing rate, which is generally assumed to be between 10 to 100 Hz. Spikes are missed if many channels are simultaneously and consecutively active above this rate, such as during a seizure [76, 143].

Compressive Sensing

The Shannon-Nyquist sampling theorem states that a continuous signal can be fully recovered from its sampled representation if the sampling rate is twice its bandwidth [201]. This theorem forms the basis of standard data acquisition systems, but it assumes that the whole bandwidth is densely occupied by the signal. In reality, many natural signals are sparse or compressible, meaning that few large coefficients capture the most of the information [202].

Compressive sensing theory [203] posits that a sparse signal $\mathbf{x} \in \mathbb{R}^N$ with only K non-zero elements (K-sparse) can be recovered from its compressed representation $\mathbf{y} \in \mathbb{R}^M$, measured through a sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M < N$ as

$$\mathbf{y} = \mathbf{\Phi x}\,, \tag{1.3}$$

where the sensing matrix $\mathbf{\Phi}$ satisfies the restricted isometry property (RIP). Random matrices with Gaussian or Bernoulli distributions satisfy the RIP with high probability if $M \geq cKlog(N/K)$ where c is a small constant [202]. The compression ratio (CR) can then be defined as $CR = N/M$. Figure 1.13 illustrates a generic measurement with CS.

Some natural signals like electrocardiography (ECG) are already sparse in time domain. Other signals like neural potentials may not be sparse in time domain, but a sparse representation can generally be found in another basis $\mathbf{\Psi}$ as

$$\mathbf{y} = \mathbf{\Phi\Psi x} = \mathbf{\Theta x}\,. \tag{1.4}$$

Choosing $\mathbf{\Psi}$ as an orthonormal basis ensures the RIP of $\mathbf{\Theta}$ if $\mathbf{\Phi}$ satisfies the RIP. Common choices for $\mathbf{\Psi}$ are spectral transforms such as discrete Fourier transform (DFT), discrete cosine transform (DCT), or Hadamard transform.

CS has been applied to neural signals with on-chip implementations. Liu *et al.* [91] presented a digital CS processor that was shared among 16 channels. In simple mode, the coefficients

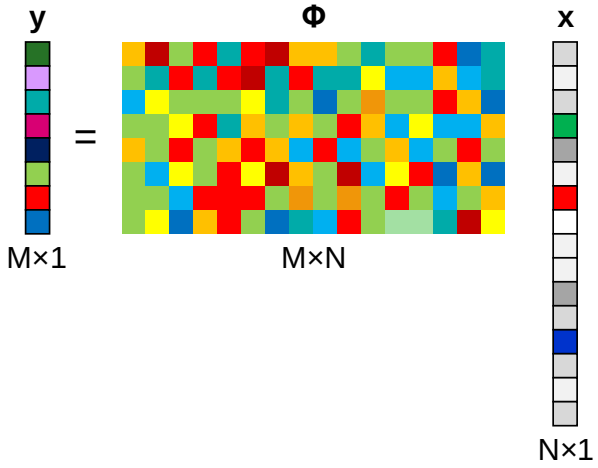


Figure 1.13 – Illustration of a generic CS measurement.

of the sensing matrix **Φ** were ternary, corresponding to a simple sign inversion. In high-resolution mode, the coefficients had 4-bit resolution and multiplications were implemented by changing the gain of the analog amplifier to avoid digital multipliers. The coefficients were generated randomly offline and loaded in the beginning before each measurement to avoid on-chip storage. The signal was recovered by finding the sparsest estimation in its own learned dictionary through orthogonal matching pursuit. The post-reconstruction SNRs were 9.04 dB, 4.85 dB, and 3.78 dB for CRs 4, 8, and 16, respectively. Spike detection on the reconstructed signal achieved near-lossless performance for CRs 8 or below.

Shoaran *et al.* [131] approximated a Bernoulli matrix with an on-chip pseudorandom binary sequence (PRBS) generator. The coefficients implied summation or subtraction of the amplified neural potentials from all channels in analog domain before digitizing, which was named multichannel compressed sensing (MCS). Reconstruction was performed by finding the sparsest representation in Gabor domain through $\ell_{1,2}$ mixed norm minimization. This method achieved up to 16 CR for the 16-channel prototype with an average 21.8 dB post-reconstruction SNR when CR was 4.

The main drawback of CS is that the recovery requires solving minimization problems. This requires a powerful and fast processor at the receiver side for real-time decoding. Moreover, the recovered signal's SNR is usually below 10 dB at high compression rates. The frequency spectrum experiences significant distortion [91] which could affect the application performance if the spectral band powers are important.

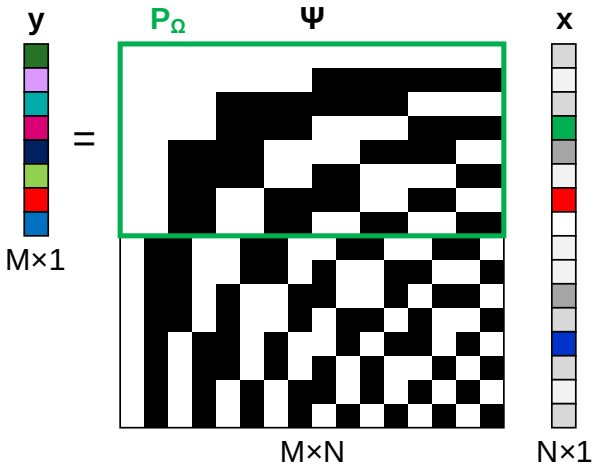


Figure 1.14 – Illustration of the LBCS concept with a subsampled Hadamard measurement matrix.

Learning-Based Compressive Subsampling (LBCS)

Although random matrices satisfy the theoretical constraints of CS, matrices that are aware of the signal structure have been found to perform better in practice [204, 205]. The structured subsampling problem can be written similarly as

$$\mathbf{y} = \mathbf{P}_\Omega \mathbf{\Psi} \mathbf{x}\,, \tag{1.5}$$

where $\mathbf{P}_\Omega$ is a subsampling operator that selects the rows of $\mathbf{\Psi}$ indexed by the set Ω. The subsampling concept is illustrated in Figure 1.14 on a Hadamard matrix. The signal can be recovered by using standard CS algorithms, or by a fast linear decoder as

$$\hat{\mathbf{x}} = \mathbf{\Psi}^* \mathbf{P}_\Omega^T \mathbf{y}\,. \tag{1.6}$$

Choosing the best set Ω depends on the signal structure and the recovery algorithm used. One possible solution is to randomize Ω based on signal properties known *a priori*. Baldassarre *et al.* [206] formed $\mathbf{\Theta}$ by subsampling a Hadamard matrix randomly according to a probability function that favors the low frequencies since they are known to contain more energy. This method, named structured Hadamard sampling (SHS), achieved an average 22.2 dB post-reconstruction SNR with tree norm recovery when CR was 16.

An alternative and better solution was proposed again by Baldassarre *et al.* [207], called LBCS. Briefly, LBCS minimizes the mean square error of the reconstructed signal by learning the set of indices Ω that captures the highest average energy, by training on an example dataset of

m measurements $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_m\}$:

$$\hat{\Omega} = \underset{\Omega, |\Omega|=M}{\arg\max} \frac{1}{m} \sum_{j=1}^{m} \sum_{i \in \Omega} |\langle \boldsymbol{\Psi}_i, \boldsymbol{x}_j \rangle|^2 . \tag{1.7}$$

Aprile *et al.* applied this concept to neural signals in [208]. Two studies from the iEEG.org dataset portal were used to learn the set Ω that subsamples a Hadamard matrix. The learned set Ω was used to subsample and recover the signals using a linear decoder as in (1.6). This method achieved 28.48 and 18.06 dB SNR for 16 and 64 CR, respectively.

Not only the choice of Ω but also the choice of the orthonormal basis $\boldsymbol{\Psi}$ is important for post-reconstruction performance of LBCS. Aprile *et al.* [209] implemented a DCT matrix instead of a Hadamard matrix, achieving 28.48 and 18.06 dB SNR for 16 and 64 CR, respectively. The main drawback of choosing DCT over Hadamard was the power and area overhead of the on-chip memory used to store the 8-bit quantized DCT coefficients. Therefore, there exists a trade-off between reconstruction quality and hardware resource utilization [210].

The main issue with LBCS is that the learned set of indices Ω always favors the lowest indices of $\boldsymbol{\Psi}$. This is not surprising as neural signals contain the most energy at low frequencies and the performance metric for reconstruction is SNR. In this case, $\mathbf{P}_\Omega \boldsymbol{\Psi}$ essentially implements a low-pass filter and discards the high-frequency information that could be useful as biomarkers.

1.4.2 On-Chip Classification

The relationship between the neural signals and the target outcome is often implicit and cannot be programmed, especially when the dimensionality of multichannel recording is considered. Therefore, many applications mentioned in Section 1.1.2 use a machine learning (ML) algorithm which learns a model of the complex relationship by training on example data.

Machine Learning for Neuroscience

A typical ML algorithm consists of two main steps: feature extraction and decoding [211]. Feature extraction aims to reduce the dimensionality of the raw data by transforming it into descriptive components that are relevant to the task. Spectral band powers[2] are the most commonly used features for neuroscience applications. Low-frequency band powers have been used in DBS placement and tuning [28, 56], speech decoding [42, 43], and cursor control [39]. High-frequency oscillations (HFOs) have been used as biomarkers for epileptic seizures [212, 213] and as predictors of arm movement [41].

A decoder predicts the outcome from the extracted features through regression (if the outcome is continuous as in cursor and prosthesis control) or classification (if the outcome is discrete

[2]The traditional EEG bands are named as delta (δ: 1-4 Hz), theta (θ: 4-8 Hz), alpha (α: 8-13 Hz), beta (β: 13-30 Hz), gamma (γ: 30-150 Hz), ripples (150-250 Hz) and fast ripples (>250 Hz).

as in speech decoding or seizure detection). Regression is often implemented as an adaptive filter that tunes its parameters over time, such as least-squares [38] or Kalman [36, 41] filters. Classification is usually performed by a decision algorithm such as a support vector machine (SVM) or an artificial neural network (ANN) [42, 44, 57]. The classifier learns its parameters from example data (training), and applies them on future signals (inference). Regression and classification may be performed simultaneously in cases like cursor control with clicking [214].

Hardware Constraints

Recent studies have focused on implementing ML algorithms on the chip to avoid high-speed data transfer between the implant and the processor. This approach has been explored especially for closed-loop neuromodulation where the recordings are used to trigger stimulation in response to identified events. However, on-chip ML faces serious challenges due to the resource constraints of the implanted chip.

First, the implemented classifiers are usually primitive compared to what can be achieved on software or unconstrained hardware. Sophisticated methods such as deep neural networks (DNNs) are currently not suitable for implantable chips [215]. Consequently, simplified versions of SVMs [158, 216, 217] and decision trees [218] have been popular choices for on-chip implementation.

Secondly, even with a simple classifier and a moderate number of channels, an on-chip ML implementation requires a large silicon area. To illustrate, the single-die demonstration by O'Leary *et al.* [158] allocated 3.31 mm^2 to classify seizures from 32 channels, which was more than half of the total chip area. The multidie solution by Wang *et al.* [217] fabricated a separate 3.5 mm^2 chip in addition to the 8-channel sensing front-end. The high cost of on-chip ML is a result of the expensive computations needed for feature extraction and classification. The common feature extractors like finite impulse response (FIR) filter banks [158], fast Fourier transform (FFT) [216] or discrete wavelet transform (DWT) [217] processors require multiply-and-accumulate operations. Moreover, the coefficients for the feature extractor and the trained classifier model need to be stored on an on-chip memory.

Finally, the hardware needs to be optimized for a specific algorithm. ML for neuroscience is still youthful and fixing the algorithm at the implant level may limit the adaptability to future progress. Moreover, neural signals are notorious for their long-term and trial-to-trial variability, thus the algorithms generally require frequent re-adaptations [214]. Furthermore, although the reported seizure detection performances for on-chip classifiers have been high on hand-picked datasets, medical applications demand generality and robustness which currently have not been addressed due to the limited availability of patient data [219].

1.5 Summary

The quest for a better neural interface continues with compelling challenges. Recording and processing neural activity have been clinically and conceptually proven to benefit patients with neurological disorders; however, we need fully-implantable neural recording SoCs to enable widespread use of this promising technology with more advanced applications.

The number of channels that can be integrated within a chip depends on the energy and area consumption of the AFE. The vast literature on AFEs has continuously decreased the cost of recording and reached close to 1 fJ mm^2/c-s E-A FoM. However, it has to be pushed down further to allow thousand-channel systems within mm-scale footprints and mW-level powers. Chapter 2 will address this need.

Wireless power and data transfer is a necessity for safe and long-term operation in independent environments. For power transfer, energy-frugal operation is not only important for keeping the power source dimensions as small as possible, but also for guaranteeing the thermal safety of surrounding tissue. For data transfer, the increasing number of recording channels towards thousands results in data rates exceeding 100 Mbps. Wireless transmission at this rate becomes a serious bottleneck with the current standards.

On-chip compression methods have emerged as viable solutions for reducing the data throughput. However, these methods tend to partially lose or distort the frequency spectrum of neural signals, which could be detrimental to the many applications that rely on spectral content. On the other hand, we see ML algorithms more often as the end-user of the recorded data at the receiver side, due to the complex relationship between the neural signals and their outcome. Consequently, recent effort has been towards implementing on-chip ML. Although this approach eliminates the need for high-speed transmission, it consumes too much of the chip resources and does not seem to be a future-proof solution at the moment. The lessons learned from both compression and ML effort inspire a new framework presented in Chapter 3, which leverages machine learning to extract only the necessary information content.

2 An Energy-Area Efficient Wideband Neural Recording Front-End

Implantable neural recording demands the lowest energy and area consumption per channel to integrate as many channels as possible within strict resource constraints. In this chapter we address the challenge of increasing the resource efficiency of a neural recording AFE.

The conventional AFE depicted in Figure 2.1a is AC-coupled to the electrodes to reject large and variable DC offsets building on the tissue-electrode interface, as discussed in Section 1.2.2. The coupling capacitor also presents a high input impedance to the electrode, and serves as an isolation layer against static device currents and short circuits [220]. Despite its advantages, AC-coupled AFEs have limited scalability for multichannel integration. The capacitor area becomes a serious bottleneck when a high gain and a low high-pass pole is required as they both demand a very large coupling capacitor.

Several DC-coupled techniques have been proposed to eliminate bulky coupling capacitors, at the expense of other circuit qualities. Direct digitization (Figure 2.1b) by an ADC results in excessive input-referred noise [167]. Canceling the offset through a mixed-signal feedback loop (Figure 2.1c) [86] requires a well-matching digital-to-analog converter (DAC) and it has a digital power overhead proportional to the sampling rate. Delta modulating the input with switched capacitors (Figure 2.1d) can achieve rail-to-rail cancellation at the cost of reduced input impedance and increased noise at high sampling rates [110], which is also the case when the electrodes are multiplexed into a shared front-end (Figure 2.1e) [100]. Moreover, the effects of injecting switching currents into the neural tissue have not been addressed.

Due to these fundamental limitations of DC coupling, AC coupling is still favorable for wideband recording thanks to low noise, passive offset rejection, and high input impedance. This chapter describes a both energy- and area-efficient neural recording AFE that permits high integration density while keeping the benefits of AC coupling [1]. The architecture follows the conventional approach depicted in Figure 2.1a: a capacitively-coupled low-noise amplifier (CC-LNA) boosts and filters neural signals within LFP and AP bands, followed by a unit-length capacitor (ULC)-based SAR ADC. The sections of this chapter describe both parts and present the measurement results of a prototype.

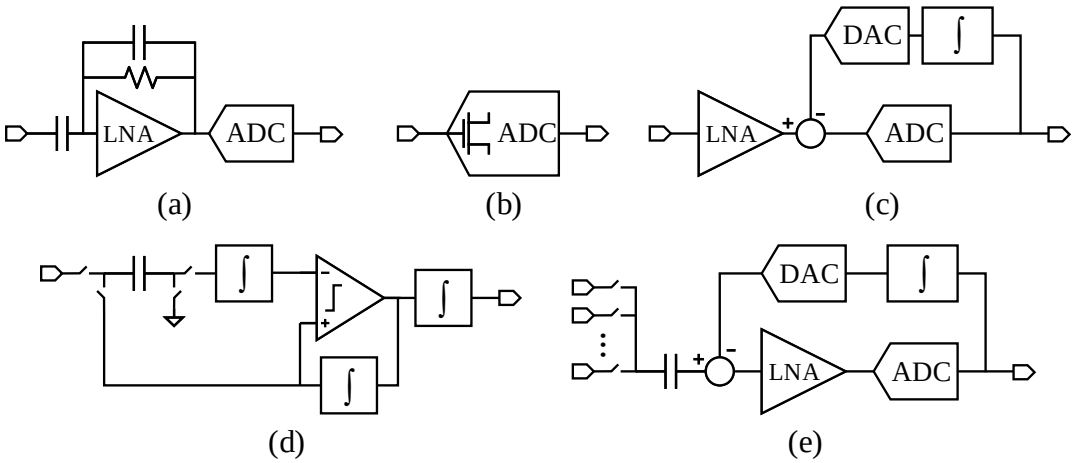


Figure 2.1 – (a) AC-coupled front-end architecture, and DC-coupled area reduction techniques using (b) direct digitization [167], (c) mixed-signal offset cancellation [86], (d) switched-capacitor delta modulation [110], and (e) electrode multiplexing [100].

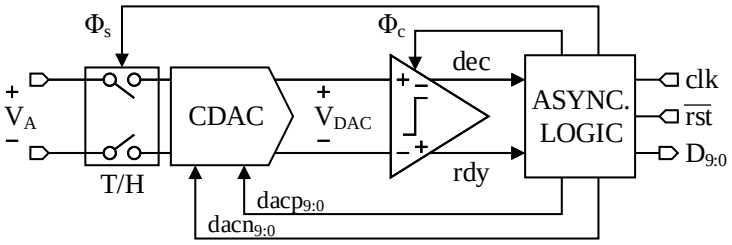
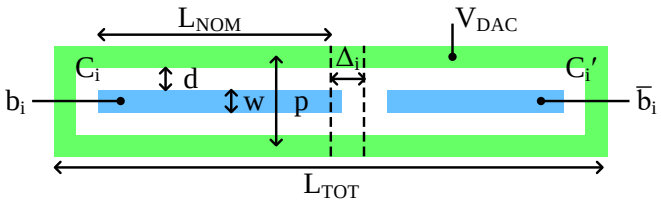


Figure 2.2 – Top level schematic of the SAR ADC highlighting the main components.

2.1 ULC-Based Asynchronous SAR ADC

We start with the ADC as it is the key element in the efficiency of the proposed AFE. We chose an asynchronous SAR ADC since it can provide the digital sample within a single clock period, contrary to its synchronous counterpart which requires N+1 clock periods for N-bit resolution. Therefore, it works at the same clock frequency as the subsequent digital stages and does not require a higher clock in the system. In addition to timing benefits, it also permits a smaller capacitor array as the sampled charge does not need to be held for longer than a fraction of the clock period.

Figure 2.2 shows the top-level block diagram of the 10-bit asynchronous SAR ADC. The track and hold switch (T/H) stores the amplifier output (V_A) on the capacitive digital-to-analog converter (CDAC). The voltage on the DAC (V_{DAC}) is controlled by the DAC control bits ($dacp_{9:0}$, $dacn_{9:0}$) which are set by the logic block. The logic block realizes the succesive approximation program based on the the decision (dec, rdy) of the comparator. It also generates the required asynchronous clocks for the sampling switch (Φ_s) and the comparator (Φ_c), and registers the final output ($D_{9:0}$). The following subsections explain each block in detail.



L_{TOT} = 16.48µm, L_{NOM} = 7.99µm, L_Δ = 200nm
p = 500nm, w = 120nm, d = 130nm, $\Delta_i = 2^{i-1} \times L_\Delta$

Figure 2.3 – Layout of a unit-length capacitor and the dimensions used in this work.

2.1.1 Unit-Length Capacitor Array DAC (ULCDAC)

The power and area of a SAR ADC strongly depend on the size of the CDAC. In a conventional N-bit CDAC, each binary weight W_i is constructed by using 2^i unit capacitors C_u in parallel, such that the total capacitance C_T is given by

$$C_T = \sum_{i=0}^{N-1} W_i = \sum_{i=0}^{N-1} 2^i C_u = (2^N - 1) C_u \,. \tag{2.1}$$

Since C_T grows exponentially with increasing number of bits, making the unit capacitor as small as possible is desirable for an efficient design. The lower bound for the unit capacitor is theoretically set by the kT/C thermal noise level. In order to keep the thermal noise below the quantization noise,

$$\frac{kT}{C_T} < \frac{V_{LSB}^2}{12} \quad \Rightarrow \quad C_u > \frac{12kT}{(2^N - 1) V_{LSB}^2} \,, \tag{2.2}$$

where V_{LSB} is the quantization step. According to (2.2), the unit capacitor has to be larger than 51 aF for a 10-bit ADC with 1 V full-scale voltage ($V_{LSB} = 1/1024 = 977\mu V$). Such value is much smaller than the library capacitors in modern CMOS processes. Therefore, a custom capacitor array has to be built to avoid overdesign.

In the context of this thesis, improving the area efficiency of the overall 10-bit AFE consists in reducing the CDAC area. The CDAC used in this work is based on the ULC concept, which was originally proposed by Harpe in [221]. A ULC relies on the difference of two metal capacitors, C_i and C_i', formed by the lateral-field fringing capacitances of three metal plates on the same layer as drawn in Figure 2.3. The encapsulating plate shown in green is the common DAC voltage V_{DAC}, and the center plates shown in blue are driven by complementary control voltages b_i and $\overline{b}_i$. If the break point of the center plates is at the center of the total length L_{TOT}, $C_i = C_i' = C_{NOM}$.

A binary weight W_i can be constructed by moving the break point of the center plates by $\pm\Delta_i$, where Δ_i is a multiple of the unit length L_Δ:

$$W_i = C_i - C_i' = C_{NOM} + C_{\Delta_i} - (C_{NOM} - C_{\Delta_i}) = 2C_{\Delta_i}, \quad \Delta_i = 2^i L_\Delta \,. \tag{2.3}$$

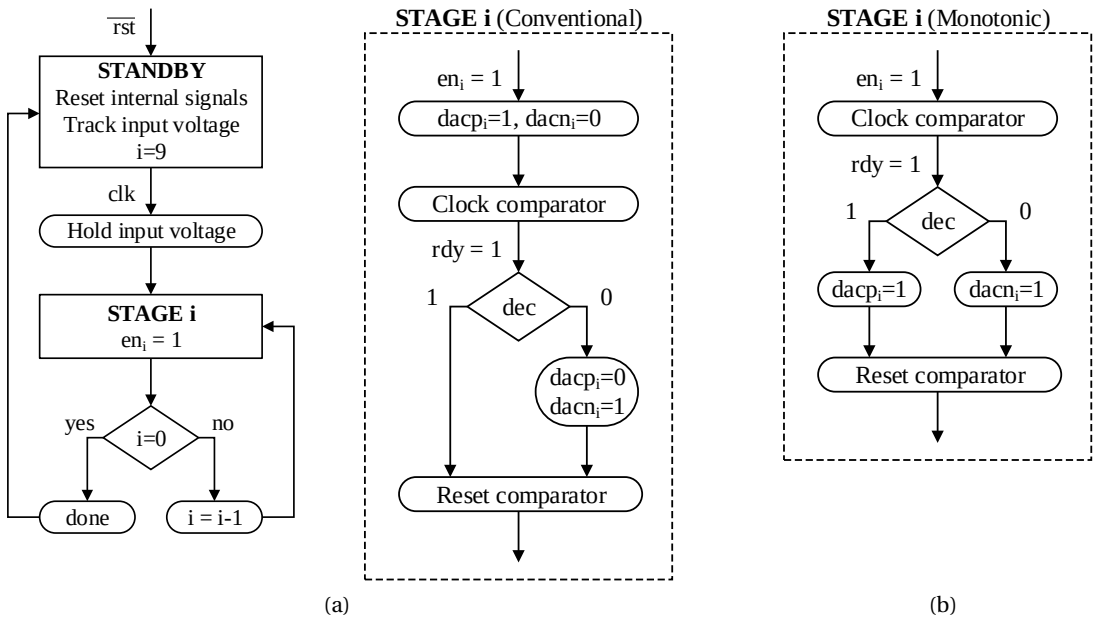


Figure 2.4 – (a) Conventional switching scheme. (b) Monotonic switching scheme.

As a result, an N-bit unit-length capacitor array digital-to-analog converter (ULCDAC) requires N ULCs instead of $2^N - 1$ unit capacitors. Moreover, the ULCDAC can be constructed using few metal layers and placed above the other circuits.

The ULCDAC implemented in [221] uses the conventional switching scheme to simplify the logic circuit driving the ULCDAC control signals. The flowchart of this scheme is given in Figure 2.4a. In the beginning of each conversion stage, the control bits are preset, and depending on the comparator decision, they are either kept or inverted. Inversion means wasted charge on both the DAC plates and the logic circuit, thus this scheme is not energy efficient.

To obtain further energy savings for multichannel neural recording, we adapted the ULCDAC concept in [221] with a more energy-efficient switching scheme. Monotonic switching described by Liu *et al.* [222] and illustrated in Figure 2.4b avoids the charge waste problem of the traditional scheme by manipulating the control voltages only after the comparator decision. Another benefit is that the first comparison is made on the sampled voltage; therefore, the capacitor with the largest weight is eliminated and the array size is halved.

The main drawback of monotonic switching is that the common-mode voltage of the differential V_{DAC} approaches to the ground as the conversion progresses which adversely effects the speed and noise of the comparator. A solution is to employ a complementary CDAC that switches in the opposite direction, as implemented by Kull *et al.* [223]. This requires a second CDAC switching in the opposite direction, but brings no area overhead as the single CDAC size is halved thanks to monotonic switching. The resulting energy efficiency of the monotonically-

switched constant common-mode ULCDAC is 10% lower than the its conventionally-switched counterpart, without an area penalty.

The schematic of the overall ULCDAC is shown in Figure 2.5a. It comprises 4 single-ended ULCDACs, named DACP+, DACP-, DACN+, and DACN-. When DACP+ switches and changes V_{DAC+}, DACP- switches oppositely and causes an equal amount of opposite change in V_{DAC-} to maintain a constant common-mode differential voltage. The same applies to DACN+ and DACN-. The capacitor pairs C_i and C_i' driven by complementary control signals represent ULC binary weights. The reference voltages are supply (V_{DD}) and ground which are delivered to the plates via simple CMOS inverters and buffers.

The layout of the ULCDAC is shown in Figure 2.5b. The total number of ULCs in the overall ULCDAC is 76. Each single-ended ULCDAC is segmented into 4-bit binary least significant bit (LSB) and 4-bit unary most significant bit (MSB) arrays to maintain a square shape and to keep mismatch under control as explained in [221]. The remaining two bits are implicit as the first comparison is performed on the sampled input, and the LSB weights are implemented single-ended.

In this work the ULC is constructed in M6 and M7 layers in parallel using the dimensions in Figure 2.3. The ULCDAC is placed above the other ADC blocks with shielding in M4 and M5 layers. The resulting C_{Δ_0} and C_{NOM} are 0.1 fF and 3.9 fF, respectively.

The total capacitance C_{TOT} of the ULCDAC on which the input voltage is sampled is given by

$$C_{TOT} = (76 \times 2C_{NOM}) + C_{PAR} = 388.04fF\,, \tag{2.4}$$

where C_{PAR} is the parasitic capacitance of the plates and the routing. However, the effective capacitance C_{EFF} differs from the total capacitance as the switching operates only on C_{Δ_i}:

$$C_{EFF} = 4 \times (2^9 - 1) \times C_{\Delta_0} = 204.40fF\,. \tag{2.5}$$

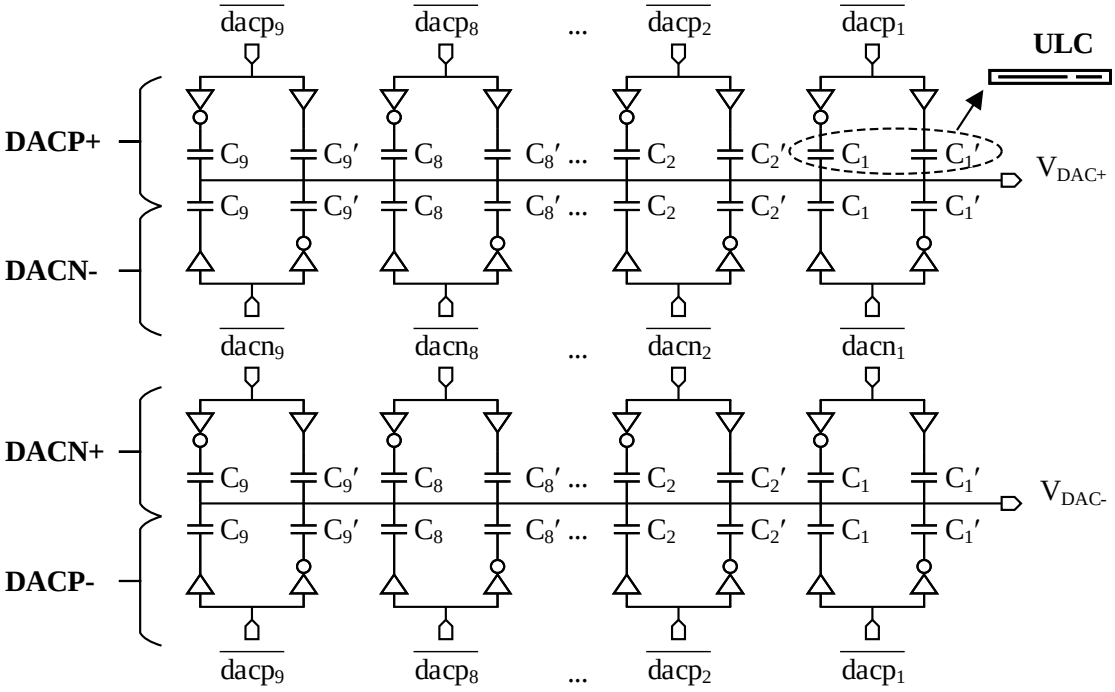
Consequently, the sampled voltage experiences an attenuation, reducing the full-scale voltage of the ADC to a fraction of the supply voltage:

$$V_{FS} = \pm V_{DD} \times \frac{C_{EFF}}{C_{TOT}} = \pm 0.527V\,. \tag{2.6}$$

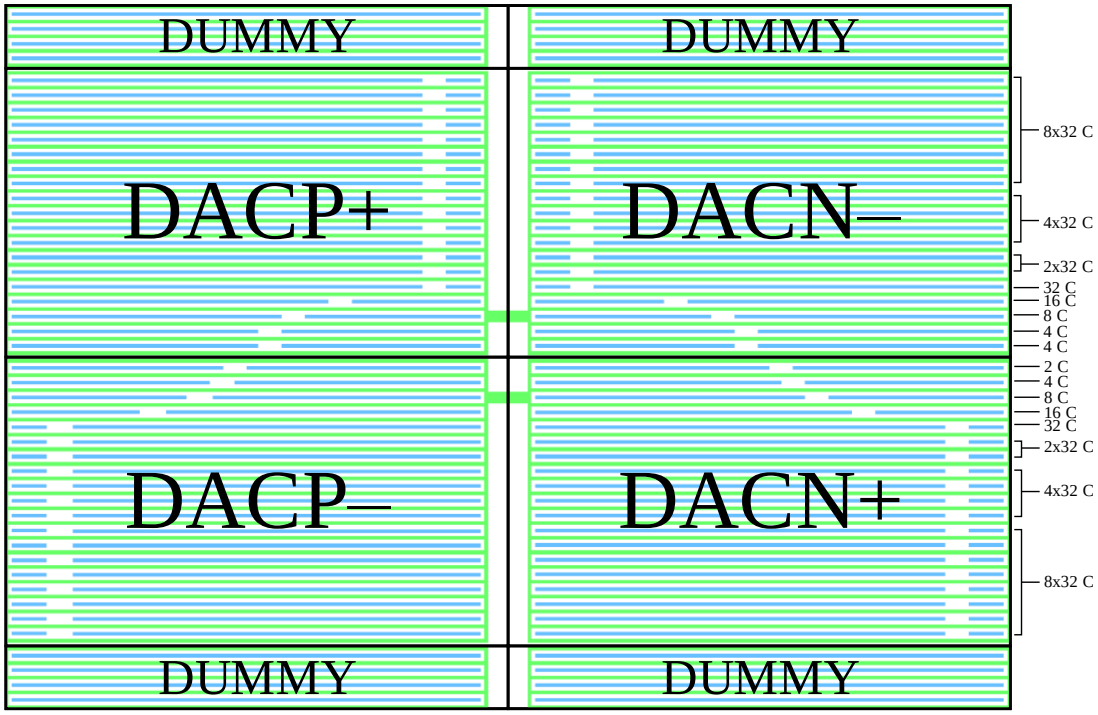
Although this reduction can be a drawback in some cases, in our specific application where the ADC is preceded by an amplifier, it has further area-saving benefits for the LNA design which will be discussed in Section 2.2.

