



Interfaces neuronales CMOS haute résolution pour l'électrophysiologie et l'optogénétique en boucle fermée

Thèse

Gabriel Gagnon-Turcotte

Doctorat en génie électrique
Philosophiæ doctor (Ph. D.)

Québec, Canada

Résumé

L'avenir de la recherche sur les maladies du cerveau repose sur le développement de nouvelles technologies qui permettront de comprendre comment cet organe si complexe traite, intègre et transfère l'information. Parmi celles-ci, l'optogénétique est une technologie révolutionnaire qui permet d'utiliser de la lumière afin d'activer sélectivement les neurones du cortex d'animaux transgéniques pour observer leur effet dans un vaste réseau biologique. Ce cadre expérimental repose typiquement sur l'observation de l'activité neuronale de souris transgéniques, car elles peuvent exprimer une grande variété de gènes et de maladies et qu'elles sont peu coûteuses. Toutefois, la plupart des appareils de mesure ou de stimulation optogénétique disponible ne sont pas appropriés, car ils sont câblés, trop lourds et/ou trop simplistes. Malheureusement, peu de systèmes sans fil existent, et ces derniers sont grandement limités par la bande passante requise pour transmettre les données neuronales, et ils ne fournissent pas de stimulation optogénétique multicanal afin de stimuler et observer plusieurs régions du cerveau. Dans les dispositifs actuels, l'interprétation des données neuronales est effectuée *ex situ*, alors que la recherche bénéficierait grandement de systèmes sans fil assez intelligents pour interpréter et stimuler les neurones en boucle fermée, *in situ*.

Le but de ce projet de recherche est de concevoir des circuits analogiques-numériques d'acquisition et de traitement des signaux neuronaux, des algorithmes d'analyse et de traitement de ces signaux et des systèmes electro-optiques miniatures et sans fil pour :

- i) Mener des expériences combinant l'enregistrement neuronal et l'optogénétique multicanal haute résolution avec des animaux libres de leurs mouvements.
- ii) Mener des expériences optogénétiques synchronisées avec l'observation, c.-à-d. en boucle fermée, chez des animaux libres de leurs mouvements.
- iii) Réduire la taille, le poids et la consommation énergétique des systèmes optogénétiques sans fil afin de minimiser l'impact de la recherche chez de petits animaux.

Ce projet est en 3 phases, et ses principales contributions ont été rapportées dans dix conférences internationales (ISSCC, ISCAS, EMBC, etc.) et quatre articles de journaux publiés ou soumis, ainsi que dans un brevet et deux divulgations.

La conception d'un système optogénétique haute résolution pose plusieurs défis importants. Notamment, puisque les signaux neuronaux ont un contenu fréquentiel élevé (~ 10 kHz),

le nombre de canaux sous observation est limité par la bande passante des transmetteurs sans fil (2-4 canaux en général). Ainsi, la première phase du projet a visé le développement d'algorithmes de compression des signaux neuronaux et leur intégration dans un système optogénétique sans fil miniature et léger (2.8 g) haute résolution possédant 32 canaux d'acquisition et 32 canaux de stimulation optique. Le système détecte, compresse et transmet les formes d'onde des potentiels d'action (PA) produits par les neurones avec un *field programmable gate array* (FPGA) embarqué à faible consommation énergétique. Ce processeur implémente un algorithme de détection des PAs basé sur un seuillage adaptatif, ce qui permet de compresser les signaux en transmettant seulement les formes détectées. Chaque PA est davantage compressé par une transformée en ondelette discrète (DWT) de type Symmlet-2 suivie d'une technique de discrimination et de requantification dynamique des coefficients. Les résultats obtenus démontrent que cet algorithme est plus robuste que les méthodes existantes tout en permettant de reconstruire les signaux compressés avec une meilleure qualité (SNDR moyen de 25 dB \pm 5% pour un taux de compression (CR) de 4.2). Avec la détection, des CR supérieurs à 500 sont rapportés lors de la validation *in vivo*.

L'utilisation de composants commerciales dans des systèmes optogénétiques sans fil augmente la taille et la consommation énergétique, en plus de ne pas être optimisée pour cette application. La seconde phase du projet a permis de concevoir un système sur puce (SoC) *complementary metal oxide semiconductor* (CMOS) pour faire de l'enregistrement neuronal et de l'optogénétique multicanal, permettant de réduire significativement la taille et la consommation énergétique comparativement aux alternatives commerciales. Ceci est une contribution importante, car c'est la première puce à être dotée de ces deux fonctionnalités. Le SoC possède 10 canaux d'enregistrement et 4 canaux de stimulation optogénétique. La conception du bioamplificateur inclut une bande passante programmable (0.5 Hz - 7 kHz) et un faible bruit referré à l'entrée (IRN de 3.2 μV_{rms}), ce qui permet de cibler différents types de signaux biologiques (PA, LFP, etc.). Le convertisseur analogique numérique (ADC) de type Delta-Sigma ($\Delta\Sigma$) MASH 1-1-1 est conçu pour fonctionner de faibles taux de sur-échantillonnage ($\text{OSR} \leq 50$) pour réduire sa consommation et possède une résolution programmable (ENOB de 9.75 Bits avec un OSR de 25). Cet ADC exploite une nouvelle technique réduisant la taille du circuit en soustrayant la sortie de chaque branche du $\Delta\Sigma$ dans le domaine numérique, comparativement à la méthode analogique classique. La consommation totale d'un canal d'enregistrement est de 11.2 μW . Le SoC implémente un nouveau circuit de stimulation optique basé sur une source de courant de type cascode avec rétroaction, ce qui permet d'accommoder une large gamme de LED et de tensions de batterie comparativement aux circuits existants. Le SoC est intégré dans un système optogénétique sans fil et validé *in vivo*.

À ce jour et en excluant ce projet, aucun système sans-fil ne fait de l'optogénétique en boucle fermée simultanément au suivi temps réel de l'activité neuronale. Une contribution importante de ce travail est d'avoir développé le premier système optogénétique multicanal qui

est capable de fonctionner en boucle fermée et le premier à être validé lors d'expériences *in vivo* impliquant des animaux libres de leurs mouvements. Pour ce faire, la troisième phase du projet a visé la conception d'un SoC CMOS numérique, appelé *neural decoder integrated circuit* (ND-IC). Le ND-IC et le SoC développé lors de la phase 2 ont été intégrés dans un système optogénétique sans fil. Le ND-IC possède 3 modules : 1) le détecteur de PA adaptatif, 2) le module de compression possédant un nouvel arbre de tri pour discriminer les coefficients, et 3) le module de classement automatique des PA qui réutilise les données générées par le module de détection et de compression pour réduire sa complexité. Un lien entre un canal d'enregistrement et un canal de stimulation est établi selon l'association de chaque PA à un neurone, grâce à la classification, et selon l'activité de ce neurone dans le temps. Le ND-IC consomme 56.9 μW et occupe 0.08 mm^2 par canal. Le système pèse 1.05 g, occupe un volume de 1.12 cm^3 , possède une autonomie de 3h, et est validé *in vivo*.

Abstract

The future of brain research lies in the development of new technologies that will help understand how this complex organ processes, integrates and transfers information. Among these, optogenetics is a recent technology that allows the use of light to selectively activate neurons in the cortex of transgenic animals to observe their effect in a large biological network. This experimental setting is typically based on observing the neuronal activity of transgenic mice, as they express a wide variety of genes and diseases, while being inexpensive. However, most available neural recording or optogenetic devices are not suitable, because they are hard-wired, too heavy and/or too simplistic. Unfortunately, few wireless systems exist, and they are greatly limited by the required bandwidth to transmit neural data, while not providing simultaneous multi-channel neural recording and optogenetic, a must for stimulating and observing several areas of the brain. In current devices, the analysis of the neuronal data is performed *ex situ*, while the research would greatly benefit from wireless systems that are smart enough to interpret and stimulate the neurons in closed-loop, *in situ*.

The goal of this project is to design analog-digital circuits for acquisition and processing of neural signals, algorithms for analysis and processing of these signals and miniature electro-optical wireless systems for:

- i) Conducting experiments combining high-resolution multi-channel neuronal recording and high-resolution multi-channel optogenetics with freely-moving animals.
- ii) Conduct optogenetic experiments synchronized with the neural recording, i.e. in closed loop, with freely-moving animals.
- iii) Increase the resolution while reducing the size, weight and energy consumption of the wireless optogenetic systems to minimize the impact of research with small animals.

This project is in 3 phases, and its main contributions have been reported in ten conferences (ISSCC, ISCAS, EMBC, etc.) and four published journal papers, or submitted, as well as in a patent and two disclosures.

The design of a high resolution optogenetic system poses several challenges. In particular, since the neuronal signals have a high frequency content (~ 10 kHz), the number of chan-

nels under observation is limited by the bandwidth of the wireless transmitters (2-4 channels in general). Thus, the first phase of the project focused on the development of neural signal compression algorithms and their integration into a high-resolution miniature and lightweight wireless optogenetics system (2.8g), having 32 recording channels and 32 optical stimulation channels. This system detects, compresses and transmits the waveforms of the signals produced by the neurons, i.e. action potentials (AP), in real time, via an embedded low-power field programmable gate array (FPGA). This processor implements an AP detector algorithm based on adaptive thresholding, which allows to compress the signals by transmitting only the detected waveforms. Each AP is further compressed by a Symmlet-2 discrete wavelet transform (DWT) followed dynamic discrimination and requantification of the DWT coefficients, making it possible to achieve high compression ratios with a good reconstruction quality. Results demonstrate that this algorithm is more robust than existing approach, while allowing to reconstruct the compressed signals with better quality (average SNDR of 25 dB \pm 5% for a compression ratio (CR) of 4.2). With detection, CRs greater than 500 are reported during the *in vivo* validation.

The use of commercial components in wireless optogenetic systems increases the size and power consumption, while not being optimized for this application. The second phase of the project consisted in designing a complementary metal oxide semiconductor (CMOS) system-on-chip (SoC) for neural recording and multi-channel optogenetics, which significantly reduces the size and energy consumption compared to commercial alternatives. This is important contribution, since it's the first chip to integrate both features. This SoC has 10 recording channels and 4 optogenetic stimulation channels. The bioamplifier design includes a programmable bandwidth (0.5 Hz -7 kHz) and a low input-referred noise (IRN of 3.2 μ V_{rms}), which allows targeting different biological signals (AP, LFP, etc.). The Delta-Sigma ($\Delta\Sigma$) MASH 1-1-1 low-power analog-to-digital converter (ADC) is designed to work with low OSR (\leq 50), as to reduce its power consumption, and has a programmable resolution (ENOB of 9.75 bits with an OSR of 25). This ADC uses a new technique to reduce its circuit size by subtracting the output of each $\Delta\Sigma$ branch in the digital domain, rather than in the analog domain, as done conventionally. A recording channel, including the bioamplifier, the $\Delta\Sigma$ and the decimation filter, consumes 11.2 μ W. Optical stimulation is performed with an on-chip LED driver using a regulated cascode current source with feedback, which accommodates a wide range of LED parameters and battery voltages. The SoC is integrated into a wireless optogenetic platform and validated *in vivo*.

To date and excluding this project, no wireless system is making closed-loop optogenetics simultaneously to real-time monitoring of neuronal activity. An important contribution of this work is to have developed the first multi-channel optogenetic system that is able to work in closed-loop, and the first to be validated during *in vivo* experiments involving freely-moving animals. To do so, the third phase of the project aimed to design a digital CMOS chip, called

neural decoder integrated circuit (ND-IC). The ND-IC and the SoC developed in Phase 2 are integrated within a wireless optogenetic system. The ND-IC has 3 main cores: 1) the adaptive AP detector core, 2) the compression core with a new sorting tree for discriminating the DWT coefficients, and 3) the AP automatic classification core that reuses the data generated by the detection and compression cores to reduce its complexity. A link between a recording channel and a stimulation channel is established according to the association of each AP with a neuron, thanks to the classification, and according to the bursting activity of this neuron. The ND-IC consumes 56.9 μW and occupies 0.08 mm^2 per channel. The system weighs 1.05 g, occupies a volume of 1.12 cm^3 , has an autonomy of 3h, and is validated *in vivo*.

Table des matières

Résumé	ii
Abstract	v
Table des matières	viii
Liste des tableaux	x
Liste des figures	xi
Remerciements	xviii
Avant-propos	xix
Acronymes	1
Introduction	4
2 A Wireless Headstage for Combined Optogenetics and Multichannel Electro-physiological Recording	22
2.1 Résumé	22
2.2 Abstract	23
2.3 Introduction	24
2.4 System Overview	26
2.5 System Design	27
2.6 Firmware Design	29
2.7 Implementation and Measured Performance	37
2.8 In-vivo Results	43
2.9 Comparison	48
2.10 Conclusion	49
3 A 0.13 μm CMOS SoC for Simultaneous Multichannel Optogenetics and Neu-ral Recording	50
3.1 Résumé	50
3.2 Abstract	51
3.3 Introduction	51
3.4 System Architecture	54
3.5 Circuit Implementation and SoC Integration	58
3.6 Experimental Results	66

3.7	<i>In-vivo</i> Results	71
3.8	Discussion	75
3.9	Conclusion	76
4	A Wireless Electro-Optic Headstage with a 0.13-μm CMOS Custom Integrated DWT Neural Signal Decoder for Closed-Loop Optogenetics	77
4.1	Résumé	77
4.2	Abstract	78
4.3	Introduction	79
4.4	System-Level Overview	81
4.5	Closed-Loop Strategy	83
4.6	ND-IC Detector Core	84
4.7	ND-IC Compression Core	86
4.8	ND-IC Classification Core	88
4.9	Wireless Headstage Integration	92
4.10	ND-IC Experimental Results	92
4.11	Headstage <i>In Vivo</i> Results	99
4.12	Discussion	106
4.13	Conclusion	107
5	Publications et Distinctions	108
5.1	Liste des publications	108
5.2	Liste des distinctions	112
	Conclusion	113
	Bibliographie	116
A	Article supplémentaire	131
B	Analyses complémentaires	135
B.1	Algorithme de seuillage adaptatif	135
B.2	Algorithme de compression	142
B.3	Algorithme de classement	153
B.4	SoC : Stimulateur optique en technologie CMOS	159
B.5	SoC : Convertisseur $\Delta\Sigma$ MASH du 3 ^e ordre	162
B.6	SoC : Filtre de décimation	165
B.7	SoC : Module de contrôle	169

Liste des tableaux

2.1	System characteristics.	40
2.2	Comparison of spike detectors using synthetic neural signals with varying firing rates from 25 to 125 spikes/s.	41
2.3	Comparison with other wireless headstages.	46
3.1	Positive node subtraction results possibilities between DAC_{x-1} and DAC_x according to the latch outputs.	62
3.2	SUMMARY OF THE EXPERIMENTAL RESULTS	71
3.3	Comparison of IC for simultaneous electrophysiological recording and optogenetics.	72
4.1	HEADSTAGE SUMMARY	92
4.2	ND-IC SUMMARY	99
4.3	COMPARISON WITH OTHER SYSTEMS FOR CL OPTOGENETICS, OPTOGENETICS WITH RECORDING OR CL ELECT. STIMULATION.	101
B.1	Score calculé pour chacun des algorithmes de compression en fonction d'un CR de 2 à 8.	153
B.2	Fonction de coût calculée pour chacun des algorithmes de compression.	153
B.3	Performance de <i>clustering</i> des algorithmes à base de DWT (<i>Symmlet-2</i> et <i>Haar</i>), DD et PCA avec 12 signaux neuronaux différents. Les algorithmes basés sur la DWT comparent différents sous-niveaux, du niveau 2 au niveau 4, et jonction de niveaux (3-4) et les coefficients sont requantifiés sur 6 bits après maximisation sur la plage dynamique. Un système de point est attribué pour le premier (rouge + 3 pts), le second (bleu + 2 pts) et le troisième (verts + 1 pts) algorithme en fonction des résultats de classification.	157
B.4	Performance de <i>clustering</i> des algorithmes à base de DWT (<i>Symmlet-2</i> et <i>Haar</i>), DD et PCA avec 12 signaux neuronaux différents. Les algorithmes basés sur la DWT comparent différents sous-niveaux, du niveau 2 au niveau 4, et jonction de niveaux (3-4) et les coefficients sont conservés avec une précision de 16 bits. Un système de point est attribué pour le premier (rouge + 3 pts), le second (bleu + 2 pts) et le troisième (verts + 1 pts) algorithme en fonction des résultats de classification.	158
B.5	La table de vérité du DAC des 2 ^e et 3 ^e branches du $\Delta\Sigma$ qui permet d'effectuer l'opération $DAC_P - DAC_C$ sans utiliser de <i>switch caps</i> . Cette technique permet de sauver de l'espace, car elle ne nécessite pas de condensateurs ni de commutateurs supplémentaires comparativement au DAC de la 1 ^{re} branche.	165

Liste des figures

1.1	Modèle d'un PA enregistré avec une électrode extracellulaire. La région dé-polarisée se déplace d'un bout à l'autre de l'axone, ce qui génère une tension positive, négative et ensuite positive vue par l'électrode, c.-à-d. un PA tripha-sique.	6
1.2	Forme d'onde d'un train de PA triphasique extracellulaire comportant trois phases (avec polarités inversées), enregistré dans le cerveau d'un rat par une des puces présentées dans cette thèse [9, 10].	7
1.3	Forme d'onde forme typique d'un signal LFP enregistré dans le cerveau d'un rat par une des puces présentée dans cette thèse [9]	9
1.4	Système d'enregistrement de signaux électrophysiologiques câblé, tirée de [27] à gauche et de [28] à droite.	10
1.5	Concept d'une chaîne d'acquisition sans fil de signaux neuronaux.	11
1.6	a) Stimulation électrique versus stimulation/inhibition optique des neurones (image tirée de [44]), b) souris transgénique en pleine expérience optogéné-tique (image tirée de [44]).	13
1.7	Concept de fonctionnement en boucle fermée qui sera abordé au Chapitre 4 de cette thèse. Un lien entre le stimulateur et le circuit d'enregistrement est établi en fonction du classement des PA et en fonction de l'activité «burst» des neurones.	14
1.8	Chronologie de l'évolution du projet.	18
2.1	Block diagram of the proposed headstage system including the foldable PCB connected to the implantable module. The headstage can stimulate optically using up to 32 LEDs and can record from up to 32 microelectrodes using spike detection and data compression thanks to an embedded digital signal proces-sor implemented inside a low-power Spartan-6 FPGA. The headstage trans-mits the neural data wirelessly to a base station which is connected to a com-puter host. The implantable module that holds the LEDs, the optical fibers and the microelectrodes can be mounted on the head of the animal.	24
2.2	Block diagram illustrating the implemented spike detector. Each new inco-ming sample is stored inside a circular buffer in dedicated RAM memory. When a detection occurs, 48 spike samples (15 samples before, and 32 samples after the detection point) are sent to the <i>Wavelet lifting core</i>	30
2.3	Multichannel threshold estimator based on a Sigma-Delta control loop. The loop locks on the ratio of sample whose amplitude is greater than the output of the integrator (σ_a). The estimated standard deviation σ_a is multiplied by G_{thr} to give the threshold value.	31

2.4	Data flow diagram of a four level DWT lifting scheme using a <i>Symmlet-2</i> basis. The wavelet transform is performed on 48 spikes samples, thus producing 48 coefficients in total.	34
2.5	System diagram for <i>Symmlet-2</i> lifting sequential implementation. Each new incoming spike sample is stored inside a FIFO, before being processed by a unique <i>Wavelet lifting core</i> . A <i>Logic core</i> handles all the arithmetic operations by sending commands along with the AP samples to a <i>Computation core</i> , and coordinates the data flow between all DWT levels as well as the storage of the coefficients inside dedicated RAM memory block.	35
2.6	a) Power consumption of each building block when all 32 channels are activated with a firing rate of 45 spikes/s per channel. The measurements were performed with one LED driven by a 150 mA current at a duty cycle of 10%, b) power consumption of the whole system and the FPGA with respect to the neuronal firing rate, without optical stimulation.	38
2.7	a) Two-sided view of the rigid-flex PCB when not folded with the referenced sub-systems, b) top system view (folded) with battery on side, c) side system view (folded) attached to the implantable module.	39
2.8	Comparison with Delta compression and with CS using an optimal sparsity matrix, a) mean SNDR and SNDR variability, b) SNDR standard deviation relative to the mean SNDR.	41
2.9	Evaluation of the quality of signal compression and reconstruction of the proposed wavelet compression module for a CR of 4.17, and for input signals of SNR between 25 dB and 0 dB.	43
2.10	Implantable module showing one LED coupled to an implantable optical fiber packaged in a polyimide tubing with 8 recording tungsten microelectrodes, one reference microelectrode and a ground wire.	44
2.11	a) The head-fixed <i>in-vivo</i> experimental setting, b) The data transmission was propagated between the headstage and a base station situated 2 meters away.	45
2.12	Reconstructed neural signal using the detected and decompressed AP superimposed on the original signal. A zoom over a detected spike shows that the reconstructed signal is nearly identical to the original waveform.	45
2.13	Detected and clustered spikes recorded simultaneously from 8 microelectrode during a 200-seconds experiment.	47
2.14	Experimental results showing extracellular recordings in the brain of a ChR2 mouse in response to local optogenetic light activation of 10 ms with a period of 100 ms. AP with amplitude of up to 1 mV were evoked by the light activation, and a decreasing amplitude is observed at the beginning of the stimulation train, until stabilization after 5 seconds.	48
3.1	System-level concept of a wireless optogenetic platform utilizing the proposed SoC, including all the building blocks, to perform simultaneous neural recording and optical stimulation. The 0.13 μm CMOS SoC includes 10 neural recording channels and 4 optical stimulation channels. Each recording channel encompasses a low-noise bio-amplifier, a $\Delta\Sigma$ MASH 1-1-1 ADC and a decimation filter. The chip also includes a 4-ch LED driver circuit with feedback to precisely set the effective forward current of each LED for robust optical stimulation. It communicates through an SPI bus for easy interfacing and scalability.	52

3.2	Schematic of the bio-amplifier design with programmable high-pass cutoff frequency. The feedback resistor uses a 3-bit pseudo-resistor bank.	54
3.3	CIC4 decimation filter with optimized binary precision between the stages, for up to 14-bit output resolution. The low-pass cut-off frequency can be programmed simply by changing the clock reduction factor M and/or by modifying the oversampling frequency.	55
3.4	Circuit schematic of the OTA used for the bio-amplifier and its continuous time CMFB. The fully-differential folded-cascode OTA with source-degeneration resistors R_1 and R_2 is used.	56
3.5	a) Schematic of the $\Delta\Sigma$ MASH 1-1-1 using a digital <i>pre-subtraction module</i> for the 2 nd and 3 rd stages, b) Conventional implementation of the subtraction between the DAC outputs using switched capacitors, c) Bias duty-cycling strategy used to turn off the integrator OTA during both clock phases.	57
3.6	a) Schematic of the current-mirror OTA used for all three $\Delta\Sigma$ integrators, b) Switched capacitor CMFB of the OTA in a), c) Bias circuit of the current-mirror OTA with a <i>Bias_enable</i> switch to turn off the bias current of the OTA during both clock phases.	60
3.7	Timing diagram illustrating the bias duty cycling strategy. The bias current of the integrator OTA is turned ON/OFF using the control signal <i>Bias_enable</i> with a duty cycle of 35%, saving $\sim 41.3\%$ in power consumption.	61
3.8	Settling time modeling of the OTA during the integration phase, (a) timing diagram representing the settling, (b) schematic diagram of the OTA during the slewing phase, and (c) during the linear settling phase.	63
3.9	Schematic of the dynamic regenerative comparator utilized in each one bit $\Delta\Sigma$ quantizer.	63
3.10	Circuit schematic of the 4-ch LED driver. This circuit allows us to precisely control the value of the forward current in each LED using a regulated wide-swing current source with feedback. The average driving current is modulated by an external <i>PWM</i> signal.	65
3.11	Die micrograph (right) and system-level concept using the SoC within an optogenetic platform (left).	66
3.12	a) Bio-amplifier BW for different high-pass cutoff frequencies, b) measured ($3.2 \mu V_{rms}$) and simulated IRN, c) measured THD (0.95%).	67
3.13	a) Post-layout and measured output spectrum of the $\Delta\Sigma$ MASH 1-1-1 for an input signal amplitude of 80% of the full dynamic range (SNDR = 60.45 dB, at an OSR of 25 and a BW of 10 kHz). An ideal simulated noise shaping envelope for a $\Delta\Sigma$ MASH 1-1-1 is also shown, b) Measured SNDR according to the input signal amplitude relative to the full dynamic range.	68
3.14	a) Power consumption of the $\Delta\Sigma$ MASH 1-1-1 w/ DF for various OSR to achieve different ENOB, b) The power consumption of the bio-amplifier, the $\Delta\Sigma$ MASH 1-1-1 and the DF is $11.2 \mu W/ch$ ($F_s=500$ kHz) when using the bias duty cycling strategy ($3.1 \mu W$ less power).	69
3.15	a) Measured regulated currents against the LED anode voltage versus a simple current mirror, b) Regulated forward current pulses of 20 mA (which perfectly overlap) utilized to drive 3 different types of LEDs with different forward voltage characteristics, c) Regulated 20 mA current & intra/inter-channel variations due to process variation for 3 different types of LEDs.	70

3.16	The head-fixed <i>in vivo</i> experimental setup. The data was propagated between the wireless platform and a base station situated 2 meters away.	72
3.17	a) Spontaneous activity recorded using a micropipette probe, and clusters of spikes collected within 3 recording sessions (bottom right), b) SoC recording (in black) superimposed over the signal recorded using the ER-1 commercial wired amplifier (in red).	73
3.18	Evoked neural activity acquired after optical stimulation through the optrode (top), spontaneous activity recorded by the SoC (bottom).	74
3.19	a) Spontaneous and evoked LFP activity recorded through the optrode, b) Spectrogram of the recorded LFP with stimulation both in the motor cortex with Chr2 expression and in the control region.	75
4.1	System-level concept of the wireless CL optogenetic headstage using the proposed ND-IC as core module : a) The headstage is connected to the optrode and mounted on the head of the animal, b) the 0.13- μm CMOS ND-IC has 3 cores for AP detection, compression and classification, c) a custom 0.13- μm CMOS SoC is used for simultaneous neural recording and optogenetic stimulation and d) the CL algorithm is performed within a low-power FPGA. . . .	80
4.2	The FPGA control sequence used for interfacing the ND-IC cores, the SoC and the wireless transceiver altogether, while running the CL algorithm.	82
4.3	Diagram of the CL algorithm (top). This algorithm counts the number of detected AP and their cluster association within a sliding time window. For example, the CL algorithm can trigger a feedback stimulation pattern if 3 similar i.e. AP from a same cluster, are within a 25-ms time window (bottom). .	84
4.4	a) The pre-processor concept, when the amplitude of the neuronal signal exceeds a threshold, retrieves a detected AP waveform window set around the peak amplitude, and b) The detector core diagram.	85
4.5	The AP compression core circuit consists of three pipelined stages : a) a 4-Level Sym-2 lifting sequential DWT is performed on each detected AP, b) the detailed coefficients are processed by a sorting tree to find the coefficient having the N^{th} and the maximum amplitude, c) compression occurs by dynamically thresholding the DWT coefficients to keep only N of them. Further compression is provided using a dynamic re-quantization operation.	87
4.6	Unsupervised classification algorithm, the circuit can handle up to five clusters per channel, and reuse the V_{L3} vector produced by the compression core and the σ_n found by the AP detector core.	89
4.7	a) Chip micrograph, printed circuits board floorplan and wireless headstage w/ ND-IC for CL optogenetics (1.05 g). b) Different views of the packaged headstage w/ the battery (3.0 g), a typical optrode is also shown.	90
4.8	a) ROC curves for a 6 dB signal w/ varying firing rate, b) subset of the signal used to compute the ROC in (a), and c) CF scores for the proposed (black) and the NEO (gray) preprocessors for different SNR. The CF scores when neglecting the algorithm complexity are shown with dotted lines.	93
4.9	Power consumption of the proposed compression core vs the implementation in [20] for five different firing rates, b) and c) current consumption of the wireless transceiver with and without compression (CR of 4.17) at 2 Mbps and 250 kbps on-air data rates, respectively.	94

4.10	a) Classification results using the L3 Sym-2 DWT (black curve) versus the other techniques, b) unsupervised classification results of the proposed algorithm (solid lines) compared with K-means (squares), c) normalized classification performance according to the threshold T_1 , and d) classification results according to the signal SNR.	95
4.11	ND-IC power consumption : a) power breakdown for each core and SRAM block, b) power consumption according to the neural firing rate (black), and power consumption according to the cores operating frequency (blue).	97
4.12	a) Head-fixed <i>in vivo</i> setup with the rat. The neural signal was collected with a micropipette and processed <i>in situ</i> by the ND-IC. b) The headstage is connected to the optrode and installed on the head of the mouse. c) The freely-behaving animal during a CL experiment.	98
4.13	a) Reconstructed signal using the detected, compressed, and sorted AP collected <i>in situ</i> in the brain of the rat. The feedback trigger signal computed by the headstage is shown (in black). The trigger was configured to occur if three AP per clusters were found within a 150-ms time window. b) Detected, compressed and sorted waveforms at two other brain depths.	99
4.14	a) Reconstructed neural signal and CL stimulation triggers (in blue) acquired with a freely-moving mouse expressing a ChR2-mCherry in inhibitory neurons of the prelimbic cortex, b) a two signal closeup showing the CL stimulation pulses and evoked silencing phase due to the ChR2-mCherry activation, c) two CL window enlargements with and without CL stim. trigger, and d) AP sorted live by the ND-IC and off-line by a PCA with K-means (97% matching).	100
4.15	Normalized distribution of the time interval between two consecutive APs for experiments with and without CL stimulation. It shows that the CL stimulation increases the number of longer AP-to-AP intervals, because of the evoked silencing phases. Moreover, the number of shorter intervals also increases, which is in accordance with [146].	102
4.16	Control experiments : a) Recording/stimulation using a commercial Fi-Wi system where silencing phases are induced, b) CL stimulation performed in a non-injected mouse, no silencing phase are induced and c) representative confocal images of a mouse at six months and 11 months post infection (same animal), showing the ChR2-mCherry expression (red)	102
4.17	Matching between the AP sorted <i>in situ</i> by the headstage vs off-line using a PCA with K-means for 12 freely-moving experiments.	103
4.18	a) Threshold computed by the ND-IC (blue), by MAD (off-line, red) superimposed over the RAW neural signal, b) threshold standard deviation from its mean value for seven freely-moving experiments.	104
4.19	a) Mean SNDR within nine experiments, b) compressed APs waveform (red) superimposed on their original counterpart (black).	105
B.1	Fonction de densité de la loi demi-normale présentée avec des multiples de son écart-type (σ).	138
B.2	Boucle de contrôle permettant de calculer le seuil de détection basé sur un multiple de l'écart-type estimé (σ_a) du signal.	139
B.3	Diagramme fonctionnel de l'algorithme permettant de calculer le seuil de détection.	140

B.4	Représentation schématique d'une transformée en ondelettes de type <i>Symmlet-2</i> de 4 niveaux de profondeur par la méthode de <i>lifting</i>	143
B.5	Coefficients obtenus après une transformée en ondelettes <i>Symmlet-2</i> de niveau 4 et seuillés afin de conserver 20 coefficients sur 48.	144
B.6	SNDR moyen obtenu avec tous les algorithmes testés pour un CR de 2 à 8. . .	147
B.7	SNDR minimum obtenu avec tous les algorithmes testés pour un CR de 2 à 8.	147
B.8	SNDR maximum obtenu avec tous les algorithmes testés pour un CR de 2 à 8.	148
B.9	Déviations moyennes du SNDR obtenu avec tous les algorithmes testés pour un CR de 2 à 8.	148
B.10	Pourcentage des PAs compressés dans le bon <i>cluster</i> avec PCA effectuée sur les PAs compressés joint aux PAs originaux.	149
B.11	Pourcentage des PAs compressés dans le bon <i>cluster</i> avec PCA effectuée en ajoutant les PAs compressés un à la fois aux PAs originaux.	150
B.12	Nombre de <i>clusters</i> trouvés après compression.	150
B.13	Nombre de <i>clusters</i> trouvés avec les PAs compressés joints aux PAs originaux.	151
B.14	Représentation de 6 PAs reconstruits en utilisant seulement les 12 coefficients (12 sur 48) produits par les niveaux 3 et 4 de la DWT. Les formes d'ondes reconstruites (en rouge) ont conservé la forme générale des PAs originaux (en bleu).	154
B.15	Représentation de 4 PAs et de leurs 6 coefficients du 3e niveau de leur DWT respective. Deux paires de PA similaires ont généré des coefficients similaires, ce qui est généralement le cas pour des formes d'ondes visuellement proches.	155
B.16	Histogramme des résultats répertoriés dans le Tableau B.3.	157
B.17	Stratégie de regroupement non supervisée utilisant les coefficients du 3e niveau de la DWT comme vecteur d'entrée.	158
B.18	Schémas du circuit de stimulation optique.	160
B.19	Concept de l'ADC $\Delta\Sigma$ MASH du 3 ^e ordre.	161
B.20	<i>Layout</i> de l'AFE amélioré et du $\Delta\Sigma$ MASH d'ordre 3. Ce circuit possède une aire totale de 0.056 mm ²	163
B.21	Concept du $\Delta\Sigma$ MASH du 3 ^e ordre avec les différentes topologies des circuits détaillés.	164
B.22	Concept du filtre de décimation CIC d'ordre 4. Il y a un filtre par $\Delta\Sigma$ et tous les filtres sont reliés à un bus unique pour communiquer avec le module de contrôle.	166
B.23	Bande passante du filtre de décimation CIC d'ordre 4. La fréquence de coupure à -3 dB du filtre est indiquée en rouge ainsi que la fréquence de <i>Nyquist</i> .	166
B.24	<i>Layout</i> des 10 filtres de décimation. La taille du circuit est de 0.27 mm ² (0.027 mm ² /filtre).	167
B.25	Schéma fonctionnel de la machine à état du module de contrôle.	168
B.26	<i>Layout</i> du module de contrôle. La taille du circuit est de 0.0055 mm ²	168

*À ma conjointe Cynthia et à mes
enfants Élizabeth et William*

Remerciements

J'aimerais tout d'abord remercier mon superviseur de recherche, Dr Benoit Gosselin de l'Université Laval pour son support tout au long de mes études graduées. Dr Gosselin m'a donné la liberté et les outils nécessaires à la réalisation de mon projet de recherche. Il m'a épaulé, avec l'apport de nouvelles idées et défis, ce qui a contribué à repousser toujours mes limites et à acquérir de nouvelles connaissances et expériences. Je suis privilégié de l'avoir eu comme directeur de recherche.

Lors de mon séjour au laboratoire de recherche sur les microsystèmes biomédicaux (BML), j'ai eu la chance de travailler avec des collègues fantastiques. J'aimerais particulièrement remercier Esmaeel Maghsoudloo, Cheikh Latyr Fall et Mehdi Noormohammadi Khiarak, collègues et amis, qui ont apporté un apport constructif à la réalisation de mon projet de recherche. De plus, j'aimerais remercier Cyril Bories, Iason Keramidis, Christian Éthier et Yves De Koninck du centre de recherche CERVO pour m'avoir permis de tester tous mes systèmes *in vivo*.

Enfin, je suis profondément reconnaissant à ma conjointe et à mes enfants pour leurs encouragements. Ce travail n'aurait pu être accompli sans leur soutien.

Avant-propos

Les chapitres 2 à 4 de cette thèse comportent du matériel déjà publié ou en révision dans des journaux techniques avec comité de révision. De plus, un article publié à la prestigieuse conférence *International Solid-State Circuit Conference* (ISSCC) est fourni en annexe. Voici mes contributions aux 3 articles de journaux :

Article 1 :

G. Gagnon-Turcotte, Y. LeChasseur, C. Bories, Y. Messaddeq, Y. De Koninck and B. Gosselin, "A Wireless Headstage for Combined Optogenetics and Multichannel Electrophysiological Recording", *IEEE Transaction Biomedical Circuits Systems*, vol. 11, no. 1, pp. 1-14, Feb. 2017. © 2019 IEEE

Dans cet article publié en 2017, je suis le concepteur du système et des algorithmes et j'ai écrit l'article au complet conjointement à B. Gosselin qui a supervisé le travail de recherche et qui a contribué aux idées. Les auteurs Y. LeChasseur, C. Bories et Y. De Koninck ont été impliqués dans les tests *in vivo* et lors de la phase de révision de l'article, alors que Y. Messaddeq a été impliqué lors de la phase de révision seulement. À l'exception du dimensionnement des figures, il n'y a pas de différence notable entre le chapitre associé à cet article et l'article original.

Article 2 :

G. Gagnon-Turcotte, M. Noormohammadi Khiarak, C. Ethier, Y. De Koninck and B. Gosselin, "A 0.13- μm CMOS SoC for Simultaneous Multichannel Optogenetics and Neural Recording", *Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3087-3100, Nov. 2018. © 2019 IEEE

Dans cet article publié en 2018, je suis le concepteur de tous les circuits et j'ai été le seul impliqué dans la production de la puce (schémas et «layout»). J'ai écrit l'article au complet conjointement à B. Gosselin qui a supervisé le travail de recherche et qui a contribué aux idées, alors que M. Noormohammadi Khiarak m'a aidé à rédiger une sous-section seulement. Les auteurs C. Ethier et Y. De Koninck ont été impliqués dans les tests *in vivo* et lors de la phase de révision de l'article. À l'exception du dimensionnement des figures, il n'y a

pas de différence notable entre le chapitre associé à cet article et l'article original.

Article 3 :

G. Gagnon-Turcotte, I. Keramidis, C Ethier, Y. De Koninck and B. Gosselin, "A Wireless Electro-Optic Headstage with a 0.13- μm CMOS Custom Integrated DWT Neural Signal Decoder for Closed-Loop Optogenetics", *IEEE Transactions on Biomedical Circuits and Systems*, 2019 (submitted). © 2019 IEEE

Dans cet article soumis au mois de février 2019, je suis l'unique concepteur du système et de tous les circuits. De plus, j'ai été le seul impliqué dans la production de la puce (schémas et «layout») et j'ai écrit l'article au complet conjointement à B. Gosselin qui a supervisé le travail de recherche et qui a contribué aux idées. Les auteurs I. Keramidis C. Ethier et Y. De Koninck ont été impliqués dans les tests *in vivo* et lors de la phase de révision de l'article. À l'exception du dimensionnement des figures, il n'y a pas de différence notable entre le chapitre associé à cet article et l'article original.

Puisque ce document contient des articles publiés dans des journaux IEEE, voici le message de copyright IEEE pour la diffusion électronique de cette thèse :

«In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of University Laval's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.»

Acronymes

ADC Analog-to-Digital Conversion

AFE Analog Front-End

AP Action Potential

ASIC Application-Specific Integrated Circuit

BGA Ball Grid Array

BW Bandwidth

ChR2 Channelrhodopsin-2

CF Cost Function

CIC Cascaded Integrator-Comb

CL Closed-Loop

CMFB Common Mode Feedback Circuit

CMOS Complementary Metal Oxide Semiconductor

CMRR Common-Mode Rejection Ratio

CS Compressed Sensing

CR Compression Ratio

COTS Commercial Off-The-Shelf

DD Discrete Derivative

DF Decimation Filter

DWT Discrete Wavelet Transform

ENOB Effective Number Of Bits

DAC Digital-to-Analog Converter

DSP Digital Signal Processing

DT Discrete Time

ENOB Effective Number Of Bits

FIFO First-In-First-Out

FIR Finite Impulse Response
FPGA Field Programmable Gate Array
FSM Finite State Machine
FRQNT Fonds de Recherche du Québec - Nature et Technologies
IC Integrated Circuit
IIR Infinite Impulse Response
IRN Input-Referred Noise
ISM Industrial, Scientific and Medical
LDO Low-Dropout Regulator
LED Light Emitting Diode
LFP Local Field Potential
LiPo Lithium-ion Polymer
LSB Least Significant Bit
MAD Median of the Absolute Deviation
MASH Multi-Stage Noise Shaping
MCU Micro Controller Unit
MF Matched-Filter
MIM Metal Insulator Metal
MSB Most Significant Bit
MTEO Multiresolution TEO
NEF Noise Efficiency Factor
NEO Non-linear Energy Operator
ND-IC Neural Decoder IC
NMOS N-channel MOSFET
NSERC Natural Sciences and Engineering Research Council of Canada
OTA Operational Transconductance Amplifier
OSPM Optical Stimulation Patterns Management
OSR Over Sampling Ratio
PA Potentiel d'Action
PCA Principal Component Analysis
PCB Printed Circuit Board
PI Proportional Integral

PID Proportional Integral Derivative
PMOS P-channel MOSFET
PMU Power Management Unit
PSRR Power Supply Rejection Ratio
PWM Pulse-Width Modulation
RAM Random Access Memory
RMS Root Mean Square
ROC Receiver Operating Characteristics
SAR Successive Approximation Register
SC Switched-Capacitor
SNDR Signal-To-Noise Distortion Ratio
SNR Signal-To-Noise Ratio
SoC System-On-Chip
SPI Serial Peripheral Interface
SRAM Static Random Access Memory
TEO Teager's Operator
THD Total Harmonic Distortion
VHDL VHSIC Hardware Description Language
VPM Ventral Posteromedial Nucleus

Introduction

L'avenir de la recherche sur les maladies du cerveau repose sur le développement de technologies innovantes qui permettront de comprendre comment le cerveau traite, intègre et transfère l'information. Par le passé, les cultures cellulaires et les préparations de tissus *ex vivo* ont permis de réaliser des avancées spectaculaires, mais il est clair que pour faire avancer davantage les connaissances, l'observation *in vivo* de cerveaux intacts sera nécessaire [1]. Par conséquent, les technologies qui feront la différence seront celles qui pourront faire le pont entre l'étude locale des microcircuits du cerveau à l'échelle cellulaire et le comportement.

Cela fait écho au *BRAIN Initiative* [2], qui soutient le développement et l'application de technologies novatrices permettant une compréhension dynamique des fonctions cérébrales. De même, le *Human Brain Project* et l'*Institut Allen* [3] représentent des efforts ambitieux afin de construire de vastes bibliothèques de données neuronales. Au Canada depuis 2011, le Fonds Canadien de Recherche sur le Cerveau [4] a accordé plus de 206 millions de dollars en fonds publics et privés à 224 projets à travers le pays dédiés à la recherche sur le cerveau. Or, une partie de ce financement est voué au développement de nouveaux outils de recherche en neurosciences. À l'Université Laval seulement, le Centre de neurophotonique CERVO a récemment bénéficié d'une importante subvention, soit plus de 60 millions de dollars, provenant de la Fondation canadienne pour l'innovation, du gouvernement du Québec, et de plusieurs autres partenaires, afin de fournir aux scientifiques canadiens les outils nécessaires pour résoudre les énigmes du cerveau. Les enjeux sont tels, que la recherche sur le cerveau jouit d'un financement sans précédent, et ce, partout dans le monde.

Dans les centres de recherches en neurosciences, l'étude du cerveau repose typiquement sur l'observation de l'activité neuronale chez les souris de laboratoires. Entre autres, celles-ci permettent de modéliser le cerveau humain, en plus d'exprimer une grande variété de gènes et de maladies, et elles sont peu coûteuses. Pourtant, la majorité des systèmes devant être utilisés avec ces petits rongeurs sont câblés, sont trop lourds et/ou trop simplistes. Dans ce contexte, le terme simpliste implique que le système déclenche une stimulation à chaque détection d'évènement (PA ou bruit) ou que l'interprétation des données neuronales est effectuée *ex situ*. Peu de systèmes sans fil existent, et ces derniers sont grandement limités par la bande passante requise pour transmettre les données neuronales. Dans les dispositifs

actuels, l'interprétation des données neuronales est effectuée *ex situ*, alors que la recherche bénéficierait grandement de systèmes légers, sans fil, et assez intelligents pour interpréter et stimuler les neurones en boucle fermée, *in situ*. Ce genre de système ouvrirait de nouvelles possibilités, tel que contourner des régions déficientes du cerveau via des liens neuronaux artificiels, vérifier l'effet d'un médicament, étudier des maladies neurodégénératives, etc.

C'est dans cette optique que cette thèse présente de nouveaux algorithmes, de nouvelles puces, de nouveaux circuits et de nouveaux systèmes sans fil qui permettent de :

- i) Mener des expériences combinant l'enregistrement neuronal et l'optogénétique multicanal haute résolution avec des animaux libres de leurs mouvements.
- ii) Mener des expériences optogénétiques synchronisées avec l'observation, c.-à-d. en boucle fermée, chez des animaux libres de leurs mouvements.
- iii) Augmenter la résolution et réduire la taille, le poids et la consommation énergétique des systèmes optogénétiques sans fil afin de minimiser l'impact de la recherche chez de petits animaux.

En résumé, notons que cette thèse présente : 1) le premier système optogénétique sans fil à être doté de stimulation optogénétique multicanal et d'enregistrement simultané haute résolution, 2) le premier système sur puce (SoC) CMOS permettant de faire de l'optogénétique multicanal et de l'enregistrement neuronal multicanal avec une seule puce, et 3) le premier système optogénétique multicanal qui est capable de fonctionner en boucle fermée et le premier à être validé lors d'expériences *in vivo* impliquant des animaux libres de leurs mouvements. Les nouvelles expériences qui pourront être menées grâce à ces avancées vont de la recherche fondamentale portant sur les communications inter neurones, à l'étude des maladies neurodégénératives en passant par la recherche de pointe en neurosciences comportementales.

Afin de clairement identifier le problème étudié et les motivations sous-jacentes à la rédaction de cette thèse, le contexte de réalisation de la recherche, les problématiques, la description du projet et les contributions apportées par ce travail sont abordés en introduction.

Contexte de réalisation de la recherche

Étude sur le cerveau

Cette sous-section est inspirée du mémoire de maîtrise [5] de l'auteur de cette thèse.

Chez les animaux, le cerveau est l'organe du système nerveux qui régule et qui contrôle l'organisme en plus de constituer le siège des fonctions cognitives menant à des processus de prise de décisions complexes. Le cerveau n'est pas un organe comme les autres, puisqu'il est constitué d'une structure extrêmement élaborée basée sur l'activité simultanée de mil-

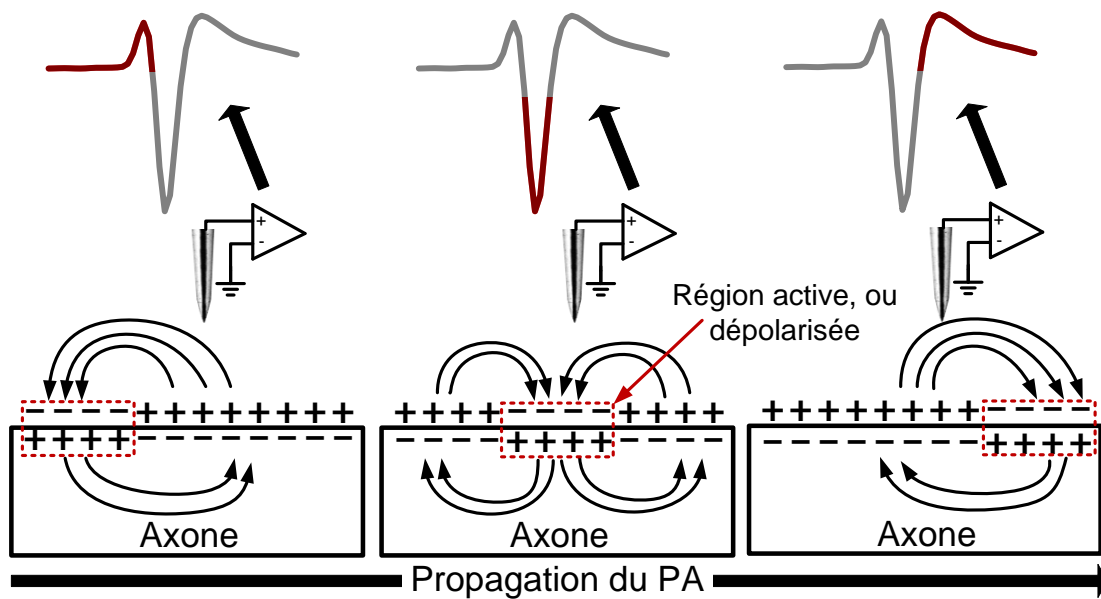


FIGURE 1.1 – Modèle d'un PA enregistré avec une électrode extracellulaire. La région dépolarisée se déplace d'un bout à l'autre de l'axone, ce qui génère une tension positive, négative et ensuite positive vue par l'électrode, c.-à-d. un PA triphasique.

liards de neurones. En fait, le cerveau est la structure biologique la plus complexe connue [6] et dont l'architecture possède des caractéristiques qui sont communes à un grand nombre d'espèces [7]. Le cerveau est divisé en sous-sections qui sont constituées de groupes de neurones dédiés à la réalisation de tâches spécifiques, conscientes ou inconscientes, comme faire battre le cœur et veiller à la motricité des membres du corps. Afin de fonctionner correctement, ces groupes de neurones communiquent entre eux, individuellement ou en groupes, en s'envoyant des signaux chimiques ou électriques dont l'information est encodée de manière temporelle et spatiale, et dont le fonctionnement n'est pas encore totalement compris [8]. Les mystères régissant les diverses fonctionnalités du cerveau sont nombreux et très complexes. Afin de mettre le lecteur en contexte, cette sous-section fait une brève revue des types de signaux neuronaux, du domaine et des méthodes d'enregistrement électrophysiologique en général ainsi que des systèmes d'acquisition de données traditionnels.

Signaux neuronaux

Deux types de signaux générés par les neurones et/ou par les groupes de neurones [11] seront abordés dans ce document. Premièrement, il y a les potentiels d'action (PA) qui sont générés par un neurone unique et qui sont issus de la dépolarisation de la membrane cellulaire. La dépolarisation de la membrane cellulaire s'effectue en suivant les phases suivantes :

- **Phase de repos** : Canaux sodiques et potassiques voltage-dépendants à sodium (Na^+) et à potassium (K^+) fermés. Potentiel membranaire stable.

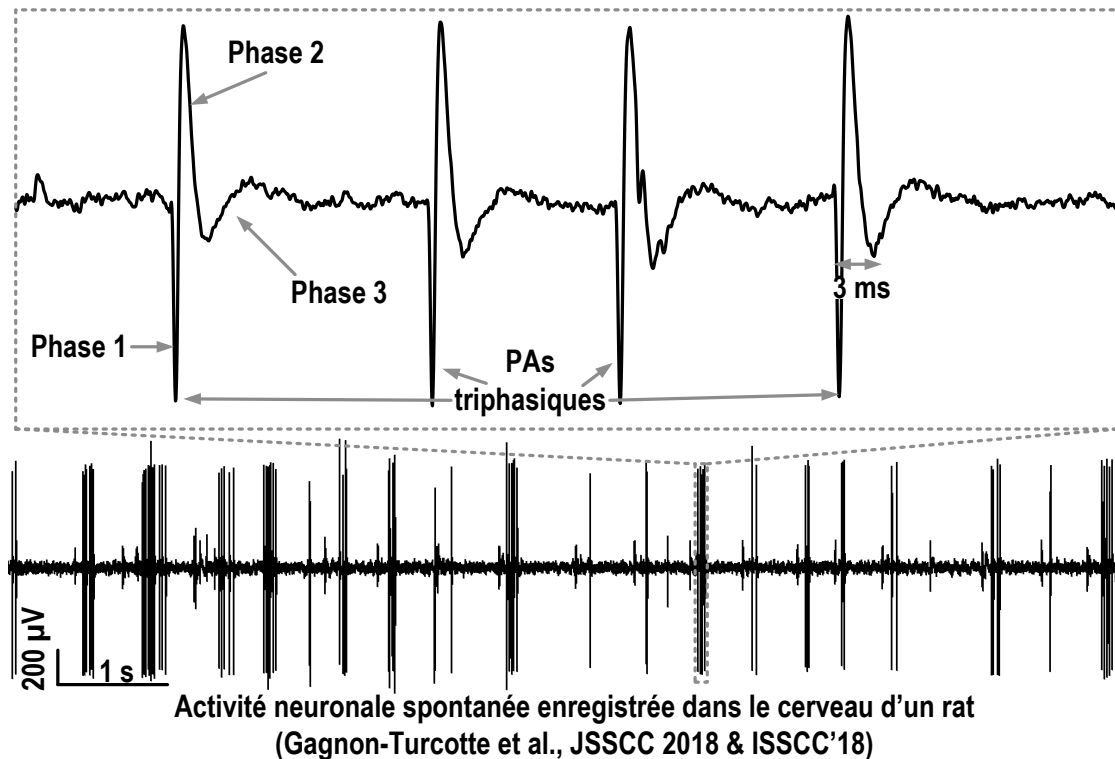


FIGURE 1.2 – Forme d’onde d’un train de PA triphasique extracellulaire comportant trois phases (avec polarités inversées), enregistré dans le cerveau d’un rat par une des puces présentées dans cette thèse [9, 10].

- **Phase de dépolarisation** : Canaux sodiques voltage dépendants activées, pénétration des ions Na^+ dans la cellule et ouverture des canaux potassiques voltage-dépendants à potassium après un court délai. Le débit entrant des ions Na^+ est supérieur au débit de sortie des ions K^+ ce qui génère une différence de potentiel entre le milieu intra et extracellulaire.
- **Phase de repolarisation** : Les pompes à sodium se referment et les pompes à potassium expulsent un maximum d’ions K^+ , ce qui tend à repolariser la membrane cellulaire.
- **Dépassement** : Les pompes à potassium expulsent un surplus d’ions K^+ , le potentiel de la membrane cellulaire se stabilise progressivement.

Différentes formes d’onde des PA peuvent être captés selon si l’enregistrement est intracellulaire ou extracellulaire. Lors de l’utilisation d’une microélectrode extracellulaire, les courants induits dans l’espace extracellulaire autour d’un neurone actif produisent la forme d’onde captée du PA, qui est différente de celle enregistrée directement dans la cellule. En effet, il est possible de modéliser un seul axone d’un neurone dans un volume conducteur à faible impédance (Ex : saline) [12]. Selon ce modèle, lorsque l’axone du neurone est au repos, le potentiel de la membrane est uniforme sur toute sa longueur (membrane polarisée). La dépolarisation de l’axone induite par la propagation du PA intracellulaire va créer une différence de potentiel le long de la membrane, ce qui va générer des courants allant des régions polarisées

(autour de la région dépolarisée) vers les régions dépolarisées. Puisque les courants circulent vers l'intérieur de la cellule au niveau de la région dépolarisée, une électrode proche de cette région verra une tension négative. Au contraire, si l'électrode est proche d'une région polarisée qui fournit un courant, cette dernière verra une tension positive. Le PA se propage tout au long de l'axone, ce implique que la région de dépolarisation se déplace progressivement le long de la membrane axonale [13]. Par exemple, si un axone se dépolarise d'un bout à l'autre et que l'électrode est placée via-à-vis le milieu de la membrane, l'électrode verra (dans l'ordre) [12] :

1. Une tension de repos (membrane entièrement polarisée, aucun courant).
2. La phase 1, soit une tension positive (vis-à-vis une région polarisée qui fournit un courant).
3. La phase 2, soit une tension négative (vis-à-vis une région dépolarisée qui reçoit un courant).
4. La phase 3, soit une tension positive (vis-à-vis une région polarisée qui fournit un courant).
5. Une tension de repos (membrane entièrement polarisée, aucun courant).

Ainsi, un PA enregistré par une électrode adjacente à un neurone isolé sera triphasique (tension positive-négative-positive). Les étapes de génération de la forme d'onde du PA enregistré avec une électrode extracellulaire sont résumées à la Figure 1.1.

La figure 1.2 présente la forme typique d'un train de PA extracellulaire avec polarité inversée enregistré dans le cerveau d'un rat par une des puces présentée dans cette thèse [9, 10]. On peut y voir les trois phases inversées d'un PA triphasique extracellulaire ainsi que sa durée. Le PA qui est généré par la dépolarisation de la membrane cellulaire se propage le long de l'axone du neurone jusqu'à ce que le signal soit transmis via des neurotransmetteurs vers un neurone voisin. L'amplitude intracellulaire crête-à-crête d'un PA est de ~ 100 mV alors que son amplitude extracellulaire crête-à-crête peut varier entre 50 μ V et quelques millivolts [14], selon le type de cellule, le type l'électrode et sa position. La durée d'un PA est de quelques millisecondes et ce type de signal possède un contenu fréquentiel variant entre 300 Hz et 5-10 kHz [14].

En plus des PA, les neurones génèrent des signaux appelés potentiels de champ locaux —ou *Local Field Potentials* (LFP), qui sont constitués de la somme des courants circulant dans les neurones environnants. La figure 1.3 présente la forme typique d'un signal LFP enregistré dans le cerveau d'un rat par une des puces présentée dans cette thèse [9] (abordé au chapitre 3). Ce type de signal possède une amplitude de quelques millivolts et un contenu fréquentiel variant entre 0.5 Hz et 200 Hz [15]. À moins d'indications contraires, l'expression «signaux neuronaux» réfère implicitement aux signaux contenant les PA dans cette thèse.

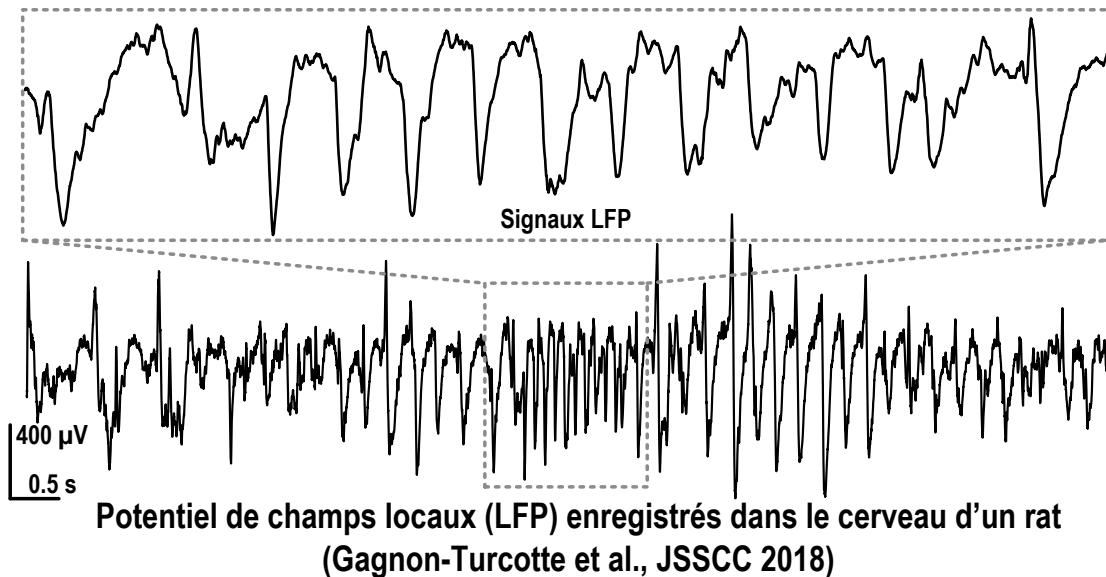


FIGURE 1.3 – Forme d'onde forme typique d'un signal LFP enregistré dans le cerveau d'un rat par une des puces présentée dans cette thèse [9]

Enregistrement électrophysiologique

L'enregistrement électrophysiologique est réalisé à partir de microélectrodes de hautes impédances ($100 \text{ k}\Omega$ - $10 \text{ M}\Omega$) implantées dans le cerveau et interfacées à l'aide d'un circuit d'amplification faible bruit [9, 10, 16–20]. Parmi les types de matériaux à haute impédance constituant les microélectrodes, le verre, le Tungsten, l'acier inoxydable, des alliages à base de Platiniun et d'Iridium ainsi que le Silicium sont utilisés. Ces électrodes peuvent être utilisées selon différentes topologies, variant selon le type d'expérience à mener, le dispositif d'enregistrement utilisé et le matériel constituant l'électrode (rigide ou flexible). Parmi les typologies existantes, notons les microélectrodes métalliques à pénétration simple [21], les microélectrodes de surface flexible [22], les micropipettes [9, 10, 23], les microélectrodes de type *linear microelectrode array* [24] ainsi que les *microélectrodes array* [25], [26].

Selon le type de microélectrode, deux méthodes d'enregistrement des PA existent, soit l'enregistrement intracellulaire et extracellulaire. L'enregistrement intracellulaire implique l'insertion de la microélectrode à l'intérieur de la cellule (typiquement une micropipette de verre pour sa haute impédance) et permet d'enregistrer le voltage au repos de la membrane cellulaire ainsi que les PA générés au niveau du segment initial de l'axone et qui reflètent une inversion de potentiel de part et d'autre de la membrane [11, 23]. Ce type d'enregistrement est plus difficile à réaliser sur de petits neurones et endommage malheureusement la cellule. L'enregistrement extracellulaire est quant à lui plus approprié pour mesurer les PA à partir de l'extérieur de la cellule et est compatible avec une plus grande variété d'électrodes. De plus, ce type d'enregistrement est très approprié pour effectuer des expériences utilisant des

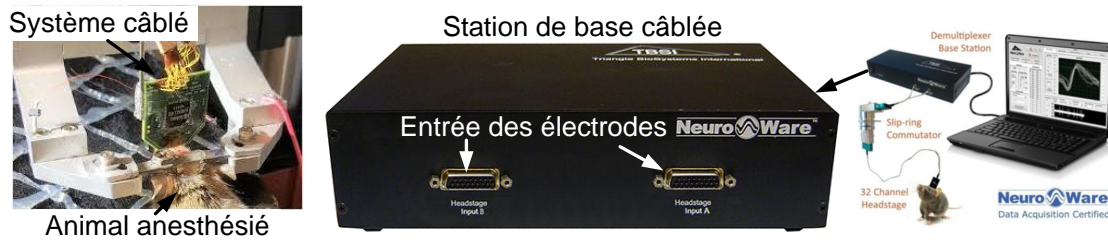


FIGURE 1.4 – Système d’enregistrement de signaux électrophysiologiques câblé, tirée de [27] à gauche et de [28] à droite.

sujects vivants, puisque la cellule en observation n’est pas endommagée par l’électrode.

Systèmes d’acquisition de données neuronales traditionnels

L’acquisition extracellulaire des signaux neuronaux est effectuée à l’aide de microélectrodes interfacées à l’aide d’un circuit électrique, soit un circuit constitué de plusieurs étages de gains à faible bruit et de filtres analogiques [9, 10, 17–20]. Par la suite, les signaux conditionnés, c.-à.-d pré-amplifiés, filtrés et transmis par câble à un amplificateur, sont convertis en signal digital avec un convertisseur analogique à numérique (ADC) pour être traité dans le domaine numérique par un processeur, soit un ordinateur, soit un microcontrôleur (MCU), un circuit de traitement de signal spécialisé (DSP) ou bien un *field programmable gate array* (FPGA). La plupart des systèmes d’acquisitions traditionnels sont câblés, c’est-à-dire que les microélectrodes sont reliées du cerveau jusqu’à une station de base par des câbles, ce qui pose énormément de problèmes, de défis et de limitations lors des expérimentations. La figure 1.4 présente ce cas de figure.

Dans le cas des dispositifs d’enregistrement sans fil, le conditionnement et la quantification des signaux sont effectués *in situ* dans le dispositif installé sur l’animal, souvent sur la tête, d’où l’appellation *headstage*, et le processeur est interfacé avec un transmetteur utilisé pour recevoir des commandes et pour transmettre les signaux neuronaux numérisés. La figure 1.5 résume sous forme conceptuelle et de schéma blocs une chaîne d’acquisition sans fil de signaux neuronaux. Indépendamment du type de système, sans fil ou câblé, la chaîne de traitement se poursuit du côté de la station de base (ordinateur) où les PA peuvent être extraits du signal original afin qu’ils soient classifiés et associés à un neurone particulier, ce sont les étapes classiques d’analyse en neurosciences comportementale [29]. Une fois ces étapes effectuées, une analyse plus poussée peut être faite selon le type d’application ou de recherche [30].

Recherche optogénétique

Cette sous-section est inspirée du mémoire de maîtrise [5] de l’auteur de cette thèse.

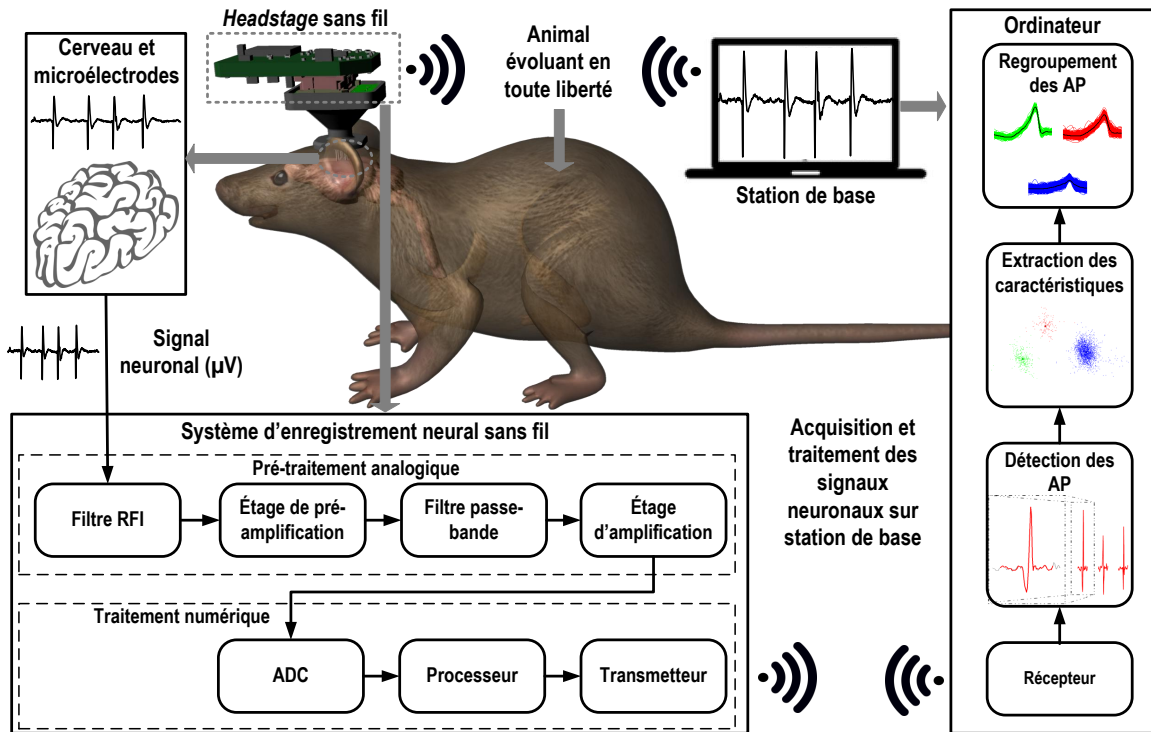


FIGURE 1.5 – Concept d’une chaîne d’acquisition sans fil de signaux neuronaux.

L’optogénétique est une technique révolutionnaire de stimulation neuronale qui permet d’activer ou de désactiver des neurones incorporant une protéine sensible à la lumière avec rapidité et grande précision temporelle et spatiale [31], tel que montré à la Figure 1.6. Ce nouveau domaine de recherche, alliant optique et génétique, a été sacrée *Méthode de Recherche de l’Année* en 2010 par *Nature Methods*, car elle permet de mener des expériences inédites sur le cerveau par l’activation ou la désactivation de groupes de neurones génétiquement modifiés ciblés par de la lumière. Contrairement à l’approche classique de stimulation électrique des neurones, qui consiste à placer une microélectrode de stimulation à proximité des tissus à stimuler avec un courant électrique diffus, la stimulation optique est robuste aux interférences de l’environnement et permet de cibler la zone à stimuler avec une plus grande précision. De plus, la stimulation optique ne produit pas d’artéfacts détectables (à l’exception de l’effet photoélectrique dans certains cas [32, 33]) et souvent dommageables pour l’équipement lors de l’enregistrement de l’activité neurale en même temps que la stimulation [34].

L’optogénétique est possible grâce à l’insertion de protéines sensibles à la lumière dans la membrane cellulaire des neurones [31]. Les protéines permettant de stimuler un neurone, comme ChR1, ChR2, ChETA, SFO, VChR1 ou VChR2, dépolarisent la cellule en laissant entrer des ions Na^+ (phase de dépolarisation) lorsqu’elles sont exposées à une longueur d’onde adéquate. Les protéines permettant d’inhiber un neurone, comme NpHR, Arch ou GtR3, hyperpolarisent la cellule en laissant entrer des ions Cl^- (NpHR) ou H^+ (Arch et GtR3), ce qui

les rend inexcitables, lorsqu'elles sont exposées à une bonne longueur d'onde. Par exemple, les protéines ChR2, ChETA et SFO sont dépolarisées par la lumière bleue (470 nm) tandis que les protéines Arch et GtR3 sont hyperpolarisées avec la lumière jaune (575 nm) et bleue (472 nm) respectivement. Différentes techniques d'intégration de gènes existent afin que les neurones deviennent sensibles à la lumière. Entre autres, il est possible d'intégrer un virus exprimant les gènes produisant la bonne protéine à l'ADN chromosomique du sujet, ou bien d'utiliser des animaux transgéniques exprimant les gènes désirés [35]. Notamment, il est possible de se procurer ce type d'animaux transgéniques de fournisseurs privés, tels que *The Jackson Laboratory* [23, 36, 37], ou bien dans des animaleries de laboratoire de recherche en neurosciences. Plusieurs laboratoires de recherche en neurosciences utilisent l'optogénétique afin de mener des expériences sur le cerveau d'animaux transgéniques. Parmi ceux-ci, le laboratoire de recherche du Dr Yves De Koninck au centre CERVO de Québec [38] est connu pour ses travaux de recherche portant sur les causes et les traitements reliés aux maladies neurologiques, et celui du Dr Christian Éthier au CERVO qui vise la conception d'interfaces cerveau-ordinateur afin de comprendre les processus menant à la formation de nouvelles connexions neuronales [39], sans oublier celui du Dr Frédéric Bretzner spécialisé dans le contrôle et plasticité des circuits moteurs au CHU de Québec [40].

