Artifact-Free Neurostimulator with Arbitrary Waveform Generation for High-Channel, High-Density Bidirectional Neural Interfaces

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

Master of Science in Electrical Engineering

by

Hsukang Chen

to be defended publicly on the 17th of January, 2024

Author:	H. Chen (4961293)
Project duration:	September 2022 – December 2023
Supervisor:	Dr. D.G. Muratore
Daily supervisor:	B.H.B.M. Abdelgaliel
Thesis committee:	Dr. D.G. Muratore Prof.dr.ir. W.A. Serdijn Dr.ir. M.A.P. Pertijs

TUDelft

Abstract

To achieve greater specificity in neurostimulation, bidirectional neural interfaces are required to verify the recorded neural response after stimulation. The specific neural interface targeted in this thesis is the epiretinal implant. Due to the heterogeneity of the retinal ganglion cells (RGCs), high-fidelity vision is only possible when all the types of RGCs near the neurostimulator are mapped. This necessitates the bidirectionality of the implant and poses significant difficulties, as large stimulation artifacts obscure the small neural response that needs to be recorded. Furthermore, in order to cover a large area of the retina, a channel count in the order of 10^4 will be required. Scaling existing neurostimulators would lead to chips that are >30 mm^2, which is too large. The aim of this thesis is therefore to design a neurostimulator for the epiretinal implant that is capable of implementing artifact-reducing algorithms, and is smaller than the current state-of-the-art. The proposed system makes use of a mismatch-based digital-to-analog converter (DAC), and has been optimized for an output range of 0–6 µA at an effective resolution of 8-bits. Furthermore, in order to decrease the amount of stimulation units required, waveform interleaving has been proposed, where the anodic and cathodic stimulator are separated. A voltage compliance monitor is also designed to ensure proper stimulation output. The designed system has been fabricated and occupies 0.0003 mm^2 for two channels. Scaling this directly to 10^4 channels would result in an area of 1.645 mm^2. This area can be reduced even further via electrode multiplexing, which the designed system readily allows for. An output availability (i.e. how many input codes are possible after calibrating) of 99.2% and 97.3% is reported for the anodic and cathodic stimulator at an 8-bit resolution over the full output range.

Acknowledgements

To start, I would like to express my gratitude to my supervisor, Dr. Dante Muratore, for his guidance throughout this journey on not only the technical details, but also on a personal level regarding career decisions. Allowing me to work independently has helped me grow a lot, and this growth will no doubt be of use to me in the future.

Having had Bakr Abdelgaliel as my daily supervisor also tremendously helped me. Even though you had just started your PhD, you were always available if I needed help with something. I am sure that you will find joy during and after your PhD, be it as a teacher or anything else.

A big thanks is also extended to all the fellow students and staff in the Bioelectronics group for creating a nice atmosphere to work in. All the events and activities were very helpful to relax after stressful periods.

Next, I would like to give recognition to the students that started their thesis at the same time as me and in particular: Ronald and Remy. Despite all the high intensity periods we shared, it was a very fun journey. I truly believe that you will be able to find the success you want in the rest of your lives.

Key to my time at TU Delft are of course also my parents. Thank you for your support and encouragement throughout these past years. I would not have been able to do this without your help.

Special mention is finally given to my brother. Even though we now live in completely different cities, your support has been immensely valuable to me. Thank you for always providing your thoughts on the problems I face.

Contents

1

Introduction

Artificially modulating the human nervous system via external means has been a potential treatment against neurological conditions for many decades now, starting with electrical neuromodulation for the brain as early as the 1960s. Over the years, as neuroscience and semiconductor technology have advanced, other nerves have been targeted and newer, smaller neural interfaces have been proposed. One such neural interface is also the focus of this thesis: the epiretinal implant.

An epiretinal implant is placed in front of the retina, and is in direct contact with the layer of retinal ganglion cells (RGCs). Fig. 1.1 shows the anatomy of the human retina and the relative position of the implant. Normally, vision is achieved through light entering the eye and reaching the photoreceptors in the retina. The photoreceptors detect the incoming light and send an electrical signal back through the interneurons and RGCs to the brain via the optic nerve for processing, after which vision is established [1].

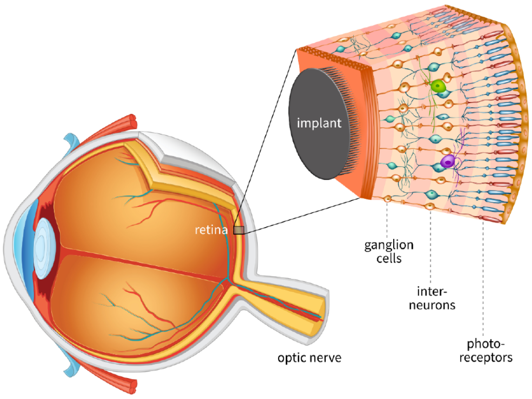


Figure 1.1: Location of the epiretinal implant inside the eye near the retina. Taken from [1].

However, if one or more of these three layers are damaged or otherwise malfunctioning, our ability to see can be severely impaired. If, for example, the photoreceptor layer is degenerated, as is the case with macular degeneration and retinitis pigmentosa, any incoming light will not be transduced into an electrical signal, and thus will result in (partial) blindness. An epiretinal implant essentially replaces the interneuron and photoreceptor layer by directly sending an electrical signal to the RGCs. With an external camera capturing light and the epiretinal implant using this information to generate the correct electrical signal, vision can be restored. Since there are more than 20 distinct types of ganglion cells, with each type responding differently to stimuli and having different sizes [2], generating the correct electrical signal becomes challenging. Sending every ganglion cell the same stimulus irrespective of their type will result in reduced vision quality [1].

Commercial epiretinal implants already exist and have even been implanted into patients. However, at the moment they only provide low resolution vision due to various reasons [3]: low (<100) channel counts, large electrode sizes, and not taking into account the heterogeneity of the ganglion cells. In order to improve upon this, many more channels (in the order of 10^4) are required, so that a large region of the RGC layer can be stimulated [4]. Smaller electrodes are also needed, so that ideally each RGC can be stimulated individually, providing single-cell resolution. Lastly, the aforementioned heterogeneity of the ganglion cells needs to be addressed. One method proposed to do so is by creating a dictionary of all the different RGC types, documenting their spontaneous activity and their response to various stimuli [4]. Then, using this dictionary, the type of ganglion cell present around each electrode can be identified and stimulation parameters can then be adjusted accordingly. The proposed method can be seen in Fig. 1.2. To be able to create this dictionary, the neural interface needs to be bidirectional, as the response of the RGC needs to be recorded directly after the stimulus has been given, since its response can occur as soon as 200 μs after stimulation.

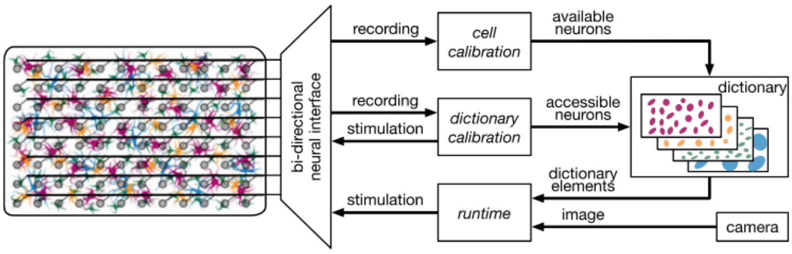


Figure 1.2: Illustration of the proposed method to address RGC heterogeneity. Cell calibration records spontaneous activity, and dictionary calibration records response to external stimuli. On the left the green dots are the electrodes, and the different colours represent different RGC types. Taken from [4].

The preferred method of stimulation is by using current. The main reason is that the injected charge is easier to control via current stimulation than via voltage stimulation. With voltage stimulation, the resulting current (and thus the injected charge) depends on the impedance of the electrode-tissue-interface (ETI), which can vary significantly, leading to inconsistent stimulation [5]. A conventional waveform for stimulation is the biphasic waveform (top part of Fig. 1.3). This waveform is often used to balance the amount of charge injected, as the charge in the positive and negative phase ideally cancel each other out. However, this is not always the case due to two reasons: (1) often the positive (anodic) and negative (cathodic) phase are not exactly matched, and (2) the ETI is non-linear. This leads to residual charge left in the tissue, which can cause irreversible damage and results in a residual artifact (bottom part of Fig. 1.3). While the direct artifact is not problematic, as there are no neural spikes elicited yet, the residual artifact presents a challenge for directly recording after stimulation. This residual artifact can be much larger in amplitude than the elicited response and thus obscure the signal that needs to be recorded. While the residual artifact decays over time, this process can take up to several milliseconds [6], which is too long as a neural response can occur within 500 μs of stimulation [7]. Even with perfectly charge-balanced stimulation waveforms, the aforementioned non-linearity of the ETI still causes the residual artifact.

To reduce the residual artifact, an algorithm called ”Fast Artifact Recovery Algorithm” (FARA) has been developed by [8]. This algorithm reduces the duration of the residual artifact by generating a non-standard triphasic stimulation waveform: it comprises a pre-correction, working, and post-correction phase. An example of the shape of this waveform can be seen in Fig. 1.4. An advantage of a triphasic waveform over a conventional biphasic waveform is that since the anodic phase is split into two, higher currents can be used during the cathodic phase, as the charge already present has to be removed first. In order to know what exact waveform is required, a model first needs to be made of the ETI. Then by using this model, different pre-correction and post-correction waveforms can be tested in an iterative manner to see if the artifact has been sufficiently reduced or not. In order to induce neuron activation, neurons need to be depolarized [9]. To do so, an outwards transmembrane current has to be generated [10], which is done in the cathodic phase. Therefore, the cathodic working phase is never adjusted in FARA so as to retain the desired neuronal response.

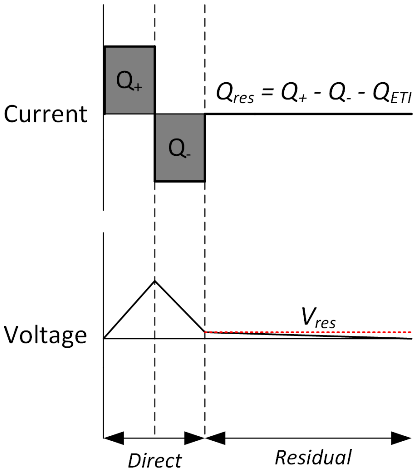


Figure 1.3: Example of a biphasic stimulation waveform. Top figure shows the waveform and its charge balancing. Bottom figure shows the recorded voltage of direct and residual artifact due to stimulation. Adapted from [8].

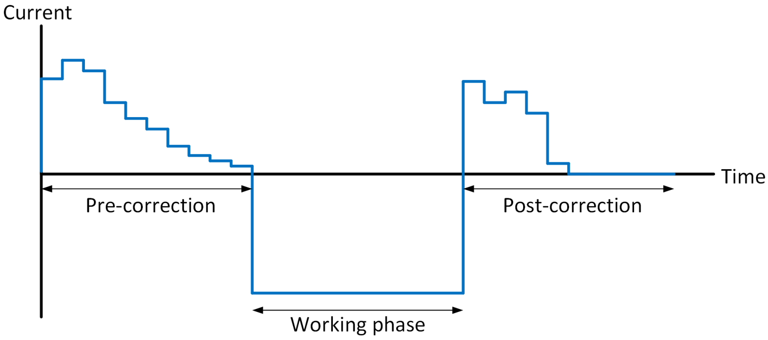


Figure 1.4: Example of a triphasic waveform that leads to reduced stimulation artifacts.

To be able to use this algorithm in an implant, an arbitrary waveform generator (AWG) is needed to produce the triphasic waveforms. This AWG can be implemented using a Digital-to-Analog Converter (DAC), and by changing its digital input, the analog output can be adjusted accordingly.

1.1. State of the Art

The neurostimulator can use either a monopolar or bipolar topology (Fig. 1.5a and Fig. 1.5b). A monopolar topology uses a sink current source and a source current source connected to the stimulation electrode, while the surrounding tissue acts as the return electrode, whereas a bipolar topology uses a single current source connected to an H-bridge (H-bridge topology) [11]. The main difference is in the use of the amount of current sources/sinks. The advantage of using only one current source (as in the case of the H-bridge topology) is that there is no current source mismatch that needs to be taken into account. Yet, this advantage only applies if the two phases (anodic and cathodic) need to be exactly the same, which is, in fact, not desired as it does not lead to perfect charge-balancing [12]. On the other hand, the source-sink topology allows for easier multi-channel implementation, as there is less cross-talk to take into consideration due to not sharing the same power supply in both phases when stimulating multiple channels simultaneously [11]. Furthermore, having a separate current source/sink for both the anodic phase and cathodic phase makes it possible to design each in a more optimized manner, potentially reducing area and power consumption. In terms of stimulation specificity, monopolar stimulation often affects a larger area due to not having a return electrode and thus having a wider charge distribution, meaning specificity is inherently lower compared to bipolar stimulation [13].

Another option is to use a source-sink topology connected to a bipolar (i.e. two electrode) set-up, as seen in Fig. 1.5c. In this case, the return electrode E2 can be either distant or local [14]. This topology allows for optimization of each source/sink, while still having relatively high specificity depending on whether a distant or local return electrode is used. Furthermore, since the return potential is set by the electrode, instead of by the surrounding tissue as in the case with the monopolar set-up, a negative potential is not needed for the current sink.

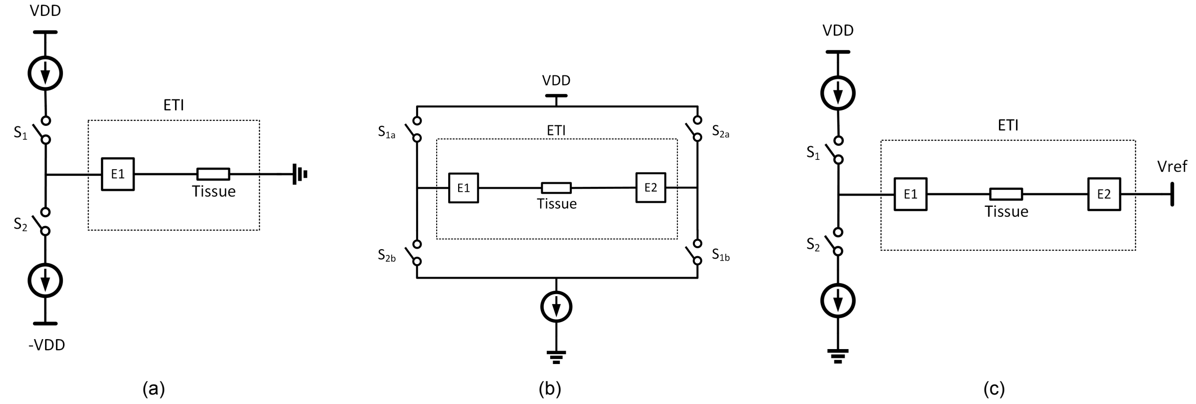


Figure 1.5: Schematic showing (a) the monopolar source-sink topology, (b) the bipolar H-bridge topology, and (c) the bipolar source-sink topology.

1.1.1. DAC architectures

There are four types of DAC architectures: voltage-domain, current-domain, charge-domain, and time-domain. Of these four, the most commonly used type that is used in neurostimulators is the current-domain DAC [3, 15–19]. This is primarily due to the low area requirement and direct current output, as will be explained further in this section.

Current-domain DACs (I-DACs) can be implemented in various architectures. Two popular examples include current-steering I-DACs [3, 17, 19] and M-βM I-DACs [15, 16].