2.1.2 Track-and-Hold Switches

The most basic sampling circuit for ADCs is the CMOS switch shown in Figure 2.6a, where the input voltage is connected to the CDAC via parallel NMOS and PMOS transistors. Although it



(a)



(b)

Figure 2.5 – (a) Schematic and (b) layout of the constant common-mode ULCDAC.

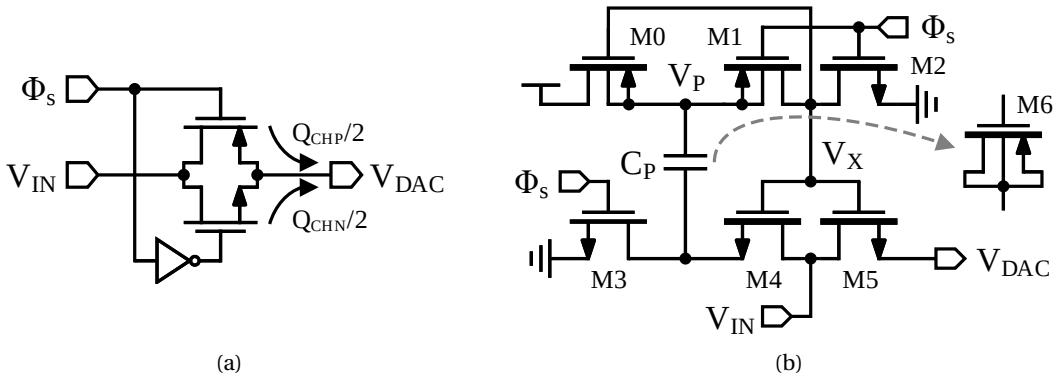


Figure 2.6 – (a) CMOS sampling switch and (b) the bootstrapped switch used in this work.

is small in size, this method suffers from a nonideality called charge injection which prevents its use above 8-bit resolution. The charge accumulated in the channels of the transistors when the switch is on, Q_{CHP} and Q_{CHN}, are given by

$$Q_{CHP} = W_P L_P C_{OX}(V_{SGP} - |V_{THP}|) = W_P L_P C_{OX}(V_{IN} - |V_{THP}|)\,, \tag{2.7}$$

$$Q_{CHN} = W_N L_N C_{OX}(V_{GSN} - V_{THN}) = W_N L_N C_{OX}(V_{DD} - V_{IN} - V_{THN})\,. \tag{2.8}$$

When the switch turns off, approximately half of the sum of these input-dependent charges is injected onto the CDAC. This causes a significant distortion ΔV_{DAC} on the sampled voltage especially when the CDAC has a low capacitance:

$$\Delta V_{DAC} = \frac{Q_{CHP} + Q_{CHN}}{2C_{DAC}}\,. \tag{2.9}$$

A solution to the input-dependent charge injection problem is the bootstrapped switch shown in Figure 2.6b. During hold mode, Φ_s is high, the storage capacitor C_P is precharged to V_{DD}, and M5 is open. During tracking mode, Φ_s is low and C_P is connected in between V_{IN} and V_X such that M5 opens with a constant V_{GS} of V_{DD} independent of V_{IN}. In this case, the charge injected onto V_{DAC} is constant and it is canceled thanks to differential operation.

All transistors in Figure 2.6b are high-voltage type in order to withstand internal voltages exceeding the nominal supply voltage. The storage capacitor C_P is generally implemented with a metal-oxide-metal (MOM) capacitor to reduce leakage. However, this would require several layers of metal and prevent the switches to be placed under the ULCDAC, which already takes 4 of the 7 available layers. This problem was solved here by using a PMOS transistor (M3) as C_P since it has superior capacitance density compared to other available passive devices. The leakage was verified to allow down to 500 Hz sampling rate and the resulting linearity is above 15 bits across PVT corners and mismatch.

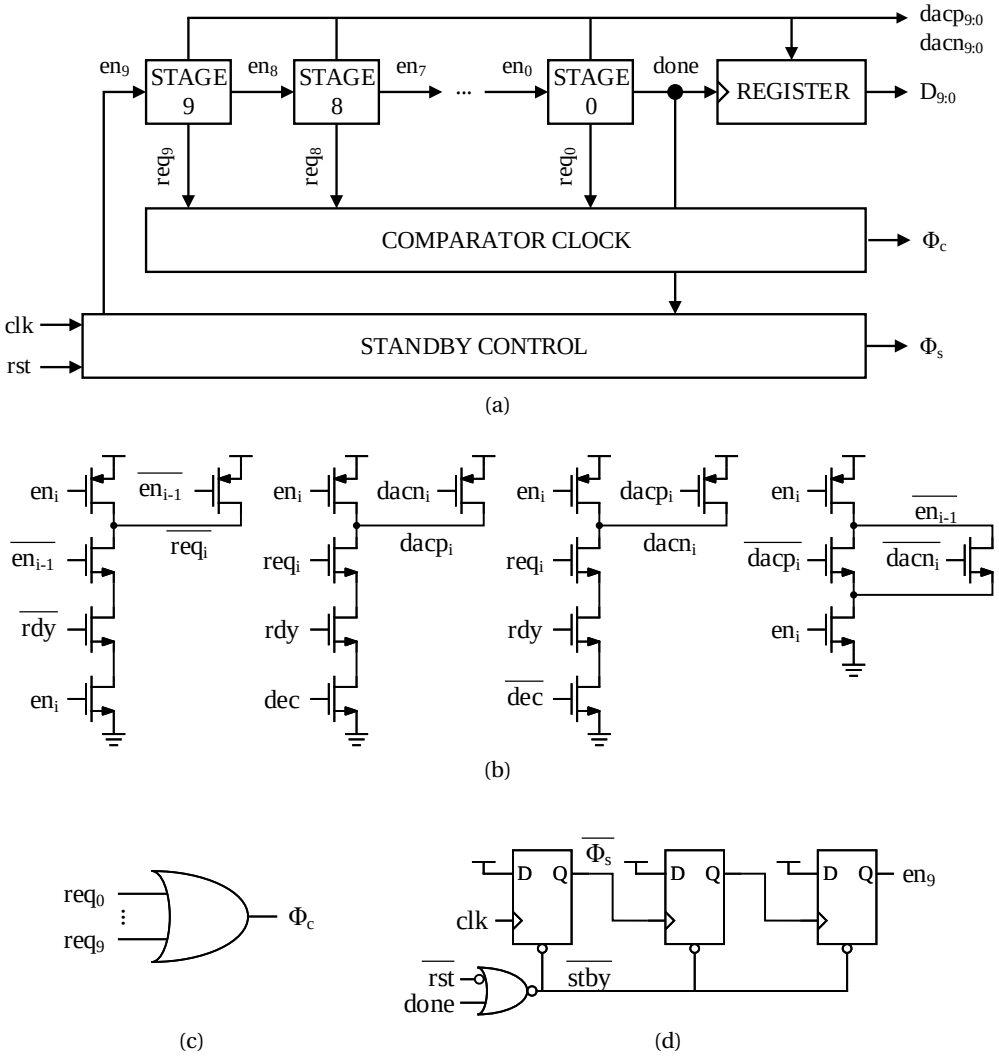


Figure 2.7 – (a) Asynchronous logic block and its main components: (b) standby controller, (c) comparator clock generator, and (d) unit conversion stage.

2.1.3 Asynchronous Logic

The asynchronous logic essentially realizes the flowchart given in Figure 2.4b. It produces the internal clocks, adjusts the ULCDAC, and registers the comparator decisions. The block diagram in Figure 2.7a denotes four main components: a standby control, a comparator clock generator, 10 conversion stages representing 10 bits, and a register to store the output. The circuit is a hybrid implementation of dynamic and static CMOS logic techniques to achieve low-power and robustness.

The ADC starts at the standby mode where it tracks the input voltage (Φ_s low). Upon a rising clock edge, the ADC holds the input voltage (Φ_s high), leaves the standby mode, and executes the chain of 10-unit stages. Each unit stage is realized by the circuit shown in Figure 2.7b. When

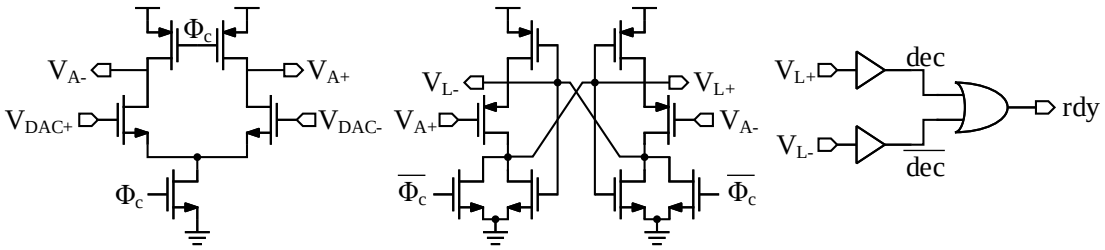


Figure 2.8 – Dynamic double-tail comparator.

a stage is enabled (en_i), a comparison request (req_i) is asserted. The comparator is clocked (Φ_c high) whenever a request is asserted as realized by the OR gate shown in Figure 2.7c. When the comparator decision (dec) is ready (rdy high), the appropriate ULCDAC plate ($dacp_i$ or $dacn_i$) is set, and the next stage (en_{i-1}) is activated. The last stage asserts the done signal which saves the final output on the register and instructs the standby controller in Figure 2.7d to reset the internal signals until the next conversion.

2.1.4 Dynamic Double-Tail Comparator

The comparator employs the well-known dynamic double-tail architecture whose schematic is shown in Figure 2.8. The input transistors and the tail switch of the first stage are sized to be in the weak inversion regime for keeping the input noise level low. The simulated gain of the first stage is 12. The second stage uses high-Vt PMOS and low-Vt NMOS transistors to ensure a low switching threshold for the cross-coupled latch. This provides enough latency for the first stage output difference to grow. The output of the latch is buffered, and a ready signal is issued after each decision. The two latch outputs are at high state during pre-charge, therefore a simple OR gate is able to produce the ready signal.

2.2 Capacitively-Coupled Low-Noise Amplifier (CC-LNA)

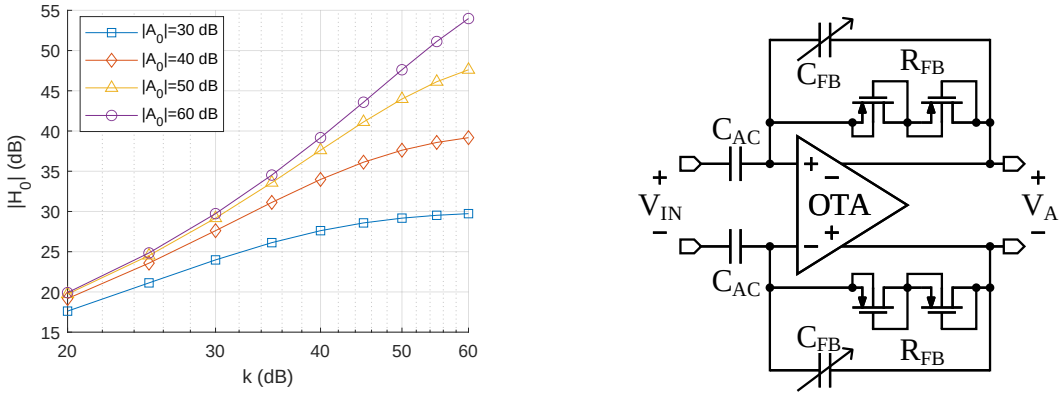
For this work, the signal swing at the electrode is expected to be 10 mV, following the commercial specifications [224]. Given the ~530 mV input range of the ADC, the required closed-loop gain for the CC-LNA is 35 dB.

2.2.1 Gain Requirement

The closed-loop transfer function H(s) of a feedback amplifier is given by

$$H(s) = \frac{A(s)}{1 + \beta(s)A(s)}, \tag{2.10}$$

where $A(s)$ is the open loop transfer function of the amplifier, and $\beta(s)$ is the feedback transfer function. For the CC-LNA, $A(s)$ can be approximated as a single-pole system with DC gain A_0



(a) (b)

Figure 2.9 – (a) Required k to achieve a target H_0 for various A_0. (b) Fully-differential capacitively-coupled LNA.

and and cutoff at ω_{OL}, and $\beta(s)$ is the capacitor ratio $C_{in}/C_{fb} = 1/k$:

$$A(s) = \frac{A_0}{1+\frac{s}{\omega_{OL}}}, \; \beta(s) = \frac{1}{k}. \tag{2.11}$$

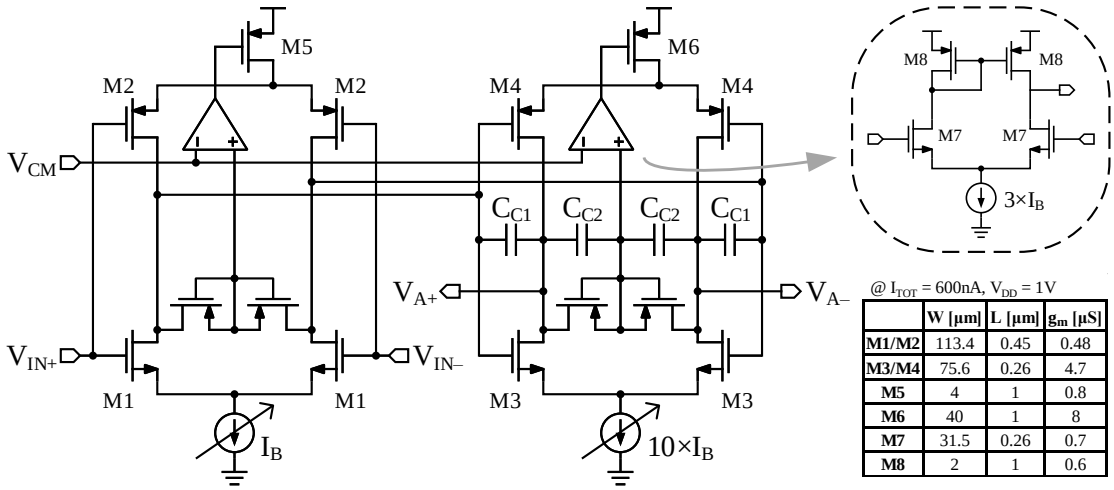
Substituting (2.11) in (2.10), we can obtain a representation

$$H(s) = H_0 \frac{1}{1+\frac{s}{\omega_{CL}}}, \text{ where } H_0 = \frac{k}{1+\frac{k}{A_0}}, \; \omega_{CL} = \omega_{OL}(1+\frac{A_0}{k}). \tag{2.12}$$

(2.12) suggests that as A_0 increases, the closed-loop gain will converge asymptotically to k. From this relation, we can determine A_0 and k based on a target H_0. According to Figure 2.9a, the open-loop gain of the OTA A_0 needs to be greater than 50 dB to be able to set the closed-loop gain H_0 accurately with k.

AC-coupled amplifiers rely on capacitive feedback to realize the desired gain. There are two different approaches to implement a capacitive feedback network: a single capacitor, or a capacitive T-network. As analyzed both theoretically and experimentally in [160], the capacitive T-network brings significant area savings, at the expense of higher input-referred noise. Consequently, the choice of the method depends on what the specific application favors. In our case, area gains in the ADC make it easier to favor low input-referred noise to achieve lower NEF. Therefore, we have chosen to implement the single capacitor feedback structure, which is also easier to place and has less routing parasitics.

Figure 2.9b shows the CC-LNA schematic. The nominal gain of the amplifier is set by the capacitor ratio C_{AC}/C_{FB}. C_{AC} was chosen as 2 pF, and C_{FB} can be configured from 15 fF to 65 fF corresponding to 30 to 40 dB programmable gain. Diode-connected PMOS feedback resistors (simulated nominal R_{FB} = 50 TΩ) ensure that the high-pass cut-off at $1/(2\pi R_{FB}C_{FB})$



@ I_{TOT} = 600nA, V_{DD} = 1V

	W [µm]	L [µm]	g_m [µS]
M1/M2	113.4	0.45	0.48
M3/M4	75.6	0.26	4.7
M5	4	1	0.8
M6	40	1	8
M7	31.5	0.26	0.7
M8	2	1	0.6

Figure 2.10 – Transistor-level schematic of the two-stage inverter-based OTA.

stays below 1 Hz for all configurations of C_{FB} across corners. The low-pass cut-off can be adjusted by changing the bias current of the OTA.

2.2.2 Operational Transconductance Amplifier (OTA)

The fully-differential two-stage OTA is shown in Figure 2.10. Based on the review of candidate topologies in Section 1.2, an inverter-based topology was chosen. This topology has high noise efficiency thanks to current being reused by the complementary NMOS and PMOS differential pairs.

All pairs are constructed with thick-oxide transistors to operate in weak inversion regime. Using the weak inversion expressions of g_m and g_{ds}, the gain of a single stage can be written as

$$A_V = \frac{g_m}{g_{ds}} = \frac{I_D}{nU_T}\frac{\lambda L}{I_D} = \frac{\lambda L}{nU_T}\,, \tag{2.13}$$

which depends mostly on the length of the transistors.

The total gain of the two stages is 63 dB. The first stage provides 38 dB gain, the second stage adds 25 dB and drives the ADC input. The output common-modes of two stages are sensed via diode-connected transistors and set nominally to $V_{DD}/2$ by individual common-mode feedback (CMFB) amplifiers. Compensation capacitors (C_{C1} = 240 fF, C_{C2} = 35 fF) and the ratio of tail currents ensures stability in all corners.

The gain-bandwidth of the amplifier is expressed by

$$GBW = \frac{g_{m1}}{C_{C1}} = \frac{I_D}{nU_T C_{C1}}\,. \tag{2.14}$$

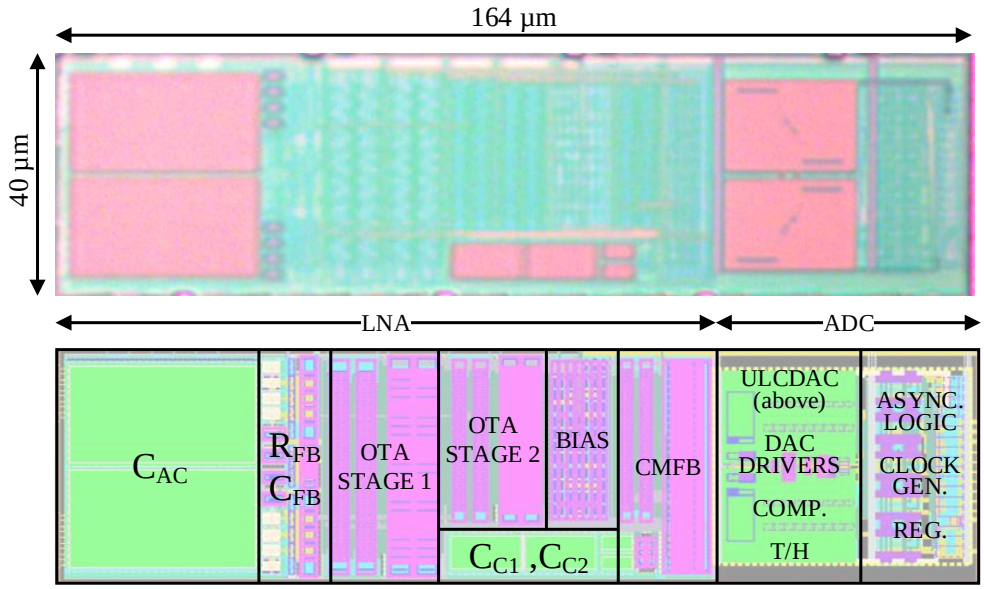


Figure 2.11 – The annotated layout and the micrograph of the analog front-end fabricated in TSMC 65nm LP process.

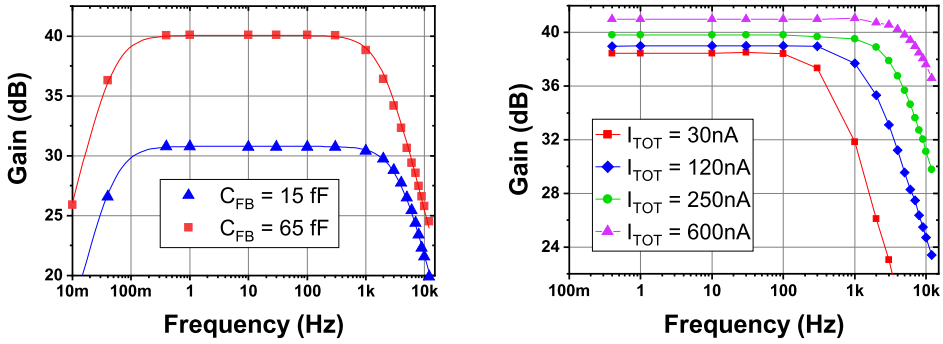
Therefore, the bandwidth of the amplifier can be changed by adjusting I_D with minimal effect on the other parameters. When the total current is 600 nA, the gain-bandwidth of the two-stage amplifier is approximately 2 MHz.

2.3 Measurement Results

Figure 2.11 displays the annotated micrograph of the prototype fabricated in TSMC 65nm 6X1Z1U LP CMOS. The total channel area is 164 µm×40 µm (6560 µm^2) and the ULC-based SAR ADC takes 35 µm×45 µm (1575 µm^2). The AFE was characterized first electrically, then connected to a soft µECoG array [225] for *in vitro* validation.

2.3.1 Bench Characterization

The measured LNA responses for different gain and bandwidth configurations are given in Figures 2.12a and 2.12b. The gain can be modified between 30.8 dB to 40.1 dB by trimming the feedback capacitor C_{FB}. The high-pass cut-off frequency is around 0.05 Hz. The bandwidth can be adjusted between 500 Hz and 10 kHz by adjusting the bias current, which correspond to 30 nA and 600 nA total supply current from 1 V supply, respectively. The total harmonic distortion (THD) of the amplifier is 1.1% when 8 mV peak-to-peak differential input applied with maximum gain, and the common-mode rejection ratio (CMRR) is 56 dB. As shown in Figure 2.13, the input-referred noise (IRN) is 1.9 µV$_{rms}$ over the LFP band (1 Hz-500 Hz), 2.5 µV$_{rms}$ over the AP band (500 Hz-10 kHz), and 3.1 µV$_{rms}$ over the full bandwidth (1 Hz-10 kHz).



(a) (b)

Figure 2.12 – Measured LNA Bode plots for different (a) gain and (b) bandwidth configurations.

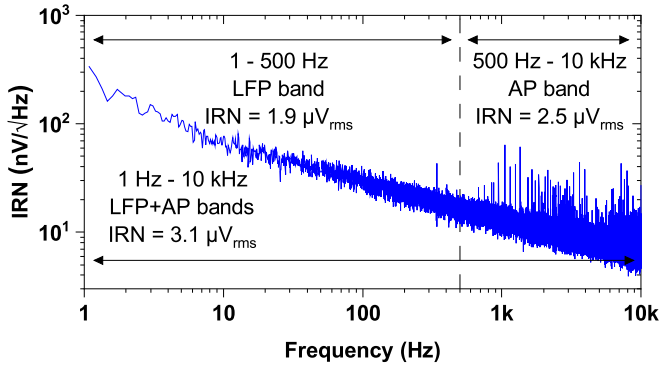
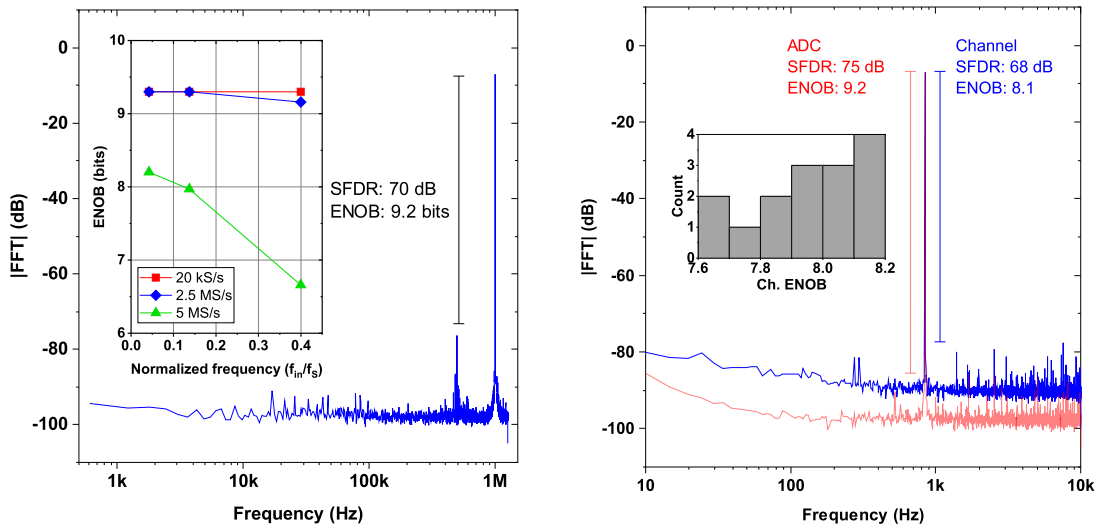


Figure 2.13 – Measured input-referred noise spectrum of the front-end.

Figure 2.14a shows the measured output spectrum of the standalone ADC and the variation of effective number of bits (ENOB) across sampling rates. The ADC is able to achieve 70 dB spurious-free dynamic range (SFDR) and 9.2 ENOB over the entire bandwidth up to 2.5 MS/s, after which the performance drops. The ADC consumes 47 nW when operating at 20 kS/s, and the consumption scales linearly with the sampling rate. The input range of the ADC was measured as 0.69 V_{ppd}.

The output spectrum of the complete channel in comparison with the standalone ADC response is given in Figure 2.14b. The channel performance reduces to 8.1 ENOB and 68 dB SFDR when the LNA is connected. This is due to the LNA increasing the overall noise floor, and the increased nonlinearity due to the MOS feedback resistors (R_{FB}) of the LNA. The histogram in the inset of Figure 2.14b depicts the intradie variation of ENOB across 15 measured samples which reflects the ULCDAC mismatch. Although this variation could be improved with a small resource penalty by increasing the ULC L_Δ if necessary, the worst-case performance still facilitates the 50 dB dynamic range of neural signals [86].



(a) (b)

Figure 2.14 – Measured output frequency spectrum of (a) the standalone ADC and (b) the overall channel.

2.3.2 In Vitro Characterization

Figure 2.15 shows the test setup for *in vitro* measurements. The prototype was connected to a custom soft µECoG array, provided by Dr. Florian Fallegger of EPFL Laboratory for Soft Bioelectronic Interfaces. It was microfabricated using thin-film and silicon processing technology inspired from the e-dura process [225, 226]. The array was submerged into a phosphate-buffer saline (PBS) solution, together with electrodes driven by a signal generator. The board and the solution were placed inside a Faraday cage to prevent interference from affecting the measurements. Two tests were conducted using different signals. The first test was to validate the detection of intermittent pulses. 1 ms-long current pulses with 5 ms interpulse duration were applied from an A-M Systems Model 2100 pulse stimulator. Figure 2.16a displays the recorded pulse waveforms (average of 10, then superimposed) with respect to the applied current pulse amplitudes. As plotted in Figure 2.16b, the recorded amplitudes exhibit good linearity, except for the point at 5 mA which is an error due to premature sampling before the generator output stabilized.

The second test was to validate the spectral quality of the recordings. Sinusoidal voltages with various frequencies were applied to the solution from Analog Discovery 2. Figure 2.17a and Figure 2.17b show the recorded waveforms and their FFTs, respectively. The peaks of the FFT match the applied frequency. The dominance of low-frequency noise ($1/f^2$) below 500 Hz, the high-pass behavior observed from the amplitude of the peaks, and the strong harmonics are caused by the noise and impedance characteristics of the electrode and the electrode-electrolyte interface [227].