Une des limitations de l'optogénétique concerne la génération d'artéfacts photoélectriques [41] lorsque combiné avec l'enregistrement électrophysiologique. L'effet photoélectrique découle de la génération d'un courant ou d'une tension électrique dans un matériau exposé à de la lumière. Malheureusement, l'effet photoélectrique peut générer des signaux qui s'apparentent à de l'activité neuronale. Différentes stratégies sont proposées dans la littérature pour atténuer cet effet, comme l'utilisation d'une source de lumière non corrélée, comme une LED par exemple, mettre la source lumineuse loin des électrodes, ou encore utiliser un conducteur ionique, telle une micropipette de verre. En effet, les conducteurs ioniques ne sont pas autant affectés par les artéfacts de stimulation que les électrodes de métal [42]. À cause de ce phénomène, un des aspects à améliorer dans le futur sera la création d'optrodes moins sensible à l'effet photoélectrique que ce qui existe actuellement, en utilisant des électrodes en graphènes par exemple [43].

Stimulation optogénétique versus stimulation électrique

La stimulation électrique dans le cerveau consiste à injecter des charges, c.-à-d. des courants, dans les tissus neuronaux afin d'exciter les neurones et déclencher la création de PAs artificiellement. Plusieurs méthodes de stimulation existent (courants, tension, etc.), mais la technique la plus couramment utilisée est basée sur l'injection de courants constants, ce qui permet notamment de contrôler la quantité de charges injectés dans le cerveau. Cette technique est généralement constituée de deux phases : une phase cathodique durant laquelle les potentiels d'action sont déclenchés par l'injection de courants, suivie d'une phase anodique

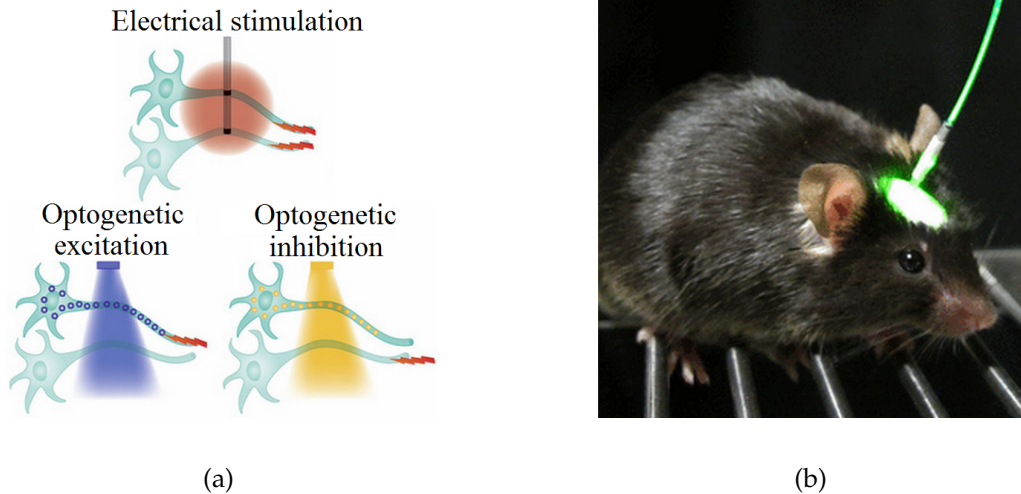


FIGURE 1.6 – a) Stimulation électrique versus stimulation/inhibition optique des neurones (image tirée de [44]), b) souris transgénique en pleine expérience optogénétique (image tirée de [44]).

qui permet de retirer les charges accumulées par les électrodes lors de la phase cathodique [45]. Or, l'injection de courants dans un volume de tissus est très diffuse, ce qui a le défaut évident d'activer toute une zone et de ne pas être très sélectif. Pour ainsi dire, la stimulation électrique ne permet pas de cibler avec précision un groupe de neurone, ou un type de neurone spécifique. De plus, si les charges injectées dans le cerveau ne sont pas bien balancées lors des phases anodiques et cathodiques, il peut se créer de l'oxydation sur les électrodes métalliques ce qui peut endommager les tissus cérébraux.

Un des avantages principaux de l'optogénétique versus la stimulation électrique est qu'il est possible de cibler certains types neurones de manière très sélective. En effet, il est possible avec l'utilisation d'un vecteur viral et d'un promoteur approprié de faire en sorte que seulement certains types de neurones fabriquent les protéines sensibles à la lumière. De cette façon, les chercheurs peuvent cibler le type des neurones qu'ils veulent stimuler en fonction de leur expérience, ce qui n'est pas possible avec la stimulation électrique. De plus, grâce à des techniques nouvellement développées [23], il est possible avec l'optogénétique de stimuler quelques cellules à la fois, voir même une cellule unique, ce qui n'est pas possible avec la stimulation électrique à cause de la diffusion des courants. Enfin, les matériaux optiques ne sont pas sujets à l'oxydation, comparativement aux électrodes utilisées lors de la stimulation électrique.

Fonctionnement en boucle fermée

Une nouvelle technique qui possède le potentiel de révolutionner la recherche sur le cerveau consiste à fermer la boucle entre le stimulateur optique et le circuit d'acquisition neuro-

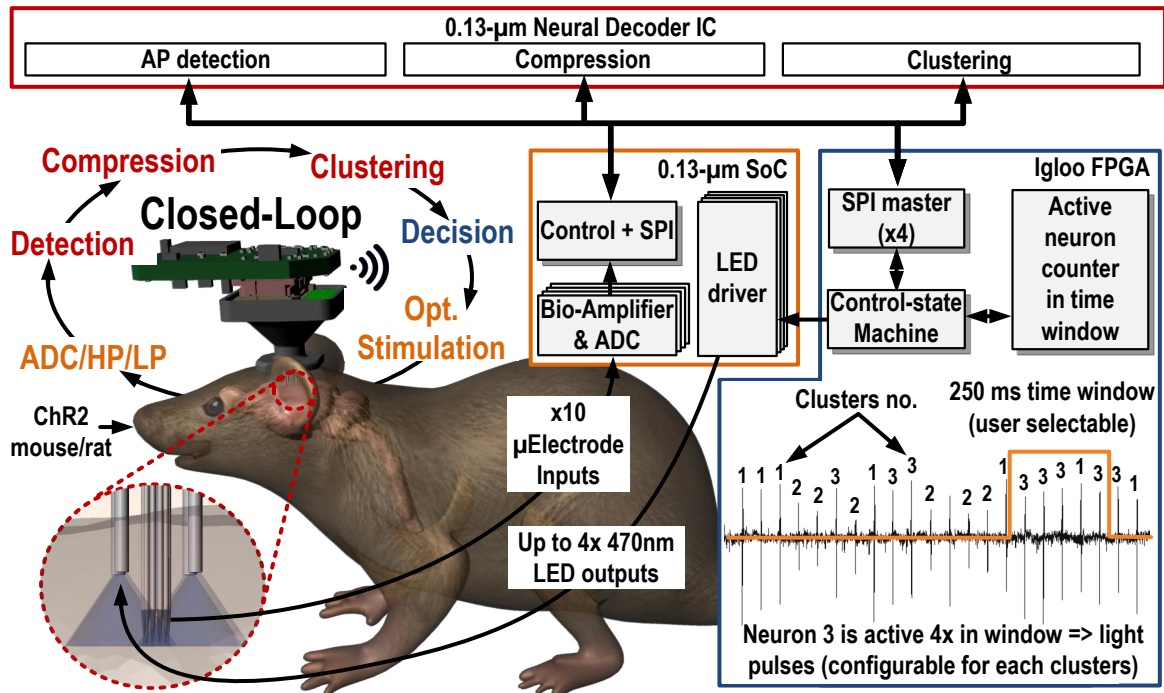


FIGURE 1.7 – Concept de fonctionnement en boucle fermée qui sera abordé au Chapitre 4 de cette thèse. Un lien entre le stimulateur et le circuit d’enregistrement est établi en fonction du classement des PA et en fonction de l’activité «burst» des neurones.

nal [46–48]. Cette nouvelle technique, communément appelée stimulation optogénétique en boucle fermée ou stimulation optique synchronisée avec l’enregistrement neuronal, consiste à stimuler certaines régions du cerveau en fonction de l’activité de certains neurones ou groupes de neurones. Dans ce document l’expression optogénétique en boucle fermée fait explicitement référence à la stimulation optique induite par l’enregistrement neuronal, telle que décrite dans la littérature [48], en contraste avec la définition classique consistant à contrôler un procédé avec un contrôleur en boucle fermée. Entre autres, cette technique permet de recruter un réseau neuronal à court terme entre différentes régions du cerveau. À long terme, une plasticité structurale peut apparaître, entraînant une réorganisation des circuits et créer de nouvelles connexions physiques [27, 49, 50]. Cette approche permettra de faire avancer l’étude et la compréhension du cerveau, ainsi que de plusieurs maladies neurologiques, dont les maladies d’Alzheimer et de Parkinson, l’épilepsie, les douleurs chroniques, etc., soit des maladies de plus en plus présentes avec le vieillissement de la population.

Un problème de la stimulation en boucle fermée consiste à analyser en temps réel une quantité importante de données pour générer les pulses de stimulation. Pour cette raison, la plupart des systèmes qui permettent de synchroniser la stimulation avec l’enregistrement sont câblés et limités à être utilisés la plupart du temps avec des animaux anesthésiés (des expériences câblées avec des rats, des primates et des humains non anesthésiés ont été rapportées

[51, 52]). Les systèmes les plus complexes vont détecter la présence de PA, les classer en fonction de leur forme d'onde, et vont ensuite stimuler en fonction de ce classement [50]. Puisque les neurones sont connus pour générer des PA de forme similaire dans le temps [53], cette technique permet d'estimer combien de neurones sont près des sites d'enregistrement et quand ces neurones sont actifs. De ce fait, les utilisateurs peuvent sélectionner quel neurone va déclencher la stimulation, ou non. Une autre méthode consiste à compter le nombre de PA dans une fenêtre temporelle, afin de détecter la présence de bouffées de PAs ou «burst» [49]. Or, les chercheurs sont persuadés que la présence de «burst» est primordiale dans la communication entre les neurones, et donc, importante afin de recruter un réseau neuronal [54–56]. Évidemment, la méthode la plus simple consisterait à stimuler pour chaque PA détecté [27, 57, 58]. Bien que simple, cette technique est très sensible au bruit et aux artéfacts de mouvements, plus spécialement pour une utilisation chez des animaux éveillés, et elle ne permet pas de discriminer des patrons complexes d'activité neuronale.

Enfin, il est possible d'analyser les signaux de type LFP, au lieu des PA, pour synchroniser la stimulation avec l'enregistrement neuronal. Des régulateurs de type proportionnel intégral (PI) ou proportionnel intégral dérivé (PID) sont parfois utilisés pour contrôler la phase et l'amplitude des signaux LFP [59–61]. Ces systèmes ont des applications dans la maladie de Parkinson, l'Épilepsie ou lors de traumatismes médullaires, en utilisant la stimulation électrique [52, 59–61], mais sont moins appropriés afin de recruter précisément un circuit neuronal avec l'optogénétique. La Figure 1.7 présente le concept de fonctionnement en boucle fermée qui sera abordé au Chapitre 4 de cette thèse. Dans la technique proposée, un lien entre le stimulateur et le circuit d'enregistrement est établi en fonction du classement des PA et en fonction de l'activité «burst» des neurones, combinant les deux techniques clés présentés dans [50] et [49] à l'intérieur d'un système sans fil.

Problématique

La majorité des appareils de mesures disponibles ne sont pas appropriés, car ils utilisent des câbles (mouvements de l'animal limités), ils sont trop lourds, et/ou ils offrent peu de fonctionnalités, ce qui réduit les possibilités d'expériences clés, comme l'étude de l'activité cérébrale de petits animaux se comportant librement [9, 10, 20, 62]. Des systèmes sans fil existent, mais ces derniers sont grandement limités par la bande passante requise pour transmettre les données neuronales [62–65] et ils ne fournissent pas de stimulation optogénétique multicanal en parallèle à l'enregistrement, une fonction pourtant essentielle afin de stimuler et observer simultanément plusieurs régions du cerveau. La plupart des systèmes commerciaux utilisent des composantes commerciales et/ou des transmetteurs énergivores qui consomment beaucoup d'énergie et qui ne sont pas optimisés pour mener des études à long terme ni pour des installations permanentes sur l'animal [58, 63–65].

À ce jour, aucun dispositif commercial sans fil dédié à l'optogénétique n'est capable de synchroniser l'enregistrement avec la stimulation, c.-à-d. à dire fonctionner en boucle fermée, et capable de transmettre simultanément les données neuronales. Dans la majorité des expérimentations, le traitement et l'interprétation des données neuronales est effectué hors-ligne, dans un ordinateur, alors que la recherche bénéficierait de systèmes miniatures et sans-fils capables d'observer et de stimuler les neurones en temps réel (latence minimale) et sans passer par une station de base [50], [39].

Un système sans fil idéal serait capable de détecter *in situ* la présence de «bursts» [54–56], ou du moins capable de détecter quand certains neurones spécifiques sont actifs [50], et stimuler en rétroaction à cette information. Un des défis consiste donc à isoler les «bursts», ou l'activité de neurones spécifiques, parmi les signaux détectés qui peuvent être du bruit (neuronal, thermique, etc.), ou bien des PA provenant de plusieurs neurones simultanément. Or, les neurones sont connus pour générer des PA de forme similaire dans le temps [53]. Conséquemment, classer les PA détectés permettrait de déterminer de quel neurone ils proviennent, alors que suivre leur occurrence temporelle et leur classement permettrait de détecter la présence de «bursts» générés par certains neurones prédéterminés. Combiner ces deux fonctionnalités permettrait de déclencher une stimulation seulement si certains patrons neuronaux sont présents. Malheureusement, pour des raisons de complexité, de poids et de consommation énergétique, aucun système sans fil actuel ne possède ces fonctionnalités.

Certains prototypes récents sont capables de fonctionner en boucle fermée, mais ces derniers ne sont pas sans fil [27, 58], ne font pas d'optogénétique multicanal [27, 50, 57–61], ne transmettent pas les données neuronales simultanément [27, 50, 57, 58], ne visent pas les signaux au niveau cellulaire [59–61] et sont beaucoup trop simplistes [27, 57, 58], pour mener des expériences pertinentes.

Projet

Le but de ce projet de recherche est de concevoir des circuits analogiques d'acquisition et de conditionnement des signaux neuronaux ainsi que des circuits numériques de traitement des signaux neuronaux numérisés, des algorithmes d'analyse et de traitement de ces signaux et des systèmes electro-optiques miniatures pour mener des expériences optogénétique en boucle fermée avec de petits animaux. Ce projet de recherche a permis de (non-exhaustif) :

1. Développer des algorithmes de traitement des signaux produits par les neurones permettant de réduire la quantité de données à transmettre via un lien sans fil, effectuer une analyse en temps-réel des communications inter-neurones, estimer le nombre de neurones près de chacun des sites d'enregistrement et déterminer quand ces neurones sont actifs.
2. Intégrer les algorithmes susmentionnés dans une puce CMOS afin de réduire la consom-

mation énergétique lié au traitement des données et la taille du système.

3. Développer une interface d'acquisition CMOS alliant la stimulation optogénétique multicanal et l'enregistrement électrophysiologique multicanal, afin d'optimiser les performances pour l'application visée et de réduire la taille et la consommation énergétique.
4. Concevoir une stratégie de stimulation en boucle fermée robuste qui utilise les algorithmes et les puces susmentionnés.
5. Concevoir des plateformes miniatures et sans fil alliant l'optogénétique multicanal et l'enregistrement électrophysiologique multicanal haute résolution utilisant les nouvelles puces CMOS et capable de faire de l'optogénétique en boucle fermée.
6. Valider les algorithmes, les puces et les plateformes sans fil lors d'essais *in vivo* impliquant des animaux libres de leurs mouvements avec nos partenaires du centre CERVO de Québec.

Synthèse et chronologie

La figure 1.8 présente la chronologie des différents systèmes, puces et articles générés depuis la genèse du projet en 2014. À la fin de sa maîtrise (en 2015), l'auteur de cette thèse a publié dans la revue *Sensors* un système optogénétique sans fil conçu avec des composantes commerciales qui a été élaboré en collaboration avec la compagnie *Doric Lenses*, Québec, Canada :

- **G. Gagnon-Turcotte**, A. Avakh Kisomi, R. Ameli, C.O Dufresne-Camaro, Y. LeChasseur, J.L Neron, P. B. Bareil, P. Fortier, C. Bories, Y. De Koninck and B. Gosselin, "A Wireless Optogenetic Stimulator Headstage with Multichannel Neural Recording Capability", *Sensors*, vol. 15, no. 9, pp. 22776-22797, 2015.

Cet article n'est pas couvert dans ce document, mais est disponible en «open access» à l'adresse suivante : <https://www.ncbi.nlm.nih.gov/pubmed/26371006>. Bien que ce système soit innovant à bien des égards et validé *in vivo* avec une souris transgénique anesthésiée, il est malheureusement trop lourd et volumineux pour être utilisé avec des souris de laboratoire éveillées et sa résolution est très limitée.

Cet état de fait a motivé la création d'une nouvelle plateforme plus petite et la conception d'algorithme de compression afin d'augmenter significativement la résolution du système. Ce système et les algorithmes ont été publiés lors de la seconde année de doctorat du rédacteur de cette thèse (travaux débutés à la fin de la maîtrise) :

- Chapitre 2 : **G. Gagnon-Turcotte**, Y. LeChasseur, C. Bories, Y. Messaddeq, Y. De Koninck and B. Gosselin, "A Wireless Headstage for Combined Optogenetics and Multichannel Electrophysiological Recording", *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 1, pp. 1-4, 2017. © 2019 IEEE

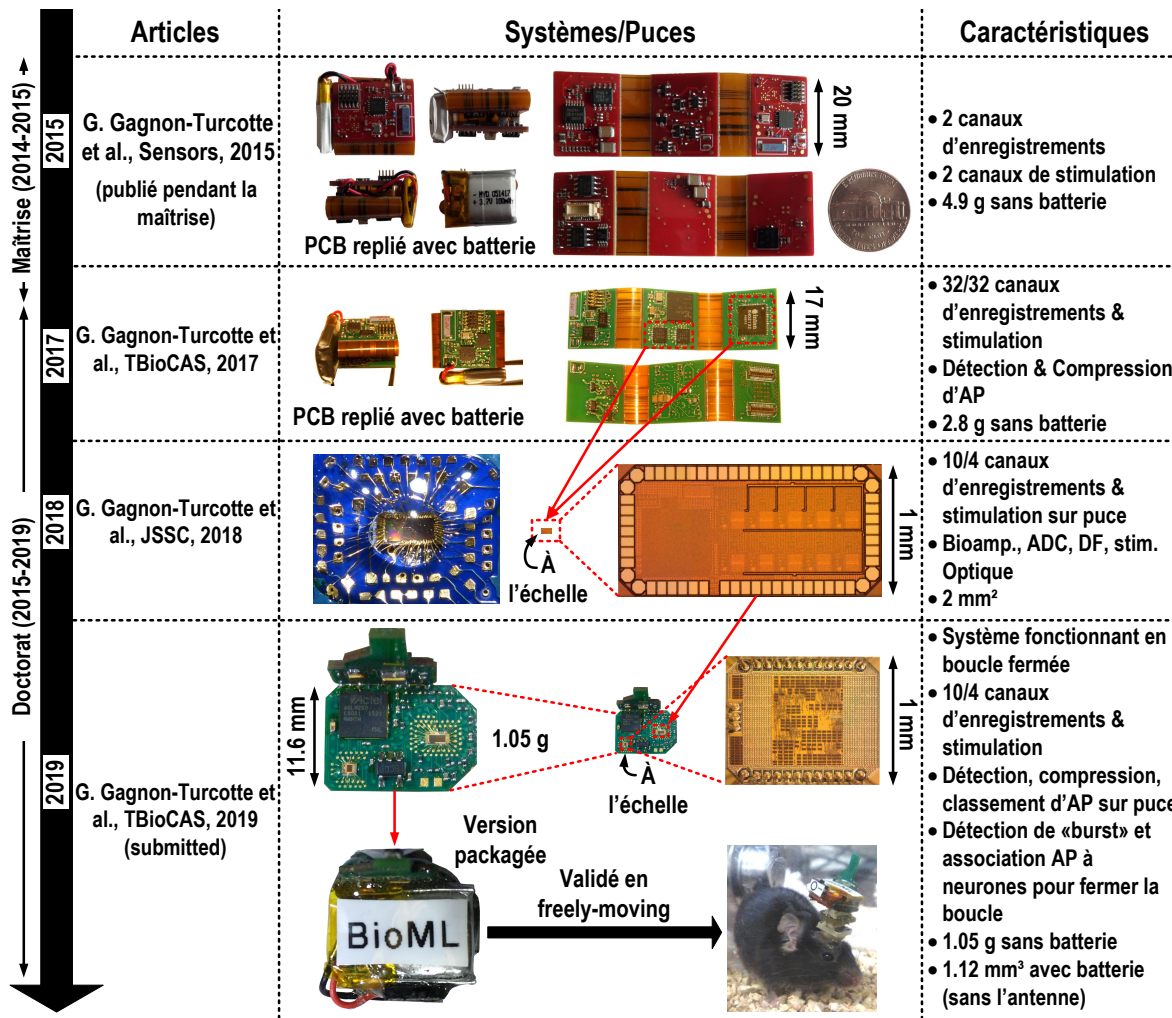


FIGURE 1.8 – Chronologie de l'évolution du projet.

Bien que ce système intègre des algorithmes innovants et qu'il soit le premier à réaliser de l'optogénétique et de l'enregistrement simultané haute résolution, il n'utilise que des composants commerciaux discrets. Or, ces composants sont volumineux et consomment beaucoup d'énergie.

Pour pallier à ces problèmes, l'auteur de cette thèse a conçu une puce CMOS d'acquisition électrophysiologique sur mesure dotée de stimulation optogénétique simultanée. Cette puce et son intégration dans une plateforme optogénétique sans fil complète ont été publiées lors de la troisième année de doctorat du rédacteur de cette thèse :

- Chapitre 3 : **G. Gagnon-Turcotte**, M. Noormohammadi Khiarak, C. Ethier, Y. De Koninck and B. Gosselin, "A 0.13- μm CMOS SoC for Simultaneous Multichannel Optogenetics and Neural Recording", *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, 3087-3100, 2018. © 2019 IEEE

Enfin, une seconde puce CMOS, appelée ND-IC, a été développée afin d'exécuter les algorithmes de compression et un nouvel algorithme de classement des signaux neuronaux. Le ND-IC est intégré aux côtés de la puce d'acquisition et de stimulation pour former un système multicanal capable de fonctionner en boucle fermée. Il s'agit de l'intégration ultime de toutes les composantes (puces, algorithmes, PCB, etc.) ayant été développées lors de ce projet de doctorat. Cette puce et le système sont décrits dans un article soumis pour publication lors de la quatrième année de doctorat du rédacteur de cette thèse :

- Chapitre 4 : **G. Gagnon-Turcotte**, C Ethier, I. Keramidis, Y. De Koninck and B. Gosselin, "A Wireless Electro-Optic Headstage with a 0.13- μm CMOS Custom Integrated DWT Neural Signal Decoder for Closed-Loop Optogenetics", *IEEE Transactions on Biomedical Circuits and Systems*, 2019 (submitted). © 2019 IEEE

Contributions

Ce travail de recherche apporte plusieurs contributions scientifiques, incluant :

1. **Création d'algorithmes pour détecter et compresser les signaux neuronaux** : L'article du chapitre 2 introduit un nouvel algorithme de détection automatique des PA, ainsi qu'un nouvel algorithme de compression des PA. Le détecteur de PA est basé sur un seuillage adaptatif utilisant une boucle de contrôle de type Sigma-Delta, ce qui permet de compresser les signaux en transmettant seulement les formes détectées. Chaque PA est davantage compressé par un algorithme basé sur une transformée en ondelette de type Symmlet-2 qui utilise une nouvelle technique de discrimination et de requantification dynamique des coefficients, ce qui permet d'atteindre des taux de compression élevés avec une bonne qualité de reconstruction comparativement aux approches existantes. Ces algorithmes permettent, entre autres, d'augmenter le nombre de canaux sous observation et de réduire la consommation énergétique du transmetteur sans fil.
2. **Création d'un système optogénétique sans fil avec enregistrement haute résolution** : L'article du chapitre 2 présente le premier dispositif optogénétique sans fil à être doté de stimulation optogénétique multicanal et d'enregistrement multicanal simultané haute résolution. Ce système possède 32 canaux d'enregistrement et 32 canaux de stimulation optique. L'enregistrement haute résolution est possible grâce à une implémentation pratique des algorithmes susmentionnés dans un FPGA embarqué à faible consommation énergétique.
3. **Développement d'un SoC CMOS d'acquisition et d'optogénétique multicanal** : L'article du chapitre 3 introduit un nouveau SoC CMOS qui permet de faire de l'optogénétique multicanal et de l'enregistrement neuronal multicanal avec une seule puce. Il s'agit du premier système intégré à être doté de ces deux fonctionnalités, ce qui permet de réduire significativement la taille (2 mm^2) et la consommation énergétique (11.2

$\mu\text{W}/\text{Canal}$) comparativement au système présenté au Chapitre 2 (1 seule puce versus plusieurs composantes commerciales discrètes). La conception du bioamplificateur inclut une bande passante programmable, permettant de cibler les PA et les LFP, et un faible bruit referré à l'entrée. Le convertisseur analogique-numérique (ADC), de type Delta-Sigma ($\Delta\Sigma$) MASH 1-1-1 à faible consommation, est conçu pour fonctionner avec de faibles taux de sur-échantillonnage (OSR). Cet ADC exploite une nouvelle technique réduisant la taille du circuit en soustrayant la sortie de chaque branche du $\Delta\Sigma$ dans le domaine numérique, plutôt que dans le domaine analogique, comme cela est fait de manière classique. La stimulation optique utilise un nouveau circuit de commande de LED basé sur une source de courant de type cascode avec rétroaction, ce qui permet d'accommoder une large gamme de LED et de plus petites tensions de batterie.

4. **Développement d'une puce CMOS (appelée *neural decoder integrated circuit*, où ND-IC) pour analyser les signaux neuronaux et permettant de faire de l'optogénétique en boucle fermée** : L'article du chapitre 4 propose une version optimisée pour une implémentation CMOS des algorithmes de détection et de compression, ainsi qu'un nouvel algorithme de classement automatique des PA, dans le but de faire de l'optogénétique en boucle fermée. Cette puce possède une plus faible consommation énergétique ($56.9 \mu\text{W}/\text{Canal}$), une plus petite taille (0.08 mm^2) et plus de fonctionnalités que les solutions existantes. Le ND-IC possède 3 modules : 1) le détecteur de PA possédant une implémentation optimisée, 2) le module de compression possédant une implémentation optimisée, et 3) le module de classement automatique des PA qui réutilise les données générées par le module de détection et de compression pour réduire sa complexité. Ce nouvel algorithme de classement est basé sur une représentation vectorielle des PA, dont le coût (taille, consommation, etc.) est considérablement réduit grâce au module de compression. De plus, le circuit réutilise l'écart-type du signal estimé par le module de détection, ce qui permet de classer les PA sans supervision et en temps-réel, contrastant avec les approches existantes.
5. **Création du premier système optogénétique multicanal capable de fonctionner en boucle fermée et le premier à être validé lors d'expériences *in vivo* impliquant des animaux libres de leurs mouvements** : Le système présenté au Chapitre 4 réutilise le ND-IC et le SoC du chapitre 3 pour former un système miniature (1.12 cm^3) et léger (1.05 g). Dans ce système, un lien entre un canal d'enregistrement et un canal de stimulation est établi selon l'association de chaque PA à un neurone, grâce à la classification, et selon l'activité «burst» de ce neurone dans le temps, combinant les techniques clés proposées dans [50] et [49] à l'intérieur d'un système sans fil. Il s'agit du premier système optogénétique complexe capable de fonctionner en boucle fermée, et le premier à transmettre simultanément les données neuronales.

Analyse complémentaire

Les articles présentés dans cette thèse suivent un style d'écriture succinct afin de se conformer aux exigences des journaux ou ils ont été publiés (nombre de pages, taille des paragraphes, etc.). Étant donné que cette thèse suit une structure par article, elle n'échappe pas à cette limitation. Pour cette raison, l'Annexe B présente une analyse complémentaire de certains concepts clés abordés dans les différents articles inclus dans cette thèse.

Structure

Cette thèse est organisée comme suit : Le chapitre 2 présente un système optogénétique multicanal sans fil doté d'enregistrement multicanal, et qui incorpore un algorithme de détection et un algorithme de compression des signaux neuronaux. Le chapitre 3 présente un SoC CMOS permettant de faire de l'optogénétique multicanal et de l'enregistrement neuronal multicanal sur une seule puce. Ce chapitre décrit tous les circuits du SoC, ainsi que son intégration dans une plateforme optogénétique sans fil complète. Le chapitre 4 présente un SoC CMOS numérique intégrant une version optimisée pour une implémentation CMOS des algorithmes présentés au chapitre 2, ainsi qu'un nouvel algorithme de classement des PA. Ce chapitre présente l'intégration de cette puce avec le SoC présenté au chapitre 3 dans un système optogénétique capable de fonctionner en boucle fermée, ainsi que sa validation lors d'expériences *in vivo* impliquant des animaux libres de leurs mouvements. Le chapitre 5 présente une liste des publications ainsi qu'une liste des distinctions générées par ce projet. Finalement, le dernier chapitre présente la conclusion.

Chapitre 2

A Wireless Headstage for Combined Optogenetics and Multichannel Electrophysiological Recording

© 2019 IEEE. Reprinted, with permission, from G. Gagnon-Turcotte, Y. LeChasseur, C. Bories, Y. Mes-saddeq, Y. De Koninck and B. Gosselin, "A Wireless Headstage for Combined Optogenetics and Multichannel Electrophysiological Recording", *IEEE Transaction Biomedical Circuits Systems*, vol. 11, no. 1, pp. 1-14, Feb. 2017.

2.1 Résumé

Cet article présente un *headstage* sans fil capable de détecter et de compresser les spikes en temps réel, permettant ainsi de faire l'optogénétique et de l'enregistrement électrophysiologique multicanaux. Ce *headstage*, qui est conçu pour de faire de l'enregistrement électrophysiologique et de la stimulation optique simultané dans le cerveau de rongeurs transgéniques se déplaçant librement, inclut 32 canaux d'enregistrements et 32 canaux de stimulation optiques. Il est capable de détecter, compresser et transmettre les formes d'onde complète des potentiels d'action sur 32 canaux en parallèle, et en temps réel, en utilisant un processeur de traitement de signal embarqué de type *field programmable gate array* (FPGA) ainsi qu'un microprocesseur softcore de type *Microblaze*. Ce processeur implémente un détecteur de spike innovant doté d'un seuillage adaptatif basé sur une boucle de contrôle de type Sigma-Delta, en plus d'un module de compression basé sur une transformée en ondelette utilisant une nouvelle technique de requantification dynamique des coefficients qui permet d'atteindre des taux de compression élevés avec une plus grande qualité de reconstruction. De la stimulation optique et des enregistrements simultanés ont été effectués *in-vivo* dans le cortex somatosensoriel et l'hippocampe d'une souris transgénique anesthésiée exprimant la protéine ChannelRhodospin-2 (Thy1 : :ChR2-YFP line 4), et ce, en utilisant une optrode ayant 8 microélectrodes et une fibre optique implantable couplée à une LED de 465-nm. Les ré-

sultats expérimentaux démontrent que le *headstage* est capable de déclencher de l'activité neuronale, tout en détectant et en compressant les signaux neuronaux de l'ordre du microvolt, sur plusieurs canaux en parallèle et avec des taux de compression supérieurs à 500. Nous rapportons le premier dispositif optogénétique sans fil à plusieurs canaux qui est doté de stimulation optogénétique et d'enregistrement simultané. Les caractéristiques mesurées démontrent que le *headstage* peut générer jusqu'à 100% de détection de vrais positifs pour un ratio signal-sur-bruit (SNR) de 15 dB, tout en générant 97.28% de vrai positifs pour un SNR aussi bas que 5 dB. Le prototype possède une autonomie de 105 minutes, et est basé sur un circuit imprimé flex-rigide léger (2.8 g) et compact (17x18x10 mm³).

2.2 Abstract

This paper presents a wireless headstage with real-time spike detection and data compression for combined optogenetics and multichannel electrophysiological recording. The proposed headstage, which is intended to perform both optical stimulation and electrophysiological recordings simultaneously in freely moving transgenic rodents, is entirely built with commercial off-the-shelf components, and includes 32 recording channels and 32 optical stimulation channels. It can detect, compress and transmit full action potential waveforms over 32 channels in parallel and in real time using an embedded digital signal processor based on a low-power field programmable gate array and a Microblaze microprocessor soft-core. Such a processor implements a complete digital spike detector featuring a novel adaptive threshold based on a Sigma-delta control loop, and a wavelet data compression module using a new dynamic coefficient re-quantization technique achieving large compression ratios with higher signal quality. Simultaneous optical stimulation and recording have been performed *in-vivo* using an optrode featuring 8 microelectrodes and 1 implantable fiber coupled to a 465-nm LED, in the somatosensory cortex and the Hippocampus of a transgenic mouse expressing ChannelRhodospin (Thy1::ChR2-YFP line 4) under anesthetized conditions. Experimental results show that the proposed headstage can trigger neuron activity while collecting, detecting and compressing single cell microvolt amplitude activity from multiple channels in parallel while achieving overall compression ratios above 500. This is the first reported high-channel count wireless optogenetic device providing simultaneous optical stimulation and recording. Measured characteristics show that the proposed headstage can achieve up to 100% of true positive detection rate for signal-to-noise ratio (SNR) down to 15 dB, while achieving up to 97.28% at SNR as low as 5 dB. The implemented prototype features a lifespan of up to 105 minutes, and uses a lightweight (2.8 g) and compact (17x18x10 mm³) rigid-flex printed circuit board.

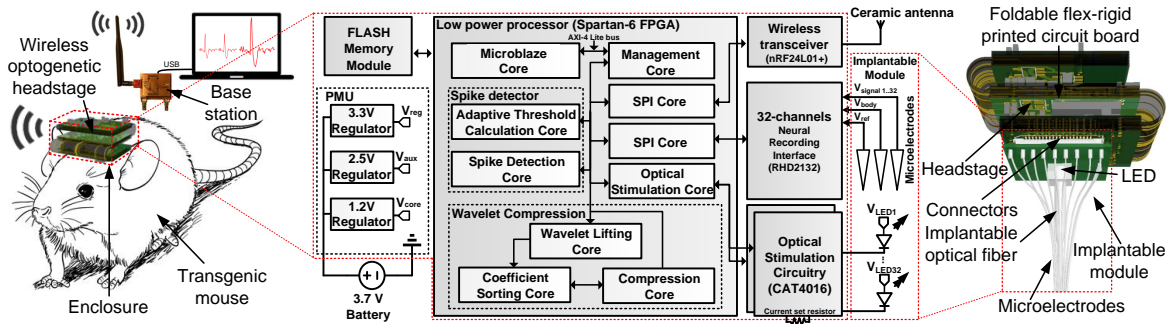


FIGURE 2.1 – Block diagram of the proposed headstage system including the foldable PCB connected to the implantable module. The headstage can stimulate optically using up to 32 LEDs and can record from up to 32 microelectrodes using spike detection and data compression thanks to an embedded digital signal processor implemented inside a low-power Spartan-6 FPGA. The headstage transmits the neural data wirelessly to a base station which is connected to a computer host. The implantable module that holds the LEDs, the optical fibers and the microelectrodes can be mounted on the head of the animal.

2.3 Introduction

Optogenetics is a new experimental approach that allows activating and/or silencing light-sensitized neurons with high temporal and spatial accuracy, offering precise control for studying the brain microcircuits. Recently, optogenetics has been used to carry various behavioral experiments with freely behaving rodents, especially mice [66], [67], [68]. Light activation/inhibition is commonly achieved using laser or LED-coupled optical fibers [14] that illuminate the targeted neural cells. Although tethered optical fibers provide strong illumination, their wired approach represents a danger of infection because of the tension in the cables attached to the implanted optical fibers, which also restrict the natural behavior of the animal. Additionally, using electrophysiology in combination with optical stimulation to collect the evoked neuronal activity has been identified as key in assessing the effect of optical stimulation in specific brain areas [37]. Therefore, wireless and light-weight optogenetic devices presenting single unit activity recording capabilities are highly sought-after.

Design of a practical wireless optogenetic system including single unit recording is very challenging since it must allow for a very high volume of monitoring channels in order to be informative. Most of the frequency content of the neural signals is located between 300 Hz and 5 kHz for action potentials (APs) [14], and between 0.5 Hz to 200 Hz for local field potentials (LFPs) [69], which necessitates a minimum sampling frequency of at least 10 ksamples/s. However, to reduce constraints on the analog filter, and to enable better recording accuracy, sampling frequencies as high as 30 ksamples/s are common in such systems. This leads to output data rates as high as 300 kbps per channel for 10-bit samples [70], which

prevent the use of large numbers of channels with most of the low-power transmitters, like those based on Bluetooth or Zigbee. In contrast, microelectrode arrays including hundreds of microelectrodes [71] are readily available, but require a bandwidth beyond several tens of Mbit/s to exploit all microelectrodes simultaneously.

Despite the many recent research efforts directed at designing new tools to collect single cell activity from hundreds of channels in parallel, existing wireless optogenetic systems including optical stimulation do not provide high-channel count neural recording capability [66], [71], [72], [73], [74]. In contrast, current wireless systems designed for neural recording applications, such as [75], [76] and [77], can provide hundreds of channels by leveraging custom integrated circuit technology, the development of which is expensive and time consuming. For instance, the systems presented in [75], [76] and [77] possess 256, 128 and 100 recording channels respectively. High-channel count systems based on commercial off-the-shelf (COTS) providing up to 96 recording channels have been reported as well for collecting neural data from non-human primates [78]. However, since their weight is as much as 17 g or more while providing only a few recording channels [78], [79], [80], these systems are normally too bulky and/or heavy to conduct behavioral experiments with small freely moving animals [73], [74]. Likewise, the commercial wireless systems, such as [72] and [71], either provide multichannel neural recording (up to 126 channels) or optical stimulation from a maximum of 2 channels, but not both functions simultaneously within a single system. Besides, the COTS based system presented in [81] weighs 4.9 g and provides both functions, but is limited to 2 high-power optical stimulation channels and 2 electrophysiological recording channels.

An alternate solution to the development of a customized transceiver consists in performing data management using dedicated signal processing approaches, such as data reduction and/or compression, to allow using a low-power transmitter. Detection and compression techniques have been previously employed for performing data management in neural recording systems. A neural wavelet compression based on an application-specific integrated circuit (ASIC) that compresses the signals of 32 channels by thresholding the wavelet coefficients compared to a fixed threshold value is proposed in [82]. A 64-channel ASIC that compresses neural data by sorting the spikes on-the-fly and by selecting and sending only a limited number of spike features is reported in [83], while [75] reports a 256-channel neural recording ASIC provided with embedded delta compression. These compression techniques have produced promising results with simulated or pre-recorded data sets, but have not yet been tested inside a complete neural data acquisition system, nor validated within *in vivo* experiments. Further, these systems do not include a dedicated wireless transceiver. Moreover, systems using data reduction often rely on detecting and sending a subset of AP features

only [79], [76], [83], without capturing and transmitting the full spike waveforms, which is not practical in most experimental settings since it does not allow further signal processing steps, such as shape-based spike sorting.

This paper presents the first attempt at high resolution neural recording with optogenetics. We are introducing the first wireless headstage system entirely built using commercial components that process and transmit 32 neural data acquisition channels in parallel in addition to controlling 32 implantable fiber-coupled optical stimulation LEDs. Compared to an early version reported by our group in [81], the proposed headstage includes 32/32 recording/stimulation channels instead of 2/2, and is using an action potential detector and a neural signal compression module implemented inside a low-power digital signal processor, while being smaller and lighter than [81] (2.8 g vs 4.9 g). A major contribution of this work is that the filters, memories, clock speeds required for spike detection and wavelet compression have been implemented on a low-power field programmable gate array (FPGA) in a practical and feasible manner. The proposed headstage performs spike detection and data compression *in situ* using an embedded digital signal processing core and a Microblaze soft processor core, both implemented inside the FPGA. The spike detector uses the absolute value operator along with a new adaptive threshold based on a Sigma-delta control loop enabling compressing and transmission of the full AP waveforms. The data compression core is using a Symmlet-2 wavelet lifting transform along with coefficient thresholding and a new dynamic coefficient re-quantization technique. The performance of the proposed system has been measured *in vitro* and demonstrated within *in-vivo* experiments with a Chr2 transgenic mouse. Results show that the system can trigger neuron activity using optical stimulation, while collecting, detecting and compressing the evoked neuronal signals from several channels in parallel with excellent reliability.

The paper is organized as follow : Section 2.4 presents a system overview of the headstage, section 2.5 presents the design of each building block, section 2.6 describes the design of the novel spike detector and the wavelet compression module, section 2.7 presents the system implementation and the measured performance, section 2.8 presents the results obtained from in-vivo optogenetic experiments, section 2.9 presents a comparison of the proposed headstage with other existing systems, and finally the conclusion in section 4.13.

2.4 System Overview

The block diagram of the proposed optogenetic headstage is shown in Figure 2.1 (as previously described in [84]). It provides real-time signal processing and wireless data transmission, and it is entirely built using COTS components. The complete system consist of two

interconnected parts. The first part includes a foldable printed circuit board (PCB), which contains the electronics for data acquisition, signal processing (including spike detection and data compression), optical stimulation and wireless data transmission. The second part is a detachable implantable module, which holds the fiber-coupled LEDs, the implantable optical fibers and the microelectrodes. Both parts are held together using two Molex connectors (part # 5024304010). The proposed design consists of five main building blocks : 1) the 32-channel neural recording interface, 2) the optical stimulation circuitry, 3) the power management unit (PMU), 4) the wireless transceiver and 5) the low-power FPGA processor. The neural recording interface allows for conditioning and sampling the low-amplitude extracellular action potentials whose bandwidth ranges between 300 Hz and 5 kHz, and amplitude between $50 \mu\text{V}$ and $500 \mu\text{V}$ [14]. A driver circuit powers the LEDs with a forward voltage close to the battery voltage to enable a precise forward current inside the LED and deliver the correct optical power to the targeted neurons (see Section 2.5.2 for more details). The PMU provides a regulated supply voltage to all building blocks. A digital radio transceiver operating at 2.4 GHz is used to transmit the neural data collected from 32 channels in parallel to a base station, and to receive the stimulation parameters inside the headstage. System control, spike detection, wavelet compression and management of optical -stimulation patterns are handled by a Microblaze microcontroller implemented within the FPGA. Using a Microblaze microcontroller considerably reduces the control logic complexity and eases the communications between each FPGA module, leading to a much more robust and flexible design. The FPGA is programmed automatically when the system powers up from an external non-volatile 16-Mbytes FLASH memory (MICRON, USA).

2.5 System Design

2.5.1 Neural Recording Interface

The neural recording interface conditions and digitizes the measured extracellular AP by up to 32 microelectrodes in parallel. The extraction of such a high number of sensitive neural signals is accomplished using a dedicated RHD2132 mixed-signal single-chip interface from *Intan Technologies*, available in a small QFN package (8 mm x 8 mm). This dedicated interface provides low input noise and high linearity to properly measure the low amplitude neural waveforms. The RHD2132 can sample 32 neural signals, the maximum amplitudes of which can lie between -5 mV to 5 mV, with a precision of 16 bits at 30 ksamples/s each, and exhibits an input-referred noise of $2.4 \mu\text{V}_{rms}$. In addition to providing low input referred noise, the RHD2132 is reconfigurable through a serial peripheral interface (SPI) bus, which allows on-the-fly reconfiguration of the band-pass filters to accommodate different input signals and/or different input referred noise levels. Each channel includes a band-pass filter that consists of a first order high-pass filter, the cutoff frequency of which is reconfigurable from 0.1 Hz to 500 Hz, cascaded with a low-pass 3rd-order Butterworth filter, the cutoff frequency

of which is reconfigurable from 0.1 kHz to 20 kHz.

2.5.2 Optical Stimulation Circuitry

The optical stimulation circuitry can drive up to 32 LEDs in parallel. A typical optogenetic setting can require delivery as high as 10 mW/mm² in the vicinity of the light sensitized neurons [23], which corresponds to an appreciable amount of forward current passing through the LED. Such a high current can generate a voltage drop almost equivalent to the battery voltage (3.7-V) across the LEDs. Therefore, we designed a 32-channel LED driver using two small low-dropout CAT4016 LED driver from *ON Semiconductor* (TQFN-packages), which has a low dropout voltage of 0.4-V when passing currents between 2-mA and 100-mA. The driver circuit output current is determined by the values of two resistors (one resistor per chip), the values of which can be set for different applications at assembly to allow for optical stimulation pulses of various intensity. Measured performance shows that the driver can power the LED with a forward current as high as 150-mA per chip without significantly increasing the dissipated heat. Thus, the maximum amount of LED current that can be handled by the optical driver without overheating is 300 mA, which allows driving several low-power LEDs in parallel (32 low-power LEDs at 9.4 mA each, for instance) or a subset of high-power LEDs (two high-power LEDs at 150 mA each, for instance). Such high-power LEDs (LB G6SP from OSRAM Opto Semiconductors) can deliver as much as 250 mW/mm² at the optical fiber output, when properly coupled, which is highly superior to the light sensitized neurons activation threshold [23]. The frequency and duty cycle of the stimulation patterns are configurable through a serial link using the CAT4016 custom bus protocol, which is controlled by the Microblaze.

2.5.3 Power Management Unit

The entire headstage is powered by a 3.7-V, 100-mAh and 2.1-g Lithium-ion battery (Model 051417, MYD Technology). The PMU provides a fixed 3.3-V regulated supply voltage to the FPGA, the neural recording interface, the transceiver and the external FLASH memory. The PMU also provides a stable 2.5-V to the FPGA auxiliary inputs and a stable 1.2-V to the FPGA core. A low-dropout voltage regulator (TLV70233, *Texas Instruments*) is used to generate the 3.3-V supply with a power supply rejection ratio (PSRR) of 68 dB at 1 kHz, which attenuates any artifacts from the optical stimulation circuitry before effecting the neural recording interface. Both the regulated 2.5-V and the 1.2-V supply voltages are supplied by a high-precision LP3990 voltage regulator. All integrated circuits in the PMU use small 3 mm x 3 mm SOT-23 packages for minimizing the size and weight of the headstage.

2.5.4 Wireless Transceiver

A nRF24l01p radio transceiver from *Nordic Semiconductors*, operating in the 2.4-GHz center frequency of the industrial, scientific and medical (ISM) band, is used to transmit the neural data to a base station, and also to receive the optical stimulation parameters. This commercial transceiver enables low-power operation, with a current consumption of 11.3 mA in transmitting mode and a current consumption of 13.3 mA in receiving mode, and is available in a small 4 mm × 4 mm QFN20 package. The maximum effective transmission rate measured for this transceiver is 1.4 Mbps.

2.5.5 Signal Processing Core / Controller

System control, spike detection, wavelet compression, generation of optical stimulation patterns and data communications are handled by a low-power LX16 Spartan-6 FPGA from *Xilinx*. The LX16 FPGA possesses all the logic resources required for this application, and is available in a small 8 mm × 8 mm CPG196 ball grid array (BGA) package, enabling appreciable miniaturization. Furthermore, the Microblaze soft microprocessor core from *Xilinx* is implemented inside the Spartan-6 to control the headstage's multiple building blocks. The FPGA mapping implements multiple processing and control cores (see Figure 2.1) : 1) a Microblaze core 2) a Spike detection core, 3) an Adaptive threshold core, 4) a Wavelet lifting core, 5) a Wavelet coefficient sorting core, 6) a Wavelet compression core, 7) an Optical stimulation patterns management (OSPM) core, 8) two SPI driver cores for communicating with the neural recording interface and with the transceiver, and 9) a Data management core, which handles the communications between the Microblaze and the other cores via a single AXI-4 Lite bus. In addition to providing internal configuration data (bandwidth and registers configuration) to the neural recording interface, the Microblaze controls all other cores and handles the communications between the RF transceiver and the other blocks. The Microblaze firmware is written in the C language, while other cores are written in the VHDL language.

2.6 Firmware Design

2.6.1 Spike Detector Core Design

The proposed spike detector is based on the absolute value operator, which yields low complexity and good performances [85], [86], described by :

$$y(n) = |x(n)|, \begin{cases} \text{if } y(n) > \text{threshold, a spike is detected} \\ \text{otherwise, no spike is detected} \end{cases}$$

where $x(n)$ is the n^{th} sample. Upon detection of a spike, several samples before and after the detection point are buffered inside the internal RAM memory. Next the data sequence

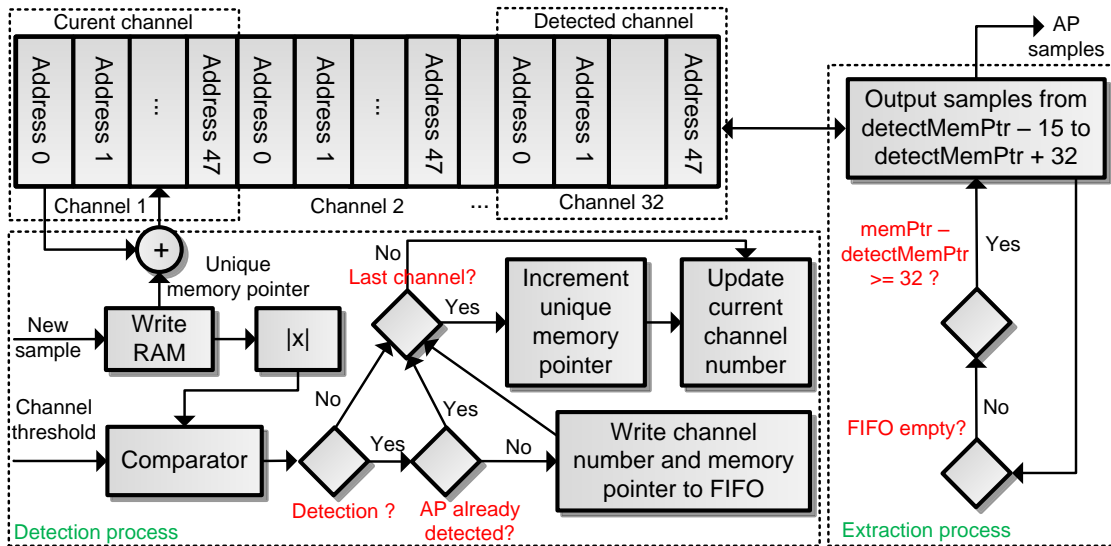


FIGURE 2.2 – Block diagram illustrating the implemented spike detector. Each new incoming sample is stored inside a circular buffer in dedicated RAM memory. When a detection occurs, 48 spike samples (15 samples before, and 32 samples after the detection point) are sent to the *Wavelet lifting core*.

representing a write complete spike is compressed, packetized and sent to a base station.

The RHD2132 chip encodes the neural data sampled at 20 ksp/s on a 16-bit resolution. For each channel, a data buffer of 2.4 ms, which corresponds to a window of 48 samples (15 samples before and 32 samples after the detection point), is kept in memory. Thus, a buffer of 3072 bytes is used for covering the worst case scenario of detecting 32 spikes simultaneously (1 spike detected on each channel within a 2.4-ms time window). Such a buffer of data is implemented inside the internal FPGA RAM memory blocks. The adopted strategy consists in splitting the RAM memory into 32 sections of 96 bytes (48 samples of 16 bits each), where each section serves a specific channel and acts as a circular buffer. Because the output of the RHD2132 consists of an ordered sequence of time-multiplexed samples each belonging to a different channel, incoming samples are stored sequentially inside the FPGA using a unique memory pointer which is added to the base address of the current memory section. After processing the 32nd sample of this sequence (which belongs to the last channel), the unique memory pointer is incremented by one position. Meanwhile, the absolute value of each new incoming sample is compared with a 16-bit threshold.

When a detection occurs, an array of 32 bits (one bit per channel) is written at the position holding the detection state of the corresponding channel, while the unique memory pointer is written along with the detected channel number into a FIFO implemented using an IP core

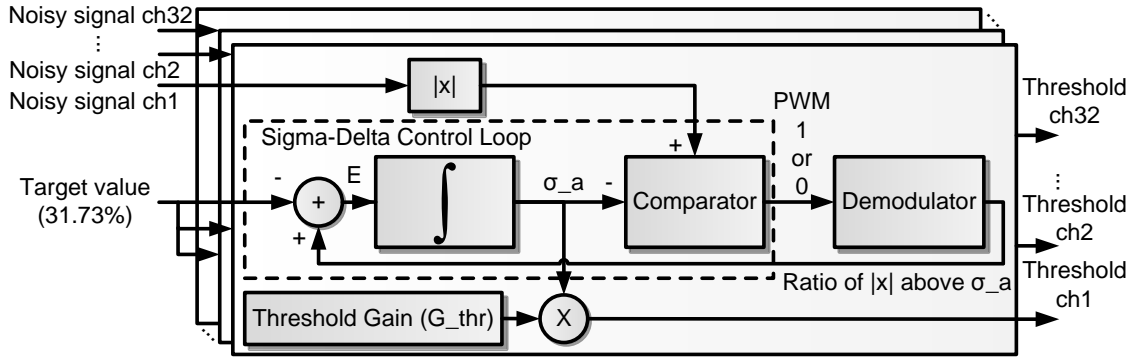


FIGURE 2.3 – Multichannel threshold estimator based on a Sigma-Delta control loop. The loop locks on the ratio of sample whose amplitude is greater than the output of the integrator (σ_a). The estimated standard deviation σ_a is multiplied by G_{thr} to give the threshold value.

from *Xilinx*. When the position in the array indicates that a detection has already occurred for a specific channel, the detected sample is deemed to belong to a spike that has previously been detected, and which is being buffered, thus it is ignored. The use of a FIFO for holding the memory pointers and the channel numbers guarantees that the detected spikes are processed in the right order.

The FIFO is poled by a parallel process, which extracts the spike samples by comparing the stored memory pointer of the former detected spike with the current unique memory pointer. After accumulating exactly 32 new spike samples (fixed amount, not programmable) inside the circular buffer following a detection event, a total of 48 spikes samples (including the 15 samples previously stored in RAM prior to a detection event) are sequentially transferred to the *Wavelet lifting core*. As soon as all samples are transferred, the corresponding position inside the detection array is reset in order to re-enable detection of new incoming spikes for that channel. All the steps mentioned above are summarized in Figure 2.2.

2.6.2 Adaptive Threshold Calculation Core Design

The absolute value operator in the *Spike detector core* works alongside a threshold, the value of which is determined adaptively according to the standard deviation of the noise. The microelectrodes used to extract the low-amplitude neuronal waveforms typically have high impedances (0.5 M Ω up to 3 M Ω), which result in poor signal-to-noise ratio (SNR). Moreover, the SNR can change over time due to electrode displacement, electrode drifting or background neuronal noise. Thus, an automatic threshold calculation strategy based on the SNR estimation is adopted in this proposed headstage.

The proposed adaptive threshold strategy relies on the statistical properties of the half-

normal distribution. The thermal noise of the microelectrode is assumed to dominate [87], thus, a white Gaussian noise is assumed in the channel [88]. After passing the absolute value operator, the amplitude of the rectified white Gaussian noise follows an half-normal distribution. Given the probability density function of the half-normal distribution, the probability that the amplitude of the rectified noise exceeds the standard deviation (σ) is 31.73%, as stated in (2.1).

$$31.73\% = 100\% - \int_0^\sigma \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2\sigma^2}} dx \times 100\%. \quad (2.1)$$

The block diagram of the proposed threshold estimator is depicted in Figure 2.3. The proposed algorithm estimates the standard deviation of the rectified noise by using a feedback control loop based on a Sigma-delta modulator which uses the integrator's output to determine the threshold value. It compares the absolute value of the noisy signal with the control signal (σ_a), and demodulates the generated PWM output in order to estimate the number of samples whose absolute value is greater than the control signal. For each iteration that is triggered by a new incoming sample, the integrator tries to reduce the error between the target value, the value of which is fixed at 31.73%, and σ_a in order to approximate the standard deviation. Once the control loop reaches its steady state, the integrator's output signal produce the estimated standard deviation of the noise, and the integrator's output is multiplied by an integer gain G_{thr} (in Figure 2.3) to provide the threshold value : $Threshold = G_{thr} \times \sigma_a$. The integer gain G_{thr} determines how many times the estimated standard deviation is multiplied to give the threshold value, which can be seen as an aggressiveness factor chosen by the user. The proposed control loop is similar to a first order Sigma-delta modulator, but differs from conventional systems since the demodulator is using a simple FIR digital filter in the feedback path, and since the comparison occurs between the new incoming samples and the integrator's output.

In contrast with the analog implementation presented in [88], the proposed digital adaptive threshold only uses simple logic blocks like RAM memory blocks, adders, comparators and shift registers, which renders its implementation flexible and highly reliable. The Integrator is based on a Riemann sum :

$$\sigma_a(n) = \sigma_a(n-1) + G \times T \times E(n), \quad (2.2)$$

where G is the gain of the integrator, T is the sampling period and $E(n)$ is the error with respect to the *Target value*. The output $\sigma_a(n)$ is the approximation of the standard deviation, the value of which is compared with the absolute value of the current sample. A bit counter, noted $P(n)$, holds the sum of all the comparison bits stored in memory which, divided by the circular buffer length ($L = 128$), represents the percentage of the samples whose amplitudes are above σ_a . Since the circular buffer length is known in advance, this design

avoids using a division circuit by setting the *Target value* to 31.73% multiplied by the circular buffer length, noted L . Thus, the operation $E(n) = TargetValue - P(n)/L$ becomes $L \times E(n) = L \times TargetValue - P(n)$, which is exempt from any operation of division. Additionally, the variable L that is multiplying the error $E(n)$ is merged with the integrator gain G and with the sampling period T . Since G , T and L are known in advance, the multiplication $G \times T \times L \times E(n)$ is approximated simply by shifting the bits of the error $E(n)$. The bit counter and the last integrated value are stored in RAM memory along with the content of the circular buffer of each channel. The proposed threshold calculation algorithm necessitates only one addition, one subtraction, two multiplications (including the calculation of the threshold gain G_{thr}), which are achieved by straightforward bit shifting operations, and a circular buffer memory of 128 bits. Each threshold calculation necessitates nine clock cycles per channel, which allows using a low clock frequency of 5.76 MHz to yield a sampling rate of 20 kHz.

2.6.3 Wavelet Compression

In order to dramatically increase the number of transmitted neural recording channels, while using a low-power commercial radio transceiver, the headstage compresses the detected AP waveforms using a four level discrete wavelet transform (DWT) applied to each of the detected waveforms' 48 samples. To reduce the DWT footprint inside the FPGA, the incoming detected APs that are extracted by the *Spike detection core* are multiplexed toward a single *Wavelet lifting core*. With a given sampling frequency of 20 kHz and a worst case scenario of 32 simultaneous detected spikes in a time windows of 2.4 ms, the processing time to compress each detected AP must be ≤ 75 s (48 samples/spikes / (20000 samples/s \times 32 channels)) to avoid any buffer overflow. Thus, a compression core frequency of 20 MHz has been selected as a reasonable trade-off between speed and power consumption, allowing up to 1500 clock cycles per compression of a full AP. The selected mother wavelet basis has a major impact on the compression ratio and on the overall quality of reconstruction [89] for this application. In fact, a mother wavelet whose shape is similar to the signal to compress can typically generate fewer significant wavelet coefficients for the same quality of reconstruction. Since the *Symmlet* wavelet basis family have demonstrated nearly optimal compression performance for neural signals [89], a sequential Symmlet-2 DWT lifting scheme is implemented inside the FPGA. For data compression, the wavelet coefficients are first thresholded to only retain the most significant N coefficients, and then, further compressed using a dynamic requantization operation, which will be described in the next few paragraphs.

Symmlet-2 Lifting Sequential Implementation

Figure 4.5 shows the data flow diagram of a four level DWT lifting scheme with *Symmlet-2* basis. $X_e(n)$ and $X_o(n)$ are the even and the odd parts of the input signal while Q_0 , Q_1 and

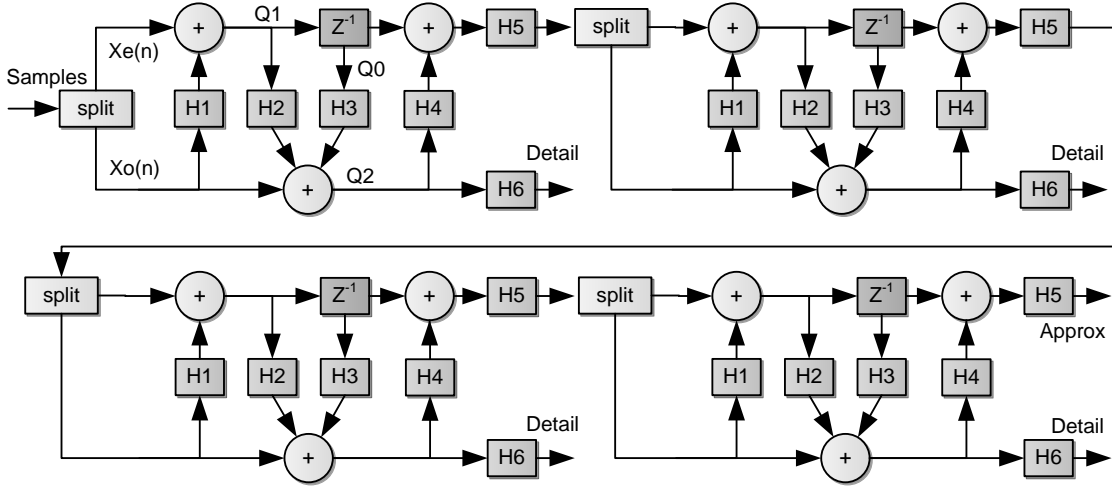


FIGURE 2.4 – Data flow diagram of a four level DWT lifting scheme using a *Symmlet-2* basis. The wavelet transform is performed on 48 spikes samples, thus producing 48 coefficients in total.

Q_2 are intermediate results. The lifting steps of each level can be described as follows :

- $Q_0 = X_o(n - 1) + H1 \cdot X_e(n - 1)$
- $Q_1 = X_o(n) + H1 \cdot X_e(n)$
- $Q_2 = X_e(n) + H3 \cdot Q_0 + H2 \cdot Q_1$ (3)
- $Approx = H5 \cdot (Q_0 + H4 \cdot Q_2)$
- $Detail = H6 \cdot Q_2$

where $H1-H4$ are the four constant filter coefficients, $\sqrt{3}$, $\frac{\sqrt{3}}{4}$, $\frac{\sqrt{3}-2}{4}$ and -1 respectively, while $H5$ and $H6$ are the $\frac{\sqrt{3}+1}{\sqrt{2}}$ and $\frac{\sqrt{3}-1}{\sqrt{2}}$ scaling factor respectively. The sequential system diagram of the proposed *Symmlet-2* DWT implementation is shown in Figure 2.5. FIFO IP Core from *Xilinx*. A *Logic core* handles all the arithmetic operations by sending commands along with the AP samples (even and odd samples) to a *Computation core* that includes the multiplier circuit, the constant filter coefficients and the scaling factors (4 x 16 bits values), the values of which are stored inside general purpose registers. The *Computation core* includes internal registers that hold the intermediate results. Additionally, the *Logic core* coordinates the data flow between all DWT levels as well as the storage of the coefficients inside four 12 x 16 bits dedicated RAM memory blocks (see section 2.6.3). The total DWT latency is of 456 clock cycles.

Threshold Calculation

After calculating the 48 wavelet coefficients, the coefficients are processed by the *Coefficient sorting core*, which sorts the coefficients by amplitude in order to find the amplitude of the N^{th} sorted coefficient, where N is the number of coefficients to retain, which value will be further used as a threshold in the *Compression core*. Instead of sorting the coefficients indivi-

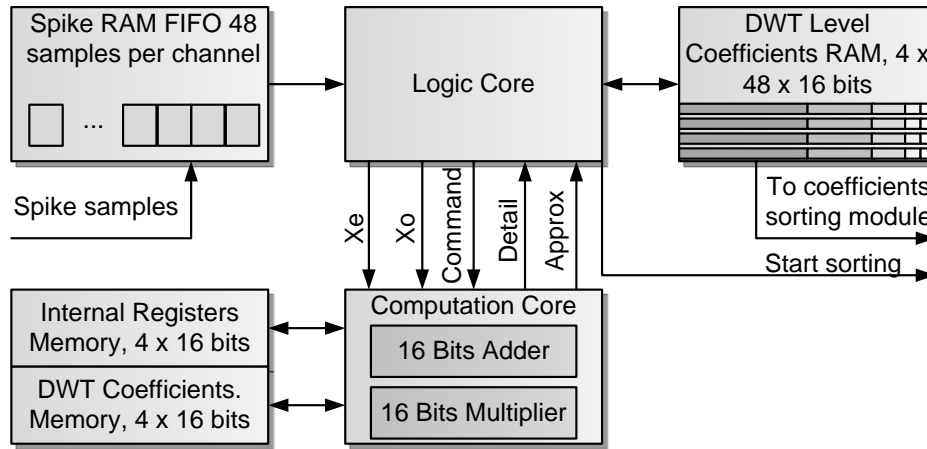


FIGURE 2.5 – System diagram for *Symmlet-2* lifting sequential implementation. Each new incoming spike sample is stored inside a FIFO, before being processed by a unique *Wavelet lifting core*. A *Logic core* handles all the arithmetic operations by sending commands along with the *AP* samples to a *Computation core*, and coordinates the data flow between all *DWT* levels as well as the storage of the coefficients inside dedicated *RAM* memory block.

dually, the adopted strategy consists of identifying this threshold by counting the number of coefficients having a higher amplitude than the current coefficient, and by repeating this task for all coefficients. Multiple simultaneous comparisons are performed per clock cycle in order to speed up the process. As soon as the count equals the predetermined number of coefficients to retain, the amplitude of the current coefficient is identified as the coefficient threshold (N^{th} sorted coefficient). Meanwhile, the coefficient with the largest amplitude is found as well, and both coefficient values are forwarded to the *Compression core*. In its basic form, the algorithm performs 2304 comparisons (48 comparisons for each of the 48 coefficients). In order to speed-up the calculations, the comparisons are made for four coefficients at a time, but at a cost of using more memory resources. Thus, the coefficient values are distributed over four different *RAM* memory blocks for enabling four simultaneous memory accesses, which finally result in a total of 576 clock cycles (12 blocks of four coefficients \times 48 coefficients).

Data Compression

Once the maximum and N^{th} wavelet coefficient values are identified, data compression occurs within the *Compression core*. The proposed *AP* compression strategy consists in combining a coefficient suppression technique based on thresholding along with a dynamic re-quantization technique :

- i) For the coefficient suppression step, the amplitudes of all coefficients are compared with

the amplitude of the N^{th} coefficient found in the previous step. The locations of the retained coefficients are marked inside an associated 48-bit tracking array.

ii) For the dynamic requantization step, a larger compression ratio is achieved by removing the least significant bits. Such an effective solution dynamically expands the coefficient values in order to represent the coefficients on a maximum number of bits before performing the re-quantization step. A direct method in achieving this goal is to shift the bits of each coefficient to the left until the most significant bit (MSB) of the maximum amplitude coefficient reaches the 16^{th} bit position. This technique is simple, but is not optimal, especially when a high compression ratio is sought over a small number of bits. For example, with a maximum coefficient amplitude of $0x1000$ (4096), another coefficient having a value of $0x0010$ (16) would be nulled when requantized on 8 bits. A more effective approach consists in normalizing all coefficient values by the maximum coefficient amplitude value and re-quantizing all remaining bits over the mantissa bits. It should be noted that fractional values are easily calculated within an FPGA using integer numbers representing the dividend on 32 bits, while the 16 original bits are located at the MSB positions, and by dividing the result by the maximum coefficient amplitude. Then, a re-quantization of all coefficients is performed on the resulting 16 LSB. Thus, for a maximum amplitude of $0x1000$, the coefficient value $0x0010$ becomes $0x0100$ (256), which is not nulled after requantization on 8 bits ($0x01$). As can be seen, the proposed method gives one more bit of precision as compared to the simple shifting operation, which is a major advantage when the coefficients are re-quantized on a small number of bits, such as 5 or 6 bits.