1.1.1.1. Current-steering I-DACs

Current-steering I-DACs use current-producing cells, often CMOS transistors, and connect these cells to the output via switches when required by the digital input code. In this way, the output current is controlled. Depending on the weight of the current cells, a unary- (thermometric-), binary- or segmented-weighted current-steering I-DAC can be designed [19]. In a unary I-DAC, each current cell produces the same current I_{LSB}. Thus, each time the digital input goes up by one, one switch connects one extra current cell to the output. While this decreases the mismatch penalty on the differential non-linearity (DNL) and thus increases the potential resolution, as the mismatch affects only one LSB each step, the complete I-DAC becomes prohibitively large at higher resolution due to the doubling of the area for each extra bit. Furthermore, since the digital input is often binary coded, a binary-to-thermometric decoder is needed to control the switches, increasing digital overhead. In a binary I-DAC, on the other hand, each subsequent current cell has twice the weight (i.e. W/L ratio) of the previous cell. An evident advantage is that this takes much less area, as each bit only adds one extra switch and requires less routing. Unfortunately however, mismatch effects are now much larger, especially for the MSB cell. Due to these advantages and disadvantages of unary and binary structures, segmented structures are often used [20]. Segmented structures combine the two, and use unary for the MSBs, and binary for the LSBs. Structuring it in this way decreases mismatch effects on the MSBs, and decreases the area taken by the LSBs, where the mismatch effects are less significant inherently. An example of a 5-bit segmented current-steering I-DAC is shown in Fig. 1.6. Tab. 1.1 summarizes the INL and DNL of each of these three topologies.

Often with current-steering I-DACs, the current cell is capable to be switched between the output and a voltage supply (i.e. a current dump), instead of only to the output. The reason for this is that shutting off the transistors in the current cell and turning it on requires discharging and charging of the parasitic capacitances every time, reducing the speed of the DAC. By dumping the current through the voltage supply, this is prevented at the cost of increased power consumption. For the epiretinal implant, the

Table 1.1: INL and DNL summary of unary, binary, and segmented I-DACs. σ_u is the relative matching of a unit element (i.e. LSB cell), N is the resolution, and N_b is the resolution realized in binary coding.

	Unary	**Binary**	**Segmented**
INL	$\frac{\sigma_u\sqrt{2^N}}{2}$	$\frac{\sigma_u\sqrt{2^N}}{2}$	$\frac{\sigma_u\sqrt{2^N}}{2}$
DNL	σ_u	$\sigma_u\sqrt{2^N-1}$	$\sigma_u\sqrt{2^{N_b+1}-1}$

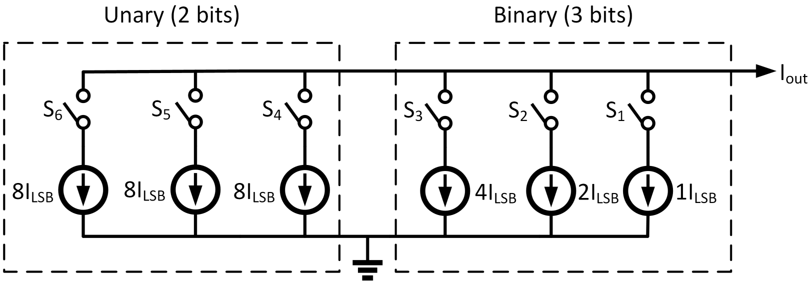


Figure 1.6: A 5-bit current-steering I-DAC, comprising a 3-bit binary part and 2-bit unary part.

reduced bandwidth is of no concern, as the bandwidth required is only in the order of 10^6 Hz [8], which can be easily achieved without the current dump, saving energy.

1.1.1.2. M-βM I-DACs

The other I-DAC architecture is the M-βM topology in Fig. 1.7. This topology is essentially the transistor version of the resistive R-2R ladder. By designing transistors M_{na} with a reduced W/L ratio relative to M_{nb}, the factor β is created. This β decides in which proportions the current splits horizontally and vertically (i.e. the currents I_{rn} and I_n). The digital input D_n determines whether the branch current I_n is connected to the dump or the load. As a result, the output current can be controlled with a radix of β. The resolution can be adjusted by adding and removing cells.

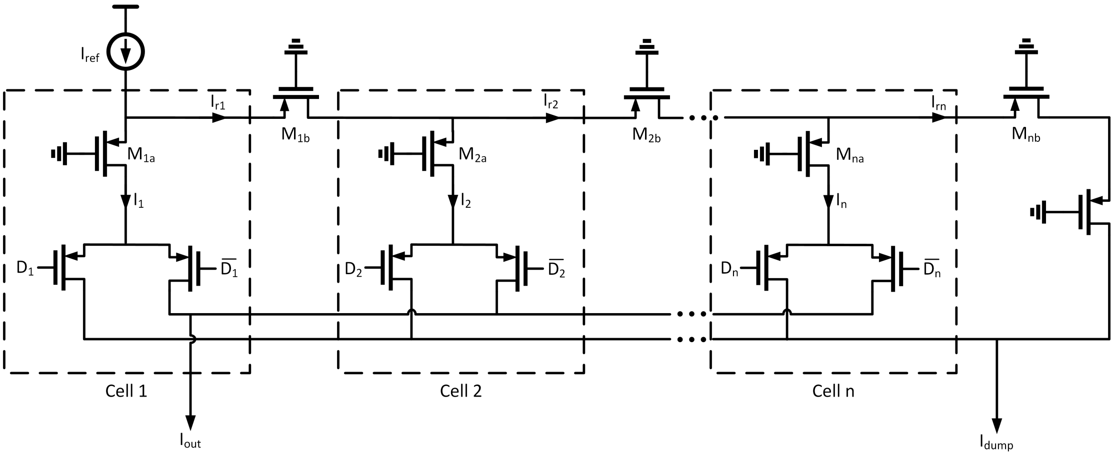


Figure 1.7: An example of an M-βM I-DAC with n bits resolution.

Compared to the current-steering I-DAC (without current dump), the M-βM I-DAC consumes more power due to the necessary presence of I_{dump}. However, it is more area-efficient, as adding one bit of resolution (e.g. if $\beta = 2$) costs only one extra cell. Even when compared to the binary current-steering I-DAC, the M-βM I-DAC performs better in terms of area-efficiency, as regardless of the resolution the same, minimum size transistors can be used in the cell, whereas with the current-steering variant a higher resolution leads to higher transistor widths or lower transistor lengths and as such lower area-efficiency [15].

1.1.1.3. Mismatch: a bug or a feature?

As briefly mentioned before, mismatch in the transistors can drastically reduce the effective resolution. Particularly when using smaller CMOS nodes (e.g. <65 nm), mismatch effects increase significantly. Since a 40 nm CMOS node will be used in this work, mismatch should be carefully considered. As mismatch increases when the transistor area is reduced, one solution to reduce mismatch effects is to increase the total area of the device. This evidently decreases area-efficiency and hence is not an optimal solution. Additionally, even with larger transistors, systematic mismatch can still occur due to layout effects. Layout techniques would then have to be applied such as common centroid or interdigitation, both of which potentially introduce dummy transistors, decreasing area-efficiency even more.

A better solution is to use the mismatch instead of trying to reduce its effects, such as with redundant crossfire (RXF) [21]. RXF creates redundancy in input codes by connecting the I-DAC to multiple current mirrors (CMs) and adding the outputs of these CMs together to generate the final output, as seen in Fig. 1.8. As a consequence, different digital input codes can result in the same output current if there is no mismatch. Conversely, some output currents are impossible to generate. However, when there is mismatch, the generated output current will randomly vary around its nominal value. By measuring the actual output current for each input configuration, a calibration look-up table (LUT) can be created for each in-out code. Fig. 1.9a and 1.9b demonstrate this diffusion effect, with no mismatch on the left, and 10% mismatch (standard deviation) on the right. With mismatch, there is less redundancy, but more output current levels are achieved. This allows for higher resolutions than the I-DAC is nominally capable of, at the cost of a calibration requirement and memory area to store the LUT. One significant disadvantage of such a system is that re-calibration is necessary every time operating conditions change (e.g., temperature). Fortunately, in the context of the epiretinal implant, this disadvantage is of no concern due to the relatively constant environment inside the eye, maintained by the human body.

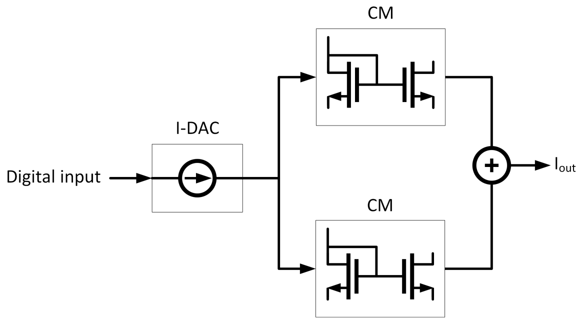
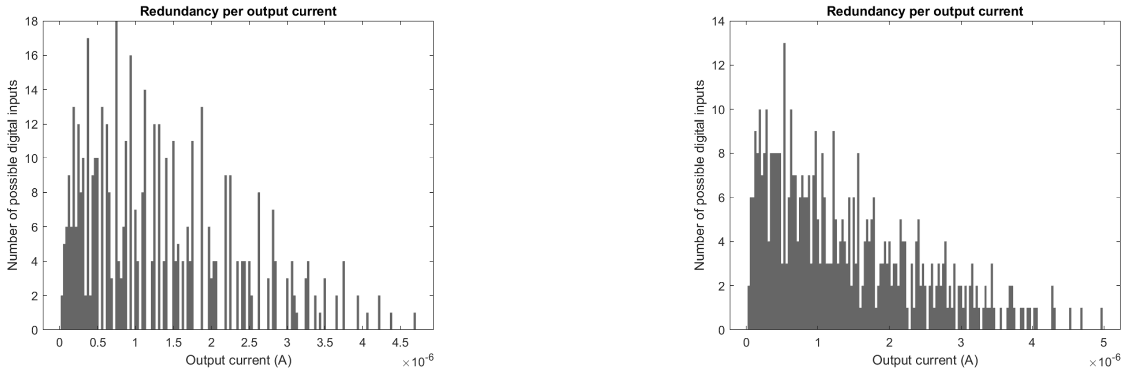


Figure 1.8: Example of RXF with an I-DAC and 2 CMs.



(a)

(b)

Figure 1.9: Redundancy with (a) no transistor mismatch, and (b) 10% transistor mismatch. Simulated is a 4-bit I-DAC, 3-bit CM, and a 2-bit CM. Displayed bin size has an output resolution of 6 bits.

1.1.2. ETI characteristics

Directly connecting the output of the DAC to the load (i.e. the electrode) would maximize the area and power efficiency. However, there are two main constraints to consider that stem from the high ETI impedance: (1) the output impedance of the DAC is typically significantly lower than of the ETI, and (2) the voltage generated across the ETI can push the DAC outside its required operation region. As a consequence of this, an output driver (OD) that can provide both a high output impedance and high voltage compliance is typically used. To be able to quantitatively characterize the required output impedance and voltage compliance of this output driver, the electrical model of the ETI will be discussed first.

A commonly used ETI model is the extended Randles model [22–26], as shown in Fig. 1.10. In this model, C_{dl} is the double-layer capacitance between the electrode and surrounding electrolyte (i.e. between electrode and tissue), R_{ct} is a voltage-dependent charge-transfer resistance, R_w and C_w together form Z_w and represents the diffusion of ions at the ETI, and is called the Warburg impedance. Finally, R_s is the spreading tissue resistance.

The main electrochemical difference between the two parallel paths is the type of processes that take place: the upper path (i.e. C_{dl}) represents the non-Faradaic processes, whereas the bottom branch (i.e. R_{ct} and Z_w) represents the Faradaic processes. The non-Faradaic path is also often called the reversible path, as charge only accumulates at the double-layer and therefore can be recovered by reversing stimulation polarity [27].

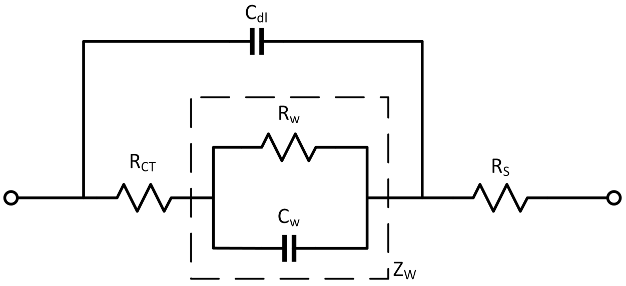


Figure 1.10: An extended Randles model of the ETI.

While exact values for the components can differ heavily depending on the geometry and material of the electrodes used, $Z_w + R_{ct}$ is often much larger than the impedance of C_{dl} at stimulating frequencies. The model can consequently be simplified to only the double-layer capacitance C_{dl} and spreading tissue resistance R_s [22, 25]. In order to quantify the impedance of C_{dl}, it can be considered a constant phase element Z_{CPE} [8, 22]:

$$Z_{CPE} = \frac{1}{A(j\omega)^{\alpha}} \tag{1.1}$$

where an α of 1 would give the impedance of a perfect capacitor, and an α of 0 would describe a pure resistance. As a consequence of the capacitive nature of the CPE, the ETI impedance dramatically increases at lower frequencies. Multiple empirical studies [5, 8, 28, 29] show that C_{dl} is the dominating impedance until 800 Hz. Directly after implantation in vivo, at a stimulating frequency of 2 kHz [8], the resistance R_s is by far the dominating aspect in the impedance of the ETI with a total absolute impedance in the order of 10^4 to 10^5 Ω for a gold/polyurethane electrode with a diameter of 30 µm. Since the electrode diameter inversely affects R_s, a smaller electrode diameter will quite significantly increase the total impedance [30]. The temporal variation in the impedance (both increasing and decreasing) is also clearly demonstrated by [5], where 10 days after implantation the measured impedance increased almost fourfold before decreasing again over time, caused by the development of scar tissue surrounding the electrode. Due to these variations, an output driver becomes necessary.

1.1.3. Output driver

In order to duplicate the output current of the DAC, a current mirror as seen in Fig. 1.11 is a natural choice. Since the drain current through M1 is forced by I_{ref}, the output current can be calculated to be:

$$I_{out} = I_{ref}\frac{(\frac{W}{L})_2}{(\frac{W}{L})_1}\frac{1+\lambda V_{DS2}}{1+\lambda V_{DS1}} \tag{1.2}$$

Equation 1.2 shows that the output current is the reference current multiplied by the gain $(W_2/L_2)/(W_1/L_1)$, with an error of $(1+\lambda V_{DS2})/(1+\lambda V_{DS1})$. This error can be fairly large, especially when using transistors with low lengths, as λ is inversely related to the transistor length. If this error is to be mitigated, it is thus important to equalize V_{DS1} and V_{DS2}. Furthermore, a standard CM unfortunately does not have a very large output impedance:

$$R_{out,CM} = r_o \tag{1.3}$$

where r_o is the small signal output impedance, which decreases as CMOS node size decreases.

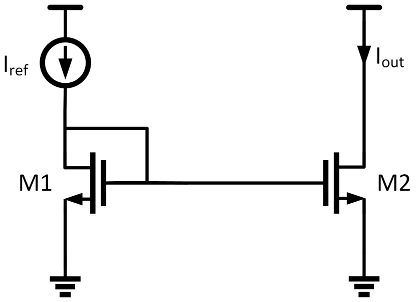


Figure 1.11: A standard current mirror.