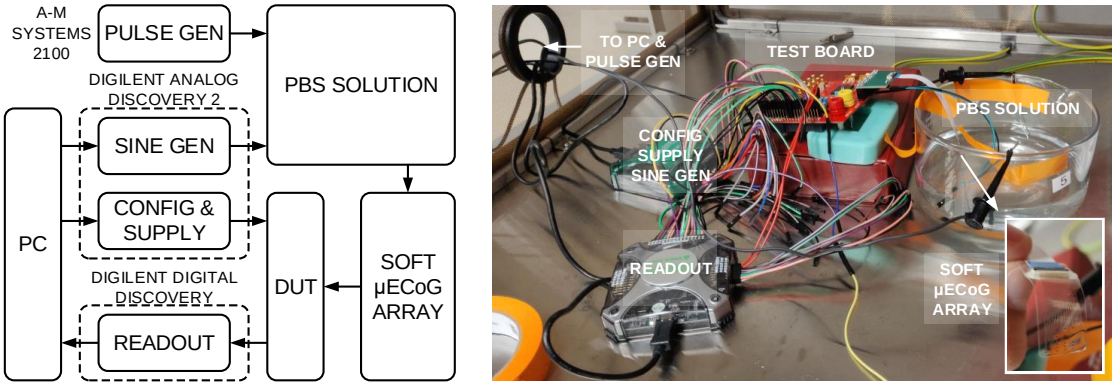
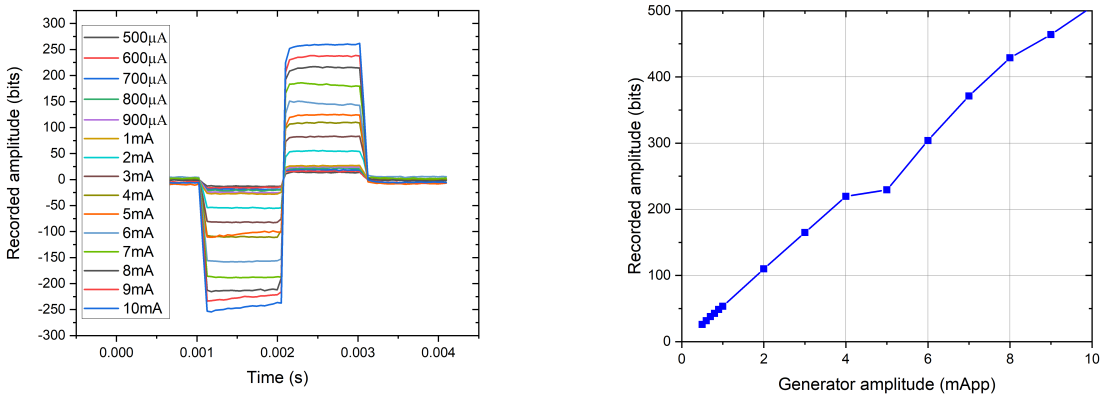
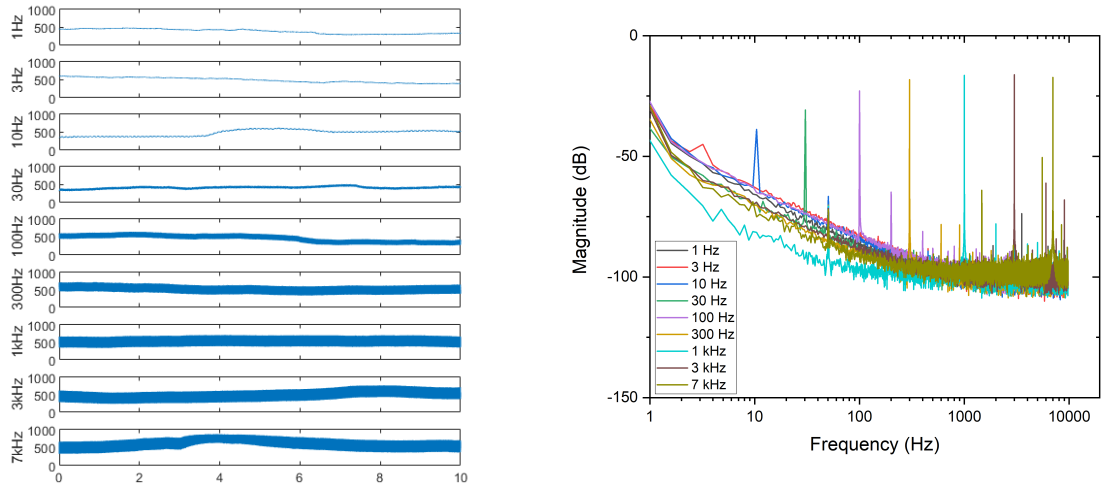


Figure 2.15 – Test setup for *in vitro* validation.



(a) (b)

Figure 2.16 – (a) The recorded pulse waveforms (average of 10 then superimposed) with respect to the applied current pulse amplitudes and (b) amplitude linearity.



(a) (b)

Figure 2.17 – (a) Time and (b) frequency domain plots of the recorded sinusoidal waveforms.

Table 2.1 – AFE performance summary and comparison with the state of the art.

	JSSC'12[86]	JSSC'15[135]	JSSC'17[110]	JSSC'18[116]	SSCL'18[152]	JSSC'18[167]	TBCAS'19[228]	TBCAS'20[100]	**This work**
Technology	65nm	65nm	130nm	180nm	65nm	180nm	180nm	65nm	**65nm**
V_{DD} [V]	0.5	1	1.2	0.5 / 1	0.6	1.8	0.5	2.5 / 0.5	**1**
Coupling	DC	AC	DC	AC	DC	DC	AC	DC	**AC**
Multiplexing	No	Yes	No	No	No	No	No	Yes	**No**
Bandwidth [Hz]	10-10k	10-8k	0.01-500	0.4-10.9k	0.1-500	0-10k	1-6.8k	1-1k	**0.05-10k**
Sampling Rate [kS/s]	20	20	1 [1]	25	1	20	31.25	2	**20**
IRN [μV_{rms}]	4.9	7.5	1.13	3.32	2.2	12.07 [1]	5.4	1.66	**3.1**
NEF/PEF	5.99 / 17.96	4.45 / 12.9	2.86 / 9.82 [1]	3.02 / 4.56	8.7 / 45.4	29.1 [1] / 1529 [1]	2.99 / 4.46	2.21 / 12.21 [1]	**0.97 / 0.94**
ADC Topology	VCO	SAR	Δ-ΔΣ	Δ-ΔΣ	VCO ΔΣ	OTA-C ΔΣ	Δ-SAR	Δ-SAR	**SAR**
Area/Channel [mm^2]	0.013	0.0258	0.013 [2]	0.058 [3]	0.01 [2]	0.0049 [2]	0.16 [3]	0.0023 [3]	**0.00656**
Power/Channel [μW]	5.04	1.84	0.63 [2]	3.05 [3]	3.2 [2]	39.14 [2]	0.88 [3]	2.98 [3]	**0.65**
Resolution [ENOB]	7.2 [4]	8.2 [4]	11.7 [1]@130 Hz	10.3@1kHz	8.18 [1]@40 Hz	8.2@1kHz	7.7 [4]	8 [5]	**8.1**
ADC FoM_W [fJ/c-s]	84	4.25 [1]	n/a	35.2	n/a	n/a	19.6	23.32 [1]	**4.0**
Channel FoM_W [fJ/c-s]	1713.9 [1]	312.86 [1]	189.36 [1]	108.83	11034 [1]	7180	135.43 [1]	5820 [1]	**118.5**
E-A FoM [mm^2×fJ/c-s]	22.28 [1]	8.07 [1]	2.46 [1]	6.34	110.34 [1]	35.182 [1]	21.67 [1]	13.39 [1]	**0.78**

[1] Estimated from published data. [2] On-chip decimation filter. [3] Off-chip decimation filter. [4] ADC only. [5] Above 200 Hz.

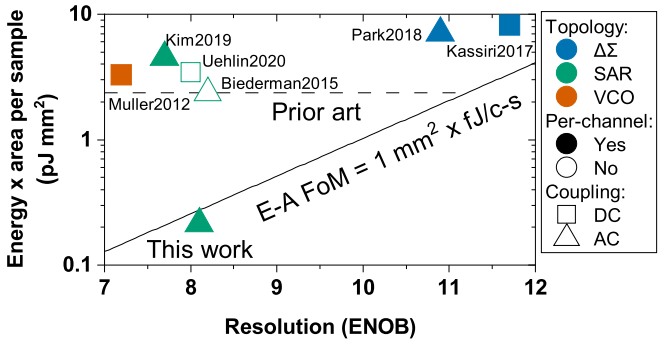


Figure 2.18 – Energy-area efficiency comparison with the state of the art.

2.3.3 Discussion

Table 2.1 summarizes the performance of the proposed neural recording front-end and compares it with the state-of-the-art works [86,100,110,116,135,152,167,228]. The silicon footprint of the front-end presented here is four times smaller than the smallest AC-coupled front-end reported, and it is comparable to the recent DC-coupled ones. The power consumption is the lowest among the wideband front-ends, which results in channel 0.97 NEF and 0.94 PEF.

The energy efficiencies (FoM_W) of the ADC and the channel align with the other architectures. However, the combined energy and area efficiency improves the state of the art in E-A FoM by three times. In other words, the energy-area cost per sample is 12 times lower than that of other similar resolution front-ends as compared in Figure 2.18. Moreover, the modularity of the presented architecture makes it easily scalable. A thousand-channel AFE would occupy less than 7 mm^2 footprint[1] and consume less than 1 mW.

[1] This estimate neglects the problem of pad arrangement. Flip-chip methods with the pads over the circuits could be a possible solution for such density.

3 A Wireless Neural Recording SoC for Machine Learning Applications

The translation of neurotechnology applications from laboratory to widespread use depends on the long-term operability in independent environments. Consequently, wireless connectivity is a desired feature for implants as it eliminates the infection-prone tethered connections. Nevertheless, transmitting the data rates reaching the 100 Mbps regime poses a grand challenge for moving towards thousand-channel systems.

We have reviewed the current approaches for data rate reduction in Section 1.4. Based on the fact that ML emerges rapidly as the end-user of neural data, the recent trend has been to explore on-chip ML. However, current demonstrations display a large energy and area need even with small number of channels, rendering this approach infeasible for massively parallel recording. Moreover, optimizing and fixing the hardware at the implant level does not lend itself to upgrades easily as the youthful field of ML progresses.

In this chapter we consider ML as a new opportunity in the context of compression and present a wireless multichannel neural recording SoC targeting ML applications [2]. Inspired by the advantages and disadvantages of the previous efforts, we adopt a new approach where feature extraction is performed on-chip to reduce the data rate, but the classifier is off-chip to reduce the die area and to maintain algorithm flexibility. As illustrated in Figure 3.1, the key concept here is that we train the on-chip feature extractor and the off-chip classifier jointly, such that the chip resources are used optimally.

Figure 3.2 shows the block diagram of the SoC. For this prototype we integrate 16 channels, but the concept can be extrapolated to any number of channels since the channel architecture is modular with an individual AFE and feature extractor. The AFE described in Chapter 2 is employed here. The feature extractor based on compressed Hadamard transform (CHT) is introduced in Section 3.1. The WPDT subsystem comprising an inductive power link and an impulse radio ultra-wideband (IR-UWB) transmitter is described in Section 3.2. Finally, Section 3.3 presents and discusses the performance of the proposed approach.

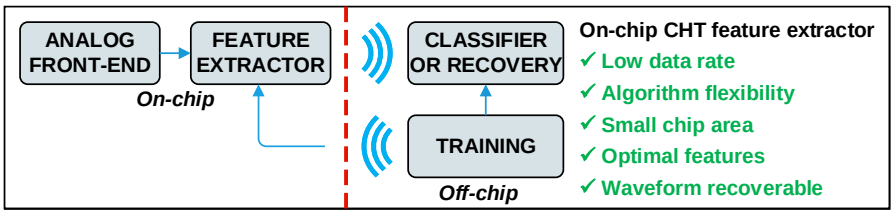


Figure 3.1 – Proposed neural recording approach where the on-chip feature extractor and the off-chip classifier are trained jointly.

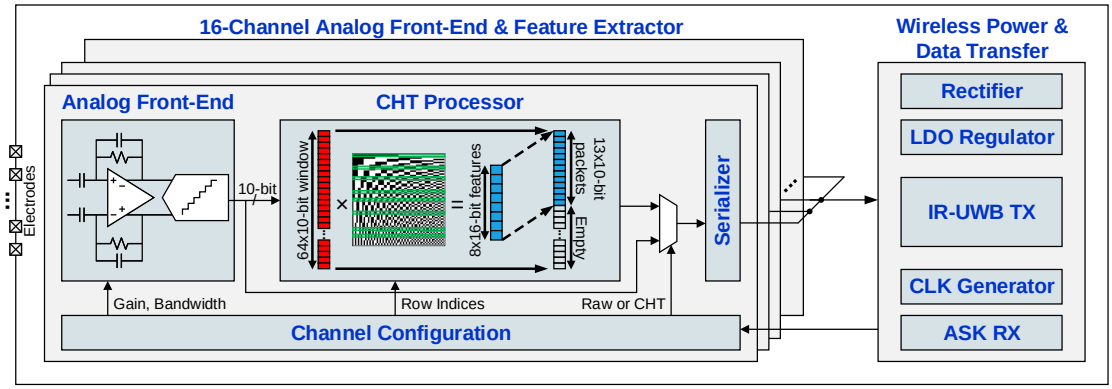


Figure 3.2 – Block diagram of the proposed neural recording system-on-chip.

3.1 On-Chip Feature Extractor

The feature extractor can be considered as an adaptation of the LBCS concept introduced in Section 1.4.1. Suppose that we have an annotated dataset where $\mathscr{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_m\}$ represents m different measurements, and $\mathscr{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_m\}$ represents their binary annotations. Instead of learning a subsampling mask Ω that captures the most of the energy in the signal, we learn the Ω that yields the best classification performance. In mathematical notation, this corresponds to solving the following minimization problem:

$$\hat{\Omega} = \underset{\Omega, |\Omega| = M}{\arg\min} \sum_{j=1}^{m} |\mathbf{y}_i - f(\mathbf{P}_\Omega \boldsymbol{\Psi} \mathbf{x}_j)| \,, \tag{3.1}$$

where f is the classifier model which transforms the selected features into a decision. In plain words, we learn the Ω that extracts for the most descriptive features.

A chief benefit of the proposed feature extractor is that the waveforms can still be reconstructed by changing the selection paradigm in (3.1) back to (1.7) which maximizes the retained energy. The reconstructed waveforms not only serve as means of diagnosis for long-term implantation, but also allow new features to be extracted on the receiver side if needed.

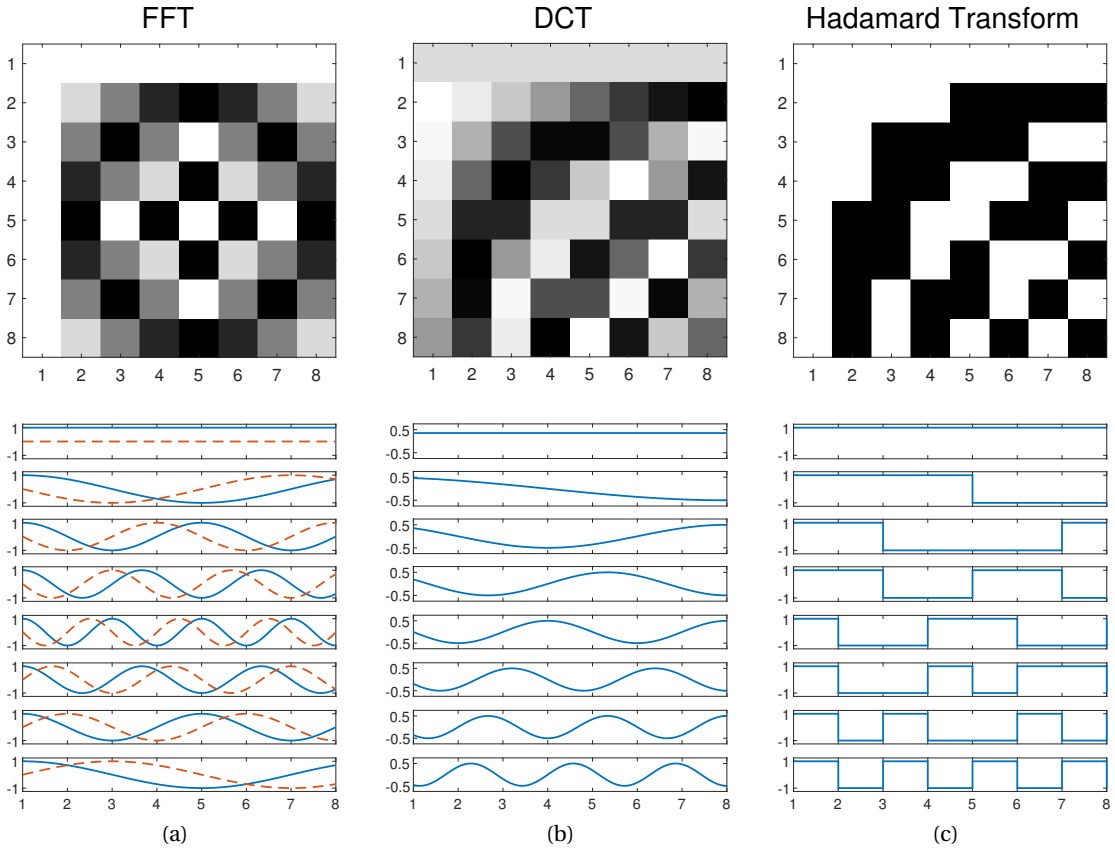


Figure 3.3 – FFT, DCT, and Hadamard transform matrices with their basis functions. 8-by-8 examples are shown for simplicity.

3.1.1 Compressed Hadamard Transform (CHT) Processor

As reviewed in Section 1.4.2, most of the neural recording applications use spectral band energies as features. Accordingly, the features we use in this work are based on Hadamard transform, which is a class of Fourier transforms like FFT and DCT. Figure 3.3 compares the transform matrices and the basis functions of the three transforms. FFT is a complex transform which decomposes the signal into magnitude and phase components. DCT operates only in the real domain, hence its computational complexity is less than FFT. Both FFT and DCT have sinusoidal basis functions, meaning that any hardware implementation needs to store the coefficients with a certain resolution. On the other hand, Hadamard transform is much more hardware-friendly compared to FFT and DCT since its coefficients are either +1 or -1.

Figure 3.4 illustrates the Hadamard transform of a time-series containing a seizure event. We can clearly see that only few indices contain significantly more energy during the seizure. Therefore, computing only these descriptive indices instead of the full transform is sufficient for classification purposes. We term the custom matrix formed by the selection of indices as CHT.

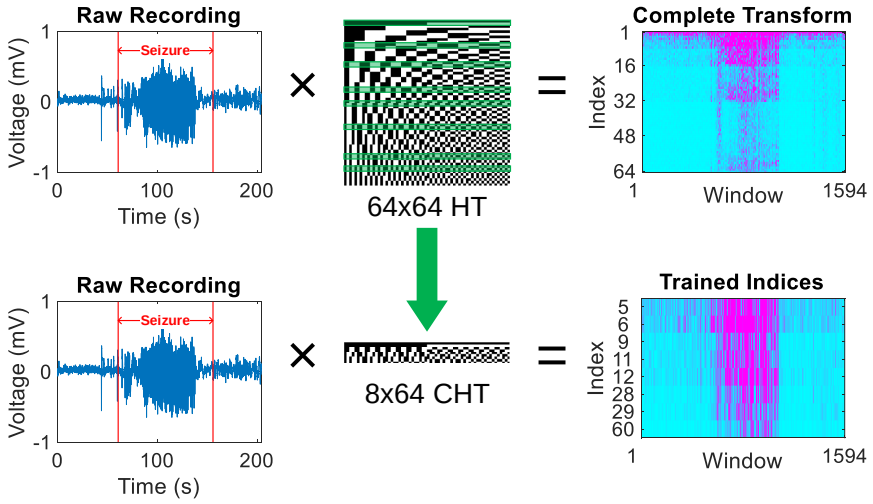


Figure 3.4 – Illustration of Hadamard transform during a seizure event. Selection of the descriptive indices form the Compressed Hadamard Transform (CHT) matrix.

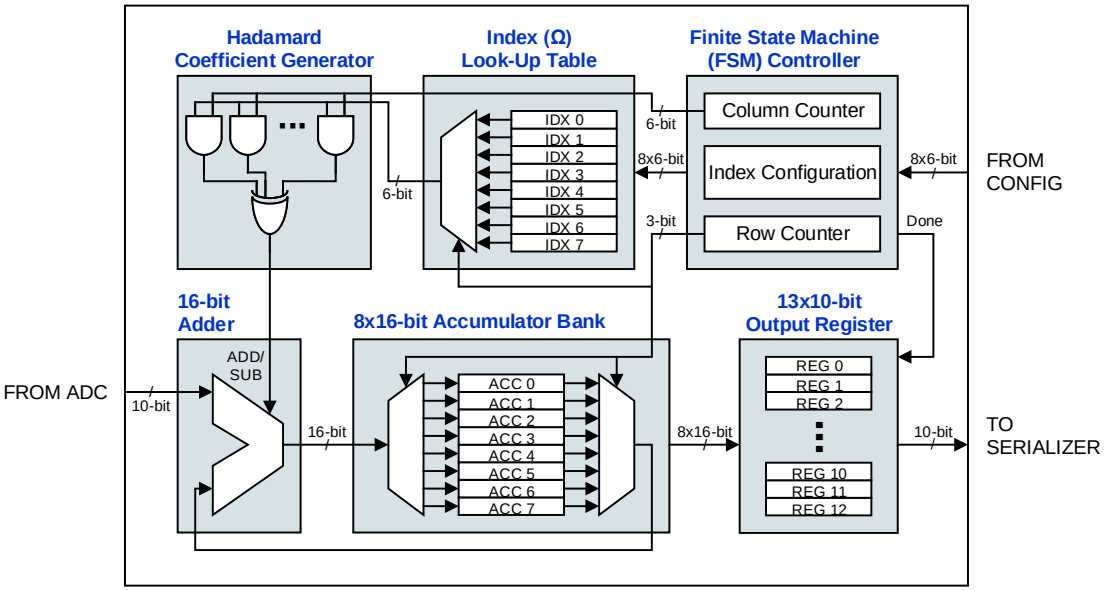


Figure 3.5 – Digital circuit implementation of the Compressed Hadamard Transform (CHT) processor.

The CHT features are computed by a processor which essentially implements the matrix product of a 64-sample input window with 8 selected rows from the 64-by-64 Hadamard matrix. Figure 3.5 shows the block diagram of the digital circuit. Since the Hadamard coefficients are ± 1, the product can be implemented by an adder. A 16-bit adder/subtractor is required since 64 successive additions are performed on the 10-bit ADC samples (10+log(64)=16). A bank of 8 16-bit accumulators, corresponding to 8 selected rows, hosts the intermediate results

during the 64 cycles. The processor runs 8 times faster than the sampling rate of the ADC such that a single adder could be shared to reduce the area. Row and column counters keep track of the operation sequence. The selected row index is fetched from the configurable look-up table based on the current row counter value. The Hadamard coefficient corresponding to the selected row index and the column counter value is generated on-the-fly by a combinational logic circuit, without a coefficient memory.

Each channel can be configured individually to output either the raw data recorded by the AFE, or the features computed by the CHT processor. When raw data is sent, the CHT processor is bypassed and the 10-bit stream is transmitted. When features are sent, the 64-sample window reduces to a feature vector of 8 features and the bit size increases from 10 bits to 16 bits. To simplify the output mapping of individually configurable channels, the 8×16-bit features are mapped into a 13×10-bit output register which is fed into the global 10-bit data stream.

The transformation of the 64×10-bit input window into the 8×16-bit feature vector yields an overall data reduction of 80%. This means that 5 times more channels can be recorded at the same data rate. To illustrate, the number of channels that could be accommodated by a state-of-the-art 40 Mbps UWB transmitter [110] could be increased from 200 to 1000.

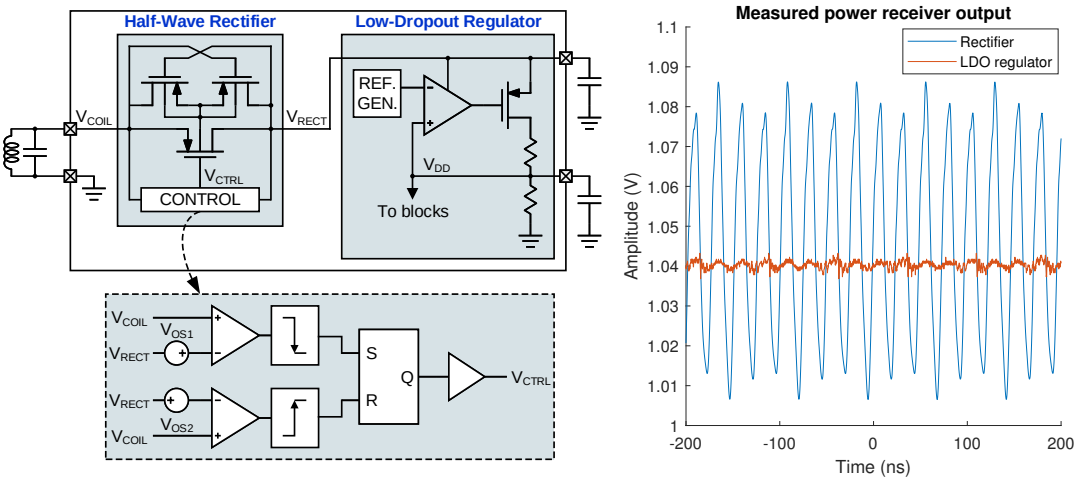
3.2 Wireless Power and Data Transfer (WPDT)

As reviewed in Section 1.3, inductive powering and UWB data transmission are the most suitable WPDT solutions for high-density implantable recording. This section briefly explains the WPDT subsystems included in the SoC prototype. The blocks described here were designed and verified by Dr. Kerim Ture of EPFL Radio Frequency Integrated Circuits Group, and full details can be found in his thesis [229].

3.2.1 Half-Wave Rectifier and Low-Dropout (LDO) Regulator

The on-chip power receiver blocks are shown in Figure 3.6a. The induced AC power on the antenna is first rectified to DC by a delay-compensated active half-wave rectifier, then regulated by a low-dropout (LDO) regulator.

The rectifier is composed of a PMOS pass transistor and control circuitry. The pass transistor charges the 400 nF off-chip reservoir capacitor only when the input is higher than the output, and turns off otherwise, hence the half-wave operation. The bulk of the pass transistor is dynamically biased by a composite PMOS diode to reduce the pass transistor's resistance and leakage. The timing of the switch is controlled by two comparators that detect the sign changes at the input and the output. The gate of the pass transistor is driven by a buffer stage. To improve the efficiency of the rectifier, the delay of the comparators and the gate driver are compensated by introducing offset to the comparators such that the control circuit is triggered before the input changes its sign.



(a) (b)

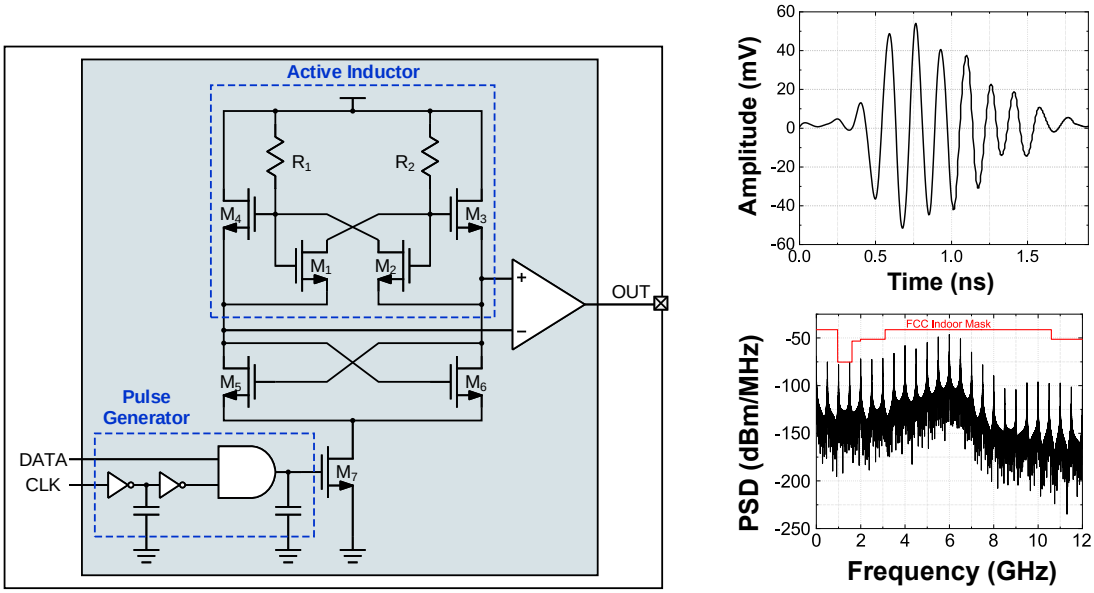
Figure 3.6 – (a) Schematic of the power receiver blocks and (b) the measured outputs.

The rectifier is followed by an LDO regulator which attenuates the ripples on the rectified voltage by 20 dB. The regulator output is stabilized by a 200 nF off-chip load capacitor. Figure 3.6b displays the measured output waveforms of the power receiver blocks. Overall, the power receiver provides 1 V supply voltage to the internal blocks.

3.2.2 Impulse-Radio Ultra-Wideband (IR-UWB) Transmitter

The raw data rate of the 16-channel, 20 kS/s, 10-bit AFE is 3.2 Mb/s. When only features are sent from all channels, the data rate reduces by 80% to 640 kb/s. When we extrapolate these numbers to a 1000-channel system, we observe the need for at least 40 Mb/s. As we have discussed in Section 1.3, UWB is the only suitable low-power method for such high data rates. Accordingly, the uplink communication for this prototype is established by an IR-UWB transmitter.

The IR-UWB pulses are generated by modulating a carrier frequency by fixed-duration pulses. The schematic of the IR-UWB transmitter is drawn in Figure 3.7a. The 6 GHz carrier frequency is generated by a an active inductor-based LC oscillator. The active inductor takes much less silicon area than a regular spiral inductor as it is built using only transistors (M1-M4) and resistors (R1-R2). The pulse generator modulates the carrier frequency by driving the tail transistor (M7) of the oscillator. If the data to be sent is logic '1', the oscillator turns on for 1.8 ns mandated by the delay on the clock line. The resulting pulse characteristics in time and frequency domains are shown in Figure 3.7b. The power spectral density (PSD) of the transmitted pulses complies with the indoor emission mask regulated by the Federal Communications Commission of the United States [230].



(a) (b)

Figure 3.7 – (a) Schematic of the IR-UWB transmitter and (b) the measured output.

The power consumption of the transmitter is data-dependent as the oscillator turns on only for transmitting logic '1'. The measured power for raw data and feature transmission modes were 33 µW and 6.6 µW, respectively. These values correspond to 10.3 pJ/b energy efficiency, which lies together with the state-of-the-art transmitters shown in Figure 1.10.

3.3 Measurement Results

Figure 3.8 displays the 16-channel prototype fabricated in TSMC 65nm 6X1Z1U LP CMOS process. The dimensions including the pad frame are 1.6 mm by 0.78 mm, and the core blocks occupy a total of 0.382 mm^2. Figure 3.9a presents the area breakdown of the chip. The area per channel including the AFE and the CHT processor is 0.021 mm^2.

