

Artifact-Free Stimulation for Next-Generation Single-Cell Resolution Epiretinal Implants

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

Masters of Science in Biomedical Engineering

by

Rohan Brash



Faculty of Mechanical, Maritime and Materials Engineering
Delft University of Technology

Netherlands

03 November 2021

Abstract

Implantable epiretinal prostheses aim to restore visual capability for patients suffering from diseases such as Macular Degeneration and Retinitis Pigmentosa by bypassing damaged photoreceptors and electrically exciting Retinal Ganglion Cells (RGCs) directly. Next-generation devices will require single-cell resolution, and bidirectional capabilities to classify and selectively stimulate different cell types in the RGC population as to produce an image closer to healthy eye operation. The classification process requires the capture of direct neural responses immediately following stimulus. A significant challenge to overcome before the next generation of devices can be realised is the stimulation artifact. Stimulation artifacts are large unwanted voltages resulting from interactions between the stimulation current and the tissue-electrode impedance that can last for tens to hundreds of milliseconds, saturating the recording electronics and obscuring the direct responses. This work aims to develop a novel stimulation artifact reduction method that can be implemented in high-resolution, high channel-count brain-machine interfaces. The algorithm requires generating a model of the tissue-electrode interface from discrete measurements in the time or frequency domain, and constructing a stimulation waveform that maximally reduces the duration of the residual artifact without the need for complex recording front-end electronics. The proposed method also includes a trimming step that corrects for small variations in the model parameters and stimulation currents. Using a custom stimulation and recording test board, the algorithm was tested for 30 μm diameter gold/polyurethane electrodes in saline solution. Anodic stimulation current pulses of -100 to -500nA and 50 to 250 μs were successfully corrected for, resulting in an average artifact recovery time of 124 μs at the stimulating electrode when measured from the end of the anodic (working) phase. This corresponds to a mean 81% improvement from the next best conventional charge-balanced stimulation waveform, and 79% improvement when compared to an active-discharge recovery method. Recovery times at non-stimulating electrodes remain largely unaffected. Future work should focus on exploring alteration of the working phase for further artifact reduction, and reducing the computational cost of the algorithm for implementation on-chip.

Dedication

This thesis is dedicated first and foremost to my parents, aunt, uncle, siblings, and to Amanda, who from the opposite side of the world encouraged and supported me endlessly during both the good times and the bad. I'm so immensely thankful to have such incredible people cheering me on. Thank you.

I also dedicate this to the amazing friends I have made here in Delft during my studies. It is an understatement to say that the past couple of years have not been an easy time to study, and having friends to share problems, celebrations, and adventures with has certainly kept me sane.

Finally, a special dedication is reserved for my Granny, who passed away during my studies and whom I know would be so proud of what I have achieved.

Acknowledgements

This project would not have been conceived, let alone completed, without the expertise, guidance and supervision of Dr. Dante Muratore, who lended advice and pushed me when needed, but who also trusted me enough to take my own approach and work independently.

I would also like to acknowledge Ron van Puffelen and Jeroen Bastemeijer for their assistance in designing, manufacturing and assembling the prototype test boards, without which my project would not have worked; and Nasim Bakhshae, for her help early on in the project obtaining electrodes for initial testing and advising me in the preparation of the final electrodes.

Finally, a big thank you to Vasiliki Giagka, Andra Velea, Anna Pak, Josh Wilson and the rest of the team at the Fraunhofer Institute in Berlin for, very graciously, taking the time to design and manufacture electrode arrays for me. Your help is hugely appreciated.

Contents

1	Introduction	1
1.1	Background	1
1.2	State of the Art	3
1.2.1	Stimulation Artifacts	3
1.2.2	Artifact Mitigation	5
1.2.3	Comparison of Existing Technologies	9
1.3	Approach	12
1.4	Thesis Outline	13
2	Algorithm Development	14
2.1	Measurement Setup	14
2.2	Algorithm Design	16
2.2.1	Electrode-Electrolyte Model	17
2.2.2	Artifact Reduction	23
3	Results	30
3.1	Artifact Characterisation	30
3.1.1	Linearity	30
3.1.2	Additivity and Time-Invariance	31
3.1.3	Spatial Variation	32
3.2	Artifact Modelling	32
3.3	Artifact Reduction	34
3.3.1	Test Criteria	34
3.3.2	Comparative Methods	35
3.3.3	Performance Comparison	36
3.4	Algorithm Cost	38
3.4.1	Chip Area	38
3.4.2	Computational Cost	38
3.4.3	Memory Requirements	41
4	Discussion and Conclusions	42
4.1	Algorithm Performance	42
4.1.1	Model Parameter Fitting	42
4.1.2	Artifact Reduction	42
4.1.3	Trimming	43
4.1.4	Voltage Accumulation	43
4.1.5	Susceptibility to Parameter Error	44
4.2	Test System Limitations	44
4.2.1	DAC Current Calibration	44

4.2.2	Saturation and Blanking	45
4.2.3	Time-Resoution	46
4.3	Recommendations and Future Work	46
4.3.1	Altering Working Phase	46
4.3.2	Scaling Up	47
4.4	Conclusions	48
Appendix		
A.1	Test Board Schematic	57
A.2	Artifact Reduction Algorithm - MATLAB Script	63
A.3	Algorithm Test Parameters	71

Chapter 1

Introduction

1.1 Background

Over the past two decades, advancements in the areas of electrophysiology and nanofabrication technologies, as well as a considerable increase in public interest, have led to a rapid and concerted effort by the scientific community to develop the next generation of closed-loop neuromodulation systems. Brain-Machine Interfaces (BMI) have proven to be effective tools in many areas of medicine. Examples include the therapeutic treatment of conditions such as Parkinson’s Disease [1] and epilepsy [2], the restoration of communicative abilities for patients with neurodegenerative conditions [3] and the control of external devices or prosthetics for paralyzed patients [4]. In order to progress understanding of human physiology and stimulation therapy, the next-generation of neural implants require higher spatial resolution than ever, permitting large scale selective stimulation and recording of individual neurons in a culture or tissue. One particular application that benefits immensely from these advancements, and the focus of this work, is the epiretinal implant.

Fig. 1.1 demonstrates the typical location of an epiretinal implant on the inner surface of the back of the eye. The retina consists of three distinct layers of cells. Detailed from posterior to anterior, they are: the photoreceptors, which convert incident light into electrical signals; interneurons, which integrate the electrical signals generated by the photoreceptors; and the retinal ganglion cells (RGCs), which transport the electrical signals to the brain, and whose long axons form the optic nerve [5].

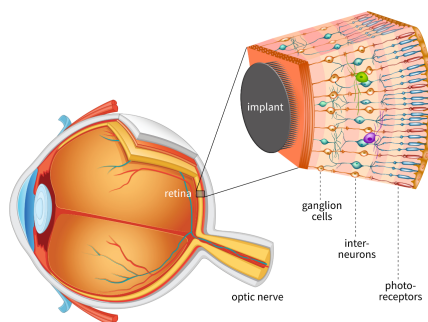


Figure 1.1: Illustrative depiction of epiretinal implant location. Source: Stanford Artificial Retina Project [6]

Diseases such as Macular Degeneration or Retinitis Pigmentosa can cause degradation of the photoreceptors within the retina, preventing the incoming light from being transduced into electrical signals for the RGCs, and resulting in severe vision impairment. Among the retinal cells, RGCs are the only cells that can be easily recorded by extracellular electrodes, so they provide a promising target for closed-loop therapeutic intervention. When paired with an external camera and a processing unit, an epiretinal implant aims to bypass the damaged photoreceptors and interneurons, stimulating the RGCs directly to emulate healthy eye operation [7].

A significant challenge that has limited the efficacy of previous generation epiretinal implants is the diversity of cell types in the retinal ganglion cell population. Rather than a homogeneous lattice of a single cell type, the RGC layer is composed of approximately 20 different cell types, all of which differ in population density, response to normal visual stimulus and destination in the brain [6]. The soma of these cells ranges from just tens to hundreds of micrometers [5]. Non-specific stimulation of multiple RGC cell types results in a convoluted image perceived by the patient. The goal of next-generation retinal implants is to utilise high-density, single-cell resolution bidirectional interfaces to classify the RGC cells, forming a dictionary of cell types and locations; and subsequently use such dictionary to generate cell-type specific stimulation patterns that result in coherent image perception. Fig. 1.2 details a block diagram of the proposed system.

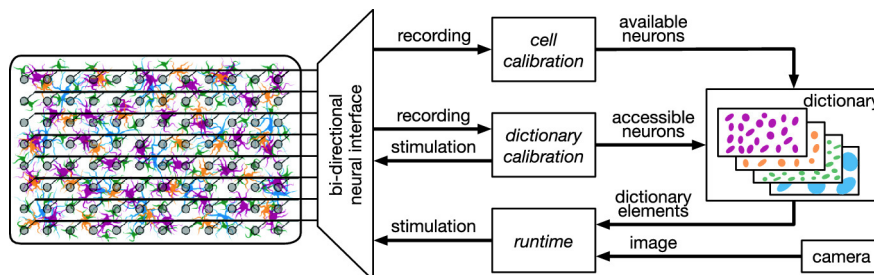


Figure 1.2: Concept block diagram of cell-type specific resolution epiretinal implant and corresponding high-density electrode array. The different cell colours in the population (left) are representative of different RGC cell types. Source: [7]

Based on the required channel pitch, which should be $\sim 60 \mu m$ to distinguish between cells, it has been proposed that in order to effectively recreate useful visual stimuli for the patient, a channel count of at least 10^4 is necessary such that a sufficient area of the visual field is covered [7]. As a result of these extreme channel counts, severe constraints on per-channel area, power consumption and computational complexity arise. Additionally, to successfully classify cell-types, direct neural response action potentials (APs) immediately following stimulation, as opposed to later network mediated responses, must be recorded. Some direct AP responses appear less than $500 \mu s$ following stimulation (*ex-vivo* monkey retina [8]). Among others, a significant hurdle to overcome before these devices become clinically feasible is the presence of stimulation artifacts, which saturate sensitive recording electronics and prevent the capture of these directly elicited action potentials [9].

The aim of this work is to develop an understanding of the origin and nature of stimulation artifacts, assess the efficacy of existing technologies and methods employed to combat the presence of stimulation artifacts, and develop an approach to minimise stimulation artifacts in large scale high resolution brain-machine interfaces. Specifically, how can stimulation artifacts be mitigated such that neural responses can be reliably captured $<500 \mu\text{s}$ post-stimulation, from up to 10^4 channels, including the stimulation channel, while maintaining strict constraints on power consumption, chip area and computational complexity?

1.2 State of the Art

1.2.1 Stimulation Artifacts

Stimulation artifacts are large, slowly decaying voltage transients that appear in the recording channels at and near the stimulating channel in an implanted micro-electrode array.

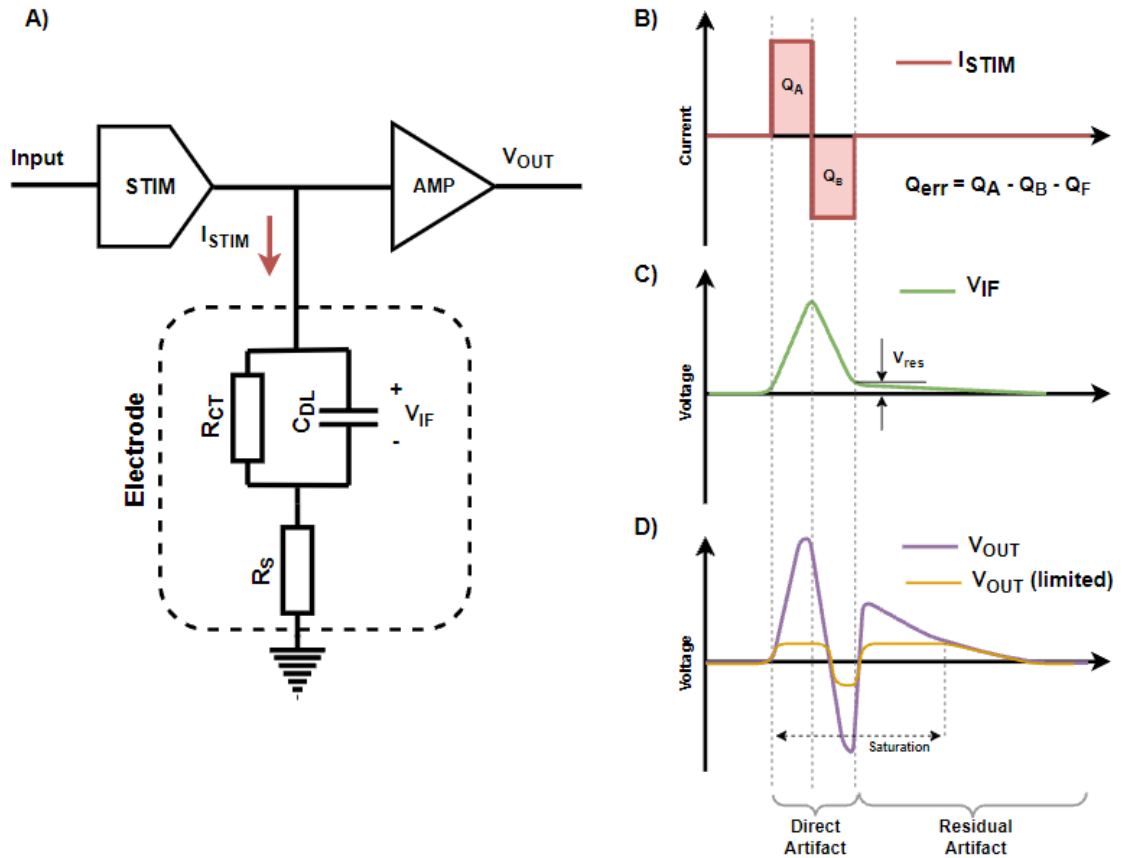


Figure 1.3: Origin of the stimulation artifact: (A) Simplified diagram of a current-mode bidirectional interface and linear tissue-electrode model with charge transfer resistance (R_{CT}), spread resistance (R_S) and double-layer capacitance (C_{DL}); (B) a biphasic charge-balanced current stimulus waveform; (C) interface voltage and residual error V_{res} due to charge balance error; (D) depiction of amplifier output with infinite range (unlimited) and with a finite output range (limited), indicating saturation of the device.

In a typical current-mode bidirectional stimulator (Fig. 1.3 (A)), charge-balanced stimulation waveforms (Fig. 1.3 (B)) are often employed to minimise excess charge accumulation at the interface. Such accumulation can result in irreversible faradaic processes that may damage tissue, produce harmful chemical species or alter local electrochemical properties [10]. Mismatches in the injected (Q_A) and removed (Q_B) charge, along with losses due to interface non-linearities (Q_F), result in a residual charge (Q_{err}) that remains at the interface. During stimulation, a potential, V_{IF} , accumulates at the interface (Fig. 1.3 (C)). After stimulation, a residual potential remains, V_{res} , resulting from the stimulation charge imbalance. There are two parts to the stimulation artifact: the direct artifact, which coincides with the stimulation waveform and has the largest amplitude; and the residual artifact, a slow-decaying component that persists after the stimulation pulse has terminated, resulting from the remaining charge at the interface (Fig. 1.3 (D)) [9, 11].

Depending on the tissue properties, electrode characteristics and placement of the electrodes, recorded neural responses of interest such as local field potentials (LFPs) and action potentials (APs) can be as low as $10 \mu V$ [12]. As such, the recording channels of neural interfaces need to be incredibly sensitive in order to digitize and analyse these signals. Stimulation artifacts can be up to tens or even hundreds of millivolts [13], so can result in significant non-linearity or saturation of the input amplifiers, distorting and obscuring important neural data. To digitize both the AP and artifact simultaneously, a dynamic range of up to 80dB is required. In an example depicting $200 \mu s$ stimulation pulses for different amplitudes (Fig. 1.4 (A)), it can be seen that while the direct artifact lasts for only $400 \mu s$, the residual artifact can saturate the amplifier for up to $40 ms$. Fig. 1.4(B-C) depict example neural signals, and the same signals corrupted by artifacts, respectively.

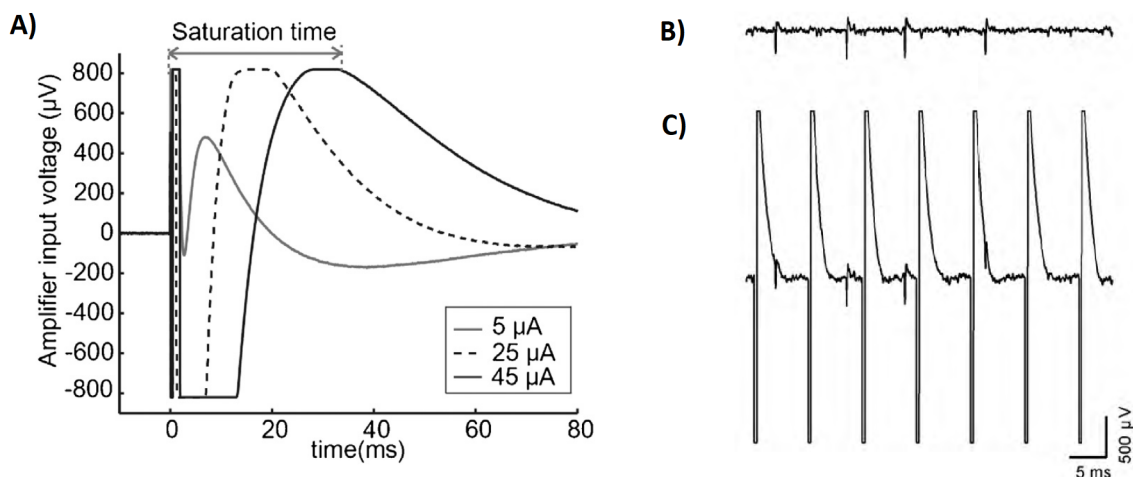


Figure 1.4: Examples of recorded stimulation artifacts: artifacts from anode-first biphasic $200 \mu s$ stimulation pulses of different current amplitudes [14]; (B) example neural signal trace; (C) neural trace corrupted by stimulation artifacts. [15]

It is important to emphasize that a perfectly charge-balanced stimulation waveform will not result in zero residual artifact. A linear model was presented in Fig. 1.3 (A), but in reality the tissue-electrode impedance demonstrates a much more complex impedance profile. Fig. 1.5 shows an Electrochemical Impedance Spectroscopy

(EIS) measurement from a cortical implant in a rat at various points before and after implantation. Although EIS is a linear measurement technique, the results demonstrate how complex and time-varying the impedance can be [16]. A systematic investigation by Mena et al. identified how stimulation artifacts vary over a number of parameters [8]. They showed that the magnitude and shape of the artifact varies significantly both spatially and temporally, while also differing in shape between the stimulating and recording electrodes.

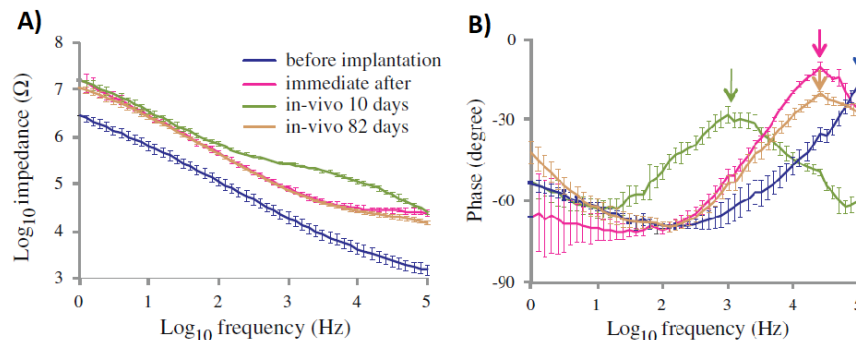


Figure 1.5: Electrical Impedance Spectroscopy magnitude (A) and phase (B) of a Pt/Ir cortical implant in a rat at various stages before and after implantation [16].

1.2.2 Artifact Mitigation

Fig. 1.6 presents a simplified overview of the bidirectional signal chain with the general location and names of common stimulation artifact mitigation methods.

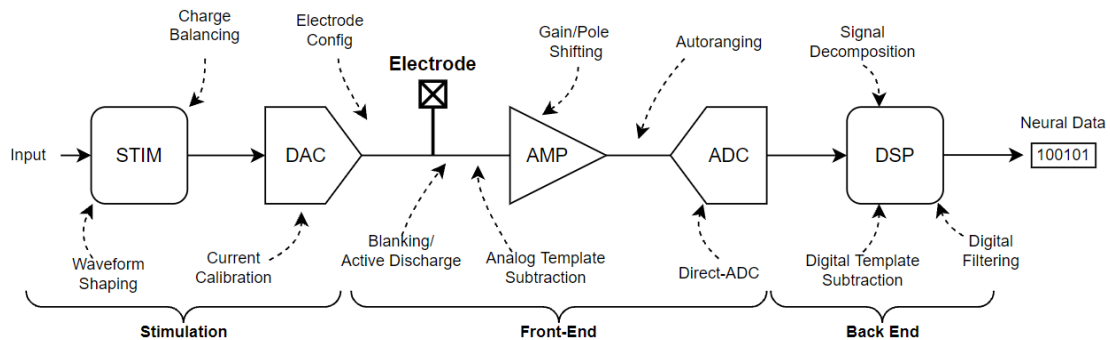


Figure 1.6: Simplified signal chain with location and types of stimulation artifact mitigation techniques.

Artifact mitigation methods can be divided into three main groups: stimulation-side mitigation, that takes place before the stimulation signal reaches the interface; front-end techniques, which are employed prior to signal digitization, and back-end processing techniques that address the artifacts in recorded data.

Stimulation-Side Mitigation

The first solution is to minimise artifact generation at the source. This involves attempting to eliminate the root cause of the residual artifact, which is the remaining charge at the interface following the stimulation pulse, or prevent the artifact from

reaching the recording electrode all together. Common methodologies to achieve this include precise charge balancing, differential stimulation topologies, stimulation cancellation patterns and stimulation shaping.

Charge balancing is implemented by most stimulator topologies, but some systems take additional steps to ensure that the charge injection error is as small as possible, such that the residual artifact is reduced. A common method is to employ an H-Bridge topology (Fig. 1.7(A)) that utilises a switch network to reuse the same current source for both the anodic and cathodic phases of current mode biphasic stimulation [17, 18, 19, 20]. Other implementations utilise separate anodic and cathodic current sources, but provide methods to automatically calibrate the current sources for precise matching [21], or actively measure and cancel the offset at the interface following the primary stimulation pulse [22]. These methods can produce very precise stimulation waveforms (0.01% mismatch) but as mentioned in the previous section, the nature of the tissue impedance means that even perfect charge balancing will not result in elimination of the residual artifact completely.

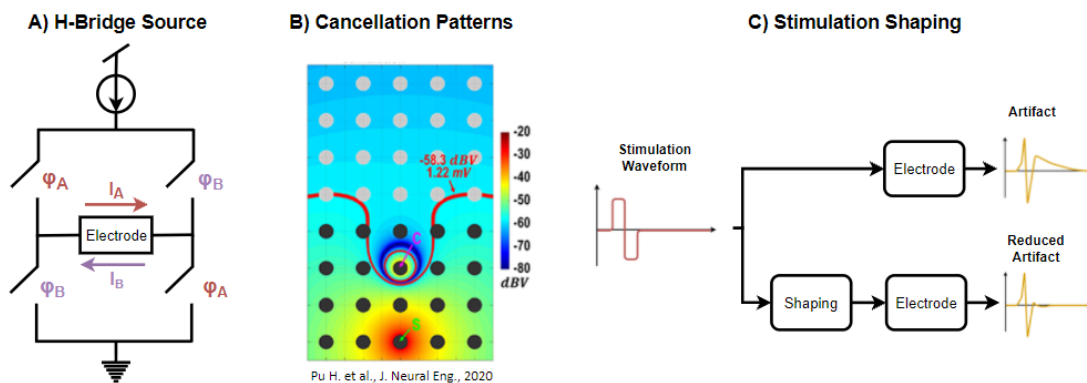


Figure 1.7: Stimulation-side artifact mitigation techniques: A) H-Bridge current reuse topology for precise charge balancing; B) complementary stimulation (electrode C) to cancel the artifact produced by the primary stimulation (electrode S) at the recording electrodes (grey electrodes) [23]; C) shaping of the stimulation pulse to reduce artifact length/magnitude.

In addition to charge balancing, the stimulation/recording electrode placement and stimulation magnitudes can be designed in such a way that the stimulation artifact is cancelled or appears as a common mode source [23, 24, 25, 26]. Pu et al. designed an optimal artifact cancellation algorithm that uses complementary additional stimulating electrodes in an array to cancel artifacts for a given set of recording electrodes (Fig. 1.7(B)). This method does not, however, eliminate the artifact on the primary stimulating electrode [23].

Some of the less explored solutions involve altering the stimulation waveform itself (Fig. 1.7(C)). For example, an approach by Kolodziej et al. [27] focuses on producing a mathematical model of the tissue-electrode impedance and applying a correction pulse to the end of the primary biphasic stimulation pulse to mitigate the trailing artifact. Dura et al. developed an algorithm to convert stimulation pulses into electrophysiologically equivalent high frequency pulse chains, the artifacts of

which can be easily filtered at the recording side [28]; and Heffer et al. demonstrated that stimulating at a higher frequency than the neural signals facilitates artifact removal through efficient linear interpolation [29]. Finally, an insightful article by Chu demonstrated that by treating the tissue-electrode interface like a communications channel, digital equalization algorithms can be implemented to pre-distort the stimulation pulse, significantly reducing the artifact length [30].

Front-End Mitigation

Once the artifact has been reduced as much as possible at the source, it is necessary to explore methods to prevent the artifact from corrupting recorded signals. It is important to note that, unless a pulse modulation method, such as that by Dura et al. discussed above, is implemented, the artifact occupies similar frequency ranges as the neural signals of interest [9]. As such, simple passive filtering does not form a complete mitigation strategy.

The goal of front-end mitigation is to prevent saturation of the recording analog front-end (AFE) and restore normal operation as quickly as possible. The simplest front-end mitigation method is to disconnect the AFE input during stimulation (blinking, Fig. 1.8(A)) [31, 32, 33, 34, 35] or hold it at a predetermined value (sample-and-hold) [36, 37].