To improve the output resistance, a cascode transistor can be added to the output branch to obtain the cascoded current mirror (CCM), as shown in Fig. 1.12a, giving the output impedance:

$$R_{out,CCM} = 2r_o + g_m r_o^2 \approx g_m r_o^2 \tag{1.4}$$

To match V_{DS1} and V_{DS2}, the bias voltage is set to $V_b = V_{GS1} + V_{GS3}$. Due to this bias voltage, the drawback of the CCM becomes apparent: the voltage compliance becomes $V_{TH} + 2V_{dsat}$.

To increase the voltage compliance without affecting the output impedance, the wide-swing CCM (WS-CCM) as shown in Fig. 1.12b is an alternative at the cost of one extra transistor. In this topology, in order to set V_{DS1} and V_{DS2} to be equal, a bias voltage of $V_b = V_{TH} + 2V_{dsat}$ is used. The wide-swing CCM thus has a voltage compliance of only $2V_{dsat}$.

It is evident that more cascode transistors can be added to improve the output impedance by a factor of $g_m r_o$ for each extra cascode used. However, doing so also reduces the voltage compliance by an additional $V_{TH} + V_{dsat}$ each time, and increases the area needed. Therefore, if a higher output impedance is needed and voltage compliance is of importance, a regulated cascode-current mirror (RCCM) is an often used choice.

RCCMs use an auxiliary amplifier to boost its output impedance by its gain (i.e. gain-boosting), while still having the voltage compliance of a CCM [31]. The most prominent drawbacks are its increased size and power consumption, due to the presence of the amplifier. An example RCCM can be seen in Fig. 1.13. In this example, transistors M1-M3 form the normal CCM, while the transistors M4-M7 make up the auxiliary amplifier. The gain of the auxiliary amplifier in this case is $A = (g_{m3} + g_{m5})(r_{o3}||r_{o5})$.

The four discussed output driver architectures and their parameters are summarized in Table 1.2.

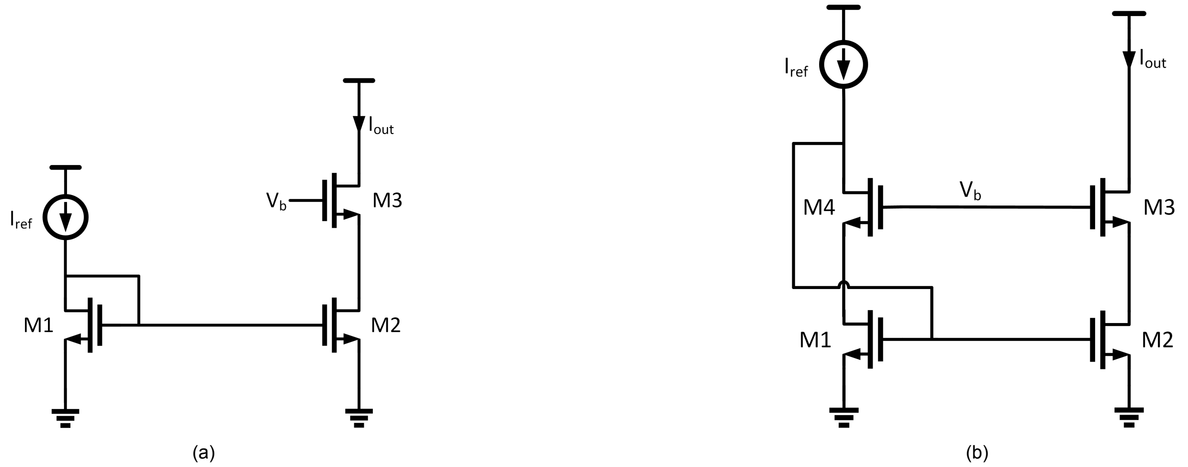


Figure 1.12: Schematic showing a) the cascoded current mirror and b) the wide-swing cascoded current mirror.

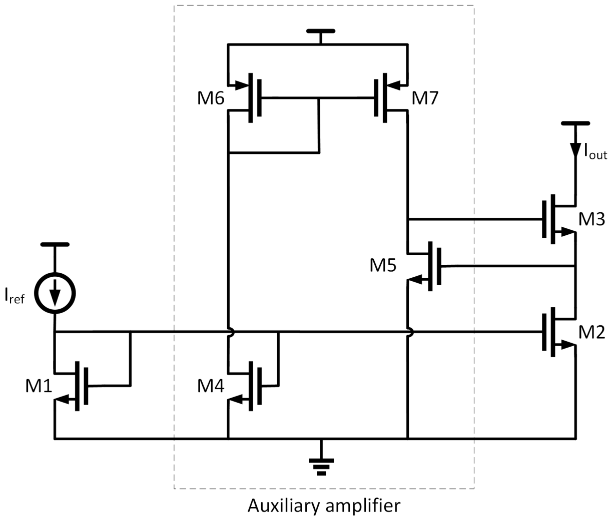


Figure 1.13: An example of a regulated cascoded current mirror. In this case, a single-ended one-stage op amp is used.

Table 1.2: Comparison between the discussed output driver topologies. Note that for RCCM different topologies are possible, and values indicated are for the topology as shown in Fig. 1.13.

	CM	**CCM**	**WS-CCM**	**RCCM**
Mirror accuracy	$\frac{1+\lambda V_{DS2}}{1+\lambda V_{DS1}}$	1*	1*	1*
Output impedance	r_o	$g_m r_o^2$	$g_m r_o^2$	$g_m r_o^2 A$
Voltage compliance	V_{dsat}	$2V_{dsat} + V_{TH}$	$2V_{dsat}$	$2V_{dsat} + V_{TH}$

* Assuming perfectly matched transistors.

1.1.4. Stimulation threshold

In reference to Section 1.1.2, the required voltage compliance depends heavily not only on the discussed ETI impedance, but also on the required stimulation current. Therefore, it is important to establish what the stimulation threshold is for RGCs, and on which parameters this threshold is dependent. From a power-efficiency perspective, it is preferable to stimulate at a current amplitude near the lower stimulation threshold. However, even from a neurological point of view, it is of interest to not stimulate with too large a current, as there also exists an upper stimulation threshold after which neural spikes no longer occur as a consequence of hyperpolarization [32, 33]. Nevertheless, as the upper stimulation threshold is frequently found to be at least one order of magnitude larger than the lower stimulation threshold [32, 33], only the lower stimulation threshold will be further elaborated.

The lower stimulation threshold is determined by several factors: electrode diameter, electrode material, duration of the stimulation pulse, shape of the stimulation pulse, spatial proximity to the stimulation site, etc. [34]. Especially electrode diameter, stimulation pulse duration, and distance between the stimulation site and the electrode can heavily alter the stimulation threshold.

Electrode diameter proportionally affects the threshold (i.e. a smaller electrode requires lower stimulation currents) [34, 35], as seen in Fig. 1.14. This proportional relation can be explained by the current density: a smaller electrode requires less current to achieve the same current density as a larger electrode, as a more dense concentration of charge leads to a stronger electrical field, which in turn increases the probability of stimulation.

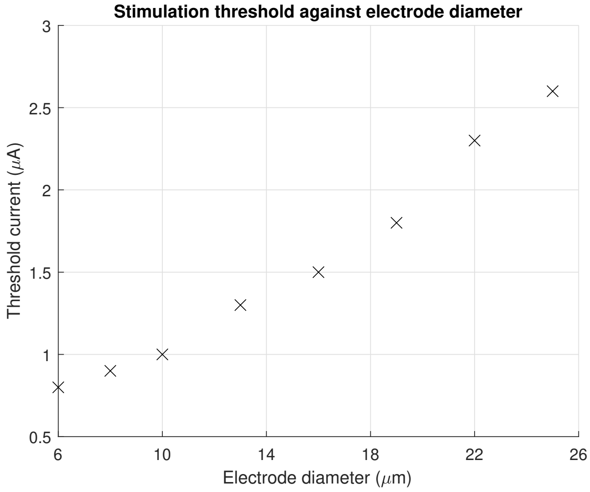


Figure 1.14: Stimulation threshold current against electrode diameter. A biphasic stimulation pulse of 0.05ms is used. Data from [34].

With the same logic, the correlation between the duration of the stimulation pulse and the threshold can be derived to be an inverse relation. A longer stimulation pulse results in more current, and hence more charge being injected into the stimulation site and consequently increases spike probability [32, 35]. However, this relationship is not a linear one. This is because the activation mechanism is dominated by different factors, depending on the pulse duration (Tab. 1.3) [36]. In short, during the rheobase regime the activation mechanism is inflow of sodium during stimulation. In the sodium-dominated regime, pulses are short enough that deactivation of the sodium channels instantly occurs after stimulation ends, leading to increased thresholds. In the potassium-dominated regime, the pulse duration is short enough that depolarization is no longer primarily due to the activation of sodium channels, but rather the deactivation of the potassium channels. Finally, in the cell polarization-limited regime the pulse has become so short that depolarization of the cell occurs after stimulation instead of during [36].

Table 1.3: Dominating activation mechanism at different pulse durations. τ_p is the cell polarization time, which depends on the target cell type. The threshold refers to the slope on a stimulation period (x-axis) vs threshold (y-axis) graph. A more negative value indicates a faster increase in threshold when decreasing period. Data from [36].

Pulse duration	**Mechanism**	**Threshold**
>5 ms	Rheobase	Constant
4 μs to 5 ms	Sodium-dominated	-0.72 (log-log scale)
τ_p to 4 μs	Potassium-dominated	-0.13 (log-log scale)
$<\tau_p$	Cell polarization-limited	-1 (log-log scale)

Finally, the closer the electrode is to the site of stimulation, the less attenuation occurs, and the more charge is injected into the stimulation site itself. Hence, stimulation thresholds decrease as distance decreases [32]. Similar to the pulse duration relation, this proximity relation is not linear, but exponential. This is primarily due to the fact that the current density decreases with the distance squared [37].

Empirical studies show that using a biphasic pulse with a period of 0.5 ms, with an electrode with a diameter of 15 µm, a stimulation threshold as low as 0.4 µA was found in vitro [32]. In another study, for an electrode with a diameter of 10 µm using a 0.05 ms biphasic pulse, a stimulation threshold of 1 µA was found in isolated pieces of rat, guinea pig, and monkey RGCs [34]. It becomes evident that the maximum stimulation current that the neurostimulator can output should be carefully designed on a case-by-case basis to fit the specific implant. Furthermore, all these studies used either a monophasic or biphasic stimulation waveform, rather than the non-standard triphasic waveform required in this work. As a result, the optimal architecture for DAC and output driver may change significantly between different neurostimulators.

It is important to emphasize that neuron activation and recruitment is not an on/off switch based on stimulation threshold, i.e. stimulation is not guaranteed to occur even when above the lower stimulation threshold [32]. It instead follows a sigmoidal shape (Fig. 1.15), where the stimulation threshold is often considered to be where the probability of activation is 50 %. Due to this, to be able to achieve single-cell resolution stimulation, it becomes increasingly important to have a range of possible stimulation output currents even if the exact stimulation threshold is known.

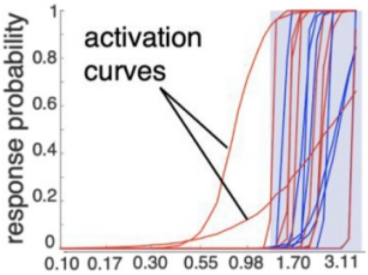


Figure 1.15: Example of sigmoidal activation curves showing probability of evoking a response against stimulation current. Taken from [38].

1.1.5. Multi-channel implementation

A channel count at least in the order of 10^4 is preferred for the epiretinal implant. Simply creating 10^4 DACs and output drivers is impossible due to area and power constraints. Table 1.4 shows other on-chip neurostimulation implants and it can be seen that scaling most of these up to 10^4 channels would require a stimulation chip in the order of cm^2, with [39] and [16] being the only two remaining under 1 cm^2. Unfortunately, the design of [39] does not incorporate an output driver yet and thus is not capable of stimulating accurately at epiretinal load impedances. Adding an output driver would inevitably increase the area needed. Furthermore, both of these stimulators only support biphasic waveforms, and not the triphasic waveform that is required. More importantly, the total estimated area of the entire epiretinal implant is 10 mm^2, with 1 mm^2 assigned to the stimulator, making it evident that scaling current state-of-the-art neurostimulators is unfeasible.

Table 1.4: Comparison between the DACs used in state-of-the-art neurostimulators.

	[40]	**[39]**	**[16]**	**[41]**	**[15]**	**[42]**
Year	2022	2020	2020	2018	2017	2016
Result type	IC	IC	Post-layout	Post-layout	IC	IC
Technology	180 nm	130 nm	130 nm	350 nm	180 nm	350 nm HV
Channels	1	32	1	1	8	4
Type DAC/OD	CS / CCM	CS / N/A	M-βM / WS-CCM	M-βM / WS-CCM	M-βM / RCCM	CS / N/A
Total chip area	1.197 mm^{2*}	0.736 mm^{2*}	0.0024 mm^2	0.0189 mm^2	2.25 mm^{2*}	10 mm^{2*}
Channel area	0.08 mm^{2**}	0.0033 mm^{2**}	0.0024 mm^2	0.0189 mm^2	0.0638 mm^2	1.2 mm^2

* Includes peripheral circuitry.
** Estimate.

Instead, what can be done is that each stimulation unit (comprising the DAC and its corresponding output driver) controls N electrodes via a multiplexer in a process also known as temporal division multiplexing (TDM) [43]. In this way, the amount of stimulation units required is reduced by a factor N. Evidently, this would mean that the implant is incapable of simultaneous stimulation of all channels. Fortunately however, it is undesirable to stimulate all channels at the same time regardless, as cross-talk between neighbouring stimulating electrodes would not only undermine the artifact-free waveform generating algorithm, but also significantly increase computational complexity for any desired precise stimulation patterns [44].

Ultimately, obtaining high-fidelity in the elicited vision is the primary objective. Since the brain does not process neuronal spikes from the RGC infinitely fast and in fact most likely has an integration time of tens of ms [7, 43, 45], time dithering does not always have to lead to reduced vision quality. Key to this is presenting all the required stimulation pulses within the brain's integration interval. Therefore increasing stimulation pulse frequency can be an option to reduce the total amount of stimulation units needed, with the drawback of having more stringent voltage compliance on the output driver requirements due to the higher stimulation current required. For example, an algorithm developed by [43], the "greedy algorithm", incorporates temporal dithering and still obtains a single perceived image for the brain. In this algorithm, the electrodes are chosen such that each subsequent time step maximally reduces the expected error in the perceived visual image. Fig. 1.16 illustrates this idea of temporal dithering.

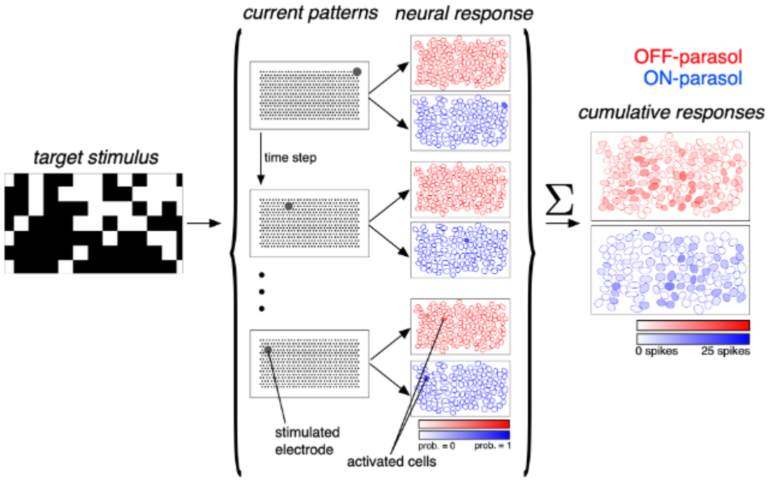


Figure 1.16: Stimulation pattern with temporal dithering, in which each time step one electrode is chosen to be stimulated. If this is done fast enough, the brain only perceives the cumulative response on the right. Taken from [43].