The total power consumption of the chip is 401.45 µW including the power and command receiver (355 µW). The pie charts in Figure 3.9b and Figure 3.9c show the power breakdown of the system for raw data and CHT feature transmission modes, respectively. First, we observe that the power overhead of the CHT processor is 1.83 µW/channel. However, the data-dependent power consumption of the UWB transmitter decreases by 5 times from 2.06 µW/channel to 0.41 µW/channel when features are transmitted, which largely compensates for the power overhead of the feature extractor. In total, the power overhead of feature extraction is only 6.7%.

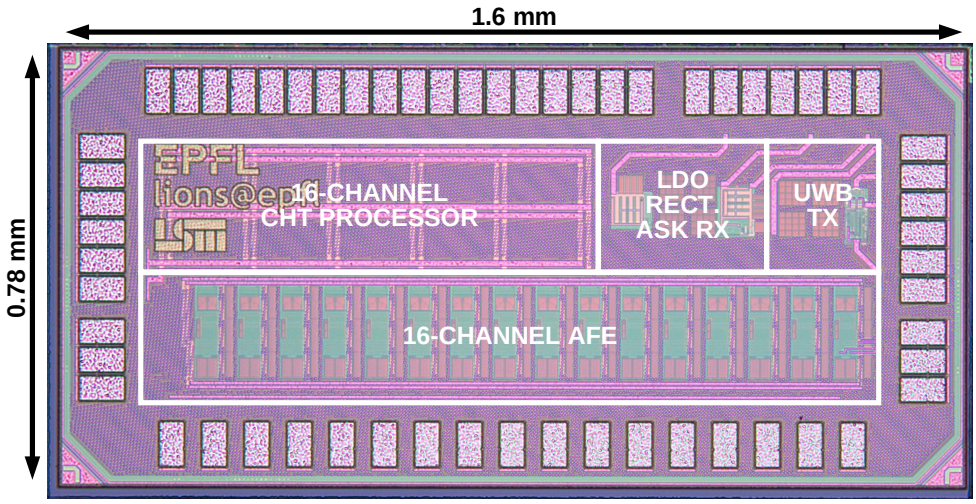


Figure 3.8 – Micrograph of the 16-channel wireless neural recording SoC fabricated in TSMC 65nm LP process.

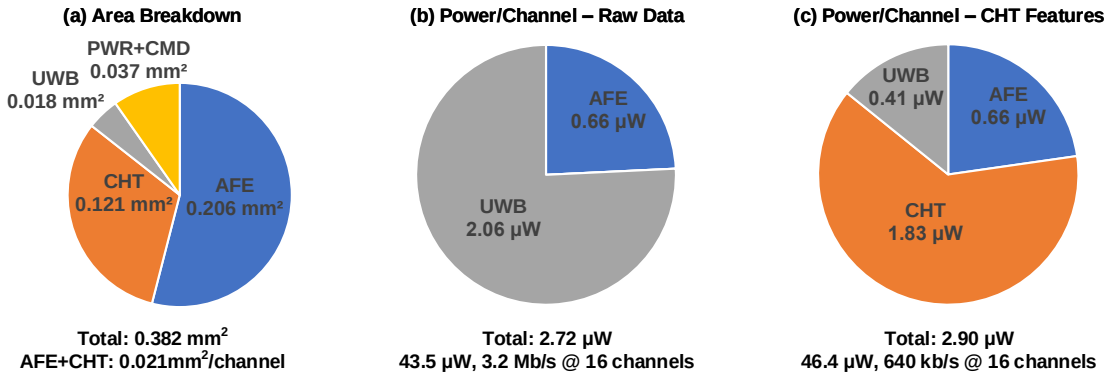


Figure 3.9 – Power and area breakdowns of the wireless neural recording SoC.

3.3.1 Offline Seizure Classification

Epileptic seizure detection has been a popular benchmark for neural signal classifiers due to the availability of large datasets containing hours of recordings from many patients. Moreover, the widespread use of spectral features for seizure detection makes it a reasonable task for demonstrating the efficacy of the proposed approach.

Dataset Description

Classification experiments were performed offline on two datasets. The first one is the CHB-MIT dataset [231]. It contains 24 EEG recording sessions collected from 22 pediatric subjects with intractable seizures at the Children's Hospital Boston. The sessions are organized in 1-hour segments, 136 of which contain one or more annotated seizure events. The total

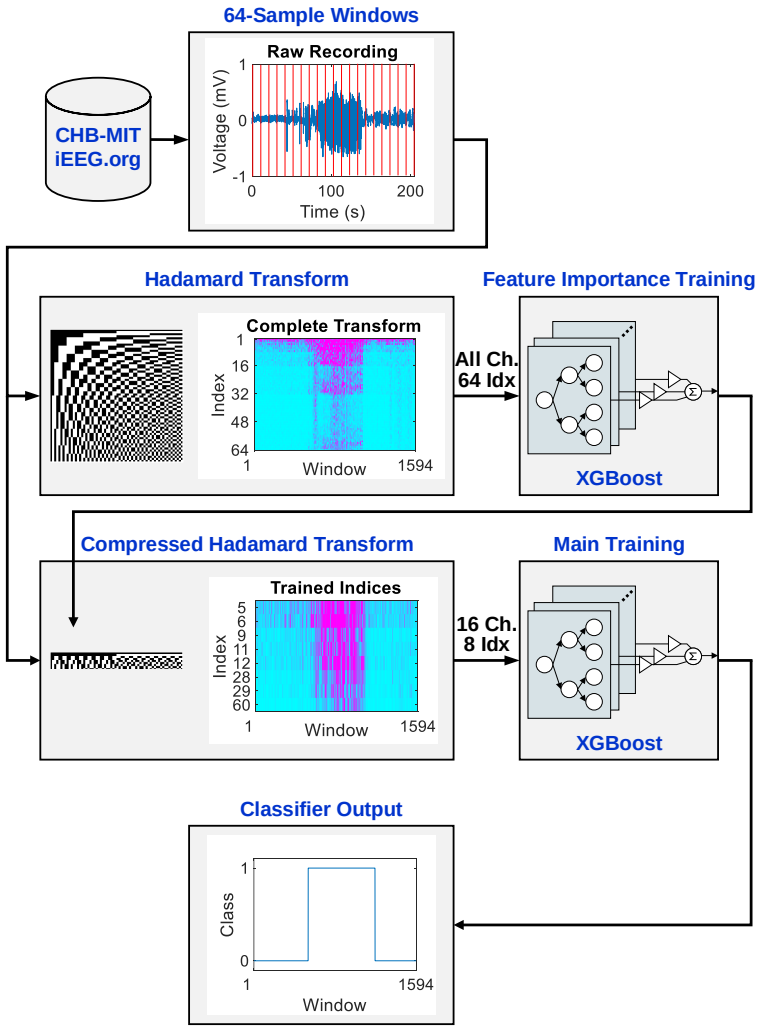


Figure 3.10 – Processing pipeline of the classification experiments.

number of seizures in all sessions is 180. The number of channels in the recordings varies between 23 and 26. The signals were sampled with 16-bit resolution at 256 S/s sampling rate. The second dataset is the iEEG.org collection [232], maintained by the University of Pennsylvania and Mayo Clinic. We selected 16 studies, each of which is a continuous multi-day recording. The selected studies contain a total of 103 seizures. The number of channels in the recordings varies between 16 and 116. The signals were sampled with 16-bit resolution at 500 S/s sampling rate.

Classifier Algorithm

The classifier was implemented in Python language using the XGBoost package [233]. We chose XGBoost based on its superior performance over other algorithms in ML competitions [233] and its training speed even with large datasets. XGBoost stands for Extreme Gradient

Boosting, as it implements a gradient-boosted decision tree algorithm. A decision tree is a classifier model that predicts the outcome of an input feature vector by successively comparing the feature elements to certain thresholds inferred from the training data. Gradient boosting forms an ensemble of decision trees incrementally by adding a new model that predicts the residuals of existing models. The final decision is the aggregate of the decisions of the trees in the ensemble.

The classifier model for the seizure detection experiments here is an ensemble of 8 decision trees with a maximum depth of 4. A new model was trained for each study to avoid underfitting due to the large patient-to-patient variations. The patient data was partitioned into sequences where each sequence was at least one hour long and contains at least one seizure. The resulting blocks were used as training and test sets for leave-one-out cross-validation.

The processing pipeline is described in Figure 3.10. The raw time-series signals were first requantized to 10 bits, then reorganized into 64-sample non overlapping windows to be compatible with the designed CHT processor. In order to standardize the different number of channels, an initial training was performed to select the most descriptive 16 channels per patient and 8 Hadamard indices per channel. The main training was performed using only those channels and indices, which simulates the chip configuration.

Classification Results

Figure 3.11 presents the results obtained on each study in the two datasets containing a total of 265 seizures from 40 patients. For most patients, all seizures could be detected with no false alarms. The average sensitivity for the CHB-MIT dataset was 92%, and reaches 97.8% if the two outliers (chb06 and chb16) are excluded. The false alarm rate (FAR) is 0.117/hour which corresponds to around 3 FARs per day. The average sensitivity for the iEEG.org dataset is 90.5% and the FAR is 0.171/hour, corresponding to around 4 FARs per day. The main reason for the performance difference between the two datasets is that the iEEG.org recordings are intracranial, as the name suggests. Since neural signals are picked up more strongly by implanted electrodes compared to EEG, strong nonseizure activity can mislead the classifier depending on the implanted location.

The most important observation here is that the descriptive indices vary significantly between patients and span the entire spectrum. Figure 3.12 visualizes the obtained results after the feature importance analysis. For the CHB-MIT dataset, the index ranges 15-20 and 40-45 have significantly higher importance. If we treat these indices as frequencies with 256 S/s sampling rate, they correspond to 30-40 Hz (low-γ) and 80-90 Hz (γ) bands, respectively. For the iEEG.org dataset, the important index ranges are 5-15, 28-30 and 57-64, which correspond to 20-60 Hz (low-γ), 110-120 Hz (high-γ), and 220-250 Hz (ripple) bands, respectively, with 500 S/s sampling rate. These findings are in accordance with the epilepsy literature suggesting the utility of HFOs for seizure detection [234]. It should be noted that selecting the exact indices manually would be very difficult, if not impossible, if we had not used joint training.

CHB-MIT Dataset					iEEG.org Dataset				
Study ID	**# Seiz.**	**# Det.**	**Sensitivity**	**FAR**	**Study ID**	**# Seiz.**	**# Det.**	**Sensitivity**	**FAR**
chb01	7	7	100%	0.143/h	**Study 004-2**	3	3	100%	0.000/h
chb02	3	3	100%	0.000/h	**Study 006**	4	2	50%	0.125/h
chb03	7	7	100%	0.140/h	**Study 011**	3	3	100%	0.000/h
chb04	4	4	100%	0.667/h	**Study 012-1**	6	4	67%	0.250/h
chb05	5	5	100%	0.000/h	**Study 014**	8	7	88%	0.000/h
chb07	3	3	100%	0.000/h	**Study 016**	7	7	100%	0.714/h
chb08	5	5	100%	0.000/h	**Study 017**	9	9	100%	0.000/h
chb09	4	4	100%	0.333/h	**Study 020**	8	7	88%	0.438/h
chb10	7	7	100%	0.000/h	**Study 021**	9	8	89%	0.722/h
chb11	3	3	100%	0.000/h	**Study 022**	7	7	100%	0.000/h
chb12	27	26	96%	0.000/h	**Study 023**	4	4	100%	0.250/h
chb13	10	9	90%	0.286/h	**Study 027**	6	6	100%	0.000/h
chb14	8	8	100%	0.000/h	**Study 028**	9	6	67%	0.111/h
chb15	20	20	100%	0.000/h	**Study 030**	8	8	100%	0.063/h
chb17	3	3	100%	0.667/h	**Study 031**	4	4	100%	0.000/h
chb18	5	5	100%	0.000/h	**Study 037**	8	8	100%	0.063/h
chb19	3	3	100%	0.000/h	**Total**	**103**	**93**	**90.5%**	**0.171/h**
chb20	8	8	100%	0.000/h					
chb21	4	4	100%	0.000/h					
chb22	3	3	100%	0.000/h					
chb23	7	5	71%	0.333/h					
chb24	16	15	94%	0.000/h					
chb06	10	2	20%	0.000/h					
chb16	8	3	38%	0.400/h					
Total	**162**	**157**	**97.8%**	**0.117/h**					

Figure 3.11 – Summary of the classification results on the CHB-MIT and iEEG.org datasets.

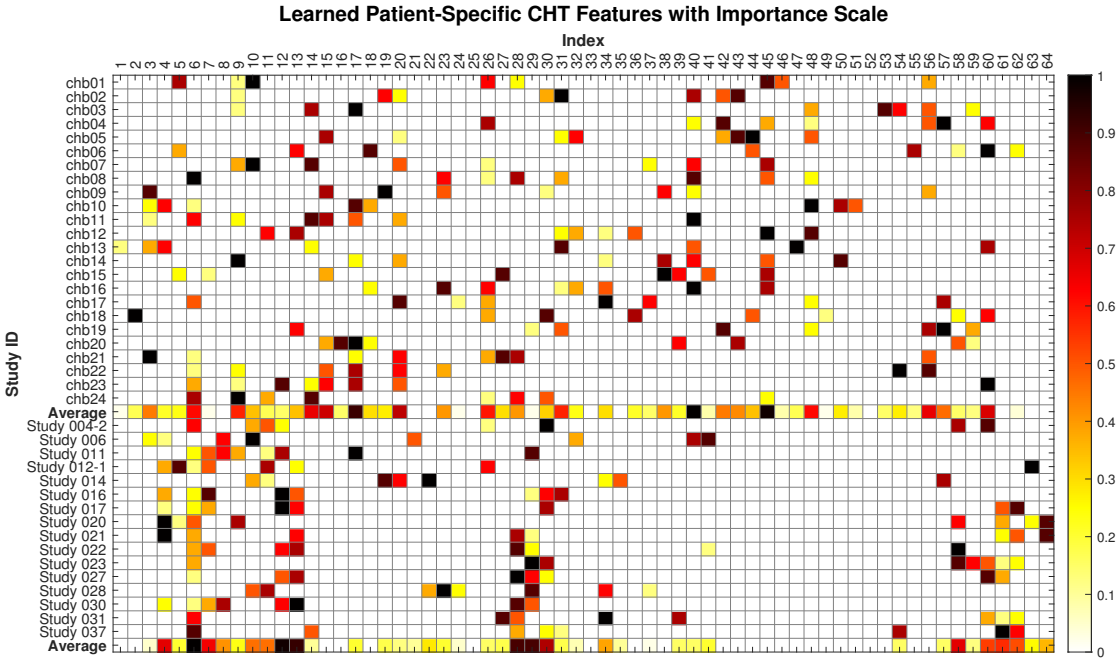


Figure 3.12 – Learned CHT indices for each study in the CHB-MIT and iEEG.org datasets. The color bar indicates the importance scale of the 8 selected indices out of 64.

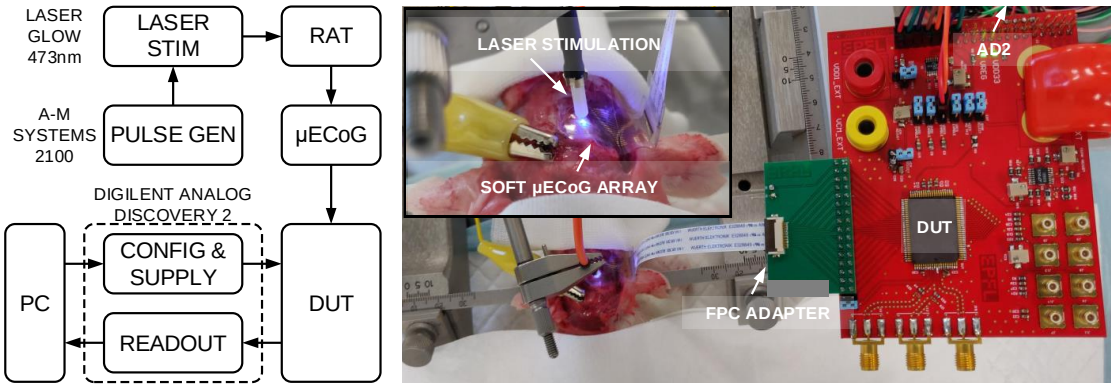


Figure 3.13 – Test setup for the *in vivo* validation of the reconstruction mode.

3.3.2 In Vivo Waveform Reconstruction

As mentioned before, an important benefit of using the CHT features is that the waveform can be reconstructed at the receiver side by taking the inverse transform, which is a linear operation that can be performed in real time. Figure 3.12 illustrates the reconstruction mode. If we configure and transmit the lowest indices, the inverse transform will give us the low-pass filtered version of the raw recording with more than 20 dB post-reconstruction SNR, as in previous LBCS demonstrations by Aprile [210]. We can also emulate a band-pass filter by selecting higher index groups.

Experiment Setup

We validated the reconstruction mode *in vivo* on an anesthetized transgenic rat. All animal experiments were approved by the Veterinary Office of the canton of Geneva in Switzerland and were in compliance with all relevant ethical regulations under animal license number GE 174_17. Figure 3.13 shows the experiment setup.

The preparation of the animal and the surgery was performed by Dr. Alix Trouillet of EPFL Laboratory for Soft Bioelectronic Interfaces. A W-Tg(Thy1-COP4/YFP) (NBRP 0685, Kyoto, Japan) transgenic female adult rat (~220 g body weight) was anesthetized with a mix of Ketamine (50-90 mg/kg) and Xylazine (5-10 mg/kg) diluted in NaCl before being head-fixed in a stereotaxic frame (David Kopf Instruments). A large craniotomy was performed, then a custom 16-channel soft µECoG array was placed epidurally over the exposed cortex.

The custom soft µECoG array was provided by Dr. Florian Fallegger of EPFL Laboratory for Soft Bioelectronic Interfaces, and it was microfabricated using thin-film and silicon processing technology, inspired from the e-dura process [225, 226]. A ground wire was fixed to the skull using a metallic screw. The array was connected to the test board using a flex cable and a custom-made FPC-to-header converter board. The test board hosting the prototype chip was interfaced to a PC using Digilent Analog Discovery 2 multifunction instrument and a Python

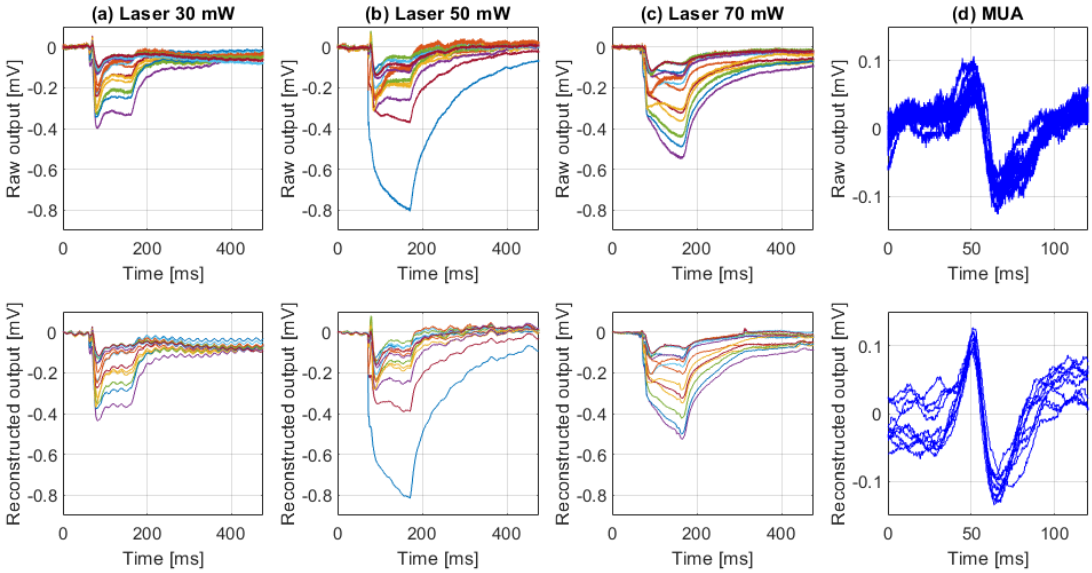


Figure 3.14 – Raw and reconstructed waveforms obtained during the *in vivo* experiment.

script for configuration and real time visualization of the recorded data.

The optical stimulation was delivered using a diode-pumped solid-state blue laser (473 nm, 100mW max power, Laserglow Technologies) coupled via a FC/PC terminal connected to a 200 µm core optical fiber (ThorLabs). Using a micromanipulator, the fiber was placed at the center of each µECoG array. The duration and frequency of the laser pulses were controlled by the A-M Systems Model 2100 pulse stimulator.

Reconstructed Waveforms

The optical pulses were delivered at 1 Hz with 100 ms pulse width with varying laser powers from 30 to 70 mW. For each power level, the chip was configured to record first raw data and then CHT features with the lowest 8 Hadamard indices.

Figures 3.14a-c plots the 16-channel waveforms recorded simultaneously. We can observe that the activity patterns differ by the applied power level. In all cases, the raw amplitude levels and pulse shapes are accurately represented by the reconstructed waveforms. Figure 3.14d shows the aligned spontaneous multiunit spiking activity (MUA) observed only on channel 8 in between trials, which verifies that the spike shapes are preserved after reconstruction.

Table 3.1 – SoC performance summary and comparison with the state of the art.

	JSSC'17[110]	JSSC'18[158]	JSSC'20[216]	ISSCC'18[217]	**This work**
Technology	130nm	130nm	40nm	180nm	**65nm**
V_{DD} [V]	1.2	1.2	0.58	1.5	**1**
Number of Channels	64	32	14	8	**16**
Blocks Included	AFE+FE+CLF+TX	AFE+FE+CLF	FE+CLF	AFE+FE+CLF	**AFE+FE+TX**
Total Area [mm^2]	3.86	7.59	2.56	5.83	**0.382**
Total Power [mW]	0.47*	0.714**	1.9	0.013**	**0.0464***
Feature Extractor	FIR+PLV	FIR+PLV/SE/CFC	FFT+SE	DWT+KDE	**CHT**
Dimensionality Reduction	-	Autoencoder	mRMR	ICA	**XGBoost**
Feature Dimension	n/r	125	16	48	**128**
Feature Computation	Fixed	Fixed	Fixed	Fixed	**Adaptive**
Waveform Reconstruction	No	No	No	No	**Yes**
Feature Extractor Area [mm^2]	1.285	0.475‡	0.618‡	0.817‡	**0.121**
Classification Rate [class/s]	n/r	4	11.1	0.074	**312.5**
Energy/Classification [μJ]	n/r	168.6	170.9	14.2	**0.149**
Classifier Algorithm	Threshold	SVM	SVM	SVM	**XGBoost DT**
Dataset	Custom	EU	CHB-MIT	CHB-MIT	**CHB-MIT**
Sensitivity	75%	100%	96.6%	97.8%	**97.8%**
False Alarm Rate/Specificity	0.5/h	0.81/h	0.28/h	99.7%	**0.12/h**

* Power management blocks excluded. ** Estimated from the reported values. ‡ Estimated from the micrograph.

3.3.3 Discussion

Table 3.1 summarizes the performance of the proposed neural recording front-end and compares it with previous works implementing on-chip classification [110, 158, 216, 217]. The results point to several improvements over the state of the art.

First, contrary to other methods, waveform reconstruction is possible with CHT. Secondly, the area occupied by the CHT processor is the smallest because Hadamard transform is more hardware friendly compared to other transforms like FFT or wavelet transform. Thirdly, the energy per each feature output is much less than on-chip classification. At 20 kHz sampling rate, the transmitted features enable 312.5 class/s at the receiver side, which corresponds to 149 nJ/class or 9.3 nJ/class/channel energy efficiency for the AFE, CHT processor and UWB transmitter combined. This means that sending multichannel data to off-chip to more advanced and flexible classifiers is more profitable than on-chip classification. Finally, while taking much less area and energy on the chip, the seizure detection performance with CHT features is similar to [216, 217] on the same dataset.

High-Speed Communications Part II

4 A Wideband Phase-Locked Loop Technique for Mitigating VCO Noise

Today's connected world increasingly relies on high-speed data communication. In just a few decades, data rates have increased by several orders of magnitude for both wireless and wireline transceivers [235, 236]. This rapid advancement can be attributed in part to the downscaling technology which has brought faster transistors. However, as we reach beyond 100 Gbps data rates, the dominant performance-limiting factor shifts from transistor speed to timing precision.

Timing uncertainty in electronics is quantified as jitter in time domain, and phase noise in frequency domain. The IEEE Standard for Jitter and Phase Noise [237] defines jitter as "the deviation of the reference instants of a sequence of events from their ideal values", and phase noise as "the random fluctuations in the phase of a periodic signal". The sources of these deviations can be random due to electronic noise of the devices, or deterministic due to interference from other circuits.

Figure 4.1 pictures the effects of timing uncertainty in three different applications. In wireline communications, bit error rate (BER) depends on jitter. Figure 4.1a illustrates a generic wireline link. The transmitter (TX) sends the bit stream to the receiver (RX) through a channel, which could be a cable or a printed circuit board (PCB) trace. The data is synchronized with the TX clock, which is then recovered at the RX side. Ideally, the transmitted symbols should construct an open eye when superposed within a unit interval. In other words, they should perfectly align with the recovered RX clock at the sampling instant. In reality, the jitter on the data causes the eye to close and increases the chances of incorrect sampling.

In wireless communications, phase noise directly affects the network capacity. As depicted in Figure 4.1b, a generic wireless TX upconverts the transmitted bit stream to the local oscillator (LO) frequency. To allow multiple users to communicate simultaneously, the total available frequency band is divided into subcarriers. Ideally, all channels should be contained within their bandwidths, but the spectral leakage due to the phase noise of the LO corrupts the adjacent channels. Reciprocal mixing at the RX side during downconversion is the manifestation of the same problem. To reduce the effects of spectral leakage and reciprocal mixing, a subcarrier

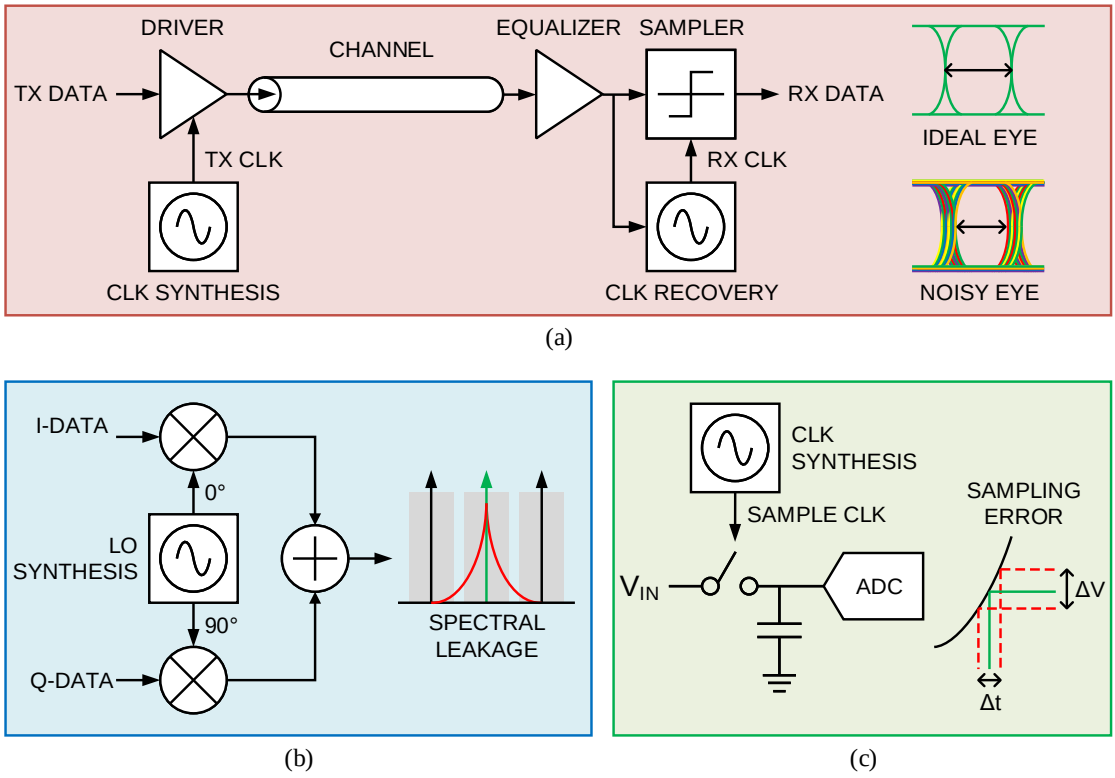


Figure 4.1 – Jitter-sensitive applications: (a) wireline communications, (b) wireless communications, and (c) high-speed ADCs.

spacing depending on the phase noise of the LOs needs to be allocated at the expense of reduced network efficiency.

In high-speed ADCs, the maximum achievable resolution is determined by jitter. Figure 4.1c shows a generic sampling operation where the input signal is captured by a sampling clock. The jitter on the sampling clock translates into a voltage error on the sampled signal. Depending on the speed and the resolution of the ADC, this error can exceed the quantization and other noise sources.

In almost every high-speed system, the fast symbol-rate clock is synthesized by a phase-locked loop (PLL) from a lower-frequency reference, such as a crystal oscillator (XTAL) or a similar low-noise generator. The output phase noise of a PLL is closely related to the type of VCO used, and the loop bandwidth. Section 4.1 briefly introduces the considerations for both aspects. A proposed solution for mitigating the effect of VCO phase noise is presented in Section 4.2, which forms the basis for the design work presented in Chapter 5.

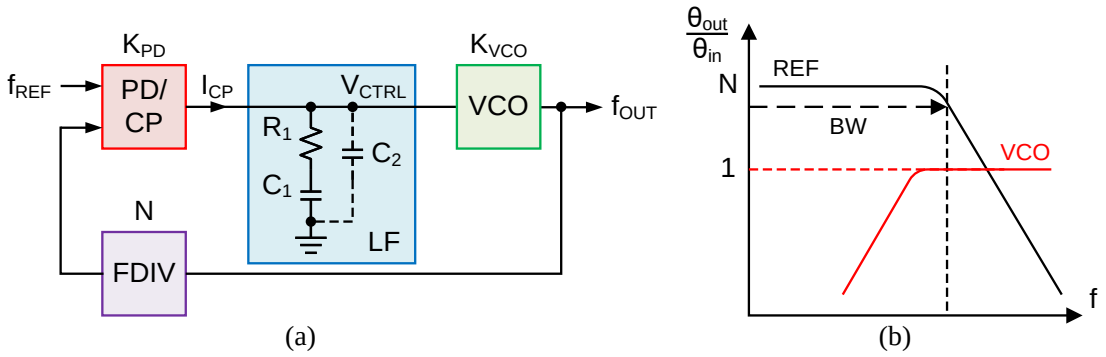


Figure 4.2 – (a) Charge-pump PLL, and (b) jitter transfer functions for reference and VCO.