Since the wavelet coefficients having the larger amplitudes contain more information on the original waveform shapes [90], the proposed approach, which consists in sending the maximum amplitude coefficient in full resolution to maximize the precision after the re-quantification step, offers much improvement. Such an approach can properly reconstruct the compressed waveform with a resolution equal to the original resolution of the waveform after reception in the base station. In the reconstruction step, the value of the non-compressed maximum amplitude coefficient represented on 16 bits is multiplied by the compressed coefficients represented on a smaller number of bits. This operation allows each recalculated coefficient to retrieve its original resolution, at the cost of transmitting 16 additional bits per AP, which is negligible for this application. The proposed compression algorithm uses 430 clock cycles, eight clock cycles per coefficient plus 46 latency cycles for the divider module, which results in a total of 1462 clock cycles per AP compression. The AP compression ratio achieved by the described approach is given by :

$$CR = \frac{48 \text{ samples} \times 16 \text{ bits}}{48 \text{ bits} + 16 \text{ bits} + N \times Q'} \quad (2.4)$$

where N is the number of coefficients kept and Q is the number of bits employed by the

re-quantization step.

2.7 Implementation and Measured Performance

2.7.1 Power Consumption and Battery Lifespan

The measured power consumption of each building blocks is depicted in Figure 2.6a. It can be observed that the stimulation LED draws most of the power (32%), followed by the FPGA (27%), the transceiver (14%) and other building blocks (27%). Since the compression core is active only when spikes are detected, the overall power consumption of the FPGA is function of the firing rate. As to quantify the potential power fluctuation, Figure 2.6b presents an histogram of the measured power consumption for both the FPGA and the whole system, without stimulation, and for different firing rates. It can be seen that increasing the firing rate increases the FPGA power consumption, and that of the overall system. In fact, the power consumption increases with $\sim 8.5 \mu\text{W}/(\text{spikes}/\text{s})$ per channel on average, which is mainly attributable to the increased power consumption of the transceiver that has to transmit more compressed spikes as the firing increases. For comparison, we measured the power consumption per channel of the system with and without detection and data compression (and without stimulation), which was found to be of 31.9 mW/channel and 3.72 mW/channel, respectively. Thus, the FPGA processor improves the overall power consumption per channel by a factor of 8.6. Note that the system w/o detection and data compression was measured at near full transmitter speed (1.28 Mbps), which corresponds to a maximum number of 4 processed and transmitted channels of RAW recorded data, whereas the system w/ detection and data compression was measured for 32 channels each processing a prerecorded input neural signal presenting a typical firing rate of 45 spikes/s.

The headstage presents a measured autonomy, without optical stimulation and in full operation mode (all FPGA cores are activated, except the OSPM core), of 105 min, and of 70 min when using a 150-mA stimulation current (with a 10% duty cycle on 1 stimulation LED and at a firing rate of 45 spikes/s on each 32 channels), which is long enough for running one or two successive typical optogenetic experiments. It should be noted that a similar autonomy can be achieved using multiple LEDs of lower forward currents, like 32 LEDs with a current 4.7-mA, for instance.

2.7.2 Foldable Printed Circuit Board

Small size and low weight are critical in this application. All headstage components are mounted on a rigid-flex PCB containing 3 rigid sections and two flexible sections which replace any bulky board-to-board connectors. The top rigid section is dedicated to the transceiver and the PMU, the middle rigid section, to the FPGA and the LED driver, and finally,

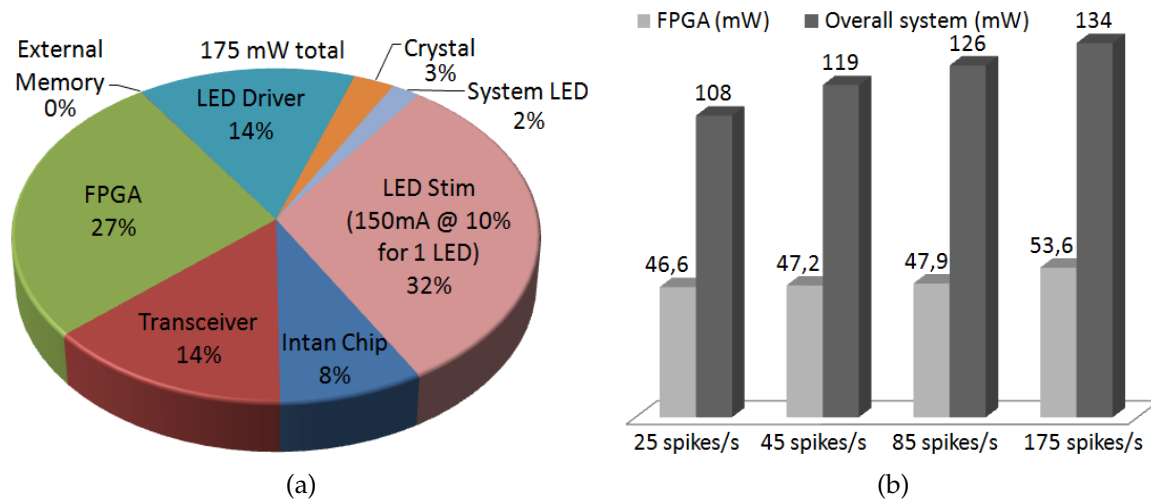


FIGURE 2.6 – a) Power consumption of each building block when all 32 channels are activated with a firing rate of 45 spikes/s per channel. The measurements were performed with one LED driven by a 150 mA current at a duty cycle of 10%, b) power consumption of the whole system and the FPGA with respect to the neuronal firing rate, without optical stimulation.

the bottom rigid section holds the neural recording interface and the two *Molex* connectors which connect the headstage with the implantable module. Since the implantable module is designed for permanent installation on the skull of the animal, the *Molex* connectors allow removing the headstage between experiments and/or to reuse the same headstage for multiple animals. The proposed foldable PCB includes 6 layers and has a minimum track width of 100 μm and minimum hole size of 200 μm . Fig. 2.7 shows the unfolded rigid-Flex PCB and the folded system with the battery on side. This miniature design results in a lightweight (2.8 g, w/o battery) and compact device (17x18x10 mm³, w/o battery). It should be noted that the battery could also be installed on the back of the animal to even further limit the weight on its head. The measured system characteristics are summarized in Table 2.1.

2.7.3 Spike Detection Performance

This section presents the measured results of the proposed spike detector, and a comparison with another threshold calculation methods using an absolute value operator. The thresholding methods were tested using synthetic neuronal signals with varying SNRs that were emitted at the headstage's inputs using an arbitrary function generator. Using multiple SNRs allows observing the impact of the noise level on each thresholding method's detection rate. The synthetic input signals were attenuated using a precise voltage divider to provide realistic AP waveforms with peak-to-peak amplitudes between 150 μV and 400 μV . The synthetic waveforms were built using a bank of five AP wave shapes taken from the University of Leicester on-line neural signal bank, and by adding white Gaussian noise to obtain the desired SNR. The SNR is measured as follows :

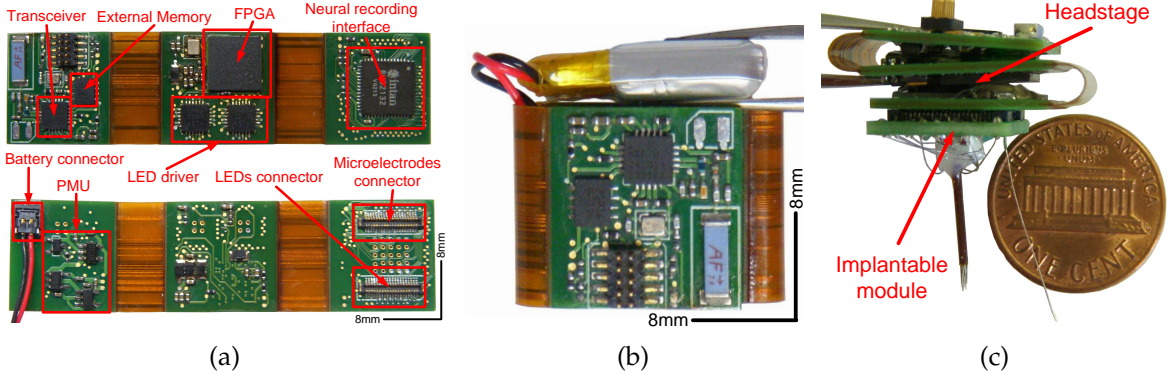


FIGURE 2.7 – a) Two-sided view of the rigid-flex PCB when not folded with the referenced sub-systems, b) top system view (folded) with battery on side, c) side system view (folded) attached to the implantable module.

$$SNR_{dB} = 10 \log \frac{P_{ap}}{P_n}, \quad (2.5)$$

where P_{ap} is the power of the spike-only signal, i.e. the power calculated over the summed signal intervals that contains AP waveforms, and P_n is the power of the noise without the APs. During the tests, all FPGA cores were activated, in order to reproduce normal operating conditions. The test was carried out on a single channel. For this channel, the detected and compressed spikes were sent to the base station along with the raw recorded signal, in order to thoroughly assess the performance of the Spike detector offline in Matlab in terms of detection rate and errors. The same raw recorded signal was used to quantify the performances of the other thresholding method.

Implementation Parameters

There is a trade-off between the size of the circular buffer employed in the demodulator (Figure 2.3) and the latency of the Sigma-delta control loop. When in a steady state, the calculated threshold is independent of the circular buffer size, which size was empirically set to 128 bits (or 16 bytes). The total integrator gain, which consists of G the integrator gain, T the sampling period and L , the circular buffer length, has been empirically set to $0.6 \times 1/20000 \times 128$ (see 2.6.2), which leads to a total gain of $\approx 2^{-8}$, implemented using a bit shifting operation. Finally, the multiplication factor G_{thr} of the threshold is set to 4, and is also implemented using a bit shifting operation.

Other Thresholding Method

The proposed algorithm is compared with the median of the absolute deviation (MAD) operator, which is employed in common spike sorting software, such as *Waveclus* an offline spike

TABLE 2.1 – System characteristics.

Parameter	Value
# of recording/stimulation channels	32/32 channels
ADC resolution	16 bits
Sampling rate per channels	20 kSamples/s
Low-pass filter	0.1 kHz - 20 kHz
High-pass filter	0.1 Hz - 500 Hz
Input referred noise	$2.4 \mu V_{rms}$
Transceiver frequency band	2.4 GHz
Transceiver measured effective data rate	1.4 Mbps
Dimensions (L x D x H, headstage w/o battery)	$17 \times 18 \times 10 \text{ mm}^3$
Weight (headstage w/o battery / w/ battery)	2.8 g / 4.9 g
Power consumption (1 LED, 150 mA @ 10%, 45 AP/s per channel)	175 mW
Autonomy (100 mAh battery, @ above parameters)	70 min
(w/o stimulation, 45 AP/s per channels)	119 mW
Autonomy (100 mAh battery, @ above parameters)	105 min

sorting freeware [91]. The MAD operator is computed as follows: $\hat{\sigma} = \text{median}\left\{\frac{|x|}{0.6745}\right\}$, where x is the band-pass filtered input signal. The threshold is determined by multiplying $\hat{\sigma}$ by a predetermined constant, which was set at four (as was the proposed algorithm). By using the 0.6745 division factor in the MAD calculation formula, $\hat{\sigma}$ act as an estimator of the signal standard deviation, making a spike sorting software like *Waveclus* well suited to assess the performance of the proposed algorithm, since both are based on the estimated signal standard deviation. However, the calculation of the median over the entire signal, is highly time and resource consuming, even for a short window of samples, rendering its implementation inside a low-power embedded system very challenging.

Comparison and Performance

Table 2.2 shows the two detector's performances for a 60 s synthetic input signal of varying SNRs at 15 dB, 10 dB, 5 dB and 0 dB respectively, and varying firing rates from 25 to 125 spikes/s. As it can be seen, the proposed method competes extremely well against the MAD operator, which is far more resource demanding. In fact, for 0 dB and 5 dB, the proposed method performs better than the other detector, as it presents a higher true detection rate, at the cost of a negligible increase in false positives of only 4.86% at 0 dB, versus 4.62% for the MAD operator. For comparison, at 0 dB, the proposed method achieved 70.23% of true positive detection, while the MAD operator achieved 69.7%. It is worthy noting that the true positive detection ratio has more importance than the false positive ratio in this application, since false detected spikes can still be discarded at the receiver by using an algorithm or by visual inspection (when such a ratio is low as with our detector), while the undetected spikes cannot be recovered.

TABLE 2.2 – Comparison of spike detectors using synthetic neural signals with varying firing rates from 25 to 125 spikes/s.

Signal		Proposed method			MAD operator		
SNR	Spikes	Detected	True	False	Detected	True	False
0	3900	2879	2739	140	2840	2718	122
5	3900	3861	3794	67	3797	3765	32
10	3901	3936	3897	39	3928	3900	28
15	3900	3917	3900	17	3913	3900	13

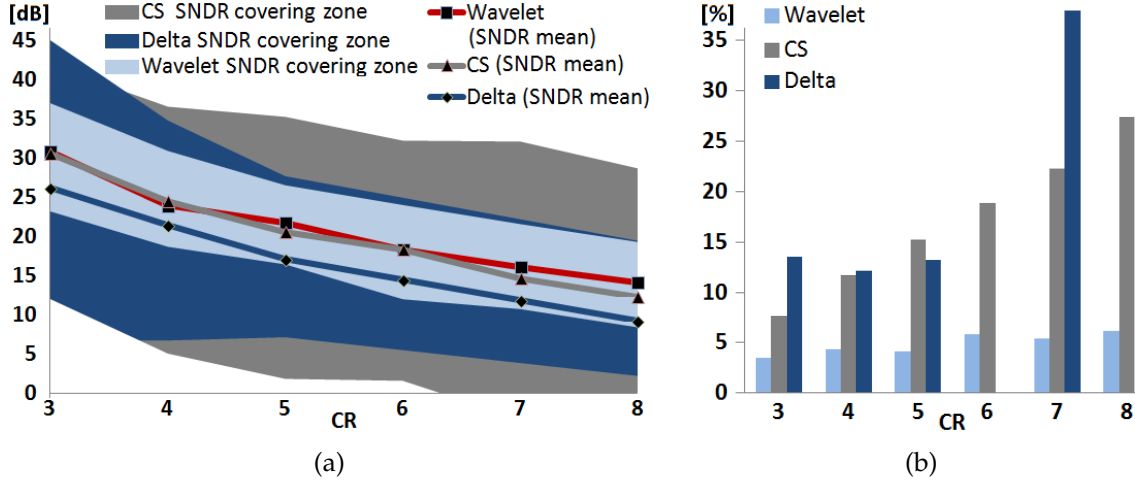


FIGURE 2.8 – Comparison with Delta compression and with CS using an optimal sparsity matrix, a) mean SNDR and SNDR variability, b) SNDR standard deviation relative to the mean SNDR.

2.7.4 Compression Performance

In order to evaluate the accuracy of the wavelet compression module, we calculated the signal-to-noise and distortion ratio (SNDR) of the compressed waveforms. The SNDR is a metric that is used to assess the compression performance and the quality of data after compression [92]. It is defined by :

$$SNDR = 20 \log \left[\frac{\|x\|_2}{\|\hat{x} - x\|_2} \right], \quad (2.6)$$

where x is the original AP waveform and \hat{x} is the reconstructed waveform.

Within Matlab simulations, which are used in this comparison to facilitate the implementation of the multiple compression modules as well as the implementation of multiple compression ratios, we used a synthetic neural signal in which 2890 spikes were detected and used to quantify the quality of the Wavelet compression module for six different compres-

sion ratios (CR) : 3, 4, 5, 6, 7 and 8. The proposed wavelet compression approach is compared with Delta compression [75] and with compressed sensing (CS), which compression methods have been previously used in biological signal acquisition systems [93], [94].

Simulations

The CS compression algorithm is implemented using a random Bernoulli matrix and the reconstruction is achieved by finding an optimal sparsity matrix using a K-SVD dictionary learning algorithm executed on 10% of the AP waveforms randomly chosen from the 2890 extracted spikes [92]. The Delta compression algorithm is achieved by reducing the quantization step of the difference between each subsequent sample using the Matlab fixed point toolbox [75].

The averaged SNDRs are presented in Figure 2.8a with respect to the compression ratio (CR) employed for each method. Both the CS and the proposed algorithm perform similarly and outperform the Delta algorithm which presents the worst performance. Even if the CS algorithm produced similar averaged SNDR to the proposed approach, it has much more variability. In fact, the SNDR of the compressed/decompressed AP using CS are sometimes under 0 dB, while the SNDR obtained with the proposed approach is always above 8.4 dB, even for a CR of 8. The poor SNDR obtained with CS arises from spikes that have slightly different wave shapes than those used to build the sparsity matrix, which situation is most likely to arise in experiments because of the variability of the experimental setting and the biological environment. Updating the sparsity matrix after some time has elapsed would drastically increase the quantity of data to handle, which represents a major drawback. Similarly, the variation obtained with Delta compression is higher than the proposed method, while the worst case SNDR is always below 10 dB for CR above 4.

Figure 2.8b shows the standard deviation of the SNDR with respect to the mean SNDR for each compression method. As expected from the preceding results, CS and Delta compression obtained the highest standard deviation for all CR, up to 6 times the amount obtained with the proposed method. These results demonstrate that the proposed wavelet compression approach yielded excellent mean SNDR for each compressed/decompressed AP leading to higher reconstruction quality. Moreover, the performance of the Wavelet approach is independent of the AP shape which is quite convenient, while other methods varied with respect to AP shapes.

Headstage Compression

In addition to the previous simulated results, we assessed the performance of the Wavelet compression module within the headstage by transmitting the detected/compressed AP

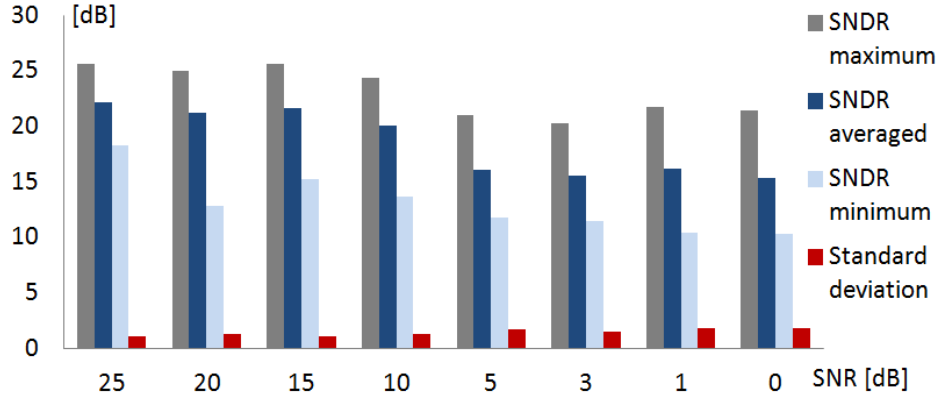


FIGURE 2.9 – Evaluation of the quality of signal compression and reconstruction of the proposed wavelet compression module for a CR of 4.17, and for input signals of SNR between 25 dB and 0 dB.

along with the raw signal. Each SNDR measurement was performed using 1500 neural spikes having varying SNRs of 25 dB to 0 dB and using a compression ratio of 4.17 (which corresponds to 20 retained coefficients on 6 bits). The calculated results are presented in Figure 2.9, and suggest that the quality of reconstruction decreases as the SNR of the signal decreases. The maximum, minimum and averaged SNDR follow the same trend while the standard deviation increases slightly as the SNR decreases. The averaged SNDR has a variation of 6.82 dB (30.7%) for input signals of SNR between 25 dB and 0 dB, respectively, which demonstrates that the proposed compression module can successfully compress the AP, even when the SNR is very low.

2.8 In-vivo Results

In order to demonstrate the functionality of the proposed headstage in a real experimental setting, we performed optical stimulations and electrophysiological recordings *in vivo*, with spike detection and wavelet data compression, with a Chr2 transgenic mouse. The experiments were performed using a test optrode featuring 8 microelectrodes and 1 high-power LED. A compression ratio of 4.17 was used, and the threshold multiplication factor (G_{thr}) was set to 4 times the estimated standard deviation. It can be shown that using a compression ratio of 4.17 along with an effective wireless data rate of 1.4 Mbps enable collecting 32 compressed neural signals simultaneously at a maximum firing rate of 210 spikes/s without loss, which fulfills the requirements of this application considering that the typical peak firing rate of a neuron is usually below 100 spikes/s while rarely exceeding 200 spikes/s [95].

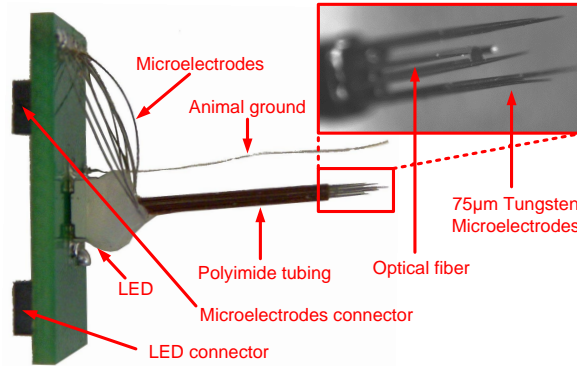


FIGURE 2.10 – Implantable module showing one LED coupled to an implantable optical fiber packaged in a polyimide tubing with 8 recording tungsten microelectrodes, one reference microelectrode and a ground wire.

2.8.1 Implantable Module

The implantable module was designed to be used for *in-vivo* experiments on a head-fixed animal. The implantable module is a mechanical part, which attaches to the headstage with two *Molex* connectors (one for the LEDs and one for the microelectrodes), mounted on the head of the rodent by surgery. The module was not cemented to the skull, as to facilitate the validation, but held in place with the headstage using an external attachment tool. A custom optrode manufactured by *Doric Lenses*, Canada, has been designed to validate the headstage within *in-vivo* trials. The optrode consists of nine 75 μm Tungsten microelectrodes (1 $M\Omega$), including one reference, a grounding wire and one LED-coupled (LB G6SP from OSRAM Opto SemiConductors) 200 μm optical fiber packaged inside a tiny polyimide tubing. Figure 2.10 shows the final prototype with a closer view on the tips of the optrode. It should be noted that the minimum data rate required to transmit the signals collected from 8 microelectrodes at full resolution (16 bits) is of 2.56 Mbps, which is not reachable by most commercial low-power transceivers, at least without performing signal processing and data compression *in situ*.

2.8.2 Animal Preparation

A transgenic mouse line expressing Channelrhodopsin-2 was used to record/stimulate from the brain surface down to 1 mm depth (medio-Lateral 1.8mm, Anteroposterior -1.8mm) for somatosensory cortex and 2.5 mm depth (mediolateral 1.8mm and Anteroposterior -1.8mm) for hippocampus (Figure 2.11). The animal preparation was performed with an adult mouse (30-50 g) anesthetized with ketamine/xylazine (10 :10 mg/kg) and placed into a mouse stereotaxic apparatus. Mouse line (Thy1 : :ChR2-YFP line 4) [96] were purchased from *The Jackson Laboratory*. The mouse was maintained in deep anesthesia and its body temperature was maintained at 37.5°C. To allow access to brain tissue with the optrode, a craniotomy was

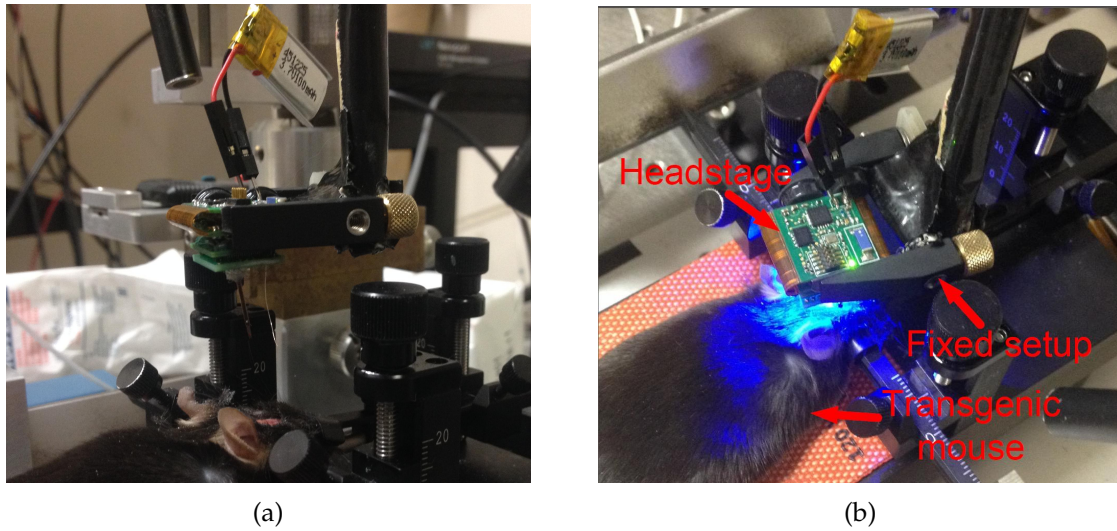


FIGURE 2.11 – a) The head-fixed *in-vivo* experimental setting, b) The data transmission was propagated between the headstage and a base station situated 2 meters away.

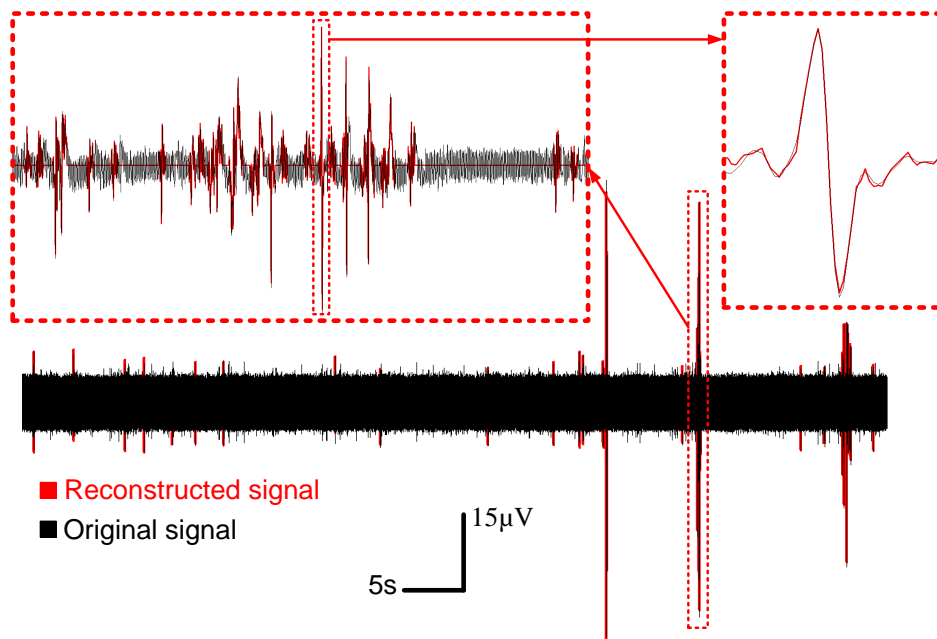


FIGURE 2.12 – Reconstructed neural signal using the detected and decompressed AP superimposed on the original signal. A zoom over a detected spike shows that the reconstructed signal is nearly identical to the original waveform.

performed and the dura was removed. The electrode and the fiber optic were lowered into the cerebral cortex using a micromanipulator. All protocols were performed in accordance with Canadian Council on Animal Care Guidelines.

TABLE 2.3 – Comparison with other wireless headstages.

System	Type	# rec./stim. channels	Spike detect.	Data compression	Precision (bits)	Sampl. (kbps)	Weight (g)	Power (mW)	Transceiver type
This work	COTS	32/32	Yes	Wavelet	16	20.0	2.8	119 (w/o stim)	Commercial
[97]	COTS	0/16	No	None	N/A	N/A	2.0	N/A	N/A
[81]	COTS	2/2	No	None	12	20.0	4.9	57.5 (w/o stim)	Commercial
[66]	COTS	2/1	No	None	10	20.0	7.4	94.5 (w/o stim)	Commercial
[78]	COTS	96/0	Yes	Low-quantiz.	8	31.3	17	100	Commercial
[79]	ASIC/COTS	8/0	Yes	AP features	12	16.7	23	4.62	Commercial
[74]	ASIC/COTS	64/0	No	None	N/A	20.0	12	645	ASIC
[73]	ASIC	15/0	No	None	16	N/A	4.0	N/A	ASIC
[77]	ASIC	100/0	No	None	12	20.0	Chip	N/A	ASIC
[76]	ASIC	128/0	Yes	AP features	9	40.0	Chip	6.00	ASIC
[83]	ASIC	64/0	Yes	AP timings	8	24.0	Chip	0.130	N/A
[82]	ASIC	32/0	No	Wavelet	12	25.0	Chip	0.0760	N/A

2.8.3 Multichannel Spike Detection and Compression Results

Spike detection and signal compression performances are validated experimentally by transmitting the detected and compressed spikes along with the raw signal for test and evaluation purposes, using the headstage implanted *in-vivo*. The simultaneous transmission of compressed spikes and original waveforms were limited to one channel/electrode at a time because of the limited transmitter data rate. Figure 2.12 shows 60 seconds of a reconstructed signal (in red) overlapped with the raw signal (in black). The averaged SNDR has been calculated for all detected, compressed and reconstructed spikes as well as for the raw waveforms. It was found that the *Wavelet compression* module yielded an averaged SNDR of 17.0 dB, which is equivalent to the results previously reported in Section 2.7.4 for a synthetic neuronal signal having an SNR varying between 10 dB and 5 dB. The spike detector was evaluated by comparing its performance with *Waveclus*, for a detection task performed offline on the raw recorded neural data. Such a comparison makes it possible to evaluate the accuracy of the standard deviation calculated by the system versus the one obtained by *Waveclus* using the whole *in-vivo* waveform. For a same threshold gain of 4 times the standard deviation for both systems, the proposed detector achieved 92.4% of matching detection compared with *Waveclus*, suggesting that the standard deviation estimated by the headstage is similar to the one calculated by *Waveclus* over the entire signal. Additionally, the proposed detector demonstrated more accuracy by detecting spikes of smaller amplitudes which were not detected by *Waveclus* in a proportion of 9.87%.

Signal detection and data compression enabled collecting neural data simultaneously from 8 microelectrodes attached to the implantable module. The detected and sorted spikes from a 200-second recording are presented in Figure 2.13, with the averaged spikes showing for

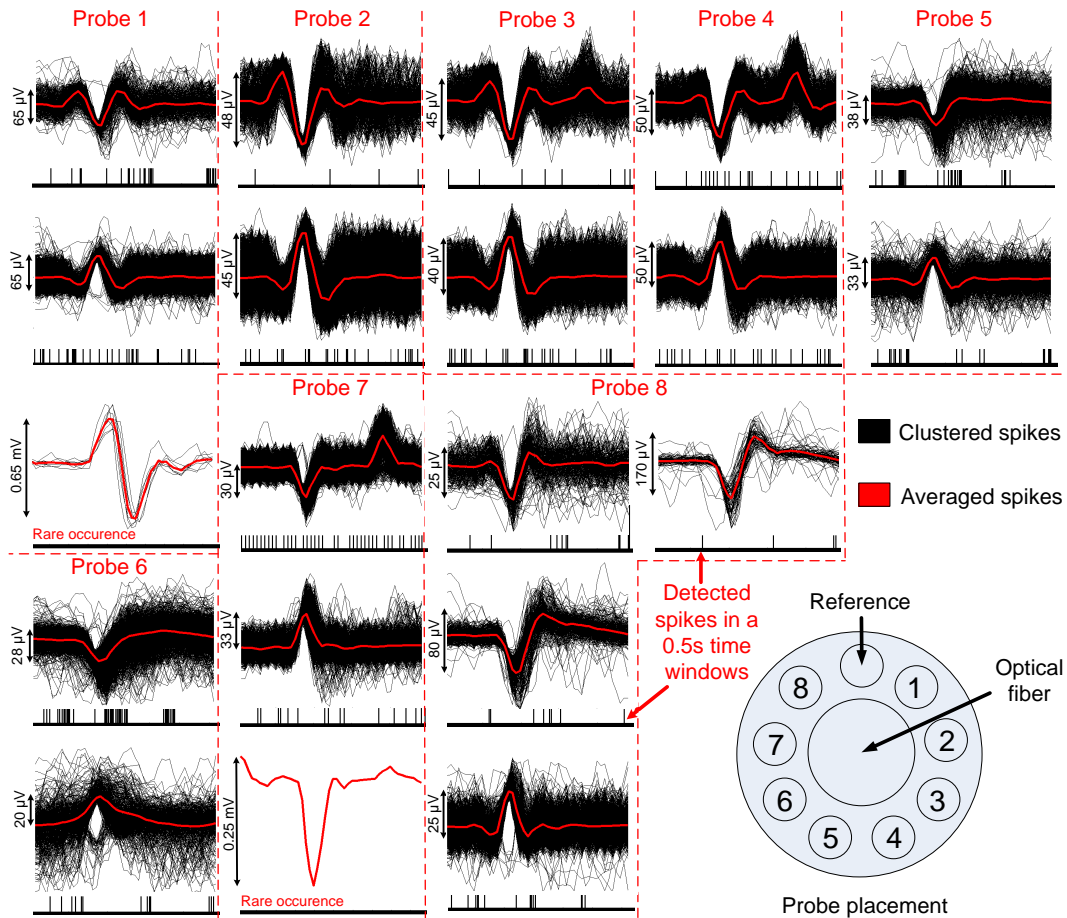


FIGURE 2.13 – Detected and clustered spikes recorded simultaneously from 8 microelectrode during a 200-second experiment.

each cluster and spike time occurrence in a 500 ms windows. Spike sorting was performed offline using a PCA followed by a k-means clustering step, which associated each spike to a specific neuron. As it can be seen, some microelectrodes, like probe 1 and 7, collected AP waveforms that occurred less frequently, while some other microelectrodes like probe 2, 3 and 4, recorded spikes of similar shapes but presenting different amplitudes and times of occurrence. Similarly, probe 5 and 6 recorded similar spike shapes of different amplitudes, but having common times of occurrence, suggesting that both these microelectrodes were closer to the firing neurons than other microelectrodes (see the optrode configuration in Figure 2.13). The overall compression ratio achieved for this recording session was 508, including the data reduction introduced by the spike detector multiplied by a compression factor of 4.17.

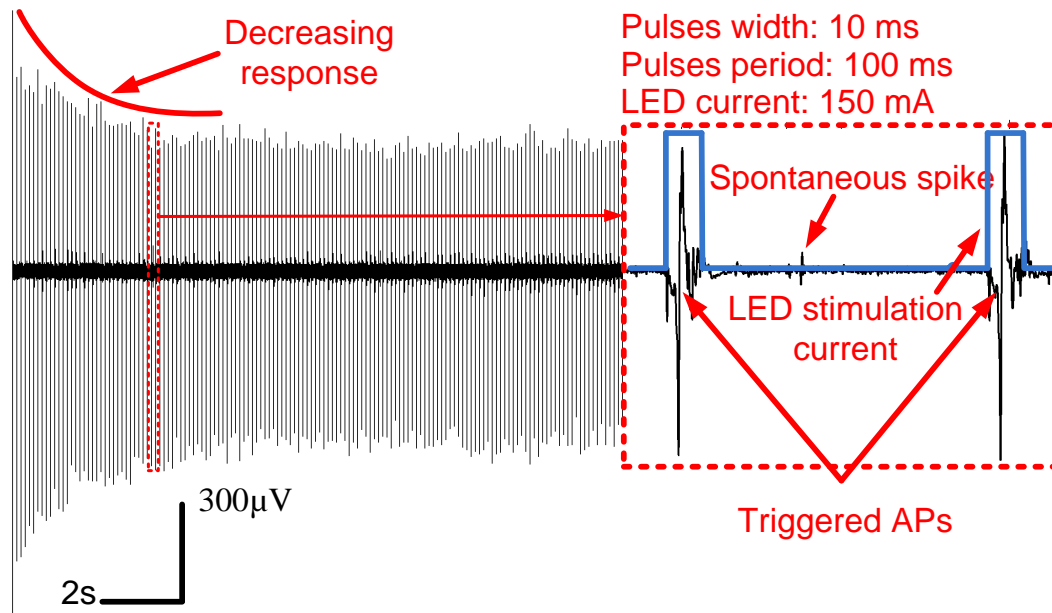


FIGURE 2.14 – Experimental results showing extracellular recordings in the brain of a ChR2 mouse in response to local optogenetic light activation of 10 ms with a period of 100 ms. AP with amplitude of up to 1 mV were evoked by the light activation, and a decreasing amplitude is observed at the beginning of the stimulation train, until stabilization after 5 seconds.

2.8.4 Optical Stimulation Results

Optical stimulation patterns were programmed using our custom software. The stimulation pulse width was set to 10 ms based on previous experimental parameters [98], [99] while the LED current was adjusted to 150 mA, giving a measured light intensity of 70 mW/mm^2 (at 465 nm) at the tip of the implantable fiber. Figure 2.14 presents the neural activity measured by one microelectrode while optical pulses were delivered to neurons in the animal's brain through the implanted fiber. Experimental results demonstrate that the proposed headstage is able to collect and transmit single spike events evoked by light activation, with amplitudes of up to 1 mV, as well as spontaneous spikes. In Figure 2.14, a decreasing AP amplitude is observed at the beginning of the stimulation train, until stabilization occurs after 5 seconds. This spike silencing phenomenon is due to a depolarization block, as described in [99].

2.9 Comparison

The proposed headstage is compared to recently published low-power wireless headstages dedicated to neural recording and/or optical stimulation applications as presented in Table 2.3. As illustrated, the proposed system possesses the largest number of optical stimulation channels, in addition to providing multichannel neural recording, and is one of the lightest.

Furthermore, the proposed system collects entire spike waveforms [78], [82], while other systems with spike detection only send minimal spike characteristics such as duration and amplitude [79], [76], [83]. The proposed headstage uses COTS components only, while other systems capable of sending complete waveforms in parallel from multiple channels use custom integrated transceivers and ASIC technology primarily [73], [74], [82], [83], [76], [77], leading to higher costs (for low and medium production volumes) and less flexibility. Moreover, the proposed headstage has one of the lowest reported weights, which is critical to enable experiments with small freely moving rodents.

2.10 Conclusion

We presented a new and unique miniature, wireless multichannel optogenetic headstage suitable for both optical stimulation and electrophysiological recording from 32 channels using spike detection and wavelet compression. This is the first attempt towards high resolution neural recording along with optogenetics. The proposed headstage can activate the neurons using light pulse trains generated by up to 32 LEDs, and simultaneously collect the evoked neural activity from 32 recording channels. Simulation and measurement results show that the spike detector and the data compression module can detect and compress the AP with a high compression ratio and excellent accuracy. Further, the proposed headstage has been validated *in-vivo* using a custom optrode including one implanted optical fiber and 8 microelectrodes implanted into the brain of a transgenic Channelrhodopsin-2 mouse. It succeeded in collecting and transmitting the neural activity elicited by one implantable optical stimulation fiber. In future research, we intend to use the proposed headstage in combination with wireless power transmission [100] as to reduce further the weight, size and limit manipulations of the animal.

Chapitre 3

A 0.13 μm CMOS SoC for Simultaneous Multichannel Optogenetics and Neural Recording

© 2019 IEEE. Reprinted, with permission, from G. Gagnon-Turcotte, M. Noormohammadi Khiarak, C. Ethier, Y. De Koninck and B. Gosselin, "A 0.13- μm CMOS SoC for Simultaneous Multichannel Optogenetics and Neural Recording", *Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3087-3100, Nov. 2018.

3.1 Résumé

Cet article présente un système sur puce (SoC) conçu en technologie 0.13 μm CMOS qui permet d'effectuer de l'optogénétique multicanal et de l'enregistrement neuronal multicanal dans le cerveau d'animaux de laboratoire se déplaçant librement. Ce système intégré possède 10 canaux d'enregistrement multimodal avec conversion analogique numérique (ADC) et 4 canaux pour la stimulation optogénétique. La conception du bioamplificateur inclut une bande passante programmable (0.5 Hz - 7 kHz), afin de collecter les potentiels d'action (AP) et/ou les potentiels de champs (LFP), un facteur d'efficacité de bruit (NEF) de 2.30 pour un bruit referré à l'entrée de 3.2 μV_{rms} dans une bande passante (BW) de 10 à 7 kHz. L'ADC Delta-Sigma ($\Delta\Sigma$) MASH 1-1-1 à faible consommation est conçu pour fonctionner avec de faibles ratios de suréchantillonnages ($\text{OSR} \leq 50$) et possède un nombre de bits effectif (ENOB) de 9.75 bits avec un OSR de 25 (BW de 10 kHz). L'utilisation d'un ADC de type $\Delta\Sigma$ est la clé afin d'optimiser le compromis entre le bruit et la consommation énergétique dans différents paramètres expérimentaux. Il exploite une nouvelle technique qui réduit sa taille en soustrayant la sortie de chaque branche du $\Delta\Sigma$ dans le domaine numérique, plutôt que dans le domaine analogique, comme cela est fait de manière classique. L'ADC est suivi d'un filtre de décimation (DF) du 4^{ème} ordre de type *cascaded integrator-comb* (CIC), implémenté sur puce. Un canal d'enregistrement complet, comprenant le bioamplificateur, le $\Delta\Sigma$

MASH 1-1-1 et le DF consomme 11.2 μW . La stimulation optique est réalisée avec un circuit de commande de DEL qui utilise une source de courant de type cascode avec rétroaction, ce qui permet d’accommoder une large gamme de paramètres pour la DEL, ainsi qu’une large gamme de tension de batterie. Le SoC est intégré dans un prototype de plateforme optogénétique et validé *in vivo* lors de deux expériences, la première dans le noyau postéro-médian ventral (VPM), et la seconde dans le cortex moteur cérébral d’un rat exprimant la protéine Channelrhodopsin-2 (ChR2) par induction virale.

3.2 Abstract

This paper presents a 0.13 μm CMOS system-on-chip (SoC) for simultaneous multichannel optogenetics and multichannel neural recording in freely moving laboratory animals. This fully integrated system provides 10 multimodal recording channels with analog-to-digital conversion, and a 4-ch LED driver circuit for optogenetic stimulation. The bio-amplifier design includes a programmable bandwidth (0.5 Hz - 7 kHz) to collect either action potentials (AP) and/or local field potentials (LFP), and has a noise *efficiency factor* (NEF) of 2.30 for an input referred noise of 3.2 μV_{rms} within a bandwidth (BW) of 10 to 7 kHz. The low-power Delta-Sigma ($\Delta\Sigma$) MASH 1-1-1 ADC is designed to work at low oversampling ratios (OSR ≤ 50) and has an effective number of bits (ENOB) of 9.75 bits at an OSR of 25 (BW of 10 kHz). The utilization of a $\Delta\Sigma$ ADC is the key to address the flexibility needed to address different noise vs power consumption trade-off of various experimental settings. It leverages a new technique that reduces its size by subtracting the output of each $\Delta\Sigma$ branch in the digital domain, instead of in the analog domain as done conventionally. The ADC is followed by an on-chip 4th order cascaded integrator-comb (CIC) decimation filter (DF). A whole recording channel, including the bio-amplifier, the $\Delta\Sigma$ MASH 1-1-1 and the DF consumes 11.2 μW . Optical stimulation is performed with an LED driver using a regulated cascode current source with feedback that can accommodate a wide range of LED parameters and battery voltages. The SoC is validated *in vivo* within a wireless experimental platform in both the ventral posteromedial nucleus (VPM) and cerebral motor cortex brain regions of a virally mediated Channelrhodopsin-2 (ChR2) rat.

3.3 Introduction

Optogenetics and multi-unit electrophysiological recording are state-of-the-art approaches in neuroscience for observing neural microcircuits *in vivo* [1]. On the one hand, optogenetics is used to activate or deactivate light sensitized neurons with high temporal and spatial resolution, for precise control. This ground breaking experimental approach can elicit or inhibit neuron depolarization using light through the insertion of light-sensitive proteins into the cell membrane using transgenic modifications or viral injections [31]. For instance,

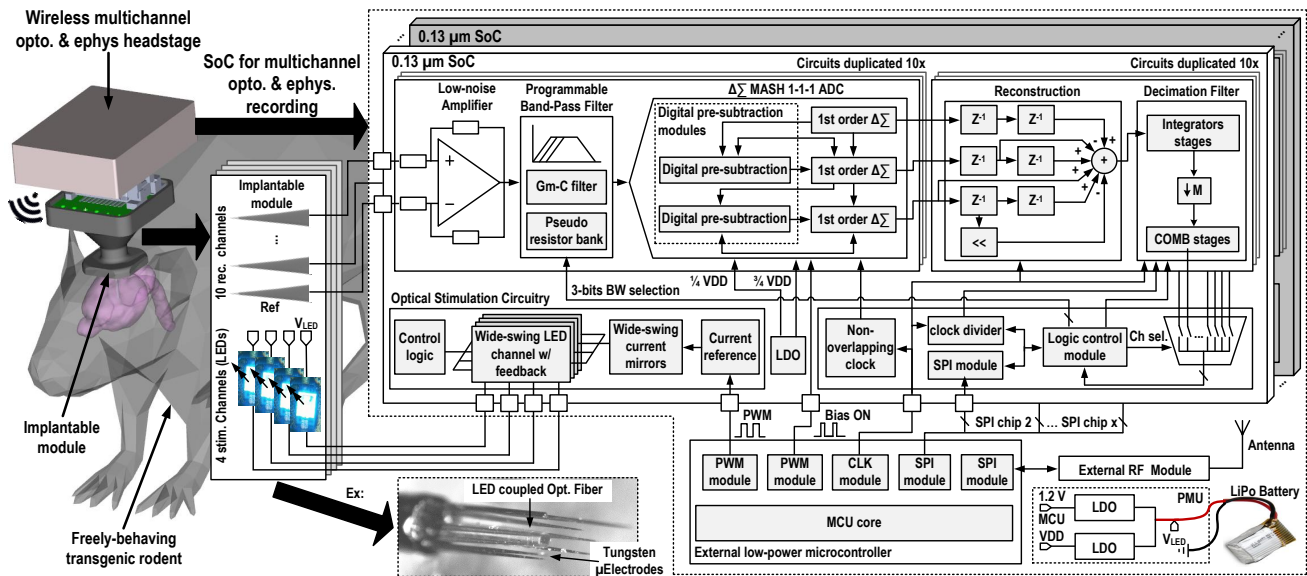


FIGURE 3.1 – System-level concept of a wireless optogenetic platform utilizing the proposed SoC, including all the building blocks, to perform simultaneous neural recording and optical stimulation. The 0.13 μm CMOS SoC includes 10 neural recording channels and 4 optical stimulation channels. Each recording channel encompasses a low-noise bio-amplifier, a $\Delta\Sigma$ MASH 1-1-1 ADC and a decimation filter. The chip also includes a 4-ch LED driver circuit with feedback to precisely set the effective forward current of each LED for robust optical stimulation. It communicates through an SPI bus for easy interfacing and scalability.

Channelrhodopsin-2 (ChR2) is a protein that controls the ionic conductance of the neurons when illuminated with a blue visible light at a wavelength of around 465 nm. On the other hand, multi-unit electrophysiological recording is used in parallel with optogenetics to monitor the electrical connections established in the brain through ionic flow of ions after optical stimulation. Combining these two approaches in a miniature wireless brain-implantable system holds substantial potential for advancing our understanding of brain functions and neurological disorders, by allowing for studies with small freely moving laboratory animals [20, 81, 101–105], without cable tethering.

Different commercial platforms and research prototypes have been dedicated to performing optogenetics and/or electrophysiological recording in freely moving laboratory animals [20, 81, 101–105]. Such platforms typically consist of a wireless electronic system (often referred to as a headstage) connected to a passive implantable module including microelectrodes and embedded light sources, such as LED [20, 81, 103, 104]. These LEDs are often coupled with implantable optical fibers to probe into deep brain structures. The design of a suitable wireless headstage faces important challenges : 1) the size and the weight must be small enough to be carried on the head of the laboratory mouse or rat, 2) it must use minimal energy to avoid carrying a bulky power source, 3) it must collect both action potentials (AP) and local field potentials (LFP), which is required to understand how the LFP relates to

spikes [106], and 4) it must include simultaneous recording and optogenetic features to be compatible with newly available optrodes [107].

Hence, brain-implantable devices incorporating optical stimulation and low-noise data acquisition circuits have been designed based on custom integrated circuits (IC) to avoid bulky commercial off-the-shelf components and to scale down the embedded electronics for looking at more neurons. Several ICs have proposed optimized neural amplifier designs [108–115], focusing on increasing the number of recording channels [108, 109], on reducing power consumption [109–115] or on improving noise performance [61, 110–118]. Optical stimulation is often provided in these systems using CMOS LED drivers that are based on current mirrors [119] or on capacitor discharging techniques [120]. ICs combining neural recording and optical stimulation are presented in [121] and [122]. The system described in [121] is limited to one optical stimulation channel and to one neural recording channel collecting AP only, while the IC reported in [122] can only collect LFP, and was not validated *in vivo*, in a real experimental setting. Hence, no demonstrated IC provides multichannel optogenetic photo-stimulation along with multiunit electrophysiological recording capability within the same die. Existing ICs [61, 108–115] and [120–123], also lack critical features : 1) they are not multichannel and/or they do not include an analog-to-digital converter (ADC) [108, 110, 111, 113, 115] and [121]; 2) they address only one signal modality [61, 109, 121] and [122], i.e., either LFP or AP, not both; 3) they have not been demonstrated within *in vivo* experiments or inside a fully functional system [121]; and 4) they are not easily interfaceable with readily available electronic systems.

In this paper, we present a new 0.13 μm CMOS system-on-chip (SoC) for simultaneous multichannel optogenetics and multimodal electrophysiological brain recording (e.g., for both LFP and/or AP signals). The proposed SoC includes 10 low-power/noise neural recording channels and a precise 4-channel LED driver for optogenetics stimulation. The chip is integrated inside a wireless optogenetic platform, which is utilized to demonstrate the performance of the proposed SoC within *in vivo* experiments. Simultaneous optical stimulation and electrophysiological recordings are performed in the brain of a rat with virally mediated Channelrhodopsin-2 (ChR2) cortical expression. The paper is organized as follows : Section 4.4 presents an overview of the proposed SoC and each building block, Section 3.5 presents the circuit implementations and the SoC integration within a wireless platform, Section 3.6 presents the experimental results, while the *in vivo* results are included in Section 4.11, Section 4.12 presents a comparison with the state-of-the-art ICs, and Section 4.13 presents the conclusions.

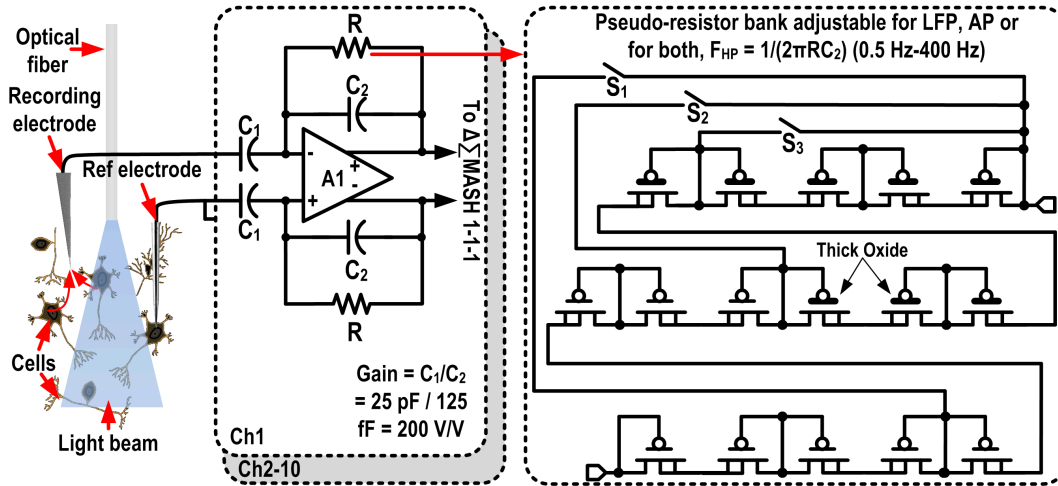


FIGURE 3.2 – Schematic of the bio-amplifier design with programmable high-pass cutoff frequency. The feedback resistor uses a 3-bit pseudo-resistor bank.

3.4 System Architecture

3.4.1 System Overview

Fig. 3.1 shows the proposed SoC for simultaneous multichannel neural recording and optogenetics, enclosed within a wireless experimental platform. This SoC includes key circuits for electrophysiological recording and optogenetics. Each multimodal neural recording channel includes a low-noise bio-amplifier with programmable bandwidth (BW), a Delta-Sigma ($\Delta\Sigma$) MASH 1-1-1 ADC with programmable resolution and power and a DF. The $\Delta\Sigma$ MASH 1-1-1 employs a bias duty-cycling strategy that decreases its power consumption and a new digital subtraction technique, located between the $\Delta\Sigma$ output branches, that simplifies its implementation. This design offers an advantageous alternate solution to the widely used successive approximation register (SAR) ADC by trading analog complexity for simple logic gates, and by enabling a programmable precision-versus-power consumption trade-off, through a simplified digital programming interface. The decimation filter (DF) is a 4th-order cascaded integrator-comb (CIC) filter with minimal binary precision between stages to reduce the circuit size. On the same die, optical stimulation is performed by a 4-ch LED driver circuit with feedback that precisely sets the effective forward current of each LED. The SoC communicates through a serial peripheral interface (SPI) bus, allowing for easy communication with the on-chip *control module*. Each building block is detailed in the next paragraphs.

3.4.2 Bio-Amplifier Design

Fig. 3.2 shows the schematic of the single-stage fully-differential AC-coupled bio-amplifier utilized in this SoC, and inspired from [113]. In this design, the ratio of capacitor C_1/C_2 sets the mid-band gain A_m , and the high-pass cutoff frequency is determined by the resistor

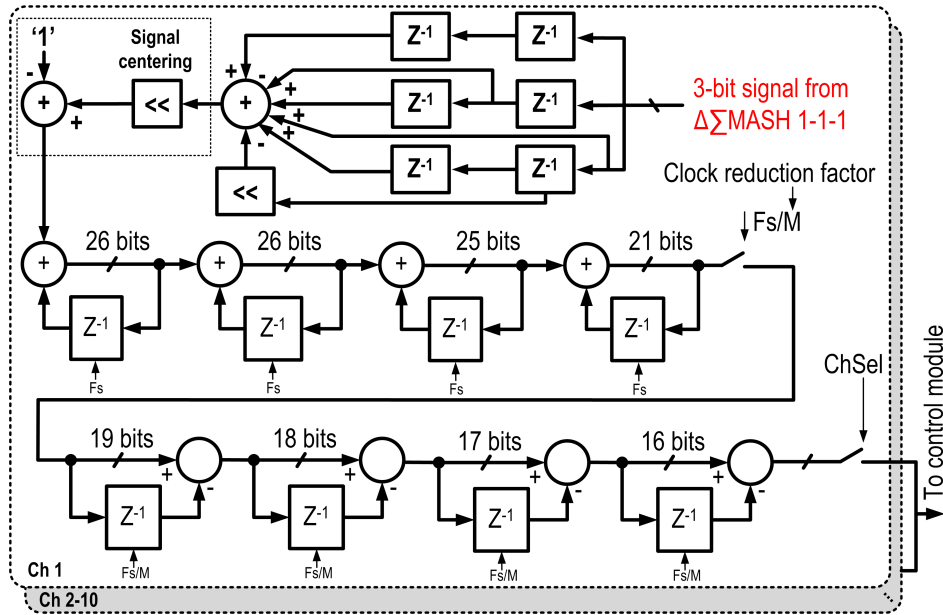


FIGURE 3.3 – CIC4 decimation filter with optimized binary precision between the stages, for up to 14-bit output resolution. The low-pass cut-off frequency can be programmed simply by changing the clock reduction factor M and/or by modifying the oversampling frequency.

R and the capacitor C_2 , so $F_{HP} = 1/2\pi RC_2$. The feedback resistors R use a 3-bit tunable pseudo-resistor bank, allowing the selection of a given high-pass BW for collecting either LFP and/or AP signals. The low-pass analog cutoff frequency is set by the loading capacitance C_L and the transconductance G_m of the operational transconductance amplifier (OTA), i.e., $F_{LP} = G_m/(2\pi A_m C_L)$. The circuit implementation details are discussed in Subsection 3.5.1

3.4.3 $\Delta\Sigma$ MASH 1-1-1

Typically, the bio-amplifier IRN, the total harmonic distortion (THD) and the unpredictable background neuronal noise put a limit on the maximum achievable resolution of the whole system. For instance, the thermal noise floor can be as high as $10 \mu V_{p-p}$, or above, for spikes recordings w/ metal electrodes [20, 121], leading to a maximum ENOB of 9 bits assuming an input dynamic range of $6 mV_{p-p}$. Previous work targeting both spikes and/or LFP reported resolutions of 8-bits [112] and 10-bits [109, 122]. Thereby, we designed a $\Delta\Sigma$ MASH 1-1-1 neural ADC topology that meets the required resolution for this application (~ 10 bits [109, 122]) with low OSR (≤ 25). Specifically, using an oversampling ADC in this application provides the following key advantages : 1) the flexibility to adjust the ENOB and the power consumption with respect to the requirements and the variability of the application (type of target signals to collect, noise level and type of electrodes (eg. : glass micropipette, metal or silicon electrodes, etc.), or SNR degradation reported in chronically implanted electrodes

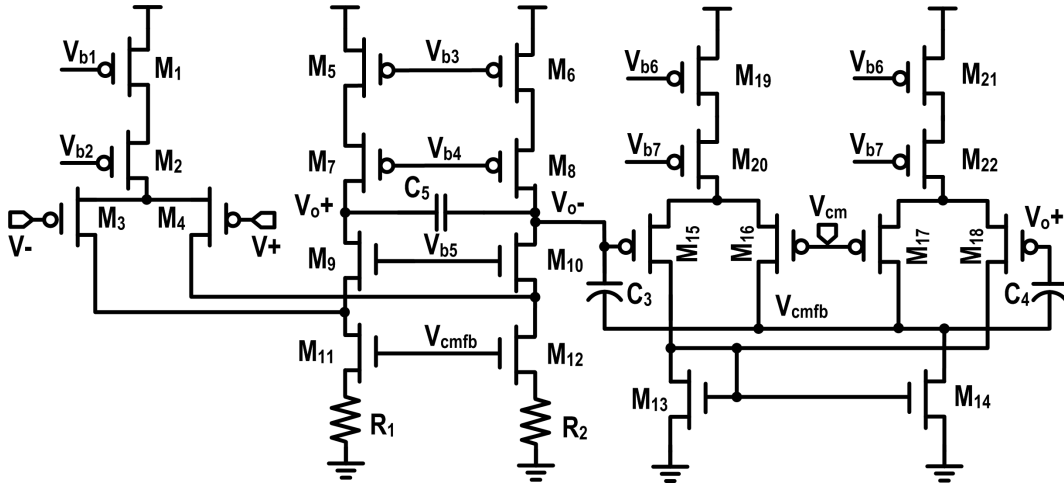


FIGURE 3.4 – Circuit schematic of the OTA used for the bio-amplifier and its continuous time CMFB. The fully-differential folded-cascode OTA with source-degeneration resistors R_1 and R_2 is used.

[124], 2) a programmable low-pass cutoff frequency, performed implicitly by the DF, 3) the relaxation of the bio-amplifier anti-aliasing filter constraints by pushing the Nyquist's frequency away from the in-band signal through oversampling, 4) the stability and robustness at full-scale, thanks to the 1st order $\Delta\Sigma$ s in the MASH, an important requirements in biomedical applications in general, and 5) the opportunity to use low OSR for enabling low-power, easy interfacing, and the utilization of an on-chip DF, thanks to the 3rd order noise shaping characteristic. The circuit implementation details are discussed in Subsection 3.5.2

3.4.4 CIC4 Decimation Filter

Fig. 3.3 shows the block diagram of the 4th order cascaded integrator-comb (CIC4) DF utilized in this design. In contrast to [61] and [123], using an on-chip DF provides programmable precision and adjustable bandwidth simply by varying the OSR and the clock reduction factor M (see Fig. 3.3) depending on the bio-amplifier performance and type of signal (i.e., LFP or AP). This filter uses a highly optimized CMOS implementation strictly based on adders and registers, avoiding power-hungry multipliers and dividers. For using as few output bits per stage as possible without degrading the precision, the minimal binary precision of each stage is determined using the Hogenauer pruning technique, to accommodate $OSR \leq 50$. Since the DF has a rather smooth cutoff frequency, it is designed to provide an attenuation of 15 dB at the Nyquist frequency (F_{NF}) after decimation, by setting M equal to the OSR value. The transfer function of the DF is :

$$H(z) = A_{PR} \cdot (1 - z^{-M})^4 / (1 - z^{-1})^4, \quad (3.1)$$

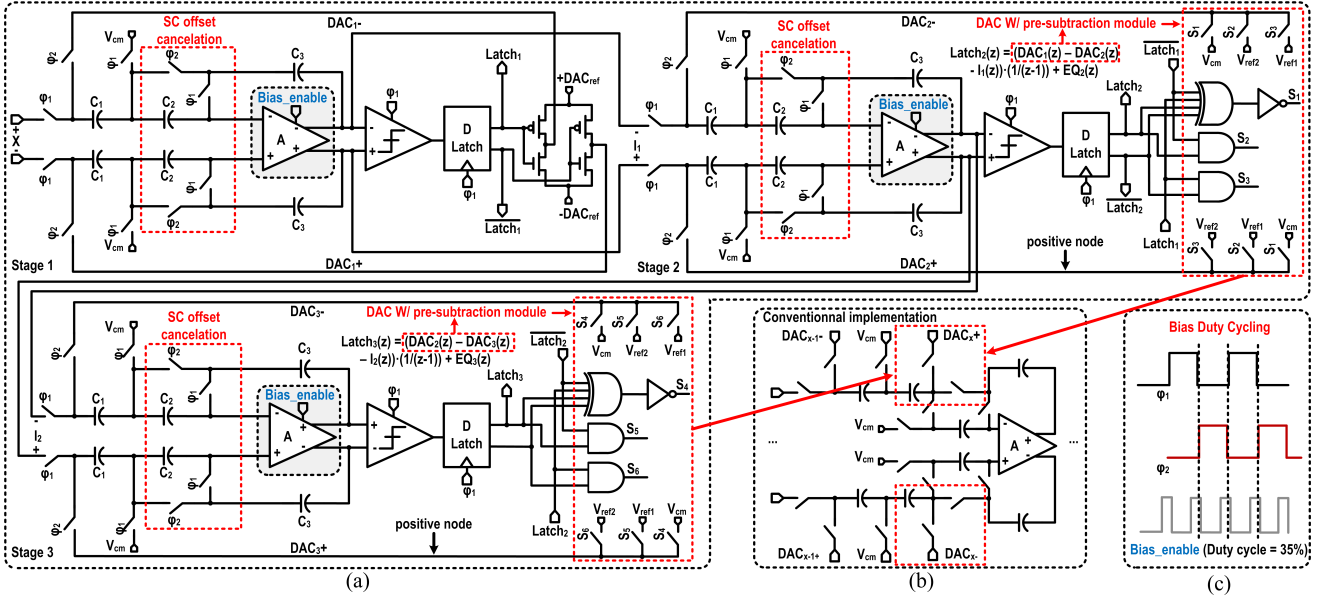


FIGURE 3.5 – a) Schematic of the $\Delta\Sigma$ MASH 1-1-1 using a digital *pre-subtraction module* for the 2nd and 3rd stages, b) Conventional implementation of the subtraction between the DAC outputs using switched capacitors, c) Bias duty-cycling strategy used to turn off the integrator OTA during both clock phases.

where M is the clock reduction factor between the integrator and comb stages, and A_{PR} is the gain reduction induced by pruning. This gain is defined as $A_{PR} = 1/2^{(\sum_{k=0}^7 B_{p,k})-1}$, where $B_{p,k}$ is the number of bits pruned at stage k , thereby $\sum_{k=0}^7 B_{p,k} = 10$. The time-domain low-pass cutoff frequency relates to the OSR by: $F_{NF} = F_s/2 \cdot OSR$. Thus, the advantage of using an on-chip DF is clear, since it allows to implement a programmable low-pass filter simply by changing F_s , the OSR or the M factor.

3.4.5 LED Driver

The LED driver is a crucial element in an optogenetic stimulation system because it determines the amount of optical power delivered to the neurons. We developed a 4-ch LED driver that precisely regulates the LED current and that is robust against the LED parameter and the battery voltage variations. The LED current is modulated by an external PWM signal, and each channel can be independently turned ON or OFF, by sending the LED configuration bits using a custom serial digital input. The circuit implementation details are discussed in Subsection 3.5.3

3.4.6 Control Logic Module

The control logic module allows the SoC to transfer the ADC data, to configure the M parameter in the decimation filter, and to program the high-pass cutoff frequency of the bio-amplifier. This circuit consists of an SPI *slave module* and a finite state machine. The SPI *module*

is designed to transfer bytes, to operate with a null clock phase and polarity, and to work at frequencies below 20 MHz. It is possible to accommodate ~ 60 recording channels (6 SoC) in parallel on the same SPI bus, providing a sampling rate of 20 kHz after decimation for each recording channel.

3.5 Circuit Implementation and SoC Integration

3.5.1 Bio-Amplifier

The bio-amplifier circuits and design choices required to meet all the requirements for this application (noise, BW, power, ect.) are described in the following paragraph.

Input Capacitor

The bio-amplifier input capacitor C_1 of Fig. 3.2 must be selected carefully to provide a large impedance for reducing the effect of the voltage divider created by the input capacitance of the OTA and for avoiding any input signal attenuation. Hence, this capacitor largely determines the input referred noise (IRN) of the closed loop bio-amplifier [113]. The mid-band gain of the bio-amplifier is set to ~ 46 dB by using $C_1 = 25$ pF, which are implemented as double metal insulator metal (MIM)-Caps having 2.05x more capacitance per area than usual MIM-Caps (i.e. double MIM-Caps benefits from additional dielectric layers and terminals), so the effect of the capacitive input divider is limited to $\sim 3.1\%$.

Pseudo-Resistor Bank

The feedback pseudo-resistor bank of Fig. 3.2 consists of PMOS pseudo-resistors whose gates are connected in a back-to-back configuration through CMOS transmission gates. These pseudo-resistors allow for equivalent resistances in the Tera-Ohm range, which is required to implement very-low high-pass cutoff frequencies down to a few mHz for LFP recording. In this design, such a low cutoff frequency is accomplished by using pseudo-resistor structures with thick oxide transistors providing small subthreshold leakage current, high threshold voltage and low oxide capacitance per unit of area compared with equally sized transistors of standard thickness.

OTA

The bio-amplifier OTA, the schematic of which is shown in Fig. 3.4, consists of a fully differential folded-cascode amplifier with source-degeneration resistors to reduce thermal noise [111], providing an open-loop gain of 66 dB. This topology has a higher gain than a single stage amplifier while offering a straightforward and robust implementation, which is critical when using a deep submicron CMOS process. Moreover, a source-degeneration resistor topology effectively reduces the IRN, while being simple to implement and bringing only

small overhead. Its main drawback is a potential reduction in the dynamic range, but it can be shown to have practically no impact in this design since the input of the $\Delta\Sigma$ is bounded between $-\text{DAC}_{\text{ref}}$ and $+\text{DAC}_{\text{ref}}$ (as seen in Section 3.4.3), where the swing reduction has nearly no impact within this range. Whereas several topologies focus on decreasing the noise efficiency factor (NEF) [113], which is often done at the expense of another design parameter (THD [117], common-mode rejection ratio (CMRR) [117], inter-channel cross-talk [116], etc.) or complexity [118], the OTA topology utilized in this design is robust and simple, while it provides a good trade-off among all the design parameters. Compared with [111], we use a fully differential design, which increases the common-mode rejection ratio (CMRR), improves the dynamic range by 6 dB, and reduces the second-order harmonics, at the expense of adding a small common mode feedback circuit (CMFB) circuit. The $1/f$ noise is reduced by using large PMOS input transistors in the differential pair, and the IRN is expressed by :

$$\overline{V_{ni,OTA}^2} = \frac{8kT}{g_{m3}^2} \left(g_{m3} + \frac{2}{R_{1,2}} + \frac{4}{3}g_{m5} \right), \quad (3.2)$$

where k is the Boltzmann constant, T is the temperature in Kelvins, and g_{m3} and g_{m5} are the transconductance of transistors M_3 and M_5 , respectively, in Fig. 3.4. The flicker and the thermal noise contribution of M_{11} and M_{12} are minimized through the use of source degeneration resistor $R_{1,2}$, as long as the condition $g_{m11,12} \cdot R_{1,2} \gg 1$ is met [111]. Thus, if $R_{1,2}$ is big enough, the noise contribution of the resistor to the output noise is better than the transistor pair M_{11} and M_{12} . In our implementation, we were able to achieve an improvement of 7% of the IRN by using two 40-k Ω resistors for R_1 and R_2 .

3.5.2 $\Sigma\Delta$ MASH 1-1-1

The circuit schematic of the $\Delta\Sigma$ MASH 1-1-1 neural ADC is shown in Fig. 3.5(a). Each $\Delta\Sigma$ stage in the proposed MASH 1-1-1 topology requires an OTA for the integrator, a one-bit quantizer, a DAC for the 1st stage and a pre-subtraction module for the 2nd and 3rd stages. These circuits are described in the following paragraphs, along with the bias duty-cycling strategy that is used to turn off the OTA bias current for saving power during both clock phases φ_1 and φ_2 (Fig. 3.5(c)).