1.2. Thesis outline

Chapter 2 provides a system-level overview of the neurostimulator. It details the important parameters of each sub-block, and defines the system architecture that is designed, alongside elaborations of the specific reasons. Chapter 3 then implements the system-level blocks on a transistor-level, and shows simulation results. Then, Chapter 4 shows the methodology of the measurements performed, and its corresponding results. Finally, Chapter 5 contains the conclusion and recommendations for future work.

2

System architecture

Based on the available state-of-the-art designs and the analysis thereof, a system-level overview of the neurostimulator is presented. A bipolar architecture is chosen to avoid needing a negative rail voltage. A negative rail voltage is undesired due to then also requiring deep N-well transistors, which are larger. Furthermore, having a return electrode improves stimulation specificity.

To reduce the complexity of the artifact-reducing algorithm FARA, a bipolar source-sink configuration is used. In addition to reduced algorithm complexity, splitting the neurostimulator into two distinct parts (an anodic part and a cathodic part) increases the neurostimulator's flexibility in design. This flexibility allows for the eventual design to be tailored to each phase of the waveform specifically, instead of having to always satisfy the stricter specifications. By splitting the neurostimulator it now also becomes possible to interleave the cathodic and anodic phases during stimulation, which will be explained next.

2.1. Stimulation interleaving

The FARA algorithm outputs a triphasic waveform, as described in Chapter 1. While the standard way of stimulation is to provide the full waveform to one electrode and then move to the next electrode, this is not the most time-efficient method. Proposed is the time-interleaving of the waveforms. Instead of waiting until each electrode has finished stimulating the full waveform, the DACs moves on a phase-by-phase basis. In other words, while one electrode is in its cathodic phase, another electrode can already begin its first anodic phase. Considering that both phases have the same pulse duration, both phases will finish at the same time, resulting in continuous stimulation. Fig. 2.1a and 2.1b show an example of this interleaving. By doing this, two electrodes can be stimulated in only four cycles, as opposed to six in the standard method. This increases the amount of electrodes that can be stimulated within the brain's integration interval by 50% (assuming all three phases are equally long), without affecting the frequency of the pulses and hence without affecting the stimulation threshold and voltage headroom. As a result,

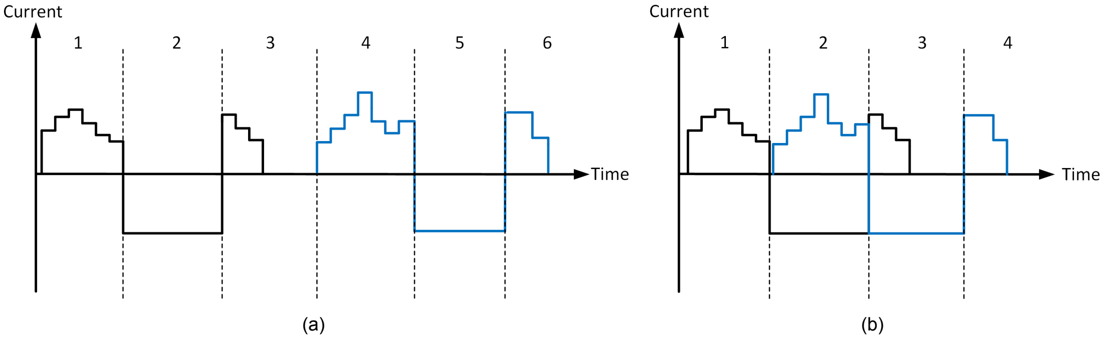


Figure 2.1: Example waveforms showing (a) the standard method of stimulation, and (b) the proposed interleaved stimulation. The black and blue lines represent different load electrodes.

fewer stimulation units are required for the same channel count, decreasing the area taken. Since this requires the stimulator to have separate current sources for the anodic and cathodic phase, the source-sink structure becomes essential. While concurrent, interleaved stimulation may present extra complexity for the artifact-reducing algorithm, as long as the two electrodes are sufficiently far away from each other, they will not affect one another and the complexity will therefore not be increased.

2.2. Neurostimulator overview

The neurostimulator is composed by the DAC, the output driver, and a multiplexer. Due to the source-sink structure, this means every stimulation unit has two DACs and two output drivers (one for the anodic phase, and one for the cathodic phase), as seen in Fig 2.2.

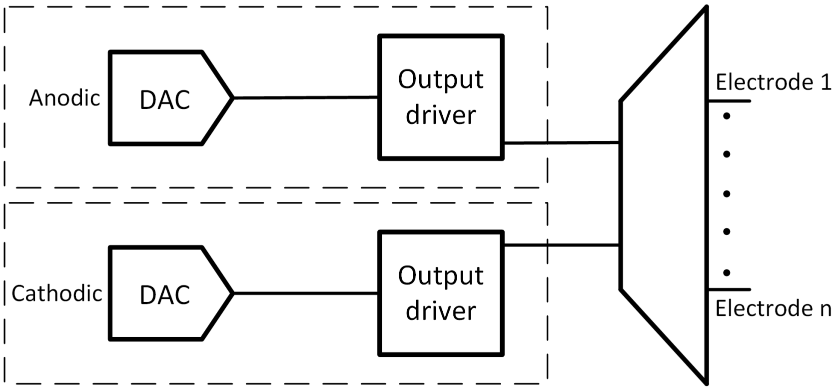


Figure 2.2: System-level of the stimulator unit with separate DACs and output drivers for the two phases.

2.2.1. DAC design

The RXF technique (Sec. 1.1.1.3) improves area efficiency by allowing for (close to) minimum size transistors and is thus used in this design. Since the native resolution of the DAC itself can now be lower due to the higher effective resolution in presence of mismatch, the area footprint of both the current-steering I-DAC and the M-βM I-DAC become almost identical, as the relative area-inefficiency of current-steering I-DACs is predominantly caused by the larger current cells present for bits at higher resolutions. Figure 2.3 shows the amount of unit transistors needed for both I-DAC architectures as resolution increases. Up to 4 bits the transistor count remains relatively equal, after which the current-steering I-DAC becomes notably larger.

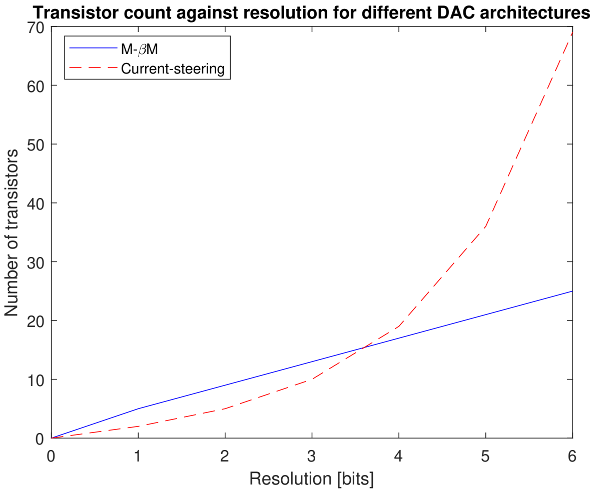


Figure 2.3: Amount of unit transistors needed for the two I-DAC architectures over increasing resolution.

Even though power consumption is not the main consideration of the design, it is still an important parameter due to the application being an epiretinal implant, which relies on wireless power transfer for its energy. Reducing the power consumption is beneficial by keeping tissue heating to a minimum, and by increasing the power budget available for the rest of the implant. The current-steering I-DAC has significantly lower power consumption, especially when not using a current dump. For these reasons, the current-steering topology is chosen as the DAC architecture.

Prior to going into detail on how to optimize the DAC for the epiretinal implant, it must first be established what exactly the requirements are in terms of the required resolution, bandwidth, and output range. Simulation data on neuron models from the continuation of the work of [8] show that the minimum resolution and current amplitude depend on the time period of the stimulation pulse. As mentioned before, FARA uses an anodic-first triphasic waveform. The two anodic phases are maximally composed of ten steps. The cathodic phase on the other hand is held constant for its entire duration of ten steps. This construction of the waveform is illustrated in Figure 2.4.

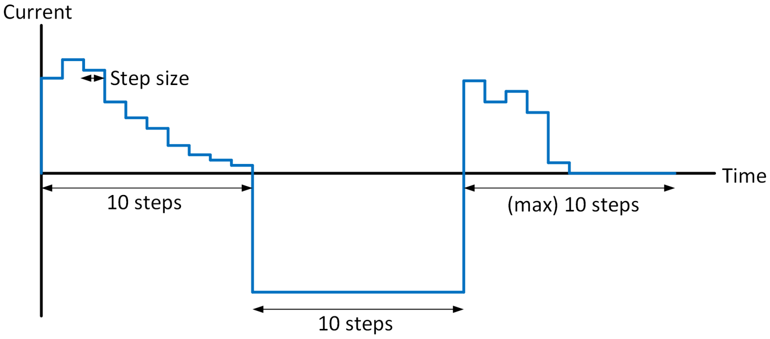


Figure 2.4: Example of the anodic-first triphasic waveform.

For a $1\,\mu s$ step duration, a resolution of 8 bits is needed, with a stimulation threshold of $5.26\,\mu A$[1]. As the step duration increases, the stimulation threshold decreases, but the necessary resolution increases (Fig. 2.5). For example, at a step duration of $5\,\mu s$, the resolution needs to be 10 bits and the threshold decreases to $2.25\,\mu A$. The resolutions are all based off of an output range of 0–$6\,\mu A$, and the stimulation thresholds are based off of the use of a $10\,\mu m$ electrode.

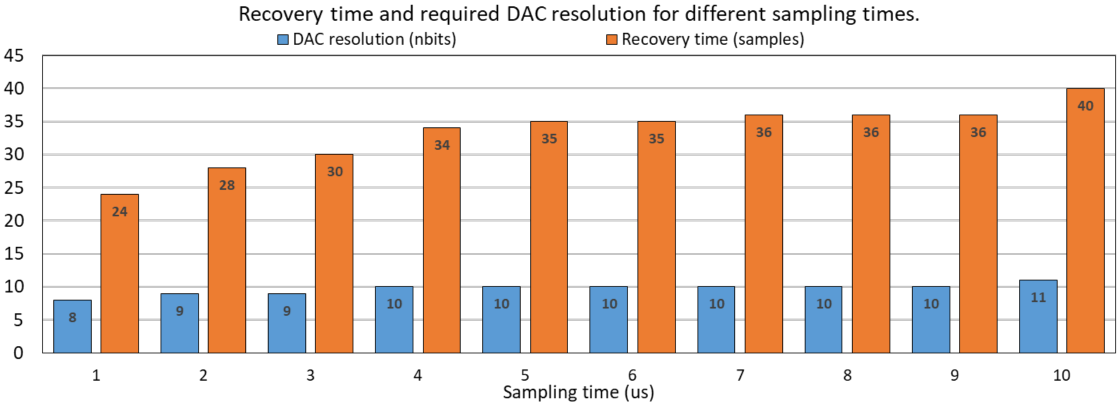


Figure 2.5: Required resolution of the stimulator for different step durations (sampling time) and resulting residual artifact duration.

For the design, there are two absolutes: (1) choosing only one step duration and thereby having the most efficient design for that specific step duration, or (2) taking into account the worst-case scenarios,

---

[1] These numbers were found using in-silico experiments, with RGC models, and were performed by another person in the group.

Table 2.1: Requirements for the DAC.

Parameter	**Requirement**
Resolution	8 bit
Minimum step duration	1 µs
Output range	0–6 µA

i.e. the highest required resolution (at longer step duration) and the highest required maximum current (at shorter step duration), and thereby having the least efficient design. Since flexibility is highly desirable to support different scenarios, the former option is not ideal. The latter option would provide the most flexibility at the cost of notably reduced area efficiency. However, only a Sith deals in absolutes. The proposed system is designed for the most stringent output range (i.e., 5.26 µA) and uses oversampling to achieve the higher resolution required for longer step duration[2]. For example, a 10-bit DAC with 4 µs step can be achieved using an 8-bit DAC with a 1 µs step. Fortunately, the stimulation of neurons and the residual artifact duration are predominantly dependent on the injected charge [8]. Therefore, due to the nature of neurostimulation, a higher output precision can be achieved by oversampling without requiring a higher resolution in the stimulator. The required DAC parameters are summarized in Table 2.1.

Due to the random nature of mismatch[3], a Monte Carlo (MC) simulation is used to obtain a representative quantification. The full output range is divided into 256 equal bins for the 8-bit resolution design. Then, for each MC run, the current output for each DAC code is checked to define which bin it falls in. The ratio of occurrence (RoO) for each bin is defined as

$$RoO_i = \frac{N_{available}}{N_{MC}} \tag{2.1}$$

where N_{MC} is the total amount of MC runs, and $N_{available}$ is the amount of MC runs that had at least one DAC code for which the output fell inside the bin width. For the standard topology seen in Figure 2.6, where the I-DAC has an LSB of 80 nA, the RoO is plotted in Fig. 2.7.

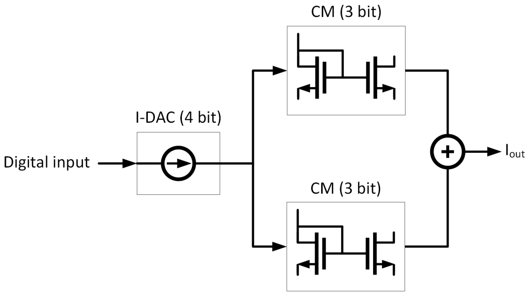


Figure 2.6: Topology using a 4-bit I-DAC and two 3-bit current mirrors.

There are two observations instantly noticeable: (1) even at the highest point in the curve, there is a RoO of only 90 %, and (2) at the tails of the output range the RoO dramatically drops. The reduced RoO at the tails is to be expected, as those are also the points where there is the least amount of redundancy possible. Overall, however, the RoO is simply insufficient for the target application. To improve this, three additional designs are considered[4] with (1) increased I-DAC resolution, (2) increased current mirror gain stages, and (3) an additional current mirror. The goal is to create more redundancy for higher output currents, and thereby increase RoO. These designs can be seen in Fig. 2.8 and the resulting RoO is plotted against the full range in Fig. 2.9.

The RoO for all three configurations is significantly higher at currents higher than 2 µA compared to the initial design. Table 2.2 summarizes the area overhead, the amount of calibration steps required

---

[2] For more details, see Appendix D.

[3] See Appendix B for more details about mismatch.

[4] More designs were tested than these three, and can be seen in Appendix C.