4.1 Phase-Locked Loop Considerations

Figure 4.2a shows a generic integer-N charge-pump PLL, denoting the five main components: phase detector (PD), charge pump (CP), loop filter (LF), VCO, and feedback divider (FDIV). The VCO output is divided by the FDIV ratio (N) and compared to the reference input via the PD. Any error between the two phases results in a proportional CP current I_{CP} injected into the LF. This adjusts the control voltage V_{CTRL} such that the reference and the VCO remain in phase.

Figure 4.2b shows the jitter transfer functions from the reference and the VCO to the output. The PLL tracks[1] the reference phase and rejects that of the VCO within its loop bandwidth (BW). For higher frequencies, the VCO phase noise is dominant. Therefore, the ideal condition to achieve low phase noise for both the reference and the VCO. While ultra-clean reference sources are already available up to the GHz range [238], achieving an on-chip low-phase-noise VCO is a grand challenge due to energy and area limitations.

4.1.1 Voltage-Controlled Oscillators

The cost of lowering VCO phase noise in terms of power consumption is proportional to the fourth power of the target jitter as pointed out by Razavi [239]. The jitter targets for the state-of-the-art communication systems have reached below 100 fs and will soon reach below 10 fs for the next generation. Providing this level of timing precision will make the frequency synthesis the power bottleneck instead of the functional blocks.

The VCO area also becomes a limiting factor when multiple PLLs have to be integrated on the same die together with other large blocks, as in the case of multilane wireline transceivers [240] or transceiver arrays for 5G beamforming [241]. The two VCO types used in integrated CMOS PLLs are the ring VCO and the LC VCO. For the majority of high-speed applications the LC VCO is preferred over the ring due to its superior phase noise performance, at the expense

[1] The reference phase noise is amplified due to the frequency multiplication factor N. Therefore, using a higher reference frequency and lower N benefits the output phase noise at low frequencies.

of much higher area occupation [242–244]. This is because the ring VCO can be constructed using only active elements in a compact fashion, whereas the LC VCO involves inductors and varactors which require a larger silicon area and additional coupling considerations.

Unfortunately, the VCO challenges are fundamental and cannot be solved easily using circuit trade-offs. Therefore, we focus on increasing the PLL bandwidth such that the impact of VCO phase noise is mitigated and the ring VCOs become feasible for jitter-sensitive applications.

4.1.2 Loop Bandwidth

The s-domain second-order transfer function of a PLL from the reference to the output is given by

$$\frac{\theta_{OUT}}{\theta_{REF}} = N\frac{2s\zeta\omega_n + \omega_n^2}{s^2 + 2s\zeta\omega_n + \omega_n^2}\,. \tag{4.1}$$

Here, ω_n is the natural frequency and ζ is the damping factor, and they are related to the block parameters as

$$\omega_n = \sqrt{\frac{K_{VCO}K_{PD}I_{CP}}{NC_1}}, \qquad \zeta = \frac{R}{2}\sqrt{\frac{K_{VCO}K_{PD}I_{CP}C_1}{N}}\,, \tag{4.2}$$

where K_{VCO} and K_{PD} are the gains of the VCO and PD, R and C_1 are the LF resistor and capacitor, and N is the multiplication factor. Together, ω_n and ζ determine the -3 dB bandwidth (ω_{-3dB}) and the jitter peaking (J_P) of the low-pass transfer function as [245]

$$\omega_{-3dB,REF} = \omega_n\sqrt{1 + 2\zeta^2 + \sqrt{(1+2\zeta^2)^2 + 1}}, \qquad J_P \approx \frac{1}{4\zeta^2}\,. \tag{4.3}$$

Similarly, the high-pass transfer function from the VCO to the output and its -3 dB bandwidth can be computed as [245]

$$\frac{\theta_{OUT}}{\theta_{VCO}} = \frac{s^2}{s^2 + 2s\zeta\omega_n + \omega_n^2}\,, \tag{4.4}$$

$$\omega_{-3dB,VCO} = \omega_n\sqrt{2\zeta^2 - 1 + \sqrt{(2\zeta^2-1)^2 + 1}}\,. \tag{4.5}$$

According to (4.3), increasing ω_n and ζ increases the bandwidth and reduces peaking. However, this does not accurately capture the reality because of three reasons.

First, the second order PLL suffers from large V_{CTRL} ripples due to the series R, which cause reference spurs in the output spectrum [242]. In practice, a parallel capacitor C_2 is added to the LF for a smoother V_{CTRL}, resulting in a third-order loop. In this case, increasing ζ does not indefinitely increase bandwidth and reduce peaking as prescribed by (4.3), but starts to increase peaking after a certain value.

Secondly, the equations neglect the loop delay, which decreases the phase margin and thus increases jitter peaking [246]. Figure 4.3 visualizes the effect of increasing ζ with and without

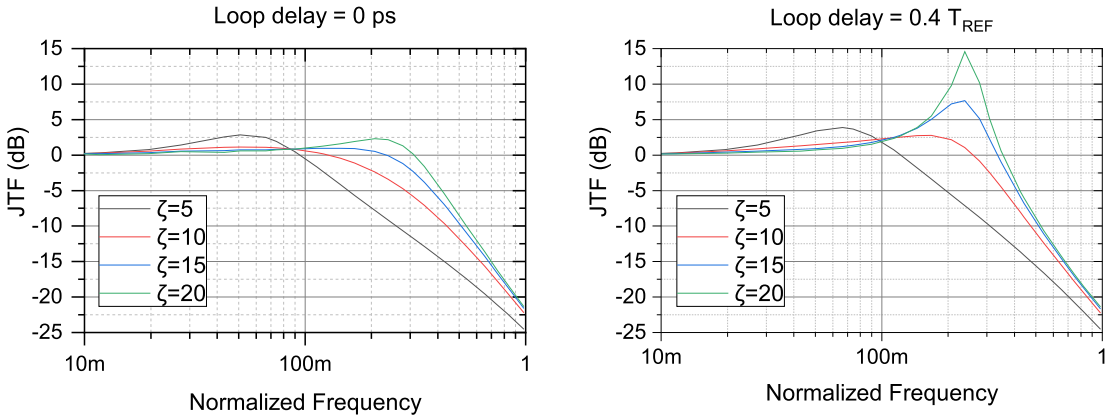


Figure 4.3 – Peaking in the third-order jitter transfer function with and without loop delay.

loop delay in a third-order PLL. The same increase in ζ creates much higher peaking in the presence of loop delay, which is inevitable in circuit implementations.

Finally, the s-domain analysis assumes a continuous-time operation, whereas in reality the loop has a discrete-time nature since the PD acts only in response to the edges of its inputs. As a rule-of-thumb, the validity of the continuous-time approximation is limited to $\omega_{-3dB} <\sim 0.1\omega_{REF}$, known as Gardner's limit [247]. For higher bandwidths, the granularity effects take over and reduce stability. In this case the s-domain equations become too optimistic to capture the actual trade-offs. The third-order z-domain transfer functions provided by Gardner [247] and Hanumolu *et al.* [248] are more accurate, but these tedious analytical expressions are less interpretable. Therefore, we will focus more on simulated responses based on an ideal model.

4.2 Multiphase Feedback Phase-Locked Loop

There has been a few previous studies that have tried to achieve bandwidths beyond Gardner's limit. Kong and Razavi [242] proposed a type-I synthesizer with a master-slave sampling loop filter and harmonic traps to counteract the reference spurs caused by large swings at V_{CTRL}. This technique achieved $0.25 f_{REF}$ bandwidth with -65 dBc reference spur level. Ting and Lee [249] proposed a type-II synthesizer with a subsampling feedforward path which created an additional zero in the VCO jitter transfer function (JTF). They reported $0.2 f_{REF}$ bandwidth, yet they did not report the spur level. Although the proposed methods could go beyond Gardner's limit, they do not address the fundamental limitation, which is the loop update rate.

In this work, we propose a new architecture called multiphase feedback (MPF) PLL to overcome this limitation. Figure 4.4 shows a 4-phase example of the concept. The main modification over the conventional PLL architecture is the XOR matrix phase detector (MPD), which compares the reference input with multiple feedback phases.

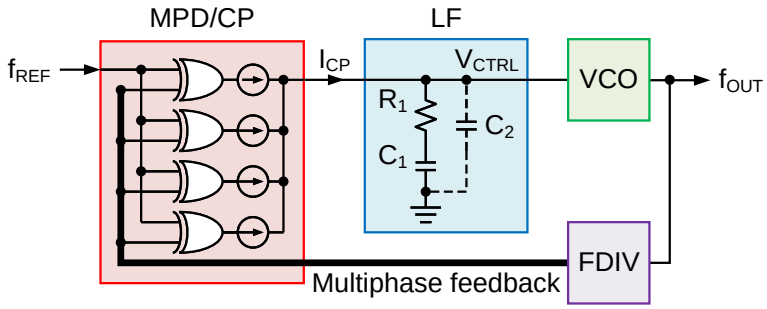


Figure 4.4 – A 4-phase example of the MPF PLL concept.

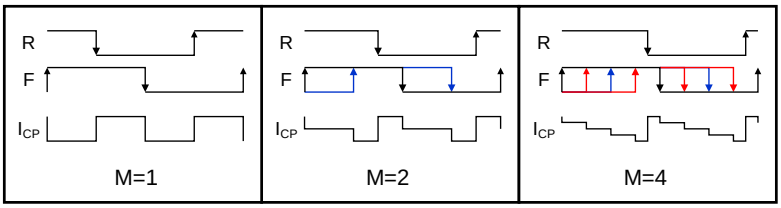
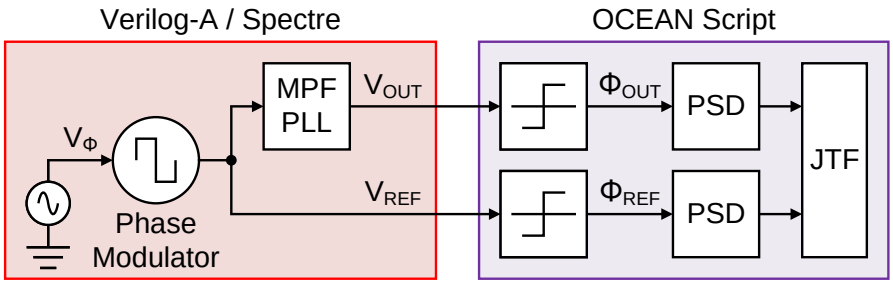


Figure 4.5 – I_{CP} for different values of active PDs.

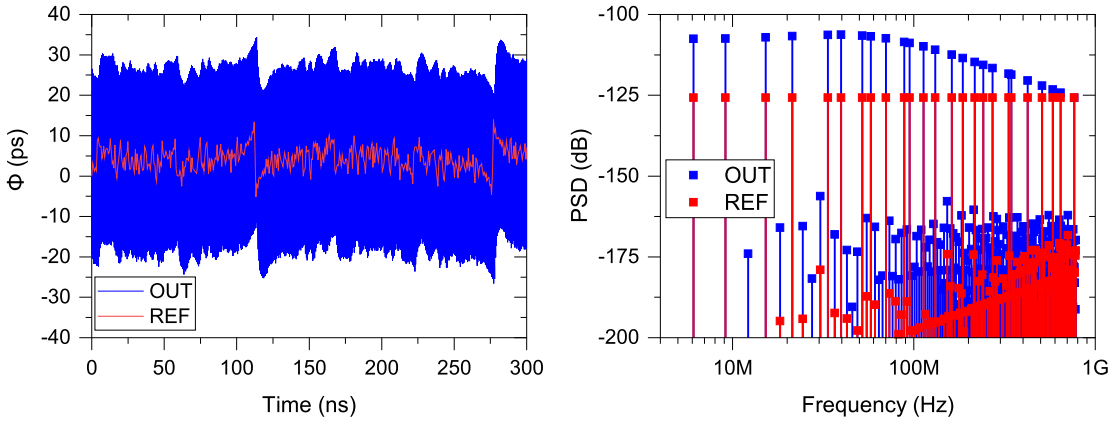
4.2.1 XOR Matrix Phase Detector

XOR PDs inherently double the loop update rate as they compare both rising and falling edges of the input and the feedback clocks, as opposed to conventional phase-frequency detectors (PFDs) which compare only a single edge. They have been used extensively for phase interpolators in clock and data recovery circuits which adjust the locked output phase by small increments. Toifl *et al.* [250] used a common-mode logic (CML)-style XOR PD where the coarse adjustment selected two of the multiple feedback phases, and fine adjustment weighed the contribution of each phase to achieve the desired phase adjustment. Tajalli *et al.* [240] used a CMOS logic XOR PD where each branch of the XOR gate was composed as an array of 15 units such that they could be weighed individually to achieve the target phase.

Unlike the conventional XOR PDs used for phase interpolation [240,250], the MPF PLL employs an XOR MPD where each feedback phase contributes equally to I_{CP}. Figure 4.5 illustrates I_{CP} for different numbers of active PDs (M). M=1 corresponds to the XOR PD case, where the loop is updated at every rising and falling edge of the reference phase and a single feedback phase. M=2 uses two feedback phases with 90° shift. As a result, I_{CP} resembles a staircase. M=4 uses all four feedback phases with 45° shift, which creates more time instances for updating the loop. The increased loop update rate allows a stable operation with a bandwidth close to f_{REF} without excessive peaking.



(a)



(b)

Figure 4.6 – (a) Behavioral JTF simulation setup with multitone sinusoidal phase input and (b) the resulting time-domain and frequency-domain plots.

4.2.2 Behavioral Simulation

Simulating the bandwidth of the MPF PLL with the conventional linearized phase-domain model is difficult since its operation cannot be approximated as a continuous-time system. Therefore, we modeled the MPF PLL in Verilog-A language and simulated it in time domain using Spectre circuit simulator. The simulation setup is depicted in Figure 4.6a. The reference input clock is phase modulated with a multitone sinusoidal signal V_ϕ which is expressed as

$$V_\phi(t) = \sum_{\substack{k=1 \\ k\in\mathbb{P}}}^{N_{PSD}} \sin\left(2\pi k \frac{f_{REF}}{N_{PSD}} t\right), \tag{4.6}$$

where N_{PSD} is the number of samples for PSD computation. The frequencies are prime multiples of f_{REF}/N_{PSD} to guarantee coherent sampling.

The voltage-domain results of the simulation V_{REF} and V_{OUT} are post-processed using an OCEAN script to extract the JTF. The script computes the reference and output phase deviations ϕ_{REF} and ϕ_{OUT} based on the zero crossings of the voltage waveforms, then computes their PSDs. The PSDs display peaks at the exact frequencies that construct V_ϕ as shown

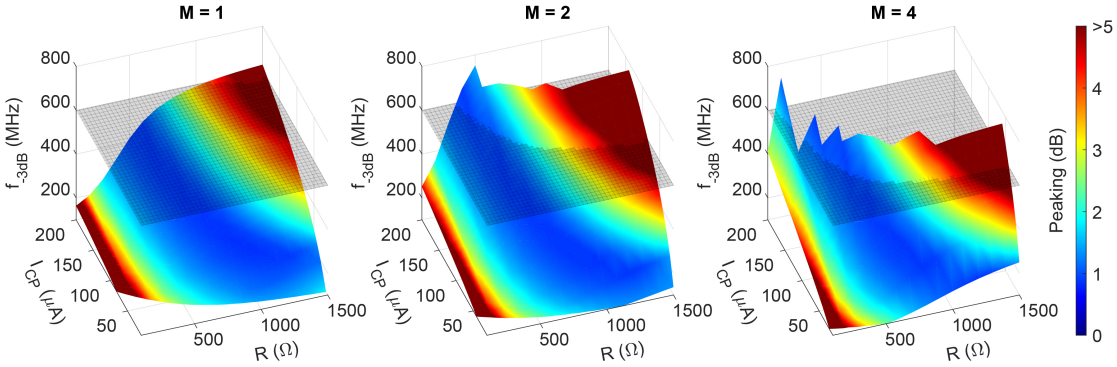


Figure 4.7 – 3D plots showing the effects of R and I_{CP} on bandwidth and peaking for different M values.

in Figure 4.6b. The difference between the reference and the output PSDs gives the JTF.

The described behavioral simulation technique was used to explore the design space for the MPF PLL. For these simulations, we chose f_{REF} as 781.25 MHz and N as 8 to produce 6.25 GHz output. Figure 4.7 displays the 3D plots showing the effects of R nd I_{CP} on bandwidth and peaking for different values of M. For M=1, it is not possible to achieve above 600 MHz bandwidth without excessive jitter peaking. However, when we start to add more phases into the loop we see an improvement. For M=2 and M=4, bandwidths above 700 MHz are possible with less than 2 dB peaking.

4.3 Summary

The growing data rates in wireline and wireless communication systems demand unprecedented levels of timing precision. However, lowering the phase noise of CMOS PLLs requires a disproportional increase in energy consumption and silicon footprint. As a result, timing uncertainty is foreseen to be the main limiting factor for next-generation transceivers.

The high-frequency local clock in a transceiver is generated from a low-frequency reference clock by a multiplying PLL. The loop tracks the reference phase noise and rejects that of the VCO within its bandwidth. Consequently, increasing the loop bandwidth could permit the use of noisy but compact ring VCOs for jitter-sensitive applications.

The discrete high-order nature of the PLL limits the stable bandwidth to a fraction of the reference frequency. Extending the bandwidth beyond this limit results in excessive jitter peaking and eventually the loss of stability. We have addressed this challenge with a MPF PLL which increases the loop update rate by comparing the reference phase with multiple feedback phases. The behavioral simulation results show that the loop bandwidth can be increased close to the reference frequency without compromising stability. Based on this premise, Chapter 5 presents a circuit implementation of the MPF PLL.

5 A 6.25 GHz Integer-N Ring PLL with $0.92f_{REF}$ Bandwidth

In this chapter we present a silicon implementation of the MPF PLL introduced in Chapter 4 [3]. Section 5.1 describes the design details, Section 5.2 provides the simulation results, and Section 5.3 discusses the measured characteristics of the fabricated prototype.

5.1 Architecture Details

The top level block diagram of the PLL is shown in Figure 5.1. The core blocks MPD, LF, VCO, and FDIV, are combined with four auxiliary blocks: bandgap reference (BGR), lock assist (LA), input reference buffer and output source-series terminated (SST) buffer. The control parameters I_{CP}, R_1, C_1 and M are configured through a serial interface (CFG). For this prototype, the multiplication ratio N is set to be 8, and f_{REF} is 781.25 MHz to synthesize 6.25 GHz. This technique, however, is applicable to any integer N. The following subsections explain each block in detail.

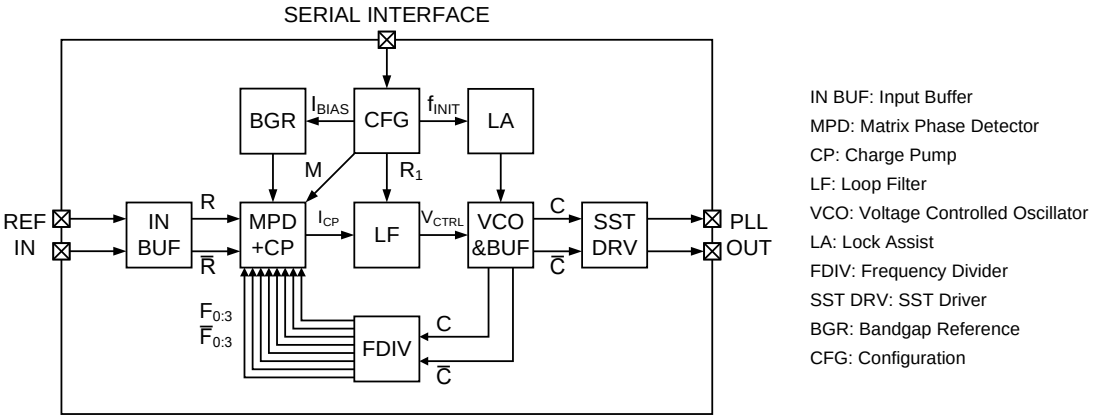


Figure 5.1 – Top level block diagram of the MPF-PLL.

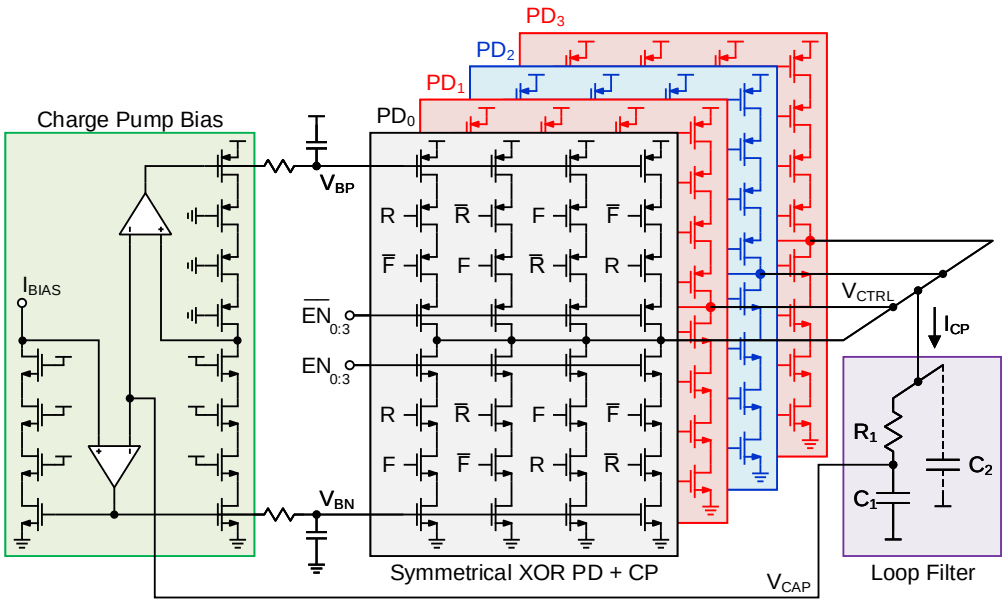


Figure 5.2 – Schematic of the MPD+CP and the loop filter.

5.1.1 Matrix Phase Detector and Loop Filter

The most important element of the proposed wideband PLL is the matrix phase detector and charge pump (MPD+CP), whose schematic is shown in Figure 5.2 together with the LF. Four integrated XOR-based phase detector and charge pump (PD+CP) circuits ($PD_{0:3}$) compare the reference input (R, $\overline{R}$) with all 4 phases of the feedback divided clock ($F_{0:3}$, $\overline{F}_{0:3}$) and inject I_{CP} into the LF. The LF consists of C_1=10pF, and programmable R_1 between 400Ω and 800Ω. C_2 is implicitly created by the parasitic capacitances at V_{CTRL} node. C_1 is implemented as a combination of metal-oxide-semiconductor (MOS) and MOM capacitors to reduce the footprint.

Each PD is integrated with its CP current source to prevent additional loop delay. The drains of the current source transistors need to switch at 6.25 GHz nominal output frequency when all PDs are enabled, thus they need to be small with short channel lengths. The switching speed comes at the expense of increased susceptibility to drain-induced current mismatch depending on the value of V_{CTRL}. To compensate for this adverse effect, the charge pump bias voltages V_{BP} and V_{BN} are driven by a replica bias circuit, which clamps the output of the replica branches to the stable capacitor voltage V_{CAP} through feedback amplifiers.

The XOR operation is implemented as CMOS logic branches. The operation of a CMOS XOR gate branch is not symmetrical as illustrated in Figure 5.3a. If the branch is turned on by input F, only one internal node has to be charged to turn on the current source transistor. If the branch is turned on by R, two internal nodes have to be charged to turn on the current source transistor. In a regular PD this would only cause a small delay mismatch; however, in a PD+CP it causes a large mismatch in I_{CP} transients.

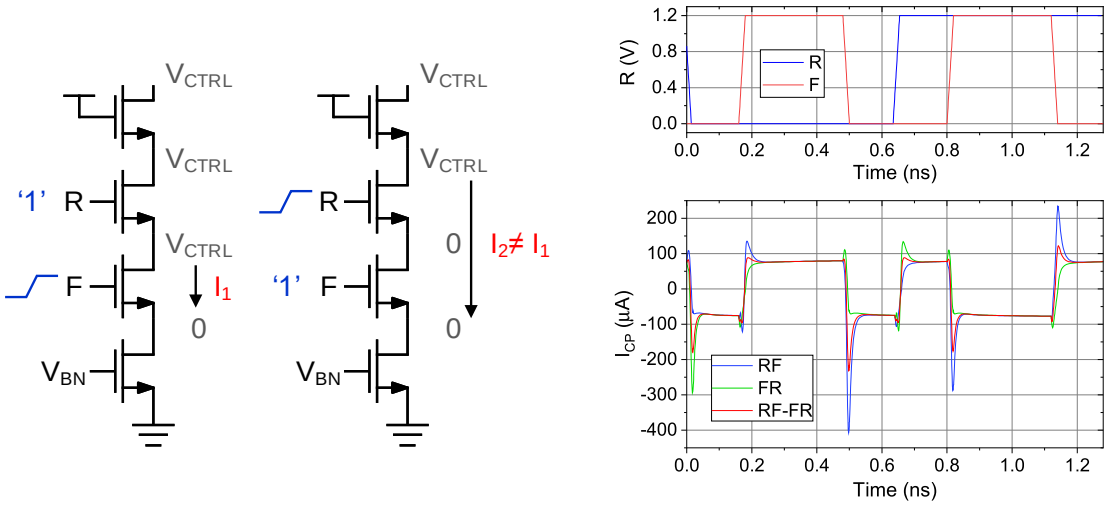


Figure 5.3 – (a) Transient current mismatch problem and (b) its mitigation using symmetrical branches.

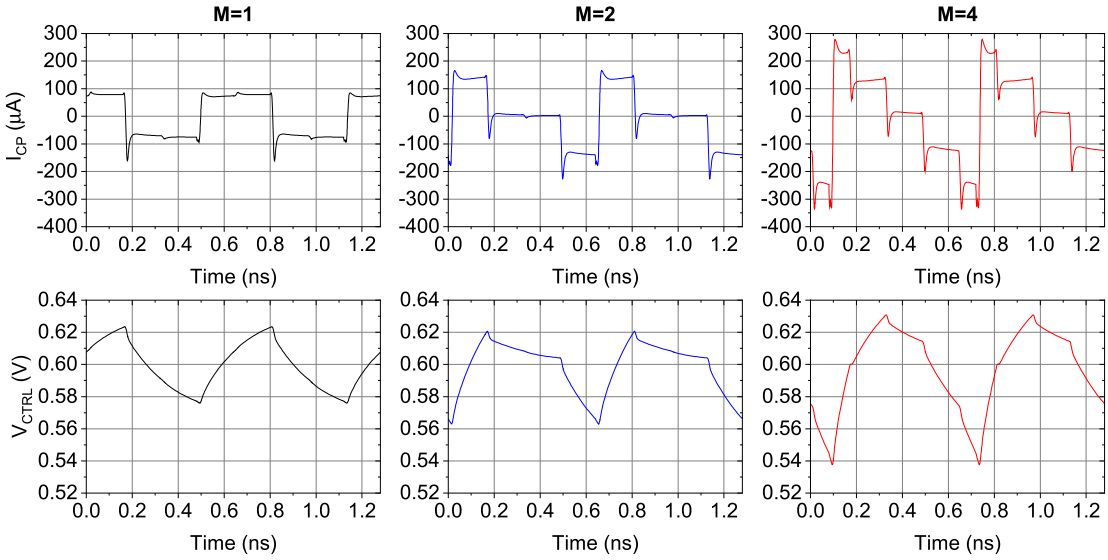


Figure 5.4 – Simulated I_{CP} and V_{CTRL} for different M values.

The solution employed here was using two identical XOR PD+CPs, where the input order of the second PD+CP is flipped compared to the first one ({R,F}→{F,R}). The transient I_{CP} differences are largely mitigated thanks to this symmetrical operation as shown in Figure 5.3b.

The PD+CPs can be enabled or disabled individually ($EN_{0:3}$,$\overline{EN}_{0:3}$) to assess the effect of increasing number of active PD+CPs (M) on the bandwidth and phase noise. Figure 5.4 depicts I_{CP} at locked state for different M values.

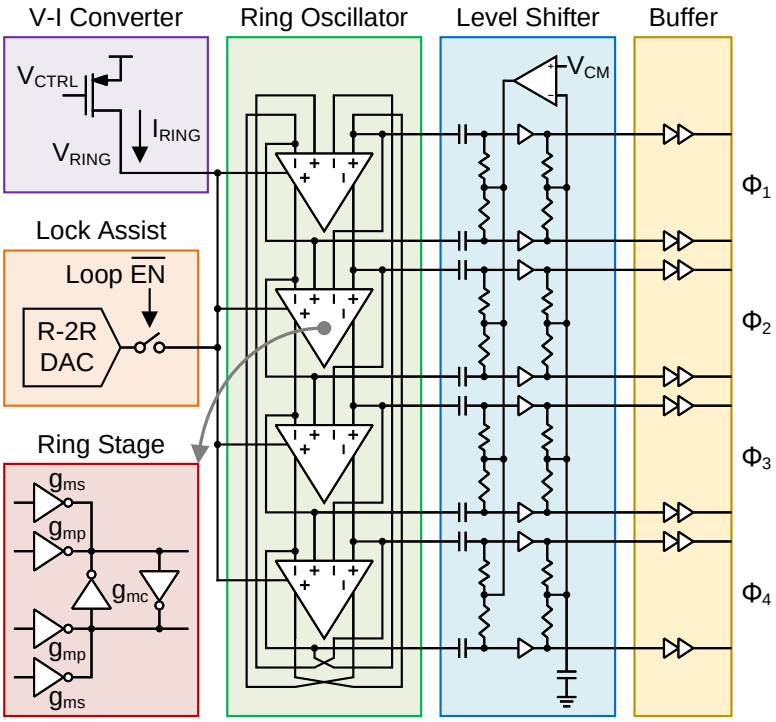


Figure 5.5 – Schematic of the multiphase multiloop VCO.