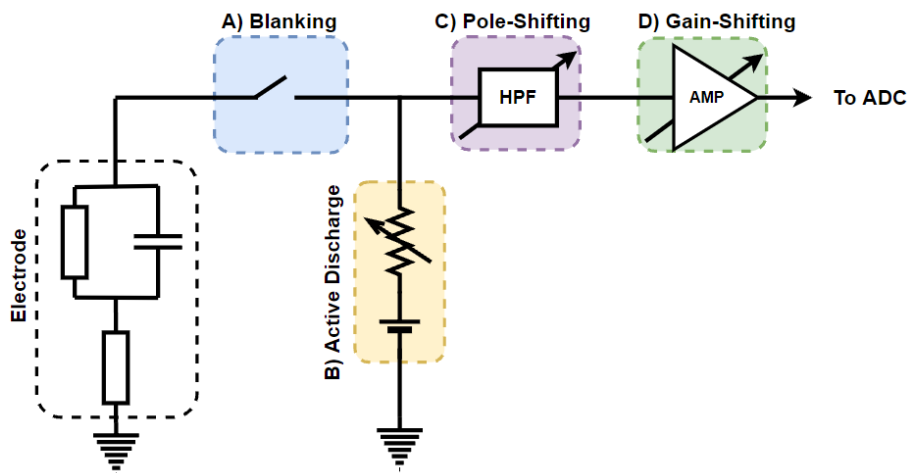


Figure 1.8: Example of front-end mitigation techniques.

These methods are possibly the least costly to implement, both in area and power consumption, and successfully prevent saturation of the AFE. However, they do nothing to reduce the length of the artifact, so it is necessary to wait until the interface discharges before reconnecting, which can take tens of milliseconds [12]. Active discharging (Fig. 1.8(B)), normally implemented as a switchable low-impedance path to ground or fixed potential, can be employed to remove the excess charge on the interface, thereby reducing the recovery time following stimulation [13, 38, 39, 40]. The recovery time of the input stage can further be reduced by implementing an adjustable high-pass filter (Fig. 1.8(C)), of which the cutoff frequency can be temporarily raised during stimulation to decrease the magnitude of the artifact. This method is known as pole-shifting [11, 41, 42, 43]. Similarly, the gain of the

input stage can also be temporarily reduced, sometimes to unity, for the duration of the artifact [12, 44] to prevent saturation at the cost of signal resolution (Fig. 1.8(D)).

A more costly solution is to design the front end with a lower gain and higher resolution ADC such that it achieves successful digitization of both the signal of interest and the artifact [17, 45, 46, 47, 48, 49, 50], or implement a sigma-delta topology to remove the need for an amplifier altogether [51, 52]. The benefit of this method is that separation of the artifact and neural signal can be shifted to the back-end digital domain, where there are many algorithms available. Of course, higher resolution ADCs come at the cost of increased power consumption and chip area.

Back-End Mitigation

If one of the methods detailed above is used to achieve digitization of all or part of the artifact, then back end processing techniques can be used to recover the neurological signals from the recorded data.

One method is to simply detect the peaks in data associated with the artifacts and interpolate between the start and end of each artifact to remove it (Fig. 1.9(A)) [29, 53, 54]. This is an appropriate method when the signal of interest is lower frequency than the artifact, such as LFPs. On the other hand, multiple APs can occur during a single artifact, so interpolating discards important data.

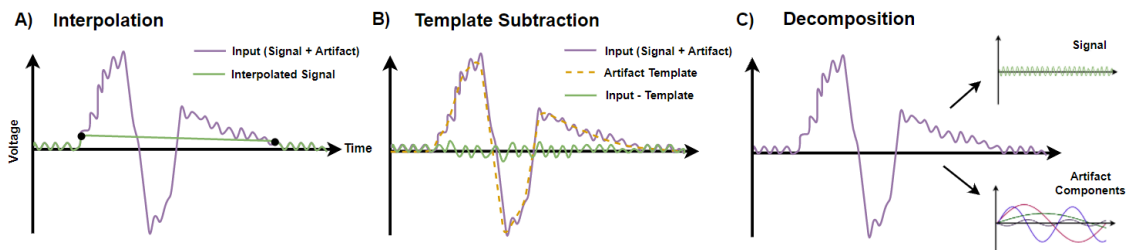


Figure 1.9: Back-end digital artifact removal methods: A) interpolation, B) artifact template formation and subtraction, and C) signal decomposition.

Digital template subtraction algorithms (Fig. 1.9(B)) are much more versatile. They involve generating a template of the artifact by averaging over multiple stimulation pulses, from the stimulation parameters and an electrical model of the interface, or through a polynomial fitting function. The artifact is then subtracted from the recorded data to recover the neural data [8, 15, 55, 56, 57, 58, 59, 60, 61, 62]. To increase computational performance, Caldwell et al. pre-generate a dictionary of templates based on different stimulation parameters and electrode locations. The incoming data is matched with the closest template and subtracted as above [63].

Finally, signal decomposition techniques (Fig. 1.9(C)) such as Principle Component Analysis (PCA) [17], Independent Component Analysis (ICA) [64] and Empirical Mode Decomposition (EMD) [65, 64], along with adaptive noise cancellation

[66] and other adaptive digital filtering methods [14, 20, 32, 48, 67, 68, 69, 70, 71] can also be used to separate the artifact and signal.

Hybrid Mitigation

Some methods utilise a combination of analog front-end and digital-back end processing to achieve stimulation artifact reduction. Such methods tend to be the more sophisticated and ultimately better performing than any single-domain implementation. Two such methods are digital auto-ranging and analog template subtraction.

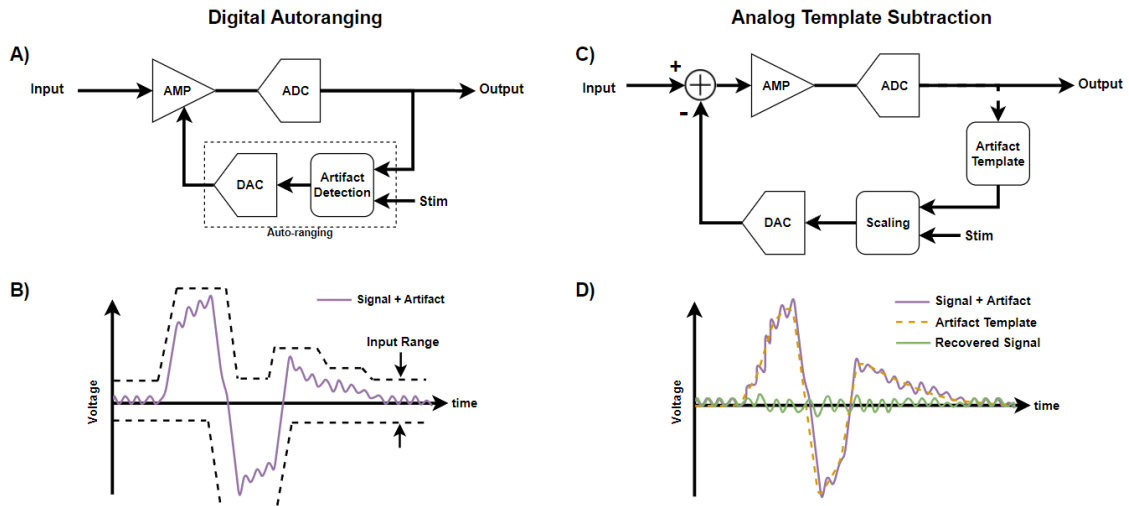


Figure 1.10: Hybrid digital-analog artifact mitigation techniques: A), B) digital auto-ranging/tracking and C) ,D) analog template subtraction. The purple trace in the lower figures represents a low-amplitude sinusoidal signal corrupted by a high-amplitude artifact.

To reduce the resolution requirements of the input ADC, some solutions employ digital auto-ranging to track rapid signal changes indicative of an artifact and use a control signal to scale the AFE range. Examples include the Track-and-Zoom approach by Pazhouhandeh [72] and the Predictive Digital Auto-Ranging method by Kim et al. [73]. These methods result in a faster artifact recovery and little to no loss of signal data. Fig. 1.10 (A) and (B) depict a simplified circuit and input range response to a signal corrupted with an artifact, respectively.

Another method that utilises both digital and analog domains is analog template subtraction. Similar to the digital equivalent mentioned in the previous section, this method subtracts a template from the input signal, but uses a feedback digital to analog converter (DAC) to perform the subtraction prior to amplification, both preventing saturation and recovering the neural signal. Because of the efficacy, reduced ADC requirements and minimal digital processing, this is one of the more popular methods in recent implementations [20, 56, 66, 74, 75, 76, 77, 78]. A simplified schematic and response is detailed in Fig. 1.10 (C) and (D) respectively.

1.2.3 Comparison of Existing Technologies

Table 1.1 below summarises the methods discussed in the previous section, specifically pertaining to qualitative stimulator complexity, recording front-end complexity,

computational complexity and overall efficacy. It can be seen that some implementations utilise multiple methods in parallel. The best results are achieved by the hybrid methods, but at the cost of increased computational and front-end complexity.

Table 1.1: Comparison table of artifact mitigation strategies. A '-' sign indicates a negative effect on the column aspect, whereas '+' indicates a positive effect.

	Method	Publications	Simulator Complexity	Recorder Complexity	Computational Complexity	Efficacy
Stimulator	Precise Balancing	[17, 18, 19, 20, 21, 22]	--		-	+
	Stim Topology	[23, 24, 25, 26]	--			++
	Waveform Shaping	[27, 28, 29, 30]	-		--	+
Front-End	Blanking/Sample-and-Hold	[31, 32, 33, 34, 35, 36, 37]		-		+
	Active Discharge/Pole-Shifting	[38, 39, 13, 40, 41, 42, 43, 11, 12, 44]		-		++
	Direct-ADC	[17, 45, 51, 52, 46, 47, 79, 48, 49, 50]		--		
Back-End	Interpolation	[29, 53, 54]			-	+
	Template Subtraction	[15, 55, 8, 56, 63, 57, 58, 59, 60, 61, 62]			--	++
	Signal Decomposition	[17, 64, 65, 64]			---	+++
	Adaptive Filtering	[20, 67, 48, 14, 32, 68, 69, 70, 71]			---	+++
Hybrid	Auto-Ranging	[72, 73]		---	-	++
	Template Subtraction	[74, 75, 20, 66, 56, 76, 77, 78]		--	--	+++

Due to the diverse applications and nature of neural interfaces, the requirement for stimulation artifact reduction varies from case to case. Of the almost 120 articles reviewed, 74 provided no numerical measure of artifact improvement. However, for those that do provide numerical results, there are two prominent measures: suppression and recovery time.

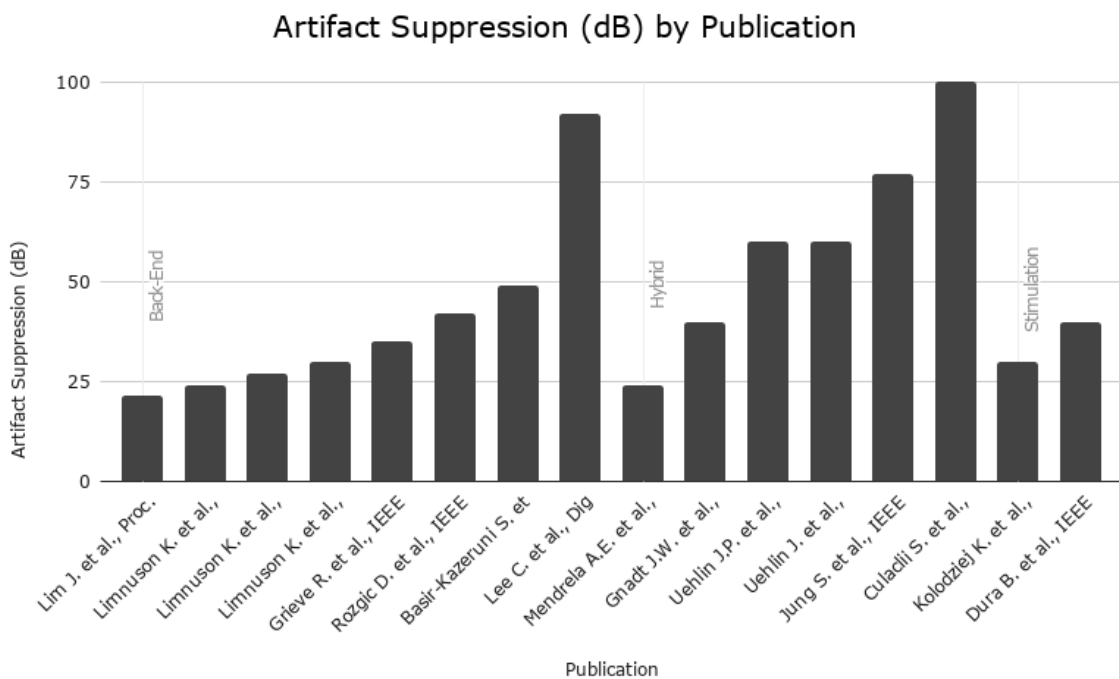


Figure 1.11: Artifact suppression in decibels by paper, organised by primary mitigation category.

Artifact suppression gives the decibel reduction in the magnitude of the artifact before and after mitigation, and the performance of publications with focus on suppression are summarised by Fig. 1.11. The most suppressive strategies are the back end methods by Lim [67] and Limnusun [80] and the hybrid solution by Mendrela [66].

Recovery time details the time from the end of stimulation waveform until the input signal returns to a level from which neural data of interest can be distinguished from the noise floor. As the neural data and technology varies from application to application, so does the voltage threshold by which recovery time is determined. As the epiretinal implant is most interested in fast-response direct APs, recovery time is the metric of importance. Performance of publications with respect to recovery time are detailed in Fig. 1.12. Lack of consistent aims and measures across publications makes demonstrations of other parameters such as power-efficiency, noise, dynamic range etc. of limited use. As such they are not depicted here.

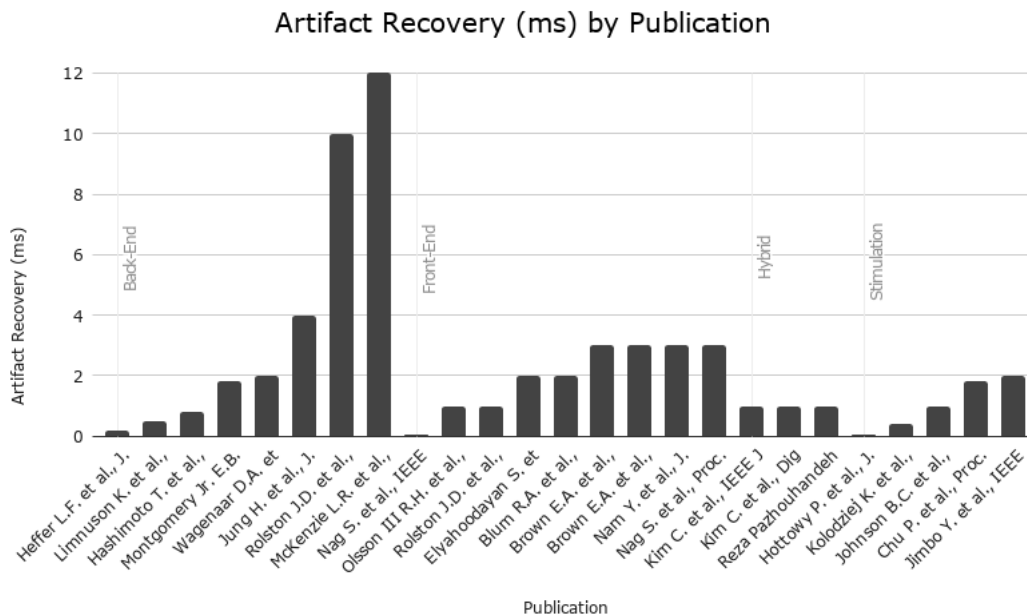


Figure 1.12: Artifact recovery time in milliseconds by paper, organised by primary mitigation category.

Twelve of the methods methods shown in Fig. 1.12 report recovery times less than or equal to one millisecond (in some cases only ' $<1\text{ms}$ ' was reported). Three methods approach the performance required by the epiretinal implant. They are the methods by Heffer [29], Nag [81] and Hottowy [39]. Heffer uses simple off-line digital interpolation requiring full digitization, so alone is inappropriate for epiretinal implant application. Nag uses a differential stimulation and recording topology with multiple amplifiers per channel to remove the artifact, limiting its usefulness in a high-density array. Finally, Hottowy, which implements a combination of stimulation waveform design and front-end methods, performed the best with a recovery time of $55 \mu\text{s}$. As such, it is of particular interest to our application.

With limited comprehensive data, it is difficult to form an accurate comparison of the methods and assess their feasibility. However, a simple test is to scale these implementations up to the magnitude of channels which the epiretinal implant hopes to achieve (10000) [7]. The best performing method (Hottowy), when scaled up, would occupy 2800 mm^2 , which is evidently far too large to be placed within the eye. The next best solutions are by Kim et al. These hybrid digital auto-ranging AFEs consume much less area and the least power of any solution. In addition to this they have a remarkable power efficiency factor (PEF) of 2.6, and 10.7 effective bits of input resolution. However, when scaled up they still require 240 mm^2 and 8 mW , not including the stimulation circuitry. This is still far in excess of what would be appropriate for an epiretinal implant.

1.3 Approach

From the findings above, it is clear that the existing methods for stimulation artifact reduction, while effective, are not well suited for scaling to the large high density arrays necessary for the next generation of brain-machine interfaces. Many of the strategies discussed employ front-end multiplexing to reduce the area required for large scale micro-electrode arrays. However, for successful cell classification and neural feedback, it is desirable to have simultaneous recording of as many channels as possible. As such, it follows that increasing complexity on the recording side to achieve artifact reduction is not the most viable way forward. It is necessary, therefore, to focus on improvements to the stimulator-side design to reduce the residual artifacts. Healthy spiking of the retinal ganglion cells is relatively sparse in time, in the order of Hz [82], so it is feasible that a few stimulator channels can be temporally multiplexed without affecting therapeutic efficacy. The overhead on the AFE is then minimal to implement blanking or unity gain switch to suppress the direct artifact.

From the stimulation-side artifact prevention methods discussed, stimulation shaping has highest potential to shift the burden of complexity away from the recording end and enable scaling to large arrays. Stimulation shaping introduces an overhead only on computational complexity as long as the stimulation circuitry is capable of arbitrary waveform generation. Stimulation shaping was also explored by Chu [30] and Kolodziej [27]. The former used digital equalisation techniques to generate a pre-distorted waveform, which resulted in a 73% reduction in artifact length. The latter developed a fitted-model from discrete frequency data points, which is then used to calculate a corrective pulse at the end of the stimulation waveform. Both of these approaches presented promising results, but suffered from model sensitivity, were not implemented outside of simulations, or did not consider maintaining stimulation efficacy.

This work will progress along a similar trajectory as Kolodziej and Chu, developing a numerical model and using it to generate an optimized stimulation waveform, while aiming to address the limitations that they encountered.

1.4 Thesis Outline

Chapter 2 outlines the method and development of an algorithm for optimum stimulation shaping to reduce the duration of the residual artifact. A numerical model of the tissue-electrode interface is developed. The model is constructed in-situ using discrete measurements, then an algorithm generates an optimum stimulation shape such that the residual artifact is reduced. We also consider how to maintain stimulation efficacy while altering the waveform. To address errors in the model's prediction and the time-varying nature of the channel, we implement adaptive 'trimming' such that the effective reduction in residual artifact length is maintained. An electrochemical test bench is also prototyped to validate the algorithm.

Chapter 3 presents the results of the algorithm testing. We begin by characterising the stimulation artifact to assess linearity and time-invariance, which is a crucial step in asserting the validity of the model. Next, the performance both in artifact reduction and computational complexity, relative to other standard methods, is assessed for a number of different pulse widths and amplitudes. Specific relationships between algorithm parameters and reduction performance are explored. The computational complexity of the chosen algorithm is also measured with varying operating conditions.

Finally, Chapter 4 presents a discussion of the results, exploring interesting findings and limitations of the algorithm and test equipment. We conclude by presenting recommendations for future work and implementation into the next generation of brain-machine interfaces.

Chapter 2

Algorithm Development

2.1 Measurement Setup

To characterise stimulation artifacts within the expected operating ranges of an epiretinal device and drive the development of a novel artifact reduction algorithm, we built the measurement setup depicted in Fig. 2.1.

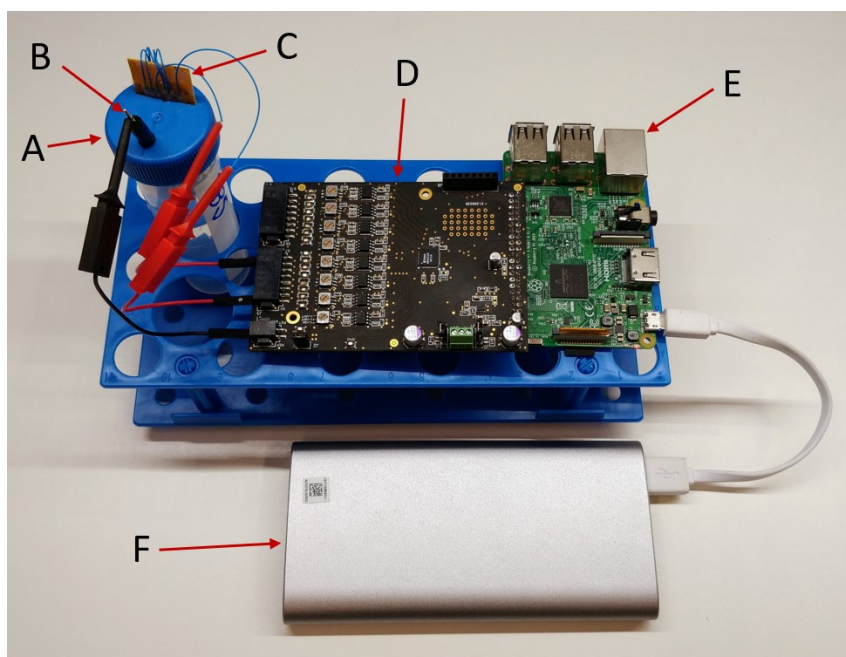


Figure 2.1: Stimulation artifact test system: A) 45ml Phosphate Buffered Saline (PBS) solution, B) Ag/AgCl reference/return electrode, C) 15 channel 30um electrode array, D) custom 16ch bidirectional stimulation/recording test board PCB, E) Raspberry Pi 3B+ Single Board Computer, F) battery power bank.

The proposed architecture is based on the Intan RHS2116 16-channel bidirectional stimulation and recording IC, the relevant specifications of which are summarised in Table 2.1. The chip is capable of 10 nA to 10 μA LSB 8-bit bipolar stimulation on each channel. This facilitates stimulation within the ranges that would be expected for single-cell stimulation ($< 1 \mu A$). For an estimated maximum electrode impedance of 500 $k\Omega$, and a maximum stimulation current of 5 μA , a maximum artifact magnitude of $\pm 100 mV$ is possible. This is well above the range

of the AC amplifiers, and with an LSB resolution of ~ 16 mV, artifacts recorded with the DC channels would be poorly digitized. As such, the test board PCB also features optional low-noise attenuators that reduce the input by a factor of approximately 20, allowing large artifacts to be recorded with the high gain AC inputs, while keeping the input-referred LSB to a reasonable value ($\text{LSB}_{\text{IN}} = 3.9 \mu\text{V}$). A simplified schematic of the test board front end is depicted in Fig. 2.2. For reference, Appendix A.1 details the schematics in their entirety.

Table 2.1: RHS2116 Stimulation and Recording IC Specifications [83]

Specification	Value	Units
Num. Channels	16	-
Max. Stim. Current	2.55	mA
Min. Stim. Current	10	nA
Max. Current Step	10	μA
Min. Current Step	10	nA
Stim. Voltage Comp.	± 4.5	V
AC Amplifier Gain	192	V/V
AC Input LSB	0.195	μV
AC Input Range	± 5	mV
DC Amplifier Gain	0.125	V/V
DC Input LSB	19.23	mV
DC Input Range	± 4.5	V

A major benefit of using the RHS2116 IC is that the chip also features multiple optionally configurable stimulation mitigation strategies such as active discharge, pole-shifting and amplifier (output) blanking, providing a baseline to compare with novel solutions. The custom-made test board interfaces with a Raspberry Pi 3B+ [84] which provides the RHS chip with power and communication. Arbitrary stimulation and recording sequences can be sent wirelessly from a PC running Matlab to the Raspberry Pi, which are translated and executed by the Intan chip. Recorded data is returned wirelessly back to the PC. By powering the setup from an USB power bank and using wireless communication, the entire test setup can be electrically isolated from the mains supply, mitigating a large amount of low-frequency noise.

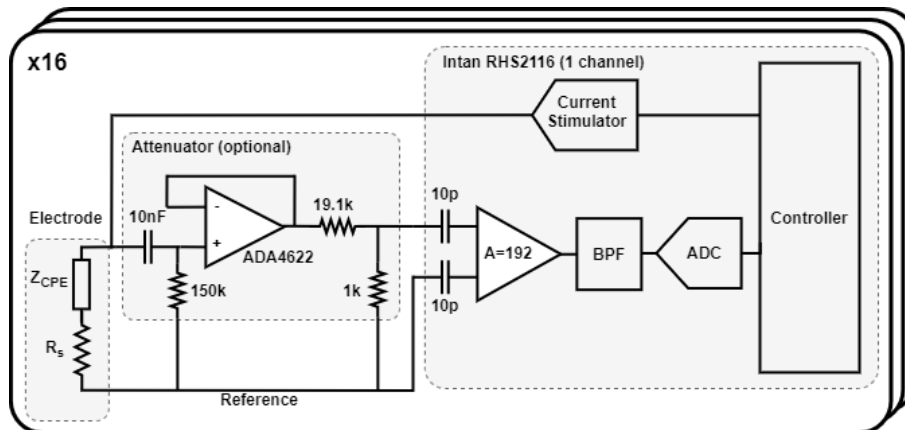


Figure 2.2: Test system channel schematic diagram.

The electrode arrays were microfabricated at the Fraunhofer Institute in Berlin, Germany. They are flexible polyurethane/gold electrodes mounted to a rigid substrate for our testing purposes. The array consists of 15 $60\ \mu\text{m}$ diameter electrode openings in a 'plus' pattern. There is a single central electrode, and 14 additional electrodes mounted at various distances from the center electrode ranging from $240\ \mu\text{m}$ to $840\ \mu\text{m}$ in increments of $120\ \mu\text{m}$. While this is not entirely representative of the $30\ \mu\text{m}$ pitch potentially required by an epiretinal implant, the electrode impedance should be sufficiently comparable. The design and manufacturing steps for the electrodes are detailed in Fig. 2.3.