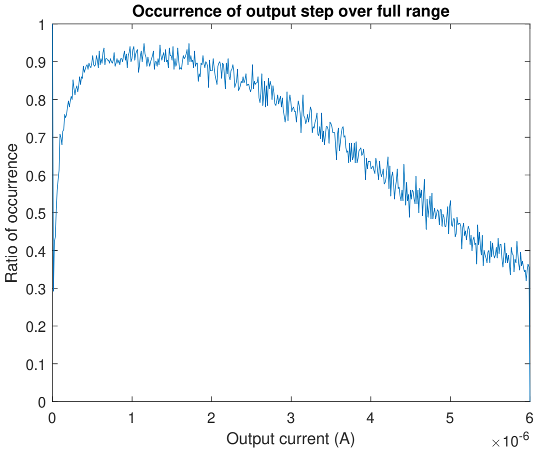
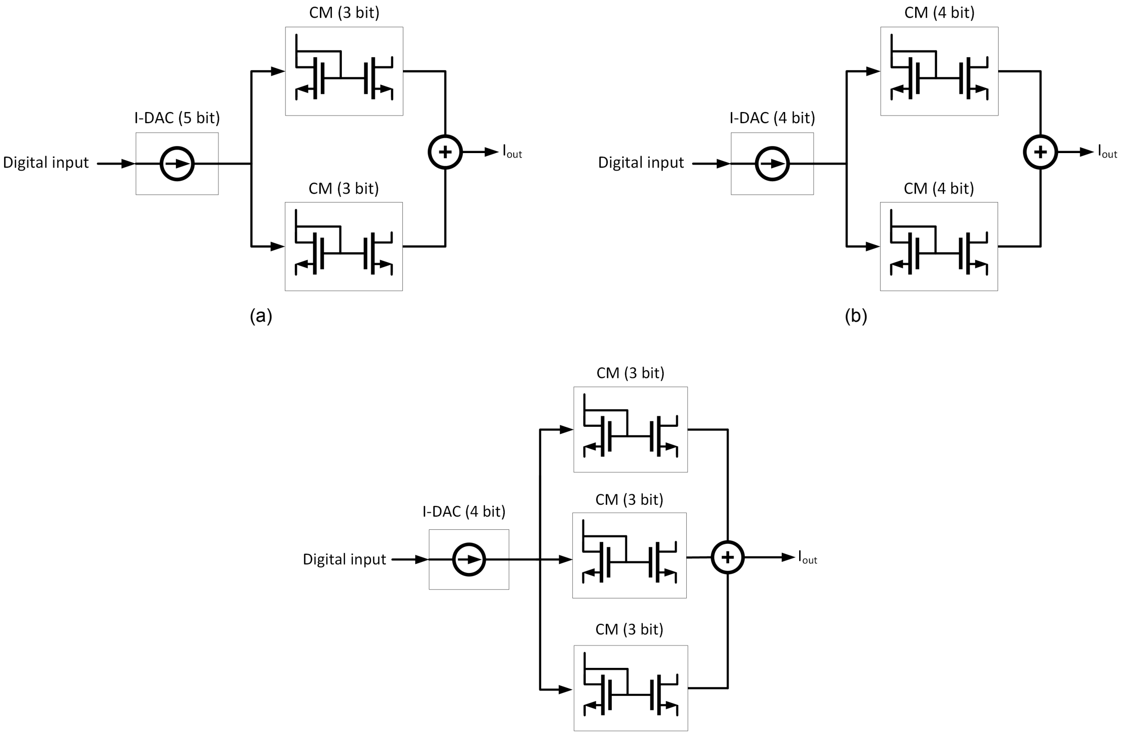


Figure 2.7: Ratio of occurrence over output range for the 4-bit I-DAC, in combination with two 3-bit current mirrors.



(c)

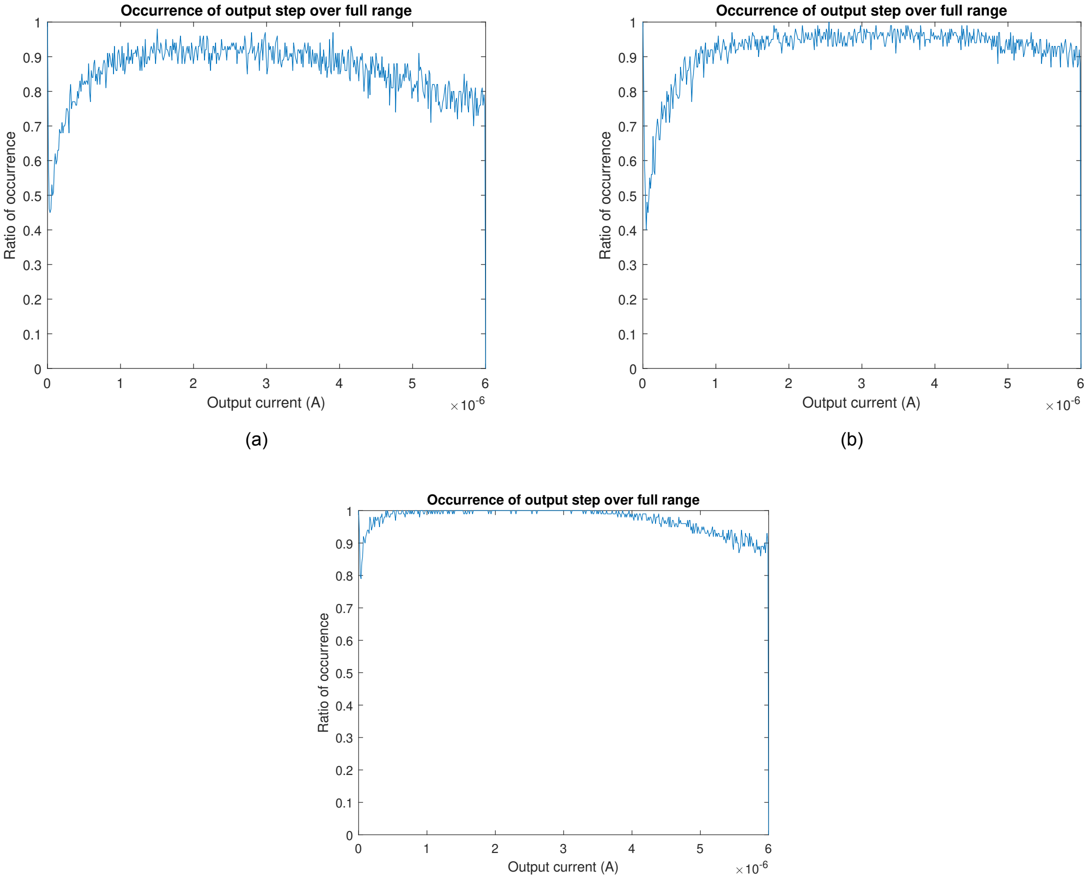
Figure 2.8: Topologies with (a) increased I-DAC resolution, (b) increased current mirror resolution, and (c) an additional current mirror.

(i.e. the total amount of configurations possible), and the RoO at three points for all three designs. The 4b DAC + 3x 3b CM has the best RoO over the full range, while requiring only 10 extra transistors. This comes at the cost of a longer calibration phase. However, infrequent foreground calibration is acceptable for the neurostimulator. Hence, the proposed design uses the 4b DAC + 3x 3b CM topology.

Unfortunately, the RoO at low currents (i.e. below 1 µA) is still very low. For the anodic phase, this will present problems as during this corrective phase small amplitudes are commonly used to make small adjustments to the residual artifact. On the other hand, for the cathodic phase, currents below 1 µA are often not used since they are insufficient to elicit a response from neurons. To address the low RoO for the anodic phase, two adjustments are made[5]: (1) a second, lower bias current is introduced to the

---

[5] More adjustments were tested, but are only shown in Appendix C.



(a)

(b)

(c)

Figure 2.9: RoO of the topologies with (a) increased I-DAC resolution, (b) increased current mirror resolution, and (c) an additional current mirror.

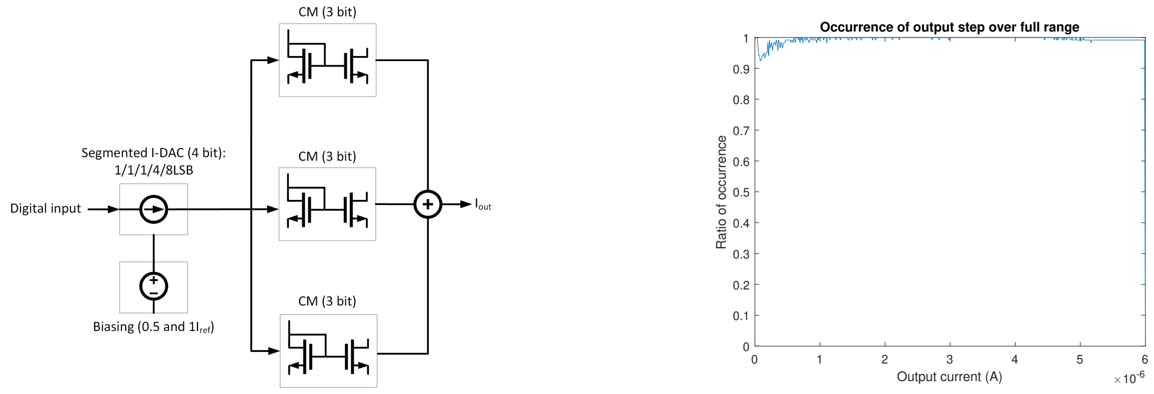
Table 2.2: Summary of the discussed three topologies and their parameters.

Topology	Added area (rel. to standard)	Calibration steps	RoO at [0, 0.5, 1] full range
4b DAC + 2x 3b CM	-	1024	[0.30, 0.78, 0.35]
5b DAC + 2x 3b CM	+17T	2048	[0.45, 0.89, 0.78]
4b DAC + 2x 4b CM	+18T	4096	[0.40, 0.94, 0.89]
4b DAC + 3x 3b CM	+10T	8192	[0.79, 0.99, 0.91]

I-DAC, and (2) the I-DAC is changed from binary to segmented (unary for the first two bits, binary for the last two bits).

The first adjustment shifts the RoO characteristic towards lower output currents. The second adjustment adds additional configurations at low output currents, increasing redundancy and RoO for low output currents. The proposed topology with both adjustments can be seen in Fig. 2.10a. The resulting RoO is shown in Fig. 2.10b and shows significantly better performance at low output currents.

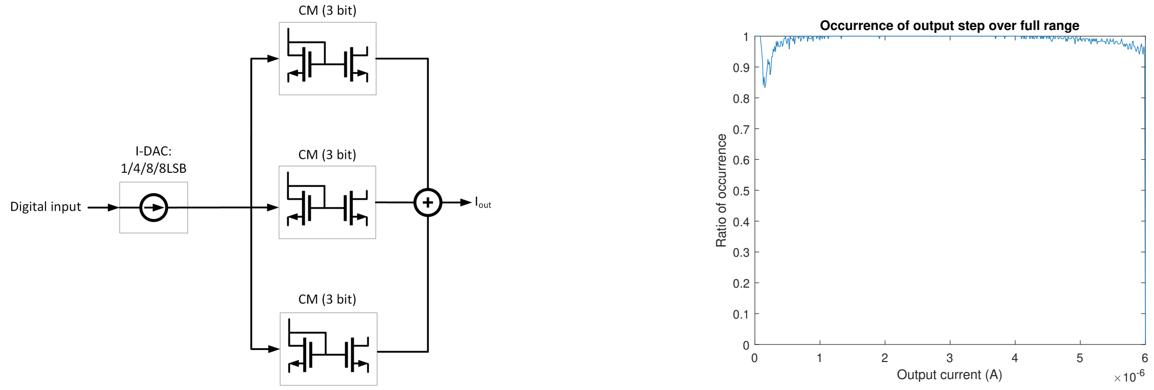
As mentioned before, the lower RoO at low output currents is not problematic for the cathodic working phase. However, high output currents are extremely important as insufficient amplitude step sizes can lead to stimulation of not only the target cell, but also the neighbouring cells. This can reduce the stimulation specificity, especially when taking into account that the electrode is not always directly above the target neuron. Hence, the topology should be adjusted to improve RoO also at high output currents. To do so, somewhat similar to the anodic phase, the binary I-DAC can be changed into a inverse-segmented one. Instead of making the LSBs unary, the MSBs are now made unary at the cost



(a) (b)

Figure 2.10: (a) Topology with an additional bias current and segmented I-DAC, and (b) its corresponding RoO.

of losing the 2LSB current cells. That is, the I-DAC now has the following cells: LSB, 4LSB, and 2x 8LSB (MSB). While this will increase the RoO at high outputs at the cost of RoO at low outputs, this is not a problem due to the reasons discussed earlier. The topology can be seen in Fig. 2.11a, and its resulting RoO can be seen in Fig. 2.11b.



(a) (b)

Figure 2.11: (a) Topology with a pseudo-segmented I-DAC, and (b) its corresponding RoO.

In summary, Table 2.3 shows the chosen topologies for each phase and their parameters. For comparison, a standard 8 bit current-steering I-DAC is also shown. However, important to emphasize here is that a standard I-DAC has to take mismatch into account, which inevitably increases the area even more, whereas with RXF (close to) minimum size transistors can be used.

Table 2.3: Summary of the chosen topologies for the DAC of each stimulation phase.

Phase	**Topology**	**Area**	**Calibration steps**	**RoO at [0, 0.5, 1] full range**
Anodic	Double-bias segmented 4b DAC + 3x 3b CM	54T	32768	[0.95, 0.99, 0.98]
Cathodic	Pseudo-segmented 4b DAC + 3x 3b CM	55T	8192	[0.85, 1.00, 0.94]
Standard	8b CS-DAC	263T	0	[1.00, 1.00, 1.00]

2.2.2. Output driver design

The output driver needs to have an output impedance (significantly) larger than the impedance of the load to reduce loading effects. An equivalent circuit model of a current source with its load can be drawn as seen in Fig. 2.12, where Z_{DAC} is the output impedance of the current source. From this circuit, the current through Z_{Load} (I_{Load}) can easily be derived to be

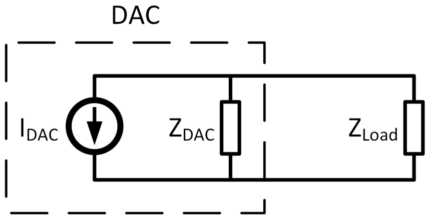


Figure 2.12: Norton equivalent circuit representing a current source with its output impedance and a load.

$$I_{Load} = \frac{I_{DAC} Z_{DAC}}{Z_{DAC} + Z_{Load}} \tag{2.2}$$

It is evident that the higher the output impedance Z_{DAC} becomes, the closer I_{Load} gets to I_{DAC}. In order to determine the minimum output impedance required, Eq. 2.2 can be rewritten to

$$Z_{DAC} = \frac{Z_{Load} I_{Load}}{I_{DAC} - I_{Load}} \tag{2.3}$$

Since the maximum output current is 6 µA, and a resolution of 8 bits is desired for the neurostimulator, the output impedance Z_{DAC} necessary is found to be at least $512 * Z_{Load}$.

In the case of this project, the load is considered to be the ETI. The true impedance of this load, as mentioned before, is dependent on several factors, such as the electrode used, the site of stimulation, and many more. Since the neurostimulator will be tested on a microelectrode array (MEA) that is already available in the group, the values of this MEA will be used for the design. The model used, and its corresponding electrical parameters can be found in Fig. 2.13 and Table 2.4. The impedance can be derived to be (at a frequency of 100 kHz)

$$Z_{Load} = \frac{1}{2\pi f C_{dl} + \frac{1}{R_{CT}}} + R_s = 104.4k\Omega \tag{2.4}$$

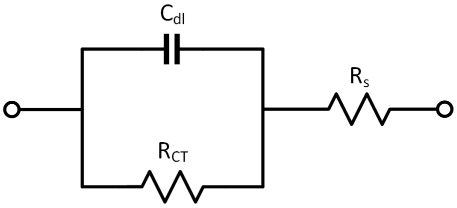


Figure 2.13: The used electrical model of the ETI.