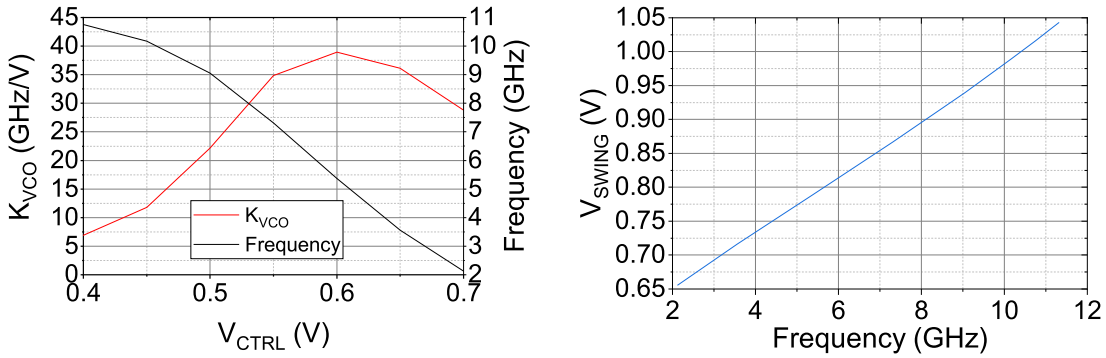
5.1.2 Voltage-Controlled Ring Oscillator

Figure 5.5 shows the schematic of the VCO. The core of the VCO is a 4-stage pseudo-differential ring oscillator with 8-phase output. The ring has a multiloop architecture [251] where each inverter-based stage has two inputs. The primary and the secondary inputs are driven by the two preceding stages. The cross-coupled weak inverters at the differential outputs ensure 180° phase difference. Although they increase the delay of the circuit by resisting the change imposed by the primary and secondary inputs, they also increase the slew rate when the positive feedback takes over which leads to better phase noise [252].

The analysis of the multiloop architecture has been provided by Sun and Kwasniewski [251] and Liu *et al.* [252]. If we denote the transconductance of the primary, secondary, and cross-coupled inverters as g_{mp}, g_{ms}, and g_{mc}, the oscillation frequency for this 4-stage multiloop ring can be written as

$$f_{osc} = \frac{1}{2\pi RC} + \frac{g_{ms} - g_{mc}}{2\pi C}, \tag{5.1}$$

where R and C correspond to the equivalent output resistance and capacitance of a single stage, respectively. Therefore, the nominal f_{osc} at the nominal V_{CTRL} can be tuned by adjusting the strengths of the secondary and the cross-coupled inverters. Based on simulations, we set them as $g_{mp} = g_{ms} = 3g_{mc}$ to achieve 6.25 GHz with $V_{CTRL} \simeq V_{DD}/2$.



(a) (b)

Figure 5.6 – Simulated (a) gain and (b) voltage swing of the VCO.

The frequency tuning characteristic of the ring VCO is plotted in Figure 5.6a. The frequency gain (K_{VCO}) at 6.25 GHz nominal operating frequency is 36 GHz/V. The swing of the current-starved ring oscillator output is not full-scale and it is frequency-dependent as seen in Figure 5.6b. Therefore, it has to be restored to full-scale by a level shifter.

Level Shifter and Buffer

The ring oscillator output is AC-coupled to the level shifter via coupling capacitors. The input bias voltage of the level shifter is set via a CMFB circuit, such that the output common-mode voltage is kept at a defined level, nominally at $V_{DD}/2$. The output common-mode voltage is sensed via two resistors, then the error between the desired V_{CM} is amplified. The amplifier output sets the input common mode voltage via two resistors. Anticipating fairly symmetrical operation, all phases share the same amplifier to reduce the footprint. The output of the level shifter is buffered by a cascade of strong inverters to drive the subsequent stages and interconnects.

Lock Assist Mechanism

Acquiring the locked state with an XOR PD requires the initial frequency of the VCO to be close to the intended frequency. To ensure this, a manual lock assist mechanism is employed. First, all PD+CPs are disabled and a 10-bit R-2R resistor ladder DAC is connected to enforce V_{CTRL}. The frequency of the VCO can then be adjusted by configuring the DAC code through the configuration interface. Then, the DAC is disconnected and the PD+CPs are enabled.

5.1.3 Feedback Divider

Figure 5.7 shows the feedback divider constructed by a cascade of three divide-by-2 flip-flops. The differential flip-flops are made of master-slave latches with tri-state inverters. The output

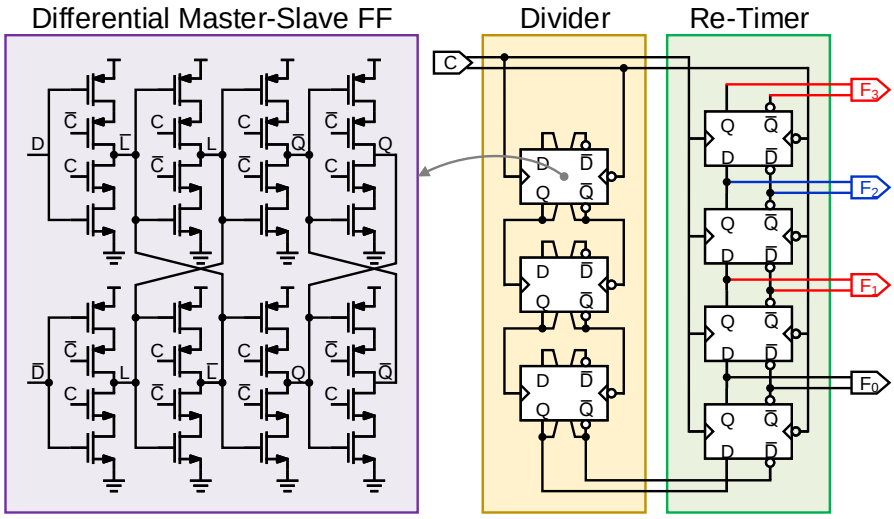


Figure 5.7 – Schematic of the feedback divider.

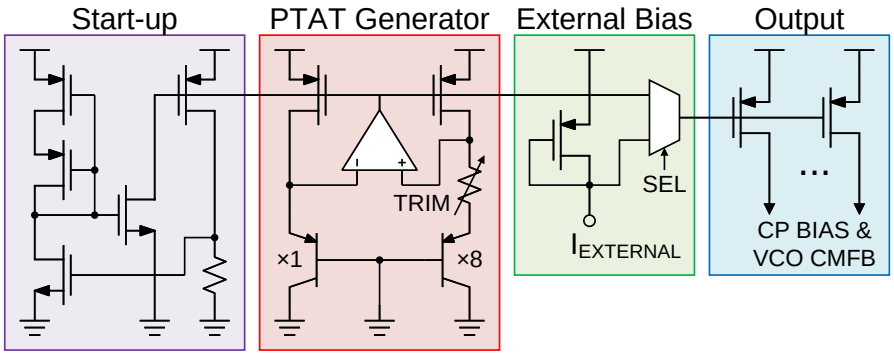
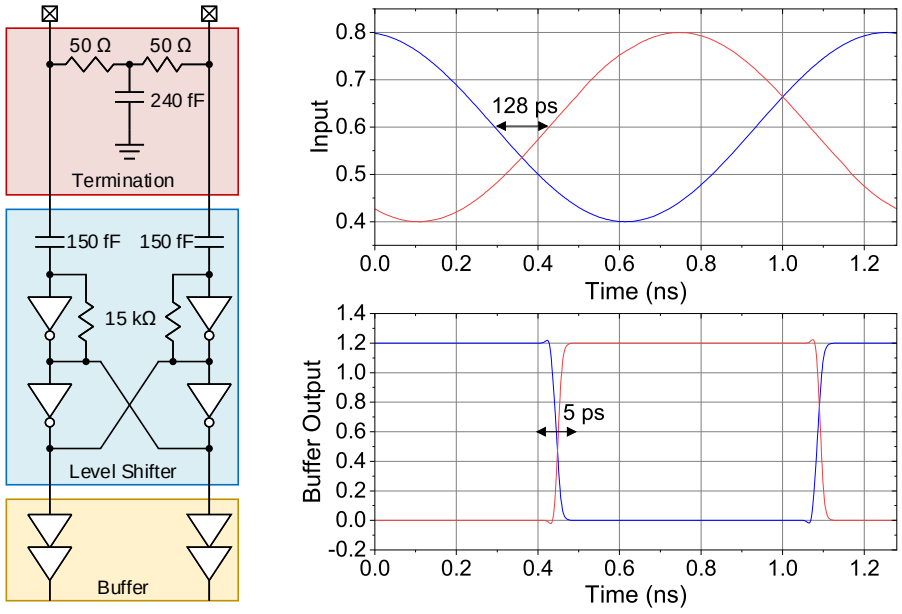


Figure 5.8 – Schematic of the bias current generator.

of the divider is re-timed by the full-speed VCO output to avoid noise and delay accumulation. The re-timer also produces 4 phases required by the MPD by sampling the divided clock in a shift register fashion.

5.1.4 Bandgap Reference

The bias currents required by the CP and the VCO CMFB amplifier are supplied by the on-chip proportional to absolute temperature (PTAT) current generator shown in Figure 5.8. It employs a basic BJT-based architecture [253]. A small startup circuit ensures that the generator is not stuck at off state after power on. The generated current is trimmable between 7.5 and 20 µA with two control bits. Other values can also be supplied with the external bias option.



(a) (b)

Figure 5.9 – (a) Schematic of the input buffer and (b) its operation in the presence of phase mismatch.

5.1.5 Input Buffer

The input buffer receives the reference input clock produced by an external signal generator. Since the generator and the coaxial cables are 50 Ω-matched, the first stage is a split differential 50 Ω termination. The output swing and the DC level of the input depend on the generator specifications. To comply with multiple instruments, the terminated signal is AC-coupled then boosted to rail-to-rail swing by a level shifter. The cross-coupled inverters at the output of the level shifter equalize the zero crossings of the differential phases, which may be skewed due to cable and PCB trace mismatches. Figure 5.9b simulates such case where the 128 ps phase mismatch reduces to 5 ps after the cross-coupled level shifter. The resulting rail-to-rail signal is buffered to drive the MPD.

5.1.6 Output Driver

The PLL output is sent off-chip by the output driver shown in Figure 5.10. The PLL output is first divided down by 8 since measuring the full speed output was not possible due to test setup limitations[1]. This division does not distort the range of interest (up to f_{REF}) and was compensated by adding $10\log(8^2) = 18.06$ dB to the measured phase noise.

[1]The available instruments were not capable of measuring the full-speed output, and 6.25 GHz requires an optimized high-speed driver such as LVDS or CML which is beyond the scope of this work.

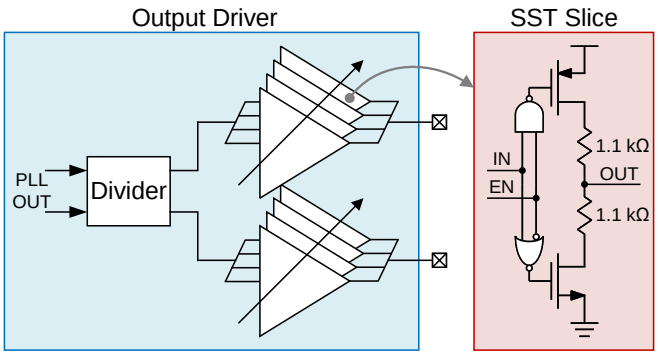


Figure 5.10 – Schematic of the output driver.

The divided output is sent out by the SST driver. The driver consists of 40 unit slices, each of which has an equivalent source impedance of 1.5 kΩ. Nominally, 30 slices would be sufficient to match 50 Ω. To counter process-voltage-temperature (PVT) variations, 15 of the slices can be switched on or off by a 4-bit control signal.

5.2 Simulation Results

5.2.1 Bandwidth Simulation

Simulating the loop bandwidth of the transistor-level schematic using the transient simulation method described in Section 4.2 takes an impractically long time. This is because an accurate result can only be obtained by simulating the parasitic-extracted netlist of the circuit which has thousands of elements. To shorten the simulation times, we used the sampled periodic transfer function (PXF) analysis offered by SpectreRF [254]. The sampled PXF can measure the instantaneous gain from a signal source to a single time point of the periodic output signal. Although this analysis is meant to be used for voltages and currents, we exploited it in the context of JTF by converting an AC voltage source v_{in} to an AC phase deviation ϕ_{ref} via a Verilog-A phase modulator model where

$$\phi_{ref} = v_{in}\,. \tag{5.2}$$

The AC output voltage variation v_{out} in response ϕ_{ref} can be converted back to output phase ϕ_{out} as

$$\phi_{out} = \frac{v_{out}}{\frac{dV_{OUT}}{dt}\Big|_{V_{OUT}=0}}\,, \tag{5.3}$$

where $\frac{dV_{OUT}}{dt}\Big|_{V_{OUT}=0}$ is the slew rate of the output at the zero crossing instant. The JTF can then be extracted as

$$JTF = \frac{\phi_{out}}{\phi_{in}} = \frac{v_{out}}{v_{in}}\left(\frac{dV_{OUT}}{dt}\Big|_{V_{OUT}=0}\right)^{-1}\,. \tag{5.4}$$

Figure 5.11a displays the JTF obtained using sampled PXF analysis for M being 1, 2 and 4. The loop bandwidths were measured as 260 MHz, 470 MHz and 695 MHz, and the jitter peaks were measured as 4 dB, 2.8 dB and 3.4 dB, respectively. The spur levels for the three cases were observed as -29.3 dBc, -40.2 dBc and -44.6 dBc as shown in Figure 5.11b.

5.2.2 Phase Noise Simulation

The phase noise of the PLL was simulated with SpectreRF's periodic noise analysis. Figure 5.12a-c display the plots obtained for each M value with individual noise contributors. At low frequencies the noise of MPD+CP is dominant for all. For M=1, VCO starts to dominate above 20 MHz. For M=2 and M=4, the contribution of the VCO noise is almost fully suppressed. Figure 5.12d plots the output phase noises and the VCO contributions in a single plot for a better comparison. The output spot noise reduction at 120 MHz for M=2 and 4 are 6 dB and 7.2 dB, respectively. The RMS jitter values integrated from 10 to 100 MHz are 162, 108, and 106 fs for M=1, 2, and 4, respectively.

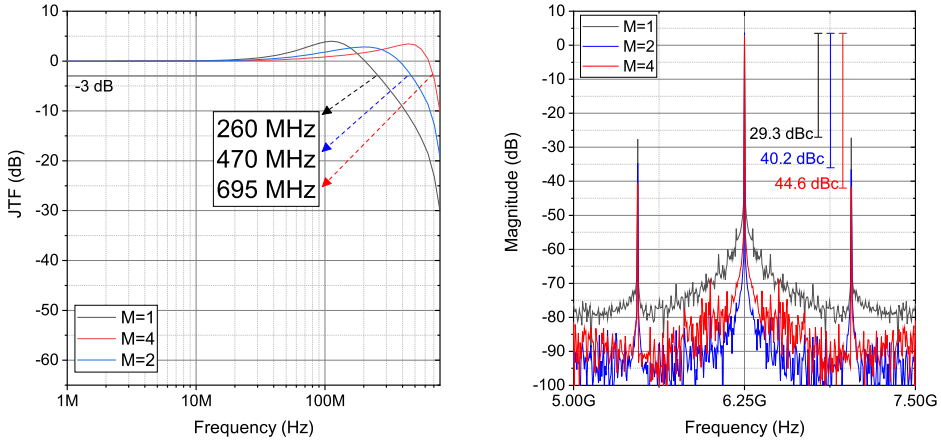
5.3 Measurement Results

The micrograph of the prototype fabricated in TSMC 40nm LP CMOS process is displayed in Figure 5.13. The core blocks highlighted in yellow occupy a total of 0.0085 mm^2 and dissipate 7 mW in total at 1.2 V supply voltage. The area and power breakdowns are given in Figure 5.14a and Figure 5.14b, respectively. The contribution of the additional PD+CPs is marginal compared to the power overhead of the VCO and the feedback divider. The fabricated prototype was characterized by two sets of measurements to verify the bandwidth extension and the phase noise performance using the measurement setup shown in Figure 5.15.

5.3.1 Bandwidth Measurements

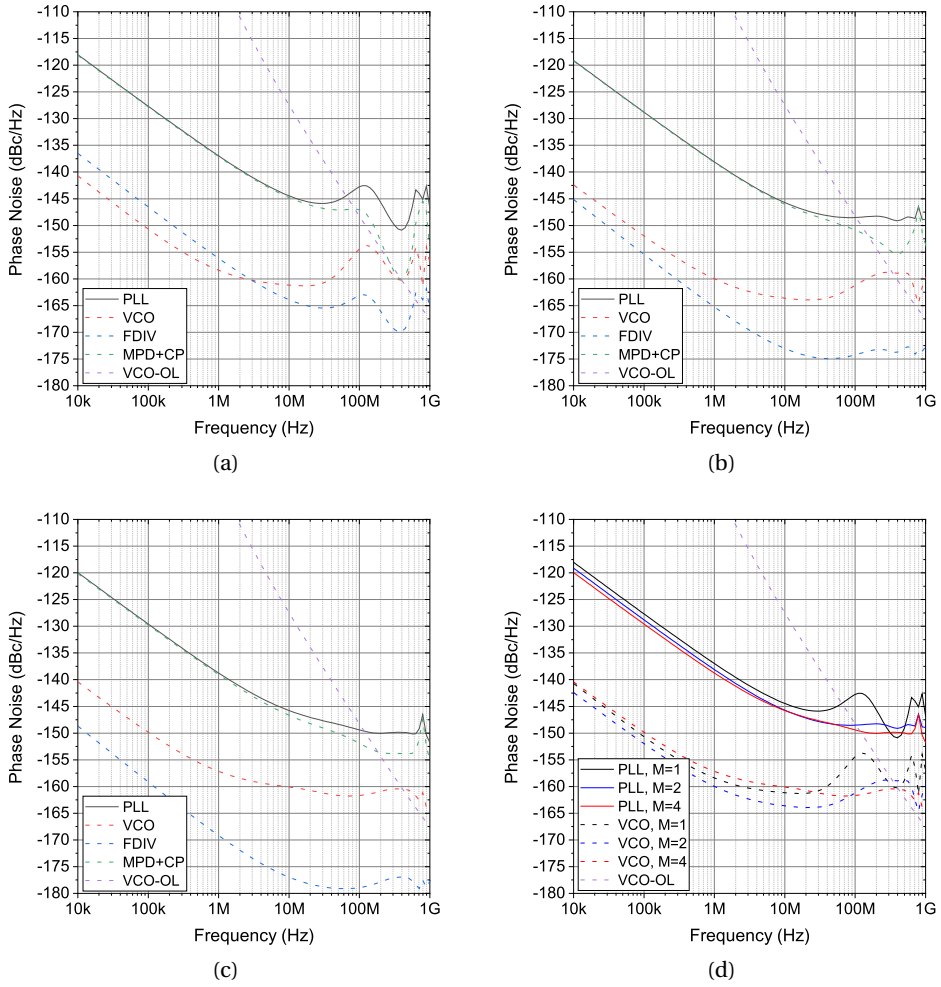
For bandwidth measurements, the 781.25 MHz reference frequency was summed with a high-frequency interferer and the peak at the output phase noise graph was registered. The interferer frequency was varied from 100 MHz to 700 MHz with 50 MHz increments to form the JTF. Figure 5.16a shows the overlaid PN plots with interference at 200, 400, and 600 MHz, measured for different values of M. The peak amplitude observed at 600 MHz indicates that high-frequency interference can pass through the loop when more than one PD is active.

Figure 5.16b shows the registered peak amplitudes for different values of M as a function of frequency. The measured -3 dB bandwidths were $0.48f_{REF}$, $0.74f_{REF}$, and $0.92f_{REF}$ for M values 1, 2, and 4, respectively. The peaking when M=4 is around 2.5 dB, which is acceptable for many applications. Lower peaking is achievable either by reducing I_{CP} or re-optimizing the loop filter. Figure 5.16c displays the same measurement with reduced I_{CP} which exhibits 1.7 dB peaking at the expense of slightly reduced ($0.83f_{REF}$) bandwidth.



(a) (b)

Figure 5.11 – Simulated (a) jitter transfer function and (b) reference spur plots.



(a) (b)

(c) (d)

Figure 5.12 – Simulated phase noise with (a) M=1, (b) M=2, and (c) M=4. (d) Comparison of the output phase noise and VCO contribution across M values.

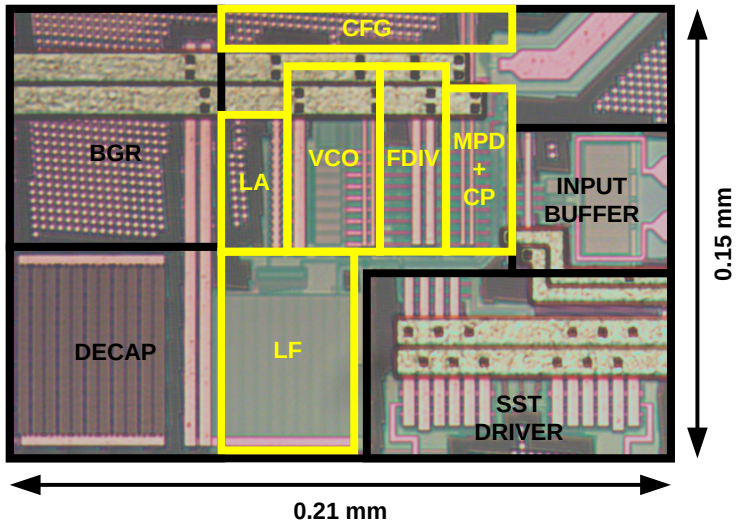
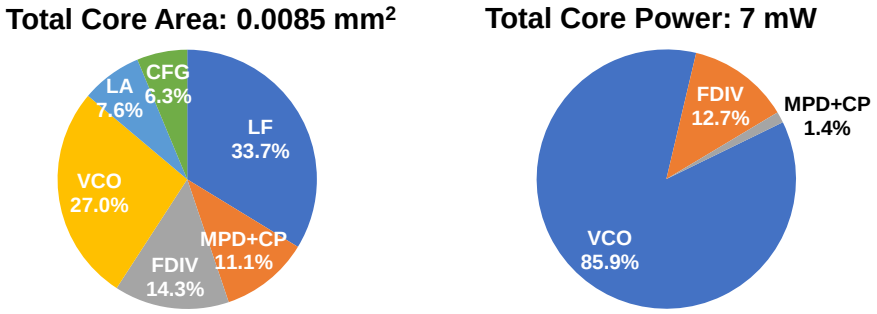


Figure 5.13 – Micrograph of the prototype fabricated in TSMC 40nm LP process.



(a) (b)

Figure 5.14 – (a) Area and (b) measured power breakdowns of the MPF PLL.

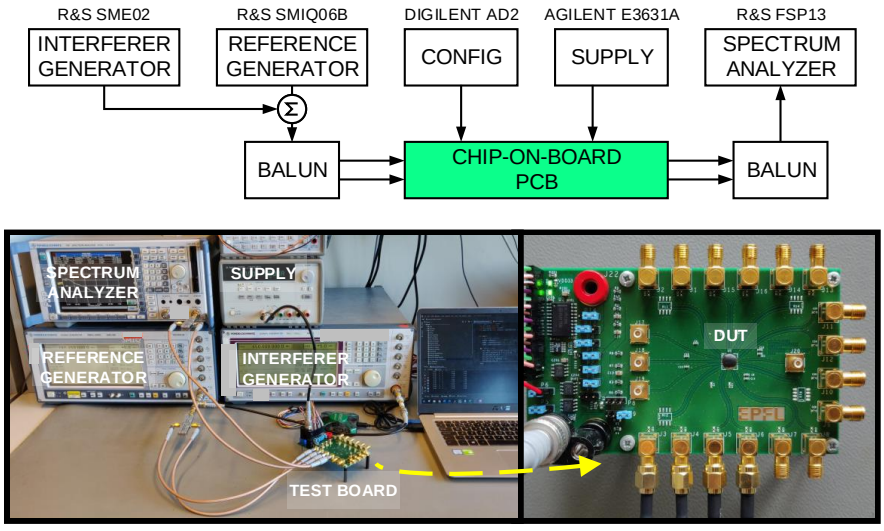


Figure 5.15 – Test setup for characterizing the PLL.

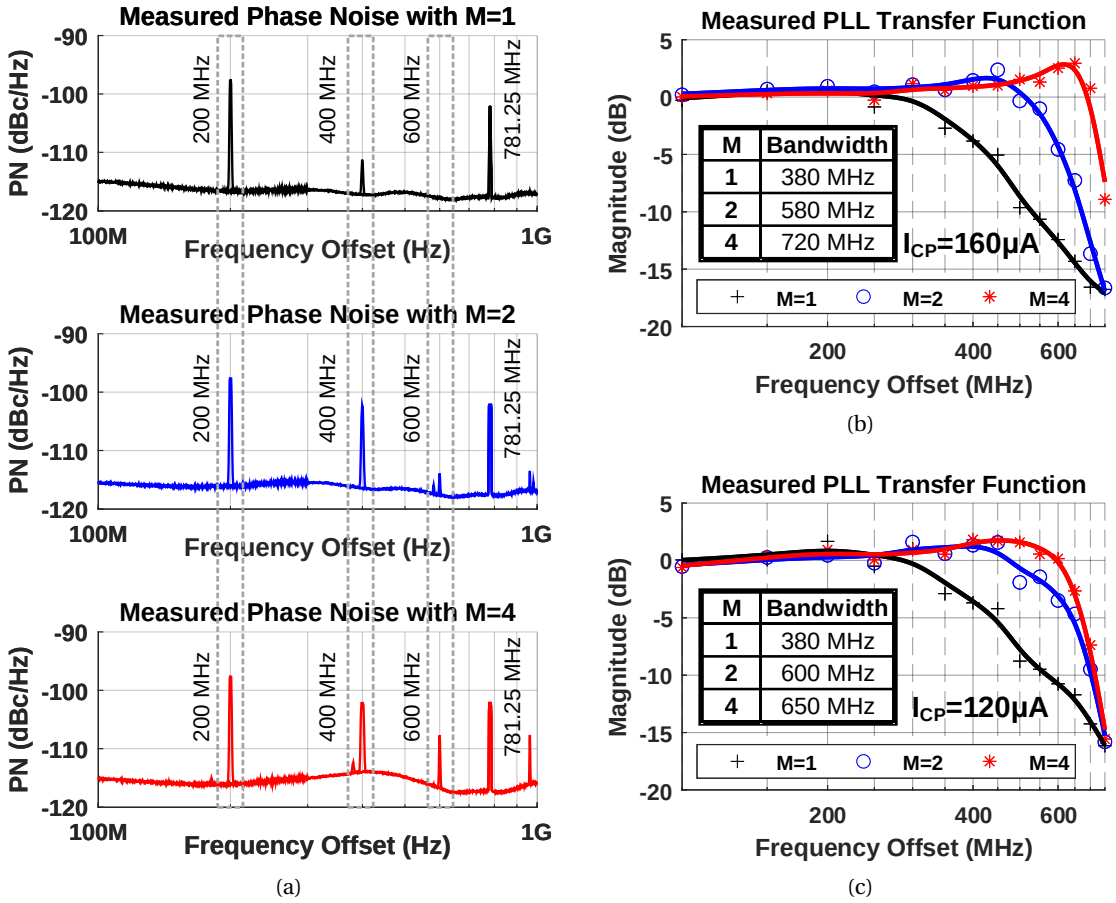


Figure 5.16 – (a) Measured phase noise with high frequency interferer, and the JTF plots for (b) I_{CP}=4 µA and (c) I_{CP}=1.5 µA.

5.3.2 Phase Noise Measurements

Figure 5.17 plots the output phase noise measured at different M values and compares them to the phase noise of the reference phase noise (normalized to 6.25 GHz) including the signal generator and the spectrum analyzer. At low frequencies, the PLL tracks the reference phase noise very closely, meaning that the measurement is limited by the instrument noise.

The jitter generation of the PLL starts to become prominent after 1 MHz. The overlaid plot verifies that the extended bandwidth results in higher suppression of the VCO phase noise as M increases. When we zoom in near f_{REF}, we observe that the output phase noise continues to track the reference for higher M. The higher jitter peaking for M=4 is also noticeable in this range.

Table 5.1 summarizes the spot noise and integrated jitter values. The comparison of the values for the 100 MHz-700 MHz band is especially interesting since the measured values converge to the reference value as M increases.

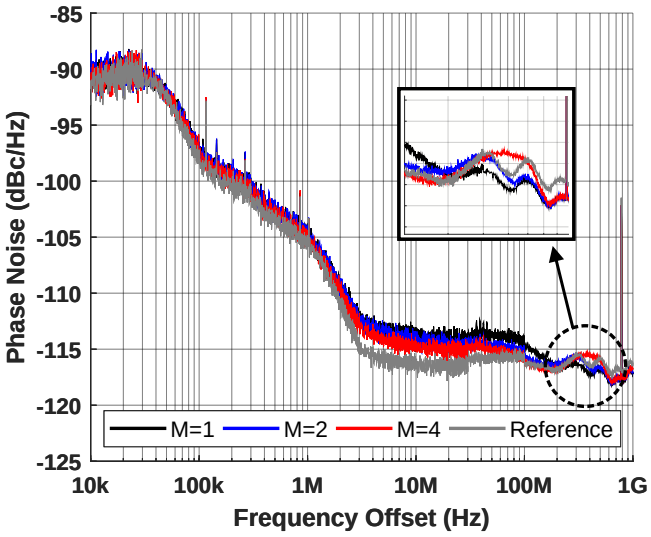


Figure 5.17 – Overlay of measured phase noise plots for different values of M.

Table 5.1 – Measured spot noise and integrated jitter values.

	Reference	M=1	M=2	M=4
Frequency Offset	Spot Noise (dBc/Hz)			
10 kHz	-91.77	-90.04	-90.28	-90.94
100 kHz	-98.21	-97.07	-97.21	-97.40
1 MHz	-105.53	-104.81	-104.68	-105.04
10 MHz	-116.07	-113.76	-114.18	-114.67
100 MHz	-116.26	-114.51	-115.44	-115.99
Integration BW	RMS Jitter (ps)			
10 kHz-10 MHz	0.42	0.47	0.47	0.45
50 kHz-10 MHz	0.37	0.42	0.42	0.40
10 kHz-100 MHz	0.69	0.84	0.79	0.75
10 kHz-200 MHz	0.87	1.02	0.96	0.92
100 MHz-700 MHz	1.34	1.28	1.30	1.34

5.3.3 Discussion

Table 5.2 summarizes the measured performance of the proposed PLL and compares it with the state-of-the-art wideband ring VCO-based PLLs [242, 249]. The loop bandwidth to reference frequency ratio obtained in this work is 4.5 and 4 times higher than what has been reported in [249] and [242], respectively. Moreover, this design achieves 1.8 times smaller silicon footprint compared to [242] in a similar process. The spot noise at 10 MHz is 8 dB lower than the other works. The integrated jitter values are also similar or lower despite the noise limitation of the measurement setup.

Table 5.2 – MPF-PLL performance summary and comparison with the state of the art.