Pre-Subtraction Module

As seen in Fig. 3.5(a), the first $\Sigma\Delta$ MASH stage uses a 1st order discrete time (DT) $\Delta\Sigma$ modulator, which is receiving the fully differential conditioned output signals of the bio-amplifier, while the 2nd and 3rd stages are two 1st order DT $\Delta\Sigma$ modulators fed by the digital-to-analog converter (DAC) output minus the integrator output from their respective previous stages ($\text{DAC}_{x-1}(z) - V_{x-1}(z)$). As a result, each stage cancels out the quantization noise induced by its previous stage, when applying the proper digital correcting functions to each $\Delta\Sigma$ output (z^{-2} , $-z^{-1}(1 - z^{-1})$ and $(1 - z^{-1})^2$ respectively) [125], as seen in Fig. 3.3. The resulting error

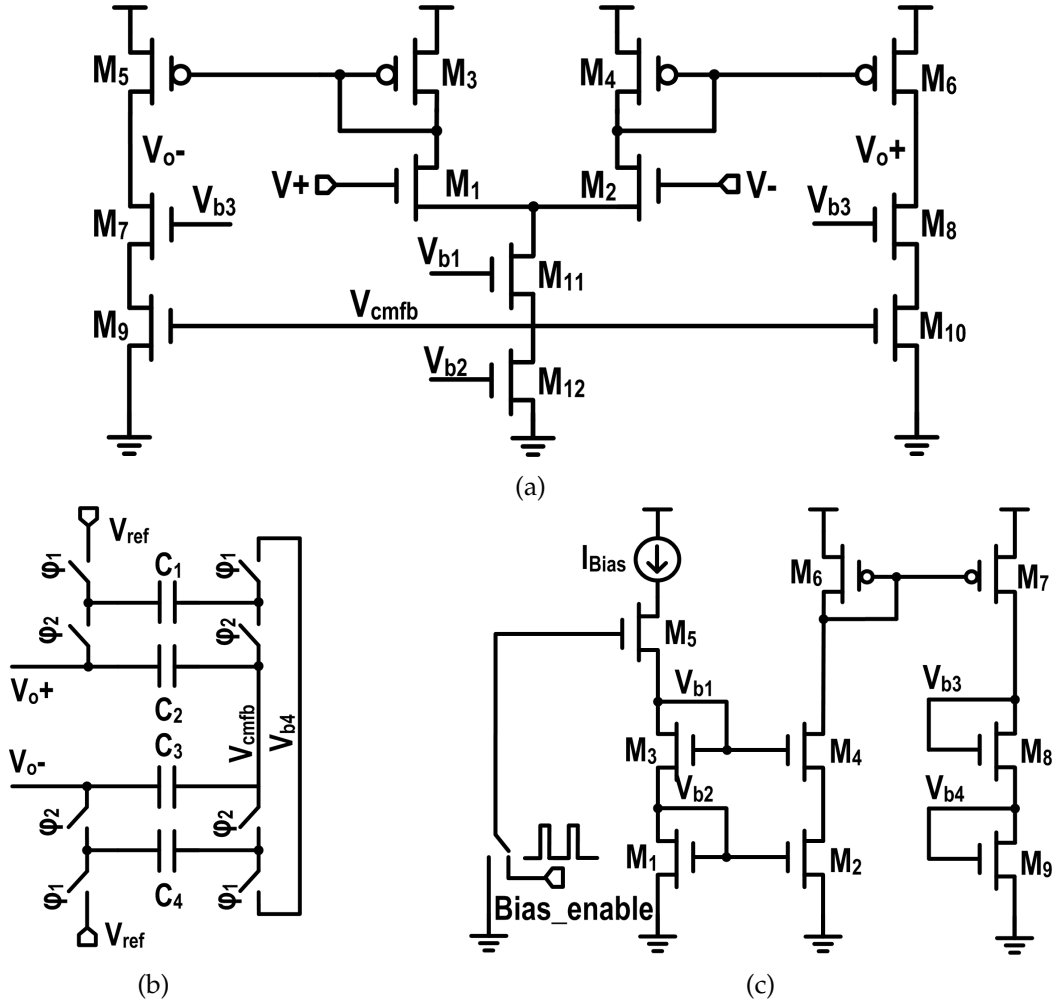


FIGURE 3.6 – a) Schematic of the current-mirror OTA used for all three $\Delta\Sigma$ integrators, b) Switched capacitor CMFB of the OTA in a), c) Bias circuit of the current-mirror OTA with a $Bias_enable$ switch to turn off the bias current of the OTA during both clock phases.

of each stage can be written as follows :

$$\begin{aligned}
 \text{Stage 1 : } E_1(z) &= x(z) - DAC_1(z), \\
 \text{Stage 2 : } E_2(z) &= DAC_1(z) - DAC_2(z) - V_1(z), \\
 \text{Stage 3 : } E_3(z) &= DAC_2(z) - DAC_3(z) - V_2(z),
 \end{aligned} \tag{3.3}$$

where $V_x(z)$ is the integrator outputs of the x^{th} stage, and $x(z)$ is the signal to be quantized. While switched-capacitor (SC) circuits are typically used to perform the subtraction between the two DAC outputs and the integrator output ($V_x(z)$) in (3.3) [123, 126], the proposed topology improves previous solutions by performing this subtraction digitally instead, which drastically simplifies the circuit implementation while being more tolerant to mismatch. Rather than performing

$$DAC_{x-1}(z) - DAC_x(z) - V_{x-1} \tag{3.4}$$

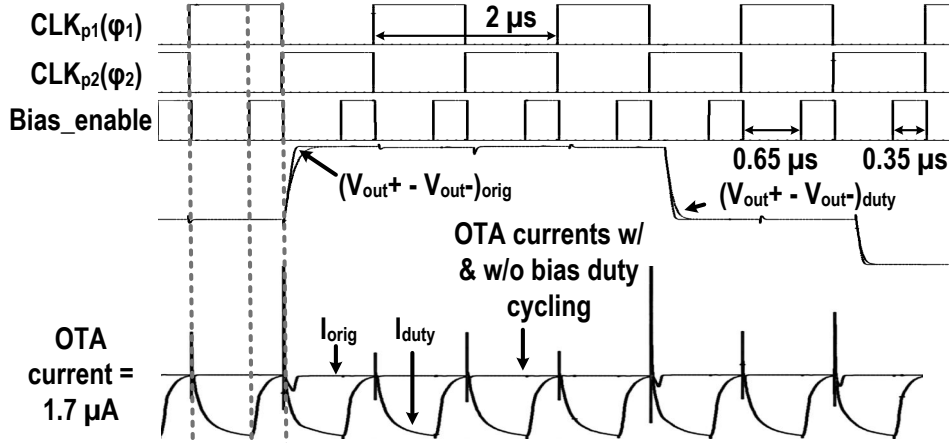


FIGURE 3.7 – Timing diagram illustrating the bias duty cycling strategy. The bias current of the integrator OTA is turned ON/OFF using the control signal *Bias_enable* with a duty cycle of 35%, saving $\sim 41.3\%$ in power consumption.

using SC circuits,

$$DAC_{x-1}(z) - DAC_x(z) \quad (3.5)$$

is performed using the digital outputs of the current ($Latch_x$) and the previous ($Latch_{x-1}$) comparators (i.e., the output of the D-Latches in Fig. 3.5) by the pre-subtraction module in Fig. 3.5. The result of $DAC_{x-1}(z) - DAC_x(z)$ can take only three possible values: V_{ref1} , V_{ref2} or V_{cm} , which are selected digitally using analog switches in this design. Table 3.1 shows all the possibilities that the *positive node* of a fully differential topology can take, while the negative node can be found by symmetry. Where $+DAC_{ref}$ and $-DAC_{ref}$ are the differential DAC output voltages of the 1st $\Delta\Sigma$ stage, while the subsequent stages require the voltages $-2 \cdot DAC_{ref}$ (i.e. V_{ref1}), $+2 \cdot DAC_{ref}$ (i.e. V_{ref2}) and V_{cm} , selected by the pre-subtraction module. The corresponding logic circuit has 3 control bits to select the proper subtraction output :

$$\begin{aligned} S_1 &= \overline{Lacth_{x-1}} \oplus Lacth_x \\ S_2 &= \overline{Lacth_{x-1}} \wedge Lacth_x \\ S_3 &= Lacth_{x-1} \wedge \overline{Lacth_x} \end{aligned} \quad (3.6)$$

Signals V_{ref1} and V_{ref2} can take any values, as long as they are equal to $-2 \cdot DAC_{ref}$ and $+2 \cdot DAC_{ref}$. However, to maximize the input dynamic range, and to reduce circuit complexity, V_{ref1} and V_{ref2} are set to VSS and VDD, respectively, in this design, leading to $+DAC_{ref} = \frac{3}{4}VDD$ and $-DAC_{ref} = \frac{1}{4}VDD$. These two reference voltages are generated using two on-chip low-dropout regulator (LDO) circuits. It can be shown that a variation of 1% of both reference signals can degrade the SNDR by approximately 2 dB, thus we use off-chip resistors to com-

TABLE 3.1 – Positive node subtraction results possibilities between DAC_{x-1} and DAC_x according to the latch outputs.

Latch _{x-1}	Latch _x	(DAC _{x-1}) ⁺	(DAC _x) ⁺	(DAC _{x-1} - DAC _x) ⁺
0	0	-DAC _{ref}	-DAC _{ref}	V _{cm}
0	1	-DAC _{ref}	+DAC _{ref}	-2 · DAC _{ref} (V _{ref1})
1	0	+DAC _{ref}	-DAC _{ref}	+2 · DAC _{ref} (V _{ref2})
1	1	+DAC _{ref}	+DAC _{ref}	V _{cm}

pensate for any process variations. The LDO voltages are shared between all $\Delta\Sigma$ channels. Fig. 3.5(b) shows a conventional circuit implementation of the subtraction between the DAC outputs, using a SC circuit, as in [123]. Compared to conventional SC circuit implementation, the proposed solution reduces the footprint size of the analog circuits by $\sim 11\%$, requiring two fewer capacitors in each 2nd and 3rd $\Delta\Sigma$ stage. Also, no mismatch is induced by the pre-subtraction module, in contrast to the SC implementation.

Integrator OTA

As shown in Figs. 3.6a and 3.6b, the integrator uses a fully differential current mirror OTA with a switched capacitor CMFB, which provides a good trade-off between complexity, power consumption and output swing. Utilizing a low OSR in this $\Delta\Sigma$ MASH 1-1-1 design allows for a longer clock period, so the OTA of each $\Delta\Sigma$ integrator can remain idle for most of the sampling period. Therefore, an efficient bias duty cycling scheme is used to turn off the OTA bias current between each sampling cycle, through the switch controlled by *Bias_enable* (Fig. 3.6c), allowing to further decrease the optimal power consumption required for the integrator output to settle within a given period. The optimal current consumption of the OTA can be obtained based on the following settling time analysis. The OTA settling time during the integrating phase can be divided into two distinct periods including *slewing* and *linear settling* (Fig. 3.8(a)). During the slewing period, the negative feedback is broken and the integrator can be modeled by the circuit shown in Fig. 3.8(b). The voltage difference between when the OTA enters the slewing region and the voltage at which the input differential pair enters the saturation region is expressed by :

$$\Delta V_x = V_i \cdot C_1 / (C_{eq} + C_1) - V_{eff1}, \quad (3.7)$$

where V_i is the maximum input voltage and $C_{eq} = C_p + C_3 \parallel C_L \approx 0.5C_L$, since $C_{1,2,3} = C_L$ in this design. As can be extracted from Fig. 3.8(b), the feedback factor f is defined as $f = C_3 / (C_3 + C_1 + C_p) \approx 0.5$. The output voltage of the integrator during the slewing phase is defined by $\Delta V_{out} = \Delta V_x / f$. Thus, t_{slew} in Fig. 3.8(a) is :

$$t_{slew} = \Delta V_{out} / SR = (\Delta V_x C_{Leff}) / (f \cdot I_{ss}), \quad (3.8)$$

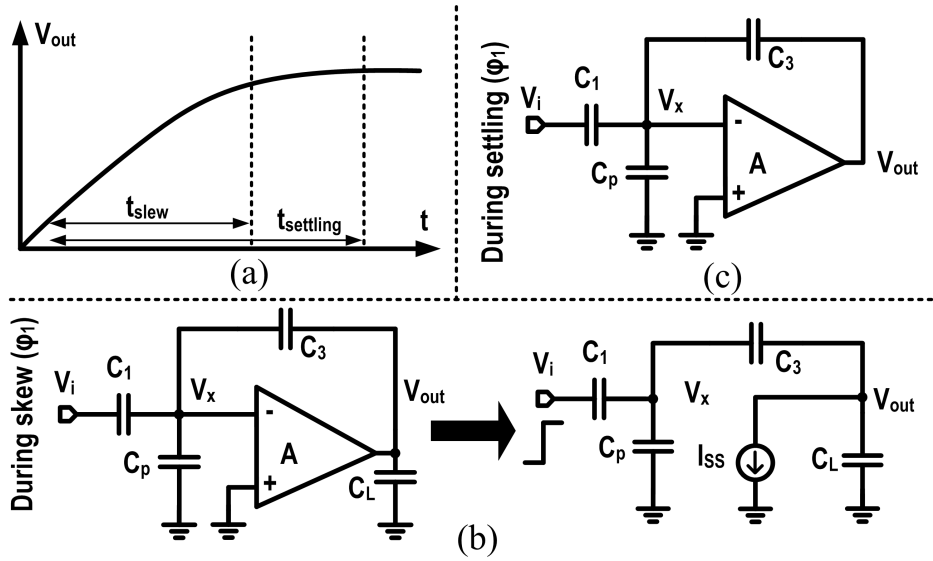


FIGURE 3.8 – Settling time modeling of the OTA during the integration phase, (a) timing diagram representing the settling, (b) schematic diagram of the OTA during the slewing phase, and (c) during the linear settling phase.

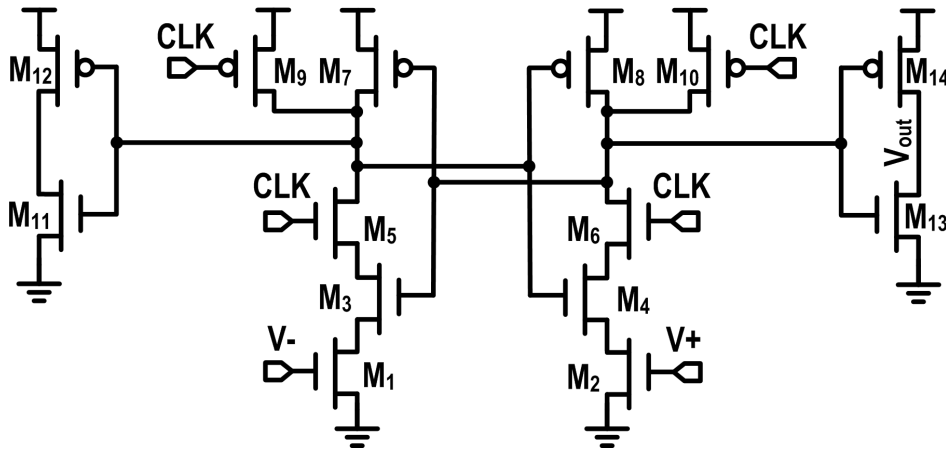


FIGURE 3.9 – Schematic of the dynamic regenerative comparator utilized in each one bit $\Delta\Sigma$ quantizer.

where the slew rate of the OTA is defined as I_{SS}/C_L , in which I_{SS} is the current flowing in the output branch of the OTA (Fig. 3.6a) and $C_{L_{eff}}$ is the total capacitance at the output of the OTA, defined by $C_{L_{eff}} \approx C_L + C_3 \parallel C_1 = \frac{3}{2}C_L$. During the linear settling phase shown in Fig. 3.8(c), the output of the OTA can be defined as :

$$V_{out}(t) = (1 - e^{-\omega - 3dBt})(C_1/C_L), \quad (3.9)$$

where $\omega_{-3dB} = \omega_u \cdot f = \frac{g_m}{C_{Leff}} \cdot f$, thus the BW variation is :

$$\Delta_{BW} = e^{\frac{g_m}{C_{Leff}} \cdot f \cdot (T - t_{slew})}, \quad (3.10)$$

which in turn is required to be smaller than the settling error i.e. $\Delta_{BW} \ll \alpha_e \cdot V_{FS}$, where V_{FS} is the full scale voltage and α_e is the error ratio (set to 0.1% for a settling error in the range of the bio-amplifier output noise). We can relate the OTA minimum bias current to the period by :

$$g_m \geq (f \cdot (T/2 - t_{slew}))^{-1} \cdot C_{Leff} \cdot \ln(\alpha_e \cdot V_{FS}) \quad (3.11)$$

Replacing g_m by $I_{ds1} / (0.5 \cdot V_{eff1})$ (square-law model) and I_{SS} in t_{slew} by $K \cdot I_{ds1}$ in (3.11) allows us to find the minimum bias current. Using this equation, we found that a static bias current of 1.7 μA per OTA (Fig. 3.6a) allows decreasing the half-period down to 0.4 μs . However, when using our bias duty-cycling strategy, the bias current behaves with a 1st order time constant, as seen in Fig. 3.6a (bottom), so we used 0.35 μs instead. The resulting averaged current per OTA is 20% lower than the theoretical minimum bias current (calculated using (3.11)) for a 2 μs period, since the OTA has less static current consumption. Compared with a 1.7 μA static bias current, this strategy saves around 41.3% of the power consumption of the whole $\Delta\Sigma$, while the steady-state output of the integrator is not affected.

One-bit quantizer

Each $\Delta\Sigma$ stage shown in Fig. 3.9 uses a dynamic regenerative comparator, followed by a D-latch as one-bit quantizer. The dynamic regenerative comparator enters in the *reset mode*, i.e., M_{10} keeps the output to VSS, when the clock phase φ_1 is low. On the rising edge of φ_1 , transistors M_3 to M_8 form a positive feedback loop that amplifies the input difference sensed by M_1 and M_2 into a full-scale output. Once the comparator makes a decision, cross-coupled transistors M_3 - M_4 and M_7 - M_8 shut down all connections between VDD and VSS, which causes null quiescent current. The comparator offset is compensated by the $\Delta\Sigma$ feedback loops, so no additional offset cancellation technique is required in this design. However, since the output is reset when φ_1 becomes low, a D-Latch is added to keep the comparator value constant during the φ_2 clock phase.

3.5.3 LED driver

The circuit schematic of the 4-ch LED driver is presented in Fig. 3.10. Compared to previous designs [119], [122], [127] and [128], this circuit allows to precisely control the value of the forward current into each LED using a regulated cascode current source with feedback. The current can be adjusted independently of the LED parameters (forward voltage, etc.) and the battery voltage, guaranteeing that the right optical power is delivered by each LED into the implanted fibers to properly activate the ChR2 neurons [1] ($>0.1\text{mW}/\text{mm}^2$ at the tip). Each

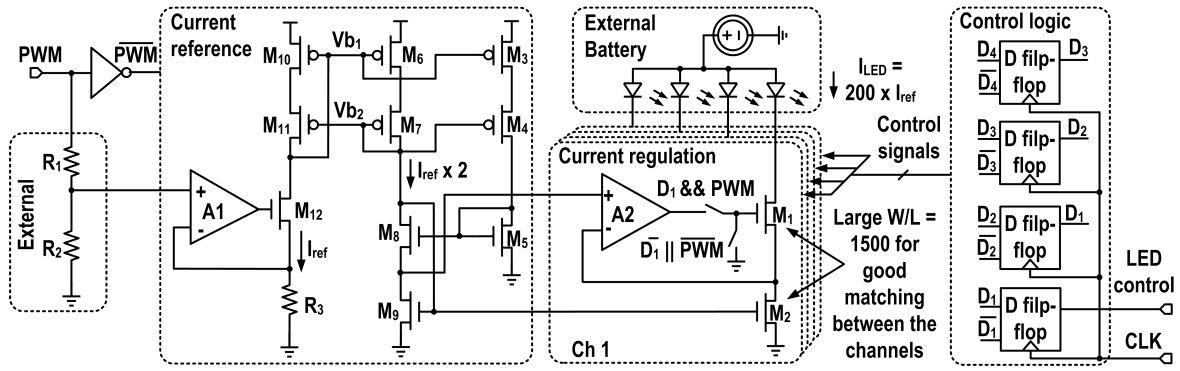


FIGURE 3.10 – Circuit schematic of the 4-ch LED driver. This circuit allows us to precisely control the value of the forward current in each LED using a regulated wide-swing current source with feedback. The average driving current is modulated by an external *PWM* signal.

channel can be activated by writing a 4-bit word in the shift register using the *LED control* and *CLK* signals. The driving current is modulated by an external *PWM* signal.

The proposed LED driver circuit works as follows : First, a precise current source generates the reference current I_{ref} based on external resistor value ratios of R_1 and R_2 . To save power, I_{ref} is also modulated by the *PWM* signal. With a 1.2-V *PWM* signal, the I_{ref} value when the *PWM* is high is :

$$I_{ref} = 1.2V \cdot R_2 / (R_3 \cdot (R_2 + R_1)). \quad (3.12)$$

By using off-chip resistors for R_1 and R_2 , I_{ref} can be easily tuned to regulate the desired current inside the LEDs. Current I_{ref} is duplicated and multiplied by 2 using a PMOS wide swing current mirror, and it is further duplicated and multiplied by 100 using an NMOS wide swing current mirror. More details about the regulated NMOS cascode current mirror can be found in [129]. However, it is worth nothing that this wide swing current mirror keeps the drain-source voltage of M_2 at the edge of saturation ($V_{ds2} \approx V_{eff}$), and M_1 can enter the triode region without significantly impacting the drain current, which both minimizes the voltage drop inside the chip to accommodate smaller battery voltages. The ratios $W_{1,2}/L_{1,2}$ are taken large (1500) to improve matching between the LED driver channels, and to reduce the internal voltage drop. When the *PWM* signal is at 0 the gate of M_1 is set to the ground to ensure that the LED is shutdown.

3.5.4 Integration Inside a Wireless Platform

Fig. 3.11(a) shows a representation of the head-mountable wireless neuroscience platform employed to characterize the chip *in vivo*. This platform is designed to perform simultaneous optogenetic stimulation and electrophysiology recording in the brain of small freely behaving rodents. It consists of six main building blocks : 1) the proposed SoC wire bonded directly to the platform's PCB, 2) a 2.4-GHz nRF24L01+ wireless transceiver from Nordic Se-

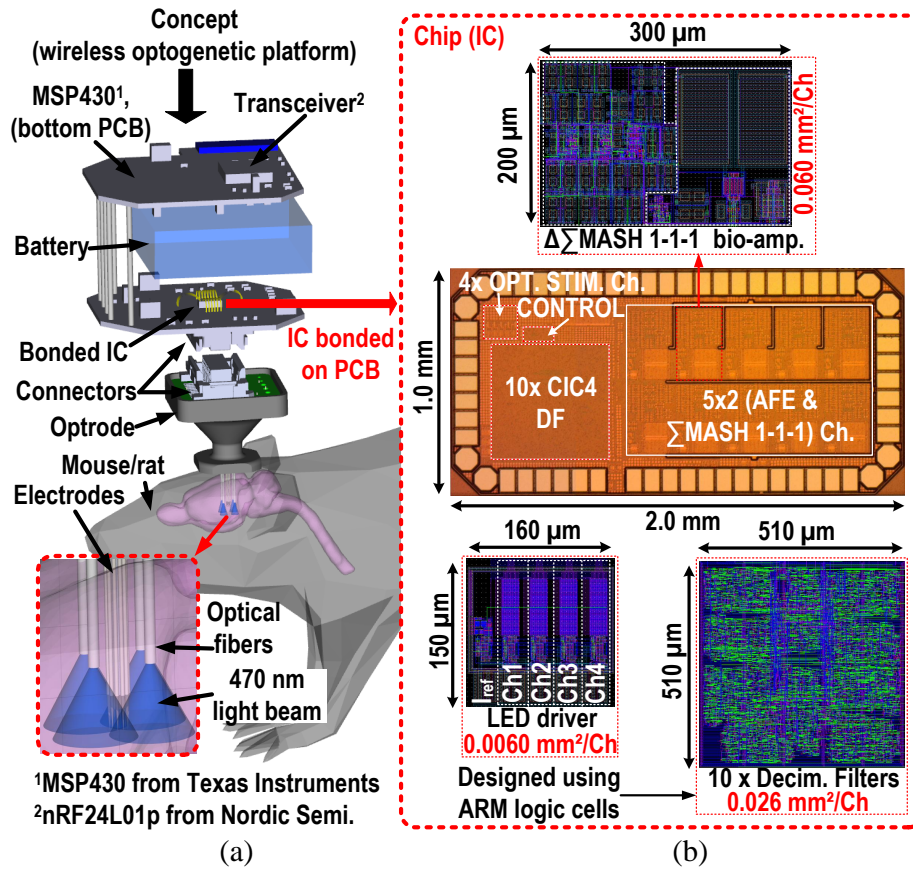


FIGURE 3.11 – Die micrograph (right) and system-level concept using the SoC within an optogenetic platform (left).

micrometers, Norway, with a maximum data rate of 2 Mbps, 3) a power management unit (PMU) consisting of a 1.2-V LDO for the SoC and a 1.9-V LDO for the MCU and the wireless transceiver, 4) a small ($5 \times 5 \text{ mm}^2$ Micro BGA package), low-power MSP430F5328 MCU from *Texas Instruments*, USA, an MCU family used in many low-power biomedical applications [18, 79, 130], 5) the implantable module holding the microelectrodes, the LEDs and the optical fibers, and 6) a 3.7-V, 100-mAh and 2.1-g Lithium-ion polymer (LiPo) battery (MYD Technology, China).

3.6 Experimental Results

Fig. 3.11(b) shows the SoC micrograph. The chip is designed and fabricated in the GF (IBM) $0.13 \mu\text{m}$ CMOS process technology and occupies a size of 1.2 mm^2 . It uses a single 1.2-V power supply for all the analog and the digital circuitry. The circuit layout of the bio-amplifier, the $\Delta\Sigma$ MASH 1-1-1, the DF and the LED driver, including their respective sizes, are shown in Fig. 3.11(b). A whole recording channel occupies a silicon area of 0.086 mm^2 of space, and

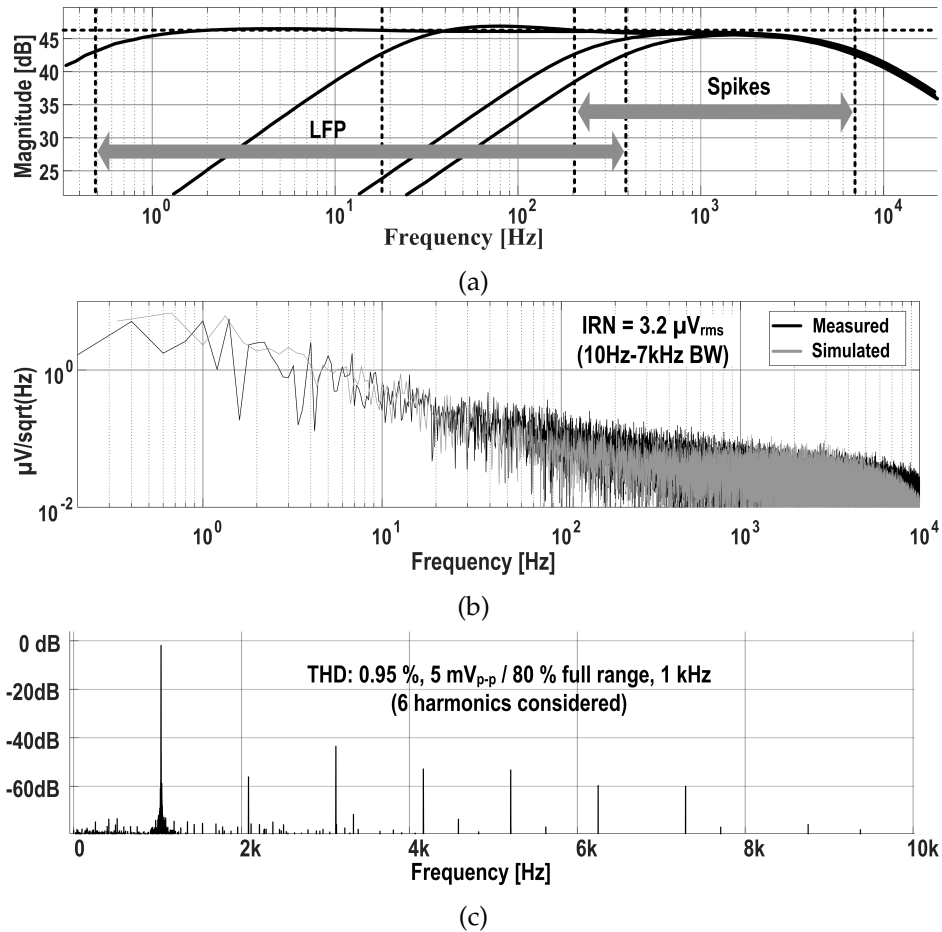


FIGURE 3.12 – a) Bio-amplifier BW for different high-pass cutoff frequencies, b) measured ($3.2 \mu\text{V}_{rms}$) and simulated IRN, c) measured THD (0.95%).

each optical stimulation channel occupies 0.006 mm^2 .

Fig. 3.12a shows the measured frequency response of the bio-amplifier. The achieved mid-band gain is of 46 dB with a 3-dB low-pass cutoff frequency of 7 kHz. The pseudo-resistor bank allows to select a suitable high-pass cutoff frequency between 0.5 Hz and 400 Hz to record either LFP, AP or both. The measured inter-channel standard deviation from the mean cutoff frequency, caused by the pseudo-resistor mismatch, is of 5% (measured across 2 chips), with a worst case variation between two channels of 13%. Fig. 3.12b depicts the measured and simulated IRN of the bio-amplifier. The IRN is calculated by measuring the output noise divided by the mid-band gain. The resulting measured power spectrum is integrated from 10 Hz to 7 kHz and is equal to $3.2 \mu\text{V}_{rms}$ with a power consumption of $2.3 \mu\text{A}$ drawn from a 1.2-V power supply. The resulting NEF is 2.30. Fig. 3.12c shows the bio-amplifier output spectrum for a 1-kHz/5.0 mV_{p-p} input signal, resulting in a THD of 0.95% when considering the first 6 visible harmonics. The bio-amplifier has a measured mid-band common-mode

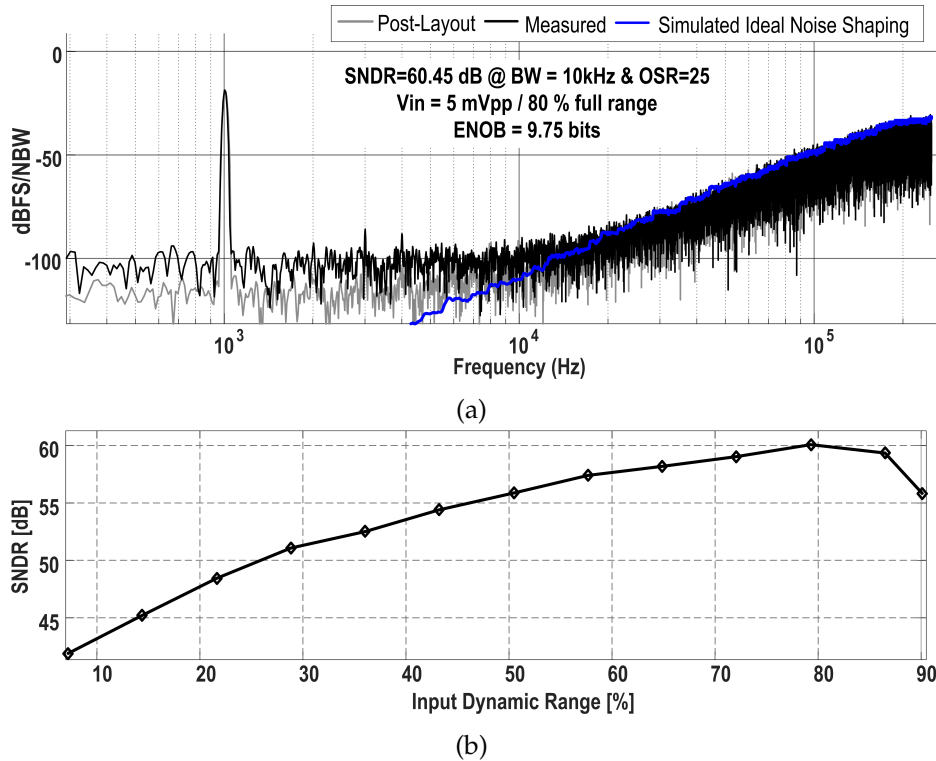


FIGURE 3.13 – a) Post-layout and measured output spectrum of the $\Delta\Sigma$ MASH 1-1-1 for an input signal amplitude of 80% of the full dynamic range (SNDR = 60.45 dB, at an OSR of 25 and a BW of 10 kHz). An ideal simulated noise shaping envelope for a $\Delta\Sigma$ MASH 1-1-1 is also shown, b) Measured SNDR according to the input signal amplitude relative to the full dynamic range.

rejection ratio (CMRR) is of 71.7 dB at 1 kHz.

Fig. 3.13a presents the post-layout and measured output spectrum of the $\Delta\Sigma$ MASH 1-1-1 with an OSR of 25 and for an input signal having an amplitude of 80% of the full dynamic range. The simulated ideal noise shaping envelope for a $\Delta\Sigma$ MASH 1-1-1 is also displayed. The measured signal to noise and distortion ratio (SNDR) is 60.45 dB, yielding an ENOB of 9.75 bits for a bandwidth of 10 kHz, with a power consumption of 3.6 μA drawn from a 1.2-V power supply, and when using the bias duty cycling strategy. Fig. 3.13b shows the SNDR according to the input signal amplitude relative to the full scale. The SNDR increases almost linearly to 80% and starts dropping significantly after 86%. The power consumption of the $\Delta\Sigma$ MASH 1-1-1 w/ DF is presented in Fig. 3.14a for five typical OSRs and their respective measured ENOB. As can be seen, the desired resolution and power consumption of the proposed SoC can be selected according to the type of signal to be recorded, the power or lifetime requirements of the application and to the performances of the bio-amplifier. As detailed in Fig. 3.14b, the power consumption of the bio-amplifier, the $\Delta\Sigma$ MASH 1-1-1 and the DF is 11.2 $\mu\text{W}/\text{ch}$ with an OSR of 25 ($\Delta\Sigma$ BW of 10 kHz).

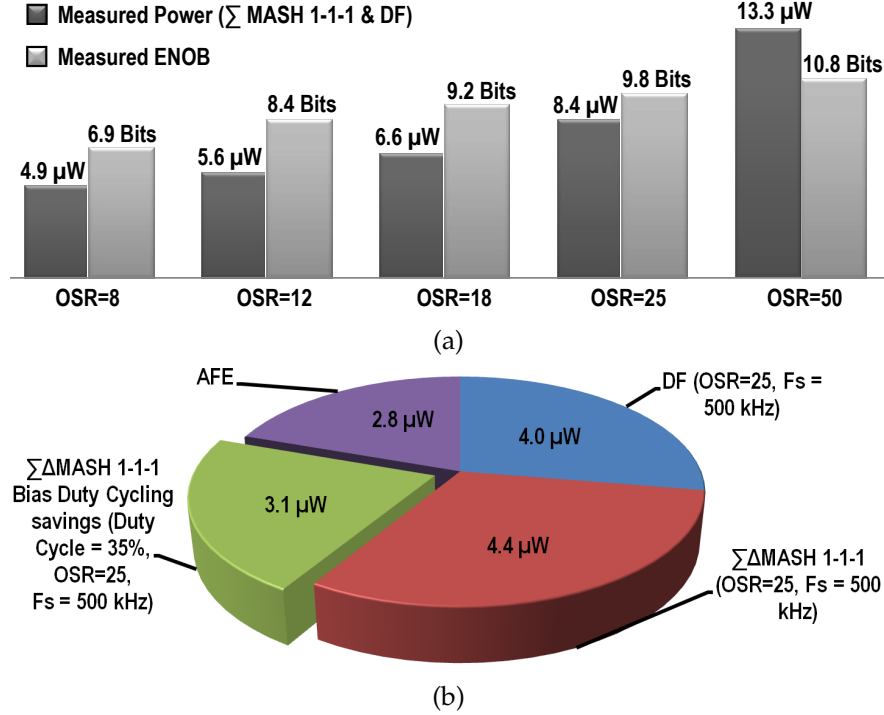


FIGURE 3.14 – a) Power consumption of the $\Delta\Sigma$ MASH 1-1-1 w/ DF for various OSR to achieve different ENOB, b) The power consumption of the bio-amplifier, the $\Delta\Sigma$ MASH 1-1-1 and the DF is 11.2 μ W/ch ($F_s=500$ kHz) when using the bias duty cycling strategy (3.1 μ W less power).

Fig. 3.15a shows the current flowing in a LED driver channel for various regulated currents and battery voltages applied at the anode of the LED. In these tests, an LB G6SP LED from OSRAM, which is often used in commercially available optrodes [20, 81, 103], is employed. The LED driver can regulate a 10-mA forward current until the battery voltage reaches 3.1 V ($V_{forward} = 2.82$ V), yielding a voltage drop of 280 mV across the current source inside de chip. Also, the LED driver can precisely regulate 20-mA and 30-mA currents until the battery voltage reaches 3.4 V and 3.5 V, respectively. These voltages are below the typical depletion voltages of small LiPo batteries often used in wireless neural recording and optogenetic systems [81] and [20]. Fig. 3.15a shows the simulated regulated currents (gray dotted-lines) obtained for a simple current mirror topology for comparison i.e. by putting M_9 and M_2 in current mirror configuration [119, 121] and by removing M_1 , M_5 , M_8 and A2 in Fig. 3.10. As can be seen, the presented topology performs better than the simple current mirror ($\sim 0\%$ vs $\sim 10\%$ of current deflection respectively at $V_{bat} = 3.6$ V), even if M_1 adds a small voltage drop at the output of the LED driver. Indeed, since $W_1/L_1=1500$ is big, r_{ds1} equals only $\sim 3 \Omega$ when V_{g1} falls in triode. In the proposed circuit, the output current vs voltage curves remain flat until both output transistors of the LED driver stay in triode (M_1 and M_2), which contrasts with the current vs voltage curves obtained for a simple current mirror, which are

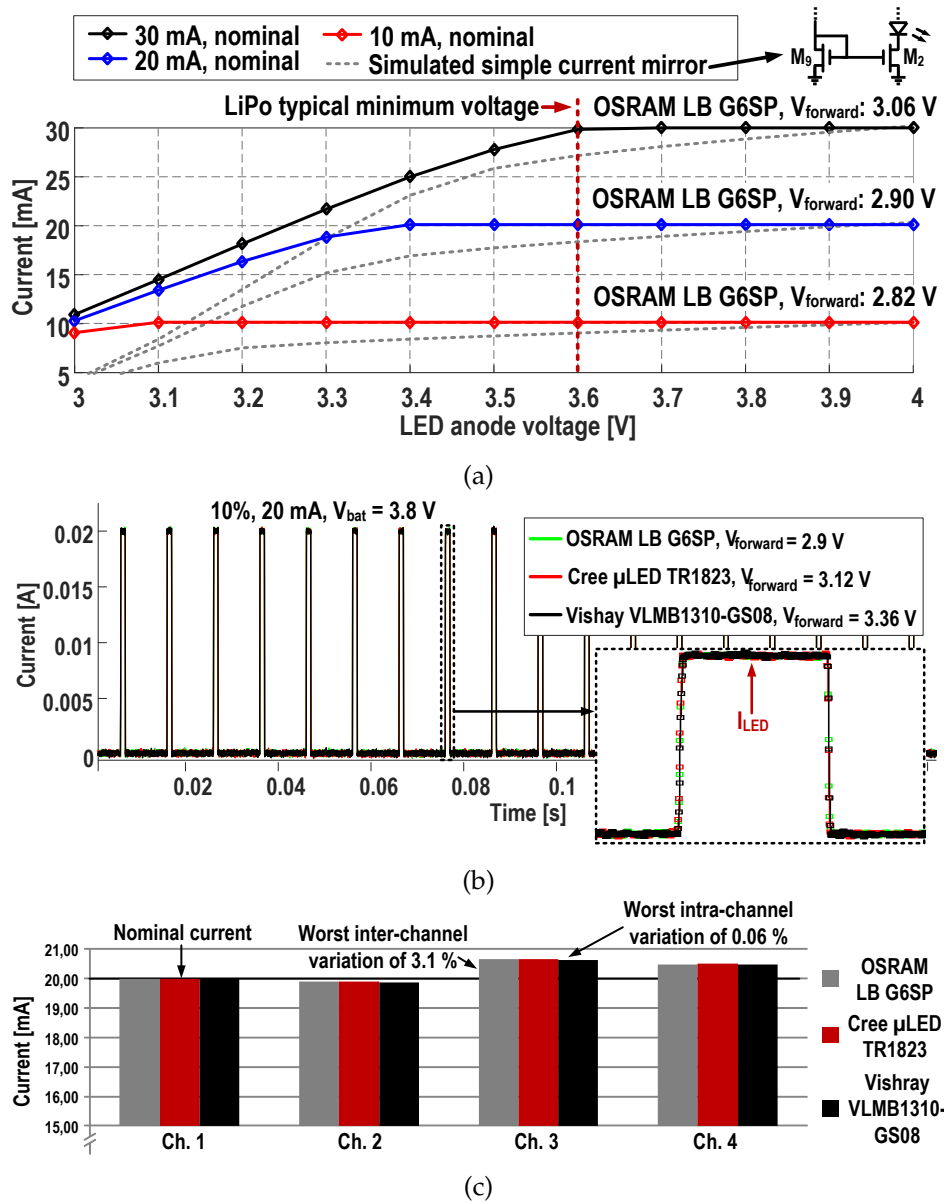


FIGURE 3.15 – a) Measured regulated currents against the LED anode voltage versus a simple current mirror, b) Regulated forward current pulses of 20 mA (which perfectly overlap) utilized to drive 3 different types of LEDs with different forward voltage characteristics, c) Regulated 20 mA current & intra/inter-channel variations due to process variation for 3 different types of LEDs.

constantly decreasing. Fig. 3.15b shows current pulses of 20-mA regulated by the LED driver using 3 blue LEDs often utilized in optogenetic settings (OSRAM [20, 81] and [103], Cree μ LED [105] and Vishay). As can be seen, the forward current curves obtained for the three LED are nearly identical and they perfectly overlap with each other, showing that the LED driver compensates for any varying LED parameters. This is also demonstrated in Fig. 3.15c,

TABLE 3.2 – SUMMARY OF THE EXPERIMENTAL RESULTS

SYSTEM-ON-CHIP:	
Supply voltage	1.2 V
Die Dimension	1.2 mm ²
Area Per Channel Rec.(w/ DF)/Stim.	0.086/0.006 mm ²
RECORDING CHANNEL:	
Bandwidth (tunable)	0.5 Hz -7 kHz
Dynamic Range*	5 mV _{p-p}
Power/Ch (w/ DF, BW=10 kHz, OSR=25)	11.2 μW/Ch
THD (1 kHz, 5 mV _{p-p} input)	0.95%
IRN (10 Hz -7 kHz)	3.2 μV _{rms}
NEF (bio-amplifier)	2.30
CMRR (mid-band at 1 kHz)	71.7 dB
ADC ENOB (BW=10 kHz, OSR=25)	9.75 bits
OPTICAL STIMULATION CHANNEL:	
Type	PWM
Optical Stimulation Current Regulation	0-35mA

*Before SNDR degradation, which is ~80% of the maximum of 6 mV_{p-p}

where the regulated current values of the 3 LEDs are shown for each stimulation channel. We can also see that the worst measured intra-channel current variation is of 0.06%, while the worst measured inter-channel variation is of 3.1%. The SoC performance is summarized in Table 4.2.

3.7 *In-vivo* Results

3.7.1 Surgical Procedure

A Long-Evans rat with virally mediated ChR2 cortical expression (four sites in primary motor cortex injected with 0.3 μl of AAV2/9-CaMKIIa-hChR2(H134R)-EYFP, with a titer of 1.4x10¹³ GC/ml) was used to stimulate and record within two *in vivo* experiments wherein the rat was anesthetized with ketamine/xylazine. In the first experiment, a glass micropipette (3 μm tip) filled with saline solution was lowered into the ventral posteromedial nucleus (VPM) region of the brain (4.2 mm depth). In the second experiment, an optrode consisting of 8×0.5 MΩ electrodes and one LED-coupled optical fiber (250 μm diameter) was lowered into the cerebral motor cortex (1.5 mm depth). The test setup, the system and the optrode are depicted in Fig. 3.16. All protocols were performed in accordance with *Canadian Council on Animal Care* Guidelines.

3.7.2 Micropipette Recordings

The SoC was connected to the micropipette using a silver wire with one extremity inserted into the capillary w/ saline solution. Simultaneously, the silver wire was connected to a

TABLE 3.3 – Comparison of IC for simultaneous electrophysiological recording and optogenetics.

	Optical Stimulation			Recording			Stimulation & Recording				
	ISSCC'10	JSSC'15	ISSCC'18	CICC'11	ISSCC'15	TBioCAS'18	BioCAS'15	JSSC'15	JSSC'17	TBME'17	This Work
	[119]	[120]	[105]	[123]	[108]	[118]	[122]	[131]	[61]	[132]	
TECHNOLOGY (μm)	0.8	0.35	0.35	0.13	0.065	0.18	0.35	0.065	0.13	0.18	0.13
Nb. Ch. (OStim./Rec. ¹)	2	4/0	16/0	0/1	0/16	0/4	6/4	0/64	0/64	1/1	4/10
TARGETED SIGNALS	-	-	-	-	AP/LFP	AP/LFP	LFP	AP/LFP	LFP	AP	AP/LFP
ON-CHIP DECIMATION	-	-	-	NO	-	-	-	-	NO	-	YES
Bio-Amp. BW (Hz)	-	-	-	-	1-8.2k	0.7-9.3k	0.08-N/A	N/A-10k	0.01-500	114-14.6k	0.5-7k
POWER/CH (μW)	-	-	-	18($\Delta\Sigma$)	2.8 ²	4.5 ²	-	1.84	0.63 ³	-	11.2 ⁴
NOISE BW (Hz)	-	-	-	-	1-8.2k	0.7-9.3k	-	100-10k	0.1-500	300-5k	10-7k
IRN (μV_{rms})	-	-	-	-	4.2	3.2	-	7.5	1.13	4.57	3.2
NEF/PEF	-	-	-	-	2.93/8.58 ²	1.94/6.77 ²	-	3.6/12.9 ²	2.86/9.81 ³	N/A	2.30/6.35 ²
AREA (AFE/ADC, mm^2/Ch)	-	-	-	-/0.06	0.042/-	0.072/-	N/A	0.026	0.013	0.20/-	0.03/0.056 ⁵
OStim. CURRENT (mA)	0-51.2	0-4	0-10	-	-	-	0-1	- ⁶	- ⁶	0-330	0-35
ENOB/OSR	-	-	-	11.5/64	-	-	10/N/A	8.2/N/A	11.7/1000	-	9.75/25
ADC FOM (pJ/conv.-step)	-	-	-	0.15	-	-	N/A	N/A	0.19	-	0.19 ⁷
IN VIVO VALIDATION	YES	YES	YES	NO	YES	YES	NO	YES	YES	YES	YES

¹Optical Stim./Rec., ²Bio-Amplifier only, ³Bio-Amplifier+ $\Delta\Sigma$, ⁴Bio-Amplifier+ $\Delta\Sigma$ +DF, ⁵ $\Delta\Sigma$ & DF, ⁶Elect. Stimulation, ⁷Calculated w/ an OSR of 50

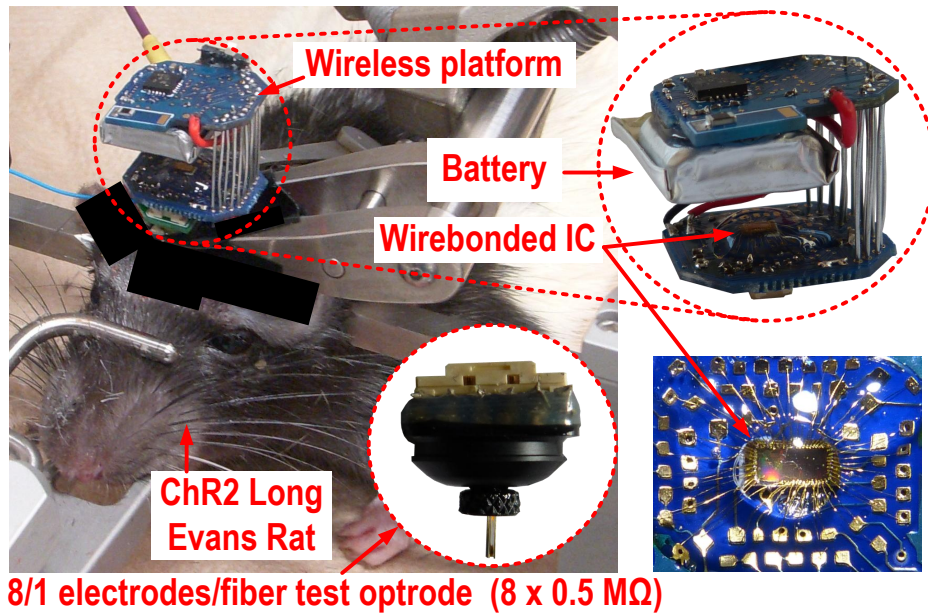


FIGURE 3.16 – The head-fixed *in vivo* experimental setup. The data was propagated between the wireless platform and a base station situated 2 meters away.

commercial extracellular neural amplifier, model ER-1 from *Cygnus Technology*, to compare the recorded waveforms with the proposed SoC. As Fig. 3.17a (top) shows, the chip enclosed within the wireless test platform was able to record spontaneous activity with a maximum amplitude of 1.3 mV_{p-p} with a noise floor standard deviation of $\sim 17 \mu\text{V}_{rms}$ using the glass micropipette. Fig. 3.17a (bottom right) presents several clusters of spikes collected within 3 recording sessions and classified and clustered using a *principal component analysis* (PCA). Fig. 3.17b shows the SoC recording in black superimposed over the ER-1 amplifier recorded signal. Both signals almost perfectly overlap, which demonstrates that the proposed SoC can collect and digitize spikes inside a wireless system with an accuracy that is comparable to a

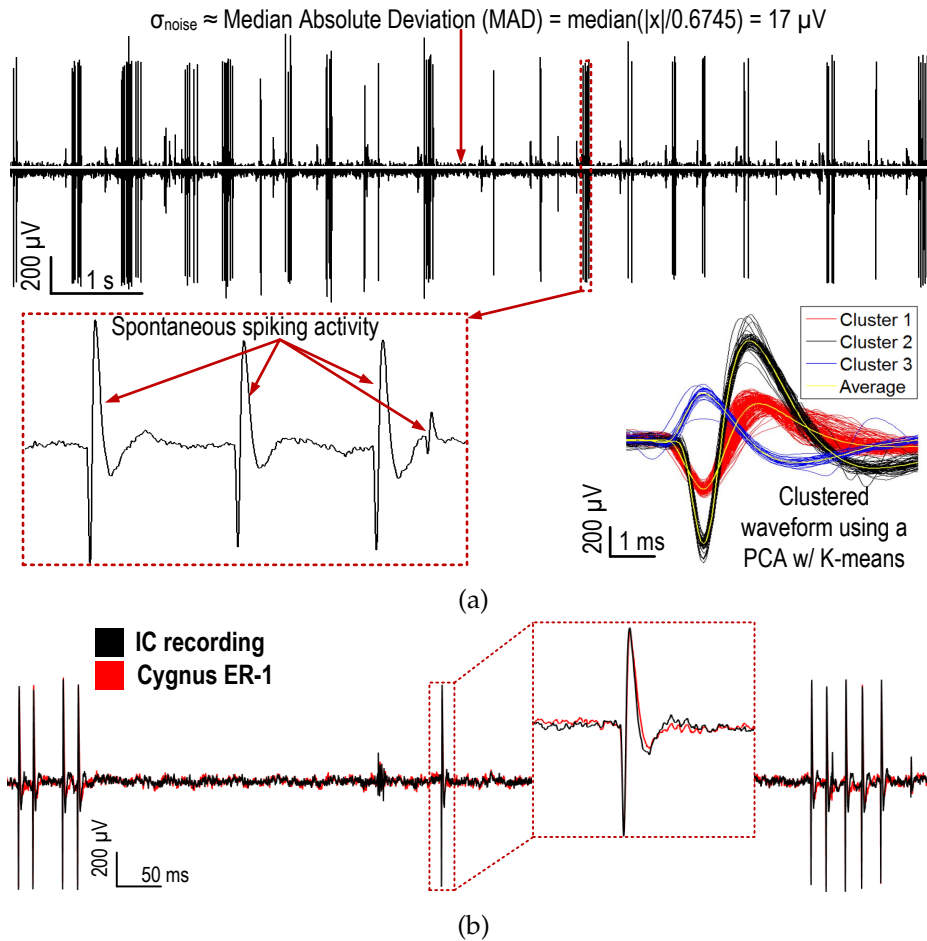


FIGURE 3.17 – a) Spontaneous activity recorded using a micropipette probe, and clusters of spikes collected within 3 recording sessions (bottom right), b) SoC recording (in black) superimposed over the signal recorded using the ER-1 commercial wired amplifier (in red).

commercial hard-wired system.

3.7.3 Optrode Stimulation and Recordings

Fig. 3.18 shows both the spontaneous and evoked spiking activity acquired while performing optical stimulation through the implanted optrode. As can be observed, the SoC was able to trigger brain activity with light, and collect the evoked neural activity. The evoked activity was triggered ~ 2 ms after the beginning of the stimulation pulses, which corresponds to what can be expected from the ChR2 protein [98]. In Fig. 3.18, it can be seen that the spikes triggered by stimulation have a higher amplitude than the spontaneous activity. Since a fiber diameter of $250 \mu\text{m}$ is used, the light beam at the tip of the fiber can trigger several neurons at the same time, generating simultaneous spikes that are collected and summed by the same electrode, thus increasing the amplitude of the aggregated signal compared to single unit

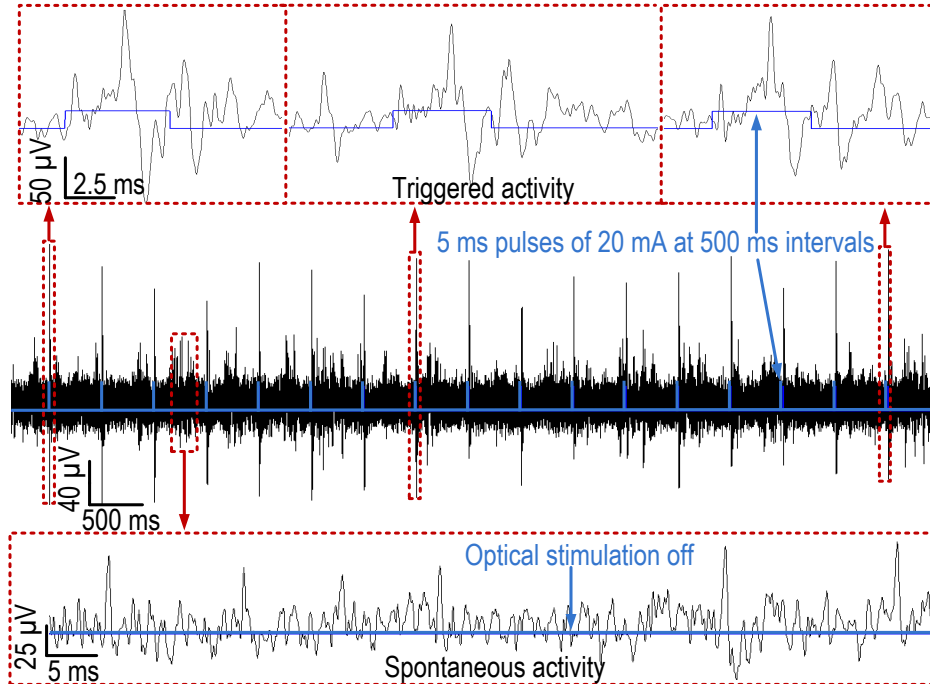
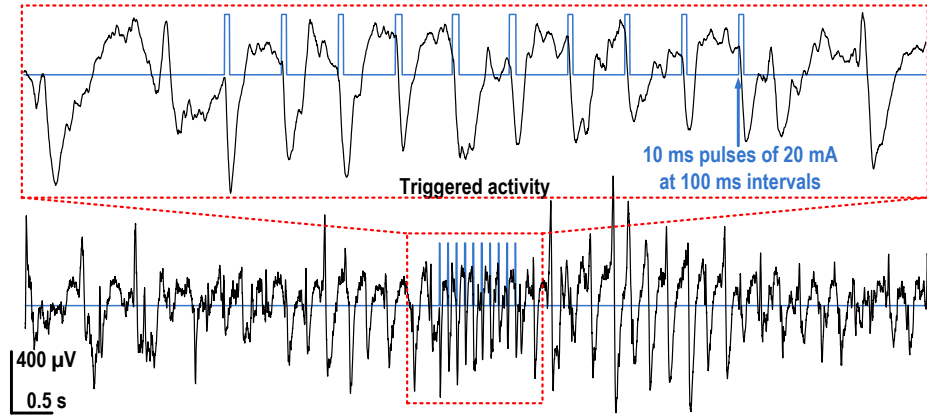
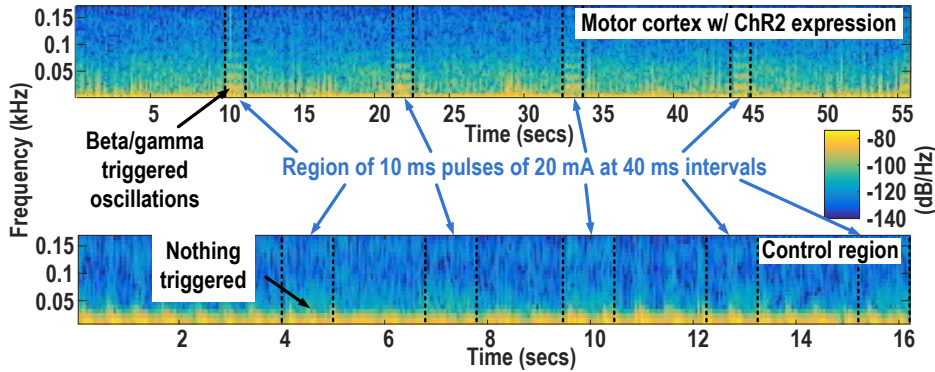


FIGURE 3.18 – Evoked neural activity acquired after optical stimulation through the optrode (top), spontaneous activity recorded by the SoC (bottom).

activity. As shown in Fig. 3.18 (bottom), the SoC recording circuitry has provided all the necessary precision to capture small spontaneous activity with less than $50\ \mu\text{V}$ of amplitude. Fig. 3.19a shows both spontaneous and evoked LFP activity after optical stimulation acquired over a 3-dB BW of 0.5-500 Hz through the optrode. Low-pass filtering was performed by setting the oversampling frequency of the DF to 50 kHz, which represents a sampling rate of 2 kHz after decimation with an OSR of 25. The LFP activity triggered and collected by the SoC has an amplitude of around $1.5\ \text{mV}_{p-p}$. Control recording along with stimulation was performed in the premotor cortex region of the rat's brain $\sim 4\ \text{mm}$ away from the injection sites in order to validate the testing setup. Fig. 3.19b shows the spectrogram of the recorded LFP waveform evoked with stimulation, both in the motor cortex with ChR2 expression and in the control region. As can be observed, the light stimulation in the motor cortex was able to trigger beta/gamma oscillation (12-100 Hz) [133], in contrast to the control region where no noticeable spectrum variation appears, which reinforces our observation that the stimulation-induced shifts in LFP signals are due to neuronal excitation rather than artifacts. It is worth noting that the utilized $0.5\ \text{M}\Omega$ electrodes were coated with polyimide and had a flat ending, so no metal was directly exposed to light, generating very low light artifacts, as Figs. 3.18 and 3.19 show.



(a)



(b)

FIGURE 3.19 – a) Spontaneous and evoked LFP activity recorded through the optrode, b) Spectrogram of the recorded LFP with stimulation both in the motor cortex with ChR2 expression and in the control region.

3.8 Discussion

Table 4.3 summarizes the performance of the proposed SoC compared with the recently published state-of-the-art systems. The FOM employed to compare all ADCs together is defined as $\text{Power}/(2 \times \text{BW} \times 2^{ENOB})$ [123]. Compared with the proposed SoC, the ICs presented in [105], [119] and [120] are dedicated to optogenetic stimulation only, without any recording capability. In [119], the LED driver is based on parallel current mirrors working in open-loop without any current feedback, while the circuits presented in [105] and [120] rely on capacitor discharging techniques providing current pulses of varying values. In contrast, the proposed LED driver provides precise and constant current pulses with low-voltage drop across the integrated current source using a circuit based on wide-swing current mirrors. The current is sensed and regulated directly in the path of the LED current, providing precise output current regulation [105, 119] and [120]. The IC presented in [61] provides electrical stimulation only and uses a $\Delta\Sigma$ converter to obtain a quadrature-phase output of the recorded neural

signal. Thus, in contrast with the proposed SoC, no decimation is performed to retrieve the original signal. This circuit achieves a slightly higher ENOB, but has the same ADC FOM as the proposed solution while targeting a bandwidth that is 20 times smaller. The IC presented in [108] can also collect both AP and LFP signals. However, it does not provide digitization, or optical stimulation, and its NEF is higher than the proposed bio-amplifier. The only serious attempt for a system including both optical and recording capabilities on the same IC is presented in [121]. However, this latter chip has a much lower optical and recording resolution than the proposed SoC. Additionally, the noise performance is worse with less BW (0.3-5 kHz), it can collect only AP, it has no ADC, and it performs optical stimulation using an open-loop current mirrors w/o regulation.

3.9 Conclusion

We presented a new SoC for simultaneous multichannel optical stimulation and electrophysiological brain recording. The proposed SoC can activate neurons using light pulse trains generated by up to 4 LEDs, and simultaneously collect the evoked neural activity from up to 10 recording channels. It can collect both AP and/or LFP signals with a programmable BW. It has on-chip low-noise bio-amplifiers, $\Delta\Sigma$ MASH 1-1-1 ADCs working at low and programmable OSR, CIC4 decimation filters and precise LED drivers. The ADC resolution can be programmed to accommodate different types of experiments and power budgets. It achieves a NEF of 2.30, an ENOB of 9.75 bits and a power consumption of 11.2 $\mu\text{W}/\text{ch}$ w/DF. The $\Delta\Sigma$ MASH 1-1-1 is enhanced by a bias duty-cycling strategy reducing its power consumption and by a digital subtraction technique decreasing its footprint size. The SoC has been integrated into a wireless neuroscience platform and validated *in vivo* in the brain of a virally mediated ChR2 rat.

Acknowledgments

We thank CMC Microsystems for supporting the chip fabrication. This project is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), the *Fonds de Recherche du Québec - Nature et Technologies* (FRQNT) and the Weston Brain Institute.

Chapitre 4

A Wireless Electro-Optic Headstage with a 0.13- μm CMOS Custom Integrated DWT Neural Signal Decoder for Closed-Loop Optogenetics

© 2019 IEEE. Reprinted, with permission, from G. Gagnon-Turcotte, I. Keramidis, C Ethier, Y. De Koninck and B. Gosselin, "A Wireless Electro-Optic Headstage with a 0.13- μm CMOS Custom Integrated DWT Neural Signal Decoder for Closed-Loop Optogenetics", *IEEE Transactions on Biomedical Circuits and Systems*, 2019 (submitted).

4.1 Résumé

Cet article présente un *headstage* électro-optique sans fil conçu pour faire de l'optogénétique en boucle fermée, et qui est démontré *in vivo* avec une souris libre de ses mouvements. Ce *headstage* miniature est basé sur un circuit intégré (IC) en technologie 0.13- μm CMOS qui implémente un décodeur neuronal digital (ND-IC) pour faire de l'optogénétique en boucle fermée et en temps réel. Le ND-IC traite les données issues de l'activité neuronale provenant de 10 canaux simultanément grâce à 3 modules digitaux : 1) Le module de détection détecte et extrait les potentiels d'action (AP) des neurones individuels en utilisant un seuil adaptatif, dont la valeur optimale est trouvée grâce à une analyse en temps réel des données, 2) le module de compression compresse chacun des AP détectés grâce à un processeur de transformé en ondelette discrète (DWT) de type Symmlet-2, et ce, afin de réduire la quantité de données à transmettre par le transmetteur sans fil à faible consommation énergétique, 3) le module de classification permet de classer les AP détectés dans des *clusters* en fonction de leur forme d'onde. Le ND-IC incorpore plusieurs innovations : 1) le module de compression réduit la complexité de $O(n^2)$, à $O(n \cdot \log(n))$ comparé aux solutions précédentes, tout en utilisant deux fois moins de mémoire grâce à l'utilisation d'un arbre de tri des coefficients.

2) Le module de classification réutilise les coefficients compressés de la DWT afin de faire une réduction implicite de la dimensionnalité, ce qui permet de réduire la consommation énergétique et la complexité du module. De plus, ce module réutilise l'écart-type du signal préalablement calculé par le module de détection des AP comme seuil afin de faire de la classification d'AP automatique. La *headstage* introduit également des innovations en introduisant une nouvelle technique permettant de fermer la boucle entre le module d'acquisition de données neuronales et le stimulateur optique. En effet, cet article démontre comment le *headstage* peut fermer la boucle entre le module analogique d'acquisition des données et le stimulateur optique en utilisant les temps d'occurrence des AP ainsi que leurs données de classification. Dans un premier temps, le *headstage* et le ND-IC sont validés *in vivo* dans le noyau laterodorsal du thalamus (c.-à-d. LD Thalamus) d'un rat de type Long-Evans, et ensuite, lors d'expériences optogénétiques en boucle fermée impliquant une souris libre de ses mouvements et exprimant la protéine ChR2-mCherry dans les neurones inhibiteurs du cortex prélimbique.

4.2 Abstract

We present a wireless electro-optic headstage which uses a 0.13- μm CMOS custom integrated circuit (IC) implementing a digital neural decoder (ND-IC) for enabling real-time closed-loop (CL) optogenetics. This paper also sets a milestone by reporting the first CL system results obtained with a freely-moving animal. The ND-IC processes the neural activity data using three digital cores: 1) The *Detector core* detects and extracts the action potential (AP) of individual neurons using an adaptive threshold, 2) the data *Compression core* compresses the detected AP using an efficient Symmlet-2 discrete wavelet transform (DWT) processor for decreasing the amount of data to be transmitted by the low-power wireless link, 3) the *Classification core* sorts the compressed AP into separated clusters on the fly according to their wave shapes. The ND-IC encompasses several innovations: 1) the Compression core decreases the complexity from $O(n^2)$, to $O(n \cdot \log(n))$ compared to previous solutions, while using two times less memory, thanks to the use of a new coefficient sorting tree. 2) The AP Classification core reuses both the compressed DWT coefficients to perform implicit dimensionality reduction, which allows for performing intensive signal processing on-chip, while increasing power and hardware efficiency. This core also reuses the signal standard deviation already computed by the AP Detector core as threshold for performing automatic AP sorting. The headstage also introduces innovations by enabling a new wireless CL scheme between the neural data acquisition module and the optical stimulator. Our CL scheme uses the AP sorting and timing information produced by the ND-IC for detecting complex firing patterns within the brain. The headstage is also smaller (1.13 cm³), lighter (3.0 g with a 40 mAh battery) and less-invasive than previous solutions, while providing a measured autonomy of 2h40, with the ND-IC. The whole system and the ND-IC are firstly validated *in vivo* in

the LD thalamus of a Long-Evans rat, and then, in freely-moving CL experiments involving a mouse virally expressing ChR2-mCherry in inhibitory neurons of the prelimbic cortex.

4.3 Introduction

With the rise of optogenetics in experimental neuroscience methods, it is now possible to use light to study the microcircuits deep inside the intact brain [1]. This new groundbreaking approach can activate specific neurons in the cortex of transgenic animals to observe their role in vast biological networks. Recently, optogenetics has been used to carry various behavioural experiments with freely behaving transgenic rodents, especially with mice [1, 37]. These experiments are performed using predefined stimulation patterns, with [20, 121], or without [102], simultaneous electrophysiological recording, and often using a bulky hardwired setup [40]

A promising paradigm with potential to accelerate the development of new therapeutics against brain diseases consists of closing the loop between the optical stimulator and the neural sensing device [46–48]. This approach, known as *closed-loop (CL) optogenetics*, can stimulate select brain areas according to the level of activity detected from a specific group of neurons or brain areas. Mainly, it requires three essential components, including a neural recording interface, an optical stimulator, and a neural analyzer. One major challenge facing CL optogenetic systems consists of analyzing dense neural activity in real time to quickly issue proper stimulation feedback commands inside specific brain areas. Most state-of-the-art systems designed for this purpose use a PC-based system and a wired experimental setting, which usually limits their use to anesthetized animals models [49, 50, 134]. Although these systems are convenient for performing complex neural analysis, they are not suitable for freely-moving experiments seeking natural behaviour, since the animals are tethered.