Component	Value
C_{dl}	361 pF
R_s	100 kΩ
R_{CT}	39.7 MΩ

Table 2.4: The used electrical parameters for the ETI model.

The minimum output impedance required for the driver is then 53.5 MΩ. Using the same model and the stimulation parameters described previously, the minimum voltage compliance of the driver can be calculated to be

$$V_{compl} = V_{cm} - I_{out}(R_s + \frac{t_{stim}}{C_{dl}}) = 883.8mV \tag{2.5}$$

where t_{stim} is the pulse duration of a single phase (10 µs), and V_{cm} = 1.65 V is set at half of the supply voltage (3.3 V).

Regardless of topology, the output impedance will improve when using larger lengths, as r_o is inversely proportional to the length. By keeping the W/L ratio and drain current constant, the transconductance

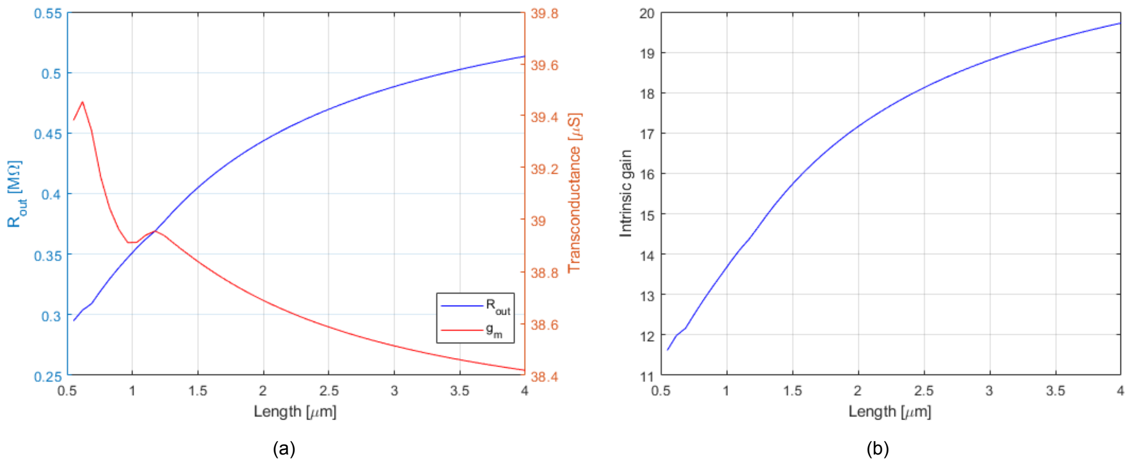


Figure 2.14: Graphs showing (a) the impedance and transconductance over increasing length, and (b) the resulting intrinsic gain over increasing length.

g_m also remains relatively constant. A sweep is performed with increasing transistor length (and equal W/L ratio). Fig. 2.14a and 2.14b show the results of this sweep.

As expected, the transconductance stays relatively constant (+-10%), whereas the output impedance increases with increasing length. As a consequence, the intrinsic gain also increases with increasing length. Given these values, using a cascoded current mirror (standard or wide-swing) will require transistor lengths well beyond $4\,\mu$m to achieve the required output impedance. For this reason, the regulated cascoded current mirror is chosen as the topology for the output driver. For convenience, the topology is shown again here in Fig. 2.15.

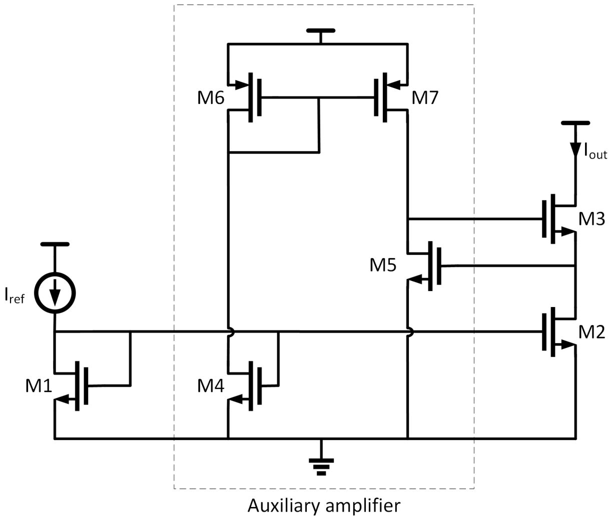


Figure 2.15: An example of a regulated cascoded current mirror. In this case, a single-ended one-stage op-amp is used.

Important to realize with this topology is that the two current mirrors (made by M4, M6, and M7) are not necessarily needed to achieve the boosted output impedance. Instead, they are used to equalize the V_{DS} of M1 and M2, thanks to transistor M5 carrying the same current as M1. Due to M5 and M1 carrying the same current, given that they are matched in size, they will also share the same V_{GS}, effectively creating a diode-connection for transistor M2. As a result, a greater mirroring accuracy is achieved at the cost of reduced voltage compliance (M5 needs at least V_{GS}, which is almost always larger than V_{Dsat}), increased area, and increased power consumption.

The proposed system is calibrated anyway, so a large mirroring accuracy is not needed. Hence, these two internal current mirrors are removed in the proposed design, shown in Fig. 2.16. By doing so, the bias current of M5 can control its V_{GS}, and, hence, the V_{DS} of M2 (as it is the same node). The bias current is chosen such that the V_{DS} of M2 is the minimum voltage needed to achieve the required output impedance, thereby achieving greater voltage compliance and lower power consumption.

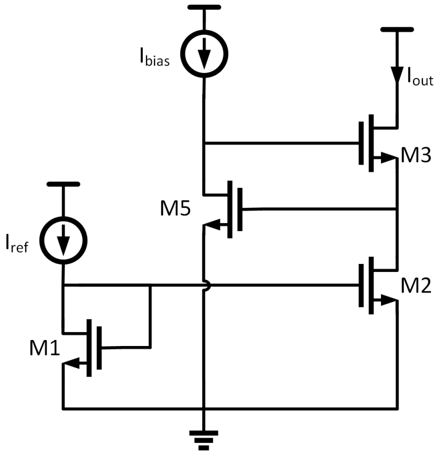


Figure 2.16: Proposed regulated cascoded current mirror without internal current mirrors.

Ideally, a feedback loop would be implemented to ensure that the absolute minimum voltage needed is always present at the V_{DS} of M2. For simplicity, a fixed bias current is chosen such that at the highest output current (i.e. 6 µA) M2 remains in saturation.

2.2.3. Peripheral circuitry

The neurostimulator also has some peripheral blocks. One such block is the voltage compliance monitor: a block that should monitor whether or not there is enough voltage headroom to satisfy the output impedance requirements. As mentioned before, due to the small node size it is not advisable to assume that the output impedance is sufficiently high as long as the transistors are in saturation. Yet, it is similarly inadvisable to compare the output voltage to a fixed, predetermined voltage above which a satisfactory output impedance is guaranteed, as this predetermined voltage would have to be taken larger than nominal to prevent false positives, resulting in the loss of voltage headroom[6].

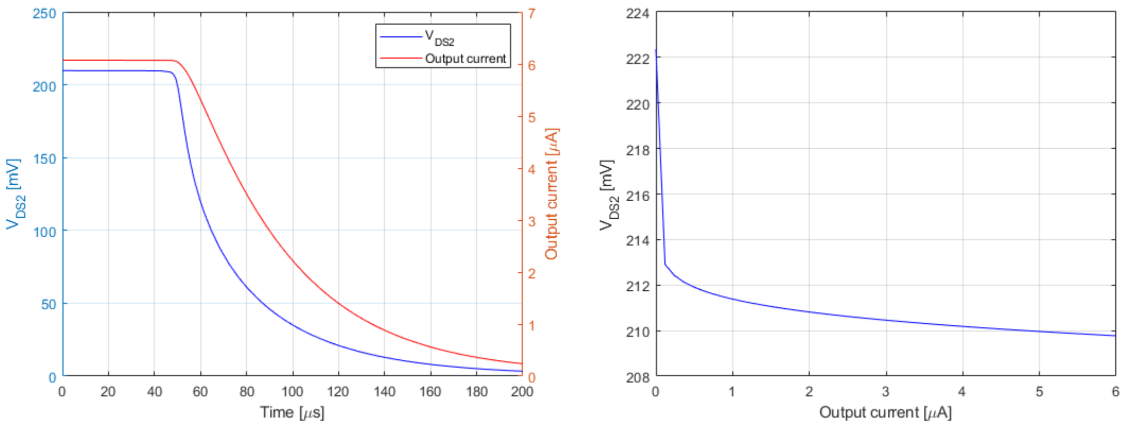
Instead, proposed is a system that directly uses the parameter that is the most important to the stimulator: the output current. In place of ensuring that the output voltage is always high enough, and thereby indirectly establishing a proper output current, the proposed monitor verifies that the possible change in output current is always within a tolerable range. By only taking into account the change in output current, the advantage of not being dependent on transistor parameters (i.e. PV variations) is gained.

Taking the proposed output driver structure of Fig. 2.16, M2 is the only transistor that ultimately determines the output current. Given that V_{GS2} is strictly set by the mirror transistor M1, V_{DS2} is the only voltage that will change depending on the output current. Therefore, it can be assumed that should the output current drop (e.g. due to not having sufficient output impedance), node V_{DS2} will also drop. This drop in voltage can then be measured, and if this drop is large enough, then a flag can be raised to indicate that something is causing the output current to drop. As an additional advantage, since this node is set to be close to the V_{Dsat} of a transistor, the comparator following hereafter does not require high voltage (HV) devices as would be the case if the output voltage were monitored, reducing the area taken. Unfortunately, node V_{DS2} is not expected to be truly fixed to a constant value and can be different depending on the starting output current. Fig. 2.17a and 2.17b show the previously described behaviour.

Zooming into Fig. 2.17a to where the monitor normally should raise a flag, obtained is Fig. 2.18. When taking an LSB of roughly 23 nA, a drop of about 5 mV is observed in V_{DS2}. Since Fig. 2.17b shows that

---

[6] The irony is not lost on me that despite being inadvisable, this approach is what was chosen nine lines prior for the bias current.



(a)

(b)

Figure 2.17: Graphs showing (a) the drop in V_{DS2} as the output currents starts to drop due to having insufficient voltage headroom, and (b) the varying V_{DS2} over the full output range in nominal conditions (i.e. with sufficient voltage headroom).

the initial V_{DS2} can already differ by 2 mV, it becomes impossible to use a fixed voltage as reference and still realize correct monitoring. Furthermore, to prevent the monitor from flagging excessively, a small offset must be maintained so that flagging only occurs when the difference becomes larger than the aforementioned 5 mV.

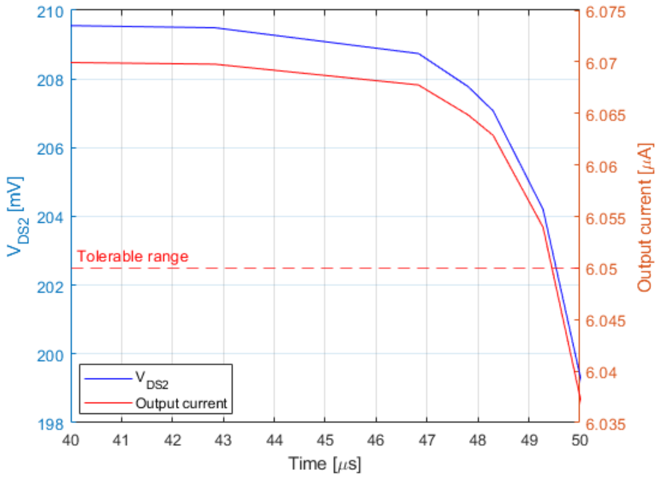


Figure 2.18: Zoomed in version of the V_{DS2} as output current starts to drop.

To overcome this issue of varying initial V_{DS2}, a sample and hold is implemented that samples this voltage at the start of the stimulation cycle and then uses this sampled voltage as a reference. The discussed system can be seen in Fig. 2.19: a comparator compares the sampled voltage to the real-time voltage and raises a flag when the discrepancy between the two inputs grows too large. This flag is subsequently fed into an OR gate to hold the flag high once triggered, until a reset is given.

Since the injected charge is almost always higher in the cathodic phase than in the anodic phase, due to the anodic phase being split into two parts, it is chosen that the monitor only monitors the cathodic phase, as during this phase the risk of running out of voltage headroom is the highest. Another reason for this is that because the anodic output driver comprises of PMOS devices that push current from 3.3 V into the ETI, area-inefficient HV transistors would need to be used for the monitoring circuitry. The clock timing of the switches and for the clocked comparator can be seen in Fig. 2.20. At the start of the stimulation for the cathodic phase, the reset switch pulls the flag low. The capacitor samples and holds V_{DS2} at the same time, and at the end of the stimulation period, the comparator is clocked to

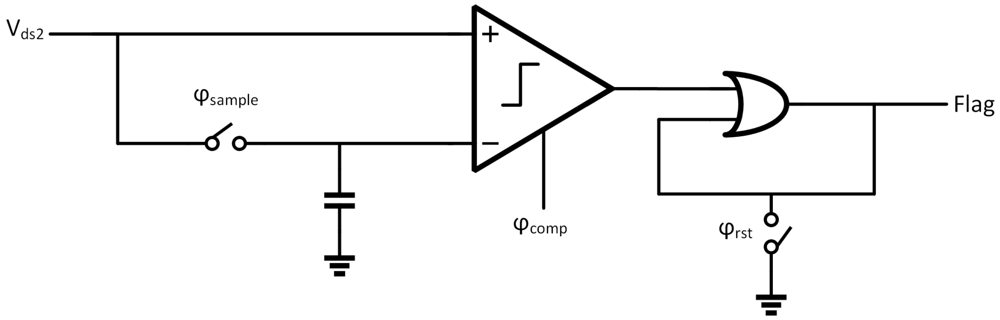


Figure 2.19: The proposed voltage compliance monitor.

determine whether or not a flag should be raised.

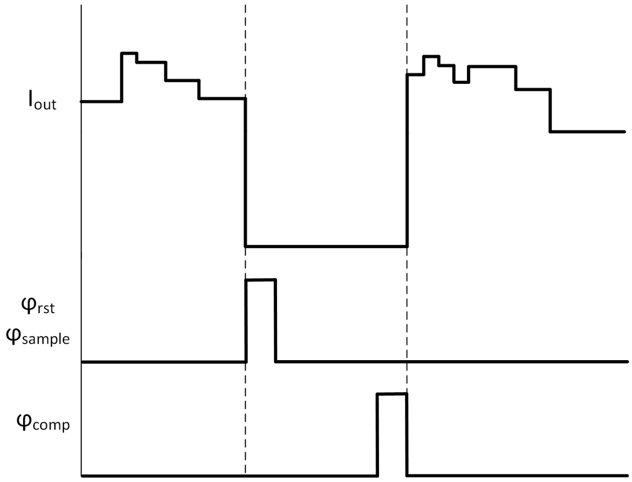


Figure 2.20: Clock timing diagram for the voltage compliance monitor.

2.3. System-level overview

In summary, Fig. 2.21 shows the system-level overview of the neurostimulator, with the blocks discussed in this Chapter. The system shown has a demultiplexer that can control two electrodes, as this will also be the set-up that will be tested. Note however, that it can easily be expanded to control more than two electrodes, as will become necessary if 10^4 electrodes need to be controlled.