	JSSC'16[242]	ISSCC'20[249]	**This work**
Technology	45nm	40nm	**40nm**
V_{DD} [V]	1	0.9	**1.2**
VCO Type	Ring	Ring	**Ring**
f_{CLK} [GHz]	2.4	5.25	**6.25**
f_{REF} [MHz]	22.6	21	**781.25**
Bandwidth [MHz]	5.6	4	**720[1] / 650[2]**
Bandwidth/f_{REF}	0.25	0.2	**0.92[1] / 0.83[2]**
Area [mm^2]	0.015	0.161	**0.0085**
Power [mW]	4	9.01	**7**
Spur [dBc]	-65	n/r	**-44.6‡**
PN @ 10 kHz [dBc/Hz]*	-86.91	-83.50	**-90.94**
PN @ 100 kHz [dBc/Hz]*	-100.87	-92.27	**-97.40**
PN @ 1 MHz [dBc/Hz]*	-105.47	-94.3	**-105.04**
PN @ 10 MHz [dBc/Hz]*	-107.99	-100.77	**-115.99**
RMS Jitter [ps]	0.97	1.95	**0.75**
Integration Range [Hz]	1k-200M	50k-10M	**10k-100M**
FoM [dB]**	-234.1	-224.6	**-234.0**

* Normalized to 6.25 GHz. ** FoM = $10\log_{10}(\sigma^2_{rms} \times P_{mW})$. ‡ Simulated.

[1] 2.5 dB peaking. [2] 1.7 dB peaking.

Conclusion

Part I: Implantable Neurotechnology

Summary

In the first part of this thesis, we addressed the contemporary challenges in implantable neurotechnology, which provides life-changing opportunities to many patients with neurological conditions. The future of this technology not only promises better treatments for a wider audience, but also aspires to enhance human capabilities by achieving symbiosis with machines. However, there is a wide technology gap between the current state of research and these ambitions which led to the present study. We extensively reviewed the current applications, limitations, and solutions for neural recording in Chapter 2 from a circuits and systems perspective. Based on this review, we identified two main research goals for future massively-parallel wireless neural recording SoCs.

First, we need to record as many neurons as possible for resolving the very implicit electrical manifestation of neural activity. As the number of channels reaches towards thousands, we see that a single-chip solution becomes infeasible in terms of both energy and area. To tackle this challenge, we presented an AFE architecture in Chapter 2 that applies various energy- and area-saving techniques to reduce the hardware resource demand of neural recording. The fabricated prototype in 65nm CMOS technology displays 12 times improvement in energy-area efficiency over the state of the art.

Secondly, we need wireless connectivity to make the brain interface technology available to a wider audience because a wired connection between the implant and the external devices is infection-prone and uncomfortable. Nevertheless, the current wireless solutions for implants do not support the ultra-high data throughput of massively-parallel recording. Inspired by the previous ideas for compression and the emerging machine learning applications in neuroscience, we introduced a new wireless neural recording approach in Chapter 3 to tackle the data rate bottleneck. This approach extracts spectral features on-chip using versatile and hardware-friendly CHT, and applies classification off-chip to maintain algorithm flexibility. A second mode of operation for the feature extractor is waveform reconstruction, which is extremely valuable for medical applications yet has been overlooked in previous studies.

The SoC prototype in 65nm CMOS containing the AFE, CHT processor, and WPDT subsystems achieved 2.9 μW power and 0.021 mm^2 area per channel, suggesting that a thousand-channel system would be feasible within a single 7 mm-by-3 mm chip. For comparison, an industry-standard 64-channel electrophysiology chip occupies 7.3 mm by 4.2 mm [255]. Moreover, the power density of the channel is 14 mW/cm^2 which is much below the 40 mW/cm^2 limit for temperature elevation.

The data rate reduction of the implemented approach is 80% without compromising the detection performance. Extrapolating to a thousand-channel system, the data rate reduction would be from 200 Mb/s to 40 Mb/s, which is within the capabilities of the current standards.

To summarize, Part I took a step towards future brain interfaces by exploiting trade-offs in circuits and systems and relaxing their resource requirements. However, the implications of this study go beyond the field of neurotechnology. In essence, any data acquisition problem with resource constraints and high throughput could benefit from the presented architectures.

Future Work

For the AFE part, the biggest room for improvement is in the design of the LNA. As seen in Figure 2.14b, the nonlinearity of the LNA causes around 1-bit reduction in the effective resolution of the overall AFE. Therefore, a more linear architecture can enhance the recording quality. A secondary consideration is the spread of the ENOB across fabricated samples. This is mainly due to the mismatch of the ULCDAC and it could be mitigated by enlarging the array with a small area penalty.

For the CHT processor part, the Hadamard indices do not fully exercise their allocated 16-bit width, especially for higher indices representing high frequencies [256]. Further data rate reduction could be achieved by statistically analyzing this redundancy.

For the software part, the next step would be to focus on highly sophisticated classifiers such as DNNs or recursive neural networks (RNNs) to fully realize the premise of the proposed hardware. A major challenge for these type of classifiers is the scarcity of patient data. Such classifiers require large amounts of high-quality data for each patient, which is usually not attainable with the currently available datasets. A recent method considers merging data from several patients which can relax this requirement [257]. Furthermore, the flexibility of the software classification could be further leveraged to create general-purpose neural interfaces, where a library of classifiers for various tasks controls CHT configurations or reconstructions.

The prototype fabricated for demonstration could include only 16-channels, whereas the projections based on the achieved results anticipate feasible integration of a thousand channels within a single die. However, reaching the stage where such a system can be fabricated and developed for chronic implantation requires the strong commitment of an interdisciplinary team and an ample budget.

Part II: High-Speed Communications

Summary

In the second part of the thesis, we focused on the timing uncertainty bottleneck in high-speed wireline and wireless communication systems. The increasing data rates and the emerging need for multitransceiver architectures necessitate the lowest energy and area for the frequency synthesis, which contradicts the current jitter-cost trade-off paradigm.

In Chapter 4, we investigated the limitations for PLL-based frequency synthesizers and identified a wide loop bandwidth as the key target, to mitigate the VCO noise and to allow the use of cost-effective ring oscillators instead of bulky LC counterparts. We introduced an MPF PLL-based synthesizer architecture as a solution to the bandwidth limitations in traditional PLLs. To accurately validate the benefit of the proposed system, we devised a time-domain simulation method instead of relying on linearized phase-domain models.

In Chapter 5, we presented the design of a 6.25 GHz MPF PLL in 40 nm CMOS. We described our circuit solutions and again used time domain simulation methods based on periodic steady-state analyses to validate the transistor-level circuit. The fabricated prototype achieved a measured bandwidth of up to 0.92 f_{REF} within a silicon footprint of 0.0085 mm^2, which improve the state of the art by 3 times and 1.8 times, respectively.

Future Work

Although the bandwidth extension was observable, the main limitation of this study was that the measured phase noise and jitter values were lower bounded by the instrument noise. We could obtain better results if we had access to a reference source and a spectrum analyzer with lower phase noise floors than the available instruments. Another setup-related limitation was that we were not able to measure the reference spur levels. This was because we had to divide the output clock by eight to be compatible with the available measurement equipment. A future prototype should include the necessary on-chip high-speed drivers for sending the full-speed clock off-chip.

The simulated and the measured jitter values of the proposed synthesizer still do not reach the levels required by next generation links. As pointed out by Razavi [239], the next-generation 224 Gbps 4-level pulse amplitude modulation (PAM-4) links, and 256-level quadrature amplitude modulation (256-QAM) for 5G radio will require below 100 fs timing precision. Now that the limiting factor is the PD noise with the virtually full bandwidth, a future work would be to focus on decreasing the contribution of the PD using techniques such as subsampling. Moreover, integrating the synthesizer in an actual transceiver could unlock system-level trade-offs to tolerate the residual jitter.

Bibliography

[1] A. Uran *et al.*, "An AC-Coupled Wideband Neural Recording Front-End With Sub-1 mm 2 ×fJ/conv-step Efficiency and 0.97 NEF," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 258–261, 2020.

[2] A. Uran *et al.*, "A 16-Channel Wireless Neural Recording System-on-Chip with CHT Feature Extraction Processor in 65nm CMOS," in *2021 IEEE Cust. Integr. Circuits Conf.* IEEE, Apr. 2021, pp. 1–2.

[3] A. Uran *et al.*, "A 6.25 GHz Integer-N Ring PLL with 0.92fREF Bandwidth Using a Multi-Phase Feedback Phase Detector in 40nm CMOS," *under Rev.*, 2021.

[4] A. L. Galvani, "The classic. The effects of artificial electricity on muscular motion. Aloysio Luigi Galvani." *Clin. Orthop. Relat. Res.*, vol. 88, pp. 2–10, 1972.

[5] E. H. Du Bois-Reymond, *Untersuchungen über thierische Elektricität.* Berlin: Verlag von G. Reimer, 1848.

[6] J. Bernstein, "Ueber den zeitlichen Verlauf der negativen Schwankung des Nervenstroms," *Pflüger, Arch. für die Gesammte Physiol. des Menschen und der Thiere*, vol. 1, no. 1, pp. 173–207, Dec. 1868.

[7] W. Einthoven, "Ein neues Galvanometer," *Ann. Phys.*, vol. 317, no. 13, pp. 1059–1071, 1903.

[8] E. D. Adrian and A. Forbes, "The all-or-nothing response of sensory nerve fibres," *J. Physiol.*, vol. 56, no. 5, pp. 301–330, Jul. 1922.

[9] H. S. Gasser and J. Erlanger, "A study of the action currents of nerve with the cathode ray oscillograph," *Am. J. Physiol. Content*, vol. 62, no. 3, pp. 496–524, Nov. 1922.

[10] E. D. Adrian, "The impulses produced by sensory nerve endings," *J. Physiol.*, vol. 61, no. 1, pp. 49–72, Mar. 1926.

[11] B. H. C. Matthews, "A special purpose amplifier," *Proc. Physiol. Soc.*, vol. 81, no. suppl, pp. 28–29, Mar. 1934.

[12] J. M. R. Delgado *et al.*, "Intracerebral radio stimulation and recording in completely free patients," *J. Nerv. Ment. Dis.*, vol. 147, no. 4, pp. 329–340, Oct. 1968.

[13] R. N. Noyce, "US2981877A Semiconductor device-and-lead structure," 1961.

[14] J. S. Kilby, "US3138743A Miniaturized electronic circuits," 1964.

[15] M. G. Dorman, M. A. Prisbe, and J. D. Meindl, "A Monolithic Signal Processor for a Neurophysiological Telemetry System," *IEEE J. Solid-State Circuits*, vol. 20, no. 6, pp. 1185–1193, Dec. 1985.

[16] K. Najafi and K. Wise, "An implantable multielectrode array with on-chip signal processing," *IEEE J. Solid-State Circuits*, vol. 21, no. 6, pp. 1035–1044, Dec. 1986.

[17] M. S. Steyaert, W. M. Sansen, and C. Zhongyuan, "A Micropower Low-Noise Monolithic Instrumentation Amplifier For Medical Purposes," *IEEE J. Solid-State Circuits*, vol. 22, no. 6, pp. 1163–1168, Dec. 1987.

[18] G. E. Moore, "Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff." *IEEE Solid-State Circuits Soc. Newsl.*, vol. 11, no. 3, pp. 33–35, Sep. 2006.

[19] R. R. Harrison *et al.*, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, Jan. 2007.

[20] E. Musk, "An Integrated Brain-Machine Interface Platform With Thousands of Channels," *J. Med. Internet Res.*, vol. 21, no. 10, p. e16194, Oct. 2019.

[21] F. A. Azevedo *et al.*, "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain," *J. Comp. Neurol.*, vol. 513, no. 5, pp. 532–541, Apr. 2009.

[22] K. R. Mills, Ed., *Oxford Textbook of Clinical Neurophysiology*. Oxford University Press, Dec. 2016, vol. 1.

[23] B. Sakmann and E. Neher, "Patch Clamp Techniques for Studying Ionic Channels in Excitable Membranes," *Annu. Rev. Physiol.*, vol. 46, no. 1, pp. 455–472, Oct. 1984.

[24] B. Gosselin, "Recent advances in neural recording microsystems," *Sensors*, vol. 11, no. 5, pp. 4572–4597, 2011.

[25] J. J. Moore *et al.*, "Dynamics of cortical dendritic membrane potential and spikes in freely behaving rats," *Science (80-.).*, vol. 355, no. 6331, p. eaaj1497, Mar. 2017.

[26] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes," *Nat. Rev. Neurosci.*, vol. 13, no. 6, pp. 407–420, Jun. 2012.

[27] K. M. Szostak, L. Grand, and T. G. Constandinou, "Neural Interfaces for Intracortical Recording: Requirements, Fabrication Methods, and Characteristics," *Front. Neurosci.*, vol. 11, Dec. 2017.

[28] A. Goyal *et al.*, "The development of an implantable deep brain stimulation device with simultaneous chronic electrophysiological recording and stimulation in humans," *Biosens. Bioelectron.*, vol. 176, p. 112888, Mar. 2021.

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

[30] NeuroNexus, "NeuroNexus 2020 Electrode Array Design Catalog," 2020.

[31] S. Wang *et al.*, "A Compact Quad-Shank CMOS Neural Probe With 5,120 Addressable Recording Sites and 384 Fully Differential Parallel Channels," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1625–1634, Dec. 2019.

[32] P. Campbell *et al.*, "A silicon-based, three-dimensional neural interface: manufacturing processes for an intracortical electrode array," *IEEE Trans. Biomed. Eng.*, vol. 38, no. 8, pp. 758–768, 1991.

[33] A. D. Degenhart *et al.*, "Histological evaluation of a chronically-implanted electrocorticographic electrode grid in a non-human primate," *J. Neural Eng.*, vol. 13, no. 4, p. 046019, Aug. 2016.

[34] D. Khodagholy *et al.*, "NeuroGrid: recording action potentials from the surface of the brain," *Nat. Neurosci.*, vol. 18, no. 2, pp. 310–315, Feb. 2015.

[35] J. del R. Millan and J. Carmena, "Invasive or Noninvasive: Understanding Brain-Machine Interface Technology [Conversations in BME," *IEEE Eng. Med. Biol. Mag.*, vol. 29, no. 1, pp. 16–22, Jan. 2010.

[36] L. R. Hochberg *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372–375, May 2012.

[37] H. M. Greiner *et al.*, "Preresection intraoperative electrocorticography (ECoG) abnormalities predict seizure-onset zone and outcome in pediatric epilepsy surgery," *Epilepsia*, vol. 57, no. 4, pp. 582–589, Apr. 2016.

[38] L. R. Hochberg *et al.*, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, pp. 164–171, Jul. 2006.

[39] G. Schalk *et al.*, "Two-dimensional movement control using electrocorticographic signals in humans," *J. Neural Eng.*, vol. 5, no. 1, pp. 75–84, Mar. 2008.

[40] A. Rastogi *et al.*, "Neural Representation of Observed, Imagined, and Attempted Grasping Force in Motor Cortex of Individuals with Chronic Tetraplegia," *Sci. Rep.*, vol. 10, no. 1, p. 1429, Dec. 2020.

[41] A. B. Ajiboye *et al.*, "Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration," *Lancet*, vol. 389, no. 10081, pp. 1821–1830, May 2017.

[42] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, Apr. 2019.

[43] D. A. Moses *et al.*, "Real-time decoding of question-and-answer speech dialogue using human cortical activity," *Nat. Commun.*, vol. 10, no. 1, p. 3096, Dec. 2019.

[44] J. G. Makin, D. A. Moses, and E. F. Chang, "Machine translation of cortical activity to text with an encoder–decoder framework," *Nat. Neurosci.*, vol. 23, no. 4, pp. 575–582, Apr. 2020.

[45] World Health Organization, "Atlas : country resources for neurological disorders 2004 : results of a collaborative study of the World Health Organization and the World Federation of Neurology." World Health Organization, Tech. Rep., 2004.

[46] R. S. Fisher *et al.*, "Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, Apr. 2005.

[47] World Health Organization, "Epilepsy: a public health imperative," World Health Organization, Tech. Rep., 2019.

[48] N. Jette, A. Y. Reid, and S. Wiebe, "Surgical management of epilepsy," *Can. Med. Assoc. J.*, vol. 186, no. 13, pp. 997–1004, Sep. 2014.

[49] M. Mohan *et al.*, "The long-term outcomes of epilepsy surgery," *PLoS One*, vol. 13, no. 5, p. e0196274, May 2018.

[50] J. Parvizi and S. Kastner, "Promises and limitations of human intracranial electroencephalography," *Nat. Neurosci.*, vol. 21, no. 4, pp. 474–483, Apr. 2018.

[51] B. Jarosiewicz and M. Morrell, "The RNS System: brain-responsive neurostimulation for the treatment of epilepsy," *Expert Rev. Med. Devices*, pp. 1–10, Sep. 2020.

[52] T. L. Skarpaas, B. Jarosiewicz, and M. J. Morrell, "Brain-responsive neurostimulation for epilepsy (RNS® System)," *Epilepsy Res.*, vol. 153, pp. 68–70, Jul. 2019.

[53] R. Balestrino and A. H. Schapira, "Parkinson disease," *Eur. J. Neurol.*, vol. 27, no. 1, pp. 27–42, Jan. 2020.

[54] V. L. Feigin *et al.*, "Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015," *Lancet Neurol.*, vol. 16, no. 11, pp. 877–897, Nov. 2017.

[55] F. L. Hitti *et al.*, "Long-term outcomes following deep brain stimulation for Parkinson's disease," *J. Neurosurg.*, vol. 132, no. 1, pp. 205–210, Jan. 2020.

[56] N. Koirala *et al.*, "Mapping of subthalamic nucleus using microelectrode recordings during deep brain stimulation," *Sci. Rep.*, vol. 10, no. 1, p. 19241, Dec. 2020.

[57] K. H. Park *et al.*, "Clinical outcome prediction from analysis of microelectrode recordings using deep learning in subthalamic deep brain stimulation for Parkinson's disease," *PLoS One*, vol. 16, no. 1, p. e0244133, Jan. 2021.

[58] S. Little *et al.*, "Adaptive deep brain stimulation in advanced Parkinson disease," *Ann. Neurol.*, vol. 74, no. 3, pp. 449–457, Sep. 2013.

[59] N. C. Swann *et al.*, "Gamma Oscillations in the Hyperkinetic State Detected with Chronic Human Brain Recordings in Parkinson's Disease," *J. Neurosci.*, vol. 36, no. 24, pp. 6445–6458, Jun. 2016.

[60] World Health Organization, "Global action plan on the public health response to dementia 2017 - 2025," World Health Organization, Tech. Rep., 2017.

[61] C. Babiloni *et al.*, "What electrophysiology tells us about Alzheimer's disease: a window into the synchronization and connectivity of brain neurons," *Neurobiol. Aging*, vol. 85, pp. 58–73, Jan. 2020.

[62] R. Ferri *et al.*, "Stacked autoencoders as new models for an accurate Alzheimer's disease classification support using resting-state EEG and MRI measurements," *Clin. Neurophysiol.*, vol. 132, no. 1, pp. 232–245, Jan. 2021.

[63] L. Scott *et al.*, "Neuronal network activity in the hippocampus of tau transgenic (Tg4510) mice," *Neurobiol. Aging*, vol. 37, pp. 66–73, Jan. 2016.

[64] B. Nouriziabari *et al.*, "Aberrant Cortical Event-Related Potentials During Associative Learning in Rat Models for Presymptomatic Stages of Alzheimer's Disease," *J. Alzheimer's Dis.*, vol. 63, no. 2, pp. 725–740, Apr. 2018.

[65] M. Stoiljkovic *et al.*, "Altered Cortical and Hippocampal Excitability in TgF344-AD Rats Modeling Alzheimer's Disease Pathology," *Cereb. Cortex*, vol. 29, no. 6, pp. 2716–2727, Jun. 2019.

[66] Blackrock Microsystems, "Cerebus Multi-Channel Neural Data Acquisition."

[67] "RZ2 BioAmp Processor | Tucker-Davis Technologies."

[68] N. V. Thakor, "Translating the Brain-Machine Interface," *Sci. Transl. Med.*, vol. 5, no. 210, pp. 210ps17–210ps17, Nov. 2013.

[69] N. Even-Chen *et al.*, "Power-saving design opportunities for wireless intracortical brain–computer interfaces," *Nat. Biomed. Eng.*, vol. 4, no. 10, pp. 984–996, Oct. 2020.

[70] A. H. Marblestone *et al.*, "Physical principles for scalable neural recording," *Front. Comput. Neurosci.*, vol. 7, 2013.

[71] Qing Bai and K. Wise, "Single-unit neural recording with active microelectrode arrays," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 8, pp. 911–920, 2001.

[72] S. F. Lempka *et al.*, "Theoretical analysis of intracortical microelectrode recordings," *J. Neural Eng.*, vol. 8, no. 4, p. 045006, Aug. 2011.

[73] B. Murmann, "Energy limits in A/D converters," in *2013 IEEE Faibl. Tens. Faibl. Consomm.* IEEE, Jun. 2013, pp. 1–4.

[74] B. E. Jonsson, "Area efficiency of ADC architectures," in *2011 20th Eur. Conf. Circuit Theory Des.* IEEE, Aug. 2011, pp. 560–563.

[75] M. Verhelst and B. Murmann, "Area scaling analysis of CMOS ADCs," *Electron. Lett.*, vol. 48, no. 6, p. 314, 2012.

[76] R. Olsson and K. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2796–2804, Dec. 2005.

[77] B. Miccoli *et al.*, "High-Density Electrical Recording and Impedance Imaging With a Multi-Modal CMOS Multi-Electrode Array Chip," *Front. Neurosci.*, vol. 13, Jun. 2019.

[78] J. D. Simeral *et al.*, "Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array," *J. Neural Eng.*, vol. 8, no. 2, p. 025027, Apr. 2011.

[79] T. Kaiju *et al.*, "High Spatiotemporal Resolution ECoG Recording of Somatosensory Evoked Potentials with Flexible Micro-Electrode Arrays," *Front. Neural Circuits*, vol. 11, Apr. 2017.

[80] L. Muller *et al.*, "Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography," *J. Neural Eng.*, vol. 13, no. 5, p. 056013, Oct. 2016.

[81] J. J. Jun *et al.*, "Fully integrated silicon probes for high-density recording of neural activity," *Nature*, vol. 551, no. 7679, pp. 232–236, Nov. 2017.

[82] K. A. Ng *et al.*, "Implantable neurotechnologies: a review of integrated circuit neural amplifiers," *Med. Biol. Eng. Comput.*, vol. 54, no. 1, pp. 45–62, Jan. 2016.

[83] A. Bagheri *et al.*, "Low-Frequency Noise and Offset Rejection in DC-Coupled Neural Amplifiers: A Review and Digitally-Assisted Design Tutorial," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 1, pp. 161–176, Feb. 2017.

[84] W. Wattanapanitch, M. Fee, and R. Sarpeshkar, "An energy-efficient micropower neural recording amplifier," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 2, pp. 136–147, Jun. 2007.

[85] C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling*. Chichester, UK: John Wiley & Sons, Ltd, Jul. 2006.

[86] R. Muller, S. Gambini, and J. M. Rabaey, "A 0.013 mm2, 5 μW, DC-Coupled Neural Signal Acquisition IC With 0.5 V Supply," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 232–243, Jan. 2012.

[87] Yen-Po Chen, D. Blaauw, and D. Sylvester, "A 266nW multi-chopper amplifier with 1.38 noise efficiency factor for neural signal recording," in *2014 Symp. VLSI Circuits Dig. Tech. Pap.* IEEE, Jun. 2014, pp. 1–2.

[88] F. Zhang *et al.*, "A low-power ECoG/EEG processing IC with integrated multiband energy extractor," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 58, no. 9, pp. 2069–2082, Sep. 2011.

[89] D. Luo, M. Zhang, and Z. Wang, "A Low-Noise Chopper Amplifier Designed for Multi-Channel Neural Signal Acquisition," *IEEE J. Solid-State Circuits*, vol. 54, no. 8, pp. 2255–2265, Aug. 2019.

[90] Dong Han *et al.*, "A 0.45 V 100-Channel Neural-Recording IC With Sub-μW Channel Consumption in 0.18 μm CMOS," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 6, pp. 735–746, Dec. 2013.

[91] X. Liu *et al.*, "A Fully Integrated Wireless Compressed Sensing Neural Signal Acquisition System for Chronic Recording and Brain Machine Interface," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 4, pp. 874–883, Aug. 2016.

[92] B. Johnson and A. Molnar, "An Orthogonal Current-Reuse Amplifier for Multi-Channel Sensing," *IEEE J. Solid-State Circuits*, vol. 48, no. 6, pp. 1487–1496, Jun. 2013.

[93] W. M. Chen *et al.*, "A fully integrated 8-channel closed-loop neural-prosthetic cmos soc for real-time epileptic seizure control," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 232–247, Jan. 2014.

[94] X. Zou *et al.*, "A 100-Channel 1-mW implantable neural recording IC," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 60, no. 10, pp. 2584–2596, Oct. 2013.

[95] Moo Sung Chae *et al.*, "A 128-Channel 6 mW Wireless Neural Recording IC With Spike Feature Extraction and UWB Transmitter," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 4, pp. 312–321, Aug. 2009.

[96] M. Rezaei *et al.*, "A low-power current-reuse analog front-end for high-density neural recording implants," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 2, pp. 271–280, Apr. 2018.

[97] M. Delgado-Restituto *et al.*, "System-Level Design of a 64-Channel Low Power Neural Spike Recording Sensor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 2, pp. 420–433, Apr. 2017.

[98] W.-S. Liew *et al.*, "A 1-V 60-μW 16-channel interface chip for implantable neural recording," in *2009 IEEE Cust. Integr. Circuits Conf.*, no. Cicc. IEEE, Sep. 2009, pp. 507–510.

[99] B. G. Do Valle, S. S. Cash, and C. G. Sodini, "Low-Power, 8-Channel EEG Recorder and Seizure Detector ASIC for a Subdermal Implantable System," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 6, pp. 1058–1067, Dec. 2016.

[100] J. P. Uehlin *et al.*, "A 0.0023 mm^2/ch. Delta-Encoded, Time-Division Multiplexed Mixed-Signal ECoG Recording Architecture With Stimulus Artifact Suppression," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 319–331, Apr. 2020.

[101] X. Zou *et al.*, "A 1V 22μW 32-channel implantable EEG recording IC," in *2010 IEEE Int. Solid-State Circuits Conf.* -, vol. 53. IEEE, Feb. 2010, pp. 126–127.

[102] G. Gagnon-Turcotte *et al.*, "A 0.13-<![CDATA[<inline-formula> <tex-math notation="LaTeX">μm </tex-math> </inline-formula>]]> CMOS SoC for Simultaneous Multichannel Optogenetics and Neural Recording," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3087–3100, Nov. 2018.

[103] M. Han *et al.*, "Bulk switching instrumentation amplifier for a high-impedance source in neural signal recording," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 62, no. 2, pp. 194–198, Feb. 2015.

[104] S. Rai *et al.*, "A 500μW Neural Tag with 2μVrms AFE and Frequency-Multiplying MICS/ISM FSK Transmitter," in *2009 IEEE Int. Solid-State Circuits Conf. - Dig. Tech. Pap.* IEEE, Feb. 2009, pp. 212–213,213a.

[105] A. Bonfanti *et al.*, "A multi-channel low-power IC for neural spike recording with data compression and narrowband 400-MHz MC-FSK wireless transmission," in *2010 Proc. ESSCIRC*. IEEE, Sep. 2010, pp. 330–333.

[106] S.-Y. Park *et al.*, "Dynamic Power Reduction in Scalable Neural Recording Interface Using Spatiotemporal Correlation and Temporal Sparsity of Neural Signals," *IEEE J. Solid-State Circuits*, vol. 53, no. 4, pp. 1102–1114, Apr. 2018.

[107] M. Azin *et al.*, "A battery-powered activity-dependent intracortical microstimulation IC for brain-machine-brain interface," *IEEE J. Solid-State Circuits*, vol. 46, no. 4, pp. 731–745, Apr. 2011.

[108] L. Lyu, D. Ye, and C.-J. R. Shi, "A 340 nW/Channel 110 dB PSRR Neural Recording Analog Front-End Using Replica-Biasing LNA, Level-Shifter Assisted PGA, and Averaged LFP Servo Loop in 65 nm CMOS," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 4, pp. 811–824, Aug. 2020.

[109] Y. Liu *et al.*, "A 64-Channel Versatile Neural Recording SoC with Activity-Dependent Data Throughput," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 6, pp. 1344–1355, Dec. 2017.

[110] H. Kassiri *et al.*, "Rail-to-rail-input dual-radio 64-channel closed-loop neurostimulator," *IEEE J. Solid-State Circuits*, vol. 52, no. 11, pp. 2793–2810, 2017.

[111] M. Mollazadeh *et al.*, "Micropower CMOS integrated low-noise amplification, filtering, and digitization of multimodal neuropotentials," *IEEE Trans. Biomed. Circuits Syst.*, vol. 3, no. 1, pp. 1–10, Feb. 2009.

[112] F. Zhang, J. Holleman, and B. P. Otis, "Design of ultra-low power biopotential amplifiers for biosignal acquisition applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 4, pp. 344–355, Aug. 2012.

[113] Y.-P. Lin *et al.*, "A Battery-Less, Implantable Neuro-Electronic Interface for Studying the Mechanisms of Deep Brain Stimulation in Rat Models," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 1, pp. 98–112, Feb. 2016.

[114] Y. Jia *et al.*, "A Trimodal Wireless Implantable Neural Interface System-on-Chip," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 6, pp. 1207–1217, Dec. 2020.

[115] C. Kim *et al.*, "A 3 mm × 3 mm Fully Integrated Wireless Power Receiver and Neural Interface System-on-Chip," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 6, pp. 1736–1746, Dec. 2019.

[116] S. Y. Park *et al.*, "Modular 128-Channel Δ - ΔΣ Analog Front-End Architecture Using Spectrum Equalization Scheme for 1024-Channel 3-D Neural Recording Microsystems," *IEEE J. Solid-State Circuits*, vol. 53, no. 2, pp. 501–514, Feb. 2018.

[117] A. Borna and K. Najafi, "A low power light weight wireless multichannel microsystem for reliable neural recording," *IEEE J. Solid-State Circuits*, vol. 49, no. 2, pp. 439–451, Feb. 2014.

[118] C. Qian, J. Parramon, and E. Sánchez-Sinencio, "A micropower low-noise neural recording front-end circuit for epileptic seizure detection," *IEEE J. Solid-State Circuits*, vol. 46, no. 6, pp. 1392–1405, Jun. 2011.

[119] T. Jang *et al.*, "A Noise-Efficient Neural Recording Amplifier Using Discrete-Time Parametric Amplification," *IEEE Solid-State Circuits Lett.*, vol. 1, no. 11, pp. 203–206, Nov. 2018.

[120] Y. Chen *et al.*, "A digitally assisted, signal folding neural recording amplifier," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 4, pp. 528–542, Aug. 2014.

[121] K. A. Ng and Y. P. Xu, "A Low-Power, High CMRR Neural Amplifier System Employing CMOS Inverter-Based OTAs With CMFB Through Supply Rails," *IEEE J. Solid-State Circuits*, vol. 51, no. 3, pp. 724–737, Mar. 2016.