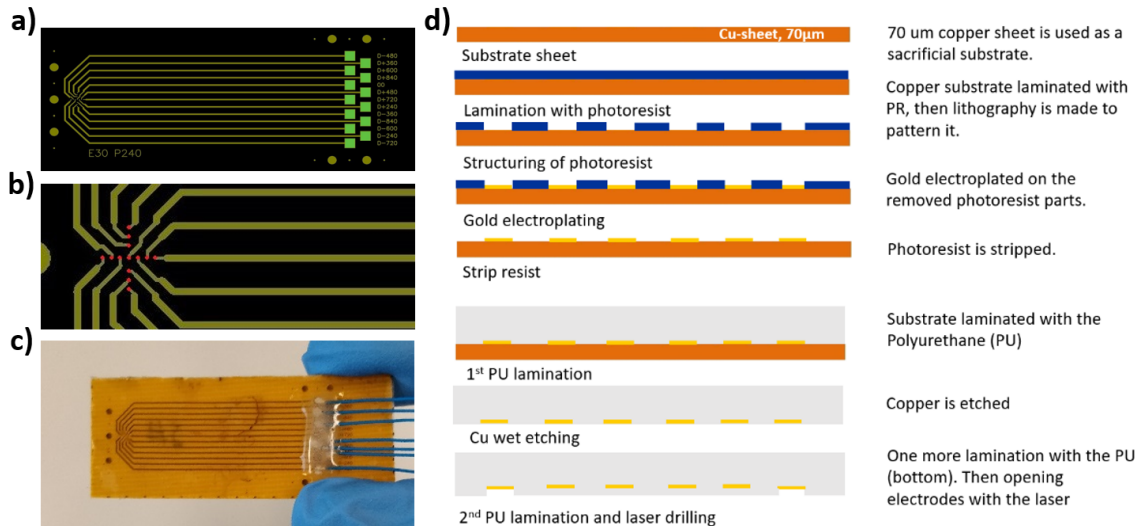


Figure 2.3: $30\ \mu\text{m}$ diameter polyurethane/gold electrodes: a) array design, b) zoomed in view of electrode openings (red), c) fabricated array, d) process steps for manufacture of the electrodes.

2.2 Algorithm Design

The key steps of the proposed model-based stimulation artifact reduction algorithm are presented in Fig. 2.4. The process involves first identifying and fitting a model of the tissue electrode interface to discrete measured data (Fig. 2.4A). Next, the model is used, along with constraints, to generate a stimulation waveform that reduces the residual artifact (Fig. 2.4B). The final step involves observing the resulting residual artifact and correcting for any remaining error during stimulation (Fig. 2.4C).

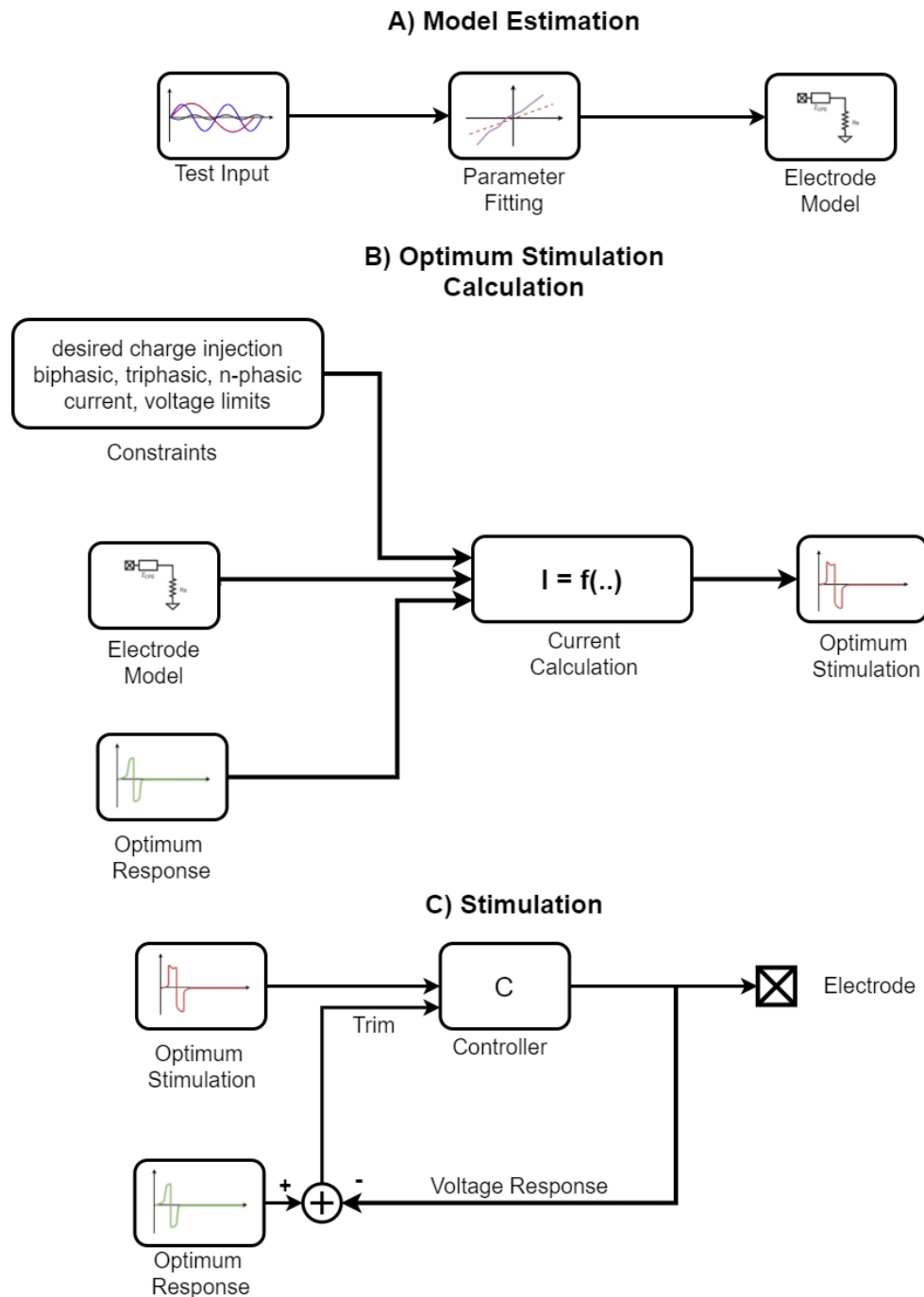


Figure 2.4: Overview of the model-based adaptive stimulation method. A) an estimate of the electrode model is formed; B) the model and stimulation constraints are used to construct an optimum stimulation waveform that produces a response with minimum residual artifact; C) the stimulation waveform is executed with error correction for disturbances.

2.2.1 Electrode-Electrolyte Model

The model selected, and the method by which it is applied, is largely based on work by Lario-Garcia et al [85]. A normal three element electrode model is depicted in Fig. 2.5a, where Z_{CPE} is a Constant Phase Element (CPE) impedance, R_e is the voltage-dependent electrochemical resistance and R_S is the tissue spreading resistance. We neglect the electrochemical resistance as it is often much larger than the impedance of the constant phase element (Fig. 2.5b) [27].

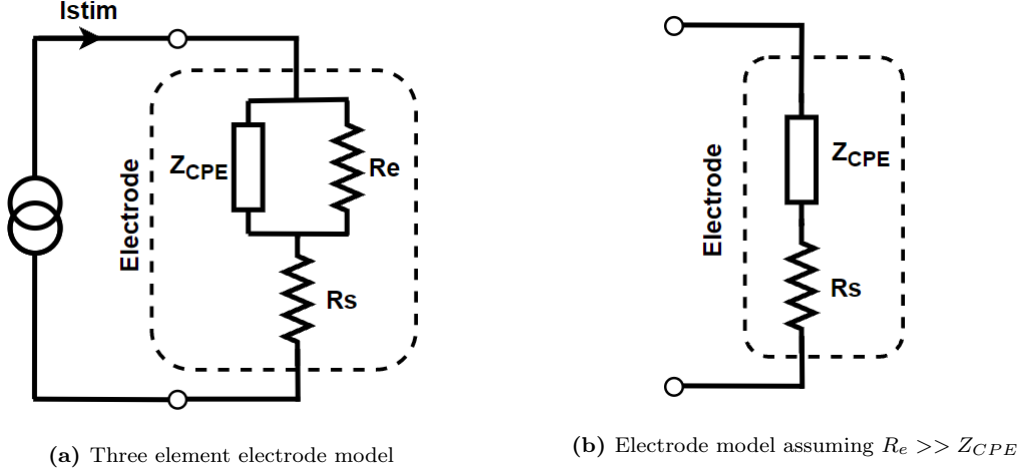


Figure 2.5: Tissue-electrode impedance models.

The total impedance of the electrode, Z_e can thus be expressed by the expression

$$Z_e \approx R_s + Z_{CPE} \quad (2.1)$$

In the s-domain, the constant phase element impedance is given by

$$Z_{CPE}(s) = \frac{1}{As^\alpha}, \quad 0 \leq \alpha \leq 1, \quad (2.2)$$

where A and α are constants of the model. If $\alpha = 0$, the constant phase element is equivalent to a pure resistance of value $1/A$, whereas if $\alpha = 1$ then it is equivalent to a pure capacitance of value A . The response to a step input current, $U(s)$, is given by

$$V_{step}(s) = Z_e(s)U(s) = \frac{Z_e(s)}{s} \quad (2.3)$$

Taking the inverse laplace transform, we find the continuous-time representation

$$V_{step}(t) = \mathcal{L}^{-1} \left\{ \frac{R_s}{s} \right\} + \mathcal{L}^{-1} \left\{ \frac{1}{As^{\alpha+1}} \right\} \quad (2.4)$$

$$= \left(R_s + \frac{1}{A} \frac{t^\alpha}{\Gamma(1+\alpha)} \right) u(t) \quad (2.5)$$

The discrete time step response, at sample n , given a sampling period T_s and step current magnitude I_0 is as follows

$$V_{step}(nT_s) = I_0 R_s + \frac{I_0}{A} \frac{T_s^\alpha}{\Gamma(1+\alpha)} n^\alpha \quad (2.6)$$

where Γ is the mathematical gamma function. Extending this, we can represent an arbitrary input current waveform as a series of current steps, as demonstrated by Fig. 2.6.

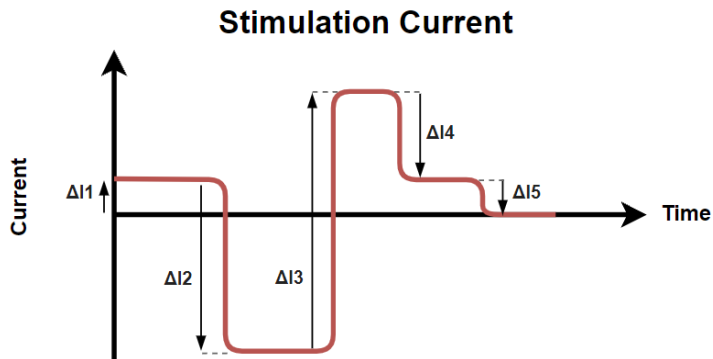


Figure 2.6: Arbitrary current signal represented as a function of discrete steps.

Given N current steps ΔI_1 to ΔI_N at samples n_1 to n_N , the electrode voltage at sample m can be expressed by the following equation, provided that $n_N \leq m$ (i.e. only prior steps contribute).

$$V(m) = I(m)R_s + \sum_{i=1}^N \left\{ \frac{\Delta I_i}{A} \frac{T_s^\alpha}{\Gamma(1 + \alpha)} (m - n_i)^\alpha \right\} \quad (2.7)$$

For a fixed sampling rate, this can be simplified to

$$V(m) = I(m)R_s + \sum_{i=1}^N \{ \Delta I_i K (m - n_i)^\alpha \}, \quad (2.8)$$

where

$$K = \frac{1}{A} \frac{T_s^\alpha}{\Gamma(1 + \alpha)}$$

Therefore, the parameters required to define the behaviour of the electrode model are R_s , K and α . These electrode model parameters can be generated in either the time domain, using impulse response characteristics; or the frequency domain, with Electrochemical Impedance Spectroscopy (EIS) measurements. Use of either method is dependent on the particular application. For example, in some cases it may not be possible to produce a sinusoidal waveform with sufficient amplitude, resolution or width to conduct reliable EIS, so low amplitude pulses are used for identification instead. On the other hand, current generation may be insufficiently stable to produce a reliable impulse characteristic, so EIS is preferred. For comparison, we implemented both methods, and the results are discussed below.

Time Domain (Impulse) Modelling

To generate the electrode parameters α , R_s and K from time-series data, it is possible to fit Eq. 2.8 to a simple current step response. If it is feasible to measure the voltage at the electrode directly, then this is a trivial process, and the exact method to achieve this is discussed in the publication by Lario-Garcia [85]. However, low-power implantable devices often do not have a low-noise wide-band AFE.

For example, the Intan RHS2116 has an input bandwidth of 0.1-20 kHz , but to remove excess noise the bandwidth is limited to 100-15 kHz . As such, the recorded voltage is distorted by the AFE transfer function, making direct fitting unfeasible.

The first step to overcoming this problem is to generate an impulse response representation of the front-end, denoted here as $h_{AFE}[n]$. This may be based on a known filter transfer characteristic (e.g butterworth) or measured by driving a known current input $x[n]$ across a fixed resistance and measuring the resulting (resistance normalised) voltage $y[n]$. With known input and output, the impulse response can be calculated with

$$h_{AFE} = (X'X)^{-1}X'y,$$

where X is the convolutional matrix of the input current signal $x[n]$. An example response of one attenuated channel is depicted in Fig. 2.7. While this is a computationally heavy method, it should only need to be done once during device commission, since the AFE transfer should remain relatively constant over time. It should be noted that the inaccuracies in the current input step directly translate to inaccuracies in the model. This issue is discussed further in Section 4.2.1.

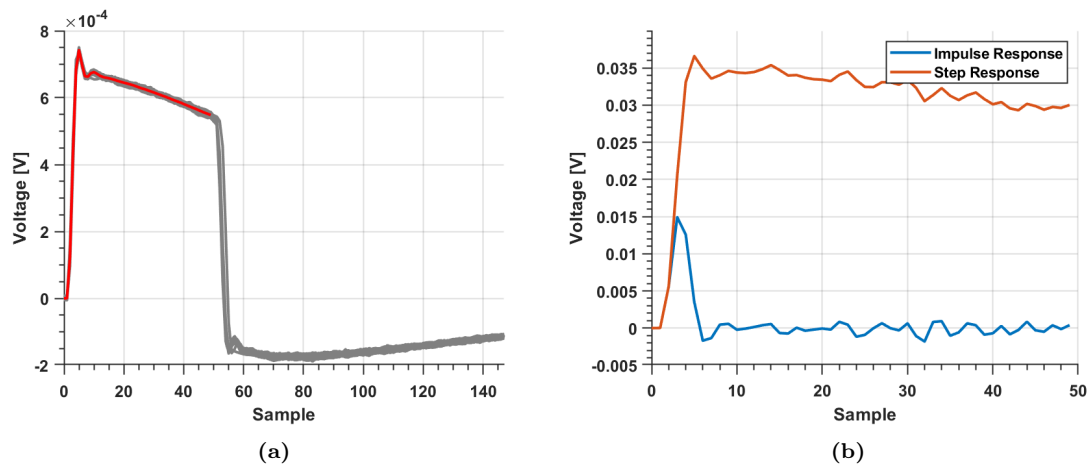


Figure 2.7: Example measurement of the impulse response for the channel 6 attenuated front-end: (a) 20x 500nA 500 μ s pulse recordings (gray) and signal average (red), (b) calculated normalised impulse response ($T_s = 10\mu$ s).

Once the impulse response of each channel is known, the model step response $d[n]$ (calculated from Eq. 2.8) can be convolved with this and equated with the step response of the electrode in solution as recorded by the amplifier $s[n]$. While there are possible deterministic approaches to generating the model parameters based on a known input, output and AFE impulse response; uncertainty in the front-end and DAC current characterisation make this an impractical approach. Instead, an iterative method is used.

Starting with parameters $R_S = 0$, $K = 0$ and $\alpha = 0.5$, Eq. 2.8 is used to generate the corresponding model step response, $d[n]$. This is then convolved with h_{AFE} to generate predicted voltages for three evenly-spaced intervals $y_p[n_1]$, $y_p[n_2]$ and $y_p[n_3]$. The choice of n_1 , n_2 and n_3 depends on the sampling rate, and should

be chosen so that the dynamics of $s[n]$ are sufficiently covered. R_s is scaled such that the first predicted output value $y_p[n_1]$ equals $s[n_1]$ (Fig. 2.8a). Next, K is incremented/decremented iteratively until $y_p[n_2]$ is close to $s[n_2]$ (Fig. 2.8b), and finally α is also adjusted until $y_p[n_3]$ is close to $s[n_3]$ (Fig. 2.8c). For each parameter change, the corresponding value of $y_p[n]$ must be recalculated. The process is repeated until all three values are within the allowable error margins. Fig. 2.8 shows an example of the iterative fitting approach.

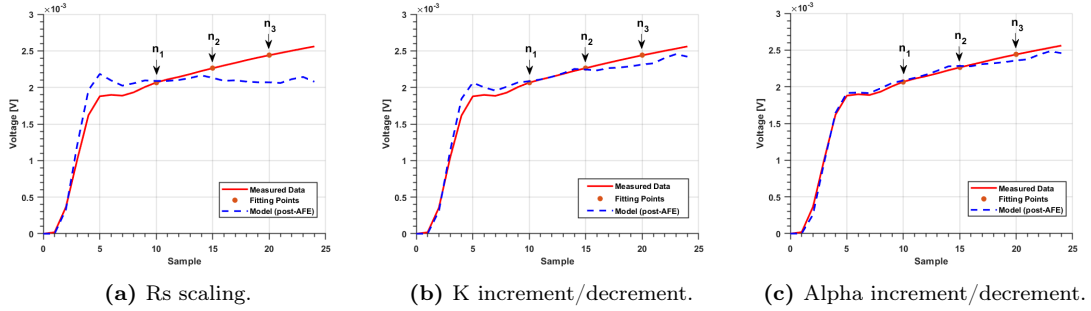
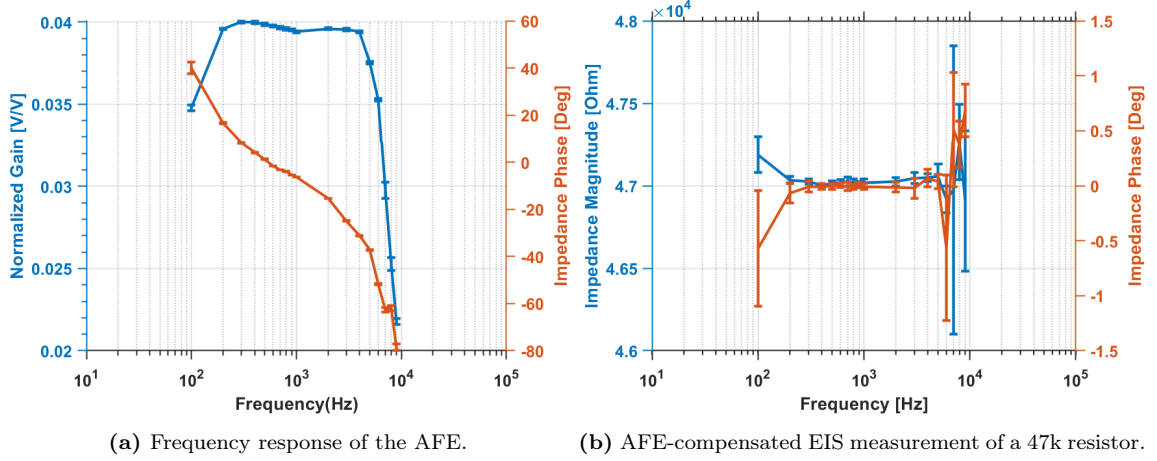


Figure 2.8: Iterative impulse-based model fitting process.

Dynamic adjustment of the K and α parameter step sizes were implemented to significantly reduce the computational cost of the fitting process. Also, since only three measurements are required, down-sampling was used to reduce the cost of the convolution steps. Additionally, the overall performance of the model was improved by repeating the fitting process at multiple current amplitudes, and taking the mean of each resulting parameter. The performance of this method is summarised in Section 3.2.

Frequency Domain (EIS) Modelling

An alternative method is to fit the model to measured data in the frequency domain. This can be done by performing Electrochemical Impedance Spectroscopy (EIS) to measure the impedance magnitude and phase spectra. As for the time-domain approach, it is required to first characterise the front-end response such that it can be accounted for during the fitting process. For this, we measure the AFE transfer function while applying a sinusoidal current input (same as used for EIS) to a linear resistor at its input. Fig. 2.9a demonstrates the frequency response of an attenuated channel front-end. In the frequency domain, it is trivial to account for the AFE transfer to refer measurements to the input of the AFE; the magnitude and phase are divided and subtracted from EIS measurements, respectively.



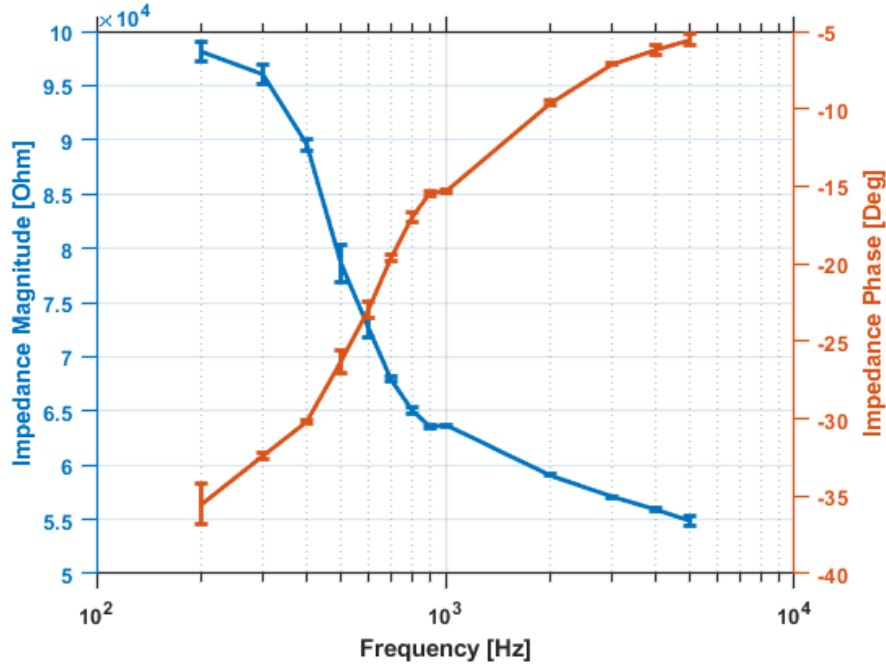
(a) Frequency response of the AFE.

(b) AFE-compensated EIS measurement of a 47k resistor.

Figure 2.9: Mean/standard deviations of EIS measurement process for 100-9kHz sweep (N=10).

Fig. 2.9b depicts an example front-end compensated EIS measurement of a 47k resistor. The details of the the acquisition of the EIS data is not particularly important, since there are many well-established on-chip EIS methods. Though unused for this work, even the Intan RHS2116 has in-built EIS measurement capability.

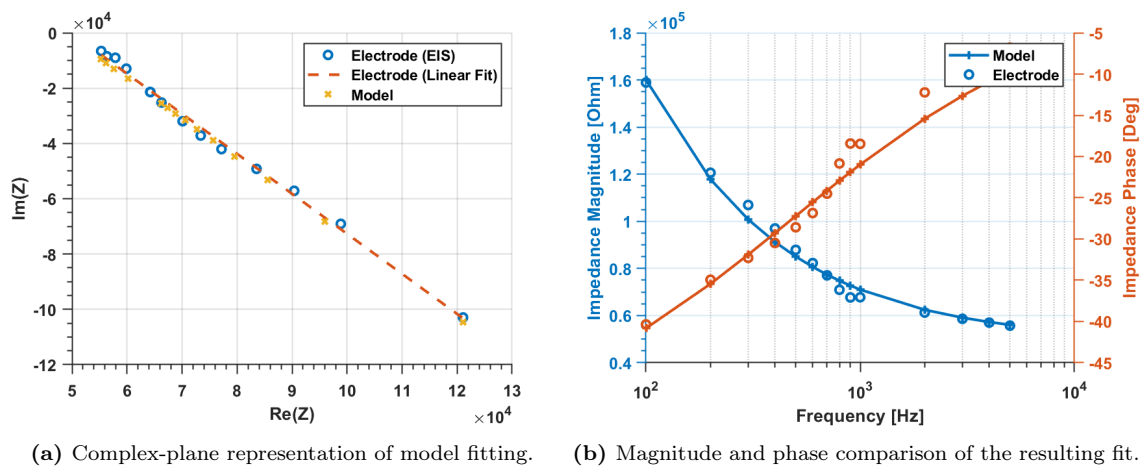
Fig. 2.10 shows an EIS sweep for the 30um gold/PU electrode that will be used fit our model.


Figure 2.10: Mean/standard deviation EIS Characterisation of the 30um Electrode for 100-9kHz sweep (N=10).

Recall, the electrode model can be expressed as

$$Z_e = R_s + \frac{1}{A S^\alpha} = R_s + \frac{1}{A(j2\pi f)^\alpha} \quad (2.9)$$

This function is linear in the complex plane, so we can make a linear fit of the electrode EIS data with respect to the real and imaginary components. The gradient of the model in the complex plane is independent of both R_s and A , so we conduct a binary search of α in the range $[0,1]$, calculating the gradient of the model impedance at each step, and equating that with the linear EIS fit. For our application, We need to resolve α to three decimal places, so we use 10 iterations to provide $1/1024$ precision. It is then trivial to calculate A , which scales magnitude of the model in the complex plane, and R_s which shifts the impedance model along the real axis. Fig. 2.11a depicts the recorded EIS data points in the complex plane, the linear fit of the EIS data and corresponding fitted model data points. Fig. 2.11b demonstrates the resulting fit with respect to impedance magnitude and phase.



(a) Complex-plane representation of model fitting. (b) Magnitude and phase comparison of the resulting fit.

Figure 2.11: EIS-based electrode model fitting.

2.2.2 Artifact Reduction

The model described above can be used to design stimulation waveforms that optimally reduce residual artifact duration.