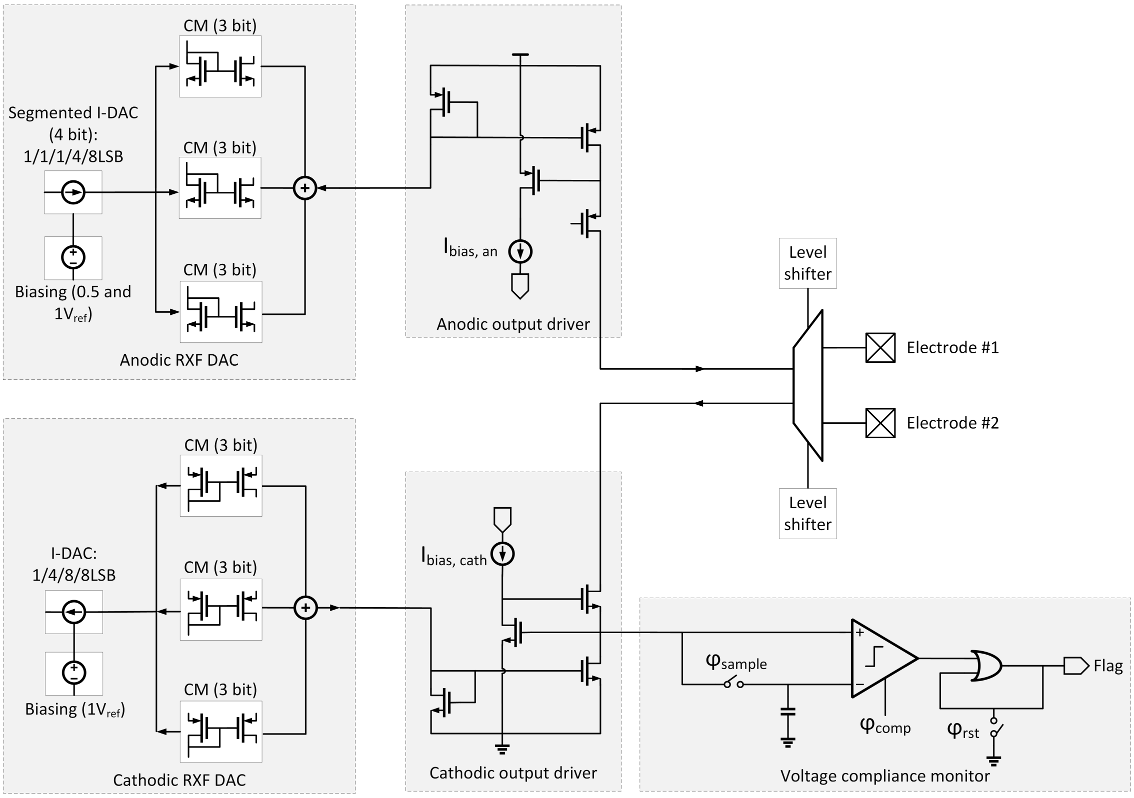


Figure 2.21: System-level overview of the neurostimulator.

3

Design

In this Chapter, the transistor-level implementation of the system detailed in the previous Chapters will be elaborated.

3.1. DAC

As mentioned in the previous Chapter, the chosen topology for the DAC is the current-steering I-DAC. Given that the main concern is the area, the design methodology is to use minimum size transistors, unless they cannot meet performance requirements (such as noise). Furthermore, as presented in more detail in App. B, mismatch must be controlled and carefully taken into account: too little mismatch, and the output may not be precise enough. Too much mismatch, and the increased output precision may end up distributed over too large a range, reducing the RoO at the higher end of the output range. Achieving the optimal amount of mismatch would normally require transistors that are circa 10x the minimum size, which is evidently not area efficient if it is not already necessary for other performance specifications. Instead, minimum size transistors will be used when possible, giving higher mismatch than what would be optimal. To still achieve the desired resolution, the full output range of the DAC is simply increased, such that the required output currents of 0–6 µA now falls in the lower part of the DAC's output range, and therefore is not affected by the increased resolution deviation.

The topologies used for the DAC in both the anodic phase and cathodic phase are shown in Fig. 3.1 and 3.2.

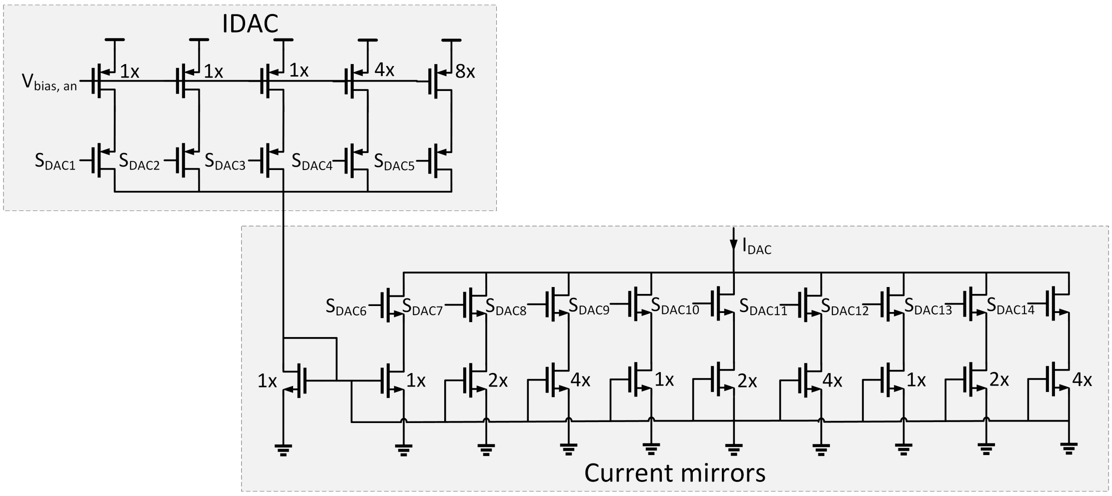


Figure 3.1: Transistor-level schematic for the DAC of the anodic phase.

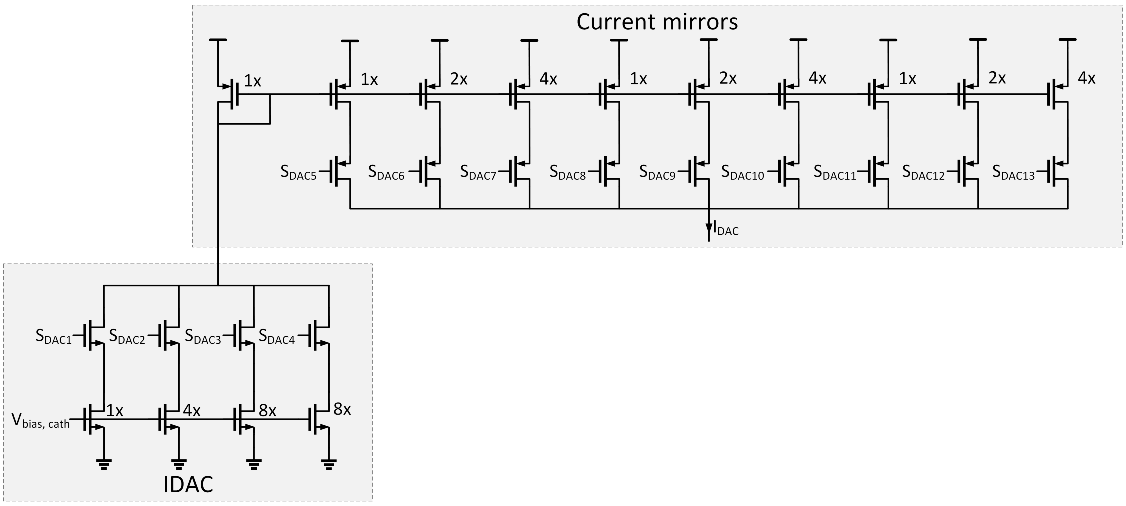


Figure 3.2: Transistor-level schematic for the DAC of the cathodic phase.

To keep the transistor count to a minimum, very simple current sources and mirrors are used. The drawback of such a simple topology is (among others) its reduced output impedance, which results in code-dependent output voltages, and therefore heavily non-linear output currents. Fortunately, this drawback is not a problem as for each configuration of DAC and current mirror the output current will remain constant over time, as the same voltage will be present each time. Additionally, in Section 2.2.1 the decision was made to not make use of current dumps, as it was assumed that the current cells were fast enough (i.e. could turn on within the signal bandwidth of 1 MHz). To verify this assumption, the parasitic capacitances of the transistors are simulated. Since the gate and source of the current cell are always well-defined (by either a voltage bias or power rail), the total capacitance on the drain of the current cell plus the total capacitance caused by the source of the switching transistor are what needs to be charged (and discharged) each time the switch toggles. Given that the charging of any capacitance introduces a transient current, this could compromise the achieved resolution. In order to establish the amount of Coulomb that is being subtracted from the load in order to charge the capacitance, the following formula can be derived:

$$\Delta Q = C_{total} V_{DD} \tag{3.1}$$

with C_{total} being the total parasitic capacitance on the node, and V_{DD} being the worst-case voltage (i.e. $1.1\,\mathrm{V}$). This ΔQ is then compared to the total injected charge, and the maximum usable frequency is at the point where

$$\Delta Q > 0.5 I_{LSB} \cdot \frac{1}{f_{max}} \tag{3.2}$$

The capacitance (of a minimum size switch, and varying size current cell) and maximum usable frequency of NMOS and PMOS devices is plotted in Fig. 3.3a and 3.3b. What can be seen is that for the highest stimulation frequency of the neurostimulator (i.e. 1 MHz), the used technology is more than sufficient in terms of speed to justify not using a current dump.

When looking at the schematic of the DACs, it can be seen that a voltage is needed that biases the current cells. Unlike the rest of the system, where mismatch and uncertainty are desired, for the biasing, a reasonably well-defined value is preferred. This is due to the fact that any variation in the LSB current caused by biasing will be amplified heavily before reaching the load, and can thus significantly shift the possible output range higher or lower. For the anodic DAC, where two biasing voltages are used (i.e. 0.5LSB and 1LSB) this is of less concern, as the extra biasing voltage will introduce overlapping redundancy. However, for the cathodic DAC, where only a single bias voltage is used, this could present issues. For these reasons, it is chosen to not rely on direct voltage biasing and instead feed a bias current into a relative large current mirror to generate the bias voltage. In this way, corner-to-corner differences can be ignored and only local mismatch plays a role, which helps improve the biasing accuracy.

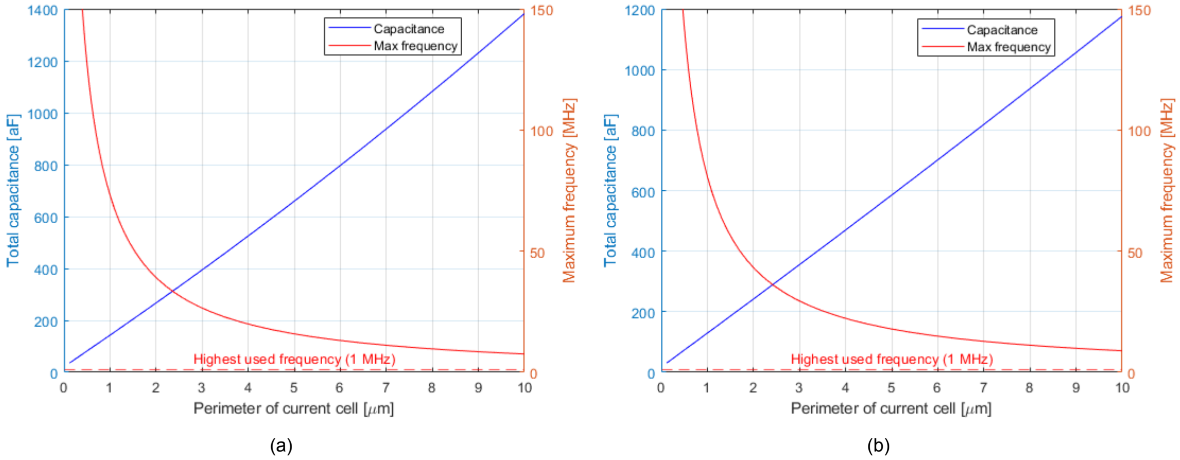


Figure 3.3: Total parasitic capacitance and maximum usable frequency against current cell perimeter of (a) NMOS devices, and (b) PMOS devices.

The transistor-level schematic of the bias generator can be seen in Fig. 3.4. It consists out of a current mirror chain, with an input bias current being supplied. A cascode current mirror is necessary for the cathodic bias generation, as otherwise the change in $V_{bias,an}$ caused by pulling $S_{bias,an}$ high or low will have an influence on the current generated by the bottom mirror via channel-length modulation. Due to this, the current at the right side of the circuit will change depending on whether or not $S_{bias,an}$ is high or low. This is unwanted considering that this would cause the ratios of the anodic bias to stray even further from the wanted 0.5LSB and 1LSB.

To prevent the switch transistor (i.e. the transistor connected to $S_{bias,an}$) from having a large effect on the resulting $V_{bias,an}$, the mirror transistors are chosen to be relatively large, and the switch is chosen as minimum size. This is desired regardless, as mismatch should be kept lower than in the rest of the system, as mentioned before. Additionally, using larger transistors helps with the noise, as will be discussed later (for the entire system altogether).

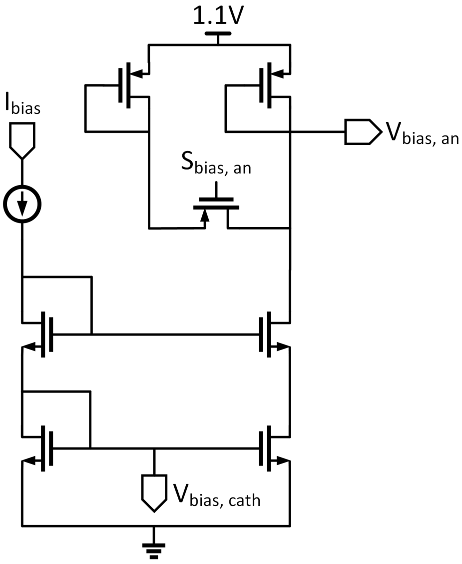
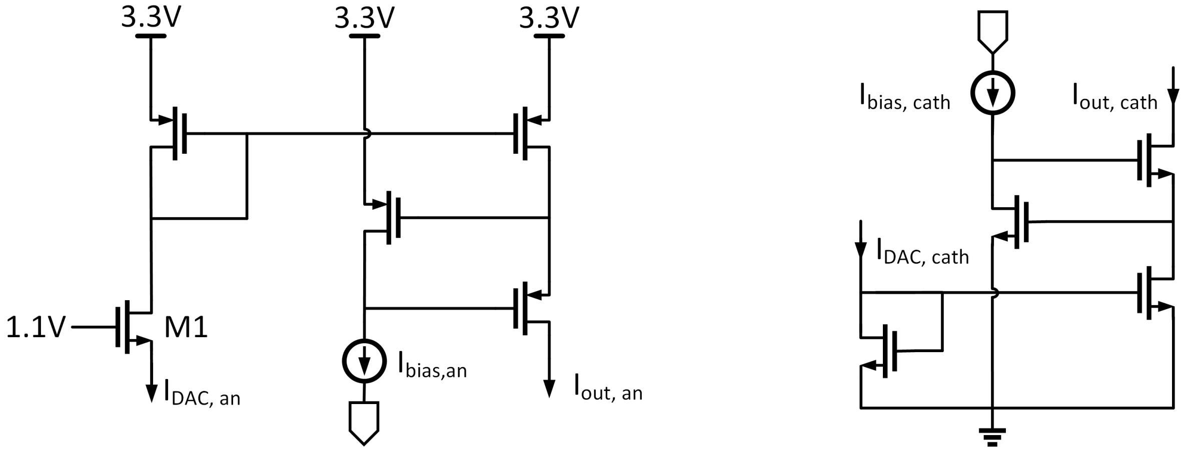


Figure 3.4: Transistor-level schematic for the bias generator.