[122] M. A. Bin Altaf, C. Zhang, and J. Yoo, "A 16-Channel Patient-Specific Seizure Onset and Termination Detection SoC With Impedance-Adaptive Transcranial Electrical Stimulator," *IEEE J. Solid-State Circuits*, vol. 50, no. 11, pp. 2728–2740, Nov. 2015.

[123] Ming Yin *et al.*, "A 100-Channel Hermetically Sealed Implantable Device for Chronic Wireless Neurosensing Applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 2, pp. 115–128, Apr. 2013.

[124] V. Majidzadeh, A. Schmid, and Y. Leblebici, "Energy efficient low-noise neural recording amplifier with enhanced noise efficiency factor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 3, pp. 262–271, Jun. 2011.

[125] S. Robinet *et al.*, "A Low-Power 0.7 μV_{rms} 32-Channel Mixed-Signal Circuit for ECoG Recordings," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 1, no. 4, pp. 451–460, Dec. 2011.

[126] R. Muller *et al.*, "A Minimally Invasive 64-Channel Wireless μECoG Implant," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 344–359, Jan. 2015.

[127] H. G. Rhew *et al.*, "A fully self-contained logarithmic closed-loop deep brain stimulation SoC with wireless telemetry and wireless power management," *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2213–2227, Oct. 2014.

[128] Honglei Wu and Yong Ping Xu, "A 1V 2.3/spl mu/W Biomedical Signal Acquisition IC," in *2006 IEEE Int. Solid State Circuits Conf. - Dig. Tech. Pap.* IEEE, 2006, pp. 119–128.

[129] R. Harrison and C. Charles, "A low-power low-noise cmos for amplifier neural recording applications," *IEEE J. Solid-State Circuits*, vol. 38, no. 6, pp. 958–965, Jun. 2003.

[130] A. Samiei and H. Hashemi, "A Bidirectional Neural Interface SoC With Adaptive IIR Stimulation Artifact Cancelers," *IEEE J. Solid-State Circuits*, pp. 1–1, 2021.

[131] M. Shoaran *et al.*, "Compact low-power cortical recording architecture for compressive multichannel data acquisition," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 6, pp. 857–870, Dec. 2014.

[132] W. Biederman *et al.*, "A Fully-Integrated, Miniaturized (0.125 mm²) 10.5 μW Wireless Neural Sensor," *IEEE J. Solid-State Circuits*, vol. 48, no. 4, pp. 960–970, Apr. 2013.

[133] K. Abdelhalim *et al.*, "915-MHz FSK/OOK wireless neural recording soc with 64 mixed-signal fir filters," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2478–2493, Oct. 2013.

[134] H. Chandrakumar and D. Markovic, "An 80-mVpp Linear-Input Range, 1.6-GΩ Input Impedance, Low-Power Chopper Amplifier for Closed-Loop Neural Recording That Is Tolerant to 650-mVpp Common-Mode Interference," *IEEE J. Solid-State Circuits*, vol. 52, no. 11, pp. 1–18, 2017.

[135] W. Biederman *et al.*, "A 4.78 mm 2 Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels With Digital Compression and Simultaneous Dual Stimulation," *IEEE J. Solid-State Circuits*, vol. 50, no. 4, pp. 1038–1047, Apr. 2015.

[136] H. Gao *et al.*, "HermesE: A 96-channel full data rate direct neural interface in 0.13 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 1043–1055, Apr. 2012.

[137] W. A. Smith *et al.*, "Exploiting Electrocorticographic Spectral Characteristics for Optimized Signal Chain Design: A 1.08 W Analog Front End With Reduced ADC Resolution Requirements," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 6, pp. 1171–1180, Dec. 2016.

[138] T. Denison *et al.*, "A 2 μW 100 nV/rtHz Chopper-Stabilized Instrumentation Amplifier for Chronic Measurement of Neural Field Potentials," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2934–2945, Dec. 2007.

[139] J. N. Aziz *et al.*, "256-channel neural recording and delta compression microsystem with 3D electrodes," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 995–1005, Mar. 2009.

[140] A. Karimi-Bidhendi *et al.*, "CMOS Ultralow Power Brain Signal Acquisition Front-Ends: Design and Human Testing," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 5, pp. 1111–1122, Oct. 2017.

[141] E. Greenwald *et al.*, "A Bidirectional Neural Interface IC With Chopper Stabilized BioADC Array and Charge Balanced Stimulator," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 5, pp. 990–1002, Oct. 2016.

[142] X. Liu *et al.*, "Design of a Closed-Loop, Bidirectional Brain Machine Interface System With Energy Efficient Neural Feature Extraction and PID Control," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 4, pp. 729–742, Aug. 2017.

[143] B. Gosselin *et al.*, "A mixed-signal multichip neural recording interface with bandwidth reduction," *IEEE Trans. Biomed. Circuits Syst.*, vol. 3, no. 3, pp. 129–141, Jun. 2009.

[144] B. Gosselin, M. Sawan, and C. A. Chapman, "A low-power integrated bioamplifier with active low-frequency suppression," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 3, pp. 184–192, Sep. 2007.

[145] C. M. Lopez *et al.*, "A multichannel integrated circuit for electrical recording of neural activity, with independent channel programmability," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 2, pp. 101–110, Apr. 2012.

[146] J. Lee *et al.*, "A 64 channel programmable closed-loop neurostimulator with 8 channel neural amplifier and logarithmic ADC," *IEEE J. Solid-State Circuits*, vol. 45, no. 9, pp. 1935–1945, Sep. 2010.

[147] J. Yoo *et al.*, "An 8-Channel Scalable EEG Acquisition SoC With Patient-Specific Seizure Classification and Recording Processor," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 214–228, Jan. 2013.

[148] Ming Yin and M. Ghovanloo, "A flexible clockless 32-ch simultaneous wireless neural recording system with adjustable resolution," in *2009 IEEE Int. Solid-State Circuits Conf. - Dig. Tech. Pap.* IEEE, Feb. 2009, pp. 432–433,433a.

[149] F. Shahrokhi *et al.*, "The 128-channel fully differential digital integrated neural recording and stimulation interface," *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 3, pp. 149–161, Jun. 2010.

[150] Z. Xiao *et al.*, "A 20μW neural recording tag with supply-current-modulated AFE in 0.13μm CMOS," in *2010 IEEE Int. Solid-State Circuits Conf.* -, vol. 53, no. 4. IEEE, Feb. 2010, pp. 122–123.

[151] N. Verma *et al.*, "A Micro-Power EEG Acquisition SoC With Integrated Feature Extraction Processor for a Chronic Seizure Detection System," *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 804–816, Apr. 2010.

[152] J. Huang *et al.*, "A 0.01-mm 2 Mostly Digital Capacitor-Less AFE for Distributed Autonomous Neural Sensor Nodes," *IEEE Solid-State Circuits Lett.*, vol. 1, no. 7, pp. 162–165, Jul. 2018.

[153] L. Shen, N. Lu, and N. Sun, "A 1-V 0.25-μW Inverter Stacking Amplifier With 1.07 Noise Efficiency Factor," *IEEE J. Solid-State Circuits*, vol. 53, no. 3, pp. 896–905, Mar. 2018.

[154] C. Enz and G. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization," *Proc. IEEE*, vol. 84, no. 11, pp. 1584–1614, 1996.

[155] C. Kim *et al.*, "A 92dB dynamic range sub-μVrms-noise 0.8μW/ch neural-recording ADC array with predictive digital autoranging," in *2018 IEEE Int. Solid - State Circuits Conf.* -, vol. 61. IEEE, Feb. 2018, pp. 470–472.

[156] W. Franks *et al.*, "Impedance Characterization and Modeling of Electrodes for Biomedical Applications," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 7, pp. 1295–1302, Jul. 2005.

[157] A. Nikas *et al.*, "A Continuous-Time Delta-Sigma Modulator Using a Modified Instrumentation Amplifier and Current Reuse DAC for Neural Recording," *IEEE J. Solid-State Circuits*, vol. 54, no. 10, pp. 2879–2891, Oct. 2019.

[158] G. O'Leary *et al.*, "NURIP: Neural Interface Processor for Brain-State Classification and Programmable-Waveform Neurostimulation," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3150–3162, Nov. 2018.

[159] R. Shulyzki *et al.*, "320-Channel Active Probe for High-Resolution Neuromonitoring and Responsive Neurostimulation," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 1, pp. 34–49, Feb. 2015.

[160] K. A. Ng and Yong Ping Xu, "A Compact, Low Input Capacitance Neural Recording Amplifier," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 5, pp. 610–620, Oct. 2013.

[161] M. R. Pazhouhandeh *et al.*, "Opamp-Less Sub-μW/Channel Δ-Modulated Neural-ADC With Super-GΩ Input Impedance," *IEEE J. Solid-State Circuits*, pp. 1–1, 2020.

[162] H. Chandrakumar and D. Markovic, "A 15.2-ENOB 5-kHz BW 4.5-<![CDATA[<inline-formula> <tex-math notation="LaTeX">μ </tex-math> </inline-formula>]]>W Chopped CT <![CDATA[<inline-formula> <tex-math notation="LaTeX">$\Delta\Sigma$ </tex-math> </inline-formula>]]>-ADC for Artifact-Tolerant," *IEEE J. Solid-State Circuits*, vol. 53, no. 12, pp. 3470–3483, Dec. 2018.

[163] A. M. Sodagar *et al.*, "An implantable 64-channel wireless microsystem for single-unit neural recording," *IEEE J. Solid-State Circuits*, vol. 44, no. 9, pp. 2591–2604, Sep. 2009.

[164] K. Abdelhalim *et al.*, "64-Channel UWB Wireless Neural Vector Analyzer SOC With a Closed-Loop Phase Synchrony-Triggered Neurostimulator," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2494–2510, Oct. 2013.

[165] H. Ando *et al.*, "Wireless Multichannel Neural Recording With a 128-Mbps UWB Transmitter for an Implantable Brain-Machine Interfaces," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 6, pp. 1068–1078, Dec. 2016.

[166] S.-I. Chang *et al.*, "BioBolt: A minimally-invasive neural interface for wireless epidural recording by intra-skin communication," *2011 Symp. VLSI Circuits - Dig. Tech. Pap.*, pp. 146–147, 2011.

[167] D. De Dorigo *et al.*, "Fully Immersible Subcortical Neural Probes With Modular Architecture and a Delta-Sigma ADC Integrated Under Each Electrode for Parallel Readout of 144 Recording Sites," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3111–3125, Nov. 2018.

[168] R. Harrison *et al.*, "Wireless Neural Recording With Single Low-Power Integrated Circuit," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 4, pp. 322–329, Aug. 2009.

[169] H. Jeon *et al.*, "A High DR, DC-Coupled, Time-Based Neural-Recording IC With Degeneration R-DAC for Bidirectional Neural Interface," *IEEE J. Solid-State Circuits*, vol. 54, no. 10, pp. 2658–2670, Oct. 2019.

[170] W. Jiang *et al.*, "A ±50-mV Linear-Input-Range VCO-Based Neural-Recording Front-End With Digital Nonlinearity Correction," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 173–184, Jan. 2017.

[171] B. C. Johnson *et al.*, "An implantable 700μW 64-channel neuromodulation IC for simultaneous recording and stimulation with rapid artifact recovery," in *2017 Symp. VLSI Circuits*. IEEE, Jun. 2017, pp. C48–C49.

[172] H. Kassiri *et al.*, "Battery-less Tri-band-Radio Neuro-monitor and Responsive Neurostimulator for Diagnostics and Treatment of Neurological Disorders," *IEEE J. Solid-State Circuits*, vol. 51, no. 5, pp. 1274–1289, May 2016.

[173] B. Lee *et al.*, "An Inductively-Powered Wireless Neural Recording and Stimulation System for Freely-Behaving Animals," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 2, pp. 413–424, Apr. 2019.

[174] C. Lee *et al.*, "A 6.5- μ W 10-kHz BW 80.4-dB SNDR G m -C-Based CT $\Delta\Sigma$ Modulator With a Feedback-Assisted G m Linearization for Artifact-Tolerant Neural Recording," *IEEE J. Solid-State Circuits*, vol. 55, no. 11, pp. 2889–2901, Nov. 2020.

[175] C. M. Lopez *et al.*, "An implantable 455-active-electrode 52-channel CMOS neural probe," *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 248–261, Jan. 2014.

[176] C. Mora Lopez *et al.*, "A Neural Probe With Up to 966 Electrodes and Up to 384 Configurable Channels in 0.13 μm SOI CMOS," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 3, pp. 510–522, Jun. 2017.

[177] V. Majidzadeh, A. Schmid, and Y. Leblebici, "A 16-channel, 359 μW, parallel neural recording system using Walsh-Hadamard coding," in *Proc. IEEE 2013 Cust. Integr. Circuits Conf.* IEEE, Sep. 2013, pp. 1–4.

[178] A. E. Mendrela *et al.*, "A Bidirectional Neural Interface Circuit with Active Stimulation Artifact Cancellation and Cross-Channel Common-Mode Noise Suppression," *IEEE J. Solid-State Circuits*, vol. 51, no. 4, pp. 955–965, Apr. 2016.

[179] Y. Perelman and R. Ginosar, "An Integrated System for Multichannel Neuronal Recording With Spike/LFP Separation, Integrated A/D Conversion and Threshold Detection," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 130–137, Jan. 2007.

[180] R. Ramezani *et al.*, "On-Probe Neural Interface ASIC for Combined Electrical Recording and Optogenetic Stimulation," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 3, pp. 576–588, Jun. 2018.

[181] D. Rozgic *et al.*, "A 0.338cm<![CDATA[<formula> <tex>^{3}</tex> </formula>]]>, Artifact-Free, 64-Contact Neuromodulation Platform for Simultaneous Stimulation and Sensing," *IEEE Trans. Biomed. Circuits Syst.*, pp. 1–1, 2018.

[182] J. Xu *et al.*, "A Wearable 8-Channel Active-Electrode EEG/ETI Acquisition System for Body Area Networks," *IEEE J. Solid-State Circuits*, vol. 49, no. 9, pp. 2005–2016, Sep. 2014.

[183] R. F. Yazicioglu *et al.*, "A 200 μW Eight-Channel EEG Acquisition ASIC for Ambulatory EEG Systems," *IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 3025–3038, Dec. 2008.

[184] B. Murmann, "The Race for the Extra Decibel: A Brief Review of Current ADC Performance Trajectories," *IEEE Solid-State Circuits Mag.*, vol. 7, no. 3, pp. 58–66, 2015.

[185] J. Voorhies and A. Cohen-Gadol, "Techniques for placement of grid and strip electrodes for intracranial epilepsy surgery monitoring: Pearls and pitfalls," *Surg. Neurol. Int.*, vol. 4, no. 1, p. 98, 2013.

[186] K. Gosalia *et al.*, "Thermal Elevation in the Human Eye and Head Due to the Operation of a Retinal Prosthesis," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 8, pp. 1469–1477, Aug. 2004.

[187] Sohee Kim *et al.*, "Thermal Impact of an Active 3-D Microelectrode Array Implanted in the Brain," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 493–501, Dec. 2007.

[188] P. D. Wolf, "Thermal Considerations for the Design of an Implanted Cortical Brain–Machine Interface (BMI)," in *Indwelling Neural Implant. Strateg. Contend. with Vivo Environ.* CRC Press/Taylor and Francis, 2008.

[189] K. M. Silay, C. Dehollain, and M. Declercq, "Numerical analysis of temperature elevation in the head due to power dissipation in a cortical implant," in *2008 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 3, no. 1. IEEE, Aug. 2008, pp. 951–956.

[190] H.-M. Lee, H. Park, and M. Ghovanloo, "A Power-Efficient Wireless System With Adaptive Supply Control for Deep Brain Stimulation," *IEEE J. Solid-State Circuits*, vol. 48, no. 9, pp. 2203–2216, Sep. 2013.

[191] D. R. Nair *et al.*, "Nine-year prospective efficacy and safety of brain-responsive neurostimulation for focal epilepsy," *Neurology*, vol. 95, no. 9, pp. e1244–e1256, Sep. 2020.

[192] C. Clément, *Brain-Computer Interface Technologies*. Cham: Springer International Publishing, 2019.

[193] A. Zebda *et al.*, "Single Glucose Biofuel Cells Implanted in Rats Power Electronic Devices," *Sci. Rep.*, vol. 3, no. 1, p. 1516, Dec. 2013.

[194] P. Bollella *et al.*, "A Microelectronic Sensor Device Powered by a Small Implantable Biofuel Cell," *ChemPhysChem*, vol. 21, no. 1, pp. 120–128, Jan. 2020.

[195] A. Apsel, "A Simple Guide to Low-Power Wireless Technologies: Balancing the Tradeoffs for the Internet of Things and Medical Applications," *IEEE Solid-State Circuits Mag.*, vol. 10, no. 4, pp. 16–23, 2018.

[196] T. Akin, K. Najafi, and R. M. Bradley, "A wireless implantable multichannel digital neural recording system for a micromachined sieve electrode," *IEEE J. Solid-State Circuits*, vol. 33, no. 1, pp. 109–118, 1998.

[197] S. B. Lee *et al.*, "An inductively powered scalable 32-channel wireless neural recording system-on-a-chip for neuroscience applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 6 PART 1, pp. 360–371, Dec. 2010.

[198] Microsemi Corporation, "ZL70103 Medical Implantable RF Transceiver Dataheet Revision 2," 2015.

[199] R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (IEEE Cat. No.03CH37439)*. IEEE, 2003, pp. 3325–3328.

[200] V. Karkare, S. Gibson, and D. Marković, "A 130-μW, 64-Channel Neural Spike-Sorting DSP Chip," *IEEE J. Solid-State Circuits*, vol. 46, no. 5, pp. 1214–1222, May 2011.

[201] C. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.

[202] E. Candes and M. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[203] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[204] R. G. Baraniuk *et al.*, "Model-Based Compressive Sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[205] B. Adcock *et al.*, "Breaking the coherence barrier: A new theory for compressed sensing," *Forum Math. Sigma*, vol. 5, p. e4, Feb. 2017.

[206] L. Baldassarre *et al.*, "Structured sampling and recovery of iEEG signals," in *2015 IEEE 6th Int. Work. Comput. Adv. Multi-Sensor Adapt. Process.*, vol. 1. IEEE, Dec. 2015, pp. 269–272.

[207] L. Baldassarre *et al.*, "Learning-Based Compressive Subsampling," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 4, pp. 809–822, Jun. 2016.

[208] C. Aprile *et al.*, "Learning-Based Near-Optimal Area-Power Trade-offs in Hardware Design for Neural Signal Acquisition," *Proc. 26th Ed. Gt. Lakes Symp. VLSI - GLSVLSI '16*, pp. 433–438, 2016.

[209] C. Aprile *et al.*, "DCT Learning-Based Hardware Design for Neural Signal Acquisition Systems," in *CF'17 Proc. Comput. Front. Conf.* New York, New York, USA: ACM Press, 2017, pp. 391–394.

[210] C. Aprile, "Learning-Based Hardware Design for Data Acquisition Systems," Ph.D. dissertation, EPFL, 2018.

[211] I. Iturrate, R. Chavarriaga, and J. d. R. Millán, "General principles of machine learning for brain-computer interfacing," in *Handb. Clin. Neurol. Vol. 168*. Elsevier B.V., 2020, pp. 311–328.

[212] J. Jacobs *et al.*, "High-frequency electroencephalographic oscillations correlate with outcome of epilepsy surgery," *Ann. Neurol.*, vol. 67, no. 2, pp. 209–220, Feb. 2010.

[213] T. Akiyama *et al.*, "Focal resection of fast ripples on extraoperative intracranial EEG improves seizure outcome in pediatric epilepsy," *Epilepsia*, vol. 52, no. 10, pp. 1802–1811, Oct. 2011.

[214] D. B. Silversmith *et al.*, "Plug-and-play control of a brain–computer interface through neural map stabilization," *Nat. Biotechnol.*, Sep. 2020.

[215] A. Reuther *et al.*, "Survey of Machine Learning Accelerators," in *2020 IEEE High Perform. Extrem. Comput. Conf.* IEEE, Sep. 2020, pp. 1–12.

[216] S.-A. Huang *et al.*, "A 1.9-mW SVM Processor With On-Chip Active Learning for Epileptic Seizure Control," *IEEE J. Solid-State Circuits*, vol. 55, no. 2, pp. 452–464, Feb. 2020.

[217] Y. Wang *et al.*, "26.3 A Closed-Loop Neuromodulation Chipset with 2-Level Classification Achieving 1.5V pp CM Interference Tolerance, 35dB Stimulation Artifact Rejection in 0.5ms and 97.8Detection," in *2020 IEEE Int. Solid- State Circuits Conf. -*. IEEE, Feb. 2020, pp. 406–408.

[218] M. Shoaran *et al.*, "Energy-Efficient Classification for Resource-Constrained Biomedical Applications," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 8, no. 4, pp. 693–707, Dec. 2018.

[219] Y. Wei *et al.*, "A Review of Algorithm and Hardware Design for AI-Based Biomedical Applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 145–163, Apr. 2020.

[220] T. Jochum, T. Denison, and P. Wolf, "Integrated circuit amplifiers for multi-electrode intracortical recording," *J. Neural Eng.*, vol. 6, no. 1, p. 012001, Feb. 2009.

[221] P. Harpe, "A Compact 10-b SAR ADC With Unit-Length Capacitors and a Passive FIR Filter," *IEEE J. Solid-State Circuits*, vol. 54, no. 3, pp. 636–645, Mar. 2019.

[222] C.-C. Liu *et al.*, "A 10-bit 50-MS/s SAR ADC With a Monotonic Capacitor Switching Procedure," *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 731–740, Apr. 2010.

[223] L. Kull *et al.*, "A 3.1 mW 8b 1.2 GS/s Single-Channel Asynchronous SAR ADC With Alternate Comparators for Enhanced Speed in 32 nm Digital SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3049–3058, Dec. 2013.

[224] Intan Technologies, "RHD2000 Series Digital Electrophysiology Interface Chips Datasheet."

[225] G. Schiavone *et al.*, "Soft, Implantable Bioelectronic Interfaces for Translational Research," *Adv. Mater.*, vol. 32, no. 17, p. 1906512, Apr. 2020.

[226] I. R. Minev *et al.*, "Electronic dura mater for long-term multimodal neural interfaces," *Science (80-.).*, vol. 347, no. 6218, pp. 159–163, Jan. 2015.

[227] A. Hassibi *et al.*, "Comprehensive study of noise processes in electrode electrolyte interfaces," *J. Appl. Phys.*, vol. 96, no. 2, pp. 1074–1082, Jul. 2004.

[228] S.-J. Kim *et al.*, "A Sub-μW/Ch Analog Front-End for Δ-Neural Recording With Spike-Driven Data Compression," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 1, pp. 1–14, Feb. 2019.

[229] K. Ture, "Wireless Power Transfer and Data Communication for Implantable Autonomous Monitoring and Detection of Epilepsy In Vivo," Ph.D. dissertation, EPFL, 2019.

[230] Federal Communications Commission, "15.517 Technical requirements for indoor UWB systems," *Code Fed. Regul. Title 47 Telecommun.*

[231] A. Shoeb, "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[232] "International Epilepsy Electrophysiology Portal."

[233] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* New York, NY, USA: ACM, Aug. 2016, pp. 785–794.

[234] M. Zijlmans *et al.*, "High-frequency oscillations as a new biomarker in epilepsy," *Ann. Neurol.*, vol. 71, no. 2, pp. 169–178, Feb. 2012.

[235] S. Mattisson, "An Overview of 5G Requirements and Future Wireless Networks: Accommodating Scaling Technology," *IEEE Solid-State Circuits Mag.*, vol. 10, no. 3, pp. 54–60, 2018.

[236] B. Razavi, "Historical Trends in Wireline Communications: 60? Improvement in Speed in 20 Years," *IEEE Solid-State Circuits Mag.*, vol. 7, no. 4, pp. 42–46, 2015.

[237] "IEEE Standard for Jitter and Phase Noise," *IEEE Std 2414-2020*, pp. 1–42, Feb. 2021.

[238] Crystek Microwave, "CRBSVS-01-1000."

[239] B. Razavi, "Jitter-Power Trade-Offs in PLLs," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 68, no. 4, pp. 1381–1387, Apr. 2021.

[240] A. Tajalli *et al.*, "A 1.02-pJ/b 20.83-Gb/s/Wire USR Transceiver Using CNRZ-5 in 16-nm FinFET," *IEEE J. Solid-State Circuits*, vol. 55, no. 4, pp. 1108–1123, Apr. 2020.

[241] S. Ek *et al.*, "A 28-nm FD-SOI 115-fs Jitter PLL-Based LO System for 24–30-GHz Sliding-IF 5G Transceivers," *IEEE J. Solid-State Circuits*, vol. 53, no. 7, pp. 1988–2000, Jul. 2018.

[242] L. Kong and B. Razavi, "A 2.4 GHz 4 mW Integer-N Inductorless RF Synthesizer," *IEEE J. Solid-State Circuits*, vol. 51, no. 3, pp. 626–635, Mar. 2016.

[243] Y. Eken and J. Uyemura, "Multiple-GHz ring and LC VCOs in 0.18 μm CMOS," in *2004 IEEE Radio Freq. Integr. Circuits Syst. Dig. Pap.* IEEE, 2004, pp. 475–478.

[244] T. Miyazaki, M. Hashimoto, and H. Onodera, "A performance comparison of PLLs for clock generation using ring oscillator VCO and LC oscillator in a digital CMOS process," in *ASP-DAC 2004 Asia South Pacific Des. Autom. Conf. 2004 (IEEE Cat. No.04EX753)*. IEEE, 2004, pp. 545–546.

[245] B. Razavi, *Design of CMOS Phase-Locked Loops.* Cambridge University Press, Jan. 2020.

[246] M. Mansuri and Chih-Kong Ken, "Jitter optimization based on phase-locked loop design parameters," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1375–1382, Nov. 2002.

[247] F. Gardner, "Charge-Pump Phase-Lock Loops," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1849–1858, Nov. 1980.

[248] P. Hanumolu *et al.*, "Analysis of Charge-Pump Phase-Locked Loops," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 51, no. 9, pp. 1665–1674, Sep. 2004.

[249] H.-H. Ting and T.-C. Lee, "25.6 A 5.25GHz Subsampling PLL with a VCO-Phase-Noise Suppression Technique," in *2020 IEEE Int. Solid- State Circuits Conf. -*. IEEE, Feb. 2020, pp. 390–392.

[250] T. Toifl *et al.*, "A 0.94-ps-RMS-jitter 0.016-mm/sup 2/ 2.5-GHz multiphase generator PLL with 360/spl deg/ digitally programmable phase shift for 10-Gb/s serial links," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2700–2712, Dec. 2005.

[251] Lizhong Sun and T. Kwasniewski, "A 1.25-GHz 0.35-μm monolithic CMOS PLL based on a multiphase ring oscillator," *IEEE J. Solid-State Circuits*, vol. 36, no. 6, pp. 910–916, Jun. 2001.

[252] Hai Qi Liu *et al.*, "A Low-Noise Multi-GHz CMOS Multiloop Ring Oscillator With Coarse and Fine Frequency Tuning," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 17, no. 4, pp. 571–577, Apr. 2009.

[253] B. Razavi, *Design of Analog CMOS Integrated Circuits*. New York: McGraw-Hill, 2001.

[254] Cadence Design Systems, "Spectre Circuit Simulator and Accelerated Parallel Simulator RF Analysis in ADE Explorer User Guide v19.1," Cadence Design Systems, Tech. Rep., 2020.

[255] Intan Technologies, "RHD2164 Digital Electrophysiology Interface Chip."

[256] W. Crowther and C. Rader, "Efficient coding of vocoder channel signals using linear transformation," *Proc. IEEE*, vol. 54, no. 11, pp. 1594–1595, 1966.

[257] D. Kostas, S. Aroca-Ouellette, and F. Rudzicz, "BENDR: using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data," Jan. 2021.

Arda Uran

IC Design Engineer

Rue de Lausanne 6, 1950 Sion
ardauran@gmail.com
linkedin.com/in/ardauran/en

EDUCATION

PhD in Microsystems and Microelectronics, EPFL 2021
MSc in Electrical and Electronics Engineering, EPFL 2017
BSc in Electronics and Communication Engineering, Istanbul Technical University 2015

PROFESSIONAL EXPERIENCE

EPFL – Lausanne, Switzerland
Doctoral Researcher – EPFL LIONS & LSM 2017 – 2021

- Thesis title: Resource Trade-Offs in Circuits and Systems: from Neurotechnology to Communications.
- Built innovative ICs for brain-machine interfaces and high-speed transceivers.
- Designed a neural sensor front-end with ultra-high power and area efficiency.
- Created an on-chip neural data compression framework for machine learning applications.
- Developed a bandwidth extension technique for suppressing VCO phase noise in ring PLLs.
- Visited California Institute of Technology (MICS) for three months as a visiting student researcher.
- Supervised three master's students, assisted analog and digital IC design courses.

Master's Thesis – EPFL LSM 2017

- Designed an interchannel mismatch calibration processor for a high-speed time-interleaved ADC.
- Implemented a JESD204B-compliant high-speed data transmission protocol for ADC readout.

EM Microelectronic-Marin SA – Neuchâtel, Switzerland
Engineering Intern – RFID Business Unit 2016

Designed an ADC for a temperature sensor from architecture research to silicon validation.

Analog Devices Inc. Design Center – **Istanbul, Turkey**
Intern – RF & Microwave Group 2014

Assisted laboratory measurement automation and datasheet characterization of HMC7044.

SKILLS

Analog / Mixed-Signal Design: ADCs, amplifiers, PLLs, mixed-signal simulation.
Digital Design: Semi-custom/full-custom flows, multi-clock-domain systems, dynamic logic.
Measurement: PCB design, RF instruments, automated test procedures.
EDA Tools: Cadence (Virtuoso, Innovus, SpectreRF, basic SKILL and Tcl scripting), Synopsys (Design Complier), Mentor Graphics (Calibre, ModelSim, Questa ADMS), Xilinx (ISE, Vivado), Altium Designer.
HDL: Verilog, VHDL, Verilog-AMS, VHDL-AMS.
PDK Experience: CMOS (40nm, 65nm), FD-SOI (28nm).
Data Analysis & Reporting: MATLAB, Python, Origin, Latex, Adobe (Illustrator, InDesign, Photoshop).
Certifications: PRINCE2 Foundation, NCEES FE, Management of Innovation and Technology Transfer.
Languages: English (fluent), Turkish (native), French (B1).

AWARDS AND HONORS

- Best Paper Award at 2018 PRIME Conference.
- EPFL Excellence Fellowship, 2015-2017.
- Top rank in ITU EE Faculty 2015 cohort.

Ce document a été imprimé au Centre d’impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.