Systems based on custom integrated circuits (IC) have been developed in order to overcome these limitations. CL proof-of-concept systems using electrical stimulation as feedback were designed and tested. For instance, [57] performs a digital AP detection that triggers an electrical stimulation pulse every time the neural signal crosses a fixed, predefined threshold. This technique is simple and straightforward, but highly sensitive to noise and motion artifacts, especially in live animals. The robustness of this technique can be improved by performing real-time AP sorting to prevent a false positive due to noise and other artifacts [50]. Because of power and area limitations, few systems can perform on-chip AP classification. The IC presented in [132] performs a principal component analysis (PCA) in real time for triggering electrical stimulation based on AP sorting. However, this IC uses a significant silicon area ($1.77 \text{ mm}^2/\text{Ch}$) and power ($\sim 0.2 \text{ mW}/\text{Ch}$), was not demonstrated inside a complete experimental setting and tested *in vivo*. In contrast, the ICs that use PI/PID controllers based on LFP energy have applications with Parkinson or Epilepsy diseases [60, 61, 135], and they

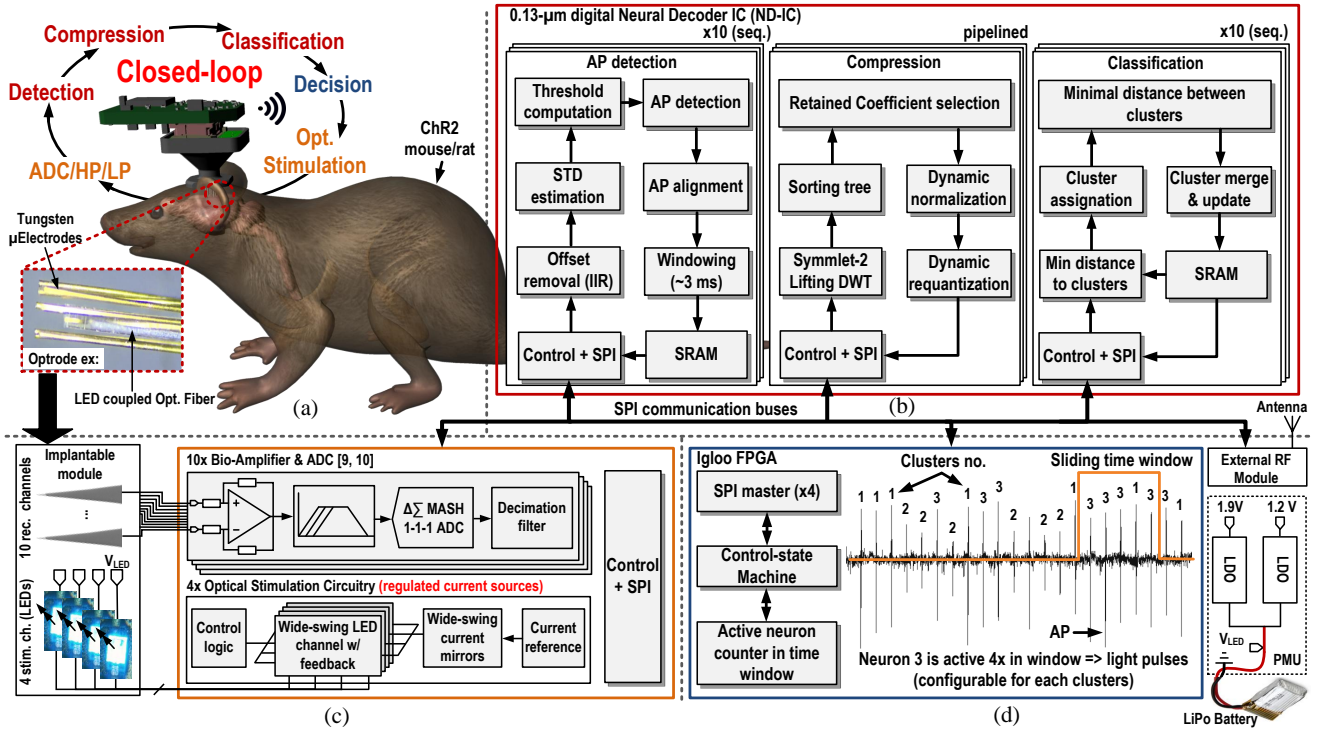


FIGURE 4.1 – System-level concept of the wireless CL optogenetic headstage using the proposed ND-IC as core module : a) The headstage is connected to the optrode and mounted on the head of the animal, b) the 0.13- μm CMOS ND-IC has 3 cores for AP detection, compression and classification, c) a custom 0.13- μm CMOS SoC is used for simultaneous neural recording and optogenetic stimulation and d) the CL algorithm is performed within a low-power FPGA.

are not intended for creating an artificial neural pathway at the neuron level using optogenetics. To date, the only system developed for wireless CL optogenetics is presented in [136], but uses a simplistic AP detection scheme to trigger feedback stimulation. Similar to [57], it presents substantial weight (11.3 g, a lab mouse is ~ 25 g), does not allow real-time neural data analysis, and was not validated live within a freely-moving experiment.

This paper presents a miniature wireless headstage for CL optogenetics based on a new 0.13- μm CMOS digital neural decoder IC (ND-IC) that detects, compresses, and sorts (i.e. clusters or classifies) the AP waveforms on-the-fly over 10 channels in parallel without supervision, which can be used to carry out complex CL schemes in live animals. The headstage is demonstrated using a freely moving mouse, which is the first reported demonstration of a complex CL optogenetics scheme implemented with a wireless system. Since AP bursting is deemed critically important for creating artificial neural pathways [54–56], the proposed CL algorithm retroactively stimulates the brain according to the AP occurrences and their cluster associations, improving the strategy introduced in [50]. Knowing that a single neuron typically generates an AP of similar shapes over time [53], the headstage estimates the

number of neurons near each recording site by looking at the clusters, and detects recognizable firing patterns by finding the number of APs belonging to a specific neuron within a sliding time window. The ND-IC also compresses each detected AP to decrease the amount of data to be handled by the low-power wireless transceiver, and allows all channels to be monitored simultaneously. The headstage and the ND-IC are first demonstrated *in vivo* with an anesthetized Long-Evans rat, and then, with a freely-moving mouse virally expressing ChR2-mCherry in inhibitory neurons of the prelimbic cortex.

This paper is the culmination of our effort devoted to the development of intelligent neural interfaces, and differs from our previous publications according to the following points :

1. This paper introduces a new 0.13- μm CMOS ND-IC, including an integrated AP detector and integrated DWT data compression algorithm, which were only prototyped in a FPGA system before [20]. This ND-IC also includes an AP sorting core that uses the DWT coefficients of the AP to perform automatic AP sorting in a very straightforward and efficient fashion.
2. We propose a new strategy to close the loop between the recorded brain activity and our optogenetic stimulator. Our system uses the AP sorting and timing information extracted by the ND-IC to recognize specific firing patterns inside the brain activity.
3. We present the integration of the ND-IC within a miniature and lightweight wireless system that is well suited for CL optogenetic experiments in real time.
4. We validate the system *in vivo* and we demonstrate our CL scheme with a freely-moving animal.

The paper is organized as follows, Section 4.4 presents the headstage architecture, Section 4.5 outlines the CL algorithm, Sections 4.6-4.8 present the ND-IC cores, Section 4.9 illustrates the headstage integration, Sections 4.10-4.11 present the experimental and *in vivo* results, followed by the discussion and conclusion in Sections 4.12-4.13.

4.4 System-Level Overview

The concept of the proposed electro-optic wireless headstage using the 0.13- μm CMOS ND-IC is presented in Fig. 4.1. The ND-IC is the most important module for CL optogenetics in this system. It provides real-time neural data decoding to establish a connection between a readout and a stimulator. The ND-IC block diagram is depicted in Fig. 4.1(b). This IC has three main cores :

1. The *Detector core* detects and extracts the AP of individual neurons. It uses an adaptive threshold based on real-time signal analysis performed through a feedback loop with a simplified implementation [20].

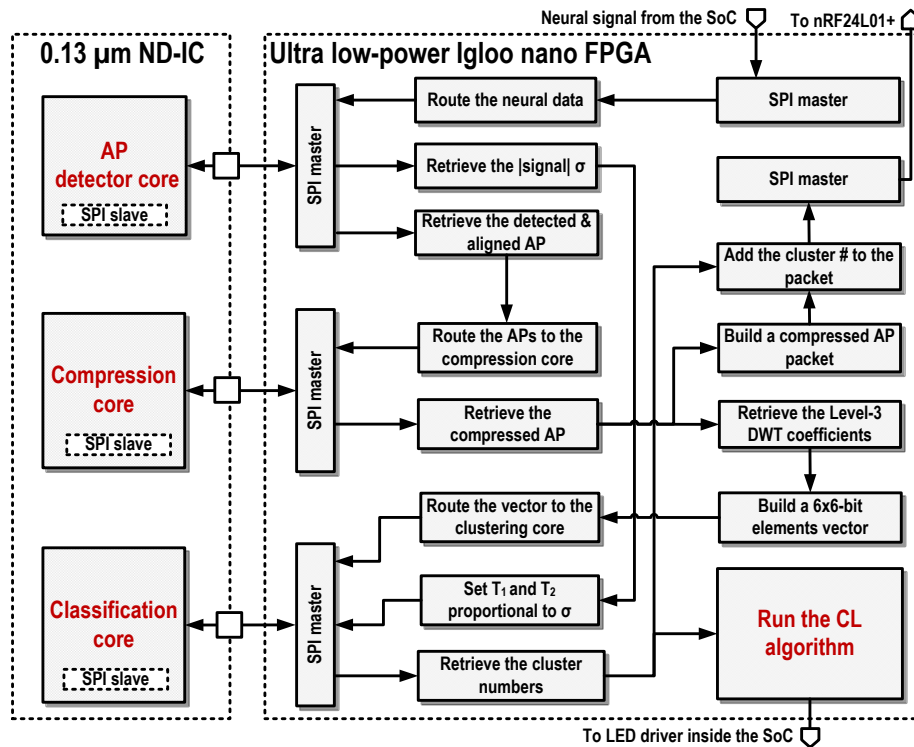


FIGURE 4.2 – The FPGA control sequence used for interfacing the ND-IC cores, the SoC and the wireless transceiver altogether, while running the CL algorithm.

2. The *Compression core* is using a Symmlet-2 (Sym-2) discrete wavelet transform (DWT) lifting scheme followed by dynamic coefficient discrimination using a sorting tree and dynamic coefficient re-quantization scheme for neural data compression [20]. Each AP waveform is compressed for saving power and bandwidth, and transmitted wirelessly along with the *classification ID* for live monitoring. Additionally, the data compression allows to operate the wireless transceiver in *low-data rate* mode, which extends the lifespan and increases the transmitting range.
3. The *Classification core* sorts the AP on-the-fly according to their wave shapes. For performing automatic classification, this core reuses i) a subset of the DWT coefficients from the compression core, and ii) the estimated signal's standard deviation calculated by the detector core.

In addition to the ND-IC, the headstage encompasses three other building blocks : i) a custom mixed-signal 0.13- μm CMOS system-on-chip (SoC) for simultaneous neural recording and optogenetic stimulation, which was presented in [9, 137] and which is depicted in Fig. 4.1(c). It includes all the circuits required to perform multichannel neural recording and optogenetic stimulation. It allows for conditioning and sampling the low-amplitude extracellular AP over 10 channels in parallel and to perform optical stimulation with a 4-ch LED driver

circuit performing precise current regulation, ii) a ultra low-power Igloo AGLN250 FPGA ($5 \times 5 \text{ mm}^2$ μ BGA) controller from Microsemi, USA, and iii) a low-power nRF24L01p wireless transceiver from Nordic Semiconductor, Norway. The wireless transceiver is connected to a 2.4 GHz ceramic antenna, and is configured to operate at 0 dBm and at either 250 kbps, 1 Mbps or 2 Mbps (user selectable), which can be set according to the required power consumption and range of the system. A 1.9-V low-dropout regulator (LDO) provides the voltage for the wireless transceiver and for one of the FPGA IO banks, while another 1.2-V LDO provides the voltage for the rest of the system.

As shown in Fig. 4.1(d), the FPGA controller is used for interfacing with the proposed ND-IC chip, the mixed-signal SoC and the wireless transceiver. The use of an FPGA provides all the necessary flexibility to ease system integration. An Igloo FPGA from Microsemi was selected for its low power consumption and its small footprint. This controller is interfaced with the ND-IC chip and with the other building blocks using dedicated master SPI modules. The FPGA control sequence is depicted in Fig. 4.2. Mainly, the neural data is routed from the mixed-signal SoC readout to the ND-IC detector core, then the detected APs are forwarded to the ND-IC compression core, and finally a subset of the compressed waveforms DWT coefficients are routed to the ND-IC classification core. After performing data decoding, the sorted neural data (*cluster number*, *timestamp* and *compressed samples*) are encapsulated into data packets and sent to the wireless transceiver. Meanwhile, the cluster number and AP timestamp are routed to the CL algorithm. When a CL stimulation is required, the FPGA controls the light pulse sequence by communicating with the mixed-signal SoC [9].

4.5 Closed-Loop Strategy

A CL link can be established between the stimulator and the neural recording circuit of the mixed-signal SoC through the ND-IC. The FPGA is counting the number of detected APs in real time within a sliding time window and with respect to their respective cluster associations (i.e. classification) established by the classification core inside the ND-IC. Each AP-to-cluster association can be set to trigger a stimulation pattern after a predetermined amount of occurrences found within a sliding time window. The duration of the sliding window can be configured to target numerous recognizable firing patterns, like AP bursts of varying lengths, while the stimulation patterns, i.e. number of pulses, pulse width, and the stimulation channel, are also configurable.

The CL scheme, whose diagram is shown in Fig. 4.3, is optimized to use only basic FIFO structures implemented inside the low-power FPGA. Each cluster pertaining to each channel is associated with a unique FIFO in SRAM memory, which contains the latest AP timestamps for that cluster. When a new AP-to-cluster association becomes available, a memory pointer to the FIFO structure belonging to that cluster is retrieved. If the FIFO is empty, the AP

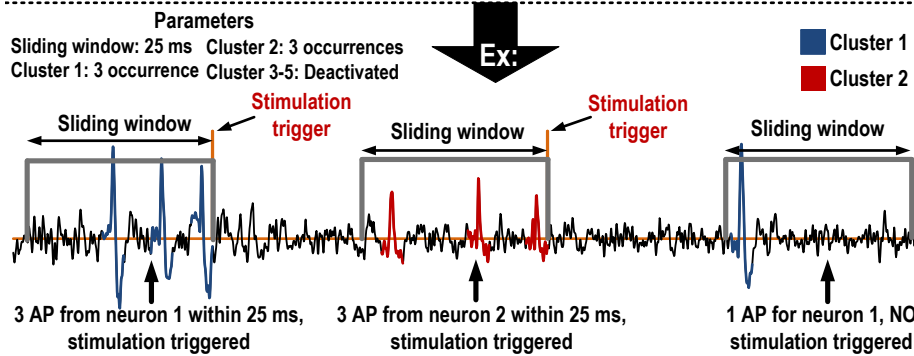
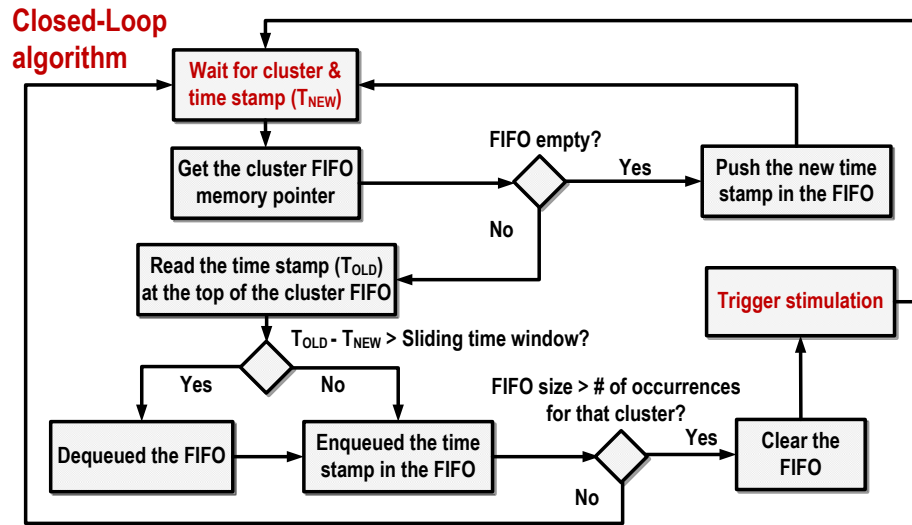
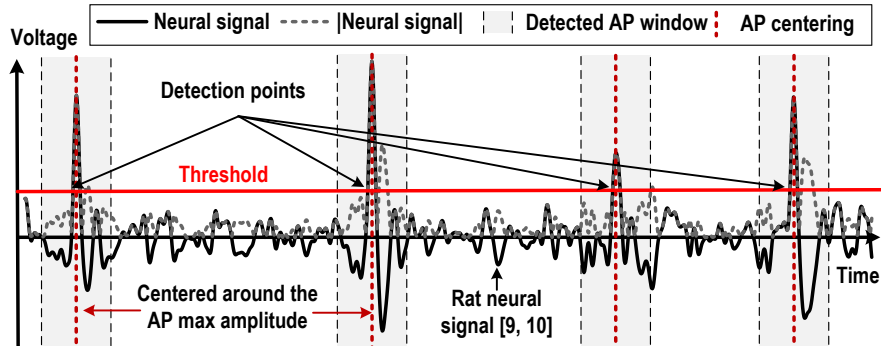


FIGURE 4.3 – Diagram of the CL algorithm (top). This algorithm counts the number of detected AP and their cluster association within a sliding time window. For example, the CL algorithm can trigger a feedback stimulation pattern if 3 similar i.e. AP from a same cluster, are within a 25-ms time window (bottom).

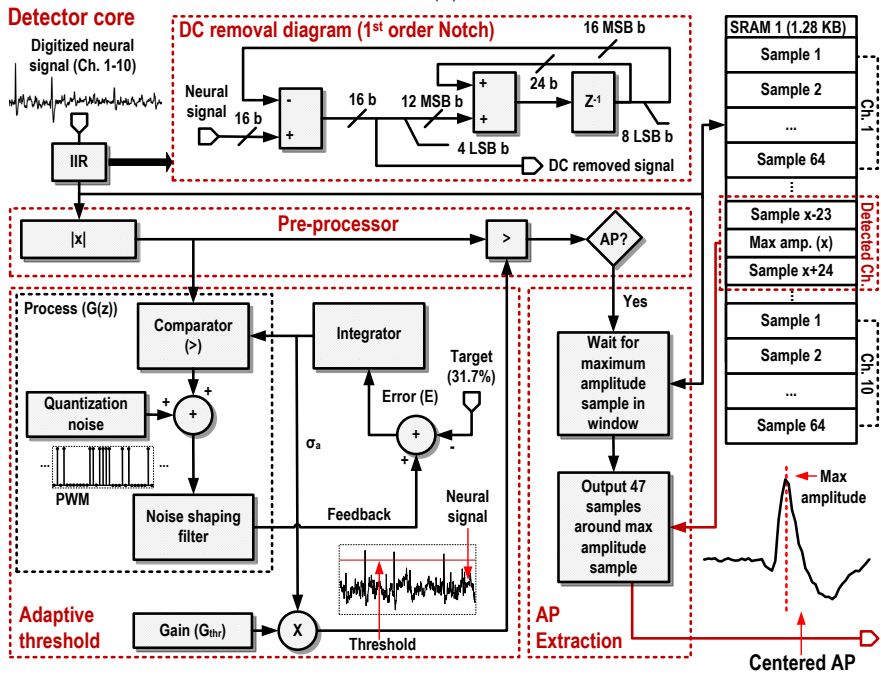
timestamp (T_{NEW}) gets enqueued. Otherwise, the module reads the older timestamp (T_{OLD}) at the top of the FIFO. If $T_{OLD} - T_{NEW}$ is greater than the sliding window length, the top of the FIFO dequeued. Then, T_{NEW} is enqueued, and if the FIFO size is equal to the number of APs required to trigger a feedback stimulation pattern, the FIFO is cleared and a trigger is issued.

4.6 ND-IC Detector Core

The proposed detector core has two main modules : the pre-processor for conditioning the neural data and the adaptive threshold for discriminating the APs from noise.



(a)



(b)

FIGURE 4.4 – a) The pre-processor concept, when the amplitude of the neuronal signal exceeds a threshold, retrieves a detected AP waveform window set around the peak amplitude, and b) The detector core diagram.

4.6.1 Pre-Processor

Methodology

The proposed pre-processor is based on the absolute value operator [86]. As previously demonstrated, this simplified operator can be as efficient as a more sophisticated energy-based operator, such as the non-linear energy operator (NEO), the smoothed NEO and the matched-filter (MF), providing an excellent trade-off between low complexity and high performance [86, 138]. The absolute value operator concept is shown in Fig. 4.4a. Upon threshold crossing, the APs are extracted and aligned on their peak amplitude sample.

Implementation

Fig. 4.4b shows the detector core diagram. First, the neural signal is high-pass filtered by an infinite impulse response (IIR) filter to remove any DC level. A power and area efficient 1st order notch filter is implemented. This filter uses bit pruning and addition/subtraction operations only. Upon signal threshold crossing, the peak sample is found across a 31-sample window immediately starting after the detection point. Then, a frame of 48 samples including the peak at the middle is forwarded to the compression core. Such an alignment around the peak value is used to facilitate AP classification, given that two APs emitted from the same neuron might appear to belong to different clusters if they were misaligned and not properly extracted.

4.6.2 Adaptive Threshold

The output of the absolute value operator is compared with a threshold, the value of which is recalculated in real time according to the characteristics of the neural signal. The proposed threshold calculation strategy is based on the estimated standard deviation (σ_n) of the noise embedded in the neural signal. This strategy is effective and robust, and is used extensively in electrophysiology [88, 139]. The detector core computes σ_n through a feedback loop that is using the statistical property of the signal's noise [20, 88], as seen in Fig. 4.4b. The adaptive threshold computation technique is similar to the FPGA-based design introduced by our group in [20], but is adapted for an efficient CMOS implementation within the ND-IC.

4.7 ND-IC Compression Core

4.7.1 Methodology

The detected APs are forwarded to the Compression core, which can increase the number of recording channels, reduce the power consumption, and increase the transceiver range, while providing a dimensionality reduction to help the classification core (covered in Section 4.8). An earlier version of this algorithm has been demonstrated by our group in [20] using an FPGA-based headstage. As shown in Fig. 4.5, the DWT-based compression technique uses three pipelined stages. First, a 4-Level Sym-2 lifting DWT is applied, followed by a coefficient sorting stage, and finally by a compression stage. The proposed CMOS implementation in this paper requires much less circuit area, memory ($2\times$), and power ($\sim 75\times$), than the system demonstrated in an FPGA.

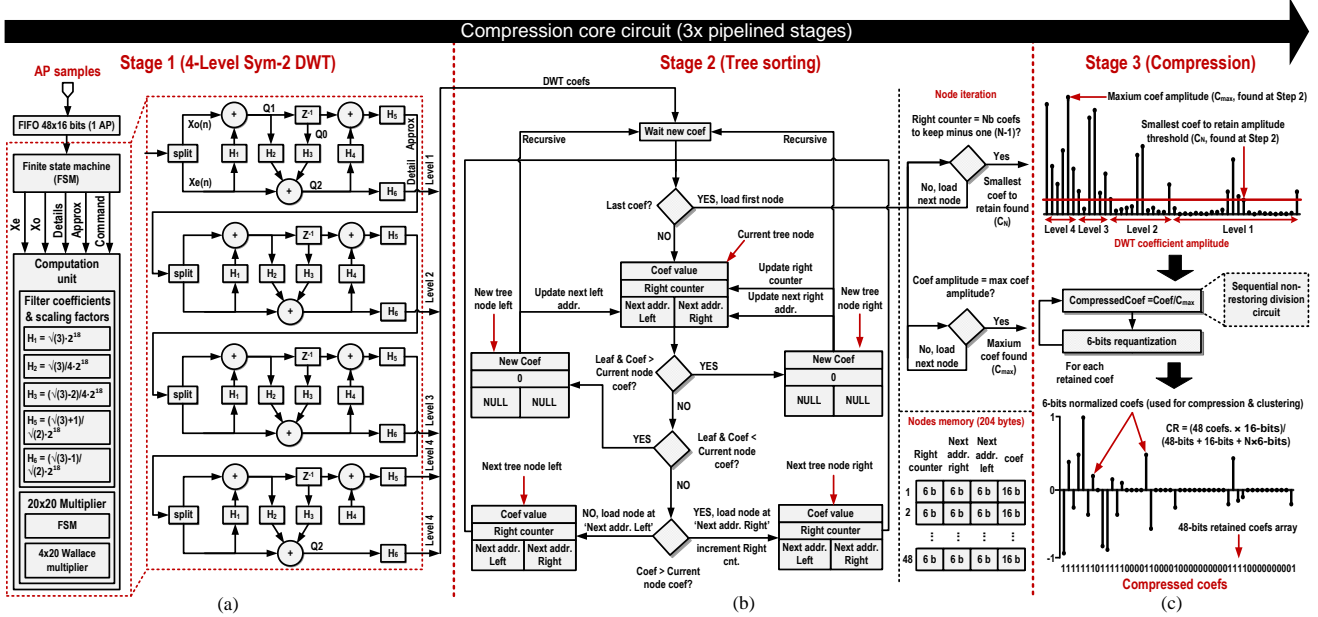


FIGURE 4.5 – The AP compression core circuit consists of three pipelined stages : a) a 4-Level Sym-2 lifting sequential DWT is performed on each detected AP, b) the detailed coefficients are processed by a sorting tree to find the coefficient having the N^{th} and the maximum amplitude, c) compression occurs by dynamically thresholding the DWT coefficients to keep only N of them. Further compression is provided using a dynamic re-quantization operation.

4.7.2 Implementation

Symmlet-2 Lifting DWT Stage

In the first stage, a Sym-2 lifting DWT is performed on each of the detected AP waveforms. The Sym-2 lifting scheme diagram is shown in Fig. 4.5(a), and is performed sequentially using a FIFO, a finite stage machine (FSM) and a Computation unit. When a new AP becomes available inside the FIFO, the FSM begins the lifting sequence by sending commands—along with the AP samples—to the computation unit. Within this unit, multiplications involving the filter coefficients are performed with a simplified custom 16x20 bits multiplier circuit, consisting of an FSM and a 4x20 Wallace tree. This topology can save significant circuit area compared to [20] (~80% smaller). The filter coefficients and scaling factors H_1 - H_6 , (values provided in Fig. 4.5(a)), are stored inside general-purpose constant registers within the computation unit. The lifting FSM routes the coefficients to the next DWT level as well as each detailed coefficient to the next coefficient sorting stage, so the coefficients can be sorted while the lifting is being performed.

Coefficient Sorting Stage

The proposed system uses a new sorting tree circuit to find the coefficient having the N^{th} amplitude (C_N). As shown in Fig. 4.5(b), the proposed sorting tree circuit has 48 nodes, each

of which encompasses different memory space, including a DWT coefficient, the number of greater coefficients, and two memory pointers. The left and right memory pointers each link to the next node with a smaller or greater coefficient amplitude, compared to the current node coefficient. The tree is gradually filled, as the coefficients are produced by the lifting circuit, and the new nodes are stored within a 204 bytes register memory. The tree works as follows :

- (i) When a new DWT coefficient is ready, the first node in memory is loaded. If the tree is empty, the root node is created at the first address in memory, otherwise step (ii) is executed.
- (ii) If the new coefficient amplitude is greater than that of the currently loaded node, the number of greater coefficients is incremented and the right node is loaded, otherwise the left node is loaded. Step (ii) is repeated until a NULL pointer address is reached. Then, a leaf node is created at the next available memory address, and the pointer of the node currently loaded (right or left) is updated with the new node address. Before overwriting the new node, the former coefficient value is sent to the next compression stage, as to perform sorting and compression in parallel. If the last coefficient is reached, step (iii) is executed, otherwise step (i) is executed.
- (iii) The circuit iterates on the nodes to find C_{max} and C_N , i.e. those having the maximum and $N - 1$ greater coefficients (*Right counter* in Fig. 4.5(b)) respectively. Meanwhile, the node pointers are reset with a NULL address. Go to step (iv).
- (iv) C_{max} and C_N are forwarded to the next compression stage, step (i) is executed.

This algorithm has a complexity of $O(n \cdot \log(n))$, an improvement compared with $O(n^2)$ in [20], and uses two times less memory.

Compression Stage

During the compression stage, the coefficients greater or equal to C_N are retained, normalized, and quantized from 16 to 6 bits, while the other coefficients are discarded [20]. The position of the retained coefficients are stored inside a 48-bits array. A coefficient normalization is performed by dividing the retained coefficients by C_{max} , while a requantization is performed by keeping the 6 MSB of the division. The 16 bit per 16 bit division is performed with a sequential *non-restoring division* circuit, which yields a simplified implementation. Only C_{max} is kept on 16 bits for allowing a resolution equivalent to the original neural data in the AP reconstruction.

4.8 ND-IC Classification Core

4.8.1 Methodology

A major burden for AP classification is the dimensionality reduction step, which consumes time, power, and area [50, 132]. Fortunately, the aforementioned compression core produces

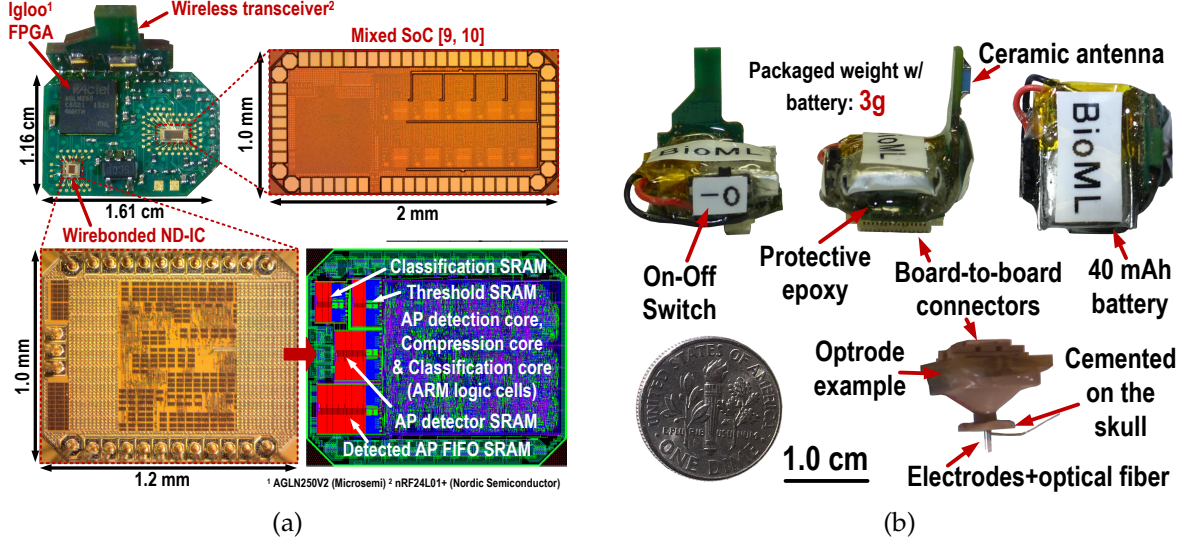


FIGURE 4.7 – a) Chip micrograph, printed circuits board floorplan and wireless headstage w/ ND-IC for CL optogenetics (1.05 g). b) Different views of the packaged headstage w/ the battery (3.0 g), a typical optrode is also shown.

$$V_{fed} = V_{AP} + V_{Noise} \quad (4.1)$$

where V_{AP} is a noiseless AP waveform, and V_{Noise} represents an additive noise vector. The Sym-2 DWT has a major advantage of being an orthonormal transformation. A property of such a transformation is that a Gaussian white noise remains Gaussian under an orthogonal transformation. Thus, if V_{Noise} is fed at the input of the Sym-2 DWT module, each level will produce a noise vector with a scaled standard-deviation compared to the original noisy vector. The Sym-2 DWT of Fig. 4.5(a) can be expressed in terms of orthogonal filter convolutions, namely with the filters $H(z)$ and $G(z)$ producing the approximative and detailed coefficients respectively. The L3 output vector ($V_{L3}(z)$) of the DWT is expressed by :

$$V_{L3}(z) = [G(z) \otimes [H(z) \otimes [H(z) \otimes V_{fed}(z)] \downarrow 2] \downarrow 2] \downarrow 2, \quad (4.2)$$

where $\downarrow 2$ represents a downsampling by 2 operation, \otimes is a convolution operation, and the filters $H(z)$ and $G(z)$ are expressed by :

$$\begin{aligned} H(z) &= H_5[H_2 + (H_1H_2-1)z^{-1} + (1-H_3)z^{-2} + H_1(1-H_3)z^{-3}] \\ G(z) &= H_6[-H_2 + (1-H_1H_2)z^{-1} + H_3z^{-2} + H_1H_3z^{-3}], \end{aligned} \quad (4.3)$$

where $H_1 - H_6$ are the filter coefficients and scaling factors of the Symm-2 DWT. Computing the amplitude of these filters at $\pi/2$ can find the gain at each DWT level : $|H(\frac{\pi}{2})| \approx 1.13$ and $|G(\frac{\pi}{2})| = 1$, meaning that if the DWT input is V_{Noise} , the L3 output vector ($V_{L3,N}$) will

also be Gaussian with an amplitude gain of 1.28, and a standard deviation 1.28 times greater than V_{Noise} . Thanks to the adaptive threshold scheme provided in the AP detector core, we already computed σ_a of $|V_{Noise}|$, which relates to the standard deviation of V_{Noise} by the scaling factor $\sqrt{1 - 2/\pi}$. Thus, an approximated value of the standard deviation of the L3 coefficient ($\sigma_{L3,N}$) is already available according to V_{Noise} .

The calculation of the L₁ distance between the cluster average vector and V_{L3} is influenced by the noise in V_{fed} , which is reflected as additive noise inside V_{L3} as stated above. This bias error can be modeled by summing the absolute value of the components of $V_{L3,N}$. Since the components of $V_{L3,N}$ are assumed to be independent, the standard deviation of the bias error (σ_e), i.e. the σ of $\sum_{n=0}^5 |V_{L3,N}(n)|$, is equal to $\sqrt{1 - 2/\pi} \cdot \sqrt{6} \cdot \sigma_{L3,N}$. Since $\sigma_{L3,N}$ is already known, we set T_1 and T_2 as multiple of σ_e , the value of which is :

$$\sigma_e = \sqrt{1 - 2/\pi} \cdot \sqrt{6} \cdot \sigma_{L3,N} = 1.28 \cdot \sqrt{6} \cdot \sigma_a \approx 3.14 \cdot \sigma_a, \quad (4.4)$$

as to reduce the classification errors due to the signal noise.

4.8.2 Circuit Implementation

As depicted in Fig. B.17, the classification core includes a FSM and a 480-bytes SRAM memory block. This memory is separated in 10 segments of 32×12 bits each, which are used for storing up to five cluster averages and the number of active clusters (N_{active}) per channel. Upon the reception of a new V_{L3} vector, the FSM computes the L₁ distance between V_{L3} and each cluster average. The cluster averages vector components are using a 12-bit fixed-point notation, in which only the 6-MSB are used to compute the L₁ distance. If a new cluster is created (i.e. $D_{min1} > T_1$), N_{active} is incremented and V_{L3} is written in the SRAM at the N_{active}^{th} cluster average address. Otherwise, V_{L3} is associated to the closest cluster before its average vector is updated using a cumulative moving average calculation. This is performed according to the following equation :

$$C_i(n) = C_i(n - 1) + (V_{L3,i}(n) - C_i(n - 1))/2^6, \quad (4.5)$$

where C_i is the i^{th} element of the cluster average vector C and $V_{L3,i}$ is the i^{th} element of the new vector V_{L3} . Equation (4.5) is computed using only simple bit shifting and addition/subtraction operations. Then, the FSM computes the L₁ distance between the updated vector C and all the other cluster average vectors. If two clusters get too close to each other (i.e. $D_{min2} < T_2$), they are merged together by taking the average of the two vectors. Then, N_{active} is decremented, and the FSM reorders the cluster averages stored in the SRAM. At worst, forty clock cycles are required to complete a classification step, which yields a minimal operating frequency of 40 kHz to handle 10 channels, assuming a maximal firing rate of 100 AP/s [95].

TABLE 4.1 – HEADSTAGE SUMMARY

# Record./Stim. Channels [9, 137]	10/4
Bandwidth (tunable for AP and/or LFP)	0.5 Hz -7 kHz
Features	CL opto./ Neural Rec.
Volume (w/ a 40 mAh bat.)	1.13 cm ³
Weight (w/ a 40 mAh bat.)	3.0 g
Power (no stim., *)	21 mW
Power (20 mA, 1x LED, *)	97 mW
CL latency **	0.823 ms (typical)
Maximum temperature increase ***	0.9 °C
Autonomy (no stim., w/ a 40 mAh bat., *)	2h40

* 1 ch. continuously transmitted (240 kbps) at 2 Mbps on-air

** Delay between AP detection and CL stimulation

*** At the bottom of the optrode, 10 ms pulses of 20 mA at 100 ms intervals

4.9 Wireless Headstage Integration

The integrated headstage and the ND-IC micrograph are shown in Fig 4.7a, along with the ND-IC layout, in which the circuits and SRAM blocks are identified. The SRAM occupies ~15% and the digital circuits ~50% of the 1.2 mm² ND-IC die area. The ND-IC is directly wirebonded on the headstage PCB along with the mixed-signal SoC to save space, next to the low-power FPGA. The headstage consists of two interconnected PCBs : i) the base PCB where the ICs are wirebonded, and ii) the PCB that holds the wireless transceiver placed at 90 degrees from the first PCB. The fully packaged system is shown in Fig. 4.7b. A drop of epoxy is applied on the dies to protect the wirebonds, and a 40-mAh LiPo battery is placed above the electronic components to fill any unused space. The maximum temperature increase that was measured at the bottom of the optrode (Fig. 4.7b) over the whole lifespan of the battery) is 0.9 °C. This temperature increase is in the range of normal physiological conditions [140], and cannot harm the animal tissues. The packaged system with the battery weighs 3.0 g, and occupies 1.13 cm³. The headstage characteristics are presented in Table 4.1.

4.10 ND-IC Experimental Results

This section presents the ND-IC measured results, while the headstage validation results are reported in Section 4.11.

4.10.1 AP Detection

This section presents simulation results of the proposed AP detector, and a comparison with other detection techniques. For these tests, we generated synthetic neuronal signals built using real, previously-recorded AP waveforms. These signals were designed to have a non-constant firing rate and a realistic SNR by adding both white Gaussian noise and neuronal

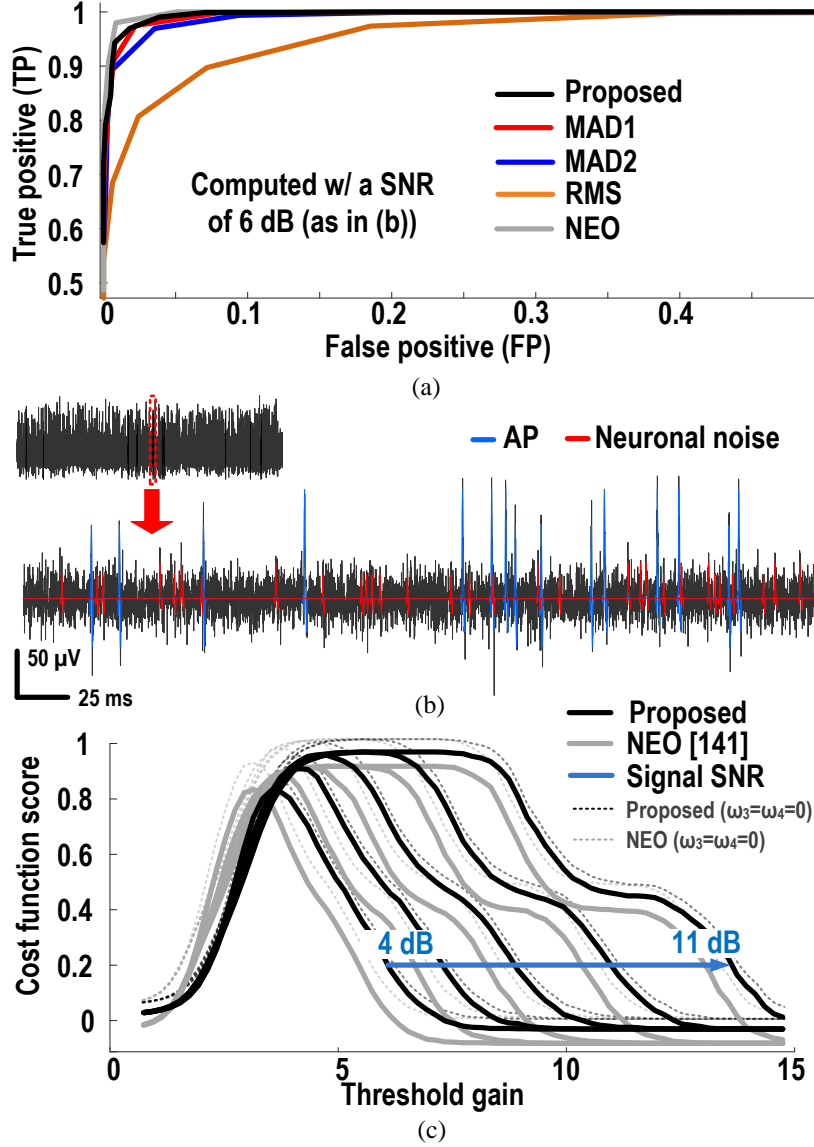


FIGURE 4.8 – a) ROC curves for a 6 dB signal w/ varying firing rate, b) subset of the signal used to compute the ROC in (a), and c) CF scores for the proposed (black) and the NEO (gray) preprocessors for different SNR. The CF scores when neglecting the algorithm complexity are shown with dotted lines.

noise. The SNR is measured as follows :

$$SNR = 10 \cdot \log(\sigma_{AP}^2 / (\sigma_{noise}^2 + \sigma_{neuronal}^2)), \quad (4.6)$$

where σ_{AP}^2 is the mean variance of the APs, σ_{noise}^2 is the variance of the added neural noise and $\sigma_{neuronal}^2$ is the mean variance of the neuronal noise.

We compared the absolute value operator used with the MAD threshold (MAD1, $\delta = \text{median}\{|x|\}/0.6745$) [91, 139, 141], the MAD threshold windowed (MAD2), the RMS threshold win-

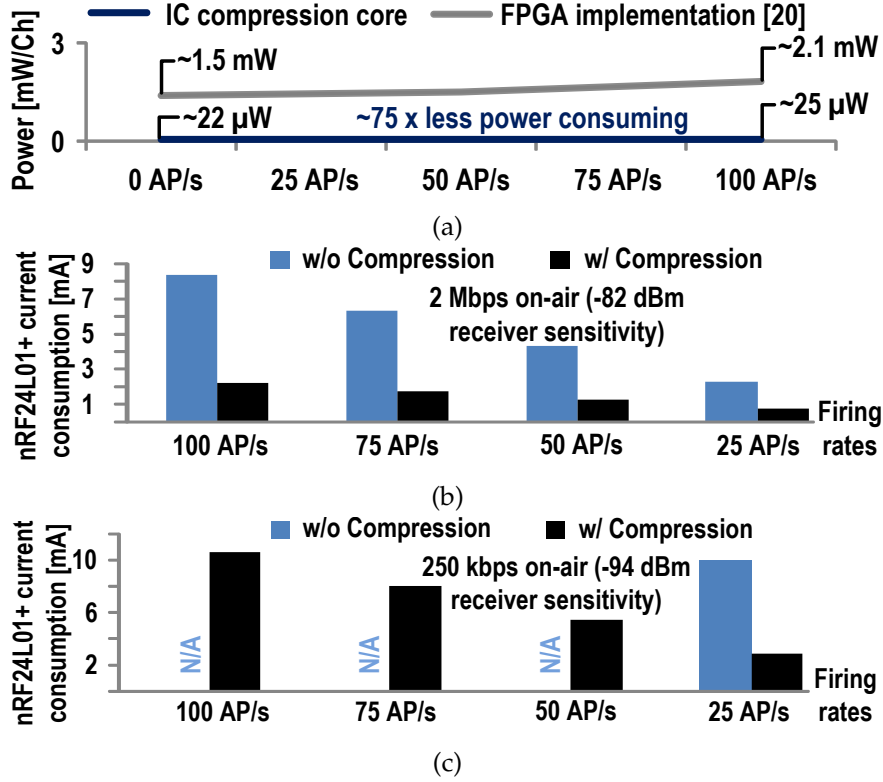
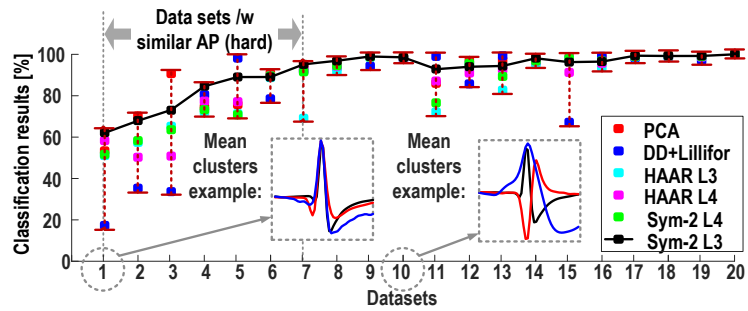


FIGURE 4.9 – Power consumption of the proposed compression core vs the implementation in [20] for five different firing rates, b) and c) current consumption of the wireless transceiver with and without compression (CR of 4.17) at 2 Mbps and 250 kbps on-air data rates, respectively.

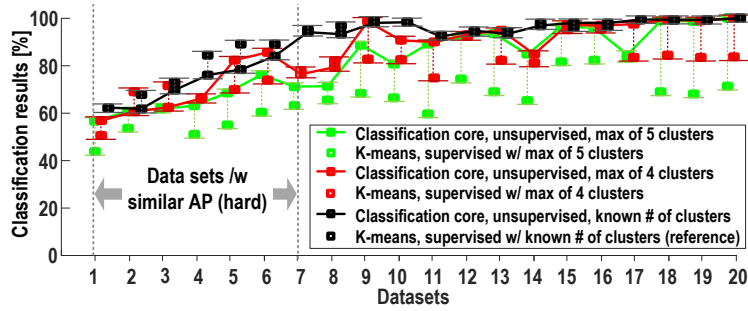
dowed [142] and the NEO operator with the MAD threshold [141]. Both windowed algorithms are calculated over a window of 50 ms, as suggested in [79]. Fig. 4.8(a) shows the receiver operating characteristics (ROC) [86] computed with a signal with a standard SNR of 6 dB, as seen in 4.8(b). As can be seen, the proposed algorithm performs better than RMS and MAD2, and has equivalent or better performance than MAD1. However, our algorithm is subtly overtaken by NEO. The trade-off between this slight ROC increase versus complexity for NEO needs to be further explored. For that purpose, we devised a simple cost-function (CF) :

$$CF = \omega_1 \cdot TP + \omega_2 \cdot FP + \omega_3 \cdot A_i / A_j + \omega_4 \cdot P_j / P_i, \quad (4.7)$$

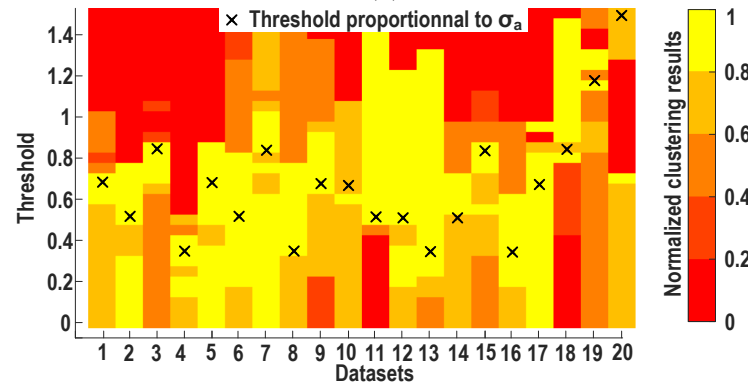
where $\omega_{1,2,3,4}$ are the weight (1, 1, 0.025 and 0.025 respectively), A_i and P_i are the synthesized estimated circuit area and estimated dynamic power consumption, respectively, of the algorithm being tested, while A_j and P_j represent the same parameters for the other algorithm. Post-synthesis power and area estimations (using Synopsis) give 30 nW ($F_{clk}=20$ kHz) and 0.013 mm² respectively for the proposed algorithm, while similar estimations using the same 0.13- μ m process reported in [141] yield 50 nW ($F_{clk}=20$ kHz) and 0.021 mm². Higher



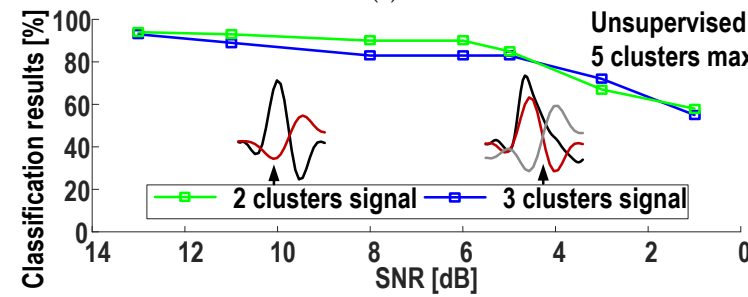
(a)



(b)



(c)



(d)

FIGURE 4.10 – a) Classification results using the L3 Sym-2 DWT (black curve) versus the other techniques, b) unsupervised classification results of the proposed algorithm (solid lines) compared with K-means (squares), c) normalized classification performance according to the threshold T_1 , and d) classification results according to the signal SNR.

area and power consumption for NEO were predictable, since it requires a minimum of two extra multiplications. The CF results are presented in Fig. 4.8(c) for an SNR ranging from 4 to 11 dB. As seen, both algorithms yielded the same CF results for an SNR of 4-dB, while the proposed algorithm scored higher for the other SNRs. The 4-dB CF is due to NEO performing well with extremely noisy signals, but since both detectors produce similar detection performances under more realistic SNRs, the NEO CF is smaller due to a higher complexity.

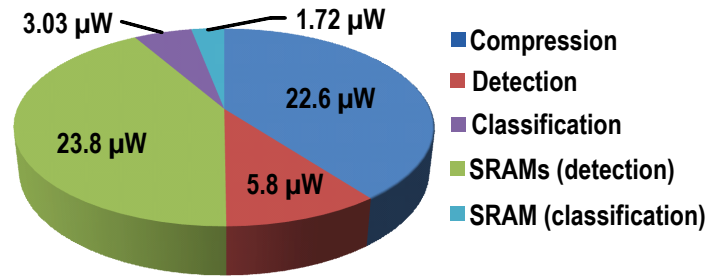
4.10.2 Compression

A signal-to-noise distortion ratio (SNDR) analysis using *in vivo* collected AP waveforms is provided in Section 4.11, while the benchmarking of the compression algorithm is provided in [20]. Fig. 4.9a shows the power consumption measurements of the proposed compression core versus the implementation in [20]. On average, the proposed IC compression core consumes $\sim 75\times$ less power, justifying the integrated approach. Fig. 4.9b and 4.9c show the current consumption of the wireless transceiver with and without compression (CR of 4.17) at 2 Mbps and 250 kbps on-air datarates respectively. As can be seen, compression significantly reduces the power consumption (by up to $3.8\times$) and increases the range by allowing lower datarates. At 250 kbps, receiver sensitivity is improved by -12 dBm (twice the range), which would not have been possible without compression at firing rates higher than 25 AP/s.

4.10.3 Classification

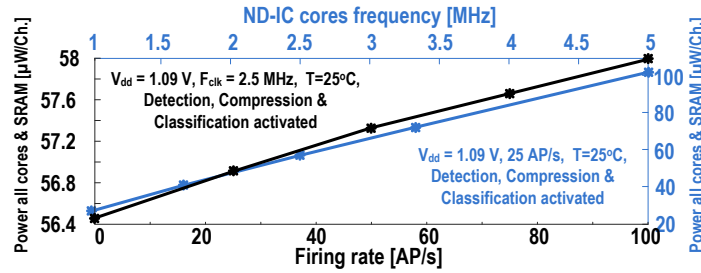
This section presents the measured results of the proposed classification algorithm when compared to other sorting techniques that were already proven in hardware implementations [132, 143, 144]. For the first tests, we used readily available synthetic neural signals built using real AP waveforms [145]. This public neural data bank is used to benchmark the Waveclust off-line neural sorting software [139], making it perfectly suited to benchmark our algorithm. We used 20 datasets, each of which encompass, on average, 3,400 different AP waveforms belonging to up to three clusters per set. Most of the datasets include overlapping AP, adding an extra level of difficulty.

The proposed dimensionality reduction technique is compared with 1) the more computationally-intensive principal component analysis (PCA) [132, 143] (w/ three components) as a reference (used as "ground truth"), 2) with the discrete derivative (DD) [143, 144] (using 10 coefficients found with the Lillifor test for normality [143]), and 3) with a DWT algorithm based on the Haar mother wavelet (which is the simplest and most popular mother wavelet). The dimensionality reduction for the Sym-2 and for the Haar DWT are validated using the L3 and L4 coefficients. We used the same K-means classification algorithm to benchmark all the dimensionality reduction techniques.



Total (25 AP/s): 56.9 μW/Ch; $V_{dd} = 1.09$ V, $F_{clk} = 2.5$ MHz,

(a)



(b)

FIGURE 4.11 – ND-IC power consumption : a) power breakdown for each core and SRAM block, b) power consumption according to the neural firing rate (black), and power consumption according to the cores operating frequency (blue).

Fig. 4.10a shows the classification results. The black curve shows the proposed Sym-2 results using the L3 coefficients, while the other results are presented by a coloured square. As can be observed, the proposed technique yielded one of the best results for almost all datasets, equal to or just a little behind the PCA. Overall, the third best results are generated by the Sym-2 (L4), followed by the Haar (L4), the Haar (L3), while the worst results are generated by the DD algorithm. These results demonstrate that our proposed dimensionality reduction technique can be superior to any existing approach, while generating similar results to those of a complex PCA.

Fig. 4.10b shows the on-line unsupervised classification results of the proposed algorithm (solid lines) compared to the ones obtained with K-means (squares). In this simulation, the maximum number of clusters is limited to : 1) the known number of clusters (black), 2) a maximum of four clusters (red) and 3) a maximum of five clusters (green). As can be seen, when the number of clusters is known, the proposed algorithm behaves almost as well as K-means, while being 100% unsupervised. When the maximum number of clusters is set to four and five, the proposed technique still produces the accurate number of active clusters, and the results remain highly competitive. This contrasts with K-means, with which the results degenerate significantly when the number of clusters is overestimated.

Fig. 4.10c shows the normalized classification performance according to the threshold T_1 (T_2

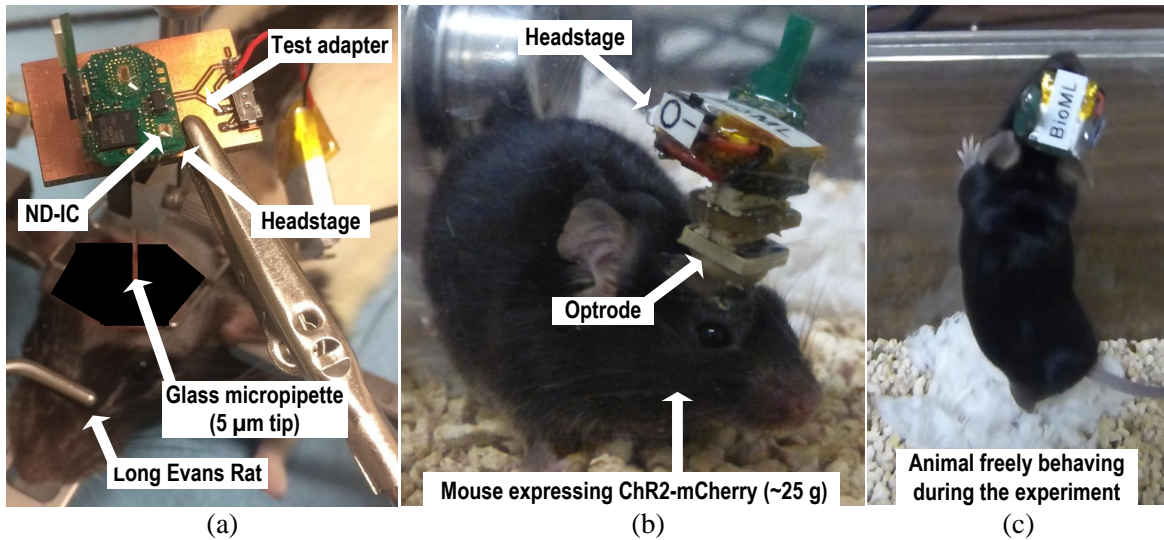


FIGURE 4.12 – a) Head-fixed *in vivo* setup with the rat. The neural signal was collected with a micropipette and processed *in situ* by the ND-IC. b) The headstage is connected to the optrode and installed on the head of the mouse. c) The freely-behaving animal during a CL experiment.

is set to $\sim 0.6 \times T_1$). The yellow area for each dataset indicates the region in which the threshold produces the optimal classification results, while the black X indicates the threshold value that was computed automatically using σ_n . We can see that the black Xs are either close to, or inside, the optimal threshold region for almost all datasets. Indeed, the calculated threshold lies in the optimal region 80% of the time, thus promising a reliable performance in real life.

Fig. 4.10d shows the unsupervised classification performances according to the signal SNR, for two neural signals encompassing AP waveforms from two and three clusters respectively, and for a maximum number of five clusters. The classification results are almost constant, and start dropping under 5 dB. Even with a low SNR of 3 dB, the classification results maintain a success rate of approximately 70%.

4.10.4 Power Breakdown & Characteristics

The power breakdown of the ND-IC is depicted in Fig. 4.11a. The detection, compression and classification cores consume 5.8 μW , 22.6 μW and 3.0 μW per channel, respectively. These measurements include the SPI modules and the communication FSMs. The power consumption of the ND-IC according to the AP firing rate is shown in Fig. 4.11b (in black). The power consumption baseline is 56.4 $\mu\text{W}/\text{Ch}$, and increases proportionally to the firing rate with a slope of 0.01 $\mu\text{W}/\text{AP}$. The AP-power relation is due to the deactivation of some circuits when no AP is being detected, compressed or sorted. Finally, the power consump-

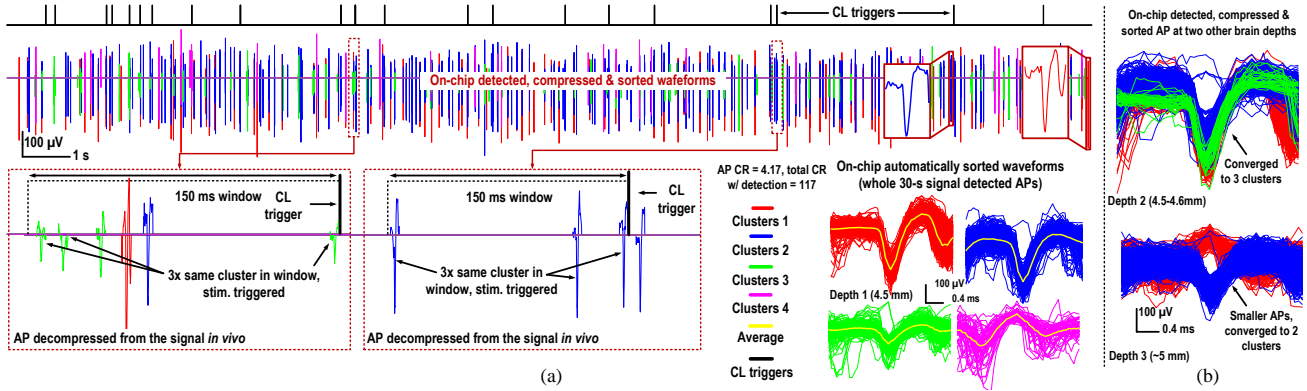


FIGURE 4.13 – a) Reconstructed signal using the detected, compressed, and sorted AP collected *in situ* in the brain of the rat. The feedback trigger signal computed by the headstage is shown (in black). The trigger was configured to occur if three AP per clusters were found within a 150-ms time window. b) Detected, compressed and sorted waveforms at two other brain depths.

TABLE 4.2 – ND-IC SUMMARY

Technology	0.13- μ m
Supply voltage	1.09 V
Die Dimension	1.2 mm ²
# of Channels	10
Features	Detection/Compression/Sorting
Input resolution	16 bits
Internal arithmetic	Fixed-Point
Area Per Channel (w/ SRAM)	0.08 mm ²
Power (25 AP/s, $F_{clk} = 2.5$ MHz/Cores)	56.9 μ W/Ch
Communication Protocol	SPI

tion of the ND-IC according to different core operating frequencies is shown in Fig. 4.11b (in blue), while the performance is summarized in Table 4.2.

4.11 Headstage *In Vivo* Results

The headstage was validated *in vivo* within two experiments, the first involving an anesthetized Long-Evans rat, and the second a freely-moving mouse. The first experiment demonstrated the system’s ability to detect, compress, sort, and generate CL trigger signals (without feedback optical stimulation) *in vivo*. Following good preliminary results, the second experiment demonstrated the CL operation of the headstage with a freely-moving mouse.

4.11.1 Surgical Procedure

For the preliminary *in vivo* trials, a Long-Evans rat was anesthetized with ketamine/xylazine solution. A glass micropipette (5- μ tip) filled with saline solution was lowered in the LD

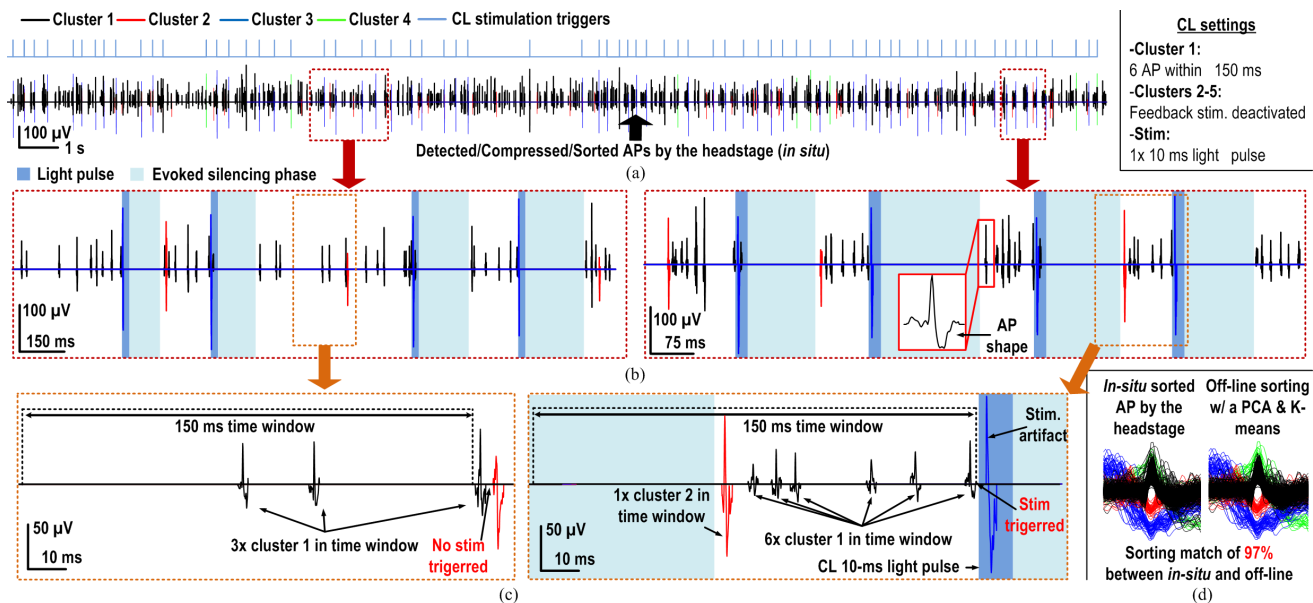


FIGURE 4.14 – a) Reconstructed neural signal and CL stimulation triggers (in blue) acquired with a freely-moving mouse expressing a ChR2-mCherry in inhibitory neurons of the prelimbic cortex, b) a two signal closeup showing the CL stimulation pulses and evoked silencing phase due to the ChR2-mCherry activation, c) two CL window enlargements with and without CL stim. trigger, and d) AP sorted live by the ND-IC and off-line by a PCA with K-means (97% matching).

brain region of the thalamus (4.5-5 mm depth). The experimental setup is depicted in Fig. 4.12(a). For the freely-moving *in vivo* trials, a B6SJLF1/J mouse (Jackson Laboratory, USA) was injected with 300 nL of AAV5-mDlx-ChR2-mCherry-Fishell-3 (Plateforme d'Outils Moléculaires, Canada ; titre 2.0E13 GC/mL), in the PL region of the medial prefrontal cortex (1.80 AP ; 0.30 ML and -1.70 DV from Bregma) followed by an implantation of a detachable optrode (Doric Lenses, Canada), consisting of one optical fiber (200- μ m core ; 0.66 NA) and four Tungsten microelectrodes (50- μ m flat tip). It should be noted that the ChR2-mCherry stimulation will result in the activation of the interneurons, an inhibitory effect is therefore expected. The recordings were carried out seven months after the viral injection, so the animal was able to fully recover from the surgical procedure and to adopt a normal behaviour with the headstage in place. Fig. 4.12(b) and 4.12(c) show the awake animal during a CL experiment. All protocols were performed in accordance with guidelines from the Canadian Council on Animal Care.

4.11.2 Anesthetized Preliminary Results

In these experiments, the threshold gain was set to 3.5, the AP CR was set to 4.17 (as in [20]) and the maximum number of clusters was set to five. As shown in Fig. 4.13(a), the ND-IC enclosed within the wireless headstage was able to automatically detect, compress, and sort

TABLE 4.3 – COMPARISON WITH OTHER SYSTEMS FOR CL OPTOGENETICS, OPTOGENETICS WITH RECORDING OR CL ELECT. STIMULATION.

	PC-based CL Optogenetics		Opto. & Rec. Syst.	CLIC/Syst. for Elect. Stim.			CLIC/Syst. for Optogenetics	
	JoNe'14 [50]	eLife'15 [49]	TBioCAS'17 [20]	JSSC'11 [57]	TBioCAS'17 [135]	ISSCC'09 [132]	Front. Neurosci.'18 [136]	This Work
TECHNOLOGY [μm]	Discrete/PC	Discrete/PC	Discrete	0.35	0.18	0.35	Discrete	0.13
Nb REC. CHANNELS	32	N/A	32	8	16	16	16	10
TARGETED SIGNALS	AP	AP	AP	AP	AP/LFP	AP	AP	AP
AP DETECTION	Yes	Yes	Yes	Yes ²	Yes (analog) ²	Yes	Yes	Yes ³
ADAPTIVE DETECTION	Yes (σ based)	N/A	Yes (σ based)	No, fixed	No, fixed	Yes (σ based)	No, fixed	Yes
SIGNAL COMPRESSION	No	No	Yes	No	No	No	No	Yes
COMPRESSION TYPE	–	–	Sym-2 DWT	–	–	–	–	Sym-2 DWT
SORTING	Yes	Yes	No	No	No	Yes	No	Yes
SORTING FEATURES	Temp. Match.	N/A	–	–	–	PCA + LUT	–	Sym-2 DWT L3
CL FEEDBACK CONTROL	Class.	Firing rate	–	AP Detect. ⁴	PID ⁵	Class.	AP Detect. ⁴	Class.+ Wind.
SYST. INTEGRATION	No	No	Yes	N/A	Yes	No	Yes (wired)	Yes
SYST. Weight [g]	–	–	2.8 (w/o bat.)	–	18	–	11.3	3.0
FREELY-MOVING TRIALS	No	No	No	No	No	No	No	Yes

¹ w/ SRAM and w/o Pads, ² Time amplitude based, ³ Amplitude based, ⁴ Simple trigger on AP detection, ⁵ LFP energy & AP occurrence

the AP on-the-fly. The overall CR achieved with detection and AP compression was 117. The detected clusters were used to generate a trigger signal *in situ* (black signal in Fig. 4.13(a)) to validate the CL scheme running in real time. The trigger was configured to occur if three AP per clusters were found within a 150-ms time window. Fig. 4.13(b) shows the automatic classification results at two different brain depths. Fig. 4.13(a) and Fig. 4.13(b) show that the classification algorithm accurately sorted most of the APs, since each cluster encompasses a different AP waveform. We also see that the algorithm converged to an appropriate number of clusters.

4.11.3 Freely-moving Closed-loop Experiment

Following the preliminary validation of the system, we performed CL optogenetic stimulation with a freely-moving mouse expressing ChR2-mCherry. In this experiment, the base station was placed beside the animal cage, approximately 1.5 feet away from the animal. The threshold gain was set to 4.2, the AP CR was set to 4.17 (as in [20]), the maximum number of clusters to five, and the feedback stimulation was set to $1 \times$ pulse of 10 ms (20 mA flowing through the LED, as in [9]). The CL algorithm was configured to trigger when 6 APs belonging to cluster 1 were found within a 150-ms time window. This configuration was selected after manually looking at close successive APs belonging to the same cluster. This simple approach was used to validate the system in freely-moving animals. However, further analysis of the neural bursting activity may be required to find the optimal configuration.