Pre-requisites

Assessment of in-vitro stimulation efficacy of different waveforms is out of the scope of this work, therefore it is important to establish which parameters will and will not be changed such that a reasonable assurance of efficacy is maintained. We define the depolarising (anodic) phase of a stimulation waveform as the 'working' phase, since it is what typically elicits neural responses in retina (but the opposite can be implemented), and set the following guidelines for the artifact reduction algorithm:

1. The width of the working phase will not be altered.
2. The magnitude of the working phase will not be altered.
3. The working phase will be the only anodic phase.

This essentially means that any corrective shaping must lie to either side of the working phase, and that the corrective pulses must have positive magnitude. The implications of allowing alteration of the working phase for artifact reduction is discussed in Section 4.3.1.

Matrix Form

Equation 2.8 can be expanded and expressed as

$$V(m) = (\Delta I_1 + \dots + \Delta I_N)R_s + K(\Delta I_1(m - n_1)^\alpha + \dots + \Delta I_N(m - n_N)^\alpha) \quad (2.10)$$

This can then be arranged in matrix form, where M voltages at samples m_1, m_2, \dots, m_M can be calculated from N current steps at samples n_1, n_2, \dots, n_N .

$$\begin{bmatrix} V_{m_1} \\ V_{m_2} \\ \vdots \\ V_{m_M} \end{bmatrix} = \begin{bmatrix} R_s + K(m_1 - n_1)^\alpha & R_s + K(m_1 - n_2)^\alpha & \dots & R_s + K(m_1 - n_N)^\alpha \\ R_s + K(m_2 - n_1)^\alpha & R_s + K(m_2 - n_2)^\alpha & \dots & R_s + K(m_2 - n_N)^\alpha \\ \vdots & \vdots & \ddots & \vdots \\ R_s + K(m_M - n_1)^\alpha & R_s + K(m_M - n_2)^\alpha & \dots & R_s + K(m_M - n_N)^\alpha \end{bmatrix} \begin{bmatrix} \Delta I_{n_1} \\ \Delta I_{n_2} \\ \vdots \\ \Delta I_{n_N} \end{bmatrix}$$

given that for any i, j , $m_i \geq n_j$, to assert causality. This matrix equation can be written as

$$V = Z\Delta I \quad (2.11)$$

It follows that in addition to calculating the voltages resulting from current steps, the matrix can be inverted to calculate the current steps required for an arbitrary voltage output.

$$\Delta I = Z^{-1}V \quad (2.12)$$

If $N = M$ (i.e Z is square), then it is possible to construct any arbitrary waveform. To apply this to artifact reduction, we simply need to calculate the current steps required to force the voltage to zero after the pulse. Fig. 2.12 describes this situation.

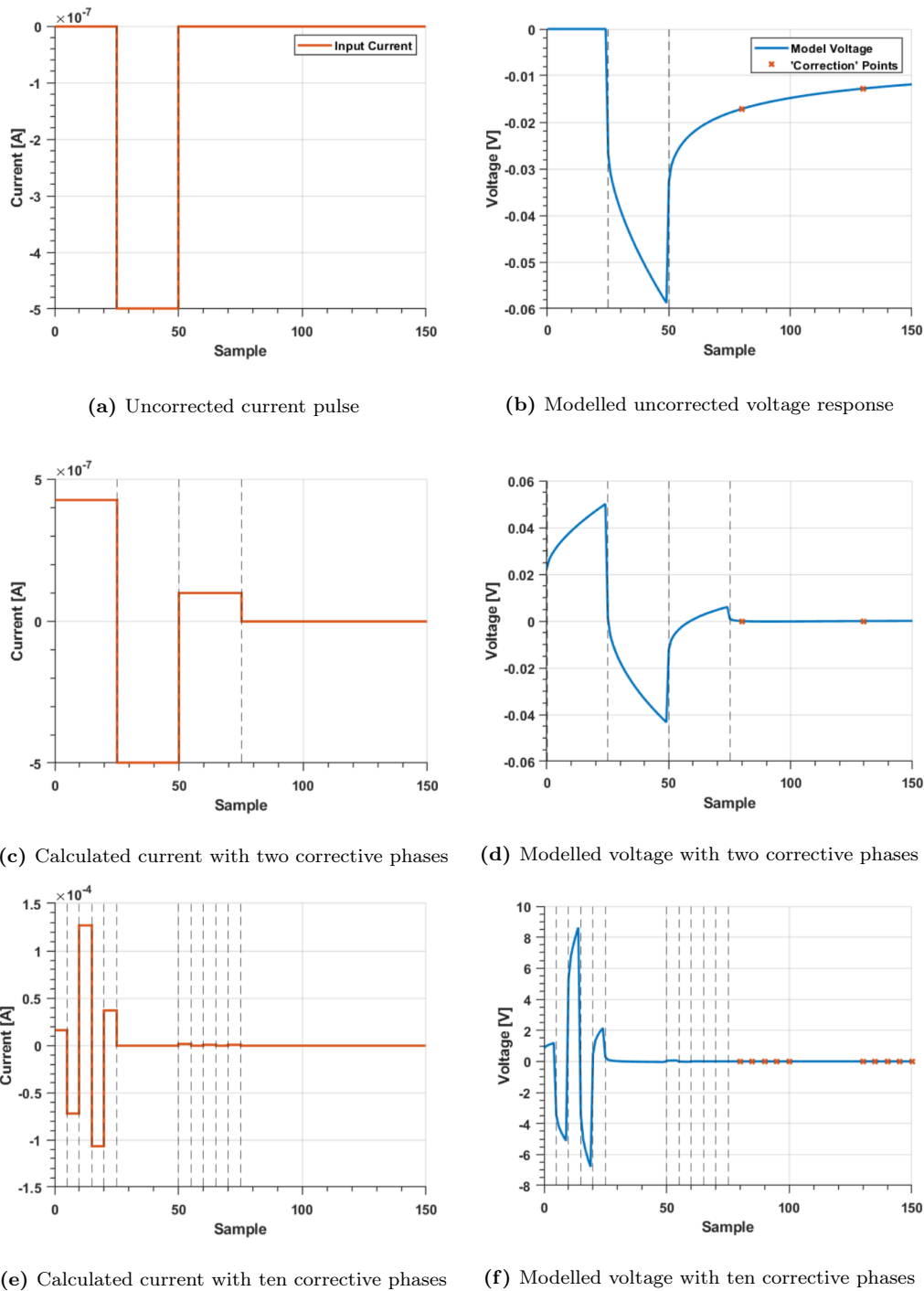


Figure 2.12: Matrix inversion for stimulation artifact reduction of a $-500nA$, $250 \mu s$ pulse ($T_S = 10 \mu s$).

Fig. 2.12a depicts a simple monophasic $-500nA$, $250 \mu s$ width current pulse, and 2.12b shows the corresponding predicted model voltage. Two 'corrective' phases are added of identical width before and after the working phase, resulting in two degrees of freedom (current steps) at $n_1 = 0$ ($t = 0 \mu s$) and $n_2 = 50$ ($t = 500 \mu s$) by which to correct for the residual artifact. The 'correction points' in Fig. 2.12b, calculated from the working phase, are the two points we want to force to zero, and will denote V_e . These points should be after the end of the pulse and are chosen by simply adding a delay of 80 samples from the correction inputs ($m_1 = 80$ $m_2 = 130$).

To calculate the required current of each correction phase, we solve the following matrix equation

$$\begin{bmatrix} -V_e(0) \\ -V_e(1) \end{bmatrix} = \begin{bmatrix} R_s + K(m_1 - n_1)^\alpha & R_s + K(m_1 - n_2)^\alpha \\ R_s + K(m_2 - n_1)^\alpha & R_s + K(m_2 - n_2)^\alpha \end{bmatrix} \begin{bmatrix} \Delta I_{n_1} \\ \Delta I_{n_2} \end{bmatrix}$$

As it is, solving this matrix would result in the correctly zeroed output, but to ensure that the working phase remains the same and that the current at the end of the waveform is zero, the added current must be removed again. We account for this by subtracting each current step at the end of the correction phase.

$$\begin{bmatrix} -V_e(0) \\ -V_e(1) \end{bmatrix} = K \begin{bmatrix} (m_1 - n_1)^\alpha - (m_1 - n_1 - 25)^\alpha & (m_1 - n_2)^\alpha - (m_1 - n_2 - 25)^\alpha \\ (m_2 - n_1)^\alpha - (m_2 - n_1 - 25)^\alpha & (m_2 - n_2)^\alpha - (m_2 - n_2 - 25)^\alpha \end{bmatrix} \begin{bmatrix} I_{n_1} \\ I_{n_2} \end{bmatrix}$$

Note that since the correction points are beyond the end of the pulse, the R_s terms have cancelled. Inverting the matrix and solving, we get the current depicted in Fig. 2.12c and the zeroed output 2.12d.

This is basically a model-based asymmetrical triphasic waveform. To produce a better result, we need to add more degrees of freedom, as depicted in Fig. 2.12e and Fig. 2.12f. However, this requires an inversion of a much larger matrix, which can be computationally expensive. Also, the nature of the model means that to force all points to zero very large oscillating currents are required that can dwarf the working phase current. This is obviously not implementable, has multiple anodic phases, and limiting the current would only result in sub-optimal performance. Since an explicit solution is not ideal, an iterative approach is considered instead.

Iterative Form

The iterative approach forces the residual artifact to zero one sample at a time, calculating each corrective phase sequentially. This avoids the need for large matrix inversions and allows for easier implementation of constraints such as quantization, current saturation and voltage limiting. Fig. 2.13 summarises the steps necessary for artifact correction of an anodic working phase (Fig. 2.13a) using 10 pre-correction phases (before the working phase), and 5 post-correction phases (after the working phase). Each correction phase is 20 μs long.

The pre-correction phases are calculated one at a time to force the long-term residual voltage to zero (Fig. 2.13b). Constraints on the maximum current output mean a voltage error may still remain after each step. However, subsequent phases are able to account for this in a way that the matrix approach was unable to. Calculation of the post-correction phases proceeds in a similar fashion to correct for the short-term residual artifact (Fig. 2.13c). As the corrective phases have mutual interaction, multiple cycles are required to achieve convergence (Fig. 2.13d).

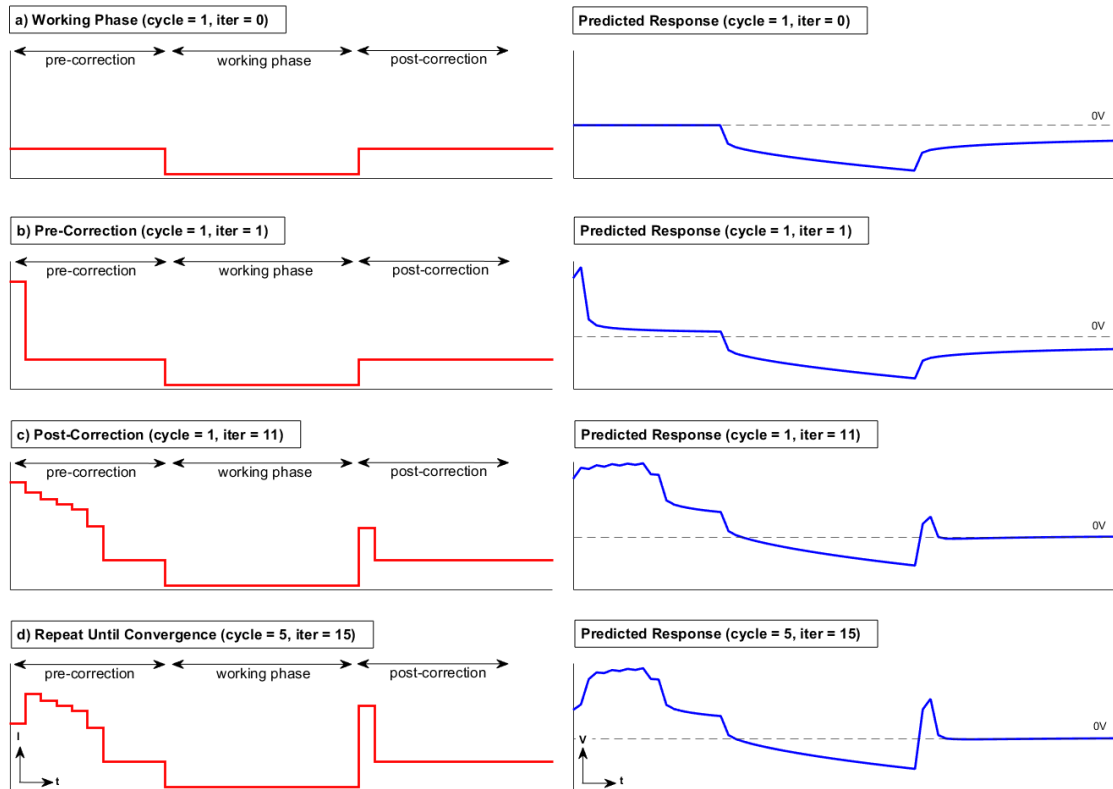


Figure 2.13: Steps in the iterative artifact correction algorithm for 10 pre-correction and 5 post-correction phases, indicating current (left) and predicted voltage artifact (right): (a) the uncorrected working (anodic) pulse, (b) pre-correction phases are added, (c) post-correction phases are added, (d) process is repeated until waveform converges.

It is often a desirable to constrain the interface voltage, since large voltages can result in faradaic processes and shorter implant lifetime. This is trivial to implement with the proposed scheme, since for each corrective phase, the modelled peak voltage at the end of the phase can be calculated. If the voltage exceeds a predetermined limit, then the corrective current amplitude is recalculated such that it is within limits.

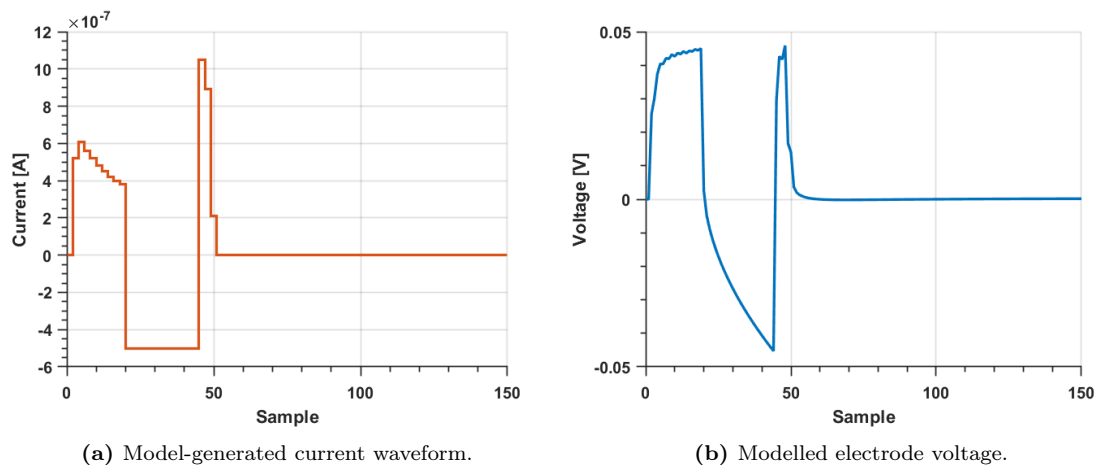


Figure 2.14: Sequential model-based stimulation artifact correction of a $-500nA$, 25 sample ($250\mu s$) pulse with a $0.05V$ upper voltage constraint.

Fig. 2.14 depicts an example optimised current waveform (Fig. 2.14a), and predicted voltage response (Fig. 2.14b) for a $-500nA$, $250\mu s$ anodic working phase, and a maximum voltage limit of $0.05V$. The reduction performance and computational cost of this method are detailed in Chapter 3. An example MATLAB implementation of the shaping algorithm is listed in Appendix A.2.

Trimming

Tissue-electrode non-linearities are excited as the voltage across the interface deviates from equilibrium. However, given the low currents required to excite single cells ($< 1\mu A$), the effects of these non-linearities should be minimal (this assumption is investigated in Section 3.1). As such, the linear model used above should be sufficient to characterise the general behaviour of the system. However, it is still inevitable that deviations in the model due to small non-linearities or inaccurate current output will result in a (small) residual error. To compensate for these errors, we introduce a 'trimming' step (Fig. 2.15). Intuitively, trimming provides a feedback path that effectively linearises the system.

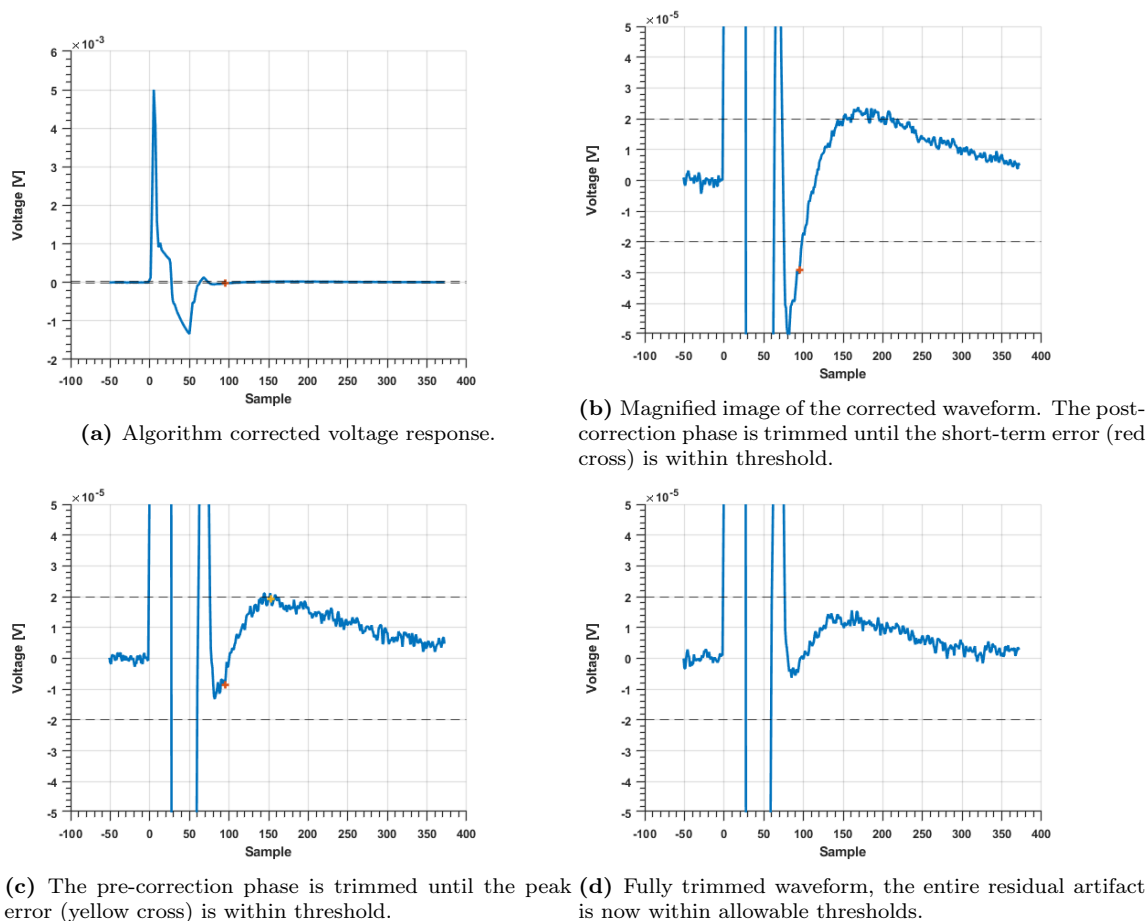


Figure 2.15: Example of trimming steps for corrected pulse.

First, short term errors are compensated for by trimming the magnitude of all post-correction phases up or down accordingly until the the voltage lies within pre-determined limits (Fig. 2.15b-c). Secondly, the maximum or minimum remaining residual voltage is identified and the pre-correction phases are trimmed up or down

to compensate until no point of the tail exceeds the thresholds (Fig. 2.15c-d).

Fig. 2.16 demonstrates correction of a -500nA , $250\mu\text{s}$ working pulse before and after the trimming process.

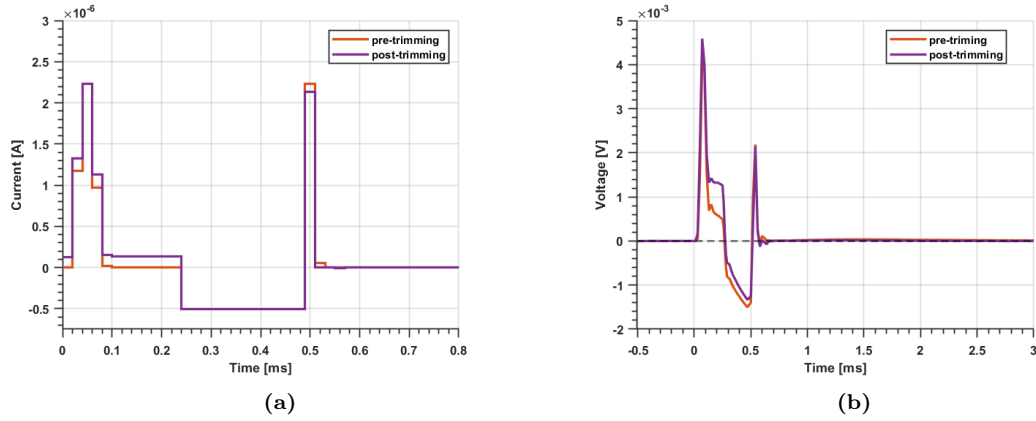


Figure 2.16: Example correction of a -500nA , $250\mu\text{s}$ working pulse before and after trimming: (a) stimulation current, (b) recorded voltage response.

Chapter 3

Results

In the following chapter, the performance of the model fitting methods and the stimulation shaping algorithm are presented. To prevent the unpredictable effects of amplifier saturation corrupting results, all tests are performed with the attenuated front end ($A \approx 1/20$). In an actual on-chip implementation, blanking or gain shifting would be used to prevent saturation from the direct artifact. The RHS chip can implement amplifier (output) blanking (called fast-settle [83]), which could be used to suppress the direct artifact and better simulate a real-world application. However, this feature resulted in unstable output even in the presence of no stimulation current, hence it was not used. This is discussed further in Section 4.2.2.

3.1 Artifact Characterisation

The proposed method assumes a linear, time-invariant (LTI) electrode-tissue interface. This is justified by the low stimulation current used ($< 1 \mu A$), which ensures mostly non-faradaic charge transfer. These assumptions have been validated experimentally.

3.1.1 Linearity

To assess artifact linearity, we use the same method as Chu et al. [30]. A $250 \mu s$, $200 - 1200 nA$ pulse is applied (Fig. 3.1a) to the electrode in solution and a normalised input-output relationship is determined (Fig. 3.1b). Using the maximum amplitude (A_0) pulse response (y_0) as the baseline, the input scale factor, $\alpha_{n,in}$ and output scale factor ($\alpha_{n,out}$) are calculated as

$$\alpha_{n,in} = \frac{A_n}{A_0}, \quad \alpha_{n,out} = \frac{\langle y_0, y_n \rangle}{|y_n|^2}, \quad (3.1)$$

where $||$ is the L^2 norm and \langle, \rangle is the inner product. We validated the linearity of our $30 \mu m$ electrodes by observing the relationship between $\alpha_{n,in}$ and $\alpha_{n,out}$, and plotting the percentage deviation from a nominal slope of 1 (Fig. 3.1c). The linearity of the output DAC was also verified as a baseline measurement, using a linear resistor as load.

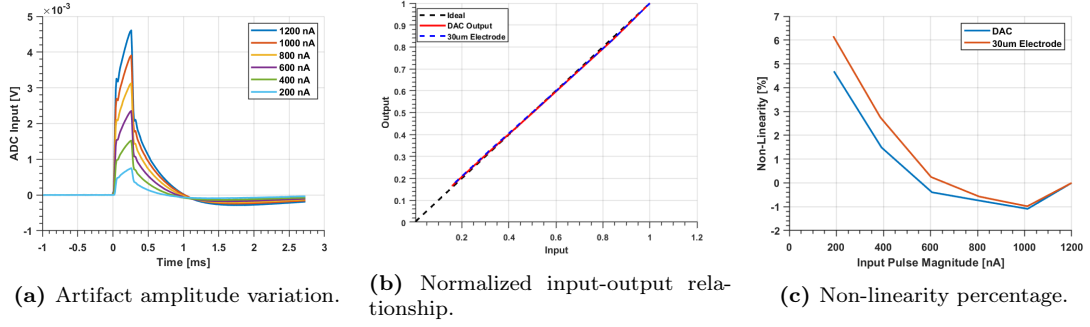


Figure 3.1: Artifact linearity assessment.

The artifact input-output relationship tracks that of the DAC very closely, indicating that the non-linearity of the artifact is minimal in the given range, and that the small deviation from ideal linearity is due to the DAC instead.

3.1.2 Additivity and Time-Invariance

Further important aspects of the LTI assumption are additivity and time-invariance. To test both of these, the response to a biphasic $250 \mu s$, $\pm 1 \mu A$ pulse is compared to the 'ideal' response, calculated by summing the responses to two shifted and inverted monophasic pulses of the same amplitude.

$$y_{bi,ideal}(t) = y_{mono}(t) - y_{mono}(t - 250\mu s) \quad (3.2)$$

The resulting waveform, when compared to the ideal calculation, is depicted in Fig. 3.2a. There is some variation in the tail of the artifact, which is likely due to small differences in the calibration of the positive and negative current sources (discussed further in Section 4.2.1).

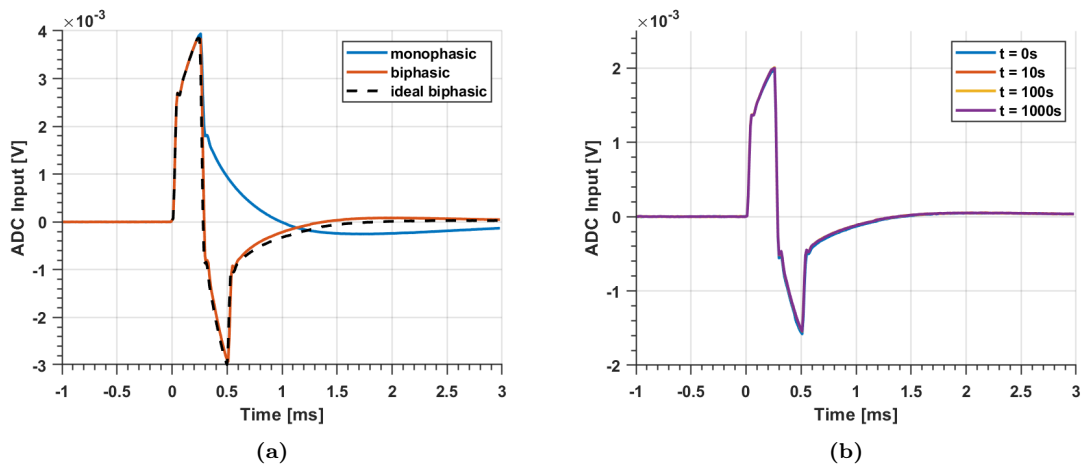


Figure 3.2: Artifact time invariance effects: (a) short-term time-invariance/additivity, (b) artifact variation for 0-1000s time range.

As discussed in 1.2.1, the long term impedance profile of an implanted electrode can vary greatly. Fig. 3.2b depicts the longer term stability of the electrode in PBS solution from 0 to 1000s. Each waveform is the average of 5 consecutive pulses to

negate noise effects. The artifact varies an insignificant amount within this time frame, which is expected as the electrochemical environment (temperature, concentration) should remain relatively constant. In a less stable environment (e.g. in-vivo), artifact stability could present a problem with a static algorithm, so this accentuates the requirement for the trimming step, which accounts for these small variations.

3.1.3 Spatial Variation

We also investigated the variation of the stimulation artifact from electrode to electrode (Fig. 3.3). The electrode response is recorded at the stimulating electrode, and at electrodes 240-840 μm from the stimulating electrode, in steps of 120 μm , for identical stimulation magnitude and pulse width.

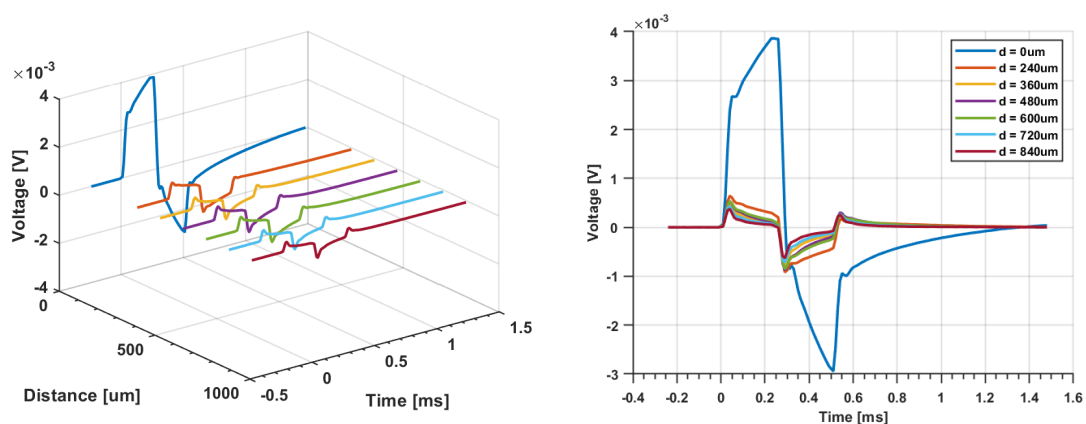


Figure 3.3: Stimulation artifact measured at the stimulating electrode ($d=0$) and surrounding electrodes ($d=240, 360, 480, 600, 720$ and $840 \mu m$).

The gradual slope of the decay between 240 and 840 μm is due to increasing tissue resistance, while most of the attenuation at neighbouring electrodes comes from the additional electrode impedance. It is important to note that the magnitude of the artifact at the non-stimulating electrodes 240 μm away is still $\sim 25\%$ that of the stimulating electrode. This suggests that artifacts at adjacent electrodes need to be taken into account when assessing the artifact reduction algorithm.

3.2 Artifact Modelling

We determine the accuracy of the proposed model in Section 2.2.1 by predicting the response to a randomly generated current waveform and comparing it to the recorded voltage (see Fig. 3.4a for the performance of both time-domain and frequency-domain approaches).

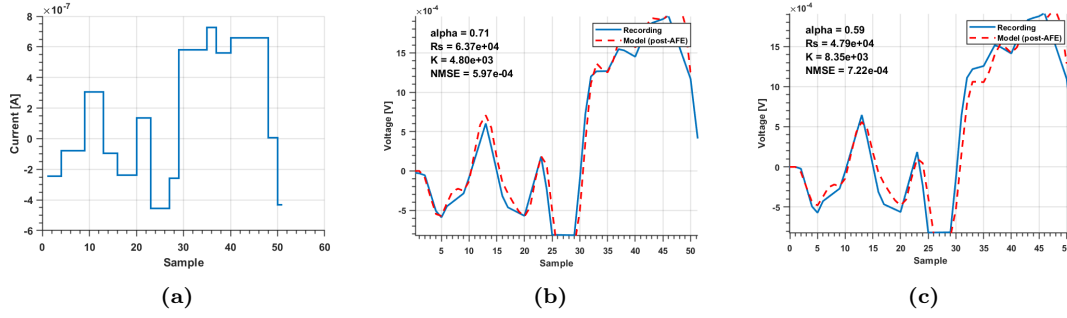


Figure 3.4: Example randomly generated input current testing for each model fitting method: (a) randomly generated input current, (b) impulse model fit recorded (blue) and predicted (red) output voltage, (c) EIS model fit recorded and predicted output voltage.

A 200 nA, 500 μ s step input was used for the impulse fitting procedure, and the frequency-domain method was implemented on EIS measurements from 700Hz to 5kHz. In both cases, the model successfully predicts the voltage at the electrode interface, as the calculated post-AFE voltage tracks that of the recording. We repeat the measurement, fitting and random stimulation tests 50 times for each method, at a rate of once per minute, recording the parameter spreads (Fig. 3.5a-c) and mean square voltage error (Fig. 3.5d) of the resulting prediction.

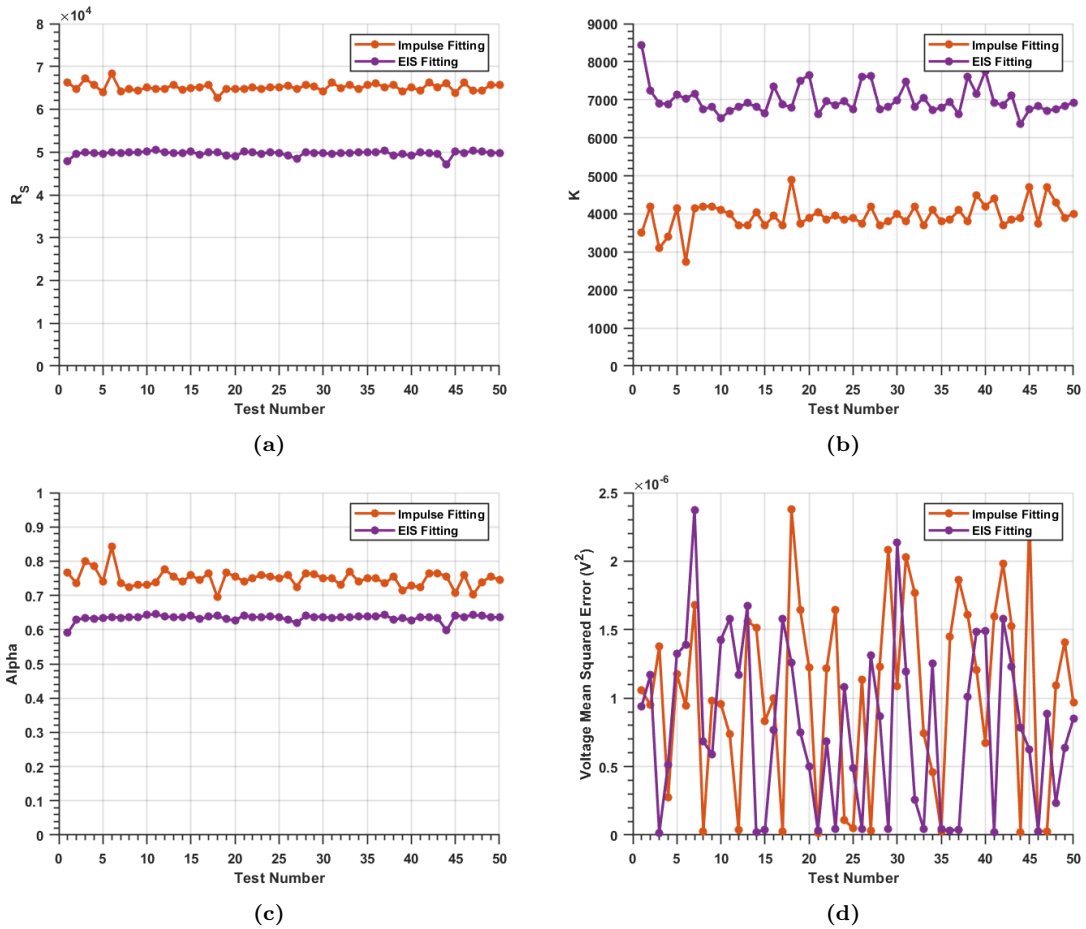


Figure 3.5: Repeat ($N=50$) impulse parameter fitting and random stimulation tests: parameter spread for (a) R_s , (b) K , and (c) α . (d) Mean-squared error between measured and predicted stimulation response.

The two methods appear to produce different parameter sets. This is likely because, with a bandwidth of 700-5kHz, the EIS method captures more of the low-frequency and less of the high-frequency behaviour than the impulse fitting procedure. Another interesting observation is that the frequency-domain fitted parameters appear to 'settle' after the first few measurements. It could be that the larger, low-frequency sinusoidal stimulation required for EIS alters the electrochemical conditions slightly. Despite the differences in parameters, the MSE performance of both methods are comparable.

3.3 Artifact Reduction

3.3.1 Test Criteria

The developed algorithm was tested for working phases of [100:100:500] nA and [50:50:250] μs for a total of $N=25$ tests. For each test scenario, (i) the artifact recovery time at the stimulating electrode (Fig. 3.6b), (ii) the recovery time at an adjacent electrode ($d=240 \mu m$, Fig. 3.6c), and (iii) the DC voltage accumulated after 50 consecutive pulses with an interpulse delay of 1 ms (Fig. 3.6d) are measured. The interpulse delay is set to account for the refractory period of neurons.

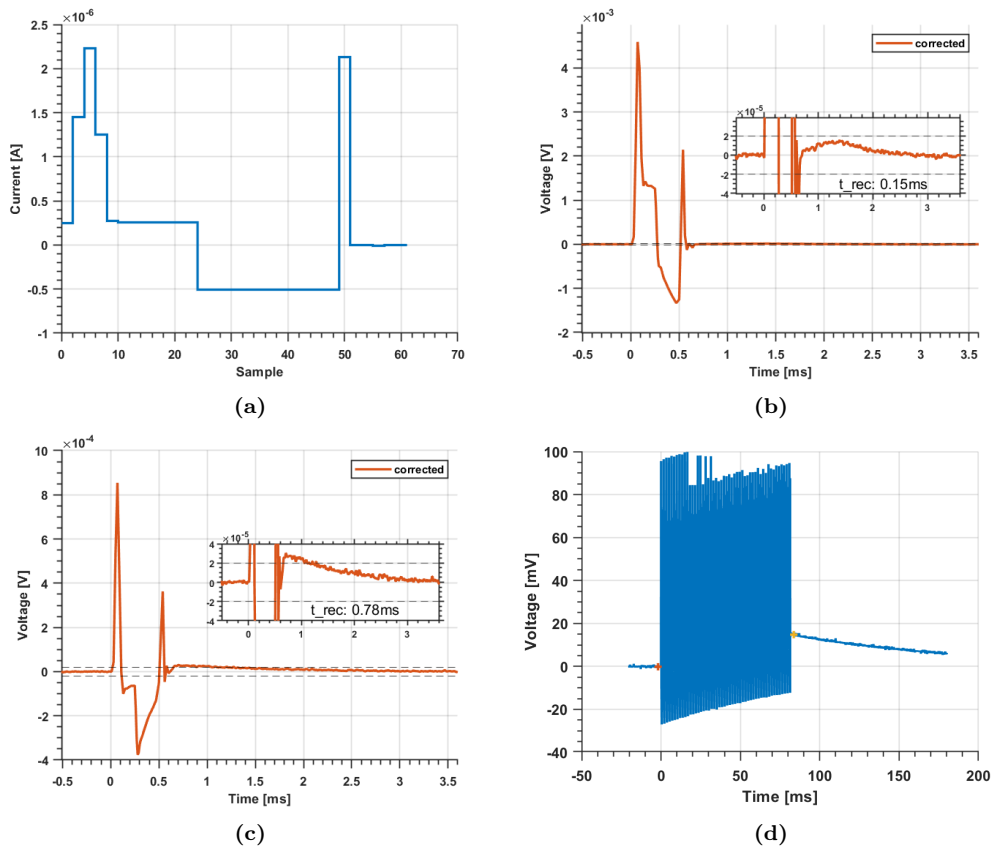


Figure 3.6: Assessment of model-based reduction algorithm: (a) generated current waveform, (b) evaluation of the correction algorithm at the stimulating electrode, (c) evaluation of the resulting waveform at the non-stimulating electrode, and (d) accumulated DC voltage after 50 consecutive pulses, measured with an external oscilloscope.

Recovery time is measured as the time from the end of the working (anodic)

phase, until the time that the residual voltage drops below $\pm 20 \mu V$ (approximately twice the input-referred peak-to-peak noise). The DC accumulated voltage is measured by connecting an external oscilloscope during the pulse train execution. The above measurements are repeated for both the trimmed and untrimmed correction waveforms. The correction algorithm parameters used for testing are detailed in Appendix A.3.

3.3.2 Comparative Methods

For comparison, the tests are also run using different waveform shapes and artifact reduction methods. As a baseline, we measure the recovery time for monophasic (Fig. 3.7a), anode-first biphasic (Fig. 3.7b), cathode-first biphasic (Fig. 3.7c) and triphasic (Fig. 3.7d) waveforms. The cathodic phase for the biphasic scenarios is of equal and opposite magnitude to the anodic phase, and the cathodic phases of the triphasic scenario are half the magnitude of the anodic phase, as conventionally used for charge balancing.

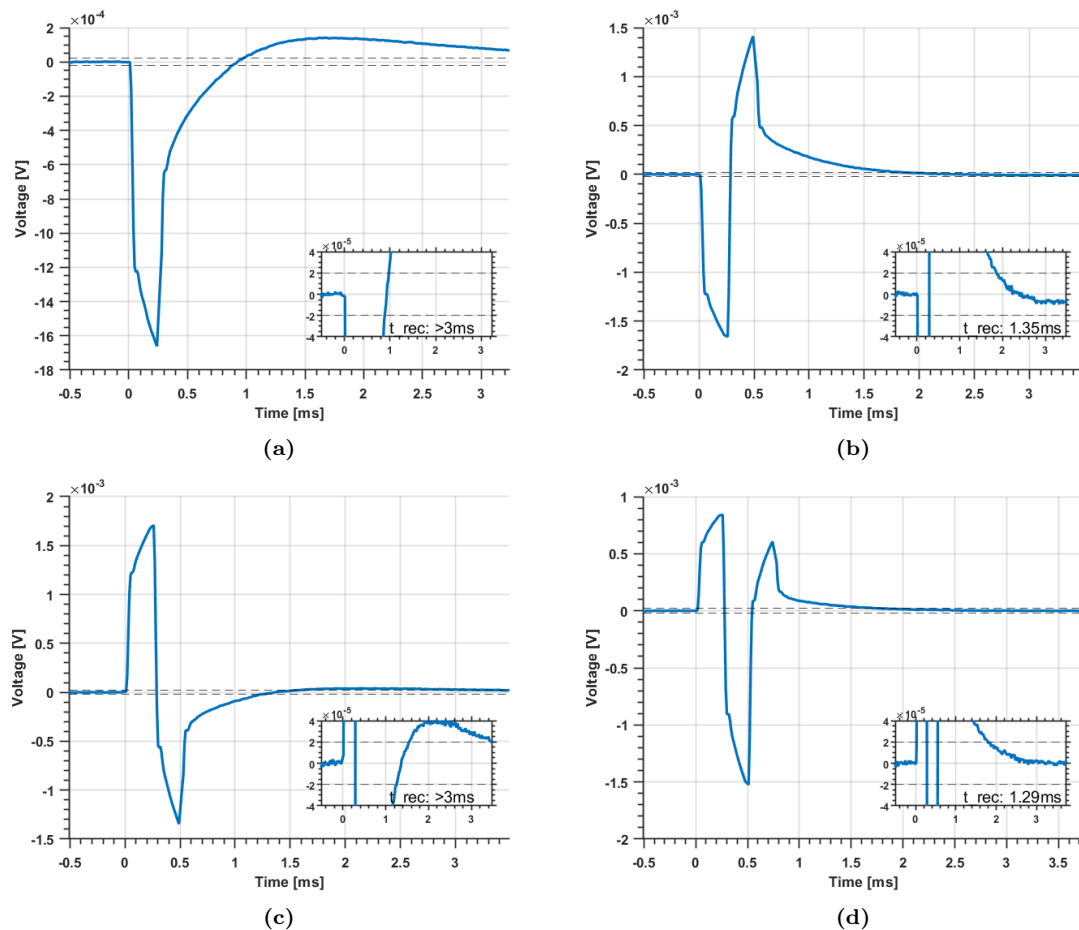


Figure 3.7: Example of standard comparative waveform voltage responses: (a) monophasic, (b) biphasic anode first, (c) biphasic cathode first and (d) triphasic. Inset figures demonstrate measurement of recovery time (from end of working anodic phase), where the dashed lines are the recovery thresholds.

Active-discharge recovery, in combination with a triphasic waveform, is also tested by making use of the 'charge recovery' functionality of the Intan IC, which

shorts the input to ground through a low impedance path ($\sim 1\text{ k}\Omega$) [83]. We adjust the discharge period for each test to identify the minimum possible recovery time (Fig. 3.8a), which is used for comparison (Fig. 3.8b).

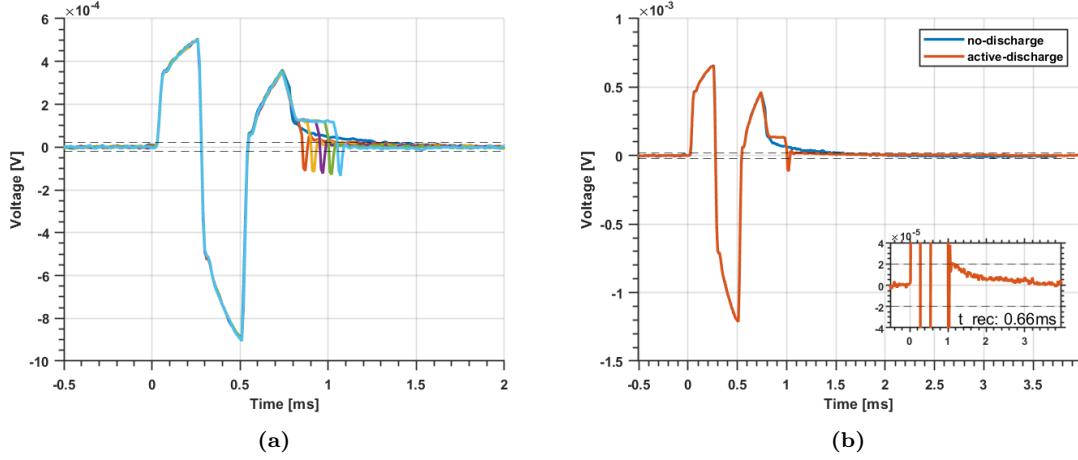


Figure 3.8: Active discharge comparative artifact recovery method: (a) multiple discharge periods are tested, (b) the discharge period of maximum reduction is selected for comparison.

3.3.3 Performance Comparison

Recovery Time (Stimulating Electrode)

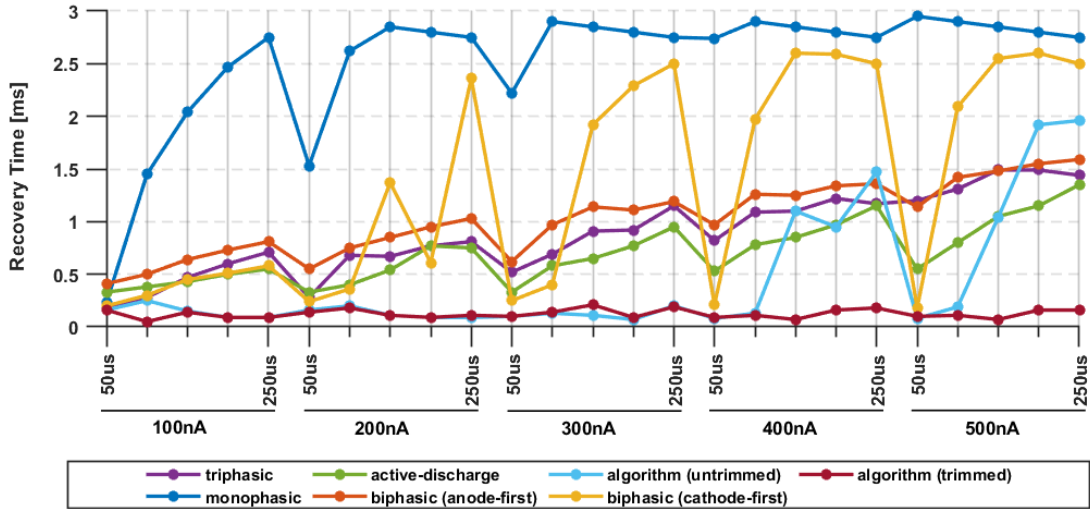


Figure 3.9: Recovery time from end of working phase, at the stimulating electrode, for each reduction strategy.

The untrimmed model-based algorithm presents a mean 50% improvement over the active-discharge recovery method and 55% improvement over the best traditional stimulation waveform (triphasic). The addition of a trimming step results in a 79% and 81% improvement over active-discharge and triphasic strategies respectively. The mean recovery time for the model-based control method, with trimming, is $124\ \mu\text{s}$. At low amplitudes, the untrimmed model successfully predicts and removes

the residual artifact, and little to no improvement is achieved by trimming. However, at larger amplitudes ($> 400 \text{ nA}$), the performance of the untrimmed model deteriorates significantly due to increasingly nonlinear behaviour. The trimming step successfully linearises the system and corrects for these errors.

Recovery Time (Non-Stimulating Electrode)

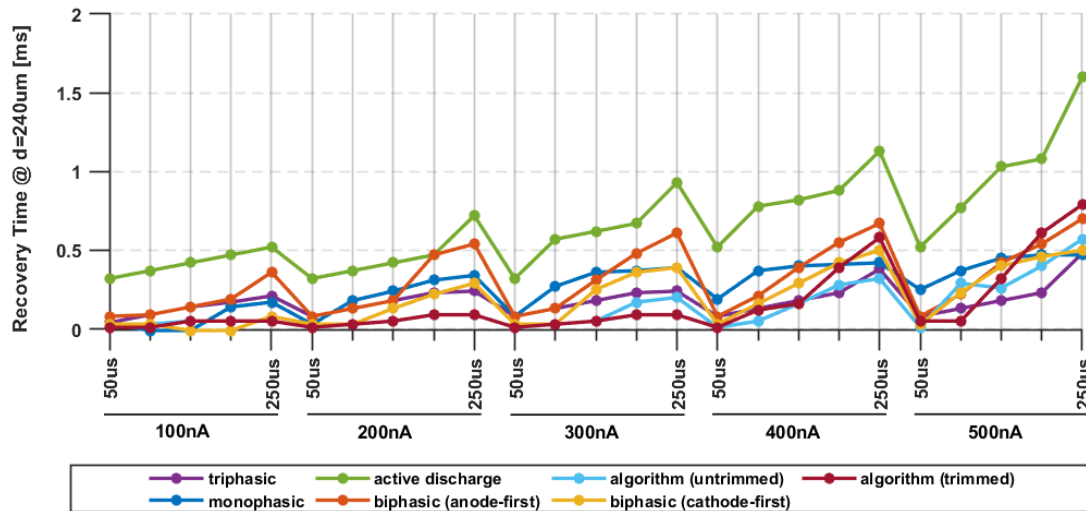


Figure 3.10: Recovery time from end of working phase, at the non-stimulating electrode ($d=240 \mu\text{m}$), for each reduction strategy.

At the non-stimulating electrode, the untrimmed algorithm is an average of 84% more effective than the active-discharge method, which is the worst performing method, and 39% better than the triphasic waveform. The trimming step degrades performance slightly at larger amplitudes, but on average it is still 83% and 34% better than active-discharge and triphasic methods, respectively. The mean trimmed recovery time at the non-stimulating electrode is $151 \mu\text{s}$.

Pulse Train Voltage Accumulation

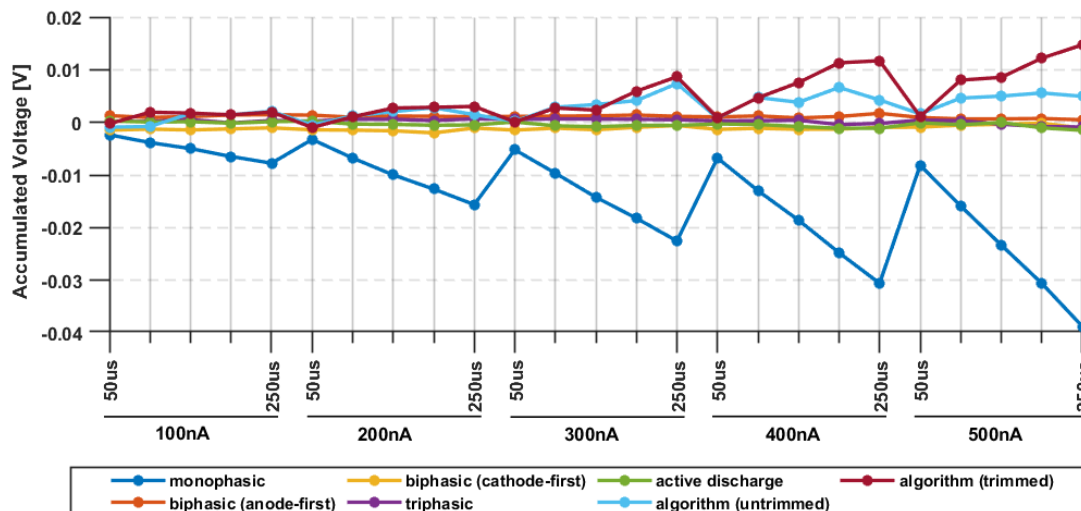


Figure 3.11: DC voltage accumulation after 50 pulses, with 1ms interpulse delay, for each reduction strategy.