Fig. 4.14(a) shows a reconstructed signal of 30s after decompression, along with some generated CL stimulation triggers (blue signal). The CL stimulation signal was of ~ 2.5 pulses/s on average, and the overall CR was of ~ 46 in this experiment. Fig. 4.14(b) shows two signal closeups in which the stimulation pulses and the evoked silencing phases, due to the inhibitory neuron activation, are visible. Fig. 4.14(c) shows two enlargements zooming in on specific AP trains. The first enlargement shows three APs belonging to cluster 1 within the 150-ms time window, thus triggering no stimulation, while the other enlargement shows 6

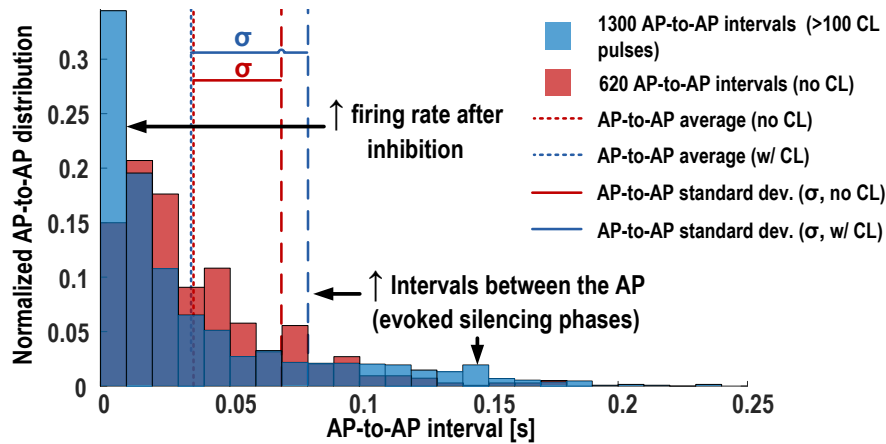


FIGURE 4.15 – Normalized distribution of the time interval between two consecutive APs for experiments with and without CL stimulation. It shows that the CL stimulation increases the number of longer AP-to-AP intervals, because of the evoked silencing phases. Moreover, the number of shorter intervals also increases, which is in accordance with [146].

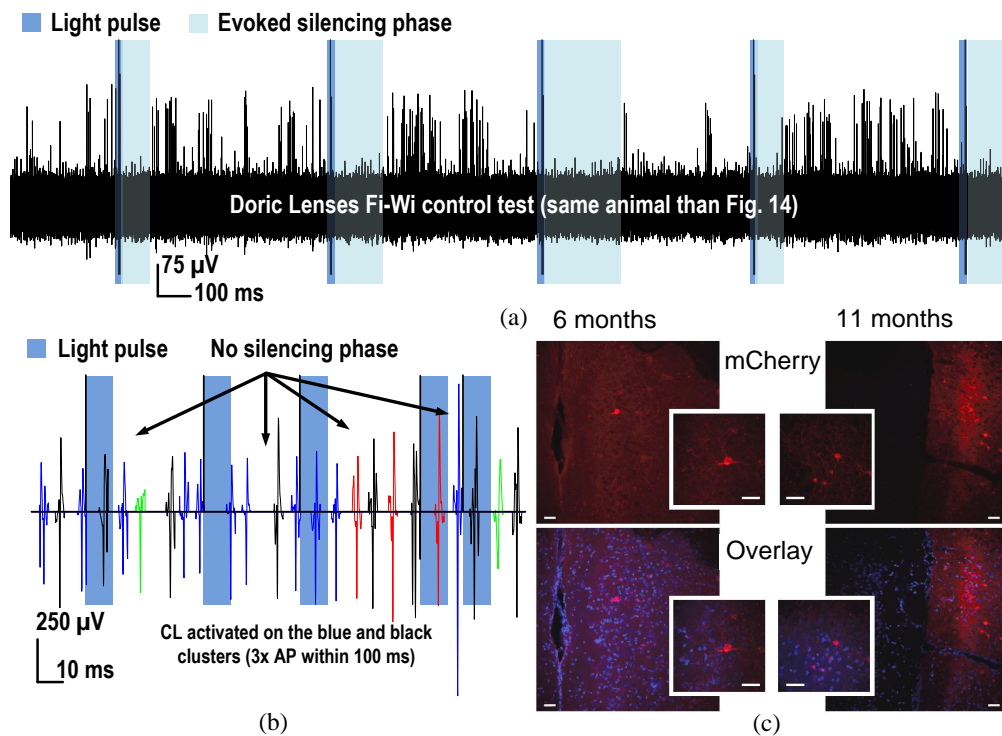


FIGURE 4.16 – Control experiments : a) Recording/stimulation using a commercial Fi-Wi system where silencing phases are induced, b) CL stimulation performed in a non-injected mouse, no silencing phase are induced and c) representative confocal images of a mouse at six months and 11 months post infection (same animal), showing the ChR2-mCherry expression (red)

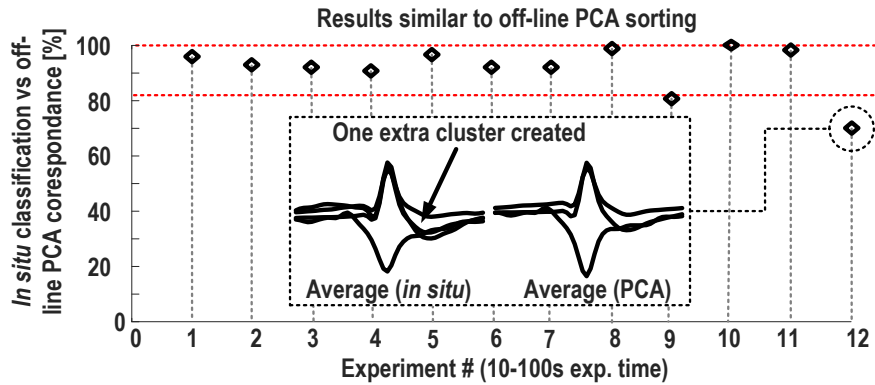


FIGURE 4.17 – Matching between the AP sorted *in situ* by the headstage vs off-line using a PCA with K-means for 12 freely-moving experiments.

APs belonging to cluster 1 within the time window, followed by a stimulation pulse and a silencing phase. Fig. 4.15 shows the distribution of the time interval between two consecutive APs for experiments with and without CL stimulation. We can see that the CL stimulation increased the amount of longer AP-to-AP intervals, which can be attributed to the evoked silence phases after each CL pulse. Moreover, we observed an increased amount of shorter intervals in the first bin of Fig 4.15. This phenomenon is widely covered in literature [146, 147], that claim the brain tends to resynchronize with higher firing rates (i.e. smaller AP-to-AP intervals) after an inhibition phase.

Stimulation artifacts are visible in Fig. 4.14. To reduce artifacts, we used blunt (opaque polyimide covered) electrodes to prevent the direct light reaching the tip of the electrode [9, 137], while the electrode tips were placed in an area with little light. However, small artifacts (caused by light or other sources) are still being detected. Assuming the optrode stays still, we expect these artifacts to have similar shapes over time. Our ND-IC, which performs on-line waveform sorting, has an important advantage to prevent noise artifacts compared to other solutions : it can process the light artifact just as any other waveshapes and sort them into a separated cluster to isolate them from the neural waveforms, similarly to the off-line strategy proposed in [41] also intended to eliminate the stimulation artifacts. The ND-IC action is illustrated in Fig. 4.14(d), where most of the artifacts were sorted within the blue cluster. It is then possible to deactivate this cluster and prevent triggering the CL stimulation upon the artifacts.

After this experiment, three control experiments were performed. The first control involved the same animal and a commercial Fi-Wi optogenetic system from Doric Lenses, Canada. Fig. 4.16(a) shows a sample signal recorded using the Fi-Wi system during stimulation. We can see that silencing phases are also observed with the commercial system after each stimulation light pulse. Fig. 4.16(b) shows CL stimulation performed in a non-injected mouse

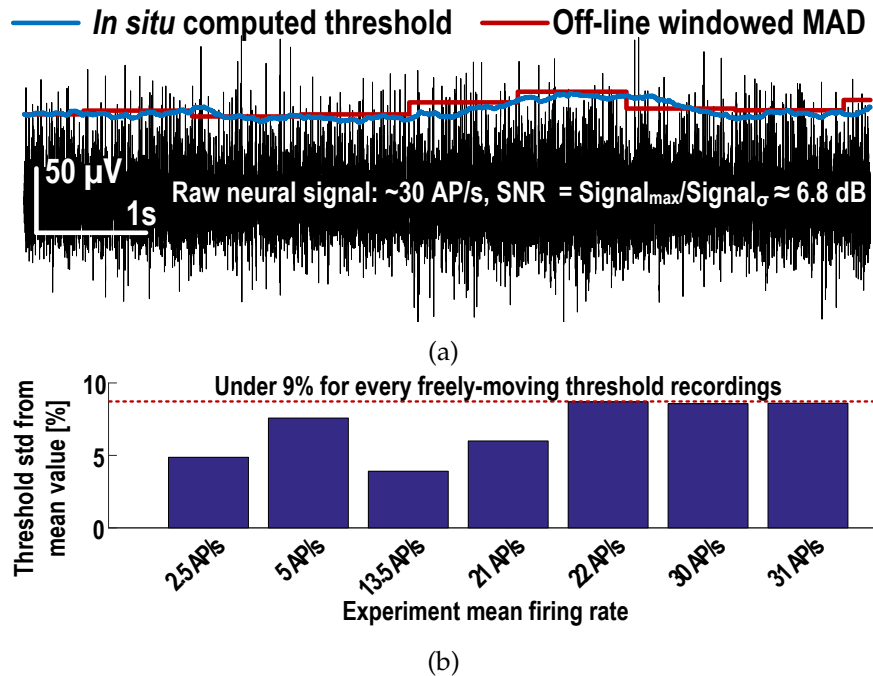


FIGURE 4.18 – a) Threshold computed by the ND-IC (blue), by MAD (off-line, red) superimposed over the RAW neural signal, b) threshold standard deviation from its mean value for seven freely-moving experiments.

(no ChR2 expression) using our headstage. No silencing phases are visible upon stimulation. Fig. 4.16(c) shows representative confocal images of mPFC sections from a mouse, six months after having performed the AAV5-mDlx-ChR2-mCherry injection (left; approximate duration post the infection of the animal at the time of the experiment), as well as those from the mouse used in the *in vivo* experiments (11 months post viral injection; right). The cells expressing the ChR2-mCherry (red) construct and the nuclei labeled with DAPI (blue) are visible, demonstrating that the ChR2 was expressed at the time of the experiment. These controls confirm that the silencing phases shown in Fig 4.14(a)-(c) were caused by inhibitory neuron activation.

The AP automatically sorted by the ND-IC are shown compared to the same AP sorted off-line by a PCA and K-means in Fig. 4.14(d). We report a classification match of 97% between the AP sorted live by the ND-IC and off-line by the PCA. To take this analysis further, Fig. 4.17 shows the classification matching between the sorted AP *in situ* by the headstage vs the off-line sorted AP using a PCA and K-means for 12 freely-moving experiments. Eleven of the twelve experiments generated a match of more than 80%, while only one produced a matching of ~70%. This lower classification matching can come from the fact that, for the same dataset, our technique have converged to an extra cluster compared to the preconfigured number of cluster used with PCA and K-means. Given that the number of clusters

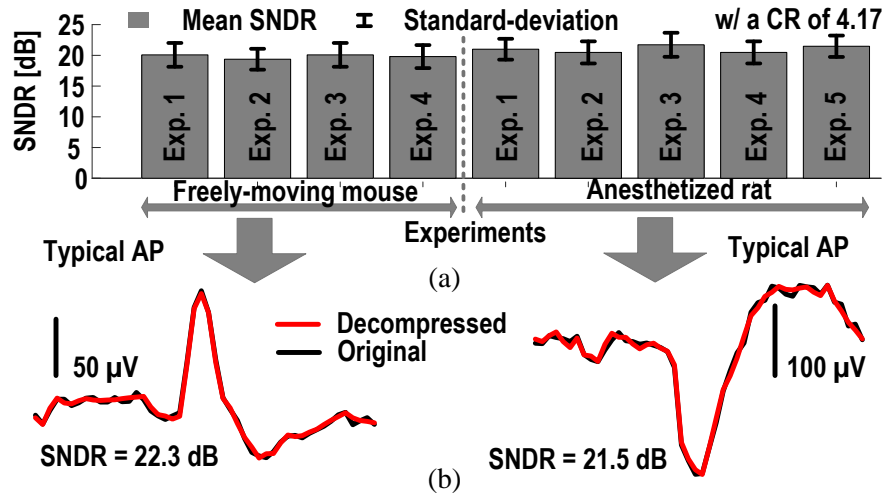


FIGURE 4.19 – a) Mean SNDR within nine experiments, b) compressed APs waveform (red) superimposed on their original counterpart (black).

is not known *a priori* for the ND-IC, the proposed DWT approach can trigger the creation of up to 5 clusters when it judges that the waveforms are slightly different, while it must be known *a priori* for the PCA and K-means. Our classifier has the advantage of using unsupervised classification without training while the PCA and K-means require the whole dataset at once. Since any additional clusters can catch a certain number of APs, classification mismatches can then occur between both classifiers. If judged necessary, it is possible to increase the accuracy of our classifier by entering the maximum number of clusters manually after data inspection, or by running a training phase on a subset of the dataset to select optimally what DWT coefficients to use in the sorting vector. This procedure would most likely increase the sorting accuracy.

We tested the adaptive threshold algorithm *in vivo* by simultaneously transmitting the threshold calculated inside the headstage to the base station along with the raw recorded signal (SNR \approx 6.8 dB). Fig. 4.18a shows the threshold calculated by the headstage (in blue), the threshold calculated using the MAD windowed operator (off-line, in red), both with a G_{thr} of 4.2. The proposed threshold is almost perfectly superimposed over the threshold calculated by the MAD operator, showing that the σ estimation and the gains are accurately computed. Since σ_n is based on a signal containing APs, it is expected that the firing rate slightly impacts the computed threshold. Fig. 4.18b shows the threshold standard deviation from its mean value for seven freely-moving experiments, having an average firing rate from \sim 2.5 AP/s to \sim 31 AP/s. We can see that the threshold varies from \sim 5% for low firing rates, and up to \sim 8% for higher firing rates, thus having a low impact on the calculated threshold value. Simulation with a synthetic signal having a realistic SNR of 6.8 dB (i.e. similar to the *in vivo* waveforms of Fig. 4.18a) shows that the AP firing rates can induce a variation of around

0.22% per AP/s in the calculated threshold. Thus, this corresponds to a variation of 2.2% for a typical firing rate of 10 AP/s, while it corresponds to a variation of 22% for a firing rate of 100 AP/s, which is less likely to occur.

Fig. 4.19(a) shows the mean compressed AP SNDR, acquired during four experiments with the freely-moving mouse and five experiments with the anesthetized rat. The SNDR is defined by : $SNDR = 20 \log(\|x\|_2 / (\|\hat{x} - x\|_2))$, where x is the original AP and \hat{x} is the reconstructed waveform. We can see in Fig. 4.19(a) that the mean SNDR is kept constant around 20 dB for all the experiments involving the two animals with different AP waveforms. Fig. 4.19(b) presents two typical compressed APs collected within the brain of the mouse and the brain of the rat, and superimposed on their non-compressed counterpart, showing they are visually indistinguishable.

4.12 Discussion

The proposed headstage and ND-IC are compared to recently published systems and experimental settings for either CL optogenetics, optogenetics with neural recording or CL electrical stimulation, as presented in Table 4.3. As illustrated, the proposed wireless headstage has features comparable to complex wired PC-based CL optogenetics settings [49, 50]. The headstage calculates its CL optogenetic triggers based on the AP classification, as in [50, 132], and also on the AP bursting [49], which are two essential features for performing neural decoding [54–56]. This contrasts with simple trigger solutions based on AP detection [57, 136], PID controller based on the LFP energy [135] or open-loop optogenetic wireless systems [20]. The proposed ND-IC integrates more features (detection, compression, and classification) than the other reported systems [57, 132, 135]. Most state-of-the-art CL optogenetic systems are based on simplistic feedback algorithms [136], and do not transmit the neural data wirelessly for providing live or post-experimental analysis. In contrast, the proposed IC is the only solution that performs AP detection and AP compression to further reduce the amount of data to transmit, which can monitor all the 10 neural channels simultaneously. Moreover, the proposed headstage has the lowest reported weight with the battery (3.0 g), which is critical to enable experiments with small freely-moving rodents. Conversely, recent publications only present ICs that are not integrated within fully working systems [132], and only tested *in vivo* with anesthetized animals [49, 50, 57, 135, 136]. To the best of our knowledge, we have reported the first integrated circuit for CL optogenetics demonstrated *in vivo* within freely-moving animal trials. Future works will focus on integrating the ND-IC, the SoC, the CL algorithm and a controller to interface with the external wireless transceiver in the same CMOS chip, to reduce further the size and power consumption.

4.13 Conclusion

We presented a miniature wireless electro-optic headstage that is the first attempt towards complex CL optogenetics in the brain of freely-moving rodents. The headstage is based on a custom 0.13- μm ND-IC that can detect, compress, and sort the neural signals on-the-fly over 10 channels in parallel without supervision, enabling a CL optogenetic strategy *in situ*. The ND-IC can detect the APs using an adaptive threshold, compress the APs using a DWT-based algorithm, and automatically classify the APs by reusing the sub-products of the compression and detector cores. The headstage can trigger complex CL optical stimulation patterns using the AP classification information along with their time occurrence within a sliding time window. Measurement and simulation results show that the ND-IC can detect, compress, and sort the AP with remarkable accuracy. The headstage and the ND-IC have been validated *in-vivo* in the brain of a rat and of a freely-moving mouse expressing ChR2-mCherry. This work reports the first wireless freely-moving CL optogenetic trials.

Acknowledgments

We thank CMC Microsystems for supporting the chip fabrication. This project is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), the *Fonds de Recherche du Québec - Nature et Technologies* (FRQNT) and the Weston Brain Institute. We acknowledge the support of the Canada Research Chair in Smart Biomedical Microsystems.

Chapitre 5

Publications et Distinctions

5.1 Liste des publications

Les contributions de ce projet de recherche ont fait l'objet de plusieurs publications scientifiques, en voici la liste exhaustive :

– **Articles de journaux (auteur principal) :**

- **G. Gagnon-Turcotte**[†], U. Côté-Allard[†], F. Laviolette and B. Gosselin, "A low-cost, wireless, 3D printed custom armband for sEMG hand gesture recognition", *Sensors*, vol. 19, no. 12, 2019 (accepted).
- **G. Gagnon-Turcotte**, I. Keramidis, C. Ethier, Y. De Koninck and B. Gosselin, "A Wireless Electro-Optic Headstage with a 0.13- μ m CMOS Custom Integrated DWT Neural Signal Decoder for Closed-Loop Optogenetics", *IEEE Transactions on Biomedical Circuits and Systems*, 2019 (submitted, minor revision).
- **G. Gagnon-Turcotte**, M. Noormohammadi Khiarak, C. Ethier, Y. De Koninck and B. Gosselin, "A 0.13- μ m CMOS SoC for Simultaneous Multichannel Optogenetics and Neural Recording", *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, 3087-3100, 2018.
- **G. Gagnon-Turcotte**, Y. LeChasseur, C. Bories, Y. Messaddeq, Y. De Koninck and B. Gosselin, "A Wireless Headstage for Combined Optogenetics and Multichannel Electrophysiological Recording", *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 1, pp. 1-4, 2017.
- **G. Gagnon-Turcotte**, A. Avakh Kisomi, R. Ameli, C.O Dufresne-Camaro, Y. LeChasseur, J.L Neron, P. B. Bareil, P. Fortier, C. Bories, Y. De Koninck and B. Gosselin, "A Wireless Optogenetic Stimulator Headstage with Multichannel Neural Recording Capability", *Sensors*, vol. 15, no. 9, pp. 22776-22797, 2015.

– **Articles de journaux (co-auteur) :**

- C. L. Fall, **G. Gagnon-Turcotte**, J.-F. Dubé, J. S. Gagné, Y. Delisle, A. Campeau-Lecours, C. Gosselin and Benoit Gosselin, "Wireless sEMG-Based Body-Machine Interface for Assistive Technology Devices", *IEEE Journal of Biomedical and Health Informatics*, vol. 21,

no. 4, pp. 967-977, 2017.

– **Articles de conférence (auteur principal) :**

- **G. Gagnon-Turcotte**, C. Ethier, Y. De Koninck and B. Gosselin, "A 13 μ m CMOS SoC for simultaneous multichannel optogenetics and electrophysiological brain recording", *IEEE International Solid-State Circuits Conference (ISSCC'18)*, pp. 466-468, San-Francisco USA, Février 2018.
- **G. Gagnon-Turcotte**, C. L. Fall, Q. Mascret, M. Biemann, L. Bouyer and B. Gosselin, "A Multichannel Wireless sEMG Sensor Endowing a 0.13 μ m CMOS Mixed-Signal SoC", *IEEE Life Sciences Conference (LSC'18)*, pp. 1-4, Montréal Canada, Octobre 2018.
- **G. Gagnon-Turcotte**, E. Maghsoudloo, Y. Messaddeq, Y. De Koninck and B. Gosselin, "A wireless photostimulator for optogenetics with live animals", *IEEE International New Circuits and Systems Conference (NEWCAS'17)*, pp. 193-196, Strasbourg France, Juin 2017.
- **G. Gagnon-Turcotte**, L. L. Gagnon, G. Bilodeau and B. Gosselin, "Wireless brain computer interfaces enabling synchronized optogenetics and electrophysiology", *IEEE International Symposium on Circuits and Systems (ISCAS'17)*, pp. 1-4, Baltimore USA, Mai 2017.
- **G. Gagnon-Turcotte**, Y. LeChasseur, C. Bories, Y. De Koninck and B. Gosselin, "An Optimized Adaptive Spike Detector for Behavioural Experiments", *IEEE International Symposium On Circuits and Systems (ISCAS'16)*, pp. 1098-1101, Montréal Canada, Mai 2016.
- **G. Gagnon-Turcotte**, Y. LeChasseur, C. Bories, Y. De Koninck and B. Gosselin, "A Wireless Optogenetic Headstage with Multichannel Neural Signal Compression", *IEEE Conference on Biomedical Circuits and System (BioCAS'15)*, pp. 1-4, Atlanta USA, Octobre 2015.
- **G. Gagnon-Turcotte** and B. Gosselin, "Multichannel Spike Detector With an Adaptive Threshold Based On a Sigma-Delta Control Loop", *IEEE Annual Conference on Engineering in Medicine and Biology Society (EMBC'15)*, pp. 7123-7126, Milan Italy, Août 2015.
- **G. Gagnon-Turcotte**, M. Sawan and B. Gosselin, "Low-Power Adaptive Spike Detector Based on a Sigma-Delta Control Loop", *IEEE Annual Conference on Engineering in Medicine and Biology Society (EMBC'15)*, pp. 2167-2170, Milan Italy, Août 2015.
- **G. Gagnon-Turcotte**, C.-O. Dufresne Camaro, A. Avakh Kisomi, R. Ameli and B. Gosselin, "A Wireless Multichannel Optogenetic Headstage With On-The-Fly Spike Detection", *IEEE International Symposium On Circuits and Systems (ISCAS'15)*, Lisbon Portugal, pp. 1758-1761, Mai 2015.
- **G. Gagnon-Turcotte**, C.O Dufresne-Camaro and B. Gosselin, "Comparison of Low-Power Biopotential Processors for On-The-Fly Spike Detection", *IEEE International Symposium On Circuits and Systems (ISCAS'15)*, Lisbon Portugal, pp. 802-805, Mai 2015.

– **Articles de conférence (co-auteur) :**

- M. Noormohammadi Khiarak, G. Gagnon-Turcotte, E. Martianova, C. Bories S. Martel, C. D. Proulx, Y. De Koninck and B. Gosselin, "A Wireless Optoelectronic Neuroscience Platform for Chronic Fluorescence Sensing in Freely Behaving Rodents", *IEEE Enginee-*

ring in Medicine and Biology Society (EMBC'18), pp. 1608-1611, 2018.

Il est à noter que les articles publiés aux conférences BioCAS'15 et LSC'18 se sont mérités les prix du meilleur article (2^{ème} place). En plus des publications scientifiques, ce projet de recherche a fait l'objet d'exposés lors de plusieurs événements d'envergure internationaux, en voici la liste non-exhaustive :

– **Exposés et démonstration techniques :**

- **G. Gagnon-Turcotte**, C. Ethier, Y. De Koninck and B. Gosselin, "Live demonstration : A 0.13 μm CMOS SoC for Simultaneous Multichannel Optogenetics and Electrophysiological Brain Recording", *IEEE International Solid-State Circuits Conference (ISSCC'18), San-Francisco USA, Février 2018.*
- **G. Gagnon-Turcotte**, C. Ethier, Y. De Koninck and B. Gosselin, "A DWT-Based Neural Signal Decoder with Spike Detection/Compression/Clustering for Closed-Loop Optogenetics in 0.13 μm CMOS", *TEXPO, Toronto, Octobre 2018.*
- **G. Gagnon-Turcotte**, C. Ethier, Y. De Koninck and B. Gosselin, "A DWT-Based Neural Signal Decoder with Spike Detection/Compression/Clustering for Closed-Loop Optogenetics in 0.13 μm CMOS", *Journée de l'innovation ReSMiQ, Montréal, Septembre 2018.*
- **G. Gagnon-Turcotte**, C. Ethier, Y. De Koninck and B. Gosselin, "A 0.13 μm CMOS Mixed SoC for Closed-Loop Multichannel Optogenetics and Neural Recording", *TEXPO, Montréal, Octobre 2017.*
- **G. Gagnon-Turcotte**, Y. LeChasseur, C. Bories, Y. Messaddeq, Y. De Koninck and B. Gosselin, "Live demonstration : A wireless headstage enabling combined optogenetics and multichannel electrophysiological recording", *IEEE International Symposium on Circuits and Systems (ISCAS'17), Baltimore USA, 2017.*
- **G. Gagnon-Turcotte**, C. Ethier, Y. De Koninck and B. Gosselin, "A 0.13 μm CMOS Mixed SoC for Closed-Loop Multichannel Optogenetics and Neural Recording", *Journée de l'innovation ReSMiQ, Montréal, Septembre 2017.*
- **G. Gagnon-Turcotte**, Y. LeChasseur, C. Bories, Y. De Koninck and B. Gosselin, "High resolution CMOS neural interface for electrophysiology and synchronized optogenetics", *Journée de l'innovation ReSMiQ, Montréal, Septembre 2016.*

Il est à noter que les démonstrations techniques et les présentations effectuées lors de la journée de l'innovation ReSMiQ ont permis d'obtenir le prix du meilleur projet aux études supérieures trois années d'affilées (de 2016 à 2018). De plus, l'auteur de cette thèse s'est vu attribuer le prestigieux prix pancanadien *Brian L. Barge Award for Excellence in Microsystem Integration* lors de l'évènement TEXPO 2017.

– **Brevets et droits d'auteur :**

- **Inventeurs : Gabriel Gagnon-Turcotte** et Benoit Gosselin, **Titre : System and Method for Detecting Spikes in Noisy Signals**, **Brevet déposé aux USA : US 2019/0012515 A1.**

- **Inventeurs : Gabriel Gagnon-Turcotte** et Benoit Gosselin, Titre : System and method for compressing action potentials in electrophysiological signals, No. de divulgation : 01691-DIV-01, Statut : Divulgation en traitement.

5.2 Liste des distinctions

Ce projet de recherche a permis à l'auteur de cette thèse d'obtenir plusieurs distinctions et prix. En voici la liste non-exhaustive :

- Prix du meilleur article (2^{ème} place) à la *IEEE Life Sciences Conference (LSC'18), Montréal Canada, Octobre 2018.*
- Prix du meilleur projet aux études supérieures à la *journée de l'innovation ReSMiQ, Montréal, 2018.*
- Prix d'excellence en communication scientifique, *journées de la recherche en santé, Québec, 2018.*
- Sélection pour soumettre au *IEEE Journal of Solid-State Circuits Special Issue of the ISSCC conference.* Seulement les meilleur articles et présentations de la prestigieuse conférence ISSCC ont été retenus pour ce *special issue*, 2018.
- Prix pancanadien *Brian L. Barge Award for Excellence in Microsystem Integration* lors de l'évènement *TEXPO, Montréal, 2017.*
- Prix spécial du jury de la bourse d'excellence aux études supérieures, décerné par l'*Ordre des Ingénieurs du Québec.*
- Prix du meilleur projet aux études supérieures à la *journée de l'innovation ReSMiQ, Montréal, 2017.*
- Sélection pour soumettre au *IEEE Transaction on Biomedical Circuits and Systems Special Issue.* Seulement les meilleur articles et présentations de la conférence BioCAS ont été retenus pour ce *special issue*, 2016.
- Médaille d'or académique du gouverneur général du Canada, 2016.
- Prix du meilleur projet aux études supérieures à la *journée de l'innovation ReSMiQ, Montréal, 2016.*
- Prix spécial IEEE Circuit and System Society (CASS) du meilleur projet à la *journée de l'innovation ReSMiQ, Montréal, 2016.*
- Lauréat de la semaine Le Soleil/Radio-Canada, semaine du 18 janvier 2016.
- Prix Mitacs pour innovation exceptionnelle, 2016.
- Mention au tableau d'honneur de la FÉSP de l'Université Laval, 2015.
- Prix du meilleur article (2^{ème} place) à la *IEEE Conference on Biomedical Circuits and System (BioCAS'15), Atlanta USA, Octobre 2015.*

Conclusion

Le but de ce projet de recherche était de concevoir des circuits analogiques-numériques CMOS d'acquisition et de traitement des signaux neuronaux, des algorithmes d'analyse et de traitement des signaux neuronaux et des systèmes electro-optiques miniatures pour mener des expériences optogénétique en boucle fermée avec de petits animaux libres de leurs mouvements. Notamment, les nouveaux systèmes développés ont permis : 1) d'augmenter la résolution et réduire la taille, le poids et la consommation énergétique des systèmes optogénétiques sans fil, 2) de mener des expériences combinant l'enregistrement neuronal et l'optogénétique à plusieurs canaux avec des animaux libres de leurs mouvements, et 3) de mener des expériences optogénétiques synchronisées avec l'observation, c.-à-d. en boucle fermée, chez des animaux libres de leurs mouvements. Ce travail a permis la publication de dix articles de conférences internationales, dont un présenté à la prestigieuse conférence ISSCC, et quatre articles de journaux publiés ou soumis (trois abordés dans cette thèse), ainsi que l'obtention d'un brevet et de deux divulgations d'invention.

Le chapitre 2 de cette thèse a abordé la conception d'un système optogénétique miniature, et sans fil haute résolution incorporant des algorithmes innovants de compression des signaux neuronaux. Au moment de la publication de l'article présenté dans ce chapitre, il s'agissait de la première tentative visant la conception d'une plateforme optogénétique sans fil haute résolution doté d'enregistrement électrophysiologique haute résolution simultané. Pour augmenter la résolution d'enregistrement, le système implémente un algorithme de détection des PAs breveté qui est basé sur un seuillage adaptatif, ce qui permet de compresser les signaux en transmettant seulement les formes détectées. De plus, chaque PA est davantage compressé par une DWT de type Symmlet-2 suivie d'une technique de discrimination et de requantification dynamique des coefficients. Grâce à son design sur PCB «flex-rigid» et l'utilisation d'un petit FPGA à très faible consommation énergétique pour exécuter les algorithmes, le système est miniature ($17 \times 18 \times 10 \text{ mm}^3$) et ne pèse que 2.8 g. Les algorithmes ont été validés par simulation et comparés à des techniques existantes. Les résultats obtenus démontrent que les algorithmes proposés sont plus robustes que les méthodes existantes testées, tout en permettant de reconstruire les signaux compressés avec une meilleure qualité. Le système et les algorithmes ont été validés lors d'expériences *in vivo* impliquant des souris transgéniques. Lors de ces expériences, il a été démontré que 1) le système peut compresser

plusieurs canaux d'enregistrement en parallèle et en temps réel (taux de compression >500), et 2) que le système est capable de déclencher une réponse biologique grâce à l'optogénétique.

Le chapitre 3 a présenté un SoC en technologie 0.13 μm CMOS permettant de faire de l'enregistrement neuronal et de l'optogénétique multicanal. Au moment de la publication de l'article présenté dans ce chapitre, il s'agit de la première puce CMOS à être dotée de ces deux fonctionnalités. Le SoC possède 10 canaux d'enregistrement et 4 canaux de stimulation optogénétique. La conception du bioamplificateur inclut une bande passante programmable et un faible bruit referré à l'entrée, ce qui permet de cibler différents types de signaux biologiques. L'ADC de type Delta-Sigma MASH 1-1-1 est conçu pour fonctionner avec de faibles taux de sur-échantillonnage pour réduire sa consommation énergétique et possède une résolution programmable. Cet ADC exploite une nouvelle technique réduisant la taille du circuit en soustrayant la sortie de chaque branche du $\Delta\Sigma$ dans le domaine numérique, comparativement à la méthode analogique classique. Le SoC implémente un nouveau circuit de stimulation optique basé sur une source de courant de type cascode avec rétroaction, ce qui permet d'accommoder une large gamme de LED et de tensions de batterie. Le SoC est intégré dans un système optogénétique sans fil et validé *in vivo*. Lors de ces expériences, il a été démontré que 1) le SoC peut enregistrer l'activité neuronale (PA et LFP) avec une qualité équivalente à un système câblé, et 2) que le SoC est capable de déclencher une réponse biologique (PA et LFP) grâce à l'optogénétique.

Le chapitre 4 a présenté le développement du premier système optogénétique multicanal qui est capable de fonctionner en boucle fermée et le premier à être validé lors d'expériences *in vivo* impliquant des animaux libres de leurs mouvements. Pour ce faire, la conception d'un SoC en technologie 0.13 μm CMOS numérique, c.-à-d. le ND-IC, est abordée. Le ND-IC possède 3 modules : 1) le détecteur de PA adaptatif, 2) le module de compression possédant un nouvel arbre de tri pour discriminer les coefficients, et 3) le module de classement automatique des PA qui réutilise les données générées par le module de détection et de compression pour réduire sa complexité. Un lien entre un canal d'enregistrement et un canal de stimulation est établi selon l'association de chaque PA à un neurone, grâce à la classification, et selon l'activité de ce neurone dans le temps. Le ND-IC est intégré dans un système optogénétique sans fil et miniature occupant un volume de seulement 1.12 cm^3 . Le système et le ND-IC ont été validés *in vivo* dans le cerveau de souris exprimant la protéine ChR2-mCherry et pouvant se déplacer librement. Lors de ces expériences, il a été démontré que 1) le ND-IC peut détecter, compresser et classer les PA avec une précision supérieure ou équivalente à des approches plus exigeantes en calculs, 2) que le système peut fonctionner en boucle fermée sans supervision, et 3) que le système est capable de modifier en temps réel le comportement des communications inter-neuronales, et ce, automatiquement.

Travaux futurs

Suite aux bons fonctionnements des différents systèmes, algorithmes et puces générés par ce travail, des expériences chroniques impliquant des souris avec des pathologies neurologiques (ex. : Parkinson, troubles de dépendances) sont prévues pour l'été 2019 au centre de recherche CERVO de Québec. Le système utilisé sera celui capable de fonctionner en boucle fermée. Entre autres, il est prévu d'utiliser des souris aux prises avec des problèmes de dépendances à l'alcool (dépendance forcée). Le système devra détecter l'envie d'alcool et stimuler rétroactivement une région du cerveau relié au relâchement de dopamine. L'objectif est maintenant de faire des percées en recherche sur le cerveau grâce aux moyens qui ont été présentés dans cette thèse. Des publications dans des revues axées sur la biologie et les neurosciences sont envisagées.

Afin d'améliorer le fonctionnement du système capable de fonctionner en boucle fermée, une avenue possible serait d'incorporer une technique de régulation du *firing rate* des neurones. Par exemple, un régulateur de type proportionnel intégral pourrait être ajouté afin de contrôler indépendamment le *firing rate* de plusieurs neurones en parallèle. Une autre amélioration consisterait à tout mettre sur une seule puce CMOS (FPGA, transmetteur, etc.), et ce, afin de rendre le système moins invasif et augmenter l'autonomie.

Bibliographie

- [1] G. Buzsáki *et al.*, “Tools for probing local circuits : high-density silicon probes combined with optogenetics,” *Neuron*, vol. 86, no. 1, pp. 92–105, 2015.
- [2] “White house brain initiative,” <https://www.whitehouse.gov/share/brain-initiative>, accessed : 2015-07-02.
- [3] “Human brain project and allen institute,” <https://www.humanbrainproject.eu/>, accessed : 2015-07-02.
- [4] “Fond canadien de recherche sur le cerveau,” <https://braincanada.ca/fr/le-fonds-canadien-de-recherche-sur-le-cerveau/>, accessed : 2019-01-15.
- [5] “Algorithmes et processeurs temps réel de traitement de signaux neuronaux pour une plateforme optogénétique sans fil,” <https://corpus.ulaval.ca/jspui/handle/20.500.11794/26339>, accessed : 2019-01-15.
- [6] G. M. Shepherd, *Neurobiology*, 3rd ed., O. U. Press, Ed. Oxford University Press, 1994.
- [7] O. Sporns, *Networks of the Brain*, M. Press, Ed. The MIT Press, 2010.
- [8] B. J. Sadock and V. A. Sadock, *Kaplan and Sadock’s Synopsis of Psychiatry : Behavioral Sciences/Clinical Psychiatry*, 10th ed., L. W. . Wilkins, Ed. LWW, 2007.
- [9] G. Gagnon-Turcotte *et al.*, “A 0.13- μm cmos soc for simultaneous multichannel optogenetics and neural recording,” *J. Solid-State Circuits*, vol. 53, no. 11, pp. 3087–3100, 2018.
- [10] —, “A 13- μm cmos soc for simultaneous multichannel optogenetics and electrophysiological brain recording,” *IEEE International Solid-State Circuits Conference (ISSCC’18)*, pp. 466–468, 2018.
- [11] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth, *Principles of Neural Science*, 5th ed. McGraw-Hill Education, 2012.
- [12] J. G. Webster *et al.*, *Medical Instrumentation*, 4th ed. Wiley, 2010.

- [13] B. W. C. M. F. Bear and M. A. Paradiso, *Neurosciences : à la découverte du cerveau*, 4th ed. Pradel, 2016.
- [14] L. Liu, L. Yao, X. Zou, W. L. Goh, and M. Je, "Neural recording front-end ic using action potential detection and analog buffer with digital delay for data compression," *IEEE Engineering in Medicine and Biology Society EMBC'13*, pp. 747–750, 2013.
- [15] B. Ulrich and O. Maurits, "A neural recorder with mixed signal feedback loop for a spectral separation of spikes and lfp," *Conference on Ph.D. Research in Microelectronics and Electronics (PRIME'12)*, pp. 1–4, 2012.
- [16] B. Gosselin *et al.*, "A low-power integrated bioamplifier with active low-frequency suppression," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 1, no. 3, pp. 184–192, 2007.
- [17] —, "A mixed-signal multichip neural recording interface with bandwidth reduction," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 3, no. 3, pp. 129–141, 2009.
- [18] R. Ameli *et al.*, "A wireless and batteryless neural headstage with optical stimulation and electrophysiological recording," *IEEE Eng. Med. Biol. Soc. (EMBC'13)*, pp. 5662–5665, 2013.
- [19] B. Gosselin, "Recent advances in neural recording microsystems," *Sensors*, vol. 11, no. 5, pp. 4572–4597, 2011.
- [20] G. Gagnon-Turcotte *et al.*, "A wireless headstage for combined optogenetics and multichannel electrophysiological recording," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 1, pp. 1–14, 2017.
- [21] "Microprobes microélectrodes métalliques," <http://www.microprobes.com/index.php/products/metal-microelectrodes>, accessed : 2015-02-17.
- [22] "Neuronexus microélectrodes flexibles de surface," <http://neuronexus.com/products/neural-probes/surface-probes>, accessed : 2015-02-17.
- [23] Y. LeChasseur, S. Dufour, G. Lavertu, C. Bories, M. Deschênes, R. Vallée, and Y. D. Koninck, "A microprobe for parallel optical and electrical recordings from single neurons in vivo," *Nature Methods*, vol. 8, pp. 319–325, 2011.
- [24] "Microprobes tableau de microélectrodes linéaire sur silicium," <http://www.microprobes.com/index.php/products/multi-channel-arrays/lma-linear-microelectrode-arrays-mc>, accessed : 2015-02-17.

- [25] "Microprobes tableau de microélectrodes métalliques," <http://www.microprobes.com/index.php/products/multi-channel-arrays/mea-micro-electrode-arrays>, accessed : 2015-02-17.
- [26] "Blackrock microsystems tableau de microélectrodes sur silicium," <http://blackrockmicro.com/neuroscience-research-products/low-noise-ephys-electrodes/blackrock-utah-array/>, accessed : 2016-06-22.
- [27] A. E. Mendrela *et al.*, "A high-resolution opto-electrophysiology system with a miniature integrated headstage," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 5, pp. 1065–1075, 2018.
- [28] "Tbsi m series system," http://www.trianglebiosystems.com/assets/tbsi_m32_manual.pdf, accessed : 2019-01-15.
- [29] K. D. Harris *et al.*, "Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements," *Journal of Neurophysiology*, vol. 84, no. 1, pp. 401–414, 2000.
- [30] G. Buzsaki, "Large-scale recording of neuronal ensembles," *Nature Neuroscience*, vol. 7, no. 5, pp. 446–451, 2004.
- [31] K. Deisseroth, "Optogenetics," *Nat. Methods*, vol. 8, no. 1, pp. 26–29, 2011.
- [32] J. A. C. *et al.*, "Targeted optogenetic stimulation and recording of neurons in vivo using cell-type-specific expression of channelrhodopsin-2," *Nature protocols*, vol. 5, p. 247–254, 2010.
- [33] J. M. *et al.*, "Photoelectric artefact from optogenetics and imaging on microelectrodes and bioelectronics : New challenges and opportunities," *Journal of Materials Chemistry B*, vol. 3, no. 23, p. 4965–4978, 2015.
- [34] J. Wells *et al.*, "Optical stimulation of neural tissue in vivo," *Optics Letters*, vol. 30, no. 5, pp. 504–506, 2005.
- [35] O. Yizhar *et al.*, "Optogenetics in neural systems," *Neurons*, vol. 71, no. 1, pp. 9–34, 2011.
- [36] "The jackson laboratory," <http://www.jax.org/>, accessed : 2015-08-17.
- [37] S. Dufour and Y. D. Koninck, "Optrodes for combined optogenetics and electrophysiology in live animals," *Neurophotronics*, vol. 2, no. 3, 2015.
- [38] "Institut universitaire en santé mentale de québec, axe de recherche en neurosciences cellulaires et moléculaires," <http://www.crulrg.ulaval.ca/fr/recherche/axes/chercheurs/chercheur.php?id=61g/>, accessed : 2015-08-17.

- [39] C. Ethier *et al.*, “Restoration of grasp following paralysis through brain-controlled stimulation of muscles,” *Nature*, vol. 485, pp. 368–371, 2012.
- [40] N. Josset *et al.*, “Distinct contributions of mesencephalic locomotor region nuclei to locomotor control in the freely behaving mouse,” *Curr. Biol.*, vol. 28, no. 6, pp. 884–901, 2018.
- [41] D. Y. K. Takashi and A. L. Vazquez, “Photoelectric artefact from optogenetics and imaging on microelectrodes and bioelectronics : New challenges and opportunities,” *J. Mater Chem. B Mater Biol. Med.*, vol. 3, no. 25, pp. 4965–4978, 2016.
- [42] J. A. C. and others, “Targeted optogenetic stimulation and recording of neurons in vivo using cell-type-specific expression of channelrhodopsin-2,” *Nature protocols*, vol. 5, pp. 247–254, 2010.
- [43] H. Lyu, X. Liu, N. Rogers, V. Gilja, and D. Kuzum, “Graphene neural interfaces for artifact free optogenetics,” *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4204–4207, 2016.
- [44] S. C. P. Williams and K. Deisseroth, “Optogenetics,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 41, pp. 16 287–16 287, 2013.
- [45] M. Sivaprakasam *et al.*, “A variable range bi-phasic current stimulus driver circuitry for an implantable retinal prosthetic device,” *IEEE Journal of Solid-State Circuit*, vol. 40, no. 3, pp. 763–771, 2005.
- [46] B. Rosin *et al.*, “Closed-loop deep brain stimulation is superior in ameliorating parkinsonism,” *Neuron*, vol. 72, no. 2, pp. 370–384, 2011.
- [47] D. A. Wagenaar *et al.*, “Controlling bursting in cortical cultures with closed-loop multi-electrode stimulation,” *J. Neurosci.*, vol. 25, no. 3, pp. 680–688, 2005.
- [48] L. Grosenick, J. H. Marshel, and K. Deisseroth, “Closed-loop and activity-guided optogenetic control,” *Neuron*, vol. 86, no. 1, pp. 106–139, 2015.
- [49] J. P. Newman *et al.*, “Optogenetic feedback control of neural activity,” *eLife*, 2015.
- [50] T. K. Nguyen *et al.*, “Closed-loop optical neural stimulation based on a 32-channel low-noise recording system with online spike sorting,” *J. Neural Eng.*, vol. 11, no. 4, 2014.
- [51] M. Capogrosso *et al.*, “Configuration of electrical spinal cord stimulation through real-time processing of gait kinematics,” *Nature Protocols*, vol. 13, no. 9, pp. 2031–2061, 2018.
- [52] C. Ehtier and L. E. Miller, “Brain-controlled muscle stimulation for the restoration of motor function,” *Neurobiology of Disease*, vol. 83, pp. 180–190, 2015.

- [53] C. Gold, D. A. Henze, K. Christof, and G. Buzáki, "On the origin of the extracellular action potential waveform : A modeling study," *J. of Neurophysiol.*, vol. 95, no. 5, pp. 3113–3128, 2006.
- [54] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis : state-of-the-art and future challenges," *Nat. Neurosci.*, vol. 7, no. 5, pp. 456–461, 2004.
- [55] J. O. Kennedy, "Neural coding," *Neuron*, vol. 26, no. 3, pp. 563–566, 2000.
- [56] E. M. Izhikevich *et al.*, "Bursts as a unit of neural information : selective communication via resonance," *Trends in Neurosciences*, vol. 26, no. 3, pp. 161–167, 2003.
- [57] A. Meysam *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, 2011.
- [58] X. Liu *et al.*, "A compact closed-loop optogenetics system based on artifact-free transparent graphene electrodes," *Frontiers in Neuroscience*, vol. 12, 2018.
- [59] L. Xilin *et al.*, "Design of a closed-loop, bidirectional brain machine interface system with energy efficient neural feature extraction and pid control," *TBioCAS*, vol. 11, no. 4, pp. 729–742, 2017.
- [60] 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, 2014.
- [61] 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.
- [62] G. Gagnon-Turcotte, A. A. Kisomi, R. Ameli, C. Dufresne-Camaro, Y. LeChasseur, J. Neron, P. B. Bareil, P. Fortier, C. Bories, Y. D. Koninck, and B. Gosselin, "A wireless optogenetic stimulator headstage with multichannel neural recording capability," *Sensors*, vol. 15, no. 9, pp. 22 776–22 797, 2015.
- [63] "Firefly e-phys+ de krs," <http://www.kendallresearchsys.com/>, accessed : 2016-11-14.
- [64] "Triangle biosystem international m-series, systèmes d'enregistrement multicanal filaire," <http://www.trianglebiosystems.com/m-series-systems.html>, accessed : 2015-07-28.
- [65] "Doric lenses fiwi," <http://www.doriclenses.com/>, accessed : 2016-11-14.
- [66] R. Ameli, A. Mirbozorgi, J.-L. Néron, Y. LeChasseur, and B. Gosselin, "A wireless and batteryless neural headstage with optical stimulation and electrophysiological recording," *IEEE Engineering in Medicine and Biology Society EMBC'13*, pp. 5662–5665, 2013.

- [67] S. K. Arfin, M. A. Long, M. S. Free, and R. Sarpeshkar, "Wireless neural stimulation in freely behaving small animals," *J. Neurophysiol.*, vol. 102, no. 1, pp. 598–6056, 2009.
- [68] T. il Kim, J. G. McCall, Y. H. Jung, X. Huang, E. R. Siuda, Y. Li, J. Song, Y. ming Song, H. A. Pao, R.-H. Kim, C. Lu, S. D. Lee, I.-S. Song, G. Shin, R. alHasani, S. Kim, M. P. Tan, Y. Huang, F. G. Omenetto, J. A. Rogers, and M. R. Bruchas, "Injectable, cellular-scale optoelectronics with applications for wireless optogenetics," *Science*, vol. 340, no. 6129, pp. 211–216, 2013.
- [69] B. Gosselin and M. Sawan, "An ultra low-power cmos automatic action potential detector," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 4, pp. 346–35, 2009.
- [70] H. Gao, R. M. Walker, P. Nuyujukian, K. A. A. Makinwa, K. V. Shenoy, B. Murmann, and T. H. Meng, "Hermese : A 96-channel full data rate direct neural interface in 0.13 μm cmos," *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 1043–1055, 2012.
- [71] "Blackrock microsystems," <http://www.blackrockmicro.com/>, accessed : 2019-01-08.
- [72] "Triangle biosystems," <http://www.trianglebiosystems.com/>, accessed : 2019-01-08.
- [73] R. Sabyasachi and X. Wang, "Wireless multi-channel single unit recording in freely moving and vocalizing primates," *Journal of Neuroscience Methods*, vol. 203, pp. 28–40, 2012.
- [74] T. A. Szuts, V. Fadeyev, S. Kachiguine, M. V. G. A. Sher, M. A. ao, P. Hottowy, W. Dabrowski, E. V. Lubenov, N. U. A. G. Siapas, A. M. Litke, and M. Meister, "A wireless multi-channel neural amplifier for freely moving animals," *Nat. Neurosci.*, vol. 14, no. 2, 2011.
- [75] J. N. Y. Aziz, K. Abdelhalim, R. Shulyzki, R. Genov, B. L. Bardakjian, M. Derchansky, D. Serletis, and P. L. Carlen, "256-channel neural recording and delta compression microsystem with 3d electrodes," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 995–1005, 2009.
- [76] M. S. Chae, Z. Y. Yang, M. R. Yuce, L. Hoang, and W. Liu, "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. 312–21, 2009.
- [77] M. Yin, D. A. Borton, J. Aceros, W. R. Patterson, and A. V. Nurmikko, "A 100-channel hermetically sealed implantable device for chronic wireless neurosensing applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 2, pp. 115–128, 2013.
- [78] M. Rizk, I. Obeid, S. H. Callender, and P. D. Wolfso, "A single-chip signal processing and telemetry engine for an implantable 96-channel neural data acquisition system," *J. Neural Eng.*, vol. 4, pp. 309–321, 2007.

- [79] S. Farshchi, A. Pesterev, P. Nuyujukian, E. Guenterberg, I. Mody, and J. W. Judy, "Embedded neural recording with tinyos-based wireless-enabled processor modules," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 134–141, 2010.
- [80] J. A. Gregory, A. Borna, S. Roy, X. Wang, B. Lewandowski, M. Schmidt, and K. Najafi, "Low-cost wireless neural recording system and software," *IEEE Engineering in Medicine and Biology Society EMBC'09*, pp. 3833–3836, 2009.
- [81] G. Gagnon-Turcotte, A. A. Kisomi, R. Ameli, C.-O. D. Camaro, Y. LeChasseur, J.-L. Néron, P. B. Bareil, P. Fortier, C. Bories, Y. D. Koninck, and B. Gosselin, "A wireless optogenetic headstage with multichannel electrophysiological recording capability," *Sensors*, vol. 15, no. 9, pp. 22 776–22 797, 2015.
- [82] A. M. Kamboh, M. Raetz, K. G. Oweiss, and A. Mason, "Area-power efficient vlsi implementation of multichannel dwt for data compression in implantable neuroprosthetics," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 2, pp. 128–135, 2007.
- [83] V. Karkare, S. Gibson, and D. Markovic, "A 130- μ w, 64-channel neural spike-sorting dsp chip," *IEEE J. Solid-State Circuits*, vol. 46, no. 5, pp. 1214–1222, 2011.
- [84] G. Gagnon-Turcotte, Y. LeChasseur, C. Bories, Y. D. Koninck, and B. Gosselin, "A wireless optogenetic headstage with multichannel neural signal compression," in *2015 IEEE Biomedical Circuits and Systems Conference (IEEE BioCAS'15)*, pp. 1–4, 2015.
- [85] G. Gagnon-Turcotte, C.-O. D. Camaro, and B. Gosseli, "Comparison of low-power bio-potential processors for on-the-fly spike detection," *IEEE International Symposium on Circuits and Systems (ISCAS'15)*, pp. 802–805, 2015.
- [86] I. Obeid and P. D. Wolf, "Evaluation of spike-detection algorithms for a brain-machine interface application," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 905–911, 2004.
- [87] Z. Yang, Q. Zhao, E. Keefer, and W. Liu, "Noise characterization, modeling, and reduction for in vivo neural recordings," *NIPS'10*, pp. 2160–2168, 2010.
- [88] R. R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," *IEEE Engineering in Medicine and Biology Society EMBC'03*, vol. 4, pp. 3325–3328, 2003.
- [89] K. E. Thomson, T. Shlien, Y. Suhail, A. Mason, and K. G. Oweiss, "B-spline vs. lifting architecture for neuroprosthetic devices," *IEEE EIT'06*, pp. 641–644, 2006.
- [90] Y. Chompusri and S. Yimman, "Energy packing efficiency based threshold level selection for dtw ecg compression," *International Journal of Applied Biomedical Engineering*, vol. 2, no. 2, pp. 19–28, 2009.

- [91] V. Shalchyan, W. Jensen, and D. Farina, "Spike detection and clustering with unsupervised wavelet optimization in extracellular neural recordings," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 9, pp. 2576–9, 2012.
- [92] J. Zhang, Y. Suo, S. Mitra, S. P. Chin, T. D. Tran, F. Yazicioglu, and R. Etienne-Cummings, "Reconstruction of neural action potentials using signal dependent sparse representations," *IEEE International Symposium on Circuits and Systems (ISCAS'13)*, pp. 1520–1523, 2013.
- [93] F. Chen, A. P. Chandrakasan, and V. Stojanović, "A signal-agnostic compressed sensing acquisition system for wireless and implantable sensor," *IEEE Conference on Custom Integrated Circuits (CICC'10)*, pp. 1–4, 2010.
- [94] C. Bulach, U. Bihl, and M. Ortmanns, "Evaluation study of compressed sensing for neural spike recordings," *IEEE Engineering in Medicine and Biology Society EMBC'12*, pp. 3507–3510, 2012.
- [95] A. H. Gittis, S. H. Moghadam, and S. du Lac, "Mechanisms of sustained high firing rates in two classes of vestibular nucleus neurons : Differential contributions of resurgent na, kv3, and bk currents," *J. Neurophysiol.*, vol. 104, no. 3, pp. 1625–1634, 2010.
- [96] B. R. Arenkiel, J. Peca, I. G. Davison, C. Feliciano, K. Deisseroth, G. J. Augustine, M. D. Ehlers, and G. Feng, "In vivo light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2," *Neuron*, vol. 54, no. 2, pp. 205–18, 2007.
- [97] C. T. Wentz, J. G. Bernstein, P. Monahan, A. Guerra, A. Rodriguez, and E. S. Boyden, "A wirelessly powered and controlled device for optical neural control of freely-behaving animals," *J. Neural Eng.*, vol. 8, no. 4, 2011.
- [98] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, and K. Deisseroth, "Millisecond-timescale, genetically targeted optical control of neural activity," *Nat. Neurosci.*, vol. 8, no. 9, pp. 1263–8, 2005.
- [99] A. M. Herman, L. Huang, D. K. Murphey, I. Garcia, and B. R. Arenkiel, "Cell type-specific and time-dependent light exposure contribute to silencing in neurons expressing channelrhodopsin-2," *eLIFE*, 2014.
- [100] S. A. Mirbozorgi, H. Bahrami, M. Sawan, and B. Gosselin, "A smart cage with uniform wireless power distribution in 3d for enabling long-term experiments with freely moving animals," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 2, pp. 424–434, 2015.
- [101] P. Gutruf and J. A. Rogers, "Implantable, wireless device platforms for neuroscience research," *Curr. Opin. Neurobiol.*, vol. 50, pp. 42–49, 2018.

- [102] K. L. Montgomery *et al.*, “Wirelessly powered, fully internal optogenetics for brain, spinal and peripheral circuits in mice,” *Nat. Methods*, vol. 12, no. 8, pp. 969–974, 2015.
- [103] “Doric lenses. fi-wi wireless headstage,” <http://doriclenses.com/life-sciences/fiberless-and-wireless/921-fi-wi-opto-electric-cannula.html>, accessed : 2019-01-08.
- [104] “Triangle biosystems wireless headstage,” <https://www.trianglebiosystems.com/iw-series-systems.html>, accessed : 2019-01-08.
- [105] J. Yaoyao *et al.*, “A mm-sized free-floating wirelessly powered implantable optical stimulating system-on-a-chip,” *IEEE Int. Solid-State Circuits Conf. (ISSCC’18)*, pp. 360–374, 2018.
- [106] B.-O. Watson, D. mingxin, and G. Buzsáki, “Temporal coupling of field potentials and action potentials in the neocortex,” *Eur. J. Neurosci.*, 2018.
- [107] K. Y. Kwon *et al.*, “Design, fabrication, and packaging of an integrated, wirelessly-powered optrode array for optogenetics application,” *Front. Syst. Neurosci.*, 2015.
- [108] K. A. Ng and Y. P. Xu, “A multi-channel neural-recording amplifier system with 90 db cmrr employing cmos-inverter-based otas with cmfb through supply rails in 65nm cmos,” *IEEE Int. Solid-State Circuits Conf. (ISSCC’15)*, vol. 58, pp. 206–207, 2015.
- [109] 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, 2007.
- [110] V. Majidzadeh *et al.*, “Energy efficient low-noise neural recording amplifier with enhanced noise efficiency factor,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 3, pp. 262–271, 2011.
- [111] W. Wattanapanitch *et al.*, “An energy-efficient micropower neural recording amplifier,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 2, pp. 136–147, 2007.
- [112] F. Zhang *et al.*, “Design of ultra-low power biopotential amplifiers for biosignal acquisition applications,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 4, pp. 344–355, 2012.
- [113] R. R. Harrison and C. Charles, “A low-power low-noise cmos amplifier for neural recording applications,” *IEEE J. Solid-State Circuits*, vol. 38, no. 6, pp. 958–965, 2003.
- [114] B. Gosselin *et al.*, “A low-power integrated bioamplifier with active low-frequency suppression,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 3, pp. 184–192, 2007.
- [115] M. Yin and M. Ghovanloo, “A low-noise preamplifier with adjustable gain and bandwidth for biopotential recording applications,” *IEEE Int. Symp. Circuits Syst. (IS-CAS’07)*, pp. 321–324, 2007.

- [116] 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, 2013.
- [117] T. Yang and J. Holleman, "An ultralow-power low-noise cmos biopotential amplifier for neural recording," *IEEE Trans. Circuits Syst. II : Express Briefs*, vol. 62, no. 10, pp. 927–931, 2015.
- [118] M. Rezaie *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, 2018.
- [119] K. Paralikar *et al.*, "An implantable 5mw/channel dual-wavelength optogenetic stimulator for therapeutic neuromodulation research," *IEEE Int. Solid-State Circuits Conf. (ISSCC'10)*, pp. 238–240, 2010.
- [120] H.-M. Lee *et al.*, "A power-efficient switched-capacitor stimulating system for electrical/optical deep brain stimulation," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 360–374, 2015.
- [121] C.-H. Chen *et al.*, "An integrated circuit for simultaneous extracellular electrophysiology recording and optogenetic neural manipulation," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 3, pp. 557–568, 2017.
- [122] H. Zhao *et al.*, "A cmos-based neural implantable optrode for optogenetic stimulation and electrical recording," *IEEE Biomed. Circuits Syst. Conf. (BioCAS'15)*, pp. 1–4, 2015.
- [123] L. Wang *et al.*, "An 18 μ w 79db-dr 20khz-bw mash delta-sigma modulator utilizing self-biased amplifiers for biomedical applications," *IEEE Custom Integrated Circuits Conf. (CICC'11)*, pp. 1–4, 2011.
- [124] V. S. Polikov *et al.*, "Response of brain tissue to chronically implanted neural electrodes," *J. Neurosci. Methods*, vol. 148, no. 1, pp. 1–18, 2005.
- [125] M. X. Song *et al.*, "A sigma-delta modulator system design with 1-1-1 mash structure," *Applied Mechanics and Materials*, vol. 389, pp. 568–572, 2013.
- [126] Y.-S. Kwak *et al.*, "A 72.9-db sndr 20-mhz bw 2-2 discrete-time sturdy mash delta-sigma modulator using source-follower-based integrators," *Asian Solid-State Circuits Conf. (A-SSCC'17)*, pp. 289–292, 2017.
- [127] H.-M. Lee *et al.*, "A wireless implantable switched-capacitor based optogenetic stimulating system," *IEEE Eng. Med. Biol. Soc. (EMBC'14)*, pp. 878–881, 2014.
- [128] H. Zhao *et al.*, "An implantable optrode with self-diagnostic function in 0.35 μ m cmos for optical neural stimulation," *IEEE Biomed. Circuits Syst. Conf. (BioCAS'14)*, pp. 244–247, 2014.

- [129] G. Gagnon-Turcotte *et al.*, "A wireless photostimulator for optogenetics with live animals," *IEEE International New Circuits and Systems (NEWCAS'17)*, pp. 193–196, 2017.
- [130] H. Park *et al.*, "A wireless magnetoresistive sensing system for an intraoral tongue-computer interface," *IEEE Trans. Biomed. Circuits Syst.*, vol. 6, no. 6, pp. 571–585, 2012.
- [131] W. Biederman *et al.*, "A 4.78 mm² fully-integrated neuromodulation soc combining 64 acquisition channels with digital compression and simultaneous dual stimulation," *IEEE J. Solid-State Circuits*, vol. 50, no. 4, pp. 1038–1047, 2015.
- [132] T. Chen, "A biomedical multiprocessor soc for closed-loop neuroprosthetic applications," *IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 434–435, 2009.
- [133] S. H. Bitzenhofer *et al.*, "Layer-specific optogenetic activation of pyramidal neurons causes beta-gamma entrainment of neonatal networks," *Nat. Commun.*, 2017.
- [134] E. Nicholson *et al.*, "Analogue closed-loop optogenetic modulation of hippocampal pyramidal cells dissociates gamma frequency and amplitude," *eLife*, 2018.
- [135] L. Xilin *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, 2017.
- [136] X. Liu *et al.*, "A compact closed-loop optogenetics system based on artifact-free transparent graphene electrode," *Front. Neurosci.*, vol. 12, 2018.
- [137] G. Gagnon-Turcotte *et al.*, "A 0.13- μ m cmos soc for simultaneous multichannel optogenetics and electrophysiological brain recording," *IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 466–468, 2018.
- [138] G. Gagnon-Turcotte, C.-O. D. Camaro, , and B. Gosselin, "Comparison of low-power biopotential processors for on-the-fly spike detection," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 802–805, 2015.
- [139] "Wave clus." <http://www2.le.ac.uk/departments/engineering/research/bioengineering/neuroengineering-lab/spike-sorting>, accessed : 2019-01-08.
- [140] S. Senova *et al.*, "Experimental assessment of the safety and potential efficacy of high irradiance photostimulation of brain tissues," *Scientific Reports*, vol. 7, 2017.
- [141] Y. Yang *et al.*, "Hardware efficient automatic thresholding for neo-based neural spike detection," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 4, pp. 826–833, 2017.

- [142] C. P. Breanne *et al.*, "Comparison of spike sorting and thresholding of voltage waveforms for intracortical brain-machine interface performance," *J. Neural Eng.*, vol. 12, no. 1, pp. 1741–2560, 2015.
- [143] S. Gibson *et al.*, "Comparison of spike-sorting algorithms for future hardware implementation," *IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5015–5020, 2008.
- [144] T.-C. Chien *et al.*, "28-channel spike sorting processor with a parallel-folding structure in 90nm process," *IEEE International Symposium on Circuits and Systems (ISCAS'09)*, pp. 1253–1256, 2009.
- [145] "Biopotential waveform bank," <http://www2.le.ac.uk/departments/engineering/research/bioengineering/neuroengineering-lab/spike-sorting>, accessed : 2019-01-08.
- [146] V. Lee and J. Maguire, "The impact of tonic gabaa receptor-mediated inhibition on neuronal excitability varies across brain region and cell type," *Front. Neural Circuits*, 2014.
- [147] I. Pavlov *et al.*, "Tonic gabaa conductance bidirectionally controls interneuron firing pattern and synchronization in the ca3 hippocampal network," *Proc. Natl. Acad. Sci.*, vol. 111, no. 1, pp. 504–509, 2014.
- [148] R. R. Harrison *et al.*, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, 2007.
- [149] M. S. Chae *et al.*, "A 128-channel 6 mw wireless neural recording ic with spike feature extraction and uwb transmitter," *IEEE Transactions on Systems and Rehabilitation Engineering*, vol. 17, no. 4, pp. 312–21, 2009.
- [150] R. H. Olsson and K. D. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2796–2804, 2005.
- [151] A. M. Sodagar *et al.*, "An implantable 64-channel wireless microsystem for single-unit neural recording," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2591–2604, 2009.
- [152] J. Mavoori *et al.*, "An autonomous implantable computer for neural recording and stimulation in unrestrained primates," *Journal of Neuroscience Methods*, vol. 148, no. 1, pp. 71–77, 2005.
- [153] R. E. Hampson, V. Collins, and S. A. Deadwyler, "A wireless recording system that utilizes bluetooth technology to transmit neural activity in freely moving animals," *Journal of Neuroscience Methods*, vol. 182, no. 2, pp. 195–204, 2009.

- [154] B. Gosselin and M. Sawan, "Adaptive detection of action potentials using ultra low-power cmos circuits," *IEEE Biomedical Circuits and Systems Conference (BIOCAS'08)*, pp. 209–212, 2008.
- [155] V. Shalchyan, W. Jensen, and D. Farina, "Spike detection and clustering with unsupervised wavelet optimisation in extracellular neural recordings," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 91, pp. 2576–2585, 2012.
- [156] H. L. Chan *et al.*, "Detection of neuronal spikes using an adaptive threshold based on the maxmin spread sorting method," *Journal of Neuroscience Methods*, vol. 172, pp. 112–121, 2008.
- [157] R. R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," *IEEE Annual International of Engineering in Medicine and Biology Society (EMBC'03)*, vol. 4, pp. 3325–3328, 2003.
- [158] C. P. Breanne *et al.*, "Comparison of spike sorting and thresholding of voltage waveforms for intracortical brain-machine interface performance," *Journal of Neural Engineering*, vol. 12, no. 1, pp. 1741–2560, 2015.
- [159] Z. Yang *et al.*, "Noise characterization, modeling, and reduction for in vivo neural recordings," *Advances in Neural Information Processing Systems*, pp. 2160–2168, 2010.
- [160] Y. Yang *et al.*, "Adaptive threshold neural spike detector using stationary wavelet transform in cmos," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, p. 1, 2015 (pending publication).
- [161] K. Qureshi and V. P. Patel, "Efficient data compression of ecg signal using discrete wavelet transform," *International Journal of Research in Engineering and Technology (IJRET)*, vol. 02, no. 4, pp. 2319–1163, 2013.
- [162] M. Abo-Zahhad, *ECG Signal Compression Using Discrete Wavelet Transform*. InTech, 2011.
- [163] M. Abo-Zahhad, S. M. Ahmed, and A. Zakaria, "Ecg signal compression technique based on discrete wavelet transform ans qrs estimation," *Signal Processing : An International Journal (SPIJ)*, vol. 4, no. 2, pp. 138–160, 2010.
- [164] Y. Chompusri and S. Yimman, "Energy packing efficiency based threshold level selection for dwt ecg compression," *International Journal of Applied Biomedical Engineering*, vol. 2, no. 2, 2009.
- [165] "Leicester university banque de signaux neuronaux," <http://www2.le.ac.uk/departments/engineering/research/bioengineering/neuroengineering-lab/spike-sorting>, accessed : 2015-02-17.

- [166] "Mean shift algorithm," <http://www.mathworks.com/matlabcentral/fileexchange/10161-mean-shift-clustering>, accessed : 2015-02-27.
- [167] C. Bulach, U. Bihl, and M. Ortmanns, "Evaluation study of compressed sensing for neural spike recording," *IEEE International Conference of Engineering in Medicine and Biology Society (EMBC'12)*, pp. 3507–3510, 2012.
- [168] G. Gagnon-Turcotte, Y. LeChasseur, C. Bories, Y. D. Koninck, and B. Gosselin, "A wireless optogenetic headstage with multichannel neural signal compression," *IEEE Conference on Biomedical Circuits and System (BioCAS'15)*, pp. 1–4, 2015.
- [169] G. Gagnon-Turcotte and B. Gosselin, "System and method for compressing action potentials in electrophysiological signals," *Divulgation d'invention en traitement*, vol. 01736-DIV-2015, 2015.
- [170] "Gagnon-turcotte, g. (2015). algorithmes et processeurs temps réel de traitement de signaux neuronaux pour une plateforme optogénétique sans fil (mémoire de maîtrise). université laval." <http://theses.ulaval.ca/archimede/meta/32117>, accessed : 2016-08-16.
- [171] R. Q. Quiroga, Z. Nadasdy, and Y. Ben-Shaul, "Unsupervised spike detection and sorting wavelets and superparamagnetic clustering," *Neural Computation*, vol. 16, no. 8, pp. 1661–1687, 2004.
- [172] T.-C. Chien *et al.*, "128-channel spike sorting processor with a parallel-folding structure in 90nm process," *IEEE International Symposium on Circuits and Systems (ISCAS'09)*, pp. 1253–1256, 2009.
- [173] Z. Yang *et al.*, "A neuron signature based spike feature extraction algorithm for on-chip implementation," *IEEE Engineering in Medicine and Biology Society (EMBC'08)*, pp. 1716–1719, 2008.
- [174] T.-C. Chien, W. Liu, and L.-G. Chen, "Vlsi architecture of leading eigenvector generation for on-chip principal component analysis spike sorting system," *IEEE Engineering in Medicine and Biology Society (EMBC'08)*, pp. 3192–3195, 2008.
- [175] "Leicester university neuronal signal bank," <http://www2.le.ac.uk/centres/csn/research-2/spike-sorting>, accessed : 2016-12-05.
- [176] U. Rutishauser *et al.*, "Online detection and sorting of extracellularly recorded action potentials in human medial temporal lobe recordings, in vivo," *Journal of Neuroscience Methods*, vol. 154, pp. 204–224, 2006.
- [177] V. Karkare *et al.*, "A 75 μ w, 16-channel neural spike-sorting processor with unsupervised clustering," *Symposium on VLSI Circuits Digest of Technical Papers*, pp. 24–25, 2011.

- [178] H. Zhao *et al.*, "A cmos-based neural implantable optrode for optogenetic stimulation and electrical recording," *IEEE Biomedical Circuits and Systems Conference (BioCAS'15)*, pp. 1–4, 2015.
- [179] H. D. Dammak *et al.*, "Design of folded cascode ota in different regions of operation through gm/id methodology," *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, vol. 2, no. 9, 2008.
- [180] S.-O. S.-D. M. in Standard CMOS Technology, "A 64 channel programmable closed-loop neurostimulator with 8 channel neural amplifier and logarithmic adc," *Serbian Journal of Electrical Engineering*, vol. 1, no. 2, pp. 37–44, 2004.
- [181] D. A. J. Tony Chan Carisone and K. W. Martin, *Analog Integrated Circuit Design*. Don Fowley, 2012.
- [182] E. Hogenauer, "An economical class of digital filters for decimation and interpolation," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. ASSP-29, no. 2, pp. 155–162, 1981.
- [183] "Design library : Arm library for gf (ibm) 0.13 μm cmos," <https://www.cmc.ca/WhatWeOffer/Products/CMC-00001-43485.aspx>, accessed : 2016-06-29.

Annexe A

Article supplémentaire

International Solid-State Circuits Conference 2018 (ISSCC'18)

© 2019 IEEE. Reprinted, with permission, from G. Gagnon-Turcotte, C. Éthier, Y. De Koninck and B. Gosselin, "A 0.13- μm CMOS SoC for Simultaneous Multichannel Optogenetics and Electrophysiological Brain Recording", *IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 466-468, 2018.

29.4 A 0.13 μm CMOS SoC for Simultaneous Multichannel Optogenetics and Electrophysiological Brain Recording

Gabriel Gagnon-Turcotte, Christian Ethier, Yves De Koninck, Benoit Gosselin

Laval University, Quebec City, Canada

Optogenetics and multi-unit electrophysiological recording are state-of-the-art approaches in neuroscience to observe neural microcircuits *in vivo* [1]. Thereby, brain-implantable devices incorporating optical stimulation and low-noise data acquisition means have been designed based on custom integrated circuits (IC) to study the brain of small freely behaving laboratory animals. However, no existing IC provides multichannel optogenetic photo-stimulation along with multi-unit electrophysiological recording capability within the same die [2-5]. They also lack critical features: they are not multichannel and/or do not include an ADC [6], or they address only one signal modality [5-6], i.e., either local field potentials (LFP) or action potentials (AP). In this paper, we report an IC for simultaneous multichannel optogenetics and electrophysiological recording addressing both LFP and AP signals all at once. This 0.13 μm CMOS chip, which includes 4/10 stimulation/recording channels, is enclosed inside a small wireless optogenetic platform, and is demonstrated with simultaneous *in vivo* optical stimulation and electrophysiological recordings with a virally mediated Channelrhodopsin-2 (ChR2) rat.

Key circuits for electrophysiological recording and optogenetics are depicted in this paper. We present a 3rd-order discrete-time (DT) $\Delta\Sigma$ ADC architecture with on-chip decimation filters (DF). This $\Delta\Sigma$ MASH 1-1-1 employs a bias duty-cycling strategy that decreases its power consumption and a digital subtraction technique between the $\Delta\Sigma$ branch outputs that reduces its footprint size. This design offers an alternate solution to the widespread use of SAR ADC by trading analog complexity for digital circuits, and by enabling a tunable precision. In particular, using a 3rd-order $\Delta\Sigma$ allows for an oversampling ratio (OSR) small enough to put the DF on-chip, unlike in previous designs [4-5]. On the same die, optical stimulation is performed by a 4-ch LED driver circuit with feedback to precisely set the effective forward current of each LEDs, which improves previous solutions based on open-loop drivers [2,6].

Figure 29.4.1 shows the schematic of the analog-front-end (AFE) circuit. It consists of a fully differential folded-cascode operational transconductance amplifier (OTA), using source-degeneration resistors to limit the input-referred noise contribution to only two transistors (M_3 and M_{11}) [3]. The ac-coupled OTA employs two capacitors and two pseudoresistor banks as feedback elements, which provides a 3b tunable high-pass filter cutoff frequency. This allows the selection of a suitable bandwidth for either the LFP or AP signal, while the low-pass analog cutoff frequency is fixed by capacitor C_5 .

Figure 29.4.2 shows the block diagram of the $\Delta\Sigma$ MASH 1-1-1 neural ADC. The 1st stage consists of a DT $\Delta\Sigma$ modulator fed by the conditioned output signal of the AFE, while the 2nd and 3rd stages are two DT $\Delta\Sigma$ modulators fed by the DAC value ($DAC_{x-1}(z)$) minus the integrator output signal ($I_{x-1}(z)$) from their respective previous stages. As a result, each stage cancels out the quantization noise induced by its previous stage. While switched-capacitor (SC) circuits are typically used to perform the subtraction between the DAC value and the integrator output [4], the reported topology improves previous solutions by performing this subtraction digitally instead, which simplifies the circuit implementation: instead of performing $DAC_{x-1}(z) - I_{x-1}(z)$ using SC, $DAC_x(z) - DAC_{x-1}(z)$ is performed using the output bits of the current and previous D-Latches by the DAC with pre-subtraction module (see Fig. 29.4.2). The result of $DAC_x(z) - DAC_{x-1}(z)$ can take only three possible values: VSS, VDD or V_{cm} that are selected digitally using analog switches, which holds for the following special case used in this design: $DAC_{ref1} = \frac{3}{4} VDD$ and $DAC_{ref2} = \frac{1}{4} VDD$. The reduced OSR needed by the $\Delta\Sigma$ MASH 1-1-1, compared with 1st- or 2nd-order $\Delta\Sigma$, allows the OTAs in each of the $\Delta\Sigma$ branches to remain idle for long periods of time. Therefore, a *bias duty cycling* technique is used to turn off the OTA bias current between each sampling cycle, allowing a decrease in power consumption of 41.3 %, for a bias duty cycle of 35%, without noticeable impact on the precision.

Figure 29.4.3 shows the block diagram of the 4th-order cascaded integrator-comb (CIC4) DF. Unlike [4-5], this $\Delta\Sigma$ with on-chip DF provides tunable precision and adjustable bandwidth by varying the OSR and the *clock reduction factor* M (see

Fig. 29.4.3) depending on the AFE performance and type of signal to record (i.e., LFP or AP). This filter uses a highly optimized CMOS implementation strictly based on adders and registers which is avoiding power-hungry multipliers and dividers. To use as few output bits per stage as possible without degrading the precision, the minimal binary precision of each stage is determined using the Hogenauer pruning technique, so as to accommodate $OSR \leq 50$, i.e., DF resolution $\leq 14b$. The DF is designed to provide an attenuation of 15dB at the Nyquist frequency, by setting M to 25.

The schematic of the 4-ch LED driver circuit is presented in Fig. 29.4.4. Compared to previous designs [2,6], this circuit allows to precisely control the value of the forward current into each LED using a regulated cascode current source with feedback. The current can be adjusted independently of the LED parameters (forward voltage, etc.) and the battery voltage, guaranteeing that the right optical power ($>0.1\text{mW/mm}^2$) is delivered by each LED to properly activate ChR2 neurons [1]. In Fig. 29.4.4, the wide-swing current mirror keeps the drain-source voltage of M_2 at the edge of saturation ($V_{ds2} \approx V_{eff}$), minimizing the voltage drop inside the chip to accommodate smaller battery voltages, and avoiding the need for small current sensing resistors with poor process accuracy and inter-channel matching. The LED driver is able to deliver constant forward currents into different LEDs typically used for optogenetics, with $V_{forward}$ varying from 2.9 to 3.36V (see Fig. 29.4.4).

The input-referred noise of the AFE is of $3.2\mu\text{V}_{rms}$ over a 7kHz BW and its THD is of 0.8% with a 1kHz/2.5mV_{p-p} signal. Using the same input signal, the AFE in series with the $\Delta\Sigma$ produces an ENOB of 7.38b with an input of 2.5mV_{p-p}, yielding an ENOB of 8.68b at full scale (6mV_{p-p}), whereas the theoretical ENOB of the $\Delta\Sigma$ MASH 1-1-1 is of 10.8b with an OSR of 25 ($F_s = 500\text{kHz}$). At this OSR, the ENOB is limited by the THD and the noise of the AFE. The power consumption of the AFE, the $\Delta\Sigma$ and the DF is of 11.2 $\mu\text{W}/\text{ch}$. The measured performance of the circuits is summarized in Fig. 29.4.5.

A Long-Evans rat with virally mediated ChR2 cortical expression (four sites in primary motor cortex injected with 0.3 μl of AAV2/6 hSyn-ChR2-EGFP-Nav1.2ts viral construction, title of 8×10^{12} GC/ml) was used to stimulate and record within two *in vivo* experiments wherein the rat was anesthetized with ketamine/xylazine. First, a micropipette (3 μm tip) was lowered in the VPM region of the brain (4.2mm depth). Secondly, an optrode ($8 \times 0.5\text{M}\Omega$ electrodes, $1 \times$ fiber) was lowered in the cerebral motor cortex (1.5mm depth). All protocols were performed in accordance with Canadian Council on Animal Care Guidelines.

As seen in Fig. 29.4.6 (top left), the chip enclosed within the wireless test platform was used to record spontaneous activity of maximum amplitude of 1.3mV_{p-p} and a noise floor of 17 μV_{rms} using the glass micropipette. Figure 29.4.6 (top right) presents several clusters of spikes collected within 3 recording sessions. Figure 29.4.6 (bottom left) presents the wireless platform, which is using the IC, during the *in vivo* experiment and Fig. 29.4.6 (bottom right) shows the evoked neural activity acquired after optical stimulation through the optrode. The platform concept and IC are shown in Fig. 29.4.7.

Acknowledgments:

We thank CMC Microsystems for supporting the chip fabrication. This project is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), the *Fonds de Recherche du Québec - Nature et Technologies* (FRQNT) and the Weston Foundation.

References:

- [1] G. Buzsáki, et al., "Tools for Probing Local Circuits: High-Density Silicon Probes Combined with Optogenetics," *Neuron*, vol. 86, no. 1, pp. 92-105, 2015.
- [2] H.-M. Lee, et al., "A Power-Efficient Switched-Capacitor Stimulating System for Electrical/Optical Deep Brain Stimulation", *IEEE JSSC*, vol. 50, no. 1, pp. 360-374, 2015.
- [3] W. Wattanapanitch, et al., "An Energy-Efficient Micropower Neural Recording Amplifier," *IEEE TBioCAS*, vol. 1, no. 2, pp. 136-147, 2007.
- [4] L. Wang, et al., "An 18 μW 79dB-DR 20KHz-BW MASH $\Delta\Sigma$ Modulator Utilizing Self-Biased Amplifiers for Biomedical Applications", *IEEE CICC*, 4 pages, 2011.
- [5] H. Kassiri, et al., "All-Wireless 64-Channel 0.013mm²/ch Closed-Loop Neurostimulator with Rail-to-Rail DC Offset Removal", *ISSCC Dig. Tech. Papers*, pp. 452-453, 2017.
- [6] C.-H. Chen, et al., "An Integrated Circuit for Simultaneous Extracellular Electrophysiology Recording and Optogenetic Neural Manipulation", *IEEE TBME*, vol. 64, no. 3, pp. 557-568, 2017.

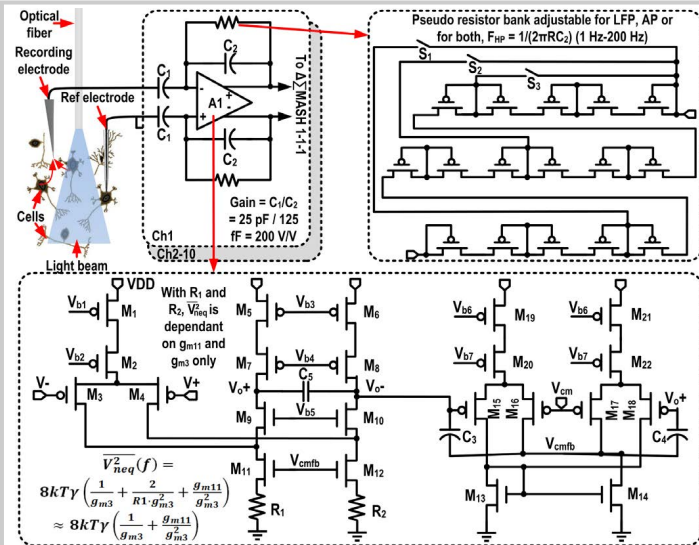


Figure 29.4.1: Circuit schematic of the AFE using folded-cascode OTA with source-degeneration resistors R_1 and R_2 .

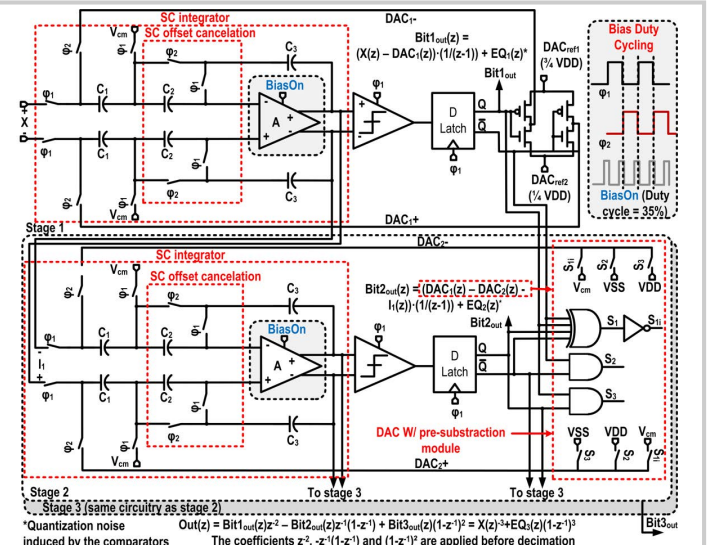


Figure 29.4.2: $\Delta\Sigma$ MASH 1-1-1 using the pre-subtraction module (2nd & 3rd stages) and a bias duty-cycling technique.

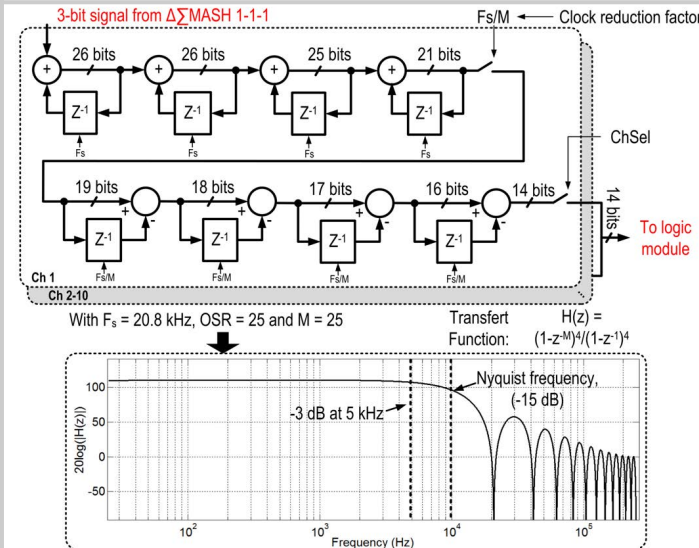


Figure 29.4.3: CIC4 DF with optimized binary precision between the stages, for up to 14b output resolution.

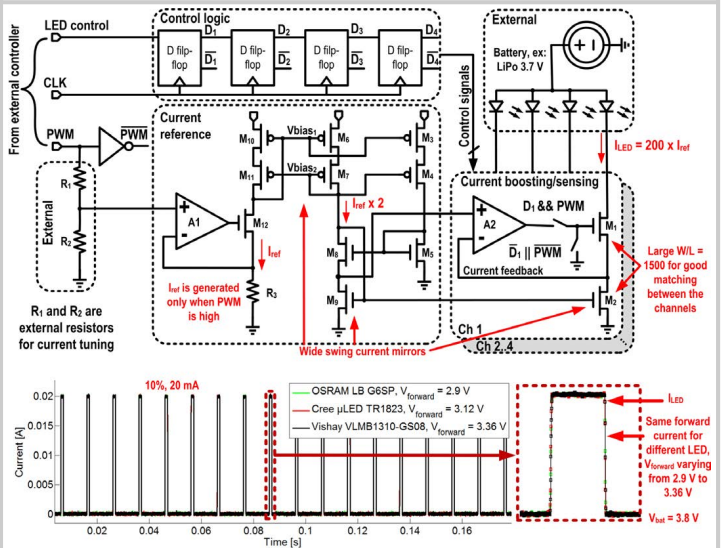
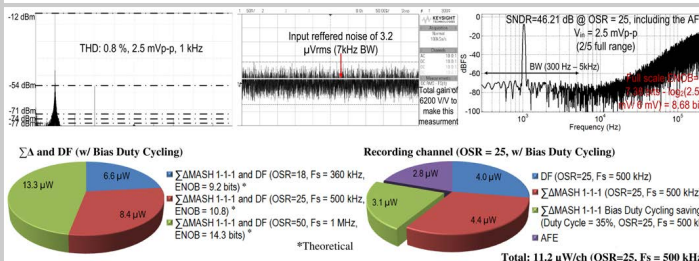


Figure 29.4.4: Circuit schematic of the LED driver (top); regulated current for 3 different types of LEDs (bottom).



	[2]	[3]	[4]	[5]	[6]	This work
TECHNOLOGY (μm)	0.35	0.5	0.13	0.13	0.18	0.13
TYPE	OSTim. ¹	Rec. ¹	Rec.	Estim./Rec.	OSTim./Rec.	OSTim./Rec.
Nb. CHANNELS (OSTim./Rec.)	4/0	0/1	0/1	0/64	1/1	4/10
TARGETED SIGNALS	-	APILFP	-	LFP	AP	APILFP
ON CHIP DECIMATION	-	-	NO	NO	-	YES
AFE BANDWIDTH (Hz)	-	0.39-5.3k	-	0.01-500	114-14.6k	0.1-7k
POWER/CH (μW)	-	7.56 (AFE)	18 ($\Sigma\Delta$)	0.63 (AFE+ $\Sigma\Delta$)	-	11.2 (AFE+ $\Sigma\Delta$ +DF)
NOISE BW (Hz)	-	45-5.3k	-	0.1-500	300-5k	10-7k
Input Ref. NOISE (μV_{rms})	-	3.06	-	1.13	4.57	3.2
NEF	-	2.67	-	2.86	2.30 ²	2.30 ²
OSTim. CURRENT (mA)	0.4	-	-	0.01-1.35 ¹	0-330	0-35
OSTim. CURRENT REGULATION	NO	-	-	NO	NO	YES
ENOB	-	-	11.5 ⁴	11.7 ⁹	-	8.68 ⁸
OSR	-	-	64	1000	-	25

Figure 29.4.5: Experimentally measured performance (top); Electrophysiological recording and optogenetic IC comparison (bottom).

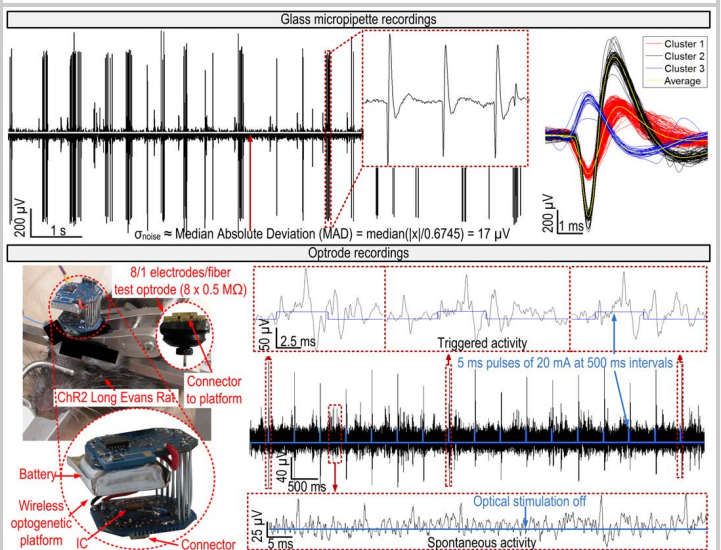


Figure 29.4.6: Neural activity collected with a glass micropipette (top) and wirelessly using an optrode (bottom).

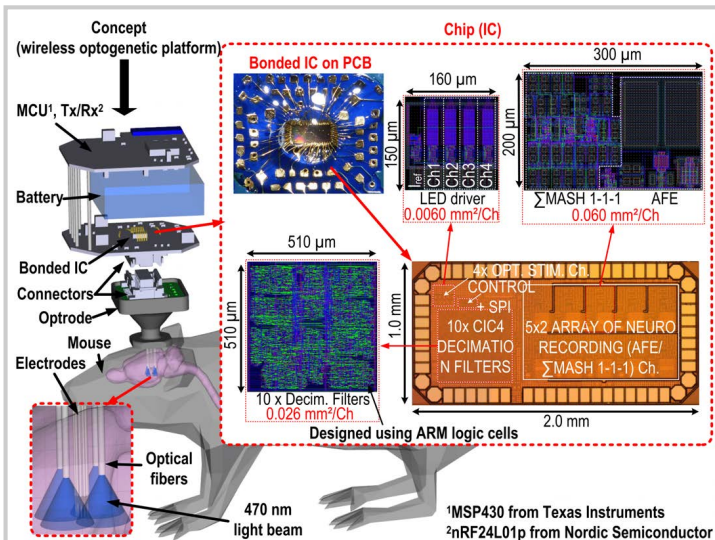


Figure 29.4.7: Die micrograph (right); system-level concept using the IC within an optogenetic platform (left).

Annexe B

Analyses complémentaires

Certaines parties de cette annexe sont inspirées du mémoire de maîtrise [5] et de la proposition de thèse de doctorat de l'auteur de cette thèse.

B.1 Algorithme de seuillage adaptatif

B.1.1 Revue de littérature

Algorithmes de détection

Une méthode simple afin de réduire la quantité de données à transmettre est d'effectuer une détection des PAs au niveau du capteur et de transmettre seulement les PAs détectés. Par exemple, le capteur développé dans [148] permet de compresser de 15Mb/s à 330kb/s les données de 100 électrodes en transmettant seulement l'état, détecté ou non, en temps réel. Le système, complètement développé avec la technologie CMOS 0.5 μm , détecte les PA en comparant les signaux avec un seuil. Chaque comparateur est interrogé à toutes les millisecondes et la valeur (détectée ou non) est transmise dans un paquet avec les données des autres électrodes.

Au lieu de transmettre l'état de détection pour chacun des canaux, certains dispositifs transmettent des caractéristiques spécifiques à chacun des PA détectés. Par exemple, [83] regroupe les potentiels en *clusters* et transmet seulement l'information reliée au *clusters*, sans transmettre le PA. [149] quant à lui transmet la valeur crête à crête maximale ainsi que la dérivée maximale et minimale des PAs détectés, tandis que [150] transmet l'amplitude maximale et minimale des PAs ainsi que le délai temporel entre ces deux valeurs.

Envoyer seulement l'information sur la détection ou certaines caractéristiques des PAs est souvent insuffisant. Transmettre la forme d'onde complète pour chacun des PA détectés est toutefois désiré. Une approche réalisée dans [14] consiste à effectuer une détection des PA analogiquement et à utiliser un filtre FIR (*switched capacitors*) comme tampon de la forme

d'onde complète. Cette technique permet d'activer l'ADC ainsi que les autres composantes numériques seulement lorsqu'il y a une détection de PA afin d'économiser de l'énergie.

Enfin, il existe plusieurs méthodes de détection des PAs dans le domaine numérique. [86] présente et compare trois groupes d'algorithmes de détection, soit à base de seuil, d'énergie et de filtres. Les méthodes incluses dans chacun des groupes sont les suivantes :

– **Seuil :**

– Seuillage absolu [86] :

$$y(n) = |x(n)|, \begin{cases} \text{Si } y(n) > \text{seuil, un AP est détecté} \\ \text{sinon, aucun AP est détecté} \end{cases}$$

– Seuillage positif [151], [148], [150], [152], [153] :

$$y(n) = x(n), \begin{cases} \text{Si } y(n) > \text{seuil, un AP est détecté} \\ \text{sinon, aucun AP est détecté} \end{cases}$$

– Seuillage négatif [151], [150] :

$$y(n) = x(n), \begin{cases} \text{Si } y(n) < \text{seuil, un AP est détecté} \\ \text{sinon, aucun AP est détecté} \end{cases}$$

– **Énergie :**

– NEO (connus parfois sous l'acronyme TEO) [83], [149], [69], [154] :

$$NEO(n) = x(n)^2 - x(n-1) \times x(n+1), \begin{cases} \text{Si } NEO(n) > \text{seuil, un AP est détecté} \\ \text{sinon, aucun AP est détecté} \end{cases}$$

– MTEO :

$$MTEO(n) = x(n)^2 - x(n-k) \times x(n+k), \begin{cases} \text{Si } MTEO(n) > \text{seuil, un AP est détecté} \\ \text{sinon, aucun AP est détecté} \end{cases}$$

– **Filtres :**

- Convolution du signal avec une banque de filtres.
- Convolution du signal généré par NEO avec une banque de filtres.
- Convolution du signal en valeur absolue avec une banque de filtres.

Algorithmes de seuillage

Certains algorithmes de détection, comme ceux à base de seuil ou d'énergie, doivent être utilisés conjointement à un seuil. Une méthode simple consiste à utiliser un seuil fixe pouvant être réglé par l'utilisateur en fonction du niveau de bruit inclus dans le signal. Cette

méthode est utilisée dans [148], [151], [152] et [153], où le seuil doit être ajusté manuellement par l'utilisateur en fonction du niveau de bruit.

Les microélectrodes utilisées pour capter les signaux neuronaux possèdent une grande impédance (jusqu'à plusieurs dizaines de $M\Omega$) ce qui peut générer un ratio signal sur bruit (SNR) faible. De plus, le SNR du signal n'est pas constant dans le temps et peut varier en fonction de plusieurs facteurs tels que le déplacement des électrodes, en fonction du phénomène de *microelectrode drifting* et à cause du bruit neuronal. Une technique de seuillage adaptative est donc hautement désirée. La majorité des algorithmes de seuillage adaptatif sont basés soit sur l'écart-type du signal, sur la valeur RMS du signal ou bien sur la valeur moyenne du signal.

Une méthode, utilisée dans le programme de *clustering* automatique *open source* nommé *Wave Clus* [155], [156], calcule une approximation de l'écart-type du signal en utilisant la déviation médiane absolue (MAD) selon cette équation :

$$\delta = \text{median}\left\{\frac{|x|}{0.6745}\right\} \quad (\text{B.1})$$

Où x est le signal filtré passe-bande. Avec cette technique, le seuil est ensuite calculé comme un multiple de δ . Une autre technique permettant de calculer une estimation de l'écart type est présentée dans [157]. Cette technique, entièrement analogique, estime l'écart-type en utilisant une boucle de contrôle et en exploitant certaines propriétés statistiques de la loi normale.

La méthode proposée dans [158] calcule la valeur RMS du signal à partir d'un nombre restreint d'échantillons :

$$X_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (\text{B.2})$$

Où x_i est le $i^{\text{ème}}$ échantillon filtré passe-bande. Le seuil est ensuite évalué comme un multiple de X_{rms} .

Une autre méthode de seuillage automatique adaptée à l'algorithme NEO est présentée dans [83]. Cet algorithme calcule la valeur moyenne de NEO selon l'équation suivante :

$$Thr = \frac{1}{N} \sum_{i=1}^N NEO(i) \quad (\text{B.3})$$

Où N est le nombre d'échantillons utilisés dans les calculs. Le seuil est ensuite évalué comme un multiple de Thr . De façon similaire, [154] calcul le seuil de l'algorithme NEO selon sa moyenne trouvée analogiquement avec un filtre passe-bas possédant une très grande constante de temps.

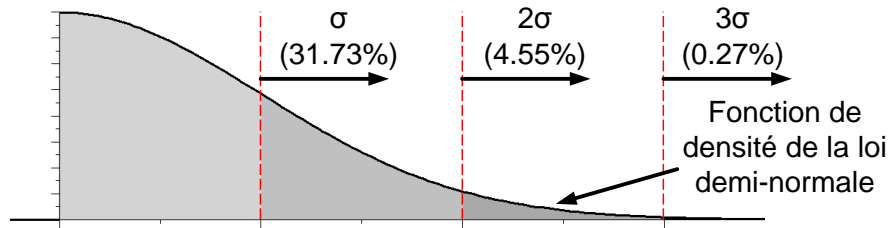


FIGURE B.1 – Fonction de densité de la loi demi-normale présentée avec des multiples de son écart-type (σ).

B.1.2 Solution retenue

Détecteur

Il a été démontré que l’algorithme de détection à valeur absolue obtient majoritairement les meilleurs résultats de détection tout en nécessitant le moins de temps de calcul parmi tous les algorithmes mentionnés à la sous-section B.1.1 [5, 85, 86]. Pour cette raison, c’est cet algorithme qui est retenu pour être implémenté de façon permanente dans tous les *headstages* présentés dans cette thèse.

Seuil adaptatif : aspects mathématiques

L’algorithme de détection des PA à valeur absolue nécessite l’utilisation d’un seuil qui doit être fixé selon l’amplitude du bruit. La nouvelle technique de seuillage automatique proposée s’inspire de [157] et consiste à estimer l’écart-type du bruit en exploitant certaines propriétés statistiques de la loi demi-normale. Puisque le bruit provenant de l’interface entre les microélectrodes à haute impédance et les neurones est considéré comme dominant [159] (en ne tenant pas compte du bruit neuronal provenant de neurones éloignés), le bruit peut être approximé par du bruit blanc gaussien possédant une distribution statistique normale [157]. Ce bruit, une fois replié par l’algorithme de détection à valeur absolue, suit une loi demi-normale. La densité de probabilité de la loi demi-normale est présentée à la figure B.1 et est définie par l’équation suivante :

$$f(x) = \begin{cases} \text{Si } x > 0, & \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2\sigma^2}} \\ \text{Sinon,} & 0 \end{cases}$$

Où σ est l’écart-type du signal. Avec cette densité de probabilité, on sait que la probabilité que la valeur absolue d’un bruit gaussien excède l’écart-type du signal est de 31.7% :

$$31.7\% = 100\% - \int_0^\sigma \frac{1}{\sigma} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2\sigma^2}} dx \times 100\%. \quad (\text{B.4})$$

Afin d’exploiter cette propriété, l’algorithme proposé estime l’écart-type de la valeur absolue du bruit à l’aide d’une boucle de contrôle. Dans cette boucle, le procédé à contrôler

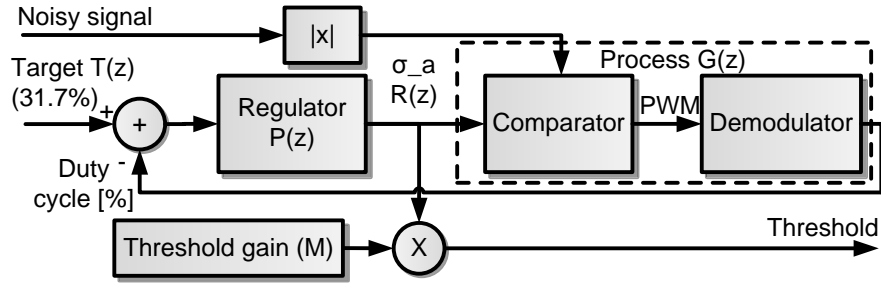


FIGURE B.2 – Boucle de contrôle permettant de calculer le seuil de détection basé sur un multiple de l'écart-type estimé (σ_a) du signal.

estime le ratio du nombre d'échantillons supérieur à l'écart-type. Ce procédé numérique compare la valeur absolue du signal bruité avec la sortie du régulateur. Cette comparaison génère un signal de type PWM, 1 pour plus grand et 0 pour plus petit, possédant un *duty cycle* exprimant le nombre d'échantillons dont la valeur absolue est plus grande que la sortie du régulateur. Le procédé applique ensuite une démodulation afin de transformer le signal PWM en ratio pouvant être soustrait à la consigne dans la rétroaction de la boucle. À chaque nouvelle itération déclenchée par l'arrivée d'un nouvel échantillon, le régulateur tente de minimiser l'erreur entre la consigne, de valeur 31.7%, afin de trouver l'approximation de l'écart-type. Une fois que la boucle a atteint son régime permanent, le signal de contrôle du régulateur représente l'écart-type estimé. Cette boucle de contrôle est présentée à la figure B.2, où σ_a est l'écart-type estimé et M est un facteur multiplicatif permettant d'ajuster le seuil ($Seuil = M \times \sigma_a$). La démonstration mathématique de cet algorithme est présentée ci-dessous. Dans cette démonstration, $T(z)$ est la consigne, $P(z)$ est le régulateur basé sur un intégrateur afin d'assurer une erreur nulle en régime permanent, $G(z)$ est le procédé et $R(z)$ est la sortie du régulateur. Premièrement, la fonction de transfert discrète du régulateur peut être exprimée comme suit :

$$R(z) = P(z) \cdot (T(z) - R(z) \cdot G(z)) \quad (B.5)$$

Ce qui devient :

$$R(z) = \frac{P(z) \cdot T(z)}{1 + P(z) \cdot G(z)} \quad (B.6)$$

Où $T(z) = \frac{0.317z}{z-1}$ et $P(z) = \frac{kz}{z-1}$ avec k le gain de l'intégrateur. Selon le théorème de la valeur finale, il est possible de calculer $R(\infty)$ avec l'équation suivante :

$$R(\infty) = \lim_{z \rightarrow 1} R(z)(1 - z^{-1}) \quad (B.7)$$

$$\Rightarrow R(\infty) = \lim_{z \rightarrow 1} \frac{\frac{0.317z}{z-1} \cdot \frac{kz}{z-1}}{1 + \frac{kz}{z-1} \cdot G(z)} \cdot (1 - z^{-1}) \quad (B.8)$$

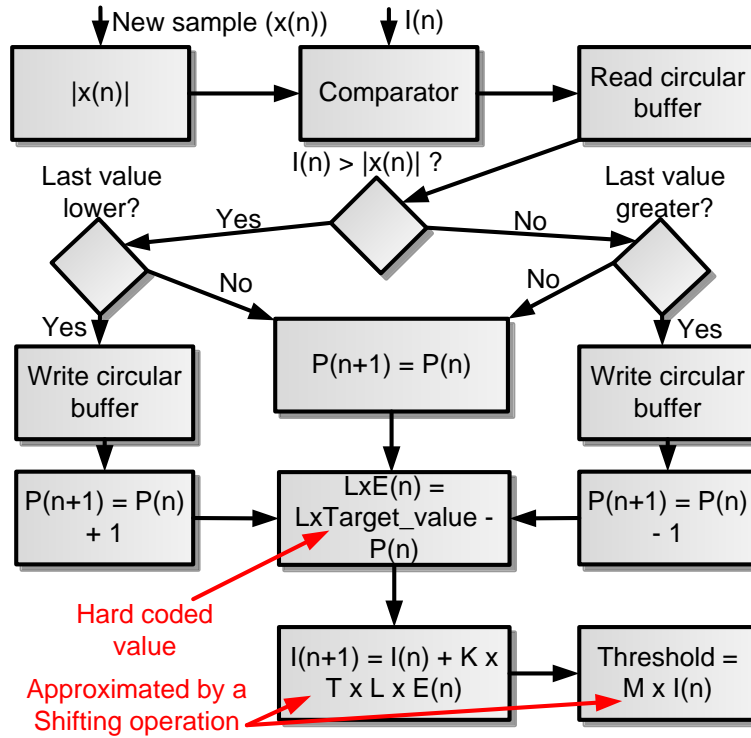


FIGURE B.3 – Diagramme fonctionnel de l’algorithme permettant de calculer le seuil de détection.

$$\Rightarrow \lim_{z \rightarrow 1} = \frac{0.317kz}{(z-1) + kzG(z)} = \frac{0.317}{G(1)} \quad (\text{B.9})$$

$$\Rightarrow R(\infty) \cdot G(1) = 0.317 \quad (\text{B.10})$$

Or, $R(z) \cdot G(z)$ peut-être exprimé par l’équation suivante :

$$R(z) \cdot G(z) = \frac{1}{N} \sum_{n=0}^{N-1} z^{-n} \begin{cases} 1 \text{ si } |x| \cdot z^{-n} > R \cdot z^{-n} \\ 0 \text{ sinon} \end{cases} \quad (\text{B.11})$$

Ainsi, $R(\infty) \cdot G(1)$ devient :

$$R(\infty) \cdot G(1) = \frac{1}{N} \sum_{n=0}^{N-1} \begin{cases} 1 \text{ si } |x| > R(\infty) \\ 0 \text{ sinon} \end{cases} = 0.317 \quad (\text{B.12})$$

En prenant $N \rightarrow \infty$, l’on obtient le ratio d’échantillons supérieurs à $R(\infty)$, soit 0.317 ce qui correspond à l’écart-type ($R(\infty) = \sigma$). En calculant l’algorithme sur une fenêtre de $N < \infty$ échantillons, l’on obtient une approximation de l’écart-type (σ_a).

B.1.3 Implémentation optimisée pour circuit CMOS ou MCU à faible puissance

Dans un premier temps, le régulateur $P(z)$ est implémenté en utilisant un intégrateur afin d’assurer une erreur nulle sur σ en régime permanent. Cet intégrateur est construit en utili-

sant une somme de Riemann :

$$I(n+1) = I(n) + K \times T \times E(n) \quad (\text{B.13})$$

Où K est le gain de l'intégrateur, T est la période d'échantillonnage et $E(n)$ le signal d'erreur. La sortie du régulateur $I(n)$, qui est aussi σ_a , est comparée avec la valeur absolue de l'échantillon courant. La sortie du comparateur est mise à 1 ou à 0 selon le résultat de la comparaison afin de générer le signal PWM. Un tableau en mémoire RAM est utilisé comme tampon circulaire afin de stocker 256 bits de comparaison par canal ($L = 256$, déterminé empiriquement). Avant de stocker le nouveau bit de comparaison, le dernier bit devant être écrasé est lu. Si le nouveau bit est à 1 et le dernier bit à 0, un compteur est incrémenté. Si le nouveau bit est à 0 et le dernier bit à 1, le même compteur est décrémenté. Dans tous les autres cas, la valeur du compteur reste inchangée. Ce compteur contient la somme de tous les bits en mémoire qui, lorsque divisé par la taille du tampon circulaire L , représente le pourcentage d'échantillons dont la valeur était supérieure à σ_a et peut être représenté mathématiquement par $P(n)$. Puisque la taille du tampon circulaire est connue à l'avance, et pour éviter d'utiliser une opération de division, la consigne de la boucle de contrôle est ajustée à 31.7% multipliée par la taille du tampon circulaire. L'opération $E(n) = \text{consigne} - P(n)/L$ devient $L \times E(n) = L \times \text{consigne} - P(n)$. Le facteur L qui multiplie l'erreur $E(n)$ peut être joint avec le gain K de l'intégrateur et avec la période d'échantillonnage T . Puisque K , L et T sont connus à l'avance, la multiplication $K \times T \times L \times E(n)$ est approximée par une simple opération de décalage de bits de l'erreur $E(n)$. Idem pour la multiplication de σ_a déterminant le seuil qui est aussi approximée par une simple opération de décalage de bits. Les étapes de l'algorithme sont présentées dans le diagramme de la figure B.3.

B.1.4 Innovations de l'algorithme de seuillage adaptatif

La méthode de seuillage adaptative proposée ne nécessite aucun circuit dédié et est entièrement numérique, soit deux avantages de taille pour diminuer les coûts, augmenter la fiabilité et accélérer le développement. Cette caractéristique avantage l'algorithme proposé, contrairement au système précédent présenté dans [157], qui utilise une implémentation analogique nécessitant un circuit dédié. Un circuit ASIC numérique exploitant certaines propriétés de la loi normale afin de seuiller les coefficients d'une DWT est présenté dans [160]. L'algorithme présenté dans [160] diffère de celui présenté dans cette section selon plusieurs points. Notamment, cet algorithme est exécuté sur une fenêtre fixe de coefficients, ce qui l'empêche d'être exécuté sur un signal continu, c.-à-d. qu'il ne possède aucune mémoire des événements précédant la fenêtre, contrairement à l'algorithme proposé qui utilise un démodulateur avec mémoire interne, lui permettant d'être exécuté en continu avec un minimum d'opérations. De plus, les systèmes présentés dans [157] et [160] exploitent les propriétés de la loi normale du bruit gaussien, alors que le nouvel algorithme exploite une loi demi-normale, ce qui constitue un net avantage étant donné que le bruit n'est jamais parfaitement

gaussien en pratique et souvent non symétrique autour de 0 volt. Avec la méthode proposée, c'est l'intégralité du bruit qui est utilisé pour déterminer le seuil de détection, alors que les méthodes précédentes [157] et [160] n'utilisent que les échantillons positifs supérieurs au seuil estimé dans ses calculs. Cette amélioration confère beaucoup plus de précision au nouvel algorithme lorsque le seuil est combiné à un algorithme de détection en valeur absolue dont le but est de détecter la présence de PA dans un signal bruité.

B.2 Algorithme de compression

Afin d'augmenter significativement le nombre de canaux d'enregistrement, les *headstages* présentés aux chapitres 2 et 4 compressent les PA détectés en calculant une DWT de 4 niveaux sur les 48 échantillons de chaque PA. La compression est effectuée en seuillant les coefficients générés par la DWT, de façon à conserver seulement ceux possédant la plus grande énergie et en requantifiant dynamiquement les coefficients conservés de façon à maximiser la plage binaire utilisée. La nouvelle technique de compression développée permet une reconstruction des PA avec la même résolution que la forme d'onde originale.

La sélection de la *mother wavelet* a un impact majeur sur la qualité des signaux reconstruits et sur les CR pouvant être atteints [89]. Tel qu'il est démontré à la sous-section B.2.5, l'ondelette de type *Symmlet-2* a produit les meilleurs résultats parmi toutes les méthodes de compression testées. Ce résultat concorde avec ceux présentés dans [89], dans lequel il est démontré que les ondelettes de type *Symmlet* permettent de générer une compression presque optimale des signaux neuronaux.

B.2.1 Implémentation séquentielle la DWT *Symmlet-2* par *lifting*

La figure B.4 présente le schéma d'une transformée en ondelettes *Symmlet-2* de 4 niveaux par *lifting*. $X_e(n)$ et $X_o(n)$ sont les échantillons d'indices pairs et impairs respectivement du signal d'entrée, tandis que Q_0 , Q_1 et Q_2 sont des résultats intermédiaires. Les étapes de *lifting* sont les suivantes :

- $Q_0 = X_o(n-1) + H1 \times X_e(n-1)$
- $Q_1 = X_o(n) + H1 \times X_e(n)$
- $Q_2 = X_e(n) + H3 \times Q_0 + H2 \times Q_1$
- Approx = $H5 \times (Q_0 + H4 \times Q_2)$
- Detail = $H6 \times Q_2$

Où $H1$ à $H4$ sont les quatre coefficients constants du schéma de *lifting* de type *Symmlet-2*, $\sqrt{3}$, $\frac{\sqrt{3}}{4}$, $\frac{\sqrt{3}-2}{4}$ et -1 respectivement, tandis que $H5$ et $H6$ sont les facteurs d'échelle de valeurs $\frac{\sqrt{3}-1}{\sqrt{2}}$ et $\frac{\sqrt{3}+1}{\sqrt{2}}$ respectivement.

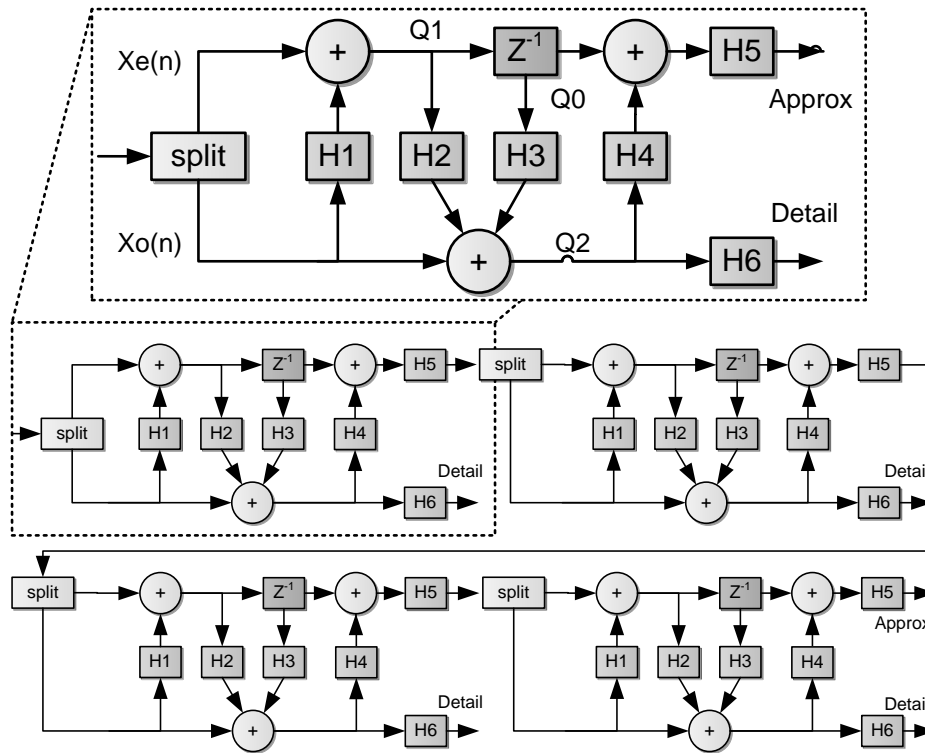


FIGURE B.4 – Représentation schématique d’une transformée en ondelettes de type *Symmlet-2* de 4 niveaux de profondeur par la méthode de *lifting*.

B.2.2 Calcul du seuil des coefficients

Une fois que les 48 coefficients de la transformée sont calculés, ils sont traités par un module de tri. La tâche de ce module consiste à classer les coefficients par amplitude afin de trouver l’amplitude du x^{ieme} coefficient trié qui sera utilisé pour seuiller les coefficients dans le module de compression, où x est le nombre de coefficients à conserver. Pendant l’exécution de l’algorithme de tri, le coefficient possédant la plus grande amplitude est trouvé et sa valeur est transférée au module de compression en même temps que la valeur du seuil des coefficients.

B.2.3 Compression des données

Une fois que le coefficient trié possédant la x^{ieme} amplitude et le coefficient d’amplitude maximale sont trouvés, la compression est effectuée par le module de compression. La stratégie de compression des PA consiste à seuiller les coefficients de la DWT pour ensuite requantifier dynamiquement les coefficients conservés sur un nombre de bits moindre. Premièrement, à l’étape de seuillage, l’amplitude de chacun des coefficients est comparée avec l’amplitude du x^{ieme} coefficient. Si l’amplitude du coefficient courant est supérieure à l’amplitude du x^{ieme} coefficient, un tableau de 48 bits est mis à jour avec un 1 à la position du

précédent et en appliquant cette méthode, avec une amplitude maximale de $0x1000$, le coefficient $0x0010$ devient $0x0100$ (256) et son information ne sera pas perdue lorsque requantifiée sur 8 bits, donnant $0x01$. Ainsi, la méthode proposée permet de conserver 1 bit supplémentaire de précision comparé à une opération de décalage de bits, ce qui est un avantage majeur lorsque les coefficients sont requantifiés sur un nombre de bits très petits, comme 5 ou 6 bits.

Dans une transformée par ondelettes, les coefficients possédant une amplitude supérieure contiennent plus d'information sur la forme d'onde originale que les coefficients de faible amplitude [90]. Une caractéristique de l'approche proposée consiste à transmettre le coefficient d'amplitude maximale en pleine résolution (16 bits), ce qui permet à tous les coefficients de retrouver leur résolution originale lors de l'opération de normalisation inverse, au prix de devoir transmettre 16 bits supplémentaires. Pour chacun des coefficients reçus à la station de base, l'opération suivante est effectuée :

$$ReconstructCoef = MaxAmplitudeCoef \times LowQuantizeCoef \quad (B.14)$$

Où *LowQuantizeCoef* est le coefficient dont la plage dynamique a été maximisée et qui a été requantifié sur un nombre de bits moindre, *MaxAmplitudeCoef* est l'amplitude originale du coefficient de plus grande amplitude en pleine résolution (16 bits) et *ReconstructCoef* est le coefficient reconstruit sur 16 bits. Le CR atteint par la méthode décrite est exprimé par l'équation suivante :

$$CR = \frac{48samples \times 16bits}{48bits + 16bits + N \times Q} \quad (B.15)$$

Où *N* est le nombre de coefficients conservés et *Q* est le nombre de bits employé dans l'opération de compression par requantification.

B.2.4 Innovations

Les méthodes conventionnelles de compression par ondelettes effectuent une compression simplement en conservant les coefficients de la transformée qui sont supérieurs à un seuil [82], [161], [162], ou qui contiennent une certaine quantité d'énergie [162], [163], [164]. La méthode développée se distingue des méthodes de compression par ondelettes conventionnelles en permettant d'atteindre de meilleurs taux de compression grâce à une diminution du nombre de bits de quantification des coefficients conservés tout en minimisant la perte de précision grâce à une maximisation de la plage dynamique de ces derniers. La méthode permet également de reconstruire les signaux compressés avec la même résolution que les données d'origine. Cet avantage essentiel survient grâce à la transmission en pleine résolution du coefficient de la transformée d'amplitude maximale (coefficient contenant le plus d'information sur la forme d'onde initiale) et qui est utilisée lors de l'opération de normalisation inverse pendant la reconstruction. Normalement, diminuer le nombre de bits de quantification des coefficients afin de compresser davantage ne permet pas de recouvrer la résolution

d'origine lors de l'étape de reconstruction des signaux. Ainsi, la méthode proposée innove en permettant de reconstruire le signal en pleine résolution.

B.2.5 Résultats de simulations versus d'autres méthodes

Certains algorithmes de compressions ont été retenus afin d'être implémentés, simulés et comparés dans *Matlab* dans le but d'identifier les algorithmes les plus prometteurs. Notamment, cette section présente des résultats relatifs à la compression par ondelettes de type *Haar* et de type *Symmlet-2*, à la compression par encodage *Delta*, à la compression par diminution du nombre de bits de quantification et à la compression de type CS avec base parcimonieuse dans le domaine des ondelettes (de type *Daubechie-8* avec matrice aléatoire de type *Bernoulli*) et avec base parcimonieuse optimale (K-SVD entraîné sur 5% des PAs).

Les algorithmes ont été implémentés de façon à compresser les PA seulement, car ces derniers sont considérés comme détectés a priori par un algorithme de détection. Les résultats sont exprimés sous forme graphique, afin d'en faciliter l'interprétation et sont présentés pour les CR suivants : 2, 3, 4, 5, 6, 7 et 8. Il est à noter que certains de ces taux ne sont pas réalisables par tous les algorithmes, par exemple, l'algorithme de compression *Delta* ne peut pas produire les CR de 6 et 8.

Les résultats présentés ont été obtenus avec des données neuronales quantifiées sur 16 bits et échantillonnées à 24 kHz issus de la banque de signaux neuronaux en ligne de l'Université Leicester [165].

SNDR

En observant les figures B.6, B.7, B.8 et B.9, on constate que tous les algorithmes subissent une réduction des SNDR moyens, minimums et maximums ainsi qu'une augmentation de la déviation moyenne du SNDR lorsque le CR augmente.

C'est l'algorithme de compressions de type CS avec matrice de parcimonie optimale qui offre le meilleur SNDR moyen pour presque tous les CR, suivi de l'algorithme à base de l'ondelette *Symmlet-2*. Les résultats sont similaires pour le SNDR maximum, à la différence que l'algorithme de compressions de type *Delta* a généré les meilleurs résultats pour un $CR \leq 3$ et qu'il se classe deuxième pour les CR supérieurs. Les bons résultats de CS avec base parcimonieuse optimale étaient prévisibles dans ces deux tests, puisque les PAs ayant une forme similaire aux PAs ayant servi à construire la matrice de parcimonie optimale ont bénéficié d'une reconstruction presque parfaite.

Les résultats sont complètement différents pour le SNDR minimum, dans lequel l'algorithme de compression à base de l'ondelette *Symmlet-2* a généré les meilleurs résultats pour tous les CR, suivi de l'algorithme de compression à base de l'ondelette *Haar*. Les deux algorithmes de type CS ont généré des SNDR négatifs pour des $CR \geq 4$, ce qui est très mauvais, alors

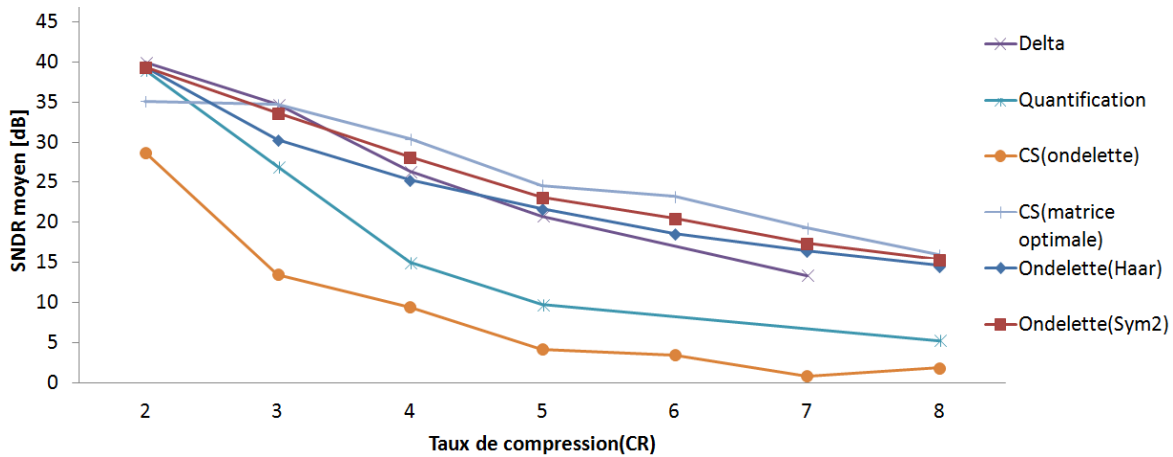


FIGURE B.6 – SNDR moyen obtenu avec tous les algorithmes testés pour un CR de 2 à 8.

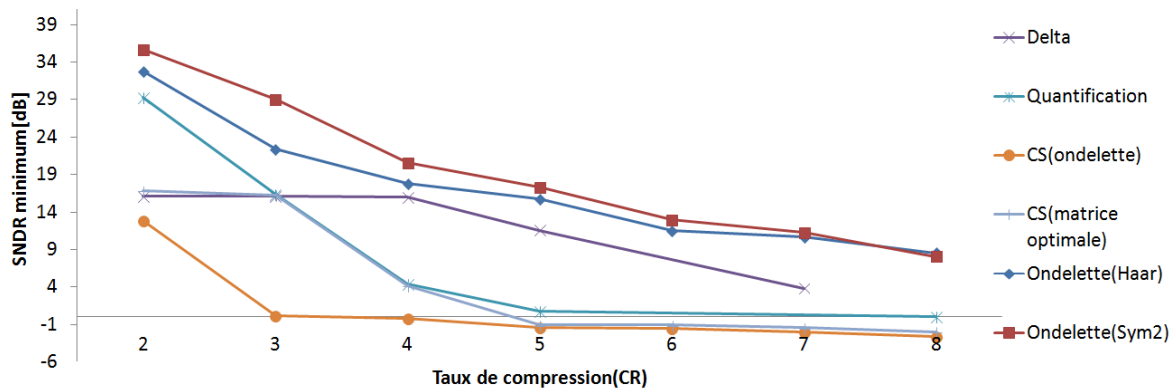


FIGURE B.7 – SNDR minimum obtenu avec tous les algorithmes testés pour un CR de 2 à 8.

que l’algorithme à faible quantification oscille autour de 0 dB pour les $CR \geq 5$. Il est à noter que l’algorithme CS avec matrice de parcimonie optimale reproduit très mal les PAs qui sont légèrement différents de ceux qui ont été utilisés pour construire la matrice de parcimonie. Cette caractéristique est un désavantage majeur, puisque dans des conditions réelles, des PAs ayant des formes différentes peuvent se produire à tout moment.

Tel que stipulé, certains algorithmes ont de la difficulté à reproduire fidèlement toutes les formes de PA. C’est le cas des deux algorithmes de type CS, de l’algorithme *Delta* et de l’algorithme à faible quantification qui ont tous générés des déviations moyennes du SNDR pouvant dépasser les 30 % dans certains cas, alors que les deux algorithmes de compression par ondelettes n’ont jamais dépassé les 10.5 % pour tous les types de PA.

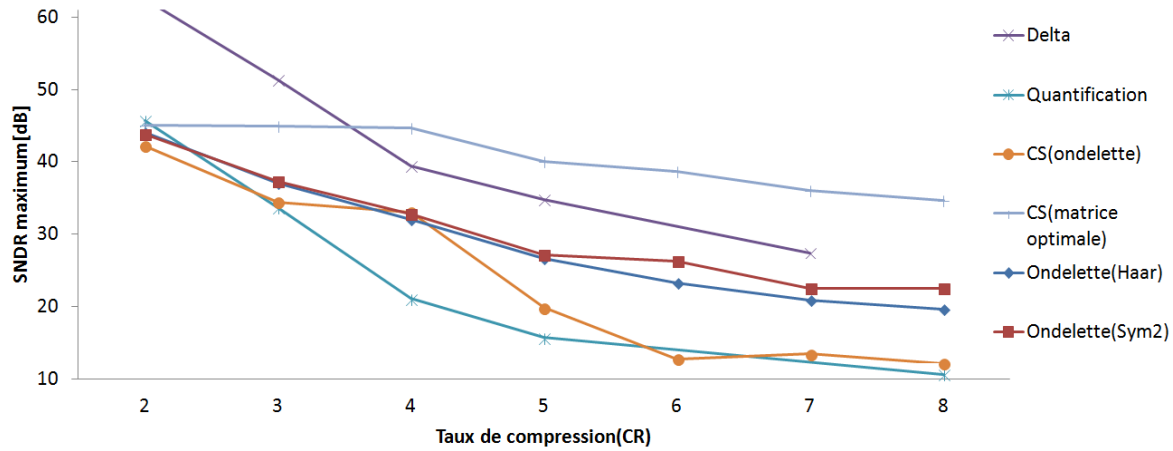


FIGURE B.8 – SNDR maximum obtenu avec tous les algorithmes testés pour un CR de 2 à 8.

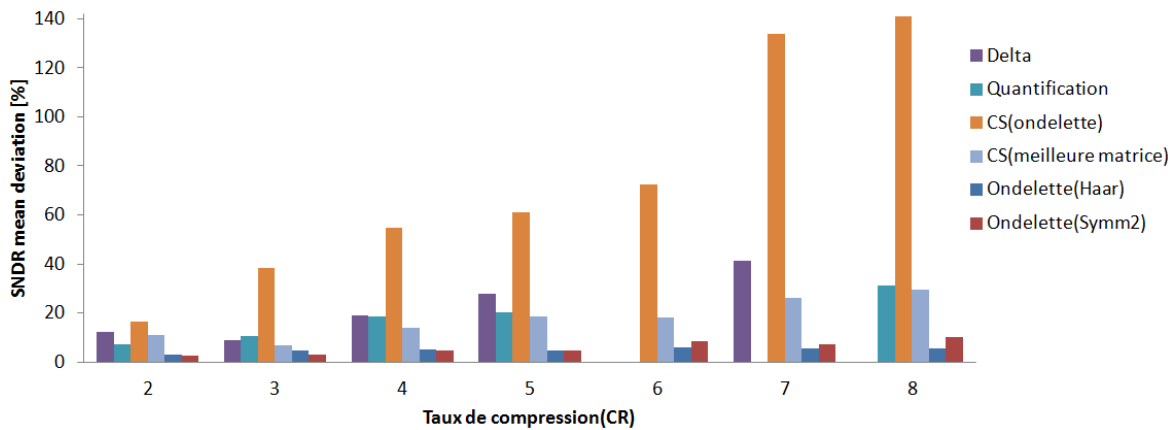


FIGURE B.9 – Déviation moyenne du SNDR obtenu avec tous les algorithmes testés pour un CR de 2 à 8.

Classification des PA et déviation moyenne

Une méthode de comparaison consiste à comparer le nombre de *clusters* issus des signaux avant et après compression. La méthode proposée consiste à identifier le nombre de *clusters* en effectuant une PCA suivie de l'algorithme de *clustering* nommé *mean shift clustering* [166]. L'algorithme *mean shift clustering* a été choisi puisqu'il détermine automatiquement le nombre de *clusters* sans intervention externe. Cette caractéristique est utile lorsque le nombre de *clusters* est inconnu au départ, ce qui est le cas des signaux compressés qui peuvent générer un nombre indéterminé de *clusters* selon la qualité de la compression. Trois itérations de *clustering* sont effectuées, la première étant sur les PAs originaux. La seconde itération effectue un *clustering* sur les PAs compressés par chacun des algorithmes de compression. La dernière itération effectue un *clustering* sur les PAs compressés par chacun des algorithmes joints aux PAs originaux. Ces deux dernières étapes de *clustering* permettent de vérifier si la

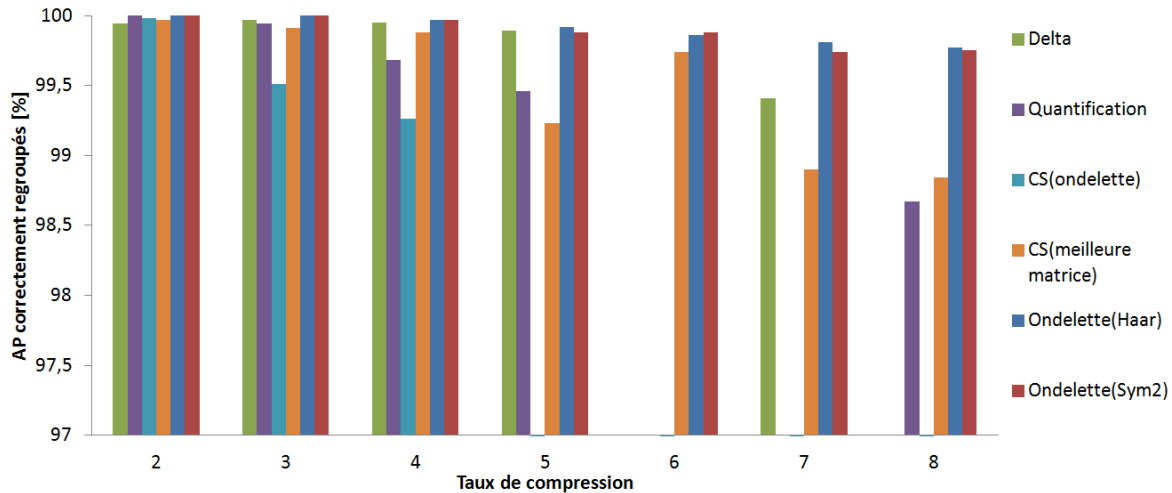


FIGURE B.10 – Pourcentage des PAs compressés dans le bon *cluster* avec PCA effectuée sur les PAs compressés joint aux PAs originaux.

compression a mené à des divergences dans la quantité de *clusters* trouvée par l’algorithme *mean shift clustering*.

Pour les deux tests de classification effectués (figures B.10 et B.11), les algorithmes de compression à base de l’ondelette *Symmet-2* et *Haar* permettent d’obtenir les meilleurs résultats pour presque tous les CR. Les performances de l’algorithme *Delta* et à faible quantification se dégradent progressivement dans les tests de la figure B.10 et sont stables dans ceux de la figure B.11. Les résultats de compression de l’algorithme CS avec parcimonie dans le domaine des ondelettes subissent une grande variation dans les deux tests, idem pour ceux avec la matrice de parcimonie optimale. La grande variabilité des résultats obtenus par l’algorithme CS, avec les deux bases parcimonieuses, peut être expliquée par le fait que certains PA ne sont pas bien représentés par la base parcimonieuse utilisée, car certains PA sont reconstruits avec fidélité alors que d’autres ne le sont pas. Ces résultats laissent supposer que l’algorithme de compression CS n’est pas un algorithme de compression approprié pour compresser les PAs et pour ensuite faire du *clustering*, ce qui concorde avec les conclusions présentées dans [167].

B.2.6 Variation du nombre de *clusters* détectés

Tous les algorithmes, à l’exception de CS, ont permis de détecter le bon nombre de *clusters* pour tous les CR. Lors des tests de *clustering* sur les PAs compressés seulement (figure B.12), l’algorithme CS avec parcimonie dans le domaine des ondelettes a généré des variations dans le nombre de *clusters* détectés pour un $CR \geq 5$, alors que l’algorithme CS avec matrice de parcimonie optimale présente des variations pour un $CR \geq 6$, allant même jusqu’à produire un *cluster* supplémentaire pour un CR de 8. Les résultats sont similaires pour les tests

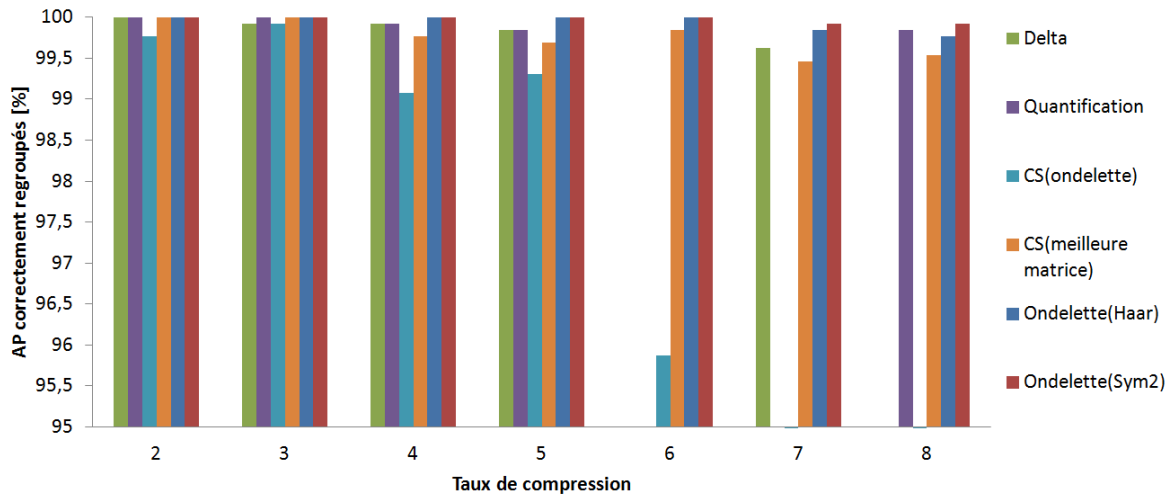


FIGURE B.11 – Pourcentage des PAs compressés dans le bon *cluster* avec PCA effectuée en ajoutant les PAs compressés un à la fois aux PAs originaux.