The algorithm results in increased charge accumulation after a train of 50 consecutive pulses when compared to the other mitigation strategies. While still being an improvement over a simple monophasic waveform, the untrimmed and trimmed methods are 20% and 11% worse respectively when compared with the next best (triphasic) waveform. The maximum accumulation of the trimmed method is 14.7 *mV*. This is expected as we are minimizing the artifact at the output of the AC-coupled AFE, which is not necessarily the same as mitigating DC charge at the input of the AFE. This is explored further in Section 4.1.2.

3.4 Algorithm Cost

3.4.1 Chip Area

The proposed algorithm has minimal impact on chip area consumption of the recording front-ends, only requiring an additional blanking switch to avoid saturation due to the direct artifact. However, arbitrary current-mode waveform generation is required to perform stimulation shaping with moderate resolution (this work used the 8-bit DAC of the Intan chip). As an example, the design by Noorsal et al. uses 0.2 *mm*² for each DAC (multiplexed to 4 channels) [86]. Both of the parameter fitting methods will require a single high dynamic range (DR) recording front-end to measure step and sinusoidal voltage responses. However, as parallel model-fitting is not necessary, a single high-DR channel can be multiplexed to many electrodes. Both fitting methods will also make use of the existing arbitrary waveform DACs at no extra cost for the stimulator area.

3.4.2 Computational Cost

With respect to the model fitting algorithms, the frequency-domain (EIS) fitting process is significantly less computationally expensive than the time-domain (im-

pulse) method, provided EIS data is available. To achieve 10-bit precision for the calculation of α , only 10 binary-search iterations are required. The calculations of R_s and K then only require a single computation each. On-chip EIS technologies are prevalent, and while the assessment of EIS methods is out of the scope of this work, we assume that its implementation is not an issue. It does, however, take a minimum of $1/f_{min}$ seconds to complete the EIS measurement, which is a concern when scaling up to many electrodes (discussed further in Section 4.3.2). The impulse method is far simpler in terms of data acquisition, only requiring a single low-amplitude step current input of $250 \mu s$, but falls short in requiring accurate, high time-resolution knowledge of the current output waveform and the impulse response of the AFE, which are both difficult to acquire, especially for an integrated device. Additionally, the fitting process requires many step iterations/convolutions to be accurate, and while there is still room for optimisation, it will always require more complexity than the EIS method.

The most computationally expensive aspect of this algorithm is calculation of the stimulation waveform. The key metrics we use for measurement of the correction algorithm complexity are shaping iterations, which are how many sweeps we need to conduct for the stimulation shape to converge; and Model Voltage Calculations (MVCs), which are how many times the model parameters are used to calculate a single voltage from a single current delta (or vice-versa). A single MVC can be described by Eq. 3.3.

$$V = \Delta I(R_s + Kn^\alpha) \quad (3.3)$$

MVCs are significant as they involve calculating an integer n (number of samples) to the power of a fraction, α , which can be in the range 0-1. We can assume a maximum width of $n=100$ samples, and limit the precision of alpha to steps of 0.01, but that still may require computing $100^{1/100}$, which is non-trivial. Using Newton's method of calculating roots, to reach 1% precision, 65 iterations are required. However, to reduce computational cost, we can pre-compute and store a 100x99 lookup table (LUT) for $n = 1:100$ and $\alpha = 0.01:0.01:0.99$. Although floating point precision was used during testing, given the required accuracy, it is feasible to encode all parameters into 16-bits without significant loss of performance. Each MVC therefore requires a single LUT enquiry, two 16-bit multiplications and one 16-bit addition.

To understand how many MVCs are required per stimulation waveform shaping, we first measure the performance of the model at each shaping iteration (Fig. 3.12). For this we observe the residual mean-squared voltage prior to trimming, as this is a good assessment of both the shaping algorithm and prediction of the actual residual voltage. While up to 20 iterations are required to fully converge, as few as 5 iterations are needed to achieve minimal residue, and it is reasonable to assume that a lower residual voltage will require fewer trimming iterations to correct.

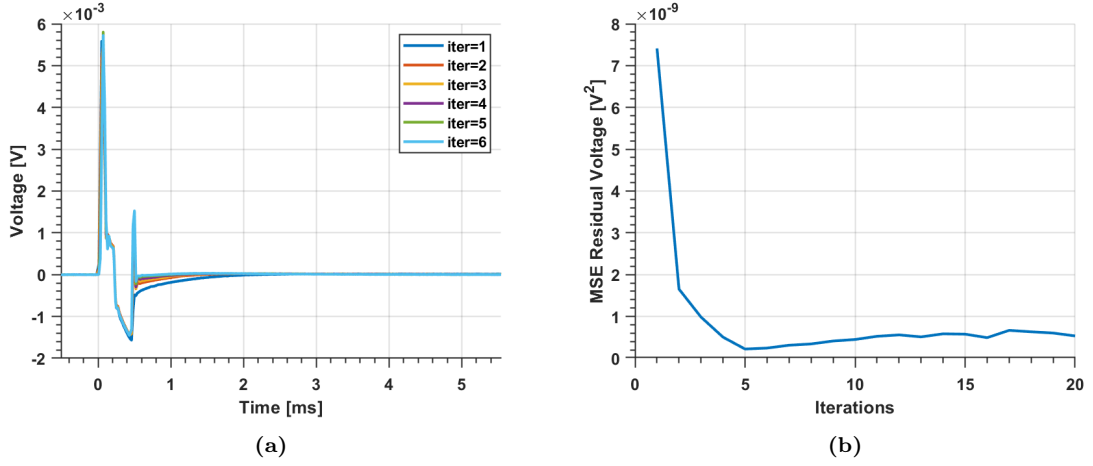


Figure 3.12: Relationship between shaping iterations and untrimmed performance for correction of a $500nA$, $250\mu s$ anodic pulse: (a) effect of increasing iterations on the voltage response, (b) MSE residual voltage with increasing shaping iterations.

To assess the performance over a range of operating parameters, we conduct Monte Carlo tests with α , R_S and K parameter variations of $\pm 10\%$. We count the number of shaping iterations until full convergence (Fig. 3.13a) and MVCs per iteration per correction phase (Fig. 3.13b), when no voltage limit is applied.

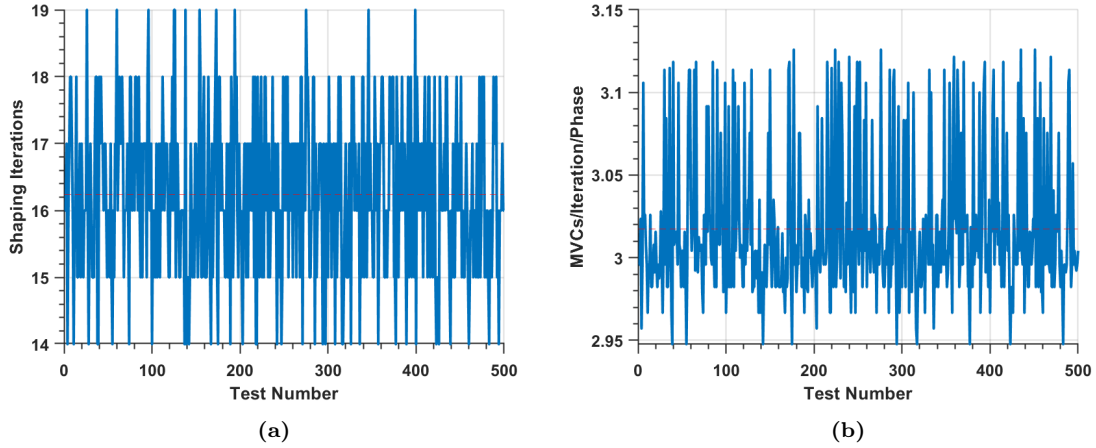


Figure 3.13: $\pm 10\%$ Monte Carlo ($N=500$) model parameter variation for $500 nA$, $250 \mu s$ anodic pulse with no-voltage limit: (a) shaping iterations ($\mu = 16.24$), (b) Model Voltage Calculations (MVCs) required per iteration per correction phase ($\mu = 3.018$).

Fig. 3.14 demonstrates the case where a $100mV$ limit on the model voltage is enforced. While it does not take longer to converge, since it has reduced degrees of freedom, significantly more MVCs are required to determine the peak voltage at the end of each phase.

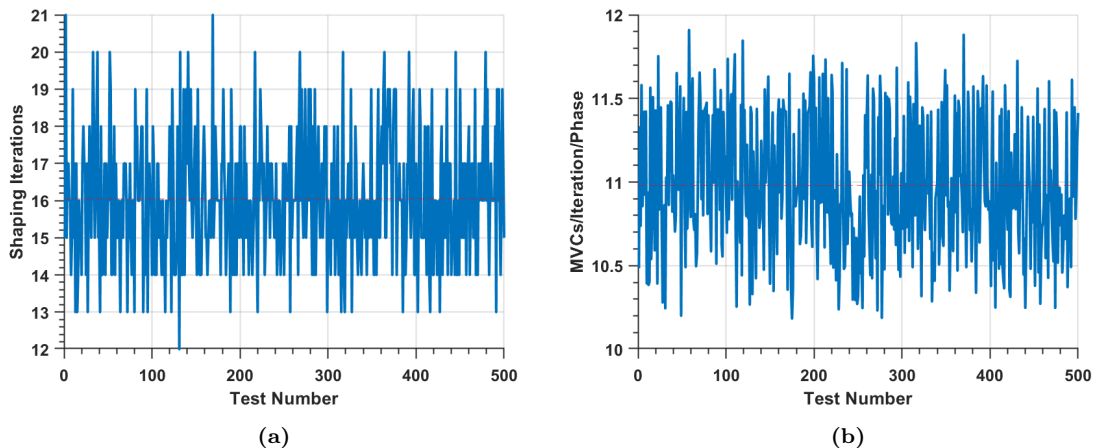


Figure 3.14: $\pm 10\%$ Monte Carlo ($N=500$) model parameter variation for 500 nA , $250\text{ }\mu\text{s}$ anodic pulse with a 100mV limit: (a) shaping iterations ($\mu = 16.05$), (b) Model Voltage Calculations (MVCs) required per iteration per correction phase ($\mu = 10.98$).

Given 10 pre-correction phases, 5 post-correction phases, and 5 shaping iterations to achieve minimal residue (from Fig. 3.12), to generate the optimum stimulation waveform, the algorithm requires ~ 225 MVCs (5 iterations \times 15 phases \times 3 MVCs/phase) for the unconstrained case, and ~ 825 MVCs (5 iterations \times 15 phases \times 11 MVCs/phase) for the voltage constrained case.

The trimming step is of minimal computational cost as it can be implemented as a simple up-down counter, adjusting the pre- and post-correction phases as necessary, depending on the recorded residual artifact voltage.

3.4.3 Memory Requirements

For the following section, we once again assume that every key parameter or measurement can be encoded into 16bits (2 bytes). EIS fitting requires ~ 10 frequency points, each consisting of a phase and magnitude, for both the electrode and front-end, utilising 40 values (80 bytes) in total. Impulse fitting requires a 25 sample step response for the AFE and test pulse, corresponding to 50 values (100 bytes) of memory for each channel. However, for both of the above methods, the measurements need only be stored temporarily in RAM while the model is calculated, so are not considered a permanent memory cost.

Three resulting model parameters are stored for each channel. Additionally, given 10 pre-correction phases, 5 post-correction phases and one working phase, the algorithm requires storage of 16 phase magnitudes and widths (64 bytes) for every working phase configuration.

Chapter 4

Discussion and Conclusions

4.1 Algorithm Performance

4.1.1 Model Parameter Fitting

As observed in Section 3.2, the frequency-domain and time-domain model fitting strategies result in differing α , K and R_S parameters. However, we also observed that both the EIS and impulse methods presented similar mean-squared error values when predicting the output of random current stimulation waveforms. This implies that the relationship between the parameters, rather than the parameters themselves, are of more importance. Since both methods produce similar results, it is largely up to preference which method is selected for use. However, the increased computational cost of the time-domain method along with the necessity to accurately characterise the current output and the AFE step response make the frequency-domain method preferable in most cases.

4.1.2 Artifact Reduction

The untrimmed algorithm performs well when compared to traditional stimulation waveforms and the active-discharge artifact recovery method. However, its performance deteriorates as the amplitudes increase. This is most likely due to the fact that higher voltages result in more non-linear effects and less accurate balancing as more charge is lost to faradaic processes. The trimming step, which corrects for these small differences, effectively linearises the system and results in consistently low (sub-250 μ s) recovery times across all tests, outperforming all other methods. This suggests that the trimming process is crucial in any real-world implementation.

At the non-stimulating electrode we observed similar, or slightly better, recovery times than that of the other methods. While we did not observe large improvements, this serves to confirm that reducing the artifact at the stimulating electrode does not degrade performance at the non-stimulating electrode, which is a very important aspect to confirm before implementation in massively parallel high density arrays.

It is necessary to state that the results presented in this work are for a single set of hand-crafted shaping parameters (correction phase widths, correction phase counts, correction points etc.). A systematic design-space exploration could result

in better performance, especially if higher time and current resolutions are possible.

4.1.3 Trimming

While effective at correcting for residual error in all test cases, the trimming method employed in this work had some notable issues. Firstly, the pre- and post-correction phases only incremented or decremented in unit steps. This meant that correction for larger magnitude pulses, with larger errors, could take upwards of 20 iterations. While this is a short settling time for a device that will be stimulating each channel many thousands of times per day, convergence could be improved significantly by implementing a smarter step algorithm (e.g. binary search) and trimming both the pre-correction and post-correction phases simultaneously.

4.1.4 Voltage Accumulation

In regards to voltage accumulation, our method performed significantly worse than the other comparative methods. There are a number of different possibilities as to why this is the case. The larger induced voltages could result in more nonlinear effects and faradaic charge losses, but it is more likely that correcting for the artifact at the output of the AC-coupled AFE does not necessarily mean correcting for the artifact at the tissue-electrode interface. In the act of trimming, we are accounting for the impulse response of the front end, which may result in an unbalanced waveform. Fig. 4.1 confirms this latter hypothesis by comparing the AC and DC recorded artifacts before and after trimming. Fig. 4.1 confirms this latter hypothesis by comparing the AC and DC recorded artifacts before and after trimming.

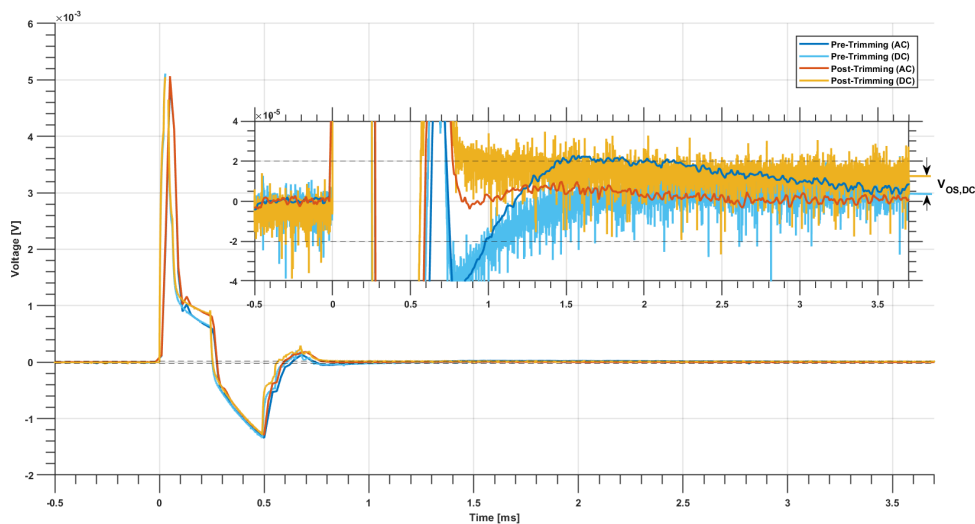


Figure 4.1: Comparison of AC (on-chip) and DC (oscilloscope) voltage recordings for a corrected -400 nA , $250 \mu\text{s}$ pulse before and after trimming. Inset figure is a zoomed-in view along the x-axis.

Before trimming, the DC voltage (light blue) returns close to zero, indicating correct charge balancing, whereas the AC artifact (dark blue) takes some time to settle. However, after trimming, the AC residual artifact (red) is now minimised, whereas the DC voltage (yellow) has accumulated a notable positive offset ($V_{OS,DC}$ in figure). The high-pass filter pole of the AC amplifier eliminates the offset at the

output of the ADC.

Although we get some charge accumulation, removing the artifact as quickly as possible from the input is much more important for our application. Recall from Chapter 1 that healthy retinal ganglion cells are excited at a very low rate (<10 Hz), so high frequency pulse trains will not be used in an implant. Even if they were used, with sub- μA currents, the accumulated voltage is still less than 15mV, which is well below safe levels to prevent harmful effects [10].

4.1.5 Susceptibility to Parameter Error

It was observed that the fitted model parameters can deviate by a small amount when using consistent measurement methods. We need to ensure that deviations in the electrode parameters will not result in poor artifact removal. Using Monte Carlo analysis, we observe the pre-trimming accuracy of the model-based algorithm for parameter deviations of up to $\pm 10\%$. The predicted model voltage fluctuates during stimulation for different parameters, but the residual component remains very stable (Fig. 4.2a). This makes sense, since with no current, the R_S component disappears, and K simply scales the residual component, so it has no effect on the artifact's length. Only α affects the decay rate. This is confirmed in the actual AC voltage response (4.2b) where minimal variation is observed. Finally, in Fig. 4.2c the mean-square residual voltage does not vary significantly from case to case. However, there is a notable increase after the first tests, the cause of which is unclear, but may be the result of stimulation-induced variations in steady-state electrochemical conditions due to insufficient settling time between tests.

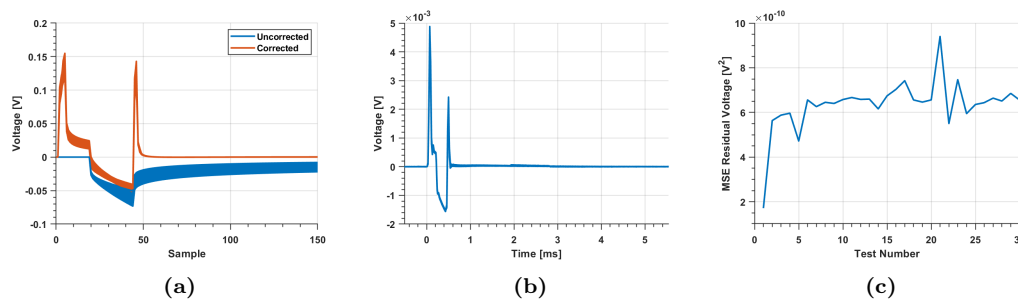


Figure 4.2: Monte Carlo $\pm 10\%$ variation in model parameters: (a) effect on the calculated waveform, (b) the recorded voltage, and (c) the mean-squared residual voltage.

4.2 Test System Limitations

4.2.1 DAC Current Calibration

The impulse-based method for model-fitting relies on an accurate characterisation of the AFE impulse response. Measuring this is not possible unless the test current pulse is sufficiently square, or well calibrated. For the RHS2116, the DAC output was nonlinear. When tested with a sinusoidal stimulation current and a linear resistor as load (Fig. 4.3), notable glitches in the sinewave can be seen in the time domain (Fig. 4.3a), likely due to changes in operating region. These appear as harmonics in the frequency domain (Fig. 4.3b).

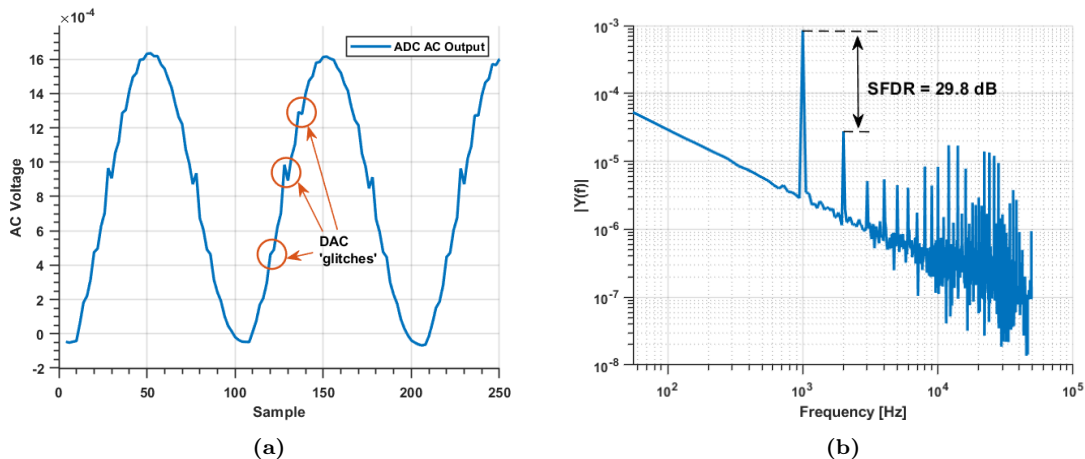


Figure 4.3: Channel 7 (a) voltage (b) FFT response to 1kHz $1.25 \mu\text{A}$ PDAC sinusoidal current stimulation across a $1 \text{ k}\Omega$ resistor (SFDR = Spurious Free Dynamic Range).

One method of overcoming this (with a limited bandwidth amplifier) is to calibrate the charge injection of the DAC for different pulse widths and amplitudes by stimulating across a capacitor of a known value (Fig. 4.4a). A lookup table of 'equivalent' delivered current is created (Fig. 4.4b), which is used as the basis of all DAC code to current conversions in this work. This may not be practical for implementation on-chip, where a precision calibration method might be preferred, as in the work by Greenwald et al. [21].

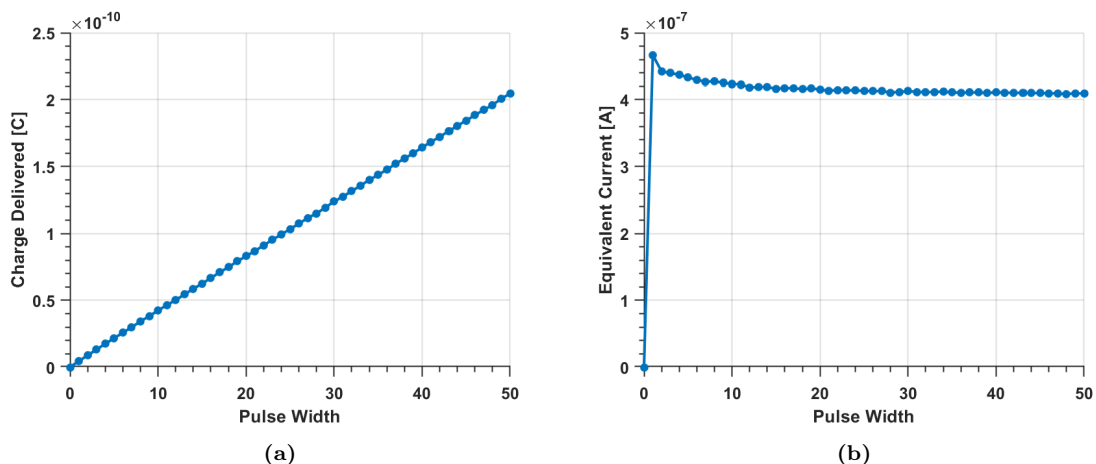


Figure 4.4: Characterisation of the Positive DAC (PDAC) current, by measuring (a) charge delivered during a constant pulse, and calculating (b) equivalent delivered current. $T_s = 10\mu\text{s}$, DAC 'setting' = 500nA .

4.2.2 Saturation and Blanking

In practice, a blanking or gain-shifting method is required to prevent saturation due to the direct artifact, however this may affect the practicability of the trimming method. Blanking would zero the voltage after the stimulation pulse, but introduce some small amount of charge injection from the switch, for which the trimming phase should be able to account. It was intended to test this hypothesis, and use the unattenuated channels ($\pm 5 \text{ mV}$) for assessing the performance of the algorithm.

The RHS2116 on-chip fast-settle feature can disconnect the output of the amplifier on command [83], however it was found that doing so results in large instabilities, even with no applied current (Fig. 4.5), that makes using this as a 'blanking' method impractical. Instead, testing was performed with the attenuated channels, so that the direct artifacts could be accommodated without saturation. Since the trimming process does not use information from the direct artifact anyway, this is equivalent to performing 'blanking' in software.

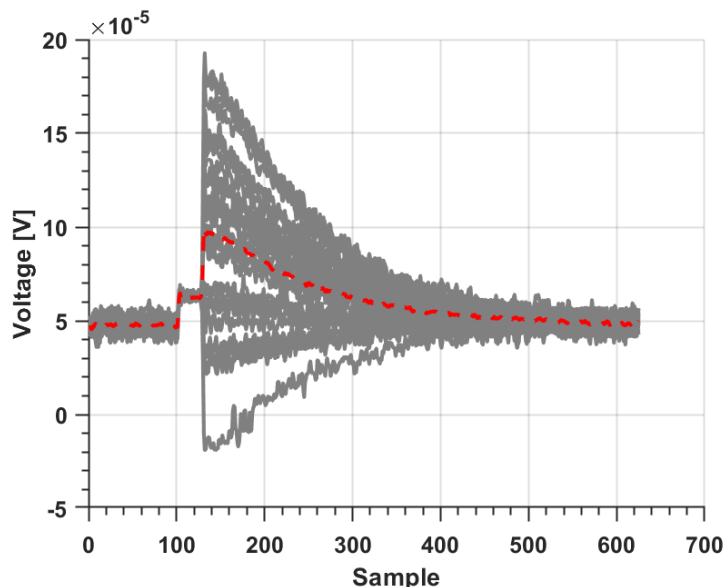


Figure 4.5: Instability of the 'fast-settle' switch feature of the RHS2116 IC. The switch is enabled for 25 samples from the 100 sample mark.

4.2.3 Time-Resolution

Communication speed was a major limiting factor in implementing more complex waveforms. The RHS2116 is technically capable of 714 kSa/s operation. However, we were only able to obtain 100 kSa/s with reasonable consistency, which corresponds to a $10 \mu s$ sampling period. As it takes two sampling periods to set both the polarity and amplitude of the current output, the resolution is effectively $20 \mu s$. For short recovery times ($< 200 \mu s$) there are limited degrees of freedom in the post-correction phase. Moving to faster speeds may produce smoother and more accurate correction, at the cost of increased computations to generate the waveform.