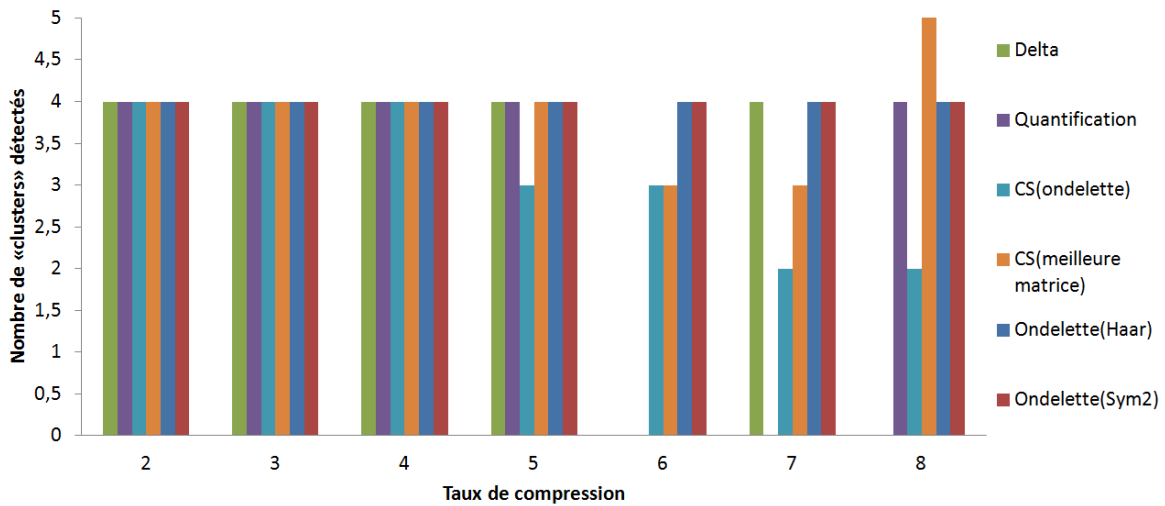


FIGURE B.12 – Nombre de *clusters* trouvés après compression.

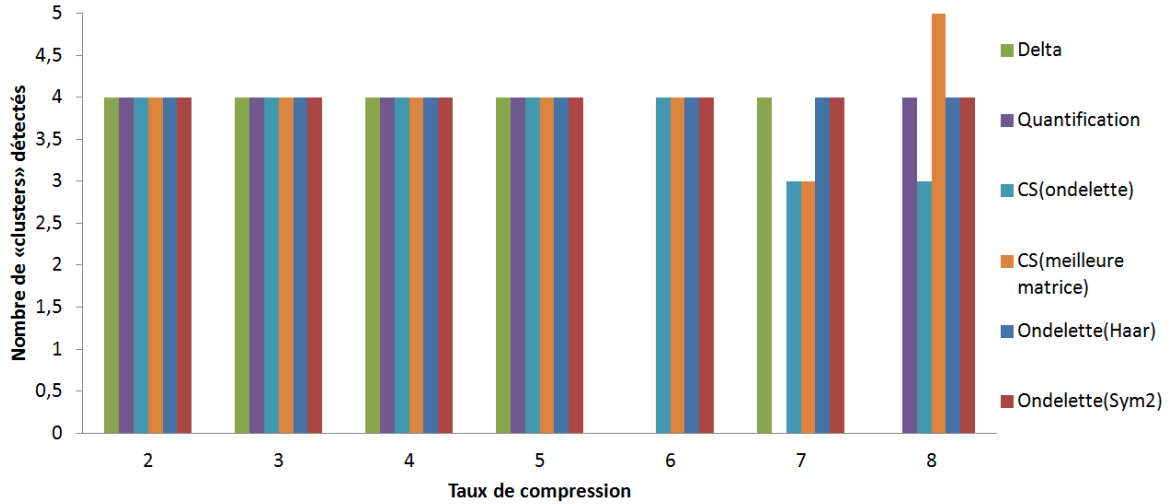


FIGURE B.13 – Nombre de *clusters* trouvés avec les PAs compressés joints aux PAs originaux.

de *clustering* avec les PAs compressés joints aux PAs originaux (figure B.13), à l'exception que des variations se produisent pour des $CR \geq 7$.

Fonction de coûts

Afin de déterminer l'algorithme de compression optimal, une fonction de coût a été créée. Premièrement, un score est calculé pour chacun des taux de compression. Ce score est pondéré à 50% en fonction des tests de SNDR et à 50% en fonction des résultats de *clustering*. Le score est construit en fonction des critères suivants :

1. SNDR :

Les critères du SNDR moyen, minimum, maximum et la déviation moyenne du SNDR se voient attribuer 12.5% du score chacun. Les fonctions $SNDR_{moy} / SNDR_{moy_max}$, $SNDR_{min} / SNDR_{min_max}$ et $SNDR_{max} / SNDR_{max_max}$ sont conçus afin de donner une pondération de 1 aux algorithmes générant les meilleurs résultats alors que la fonction $(1 - SNDR_{dev} / SNDR_{dev_max})$ est conçue afin de donner une pondération de 0 à l'algorithme générant la plus grande la déviation moyenne. Dans ces formules, les termes $SNDR_{xxx_max}$ sont les valeurs maximales du critère $SNDR_{xxx}$ obtenues parmi tous les algorithmes pour un même CR. Si la valeur de $SNDR_{xxx}$ est négative, le critère correspondant se voit accorder une pondération de 0.

2. Résultats de *clustering* :

Le ratio de PA compressés et associés dans le bon *cluster* se voit attribuer 33.3% du score à partir de la fonction $e^{-10 \times (1 - CL_{comp})} + e^{-10 \times (1 - CL_{joint})}$, où CL_{comp} est le ratio des PAs compressés dans le bon *cluster* (test de la figure B.10), CL_{joint}^i est le ratio des PAs compressés dans le bon *cluster* en les ajoutant un à la fois aux potentiels originaux (test de la figure B.11). Ces fonctions sont exponentielles avec un facteur multiplicatif de 10

afin de pénaliser sévèrement les algorithmes ayant mené à de mauvaises associations et afin de donner une note de 1 lorsque toutes les associations sont bonnes.

Le critère du nombre de *clusters* trouvés se voit attribuer 16.7% du score à partir de la fonction $(4 - |4 - NBCL_{comp}|)/8 + (4 - |4 - NBCL_{joint}|)/8$, où $NBCL_{comp}$ est le nombre de *clusters* trouvés après compression (test de la figure B.12) et $NBCL_{joint}$ est le nombre de *clusters* trouvés avec les PAs compressés joints aux originaux (test de la figure B.13). Cette fonction génère une pondération de 1 pour les algorithmes ayant trouvé 4 *clusters* dans les deux tests en plus de pénaliser de 0.125 point pour chaque *cluster* manquant ou supplémentaire trouvé.

Le score est défini par l'équation suivante :

$$\begin{aligned}
Score_{CR} = & \left(SNDR_{min}/SNDR_{min_max} + SNDR_{moy}/SNDR_{moy_max} \right. \\
& + SNDR_{max}/SNDR_{max_max} + (1 - SNDR_{dev}/SNDR_{dev_max}) + e^{-10 \times (1 - CL_{comp})} \quad (B.16) \\
& \left. + e^{-10 \times (1 - CL_{joint})} + (4 - |4 - NBCL_{comp}|)/8 + (4 - |4 - NBCL_{joint}|)/8 \right) / 6
\end{aligned}$$

Le tableau B.1 présente les scores calculés pour tous les algorithmes. Puisque les algorithmes *Delta* et à faible quantification ne peuvent pas produire tous les CR, les valeurs impossibles à réaliser sont présentées en rouge dans le tableau et sont les mêmes que ceux obtenus avec le CR le plus petit et le plus proche réalisable.

La fonction de coût est calculée avec l'équation B.17 à partir des scores du tableau B.1 et donne une pondération plus grande aux CR plus élevés. Les résultats de la fonction de coût sont présentés dans le tableau B.2.

$$\begin{aligned}
Coût = & (0.1 \times Score_2 + 0.2 \times Score_3 + 0.3 \times Score_4 + 0.4 \times Score_5 \\
& + 0.5 \times Score_6 + 0.6 \times Score_7 + 0.7 \times Score_7) / 2.8 \quad (B.17)
\end{aligned}$$

Conclusion des simulations

Selon les résultats obtenus avec la fonction de coût, il est possible de conclure que l'algorithme de compression à base d'ondelettes de type *Symmlet-2* est le meilleur algorithme identifié afin de compresser les PAs, suivi de l'algorithme à base d'ondelette de type *Haar*. De façon générale, l'algorithme *Delta* et à faible quantification ont généré des résultats moyens, presque toujours en deçà de ceux obtenus avec la transformée en ondelettes de type *Symmlet-2* et/ou de type *Haar*. Bien que l'algorithme de compression CS avec une base parcimonieuse optimale ait généré de bons résultats dans certaines circonstances, il s'agit d'un algorithme

TABLE B.1 – Score calculé pour chacun des algorithmes de compression en fonction d’un CR de 2 à 8.

CR	<i>Haar</i>	<i>Symmlet-2</i>	<i>Delta</i>	<i>DNBQ</i>	<i>CSWD</i>	<i>CSOP</i>
2	0.930	0.942	0.902	0.921	0.624	0.841
3	0.819	0.948	0.924	0.831	0.506	0.903
4	0.897	0.941	0.920	0.694	0.494	0.863
5	0.906	0.933	0.892	0.617	0.328	0.810
6	0.873	0.922	0.892	0.617	0.274	0.793
7	0.877	0.911	0.776	0.617	0.150	0.753
8	0.895	0.914	0.776	0.488	0.162	0.732

TABLE B.2 – Fonction de coût calculée pour chacun des algorithmes de compression.

	<i>Haar</i>	<i>Symmlet-2</i>	<i>Delta</i>	<i>DNBQ</i>	<i>CSWD</i>	<i>CSOP</i>
<i>Coût</i>	0.890	0.924	0.844	0.619	0.280	0.789

extrêmement gourmand en calcul et qui génère une grande variabilité des résultats c.-à-d. on ne peut pas prédire avec certitude la qualité de la compression pour tous les PAs. C’est la même chose pour l’algorithme CS avec base parcimonieuse dans le domaine des ondelettes qui a généré les pires résultats dans presque tous les tests.

B.3 Algorithme de classement

B.3.1 Réduction de la dimensionnalité

La technique de regroupement retenue est basée sur une transformée en ondelette de type *Symmlet-2*. Puisque cette transformée est déjà utilisée par l’algorithme de compression [20], [168], [169], [170], cette dernière n’a pas à être ré-implémentée, ce qui est un atout majeur qui a influencé le choix de cette méthode. De plus, comme il le sera démontré à la sous-section B.3.1, des tests comparatifs ont permis de démontrer que la technique proposée génère des résultats meilleurs ou similaires que des techniques beaucoup plus complexes.

Il est démontré dans [143] qu’une DWT permet de mettre en évidence les différences existant entre différentes formes d’ondes, ce qui permet de réduire la dimensionnalité des PA avant la phase de regroupement. Or, il faut choisir judicieusement le nombre et les coefficients à retenir afin d’obtenir une classification optimale impliquant le moins de calcul possible. Malheureusement, les techniques de sélection basées sur la déviation maximale à la normalité [171], comme le *Lilliefors test for normality* [143], nécessitent une phase d’entraînement et beaucoup de calculs. Pour un système sur puce autonome, ces techniques ne peuvent pas

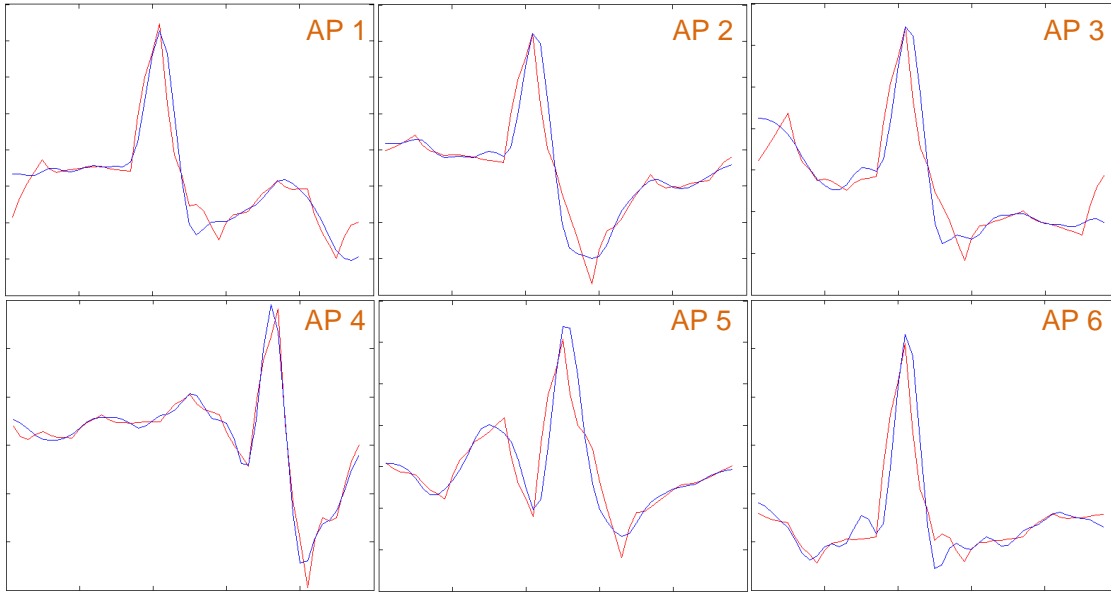


FIGURE B.14 – Représentation de 6 PAs reconstruits en utilisant seulement les 12 coefficients (12 sur 48) produits par les niveaux 3 et 4 de la DWT. Les formes d’ondes reconstruites (en rouge) ont conservé la forme générale des PAs originaux (en bleu).

être implémentées sans l’utilisation de beaucoup de mémoire et ressources matérielles et temporelles, sans compter que la phase d’entraînement doit être répétée à intervalle régulier afin de garantir l’optimalité des coefficients choisis.

Il est possible de démontrer que les détails d’un signal traité par une DWT se retrouvent principalement dans les coefficients des premiers niveaux de la transformée. Entre autres, la transformée de type *Symmlet-2* peut être calculée par des étapes de convolution et de décimation en utilisant les filtres suivants :

$$h(z) = h_5(H_2 + (H_1H_2 - 1)z^{-1} + (1 - H_3)z^{-2} + H_1(1 - H_3)z^{-3}) \quad (\text{B.18})$$

$$g(z) = H_6(-H_2 + (1 - H_1H_2)z^{-1} + H_3z^{-2} + H_1H_3z^{-3}), \quad (\text{B.19})$$

où $h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}$, $h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}$, $h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}$ et $h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$. Or, $h(z)$ et $g(z)$ forment des filtres passe-bas et passe-haut respectivement, avec $h(z)$ générant les coefficients approximatifs de la transformée et $g(z)$ les coefficients détaillés. Les coefficients approximatifs décimés d’un facteur 2 sont ensuite passés au prochain niveau, qui effectue les mêmes étapes de filtrage et de décimation et ainsi de suite jusqu’au dernier niveau de la DWT, obtenant ainsi une analyse à multirésolution. Pour ainsi dire, les coefficients approximatifs représentent une version comprimée du signal approximatif précédent alors que les coefficients détaillés représentent une version comprimée des détails du signal approximatif précédent. Or, plus la transformée progresse dans les niveaux, plus les coefficients détaillés expriment les détails importants représentant la forme d’onde, car ces derniers sont puisés d’une approximation

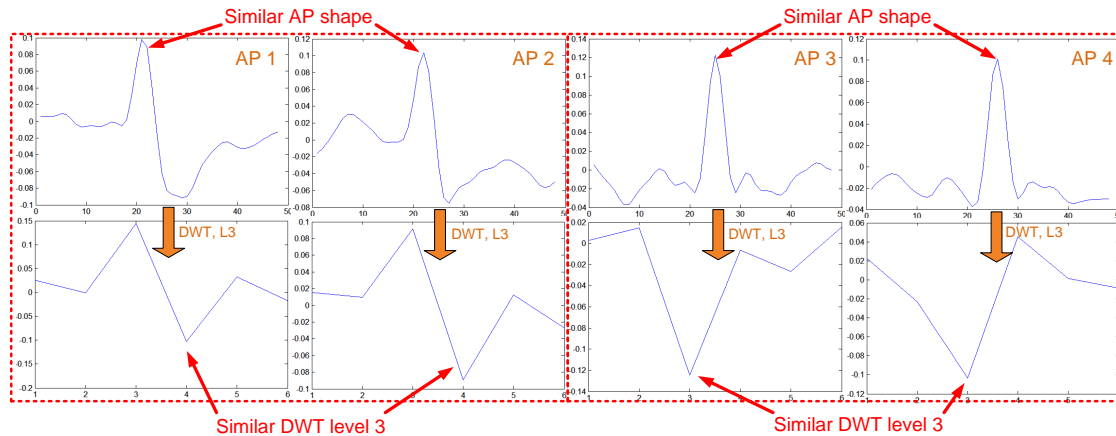


FIGURE B.15 – Représentation de 4 PAs et de leurs 6 coefficients du 3e niveau de leur DWT respective. Deux paires de PA similaires ont généré des coefficients similaires, ce qui est généralement le cas pour des formes d’ondes visuellement proches.

de plus en plus grossière du signal original. Cette caractéristique est démontrée à la figure B.14, où 4 PAs (en bleu) ont été reconstruits (en rouge) en utilisant seulement les coefficients des niveaux 3 et 4 (12 coefficients sur 48). Bien que les PAs reconstruits diffèrent quelque peu de leurs originaux, la forme d’onde générale de chacun des PAs a été conservée par les coefficients des niveaux 3 et 4 alors que certains détails ont disparu.

Ce dernier point est très important, car il signifie que des formes d’ondes similaires vont générer des coefficients similaires dans les derniers niveaux de la DWT, car leurs approximations vont se ressembler. Ce fait peut être constaté visuellement à la Figure B.15, où 4 PA sont affichés conjointement à leurs coefficients du 3e niveau suite à une DWT de type *Symmllet-2*. On peut voir dans cette figure que les formes d’onde AP_1 et AP_2 , qui ont une forme temporelle similaire, ont générés des vecteurs de coefficients du 3e niveau très similaires, de même que pour les formes d’onde AP_3 et AP_4 .

La stratégie de réduction de la dimensionnalité retenue consiste à conserver les 6 coefficients du 3e niveau de la DWT de type *Symmllet-2*. En plus, le module de compression effectue deux autres étapes importantes : il normalise et requantifie les coefficients sur 6 bits. Ceci est un atout supplémentaire, car cela permet de réduire la complexité et la quantité de logique qui est utilisée par l’algorithme de regroupement. Au lieu de traiter des coefficients de 16 bits, l’algorithme traite des coefficients de 6 bits, soit d’une résolution 2.7 fois moindre, et sans compromis sur la qualité des résultats. Les performances de réduction de la dimensionnalité en fonction des coefficients choisis et de leur quantification sont présentées à la sous-section B.3.1. Des tests comparatifs, notamment avec le même algorithme utilisant une DWT de type *haar* et avec les algorithmes DD et PCA, permettent de démontrer que la technique proposée génère des résultats similaires à PCA, meilleurs que *haar* et DD, tout en étant beaucoup

moins complexe, car la DWT est déjà calculée.

Résultats de simulation

Les performances de l'algorithme de réduction de la dimensionnalité ont été obtenues via des simulations *Matlab*. Quatre algorithmes ont été retenus dans le but d'effectuer des tests comparatifs : la DWT de type *Symmlet-2* (4 niveaux de profondeur) [20], la DWT de type *Haar* (4 niveaux de profondeur) [143], [155] et les algorithmes DD [143], [172], [173] et PCA [143], [174]. Afin d'identifier les niveaux optimaux de la DWT, les tests ont été effectués en utilisant les niveaux 2, 3, 4 et 3-4. Dans le cas de l'algorithme DD, 10 coefficients sont sélectionnés via le *Lilliefors test for normality* [143] et pour la PCA les trois premières composantes sont retenues. Il est important de spécifier que l'algorithme *Lilliefors test for normality* nécessite une phase d'entraînement afin de sélectionner les coefficients optimaux, ce qui est un désavantage majeur. De plus, l'algorithme PCA est beaucoup trop gourmand en calcul pour une implémentation matérielle et sert de point de comparaison seulement. Pour ce qui est de l'algorithme DWT de type *Haar*, il a été implémenté afin de comparer les performances de la méthode proposée avec la DWT la plus simple et la plus répandue. L'algorithme de regroupement utilisé pour les tests est *K-means*.

Le Tableau B.3 présente les résultats de *clustering* obtenus avec 12 signaux neuronaux synthétiques issus de la banque en ligne [175] de l'Université Leicester. Ces résultats sont présentés visuellement dans l'histogramme de la Figure B.16. On peut constater que les résultats varient d'un signal à l'autre et que certains algorithmes performant mieux dans presque toutes les situations. Un système de point est attribué en fonction des performances de classification. L'algorithme présentant les meilleures performances pour un signal donné se voit attribuer 3 points (rouge), le second 2 points (bleu) et le troisième 1 point (vert). L'algorithme ayant récolté le plus de points est sans surprise l'algorithme PCA, suivi de la DWT *Symmlet-2* de niveau 3 (6 coefficients) et finalement la DWT *Symmlet-2* de niveaux 3-4 (12 coefficients). Or, la méthode proposée a dépassé les performances de PCA dans 6 signaux, soit la moitié, et suit de très près dans les autres signaux. De plus, pour les signaux ayant présenté les pires résultats pour tous les algorithmes (*C_Difficult1_noise02* et *C_Difficult1_noise015*), c'est la méthode proposée qui performe le mieux.

Afin de quantifier l'impact de la quantification sur 6 bits des coefficients de la DWT, les mêmes tests ont été effectués en conservant une résolution de 16 bits sur ces derniers (*Symmlet-2* et *Haar*). Les résultats sont présentés dans le Tableau B.4 et indiquent que la réduction de la quantification n'a pas un impact significatif sur les résultats.

B.3.2 Stratégie de regroupement non supervisée

La stratégie de regroupement non supervisée implémentée est une variante de celle décrite dans [176] et [177]. Cette technique est présentée à la Figure B.17, et diffère de [176] et [177]

TABLE B.3 – Performance de *clustering* des algorithmes à base de DWT (*Symmlet-2* et *Haar*), DD et PCA avec 12 signaux neuronaux différents. Les algorithmes basés sur la DWT comparent différents sous-niveaux, du niveau 2 au niveau 4, et jonction de niveaux (3-4) et les coefficients sont requantifiés sur 6 bits après maximisation sur la plage dynamique. Un système de point est attribué pour le premier (rouge + 3 pts), le second (bleu + 2 pts) et le troisième (verts + 1 pts) algorithme en fonction des résultats de classification.

Signal	Nb PA	dB2 Level 2 [%]	dB2 Level 3 [%]	dB2 Level 4 [%]	dB2 Level 3-4 [%]	Haar Level 2 [%]	Haar Level 3 [%]	Haar Level 4 [%]	DD [%]	PCA [%]
C_Difficult1_noise02	3414	60.9	65.4	38.9	45.9	50.8	56.6	43.0	58.0	53.4
C_Difficult1_noise015	3472	63.3	74.2	43.2	53.3	52.6	52.7	48.4	36.2	69.6
C_Difficult1_noise01	3448	65.5	82.7	47.0	81.8	52.6	73.5	66.0	96.4	90.6
C_Easy1_noise04	3386	89.4	85.7	75.5	92.0	89.6	80.9	67.5	71.1	94.8
C_Difficult1_noise005	3383	65.8	88.9	90.0	94.9	48.6	80.2	93.4	77.5	97.3
C_Easy1_noise03	3475	91.5	93.8	90.4	96.0	91.1	86.7	89.6	97.5	98.1
C_Difficult2_noise02	3493	70.2	95.7	66.3	79.1	99.6	71.9	73.7	90.3	75.8
C_Easy1_noise02	3474	92.9	97.1	96.7	98.5	91.8	93.0	97.1	99.2	99.1
C_Difficult2_noise015	3440	90.9	97.5	75.7	95.1	99.5	79.4	75.8	89.2	86.0
C_Difficult2_noise01	3462	93.0	98.7	89.3	98.5	99.7	88.4	82.4	98.2	97.1
C_Difficult2_noise005	3364	93.1	98.7	97.0	98.6	99.5	94.4	92.6	97.5	98.6
C_Easy1_noise01	3522	93.5	98.4	99.0	99.3	92.8	98.4	99.0	99.2	99.5
Points	—	2	16	0	12	10	1	2	11	19

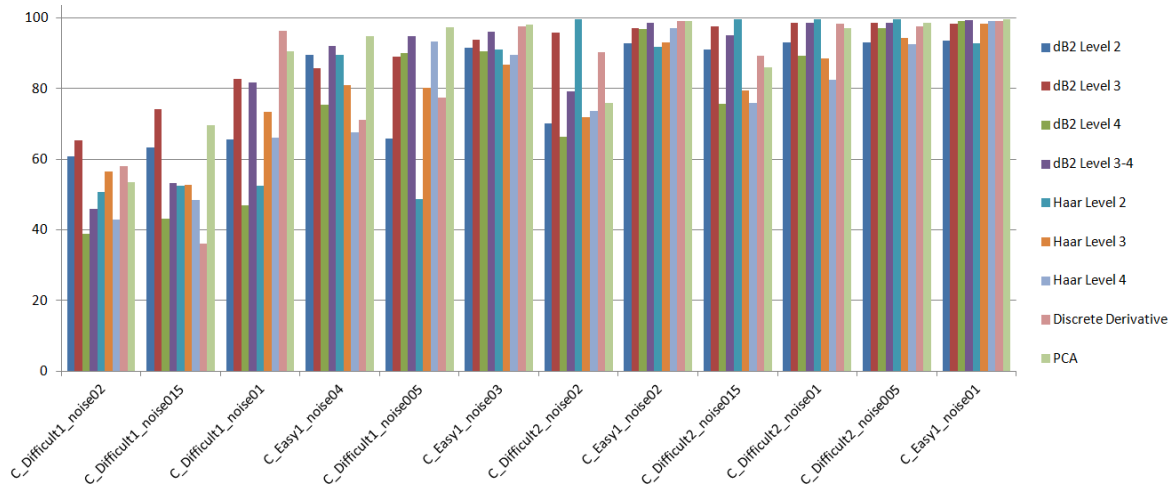


FIGURE B.16 – Histogramme des résultats répertoriés dans le Tableau B.3.

par l'utilisation des 6 coefficients du niveau 3 de la DWT comme vecteur d'entrée à l'algorithme au lieu des 48 échantillons temporels de chaque PA ($21 \times$ moins de mémoire requises). De plus, cet algorithme nécessite l'utilisation de deux seuils ($Thr1$ et $Thr2$) qui sont déterminés en fonction de l'écart-type de bruit contenu dans le signal neuronal, tel que décrit au Chapitre 4. L'algorithme s'exécute comme suit :

1. Calcul de la norme L_1 entre le vecteur d'entrée et le vecteur moyen de tous les *clusters*.

TABLE B.4 – Performance de *clustering* des algorithmes à base de DWT (*Symmetlet-2* et *Haar*), DD et PCA avec 12 signaux neuronaux différents. Les algorithmes basés sur la DWT comparent différents sous-niveaux, du niveau 2 au niveau 4, et jonction de niveaux (3-4) et les coefficients sont conservés avec une précision de 16 bits. Un système de point est attribué pour le premier (rouge + 3 pts), le second (bleu + 2 pts) et le troisième (verts + 1 pts) algorithme en fonction des résultats de classification.

Signal	Nb PA	dB2 Level 2 [%]	dB2 Level 3 [%]	dB2 Level 4 [%]	dB2 Level 3-4 [%]	Haar Level 2 [%]	Haar Level 3 [%]	Haar Level 4 [%]	DD [%]	PCA [%]
C_Difficult1_noise02	3414	61.1	65.7	38.9	45.6	50.4	56.6	42.7	58	53.4
C_Difficult1_noise015	3472	63.7	74.2	43.2	52.9	52.5	52.5	48.3	36.2	69.6
C_Difficult1_noise01	3448	65.6	82.8	46.7	82.1	52.6	73.1	74	96.4	90.6
C_Easy1_noise04	3386	89.3	85.7	74.7	91.6	89.6	81.3	67.7	71.1	94.8
C_Difficult1_noise005	3383	65.4	89.2	90	94.8	50.1	80.6	93.7	77.5	97.3
C_Easy1_noise03	3475	91.8	93.9	90.5	96	91.1	86.6	89.7	97.5	98.1
C_Difficult2_noise02	3493	70.1	95.7	66.4	80	99.7	71.6	73.8	90.3	75.8
C_Easy1_noise02	3474	92.5	97	97	98.5	91.9	92.9	97.2	99.2	99.1
C_Difficult2_noise015	3440	91.4	97.4	76.3	94.4	99.5	79.5	75.6	89.2	86
C_Difficult2_noise01	3462	92.7	98.7	98.6	98.4	99.6	88.5	82.3	98.2	97.1
C_Difficult2_noise005	3364	93.3	98.8	97.2	98.6	99.6	94.2	92.7	97.5	98.6
C_Easy1_noise01	3522	93.3	98.4	99	99.3	92.6	98.3	99	99.2	99.5
Points	—	0	15	1	13	13	0	1	12	18

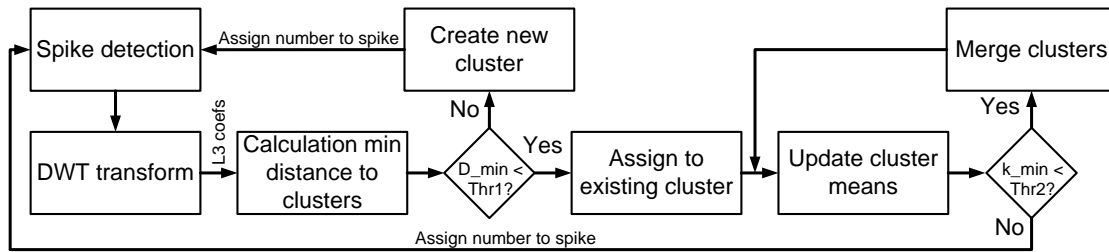


FIGURE B.17 – Stratégie de regroupement non supervisée utilisant les coefficients du 3e niveau de la DWT comme vecteur d’entrée.

2. Si la norme minimale est inférieure au seuil $Thr1$, un nouveau *cluster* est créé avec comme moyenne le vecteur d’entrée. Sinon, le vecteur est associé au *cluster* dont la norme est minimale.
3. Dans le second cas du point précédent, la moyenne du *cluster* est mise à jour en utilisant une moyenne glissante.
4. Ensuite, la norme L_1 est calculée entre le vecteur moyen du *cluster* mis à jour et les vecteurs moyens de tous les autres *clusters*.
5. Si la norme minimale est inférieure au seuil $Thr2$, les *clusters* sont fusionnés.
6. Mise à jour du vecteur moyen du *cluster* fusionné. Retour au point 5.
7. L’association d’un ‘PA à un *cluster* est terminée.

B.4 SoC : Stimulateur optique en technologie CMOS

B.4.1 Revue de littérature

Quelques stimulateurs optiques en technologie CMOS ont été répertoriés dans la littérature. Entre autres, le système présenté dans [119] a été implémenté en technologie $0.8\ \mu\text{m}$ et est capable de fournir en courant deux LED avec $51.2\ \text{mA}$, ou bien une seule LED avec $102.4\ \text{mA}$. Ce système est conçu pour générer des ondes de type PWM de fréquence variant entre $0.153\ \text{Hz}$ à $200\ \text{Hz}$ et avec un *duty-cycle* variant entre $100\ \mu\text{s}$ et $10\ \text{ms}$. Essentiellement, la méthode employée contrôle les courants avec plusieurs miroirs de courants connectés en parallèle. Le courant maximum de référence est de $5\ \mu\text{A}$, alors que le gain en courant est déterminé par un ratio fixé à 40 entre deux résistances. En utilisant 512 miroirs de courants, ils peuvent fournir un maximum de $102.4\ \text{mA}$ pour les deux canaux. Cette méthode possède un désavantage évident, il n'y a aucune rétroaction sur le courant réel circulant dans les LED. Ainsi, la puissance optique fournie aux neurones ciblés ne peut être garantie et est grandement dépendante des variations du procédé et du type de LED utilisé.

Un autre système CMOS présenté dans [178] a été implémenté en technologie $0.35\ \mu\text{m}$. Le contrôle du courant est effectué en utilisant un DAC de 8 bits connecté sur la *gate* d'un transistor PMOS servant de miroir de courant. Le courant est ensuite injecté dans deux branches d'un pont en H, dont les quatre NMOS le constituant sont contrôlés via un module logique de contrôle. Tout comme pour le système précédent, aucune rétroaction sur le courant réel circulant dans les LED n'est effectuée par le circuit.

Une méthode exclusivement basée sur des *switch caps* qui est implémentée en technologie CMOS $0.35\ \mu\text{m}$ est présentée dans [127]. Ce système à 4 canaux peut générer des pulses de voltage aux bornes des LED de l'ordre de $3.45\ \text{V}$. Malheureusement, de tels voltages ne sont pas constants et nécessitent d'énormes condensateurs externes ($1\ \mu\text{F}$ - $10\ \mu\text{F}$) afin de stocker suffisamment de charges. Par contre, en utilisant des condensateurs externes haute précision ($< 1\%$), il est possible de déterminer le nombre de charges stockées et délivrées aux LED avec une assez bonne précision (en connaissant la période de charge et de décharge). Le désavantage majeur de cette approche concerne la non-constance de la puissance optique délivrée, qui se comporte comme un circuit RC.

Une approche très simple implémentée en technologie CMOS $0.35\ \mu\text{m}$ est présentée dans [128]. Essentiellement, le courant est fourni par une source de courant PMOS, sans rétroaction. La *gate* du PMOS est soit au niveau VDD lorsque le courant circule, ou bien à VSS lorsque le courant est coupé. Il n'y a aucune rétroaction ni aucun contrôle du courant circulant dans les LED avec un tel système.

Aucun stimulateur optogénétique intégré possédant une rétroaction directe sur le courant réel circulant dans les LED n'a été identifié, ce qui constitue un problème majeur afin de

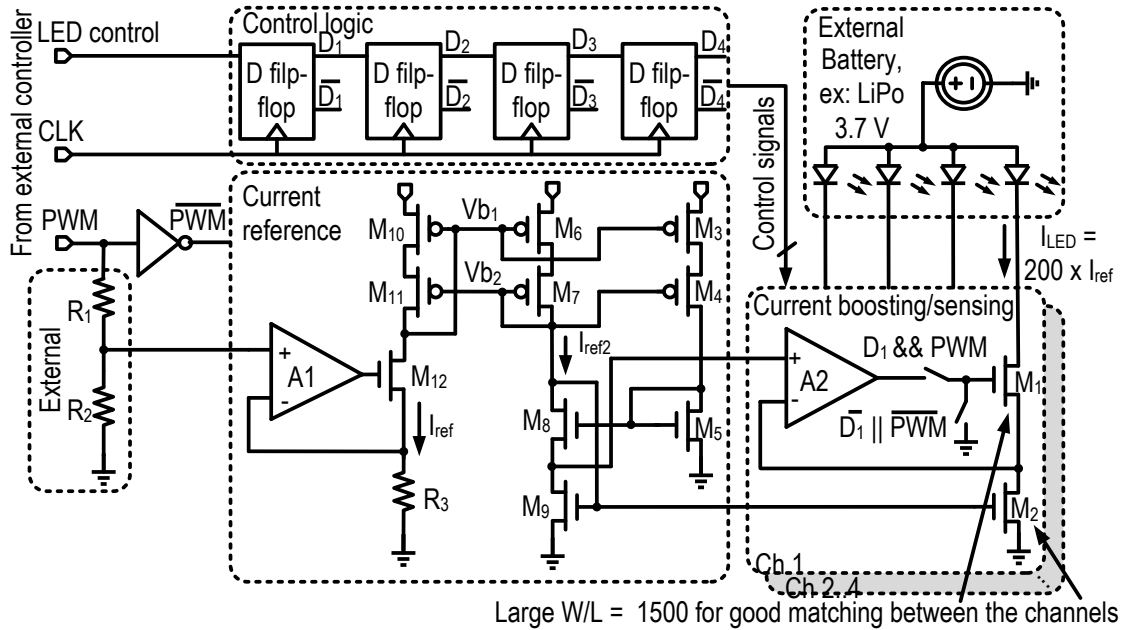


FIGURE B.18 – Schémas du circuit de stimulation optique.

garantir la puissance optique délivrée aux neurones ciblés. Entre autres, [119], [178] et [128] utilisent des sources de courant dont le courant peut varier en fonction du procédé et du type de LED connecté c.-à-d. en fonction de la chute de tension nominale à ses bornes.

B.4.2 Solution proposée

Le stimulateur optique proposé possède 4 canaux de stimulation et est présenté à la Figure B.18. Notamment, ce circuit diffère des solutions précédentes, voir section B.4.1, en offrant une rétroaction directe sur le courant réel circulant dans chacune des LED. La solution proposée est détaillée aux points suivants :

1. Une source de courant précise basée sur un *opamp* génère un courant de référence, identifié par I_{ref} dans la Figure B.18 [119]. Le courant I_{ref} est contrôlé par une tension qui est reproduite par l'*opamp* «A1» à la borne d'une résistance haute précision externe à la puce.
2. Le courant I_{ref} est reproduit avec un gain de 2 par une source de courant PMOS de type *wide-swing* pour devenir le courant I_{ref2} dans la Figure B.18.
3. Le courant I_{ref2} est reproduit avec un gain de 100 par une source de courant NMOS de type *wide-swing* pour devenir le courant I_{LED} dans la Figure B.18.
4. Un second *opamp* («A2») contrôle la *gate* du transistor «M1» en fonction de la tension du *drain* du transistor «M2» dans le suiveur de courant *wide-swing* afin que le courant soit reproduit avec fidélité dans les LED.

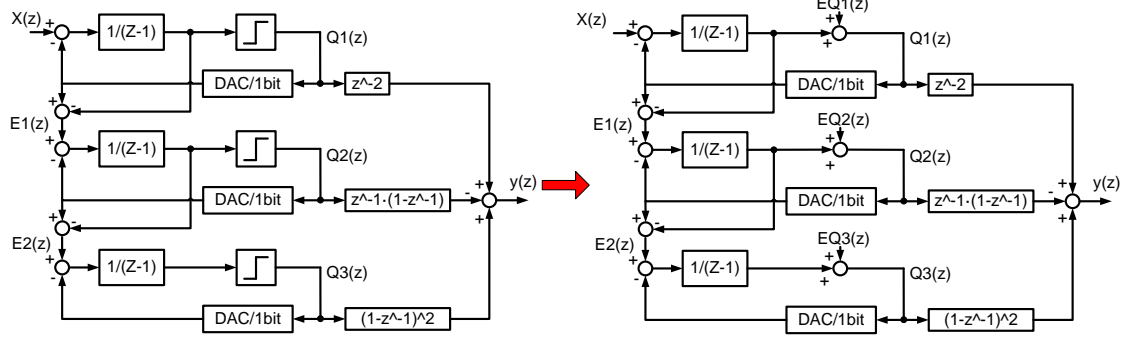


FIGURE B.19 – Concept de l'ADC $\Delta\Sigma$ MASH du 3^e ordre.

5. La source de courant générant I_{ref} est contrôlée par un diviseur de tension externe et par un signal de type PWM.
6. Lorsque le PWM est actif, I_{ref} est fonction des valeurs des résistances dans le diviseur de tension. Lorsque le PWM est inactif, I_{ref} est à 0 et la *gate* de «Q1» est mise à VSS afin de s'assurer qu'aucun courant résiduel ne circule dans les LED (les *opamps* ne sont pas *rail-to-rail*). Puisque $I_{LED} = 0$ à ce moment, le drain de «Q1» est en haute impédance.
7. Il y a un seul circuit générateur des courants I_{ref} et I_{ref2} qui est relié à 4 circuits générant les courants dans les LED. Chacun des 4 circuits peut-être activés/désactivés indépendamment des autres.
8. Un registre à décalage sériel de 4 bits mémorise l'état d'activation/désactivation de chacun des 4 canaux (contrôlé par «LED control» et «CLK»). Les sorties des bascules D contrôlent l'état d'activation de leur circuit respectif. Si la sortie d'une bascule est à 0, alors la *gate* de «M1» est mise à VSS et le courant de biais de «A2» est annulé.

Le circuit de stimulation de la Figure B.18 a été réalisé et implémenté en technologie 0.13 μm . Les ratios des transistors ont été ajustés afin de fournir des courants précis jusqu'à 30 mA. Le type d'*opamp* utilisé est de type *wide-swing folded cascode*, et a été choisit pour son gain élevé, sa faible consommation, son grand *output swing* et son bon *input common-mode range* [179].

Les avantages du circuit proposé par rapport aux solutions existantes sont les suivants :

- Rétroaction en fonction du courant réel circulant dans les LED de stimulation optique.
- L'utilisation de suiveurs de courant de type *wide-swing* permet de fonctionner avec une faible tension d'alimentation de 1.2 V. Ainsi, le circuit proposé est fonctionnel avec des technologies de type *short-channel* qui fonctionnent avec des tensions d'alimentation très basses, ce qui n'est pas le cas des solutions présentées à la section B.4.1.

B.5 SoC : Convertisseur $\Delta\Sigma$ MASH du 3^e ordre

L'ADC implémenté est de type Sigma-Delta ($\Delta\Sigma$) du 3^e ordre de type MASH. Cette topologie a été sélectionnée, car elle permet d'obtenir :

1. Une précision qui peut être ajustée en fonction des besoins (résolution et consommation énergétique).
2. Une grande précision avec de faibles OSRs (un ENOB de ~ 10 bits avec un OSR de 50).
3. Une excellente linéarité par branche (conversion sur 1 bit seulement par branches).
4. Une stabilité pour toute la plage d'entrée (versus 70% de la plage pour un 2^{eme} ordre standard).
5. Permet de passer rapidement du domaine analogique au domaine numérique, tout en relaxant les contraintes sur le filtre anti-recouvrement spectral (caractéristique propre aux $\Delta\Sigma$ en général).

En effet, les contraintes posées sur le filtre d'anti-recouvrement spectral sont relaxées, due au sur échantillonnage qui repousse dans les plus hautes fréquences la fréquence de *Nyquist*. Ainsi, il n'est pas rare de voir des filtres passe-bas anti-recouvrement du premier ordre de type RC, puisque le filtrage est reporté dans le domaine numérique avec le filtre de décimation (voir section B.6). C'est d'ailleurs la stratégie employée dans la puce, où la fréquence de coupure de l'AFE est fixée à 7 kHz via un filtrage de type Gm-C du premier ordre. Le concept de l'ADC $\Delta\Sigma$ MASH du 3^e ordre est présenté à la Figure B.19. Cette topologie est construite à partir de modulateur Sigma-Delta du premier ordre, ce qui permet de garantir sa stabilité et chacune des branches supplémentaires annule le bruit de quantification de la branche précédente. La sortie de la première branche s'exprime comme suit :

$$Q_1(z) = x(z)z^{-1} + EQ_1(z)(1 - z^{-1}), \quad (\text{B.20})$$

où $x(z)$ est le signal d'entrée et $EQ_1(z)$ est le bruit de quantification induit par le comparateur de la première branche. On peut voir dans (B.20) que la première branche permet de moduler le bruit de quantification par un filtre du premier ordre. En prenant $EQ_1(z)$ et $EQ_2(z)$ comme entrée de la deuxième et de la troisième branche respectivement, les sorties $Q_2(z)$ et $Q_3(z)$ s'expriment comme suit :

$$Q_2(z) = EQ_1z^{-1} + EQ_2(z)(1 - z^{-1}) \quad (\text{B.21})$$

$$Q_3(z) = EQ_2z^{-1} + EQ_3(z)(1 - z^{-1}) \quad (\text{B.22})$$

Or, en multipliant $Q_1(z)$, $Q_2(z)$ et $Q_3(z)$ par z^{-2} , $-z^{-1}(1 - z^{-1})$ et $(1 - z^{-1})^2$ respectivement et en les additionnant, l'on obtient l'équation suivante :

$$\begin{aligned} y(z) &= Q_1(z)z^{-2} - Q_2(z)z^{-1}(1 - z^{-1}) + Q_3(z)(1 - z^{-1})^2 \\ &= x(z)z^{-3} + EQ_3(z)(1 - z^{-1})^3 \end{aligned} \quad (\text{B.23})$$

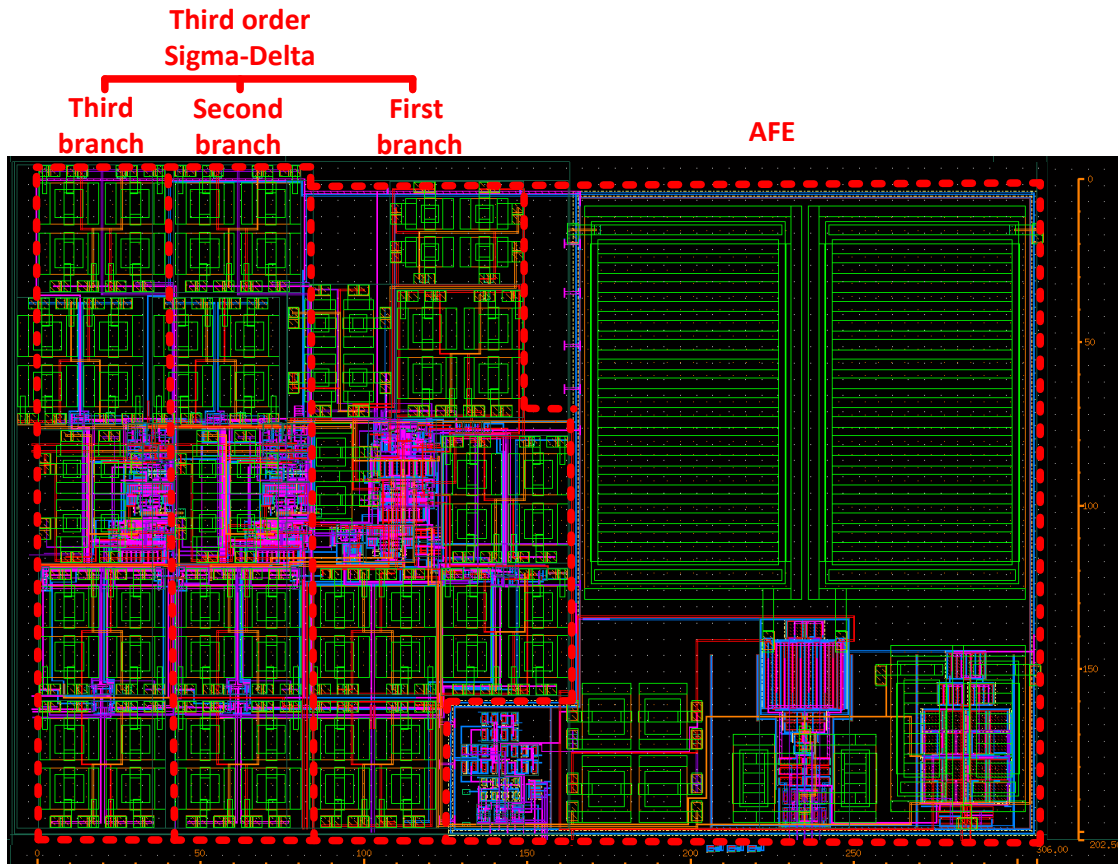


FIGURE B.20 – *Layout* de l’AFE amélioré et du $\Delta\Sigma$ MASH d’ordre 3. Ce circuit possède une aire totale de 0.056 mm².

Or, $y(z)$ correspond à $x(z)$ retardé de 3 auquel il faut ajouter le bruit de quantification de la troisième branche modulé par un filtre du 3^e ordre.

L’implémentation CMOS, 100% différentielle, de l’ADC est présentée à la Figure B.21 et son *layout* est présenté à la Figure B.20. Les fonctions d’intégrations ($\frac{1}{z-1}$) sont implémentées avec des intégrateurs de type *switch caps* différentiels avec suppression d’offset. Ces fonctions sont réalisées grâce aux condensateurs C1 et C3 qui conservent un ratio $\frac{C1}{C3} = \frac{80fF}{80fF} = 1$ afin d’avoir un gain d’intégration unitaire. Les condensateurs C2 permettent de stocker l’offset à l’entrée de l’OTA1 pendant la phase CLKp1 et de soustraire cet offset lors de la phase CLKp2 (phase d’intégration) [180]. Tous les commutateurs sont implémentés en utilisant des portes de transmissions de type NMOS-PMOS. L’OTA1 est de type *current mirror OTA* avec entrées PMOS et le circuit CMFB est de type *switch caps*. Le comparateur est de type *latch comparator* [181] afin de minimiser le temps de réponse et la consommation énergétique du comparateur (consomme seulement lors des commutations). Il est important de noter que le $\Delta\Sigma$ compense pour l’offset du comparateur. Étant donné que le comparateur maintient la valeur de comparaison seulement pendant une moitié de phase (quand CLKp1 est à son niveau

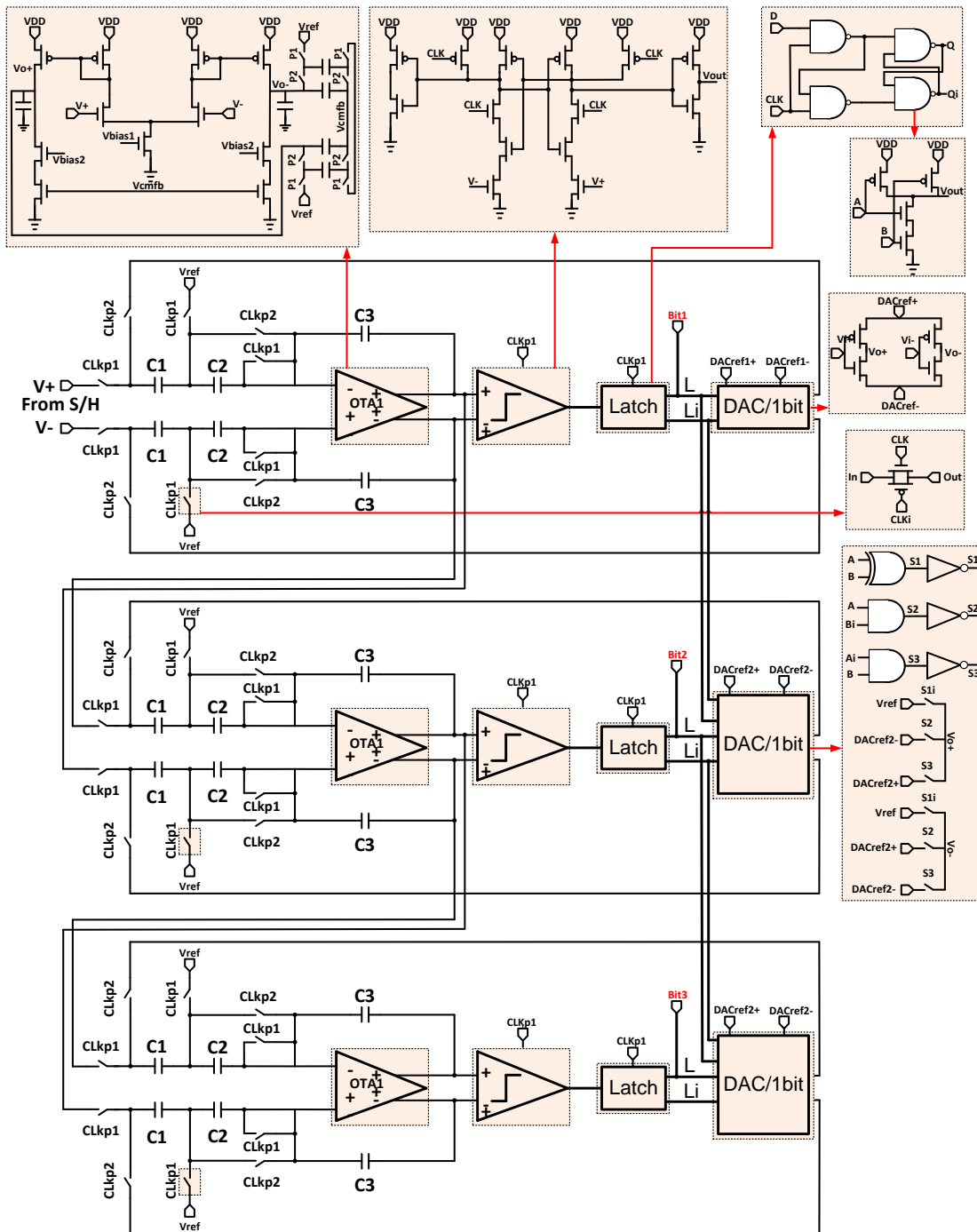


FIGURE B.21 – Concept du $\Delta\Sigma$ MASH du 3^e ordre avec les différentes topologies des circuits détaillés.

haut), un *latch* a été ajouté afin de conserver la valeur de comparaison pendant toute la période d'horloge. Ensuite, la sortie du *latch* et son inverse sont passés en argument d'un circuit de type DAC à 1 bit. Pour la première branche, le DAC convertit la sortie du *latch* en deux niveaux analogiques possibles, soit DACref1+ et DACref1-. Étant donné la nature différentielle du circuit, DACref1- peut être connecté à la référence, alors que DACref1+ doit être relié à un signal DC. Or, l'amplitude de DACref1+ permet de fixer la plage dynamique maximale à l'entrée du circuit à DACref1+/2. Une autre configuration est possible en sélectionnant DACref1- = -DACref1+ afin d'augmenter la plage dynamique maximale à DACref1+, avec l'inconvénient de devoir générer un signal DC supplémentaire.

Les DAC de la seconde et de la troisième branche sont plus complexes. Cette topologie est une innovation comparativement aux ADC de type $\Delta\Sigma$ existants et possède l'avantage d'effectuer la soustraction entre la sortie du DAC de la branche courante et la sortie du DAC de la branche précédente de façon numérique (voir Figure B.19). Les opérations arithmétiques à réaliser à l'entrée des deux dernières branches sont $V_{in} = DAC_P - Int_P - DAC_C$, ou DAC_P est la sortie du DAC précédent, DAC_C est la sortie du DAC courant et Int_P est la sortie de l'intégrateur de la branche précédente (voir Figure B.19). Or, $DAC_P - DAC_C$ peut seulement générer 3 possibilités, soit Vref, 2*DACref1+ et -2*DACref1+. Ainsi, les signaux DACref2+ et DACref2- sont générés à 2*DACref1+ et 2*DACref1- (ou Vref si DACref1-=Vref) respectivement. La table de vérité du DAC est présentée dans le Tableau B.5. La valeur de Int_P est ensuite soustraite à sortie du DAC via des transferts de charges effectués avec les *switch caps* à l'entrée de chacune des branches.

Enfin, les trois bits générés par chacune des branches ($Q_1(z)$, $Q_2(z)$ et $Q_3(z)$) sont transférés au filtre de décimation qui s'occupe d'effectuer l'opération $y(z) = Q_1(z)z^{-2} - Q_2(z)z^{-1}(1 - z^{-1}) + Q_3(1 - z^{-1})^2$ de façon numérique avant de filtrer et de décimer.

TABLE B.5 – La table de vérité du DAC des 2^e et 3^e branches du $\Delta\Sigma$ qui permet d'effectuer l'opération $DAC_P - DAC_C$ sans utiliser de *switch caps*. Cette technique permet de sauver de l'espace, car elle ne nécessite pas de condensateurs ni de commutateurs supplémentaires comparativement au DAC de la 1^{re} branche.

DAC courant	DAC précédent	Sortie+	Sortie-
0	0	Vref	Vref
0	1	2*DACref1+	-2*DACref1-
1	0	-2*DACref1-	2*DACref1+
1	1	Vref	Vref

B.6 SoC : Filtre de décimation

Le filtre de décimation est un filtre FIR de type CIC d'ordre 4 implémenté avec seulement des intégrateurs et des filtres en peigne avec gain négatif unitaire. Ce type de filtre a été

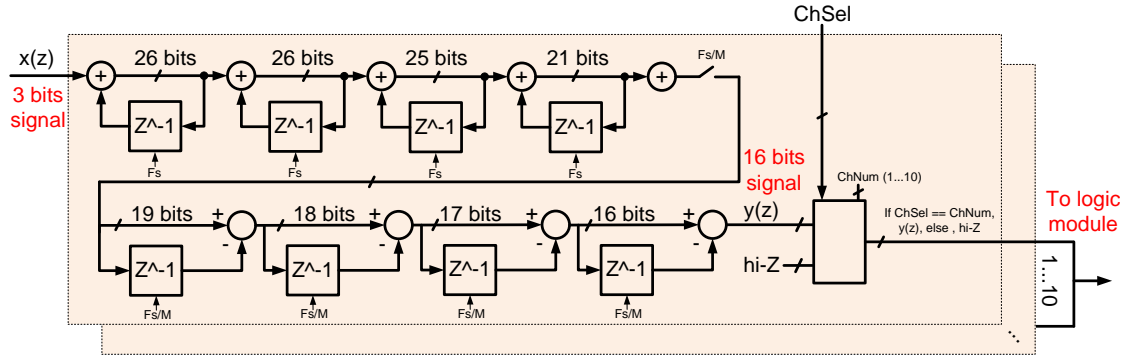


FIGURE B.22 – Concept du filtre de décimation CIC d’ordre 4. Il y a un filtre par $\Delta\Sigma$ et tous les filtres sont reliés à un bus unique pour communiquer avec le module de contrôle.

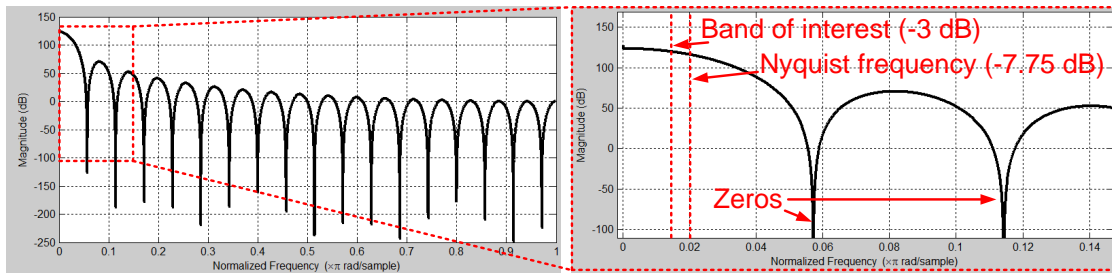


FIGURE B.23 – Bande passante du filtre de décimation CIC d’ordre 4. La fréquence de coupure à -3 dB du filtre est indiquée en rouge ainsi que la fréquence de *Nyquist*.

choisi pour une raison évidente : sa simplicité d’implémentation *on-chip*. De plus, le filtre de décimation a été conçu en suivant la procédure d’optimisation décrite dans [182].

B.6.1 Aspects mathématiques

La fonction de transfert d’un intégrateur peut être écrite comme suit $H_i(z) = \frac{1}{1-z^{-1}}$ et la fonction de transfert d’un filtre en peigne avec gain négatif unitaire et un délai différentiel de M comme suit $H_C(z) = 1 - z^{-M}$. La fonction de transfert du filtre de décimation est donc :

$$H(z) = A_{PR} \cdot \frac{(1 - z^{-M})^4}{(1 - z^{-1})^4} \quad (\text{B.24})$$

La réponse en fréquence de ce filtre est :

$$H(\omega) = A_{PR} \cdot \left[\frac{1 - e^{-jM\omega}}{1 - e^{-j\omega}} \right]^4 = A_{PR} \cdot e^{-j2\omega(M-1)} \left[\frac{\sin(\frac{\omega M}{2})}{\sin(\frac{\omega}{2})} \right]^4 \quad (\text{B.25})$$

Or, en négligeant le terme de phase $e^{-j2\omega(M-1)}$ dans (B.25), le terme $A_{PR} \cdot \left[\frac{\sin(\frac{\omega M}{2})}{\sin(\frac{\omega}{2})} \right]^4$ est un filtre passe-bas possédant des zéros à chaque $\omega = \frac{2\pi}{M}, i = 1..M/2$, un gain de $A_{PR} \cdot M^4$

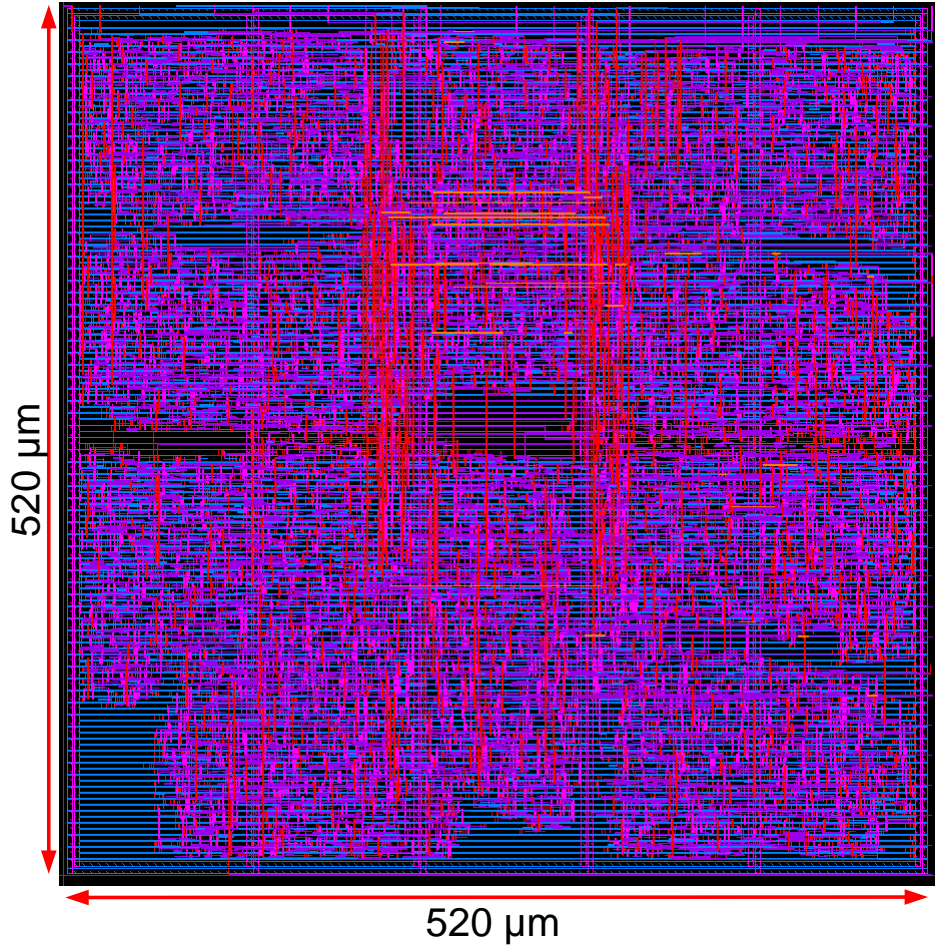


FIGURE B.24 – *Layout* des 10 filtres de décimation. La taille du circuit est de 0.27 mm^2 ($0.027 \text{ mm}^2/\text{filtre}$).

(applicable sur la valeur binaire à l'entrée $[0, 7]$) et possédant une fréquence de coupure (-3dB) à $\omega \approx \pi/(2M)$. Ici A_{pR} est un facteur de réduction de gain engendré par la limitation de la résolution binaire entre les étages afin de réduire la taille du circuit. Ce terme est défini par $A_{pR} = 1/2^{\sum_{k=0}^7 B_{p,k}-1}$, ou $B_{p,k}$ est le nombre de bits réduit à l'étage k .

B.6.2 Implémentation

Tel que susmentionné, chacun des filtres d'intégration et de peigne a été implémenté afin de minimiser la taille des registres et de la logique. La méthode utilisée est décrite dans [182] et permet d'optimiser considérablement la taille et la consommation énergétique du circuit sans en affecter les performances (troncature d'un registre à l'autre). Les tailles optimisées de chacun des registres sont indiquées à la Figure B.22.

On peut voir dans la Figure B.22 que les filtres de décimation sont reliés via un bus unique qui est relié au module de contrôle (voir section B.7). Le module de contrôle sélectionne un

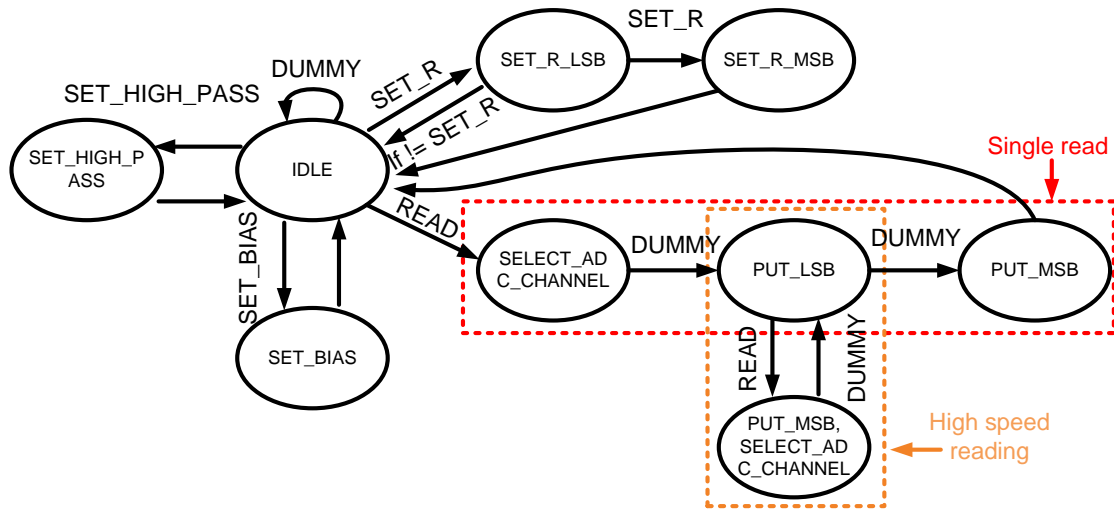


FIGURE B.25 – Schéma fonctionnel de la machine à état du module de contrôle.

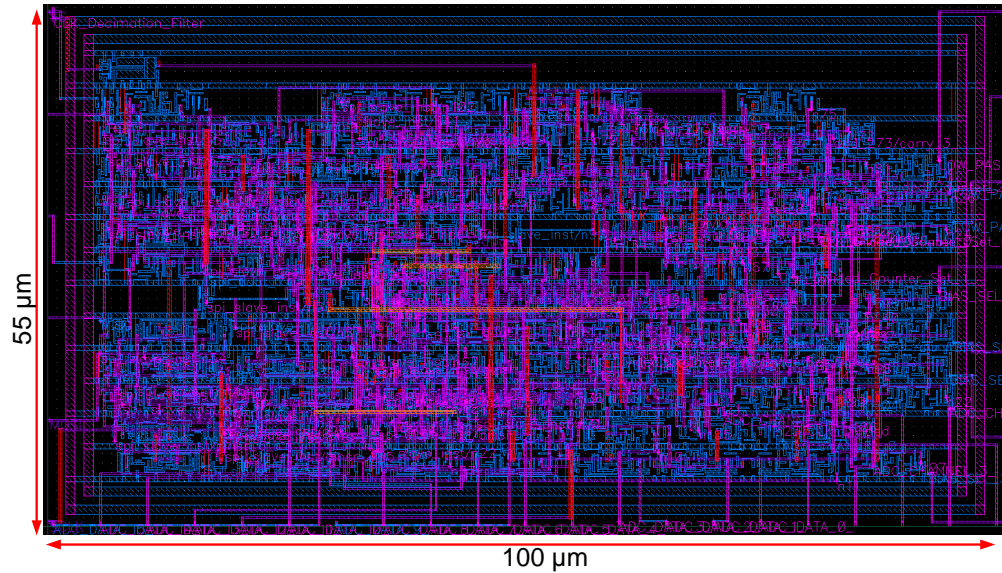


FIGURE B.26 – *Layout* du module de contrôle. La taille du circuit est de 0.0055 mm².

seul canal à la fois. Le canal sélectionné met sa sortie $y(z)$ sur le bus alors que les autres mettent leur sortie en haute impédance. Le *layout* du circuit de décimation est présenté à la Figure B.24 et possède une aire de 520 μm x 520 μm (0.027 mm²/filtre). Il est important de noter que les filtres ont été conçus en langage VHDL et importés dans *Cadence* en suivant le *design flow* de CMC, soit en utilisant les logiciels *Synopsis*, *ModelSim*, *Encounter* et *Cadence*. La librairie de cellules logiques utilisée est GF (IBM) 0.13 μm CMOS [183] de CMC.

B.7 SoC : Module de contrôle

Le circuit de contrôle permet à la puce de transférer les données de l'ADC, configurer la valeur de M dans le filtre de décimation, configurer la fréquence de coupure passe-haut des filtres analogiques et de configurer le courant de biais du AFE. Essentiellement, ce circuit est constitué d'un module SPI de type *slave* et d'une machine à états qui ont été codés en VHDL. Le module SPI est conçu afin de transférer des octets et de fonctionner avec une phase et une polarité nulle. Les octets transférés du *master* doivent être encodés de la façon suivante : xxxxyyy, où *yyy* est le type de commande et *xxxx* sont les données associées à la commande. Le schéma fonctionnel de la machine à état est présenté à la Figure B.25 et le *layout* du circuit de contrôle à la Figure B.26 (100 μm x 55 μm). Il est important de noter que le circuit de contrôle a été conçu en langage VHDL et importés dans *Cadence* en suivant le *design flow* de CMC, soit en utilisant les logiciels *Synopsis*, *ModelSim*, *Encounter* et *Cadence*. La librairie de cellules logiques utilisée est GF (IBM) 0.13 μm CMOS [183] de CMC.