4.3 Recommendations and Future Work

4.3.1 Altering Working Phase

A very interesting area of future development would be to allow the shaping algorithm to alter the working phase. The sharp edges on the anodic phase mean there is a minimum time for the amplifier to settle, resulting in a lower limit to recovery time. This would be avoided by blanking, but the effects of abrupt current changes at the electrode interface would still be a problem. Allowing the algorithm to alter the working phase to minimise the artifact, while iteratively scaling the

resulting waveform to maintain a constant charge injection, may result in superior performance. It could be possible that a well-shaped working phase removes the requirement for the post-correction phase entirely, resulting in a near-zero recovery time. However, assessing the stimulation efficacy of this method is outside of the scope of this work.

4.3.2 Scaling Up

In Section 3.4, we summarised the cost of implementing the model-based artifact reduction algorithm for a single channel. The proposed epiretinal implant, discussed in Chapter 1, will have up to 10^4 channels. Any practical artifact reduction method must be scalable to this channel count, within the constraints of chip area, memory requirements, computation time and data acquisition time. For the following discussion, we assume the worst case, in which every channel is independent. However, we have seen that small variations in model parameters do not produce significant change in the resulting waveform, so it is reasonable that groups of electrodes may share parameters if they are sufficiently similar.

An investigation by Shah et al. [87] showed that reconstruction of visual stimulus using 10kHz stimulation (biphasic, $50 \mu s$ per phase), requires an integration period of 400ms. This suggests that achieving integration times similar to the visual circuitry of the brain ($<100ms$) requires upwards of 10 stimulating channels operating in parallel. Using the design by Noorsal [86] as a basis, we will require up to $\sim 2 mm^2$ of chip area for the DACs, which is feasible for an implanted retinal device.

Based on the long-term stability of the artifact, measured in Section 3.1.2, and the minimal susceptibility to parameter variation, it is reasonable to assume that it is sufficient to update the model parameters twice a day. Scaling up the modelling processes to 10k channels requires 40s per day for EIS fitting ($f_{min} = 500 Hz$) and 10s per day for impulse fitting ($25 \mu s$ pulse, two amplitudes).

As for computational cost, we find that for 10k unique channels, we require a total of $N*2.25e6$ and $N*8.25e6$ MVCs per day for the non-voltage-constrained and voltage-constrained algorithms, respectively, where N is the number of working phase configurations (widths/amplitudes). This is significant, as each MVC requires a LUT inquiry, two multiplications and an addition. In regards to memory usage, we require a total of $N*640kB$ for all channels. Both the computational and memory cost of the algorithm are potentially infeasible for an ultra low power device implanted within the eye. However, it is certainly possible that the optimum stimulation waveforms may be computed remotely, and transmitted to the implant as required. For a 100us working phase, an 8-bit DAC, 10 pre-correction phases and 5 post-correction phases ($20 \mu s$ each), the required data rate for a single waveform would therefore be

$$R = (16ph * 8bit)/(100\mu s + 15ph * 20\mu s) = 320kbps \quad (4.1)$$

A 1024 channel epiretinal implant design by Chen et al. [88] achieves a $2Mbps$ data rate, indicating that streaming the stimulation waveforms is a viable approach. The trimming method requires only minor alteration of the stimulation waveform

based on the recorded residual artifact, which can be done on the implant itself.

Despite being technically feasible, the model-based artifact reduction algorithm designed in this work is fundamentally a support algorithm, complementary to the primary goal of the epiretinal implant, which is to provide visual capability to the patient. As such, its overhead should be minimal. Therefore, a key focus in the future development of this algorithm should be reducing the memory usage and computational complexity by optimising the numbers of pre- and post-correction phases required, and exploring ways to further reduce the iterations required for waveform generation.

4.4 Conclusions

The aim of this work was to determine how stimulation artifacts can be mitigated such that neural responses can be reliably captured $<500 \mu s$ post-stimulation, from up to 10^4 channels, including the stimulation channel, while maintaining strict constraints on power consumption, chip area and computational complexity. With a mean recovery time of $124 \mu s$, we have achieved this requirement, and managed to improve upon conventional methods by an average of 79%. In the process, we found that correcting the artifact can degrade charge balancing at the interface, resulting in 20% more voltage accumulation during pulse trains. However, with low stimulation currents this is not an issue. Achieving even lower recovery times may be possible using this method, along with careful selection of algorithm parameters, and improved time-resolution of the current output. In terms of cost, this method requires a small amount of additional chip area due to the requirement of arbitrary current generation, and although much of the computational complexity and memory requirements can be outsourced to an external device, the impact of implementing the algorithm in its current form is far from negligible. Future work should focus on refactoring the algorithm for implementation on-chip, exploring alteration of the working phase shape as a method for further residual artifact reduction, implementing blanking to deal with the direct artifact and further exploration and optimisation of the shaping parameters to maximise artifact reduction while minimising overall cost.

Bibliography

- [1] M. Sudmeyer S. Groiss L. Wojkdecki and A. Schnitzler. “Deep Brain Stimulation in Parkinson’s Disease”. In: *Therapeutic Advances in Neurological Disorders* 2.6 (2009), pp. 20–28. DOI: <https://doi.org/10.1177/1756285609339382>.
- [2] M. Sprengers et al. “Deep brain and cortical stimulation for epilepsy”. In: *Cochrane Database of Systematic Reviews* 2014.6 (2014). DOI: [10.1002/14651858.CD008497.pub2](https://doi.org/10.1002/14651858.CD008497.pub2).
- [3] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday. “Brain–computer interfaces in the completely locked-in state and chronic stroke”. In: *Progress in Brain Research* 228 (2016), pp. 131–161. DOI: [10.1016/bs.pbr.2016.04.019](https://doi.org/10.1016/bs.pbr.2016.04.019).
- [4] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday. “Brain-computer interfaces for communication and rehabilitation”. In: *Nature Reviews Neurology* 12.9 (2016), pp. 513–525. DOI: [10.1038/nrneuro.2016.113](https://doi.org/10.1038/nrneuro.2016.113).
- [5] E. Marieb and K Hoehn. “Human Anatomy ” Physiology”. In: 10th ed. Pearson Education Limited, 2015. Chap. 15, pp. 574–576.
- [6] Stanford Medicine. *Stanford Artificial Retina Project*. 2020. URL: <http://med.stanford.edu/artificial-retina.html> (visited on 11/26/2020).
- [7] Dante G. Muratore and E. J. Chichilnisky. “Artificial Retina: A Future Cellular-Resolution Brain-Machine Interface”. In: *NANO-CHIPS 2030: On-Chip AI for an Efficient Data-Driven World*. Ed. by Boris Murmann and Bernd Hofflinger. Cham: Springer International Publishing, 2020, pp. 443–465. ISBN: 978-3-030-18338-7. DOI: [10.1007/978-3-030-18338-7_24](https://doi.org/10.1007/978-3-030-18338-7_24).
- [8] G.E. Mena et al. “Electrical stimulus artifact cancellation and neural spike detection on large multi-electrode arrays”. In: *PLoS Computational Biology* 13.11 (2017). DOI: [10.1371/journal.pcbi.1005842](https://doi.org/10.1371/journal.pcbi.1005842).
- [9] A. Zhou, B.C. Johnson, and R. Muller. “Toward true closed-loop neuromodulation: artifact-free recording during stimulation”. In: *Current Opinion in Neurobiology* 50 (2018), pp. 119–127. DOI: [10.1016/j.conb.2018.01.012](https://doi.org/10.1016/j.conb.2018.01.012).
- [10] D.R. Merrill, M. Bikson, and J.G.R. Jefferys. “Electrical stimulation of excitable tissue: Design of efficacious and safe protocols”. In: *Journal of Neuroscience Methods* 141.2 (2005), pp. 171–198. DOI: [10.1016/j.jneumeth.2004.10.020](https://doi.org/10.1016/j.jneumeth.2004.10.020).
- [11] E.A. Brown et al. “Stimulus-artifact elimination in a multi-electrode system”. In: *IEEE Transactions on Biomedical Circuits and Systems* 2.1 (2008), pp. 10–21. DOI: [10.1109/TBCAS.2008.918285](https://doi.org/10.1109/TBCAS.2008.918285).
- [12] G.A. DeMichele and P.R. Troyk. “Stimulus-Resistant Neural Recording Amplifier”. In: vol. 4. 2003, pp. 3329–3332.

-
- [13] E.A. Brown et al. “Stimulation and recording of neural tissue, closing the loop on the artifact”. In: 2008, pp. 356–359. DOI: 10.1109/ISCAS.2008.4541428.
- [14] H. Jung, J. Kim, and Y. Nam. “Recovery of early neural spikes from stimulation electrodes using a DC-coupled low gain high resolution data acquisition system”. In: *Journal of Neuroscience Methods* 304 (2018), pp. 118–125. DOI: 10.1016/j.jneumeth.2018.04.014.
- [15] Y. Erez et al. “Generalized framework for stimulus artifact removal”. In: *Journal of Neuroscience Methods* 191.1 (2010), pp. 45–59. DOI: 10.1016/j.jneumeth.2010.06.005.
- [16] C. Wang et al. “Characteristics of electrode impedance and stimulation efficacy of a chronic cortical implant using novel annulus electrodes in rat motor cortex”. In: *Journal of Neural Engineering* 10.4 (2013). DOI: 10.1088/1741-2560/10/4/046010.
- [17] B.C. Johnson et al. “An implantable 700W 64-channel neuromodulation IC for simultaneous recording and stimulation with rapid artifact recovery”. In: 2017, pp. C48–C49. DOI: 10.23919/VLSIC.2017.8008543.
- [18] J.L. Valtierra et al. “A High TCMRR, Inherently Charge Balanced Bidirectional Front-End for Multichannel Closed-Loop Neuromodulation”. In: 2019. DOI: 10.1109/BIOCAS.2019.8919111.
- [19] E. Pepin et al. “A high-voltage compliant, electrode-invariant neural stimulator front-end in 65nm bulk-CMOS”. In: vol. 2016-October. 2016, pp. 229–232. DOI: 10.1109/ESSCIRC.2016.7598284.
- [20] J. Uehlin et al. “A Bidirectional Brain Computer Interface with 64-Channel Recording, Resonant Stimulation and Artifact Suppression in Standard 65nm CMOS”. In: 2019, pp. 77–80. DOI: 10.1109/ESSCIRC.2019.8902911.
- [21] E. Greenwald et al. “A CMOS neurostimulator with on-chip DAC calibration and charge balancing”. In: 2013, pp. 89–92. DOI: 10.1109/BioCAS.2013.6679646.
- [22] Y. Jimbo et al. “A system for MEA-based multisite stimulation”. In: *IEEE Transactions on Biomedical Engineering* 50.2 (2003), pp. 241–248. DOI: 10.1109/TBME.2002.805470.
- [23] H. Pu et al. “Optimal artifact suppression in simultaneous electrocorticography stimulation and recording for bi-directional brain-computer interface applications”. In: *Journal of Neural Engineering* 17.2 (2020). DOI: 10.1088/1741-2552/ab82ac.
- [24] E.J. Peterson et al. “Stimulation artifact rejection in closed-loop, distributed neural interfaces”. In: vol. 2016-October. 2016, pp. 233–236. DOI: 10.1109/ESSCIRC.2016.7598285.
- [25] W. Biederman et al. “A 4.78 mm² Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels with Digital Compression and Simultaneous Dual Stimulation”. In: *IEEE Journal of Solid-State Circuits* 50.4 (2015), pp. 1038–1047. DOI: 10.1109/JSSC.2014.2384736.
- [26] S. Nag et al. “Sensing of Stimulus Artifact Suppressed Signals From Electrode Interfaces”. In: *IEEE Sensors Journal* 15.7 (2015), pp. 3734–3742. DOI: 10.1109/JSEN.2015.2399248.

- [27] K. Kolodziej et al. “Modelling and Cancellation of the Stimulation Artifact for ASIC-based Bidirectional Neural Interface”. In: 2018, pp. 449–453. DOI: 10.23919/MIXDES.2018.8436947.
- [28] B. Dura et al. “High-frequency electrical stimulation of cardiac cells and application to artifact reduction”. In: *IEEE Transactions on Biomedical Engineering* 59.5 (2012), pp. 1381–1390. DOI: 10.1109/TBME.2012.2188136.
- [29] L.F. Heffer and J.B. Fallon. “A novel stimulus artifact removal technique for high-rate electrical stimulation”. In: *Journal of Neuroscience Methods* 170.2 (2008), pp. 277–284. DOI: 10.1016/j.jneumeth.2008.01.023.
- [30] P. Chu et al. “Equalization for intracortical microstimulation artifact reduction”. In: 2013, pp. 245–248. DOI: 10.1109/EMBC.2013.6609483.
- [31] S. Elyahoodayan et al. “A multi-channel asynchronous neurostimulator with artifact suppression for neural code-based stimulations”. In: *Frontiers in Neuroscience* 13.SEP (2019). DOI: 10.3389/fnins.2019.01011.
- [32] L.R. McKenzie et al. “Same-Electrode Stimulation and Recording With Dynamic Hardware Artefact Suppression”. In: *IFAC-PapersOnLine* 51.27 (2018), pp. 56–61. DOI: 10.1016/j.ifacol.2018.11.609.
- [33] C. Hartmann et al. “Closed-loop control of myoelectric prostheses with electrotactile feedback: Influence of stimulation artifact and blanking”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 23.5 (2015), pp. 807–816. DOI: 10.1109/TNSRE.2014.2357175.
- [34] M. Szypulska et al. “Modular ASIC-based system for large-scale electrical stimulation and recording of brain activity in behaving animals”. In: 2016, pp. 217–222. DOI: 10.1109/MIXDES.2016.7529735.
- [35] H. Dispirito et al. “A layered approach to artifact rejection for improved neural recording”. In: 2015. DOI: 10.1109/NEBEC.2015.7117157.
- [36] T.L. Babb et al. “A sample and hold amplifier system for stimulus artifact suppression”. In: *Electroencephalography and Clinical Neurophysiology* 44.4 (1978), pp. 528–531. DOI: 10.1016/0013-4694(78)90038-X.
- [37] R.J. Roby and E. Lettich. “A simplified circuit for stimulus artifact suppression”. In: *Electroencephalography and Clinical Neurophysiology* 39.1 (1975), pp. 85–87. DOI: 10.1016/0013-4694(75)90130-3.
- [38] Y. Nam et al. “A retrofitted neural recording system with a novel stimulation IC to monitor early neural responses from a stimulating electrode”. In: *Journal of Neuroscience Methods* 178.1 (2009), pp. 99–102. DOI: 10.1016/j.jneumeth.2008.11.017.
- [39] P. Hottowy et al. “An integrated multichannel waveform generator for large-scale spatio-temporal stimulation of neural tissue”. In: *Analog Integrated Circuits and Signal Processing* 55.3 (2008), pp. 239–248. DOI: 10.1007/s10470-007-9125-x.
- [40] R.A. Blum et al. “Models of stimulation artifacts applied to integrated circuit design”. In: vol. 26 VI. 2004, pp. 4075–4078.

-
- [41] X. Liu et al. “Design of a Closed-Loop, Bidirectional Brain Machine Interface System with Energy Efficient Neural Feature Extraction and PID Control”. In: *IEEE Transactions on Biomedical Circuits and Systems* 11.4 (2017), pp. 729–742. DOI: 10.1109/TBCAS.2016.2622738.
- [42] R.H. Olsson III et al. “Band-tunable and multiplexed integrated circuits for simultaneous recording and stimulation with microelectrode arrays”. In: *IEEE Transactions on Biomedical Engineering* 52.7 (2005), pp. 1303–1311. DOI: 10.1109/TBME.2005.847540.
- [43] V. Viswam et al. “2048 action potential recording channels with 2.4 vrms noise and stimulation artifact suppression”. In: 2016, pp. 136–139. DOI: 10.1109/BioCAS.2016.7833750.
- [44] S. Zanos et al. “The neurochip-2: An autonomous head-fixed computer for recording and stimulating in freely behaving monkeys”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19.4 (2011), pp. 427–435. DOI: 10.1109/TNSRE.2011.2158007.
- [45] H. Chandrakumar and D. Markovic. “An 80-mVpp linear-input range, 1.6-G input impedance, low-power chopper amplifier for closed-loop neural recording that is tolerant to 650-mVpp common-mode interference”. In: *IEEE Journal of Solid-State Circuits* 52.11 (2017), pp. 2811–2828. DOI: 10.1109/JSSC.2017.2753824.
- [46] H. Jeon et al. “A 3.9W, 81.3dB SNDR, DC-coupled, Time-based Neural Recording IC with Degeneration R-DAC for Bidirectional Neural Interface in 180nm CMOS”. In: 2018, pp. 91–92. DOI: 10.1109/ASSCC.2018.8579284.
- [47] H. Jeon et al. “A High DR, DC-Coupled, Time-Based Neural-Recording IC with Degeneration R-DAC for Bidirectional Neural Interface”. In: *IEEE Journal of Solid-State Circuits* 54.10 (2019), pp. 2658–2670. DOI: 10.1109/JSSC.2019.2930903.
- [48] D. Rozgic et al. “A 0.338 cm³, Artifact-Free, 64-Contact Neuromodulation Platform for Simultaneous Stimulation and Sensing”. In: *IEEE Transactions on Biomedical Circuits and Systems* 13.1 (2019), pp. 38–55. DOI: 10.1109/TBCAS.2018.2889040.
- [49] J.D. Rolston, R.E. Gross, and S.M. Potter. “A low-cost multielectrode system for data acquisition enabling real-time closed-loop processing with rapid recovery from stimulation artifacts”. In: *Frontiers in Neuroengineering* 2.JUL (2009). DOI: 10.3389/neuro.16.012.2009.
- [50] W. Jiang et al. “A ± 50 -mV Linear-Input-Range VCO-Based Neural-Recording Front-End With Digital Nonlinearity Correction”. In: *IEEE Journal of Solid-State Circuits* 52.1 (2017), pp. 173–184. DOI: 10.1109/JSSC.2016.2624989.
- [51] H. Chandrakumar and D. Markovic. “A 15.2-ENOB 5-kHz BW 4.5- W Chopped CT -ADC for artifact-tolerant neural recording front ends”. In: *IEEE Journal of Solid-State Circuits* 53.12 (2018), pp. 3470–3483. DOI: 10.1109/JSSC.2018.2876468.

- [52] C. Lee et al. “A 6.5W 10kHz-BW 80.4dB-SNDR Continuous-Time Modulator with Gm-Input and 300mVpp Linear Input Range for Closed-Loop Neural Recording”. In: vol. 2020-February. 2020, pp. 410–412. DOI: 10.1109/ISSCC19947.2020.9063074.
- [53] U. Hoffmann et al. “Detection and removal of stimulation artifacts in electroencephalogram recordings”. In: 2011, pp. 7159–7162. DOI: 10.1109/IEMBS.2011.6091809.
- [54] C. Waddell et al. “Deep brain stimulation artifact removal through under-sampling and cubic-spline interpolation”. In: 2009. DOI: 10.1109/CISP.2009.5301199.
- [55] D.A. Wagenaar and S.M. Potter. “Real-time multi-channel stimulus artifact suppression by local curve fitting”. In: *Journal of Neuroscience Methods* 120.2 (2002), pp. 113–120. DOI: 10.1016/S0165-0270(02)00149-8.
- [56] S. Culaclii et al. “Online Artifact Cancellation in Same-Electrode Neural Stimulation and Recording Using a Combined Hardware and Software Architecture”. In: *IEEE Transactions on Biomedical Circuits and Systems* 12.3 (2018), pp. 601–613. DOI: 10.1109/TBCAS.2018.2816464.
- [57] T. Hashimoto, C.M. Elder, and J.L. Vitek. “A template subtraction method for stimulus artifact removal in high-frequency deep brain stimulation”. In: *Journal of Neuroscience Methods* 113.2 (2002), pp. 181–186. DOI: 10.1016/S0165-0270(01)00491-5.
- [58] T. Wichmann. “A digital averaging method for removal of stimulus artifacts in neurophysiologic experiments”. In: *Journal of Neuroscience Methods* 98.1 (2000), pp. 57–62. DOI: 10.1016/S0165-0270(00)00190-4.
- [59] L. Trebaul et al. “Stimulation artifact correction method for estimation of early cortico-cortical evoked potentials”. In: *Journal of Neuroscience Methods* 264 (2016), pp. 94–102. DOI: 10.1016/j.jneumeth.2016.03.002.
- [60] T. Blogg and W.D. Reid. “A digital technique for stimulus artifact reduction”. In: *Electroencephalography and Clinical Neurophysiology* 76.6 (1990), pp. 557–561. DOI: 10.1016/0013-4694(90)90005-5.
- [61] L. Sun and H. Hinrichs. “Moving average template subtraction to remove stimulation artefacts in EEGs and LFPs recorded during deep brain stimulation”. In: *Journal of Neuroscience Methods* 266 (2016), pp. 126–136. DOI: 10.1016/j.jneumeth.2016.03.020.
- [62] K. Limnusun et al. “Real-time stimulus artifact rejection via template subtraction”. In: *IEEE Transactions on Biomedical Circuits and Systems* 8.3 (2014), pp. 391–400. DOI: 10.1109/TBCAS.2013.2274574.
- [63] D.J. Caldwell et al. “Signal recovery from stimulation artifacts in intracranial recordings with dictionary learning”. In: *Journal of Neural Engineering* 17.2 (2020). DOI: 10.1088/1741-2552/ab7a4f.
- [64] K. Zeng et al. “An EEMD-ICA Approach to Enhancing Artifact Rejection for Noisy Multivariate Neural Data”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24.6 (2016), pp. 630–638. DOI: 10.1109/TNSRE.2015.2496334.

- [65] T. Al-ani et al. “Automatic removal of high-amplitude stimulus artefact from neuronal signal recorded in the subthalamic nucleus”. In: *Journal of Neuroscience Methods* 198.1 (2011), pp. 135–146. DOI: 10.1016/j.jneumeth.2011.03.022.
- [66] A.E. Mendrela et al. “A Bidirectional Neural Interface Circuit with Active Stimulation Artifact Cancellation and Cross-Channel Common-Mode Noise Suppression”. In: *IEEE Journal of Solid-State Circuits* 51.4 (2016), pp. 955–965. DOI: 10.1109/JSSC.2015.2506651.
- [67] J. Lim et al. “Pre-whitening and Null Projection as an Artifact Suppression Method for ElectroCorticography Stimulation in Bi-Directional Brain Computer Interfaces”. In: vol. 2020-July. 2020, pp. 3493–3496. DOI: 10.1109/EMBC44109.2020.9175760.
- [68] R. Grieve et al. “Nonlinear adaptive filtering of stimulus artifact”. In: *IEEE Transactions on Biomedical Engineering* 47.3 (2000), pp. 389–395. DOI: 10.1109/10.827307.
- [69] V. Parsa, P.A. Parker, and R.N. Scott. “Adaptive stimulus artifact reduction in noncortical somatosensory evoked potential studies”. In: *IEEE Transactions on Biomedical Engineering* 45.2 (1998), pp. 165–179. DOI: 10.1109/10.661265.
- [70] H. Liang and Z. Lin. “Stimulus artifact cancellation in the serosal recordings of gastric myoelectric activity using wavelet transform”. In: *IEEE Transactions on Biomedical Engineering* 49.7 (2002), pp. 681–688. DOI: 10.1109/TBME.2002.1010851.
- [71] S. Basir-Kazeruni et al. “A blind Adaptive Stimulation Artifact Rejection (ASAR) engine for closed-loop implantable neuromodulation systems”. In: 2017, pp. 186–189. DOI: 10.1109/NER.2017.8008322.
- [72] M. Reza Pazhouhandeh et al. “Track-and-zoom neural analog-to-digital converter with blind stimulation artifact rejection”. In: *IEEE Journal of Solid-State Circuits* 55.7 (2020), pp. 1984–1997. DOI: 10.1109/JSSC.2020.2991526.
- [73] C. Kim et al. “Sub- Vrms-noise Sub- W/Channel ADC-direct neural recording with 200-mV/ms transient recovery through predictive digital autoranging”. In: *IEEE Journal of Solid-State Circuits* 53.11 (2018), pp. 3101–3110. DOI: 10.1109/JSSC.2018.2870555.
- [74] T.K.T. Nguyen et al. “Mixed-signal template-based reduction scheme for stimulus artifact removal in electrical stimulation”. In: *Medical and Biological Engineering and Computing* 51.4 (2013), pp. 449–458. DOI: 10.1007/s11517-012-1013-6.
- [75] T. Wichmann and A. Devergnas. “A novel device to suppress electrical stimulus artifacts in electrophysiological experiments”. In: *Journal of Neuroscience Methods* 201.1 (2011), pp. 1–8. DOI: 10.1016/j.jneumeth.2011.06.026.
- [76] J.W. Gnadt et al. “Spectral cancellation of microstimulation artifact for simultaneous neural recording in situ”. In: *IEEE Transactions on Biomedical Engineering* 50.10 (2003), pp. 1129–1135. DOI: 10.1109/TBME.2003.816077.
- [77] W.A. Smith et al. “A scalable, highly-multiplexed delta-encoded digital feedback ECoG recording amplifier with common and differential-mode artifact suppression”. In: 2017, pp. C172–C173. DOI: 10.23919/VLSIC.2017.8008470.

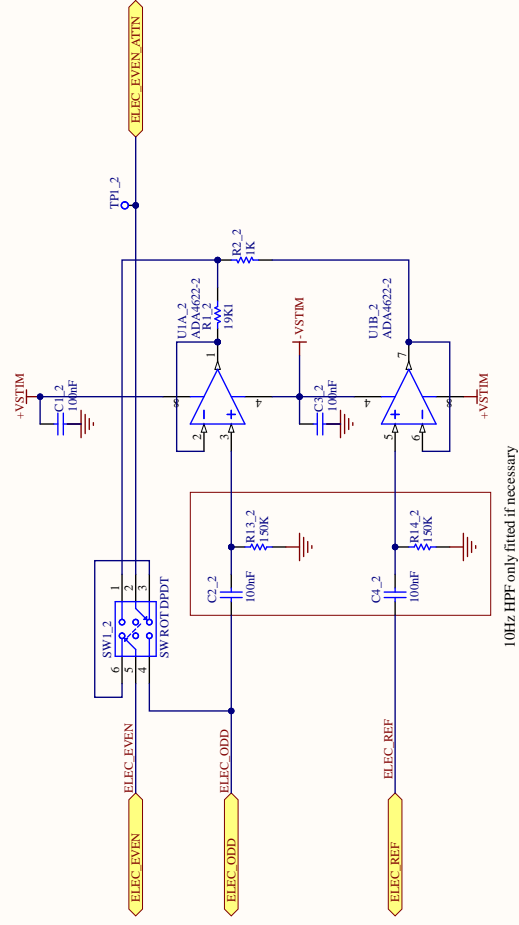
- [78] J.P. Uehlin et al. “A Single-Chip Bidirectional Neural Interface with High-Voltage Stimulation and Adaptive Artifact Cancellation in Standard CMOS”. In: *IEEE Journal of Solid-State Circuits* 55.7 (2020), pp. 1749–1761. DOI: 10.1109/JSSC.2020.2991524.
- [79] Y. Li, J. Chen, and Y. Yang. “A Method for Suppressing Electrical Stimulation Artifacts from Electromyography”. In: *International Journal of Neural Systems* 29.6 (2019). DOI: 10.1142/S0129065718500545.
- [80] K. Limnuson et al. “A bidirectional neural interface SoC with an integrated spike recorder, microstimulator, and low-power processor for real-time stimulus artifact rejection”. In: *Analog Integrated Circuits and Signal Processing* 82.2 (2015), pp. 457–470. DOI: 10.1007/s10470-015-0489-z.
- [81] S. Nag and D. Sharma. “Wirelessly powered stimulator and recorder for neuronal interfaces”. In: 2011, pp. 5612–5616. DOI: 10.1109/IEMBS.2011.6091358.
- [82] Nina Milosavljevic et al. “Photoreceptive retinal ganglion cells control the information rate of the optic nerve”. In: *Proceedings of the National Academy of Sciences* 115.50 (2018), E11817–E11826. ISSN: 0027-8424. DOI: 10.1073/pnas.1810701115. eprint: <https://www.pnas.org/content/115/50/E11817.full.pdf>.
- [83] Intan Technologies. *RHS Stim/Amplifier Chips*. 2021. URL: https://intantech.com/products_RHS2000.html (visited on 10/16/2021).
- [84] Raspberry Pi Ltd. *Raspberry Pi 3 Model B+*. 2021. URL: <https://www.raspberrypi.com/products/raspberry-pi-3-model-b-plus/> (visited on 10/16/2021).
- [85] Javier Lario-García and Ramon Pallàs-Areny. “Constant-phase element identification in conductivity sensors using a single square wave”. In: *Sensors and Actuators A: Physical* 132.1 (2006). The 19th European Conference on Solid-State Transducers, pp. 122–128. ISSN: 0924-4247. DOI: 10.1016/j.sna.2006.04.014.
- [86] Emilia Noorsal et al. “A Neural Stimulator Frontend With High-Voltage Compliance and Programmable Pulse Shape for Epiretinal Implants”. In: *IEEE Journal of Solid-State Circuits* 47.1 (2012), pp. 244–256. DOI: 10.1109/JSSC.2011.2164667.
- [87] Nishal P. Shah et al. “Optimization of Electrical Stimulation for a High-Fidelity Artificial Retina”. In: *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. 2019, pp. 714–718. DOI: 10.1109/NER.2019.8716987.
- [88] Kuanfu Chen, Yi-Kai Lo, and Wentai Liu. “A 37.6mm² 1024-channel high-compliance-voltage SoC for epiretinal prostheses”. In: *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*. 2013, pp. 294–295. DOI: 10.1109/ISSCC.2013.6487741.

Appendix

A.1 Test Board Schematic

Analog Input Channels

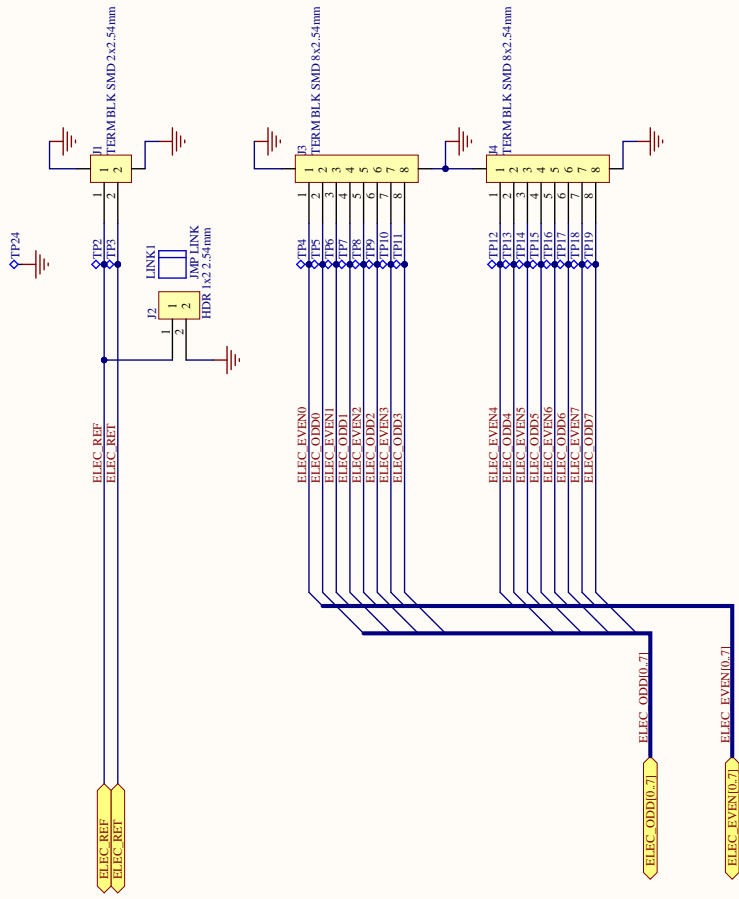
Optional 1/20.1 Attenuation



POWER CONSUMPTION (MAX):
 ADA4625: 2.54mA max
 = 8mA/channel
 Total: 64mA @ +/-4.5V = 576mW

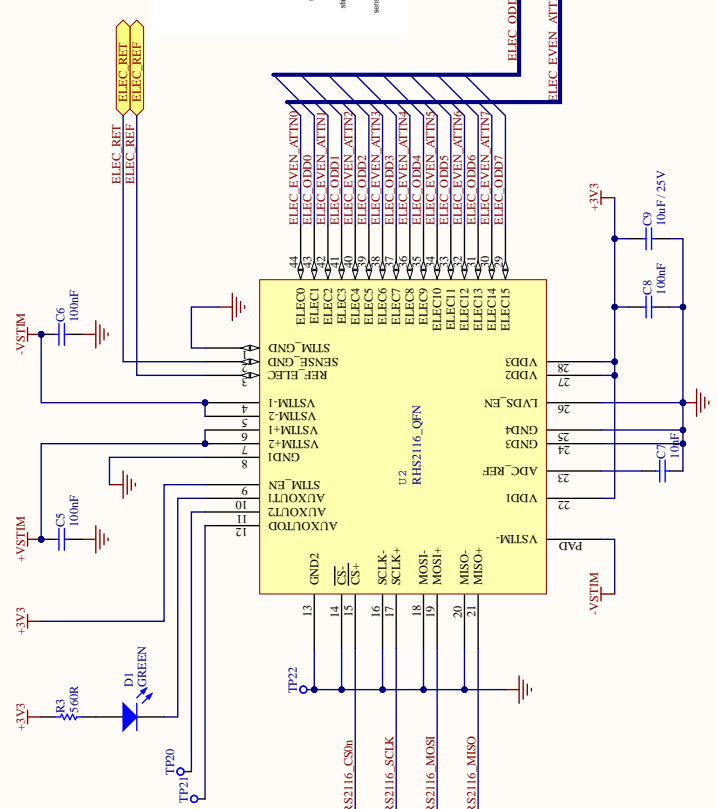
Title		AFE_Channel.SchDoc	
Size	Number	Revision	
A4		V2.1	
Date:	3/18/2021	Sheet 2 of	6
File:	C:\Users\...AFE_Channel.SchDoc	Drawn By:	Rohan Brash

ELECTRODE CONNECTIONS



Title		Electrode_Header.SchDoc	
Size	Number	Revision	
A4		V2.1	
Date:	3/18/2021	Sheet 3 of 6	
File:	C:\Users\... \Electrode_Header.SchDoc	Drawn By: Rohan Brash	

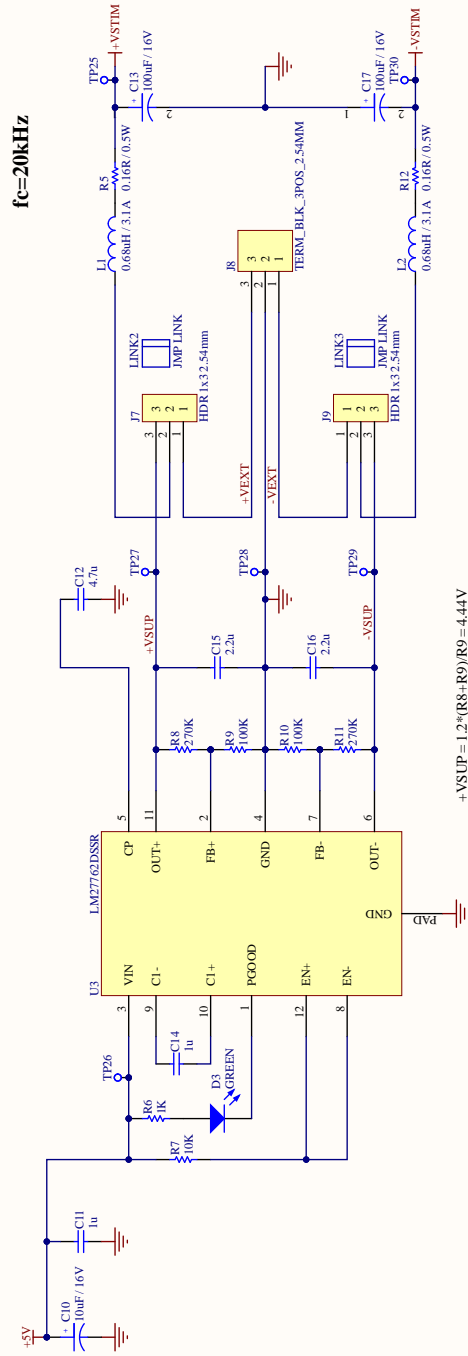
RHS2116 Stim/Record IC



POWER CONSUMPTION:
 16 Channels @ 44.6kS/s vs +2.55mA stim, max low pass cutoff (20kHz)
 Baseline current (VDD): 3.2mA
 AC amps: 9.2uA/(Hz) = 9.2uA*20 = 0.18mA
 ADC/MUX: 5.2uA/BS# = 5.2*714 = 3.712mA
 Impedance Meas: 0.12mA
 VDD subtotal = 7.216mA
 Max stim current in negative cycle:
 Stim VDD = 0.1*Isstim = 0.1*2.55 = 0.255mA/ch
 Stim VSTIM+ = 0.34*Isstim = 0.34*2.55 = 0.867mA/ch
 Stim VSTIM- = 1.31*Isstim = 1.31*2.55 = 3.3405mA/ch
 For operating channels:
 VDD: 7.471mA @ 3.3V = 24.65mW
 VS+: 0.867mA @ 4.5V = 3.9mW
 VS-: 3.3405mA @ 4.5V = 15.03mW
 For 16 stim channels:
 VDD: 11.296mA @ 3.3V = 37.0768mW
 VS+: 14.682mA @ 4.5V = 66.129mW
 VS-: 53.448mA @ 4.5V = 240.1516mW

Title		RHS2116.SchDoc	
Size	Number	Revision	V2.1
A4		Sheet of	6
Date:	3/18/2021	Drawn By:	Rohan Brash
File:	C:\Users\...RHS2116.SchDoc		

Power Supply



$f_c = 20\text{kHz}$

$$+VSUP = 1.2 \cdot (R8 + R9) / R9 = 4.44\text{V}$$

$$-VSUP = -1.22 \cdot (R11 + R10) / R10 = -4.514\text{V}$$

POWER BUDGET
 - ADS168 ADC: 0mA
 - RFS2116: 67.32mA (all channels stimulating)
 - AFEs: 80mA static current
 TOTAL: 146.32mA (absolute maximum)
 LM27762 Max operating current: 250mA
 Size'll be right.

Title		Revision	
Size	Number	V2.1	
A4	3/18/2021	Sheet of	6
File: C:\Users\...Supply_SchDoc		Drawn By:	Rohan Brash

A.2 Artifact Reduction Algorithm - MATLAB Script

```
1 %% Sequential Artifact Reduction Algorithm – General Form
2 % Rohan Brash
3
4 clear all
5 close all
6
7 I_STEP = 10e-9;
8 I_MAX = 2.55e-6;
9 I_MIN = -2.55e-6;
10
11 % Model Parameters
12 ALPHA = 0.6425;
13 RS = 4.8899e+04;
14 K = 9.2256e+03;
15
16 % Working pulse parameters
17 PULSE_AMP = -500e-9; % amps
18 PULSE_WIDTH = 25; % samples
19
20 % Pulse acquisition parameters
21 PULSE_PREFIX = 50;
22 PULSE_SUFFIX = 150;
23
24 % Pre-correction parameters
25 PRE_CORR_PHASE_WIDTH = 2; % Width of each pre-correction
    phase
26 PRE_CORR_N_PHASES = 10; % Number of pre-correction phases
27 PRE_CORR_WIDTH = PRE_CORR_N_PHASES*PRE_CORR_PHASE_WIDTH; %
    Total width of the pre-correction phase
28 PRE_CORR_POINT = PULSE_WIDTH*4; % Correct for one pulse
    width after the end of triphasic
29 PRE_CORR_I_MAX = I_MAX; % Maximum allowed pre-correction
    current
30 PRE_CORR_I_MIN = 0; % Minimum allowed pre-correction current
31 PRE_CORR_V_MAX = 0.1; % Maximum allowed pre-correction
    voltage
32 PRE_CORR_V_MIN = -1; % Minimum allowed pre-correction
    voltage
33 PRE_CORR_SKIP_FIRST = false;
34
35 % Post-correction parameters
36 POST_CORR_PHASE_WIDTH = 2; % Width of each post-correction
    phase
37 POST_CORR_N_PHASES = 5; % Number of post-correction phases
38 POST_CORR_WIDTH = POST_CORR_N_PHASES*POST_CORR_PHASE_WIDTH;
    % Total width of the post-correction phase
```

```

39 POST_CORR_POINT = PRE_CORR_WIDTH+PULSE_WIDTH+POST_CORR_WIDTH
    +5; % Correct for 5 samples after the end
40 POST_CORR_I_MAX = I_MAX; % Maximum allowed post-correction
    current
41 POST_CORR_I_MIN = 0; % Minimum allowed post-correction
    current
42 POST_CORR_V_MAX = 0.1; % Maximum allowed post-correction
    voltage
43 POST_CORR_V_MIN = -1; % Minimum allowed post-correction
    voltage
44 POST_CORR_SKIP_LAST = true;
45
46 % Maximum number of complete corrective iterations
47 MAX_CORR_ITERATIONS = 100;
48 LIMIT_VOLTAGE = true;
49
50 %% Set up required storage
51
52 % Working pulse index/delta notation
53 ni_wp = [PRE_CORR_WIDTH;PRE_CORR_WIDTH+PULSE_WIDTH];
54 di_wp = [PULSE_AMP;-PULSE_AMP];
55
56 % Pre-correction pulse
57 ni_pre = zeros(PRE_CORR_N_PHASES+1,1); % Pre-correction
    delta locations
58 di_pre = zeros(PRE_CORR_N_PHASES+1,1); % Pre-correction
    delta magnitudes
59
60 % Post-correction pulse
61 ni_post = zeros(POST_CORR_N_PHASES+1,1); % Post-correction
    delta locations
62 di_post = zeros(POST_CORR_N_PHASES+1,1); % Post-correction
    delta magnitudes
63
64 %% Calculate initial voltages due to working pulse
65
66 ni = [ni_pre;ni_wp;ni_post];
67 di = [di_pre;di_wp;di_post];
68
69 figure()
70 formatFig(gcf);
71 grid on
72 hold on
73 xlabel('Sample');
74 ylabel('Voltage [V]');
75
76 % Plot initial model voltage (for demonstration only)
77 nm = 0:PULSE_SUFFIX;

```

```

78 vm = zeros(size(nm));
79 for j = 1:length(nm)
80     for k = 1:length(ni)
81         if nm(j) >= ni(k) && di(k)~=0
82             vm(j) = vm(j) + di(k)*(RS + K*(nm(j)-ni(k))^
                ALPHA);
83         end
84     end
85 end
86 plot(nm,vm, 'Color',[0, 0.4470, 0.7410]);
87
88 di_prev = di;
89
90 mvc_count = 0;
91
92
93 %% Calculate Corrective Pulses
94 for iter = 1:MAX_CORR_ITERATIONS
95
96     % Calculate Pre-Corrective Pulse
97     if iter > 1 || ~PRE_CORR_SKIP_FIRST
98         ni = [ni_pre;ni_wp;ni_post];
99         di = [di_pre;di_wp;di_post];
100
101         ni_curr = 0;
102         i_curr = 0;
103
104         % Calculate the error voltage at the pre-correction
            point
105         nm_pre = PRE_CORR_POINT;
106         ve_pre = 0;
107         for k = 1:length(ni)
108             if nm_pre >= ni(k) && di(k) ~= 0
109                 ve_pre = ve_pre + di(k)*(RS + K*(nm_pre-ni(k)
                    ))^ALPHA);
110                 mvc_count = mvc_count+1;
111             end
112         end
113
114         for i = 1:PRE_CORR_N_PHASES
115
116             % Pre-correction current up until this point
117             i_curr = i_curr + di_pre(i);
118
119             % Calculate the correction current delta
120             % Since the correction point is (should be)
                after the end of
121             % the pulse, the Rs component is neglected.

```

```

122     di_curr = -ve_pre/(K*((PRE_CORR_POINT-ni_curr)^(
        ALPHA-(PRE_CORR_POINT-ni_curr-
        PRE_CORR_PHASE_WIDTH)^ALPHA));
123    .mvc_count =.mvc_count+1;
124
125     % We have a new current magnitude
126     i_new = i_curr + di_curr;
127
128     % Quantize and limit
129     i_new = round(i_new/I_STEP)*I_STEP;
130     i_new = max(min(i_new,PRE_CORR_I_MAX),
        PRE_CORR_I_MIN);
131
132     % Re-calculate new effective current delta
133     di_curr = i_new-i_curr;
134
135     if LIMIT_VOLTAGE
136         % Calculate the peak voltage at the end of
        the phase
137         n_peak = ni_curr + PRE_CORR_PHASE_WIDTH;
138         v_peak = 0;
139         for k = 1:length(ni)
140             if n_peak > ni(k) && di(k) ~= 0
141                 v_peak = v_peak + di(k)*(RS + K*(
                    n_peak-ni(k))^ALPHA);
142                .mvc_count =.mvc_count+1;
143             end
144         end
145
146         % Calculate the new peak voltage, and adjust
        if outside limits
147         v_peak_new = v_peak + di_curr*(RS + K*
            PRE_CORR_PHASE_WIDTH^ALPHA);
148        .mvc_count =.mvc_count+1;
149         v_adj = 0;
150         if v_peak_new > PRE_CORR_V_MAX
151             v_adj = v_peak_new-PRE_CORR_V_MAX;
152         elseif v_peak_new < PRE_CORR_V_MIN
153             v_adj = v_peak_new-PRE_CORR_V_MIN;
154         end
155
156         if v_adj ~= 0
157             di_adj = -v_adj/(RS + K*
                PRE_CORR_PHASE_WIDTH^ALPHA);
158            .mvc_count =.mvc_count+1;
159             i_new = i_new + di_adj;
160         end
161

```

```

162         % Quantize and limit
163         i_new = round(i_new/I_STEP)*I_STEP;
164         i_new = max(min(i_new,PRE_CORR_I_MAX),
165                    PRE_CORR_I_MIN);
166
167         % Re-calculate new effective current delta
168         di_curr = i_new-i_curr;
169     end
170
171     % Update with the new current
172     i_curr = i_new;
173
174     % Update the correction point voltage error
175     ve_pre = ve_pre+di_curr*K*((PRE_CORR_POINT-
176        ni_curr)^ALPHA-(PRE_CORR_POINT-ni_curr-
177        PRE_CORR_PHASE_WIDTH)^ALPHA);
178     mvc_count = mvc_count+1;
179
180     % Record current and subtract delta from next
181     % phase to ensure
182     % the final current is always zero
183     ni_pre(i) = ni_curr;
184     ni_pre(i+1) = ni_curr+PRE_CORR_PHASE_WIDTH;
185     di_pre(i) = di_pre(i)+di_curr;
186     di_pre(i+1) = di_pre(i+1)-di_curr;
187
188     % Update the total waveform
189     ni = [ni_pre;ni_wp;ni_post];
190     di = [di_pre;di_wp;di_post];
191
192     % Increment to next corrective phase
193     ni_curr = ni_curr+PRE_CORR_PHASE_WIDTH;
194 end
195
196 % Calculate Pre-Corrective Pulse
197 if iter < MAX_CORR_ITERATIONS || ~POST_CORR_SKIP_LAST
198     ni = [ni_pre;ni_wp;ni_post];
199     di = [di_pre;di_wp;di_post];
200
201     ni_curr = PRE_CORR_WIDTH+PULSE_WIDTH;
202     i_curr = 0;
203
204     % Calculate the error voltage at the post-correction
205     % point
206     nm_post = POST_CORR_POINT;
207     ve_post = 0;
208     for k = 1:length(ni)

```

```

205         if nm_post >= ni(k) && di(k) ~= 0
206             ve_post = ve_post + di(k)*(RS + K*(nm_post-
                ni(k))^ALPHA);
207             mvc_count = mvc_count+1;
208         end
209     end
210     for i = 1:POST_CORR_N_PHASES
211
212         % Post-correction current up until this point
213         i_curr = i_curr + di_post(i);
214
215         % Calculate the correction current delta
216         % Since the correction point is (should be)
217         % after the end of
218         % the pulse, the Rs component is neglected.
219         di_curr = -ve_post/(K*((POST_CORR_POINT-ni_curr)
                ^ALPHA-(POST_CORR_POINT-ni_curr-
                POST_CORR_PHASE_WIDTH)^ALPHA));
220         mvc_count = mvc_count+1;
221
222         % We have a new current magnitude
223         i_new = i_curr + di_curr;
224
225         % Quantize and limit
226         i_new = round(i_new/I_STEP)*I_STEP;
227         i_new = max(min(i_new,POST_CORR_I_MAX),
                POST_CORR_I_MIN);
228
229         % Re-calculate new effective current delta
230         di_curr = i_new-i_curr;
231
232         if LIMIT_VOLTAGE
233             % Calculate the peak voltage at the end of
                % the phase
234             n_peak = ni_curr + POST_CORR_PHASE_WIDTH;
235             v_peak = 0;
236             for k = 1:length(ni)
237                 if n_peak > ni(k) && di(k) ~= 0
238                     v_peak = v_peak + di(k)*(RS + K*(
                            n_peak-ni(k))^ALPHA);
239                     mvc_count = mvc_count+1;
240                 end
241             end
242
243             % Calculate the new peak voltage, and adjust
                % if outside limits

```

```

244     v_peak_new = v_peak + di_curr*(RS + K*
                POST_CORR_PHASE_WIDTH^ALPHA);
245     mvc_count = mvc_count+1;
246     v_adj = 0;
247     if v_peak_new > POST_CORR_V_MAX
248         v_adj = v_peak_new-POST_CORR_V_MAX;
249     elseif v_peak_new < POST_CORR_V_MIN
250         v_adj = v_peak_new-POST_CORR_V_MIN;
251     end
252
253     if v_adj ~= 0
254         di_adj = -v_adj/(RS + K*
                POST_CORR_PHASE_WIDTH^ALPHA);
255         mvc_count = mvc_count+1;
256         i_new = i_new + di_adj;
257     end
258
259     % Re-quantize and limit
260     i_new = round(i_new/I_STEP)*I_STEP;
261     i_new = max(min(i_new,POST_CORR_I_MAX),
                POST_CORR_I_MIN);
262
263     % Re-calculate new effective current delta
264     di_curr = i_new-i_curr;
265     end
266
267     % Update with the new current
268     i_curr = i_new;
269
270     % Update the correction point voltage error
271     ve_post = ve_post+di_curr*K*((POST_CORR_POINT-
                ni_curr)^ALPHA-(POST_CORR_POINT-ni_curr-
                POST_CORR_PHASE_WIDTH)^ALPHA);
272     mvc_count = mvc_count+1;
273
274     % Record current and subtract delta from next
                phase to ensure
275     % the final current is always zero
276     ni_post(i) = ni_curr;
277     ni_post(i+1) = ni_curr+POST_CORR_PHASE_WIDTH;
278     di_post(i) = di_post(i)+di_curr;
279     di_post(i+1) = di_post(i+1)-di_curr;
280
281     % Update the total waveform
282     ni = [ni_pre;ni_wp;ni_post];
283     di = [di_pre;di_wp;di_post];
284
285     % Increment to next corrective phase

```

```

286         ni_curr = ni_curr+POST_CORR_PHASE_WIDTH;
287     end
288 end
289
290 % If we have converged, stop
291 if max(di-di_prev) <= I_STEP
292     break;
293 end
294 di_prev = di;
295 end
296
297 ni = [ni_pre;ni_wp;ni_post];
298 di = [di_pre;di_wp;di_post];
299
300 % Plot final model voltage (for demonstration only)
301 nm = 0:PULSE_SUFFIX;
302 vm = zeros(size(nm));
303 for j = 1:length(nm)
304     for k = 1:length(ni)
305         if nm(j) >= ni(k) && di(k)~=0
306             vm(j) = vm(j) + di(k)*(RS + K*(nm(j)-ni(k))^
                ALPHA);
307         end
308     end
309 end
310 plot(nm,vm);

```

A.3 Algorithm Test Parameters

Parameter	Description	Value
pulse_width	Width of the anodic (working) phase	50,100,150,200,250us
pulse_amp	Amplitude of the anodic (working) phase	-100,-200,-300,-400,-500nA
corr_phase_width	Width of each corrective phase	20us (2 samples)
num_pre_phases	Number of pre-correction phases (before the working phase)	floor(pulse_width/corr_phase_width)
pre_corr_point	When the pre-correction phases force the voltage to zero	3*pulse_width
num_post_phases	Number of post-correction phases (after the working phase)	ceil(num_pre_phases)/2
post_corr_point	When the post-correction phases force the voltage to zero	2*pulse_width + 50us
max_iterations	Maximum number of shaping iterations	10
v_max	Maximum allowed model voltage	None
v_min	Maximum allowed model voltage	None
i_max	Maximum allowed correction current	2.55uA
i_min	Minimum allowed correction current	0uA