3.2. Output driver

The transistor-level schematic for both the anodic stimulator and cathodic stimulator can be seen in Fig. 3.5a and 3.5b, and are the regulated cascoded current mirror without internal mirroring as discussed in Section 2.2.2.



(a) (b)

Figure 3.5: Transistor-level schematic of the output drivers for the (a) anodic part, and (b) cathodic part.

The output driver of the anodic stimulator has transistor M1 to ensure that the circuitry before it (i.e. the anodic DAC) do not see more than 1.1 V, as they comprise of standard voltage transistors. Unfortunately, M1 drops its source to $1.1V - V_{TH}$, which results in the DAC having less voltage headroom to operate with. However, this is not a problem for two reasons: (1) V_{TH} is a parameter that does not change significantly in time for a constant temperature. As a result the voltage at the source of M1 is relatively constant, and the generated current is therefore also relatively constant, and can thus be calibrated for. Another reason (2) is that a native transistor is used, which have a V_{TH} of nearly 0 V.

As for point (2), the use of a native transistor can be considered paradoxical given the strict area requirements. They not only have 25% larger minimum widths and 118% larger minimum lengths (compared to standard HV transistors), they also have to be isolated in layout from other non-native transistors, incurring an even larger area penalty. However, the nominal V_{TH} of a standard HV transistor is simply too large (circa 600 mV), and would give rise to serious output current fluctuation when considering supply ripple. This can be seen in Fig. 3.6. Notice the slope difference in output current for the native transistor (solid red line) vs the standard threshold transistor (dashed red line), indicating a much higher supply sensitivity for the standard HV transistor.

3.2.1. Biasing current

To determine the value needed for the bias current in each output driver, the maximum V_{dsat} needs to be found for the output mirror transistor. Considering that V_{dsat} can be approximated to be

$$V_{dsat} = V_{GS} - V_{TH} \tag{3.3}$$

the highest V_{dsat} can be found when having the highest V_{GS} and thus the highest output current. For this reason, simulations are performed for when the output current is at its maximum, i.e. 6 µA. These simulations determine the V_{dsat} for all the corners, and the results can be seen in Tab. 3.1. The maximum V_{dsat} is found in the SS corner, and is 146.9 mV and 169.5 mV for the anodic and cathodic driver respectively.

In order to find the bias current needed, it can be swept against the resulting V_{DS} of the output mirror transistor. It should be noted however that there is a limit to how low the bias current can be: if the biasing current becomes too small, slewing might occur as a consequence due to insufficiently fast charging of the nodal capacitances (at the gate of the output cascode transistor). To this end, ideally low threshold devices are used as bias current transistors. This is done for the cathodic driver only, as

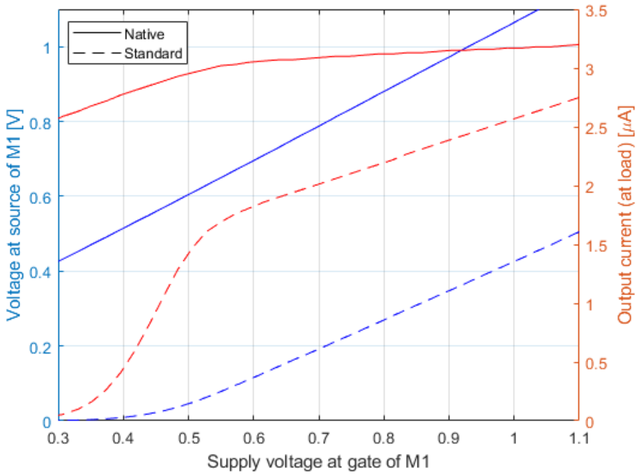
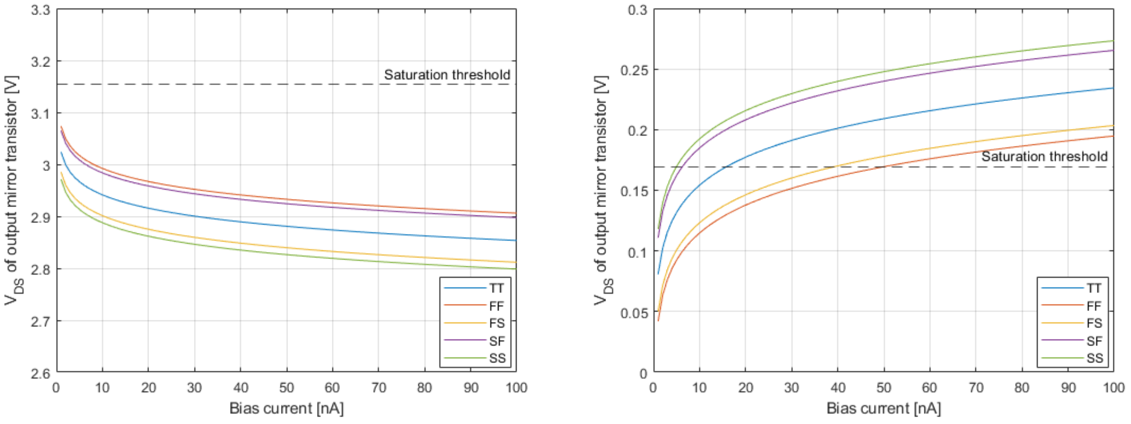


Figure 3.6: Effect of using HV native vs standard threshold transistors on the voltage headroom and output current over increasing supply voltage. The blue lines indicate the voltage at the source of M1 (so the voltage available to the RXF circuitry), and the red lines indicate the output current (after the output driver). The solid lines are for the native HV transistor, and the dashed lines are for the standard HV transistor.

Table 3.1: Drain saturation voltage over different corners for the anodic and cathodic driver.

Corner	$V_{dsat,an}$	$V_{dsat,cath}$
TT	141.3 mV	165.7 mV
FF	136.4 mV	162.8 mV
FS	145.0 mV	166.7 mV
SF	136.9 mV	165.8 mV
SS	146.9 mV	169.5 mV

there exist no HV low threshold voltage devices in the used technology. Fig. 3.7a and 3.7b show that the lack of a low threshold voltage device in the anodic driver results in extremely low bias currents (far below 1 nA).



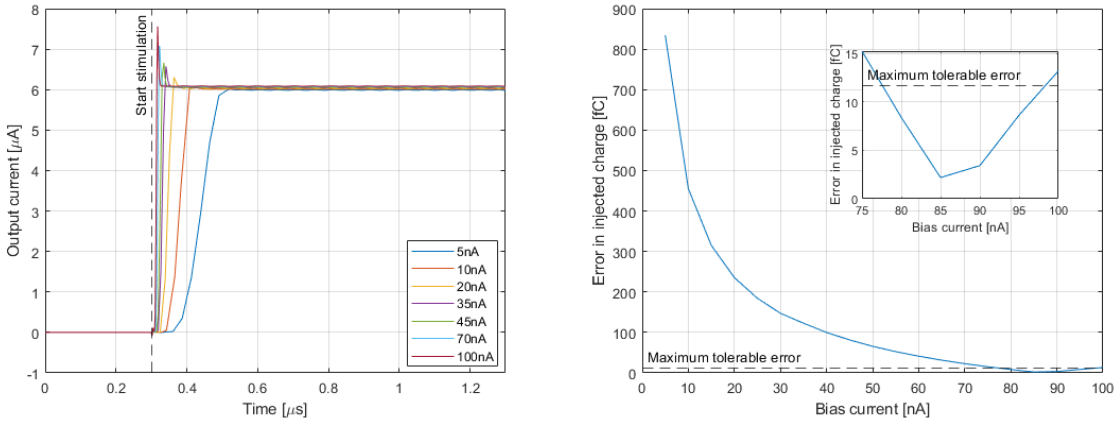
(a)

(b)

Figure 3.7: Bias current against the resulting V_{DS} of the output mirror transistor for (a) the anodic driver, and (b) the cathodic driver.

Given that the minimum stimulation time step is 1 μs for the anodic phase and 10 μs for the cathodic phase, the maximum allowable slewing can be calculated via the total injected charge. As long as the error in total injected charge remains below $0.5 I_{LSB} \cdot t_{step,min}$ (i.e. 11.7 fC for the anodic phase, and

117 fC for the cathodic phase), the error can be considered negligible. The slewing for different bias currents in the anodic driver can be seen in Fig. 3.8a, and the error in injected charge can be seen in Fig. 3.8b.



(a) (b)

Figure 3.8: Different bias currents for the anodic driver against (a) the output current over time, and (b) the error in total injected charge.

From this, the required bias current in the anodic driver can be seen to be at least 80 nA. The required bias current in the cathodic driver (i.e. 52 nA), obtained from Fig. 3.7b, is less than this. The error in the cathodic driver is found to be 82 fC, and is thus negligible due to the 10x larger error margin mentioned before.

3.2.2. Output impedance

The required output impedance, as calculated in Section 2.2.2, is at least 53.5 MΩ. The three transistors in the output driver that affect the output impedance are: (1) the output mirror transistor (M2), (2) the output cascode transistor (M3), and (3) the bias current transistor (M4). The output impedance then follows

$$Z_{out} = r_{o2}A_{M3}(A_{M4} - 1) \tag{3.4}$$

with $A_{Mn} = g_{mn}r_{on}$. Transistors in strong inversion (SI) usually have higher r_o, whereas transistors in weak inversion (WI) usually have higher g_m. For this reason, the general design strategy is to have M2 biased in SI, and M3/M4 biased in WI. An additional advantage of biasing M3/M4 in WI is that it decreases their V_{dsat}. The output impedance of the designed output drivers are seen in Fig. 3.9a and 3.9b for the anodic and cathodic phase, respectively (N=200 for both).

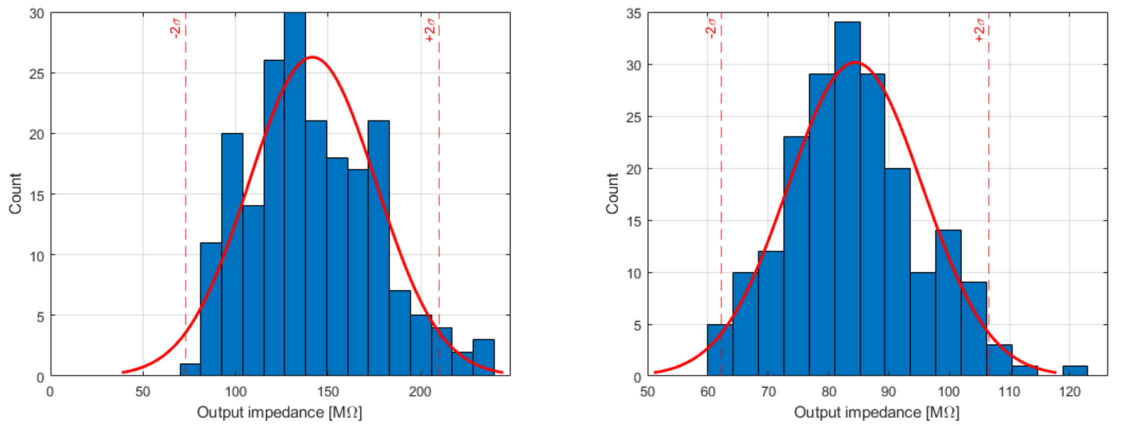
3.2.3. Voltage compliance

The voltage compliance of each driver is determined by the output voltage at which the output impedance drops below the minimum 53.3 MΩ. These graphs can be seen in Fig. 3.10a and 3.10b for the anodic and cathodic phase, respectively. The voltage compliance is therefore: 475 mV for the anodic driver, and 287 mV for the cathodic driver. The increased voltage required in the anodic driver can be explained by the bias current. Since there is no low threshold voltage device available, the V_{DS} of the output mirror transistor becomes higher than is necessary, resulting in loss of voltage headroom.

3.3. System noise

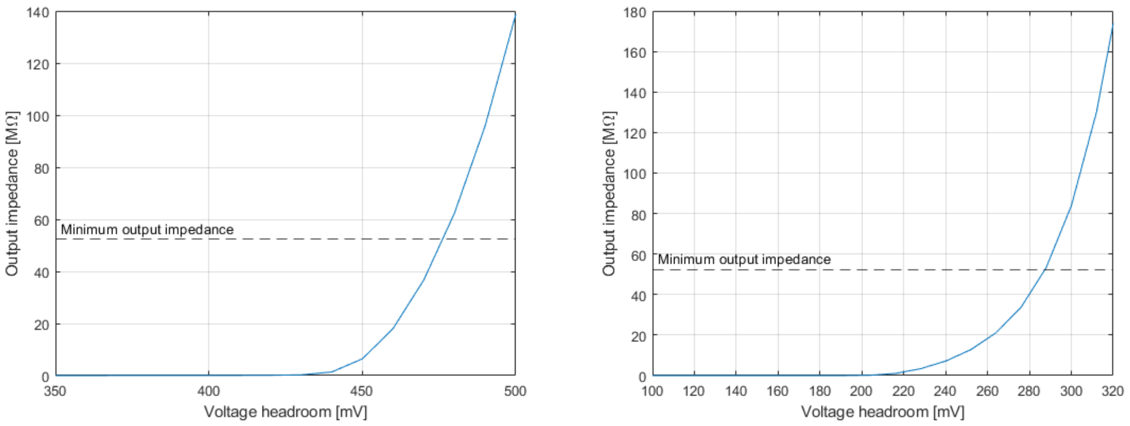
With calibration, static errors are of no concern. Dynamic errors such as noise, however, still need to be taken into account. There are two sources of noise that are dominant: (1) thermal noise, and (2) flicker noise. The spectral density of thermal noise is described in Eq. 3.5. Flicker noise cannot be accurately described, but a common empirical model for its spectral density is shown in Eq. 3.6.

$$|I_{th,noise}|^2 = 4kT\gamma g_m \tag{3.5}$$



(a) (b)

Figure 3.9: Histogram of the output impedance for (a) the anodic driver, and (b) the cathodic driver.



(a) (b)

Figure 3.10: Output impedance over reducing voltage headroom for (a) the anodic driver, and (b) the cathodic driver.

$$|I_{flicker,noise}|^2 = \frac{K_{1/f} g_m^2}{C_{ox} W L} \cdot \frac{1}{f} \tag{3.6}$$

For neurostimulation purposes, high frequency noise can be disregarded, as it does not have an effect on the functionality of stimulation (see Appendix E for the full explanation). The total integrated noise presented in the remainder of this Section will therefore use 1 MHz as the maximum frequency for the anodic stimulator, and 100 kHz as the maximum frequency for the cathodic stimulator.

In order to improve power efficiency, the output driver should not have a 1:1 gain, as this would limit the current efficiency to 50%. The allowable gain of the output driver depends on the acceptable output current noise: the larger the gain of the output driver, the lower the input bias current has to be to achieve the same output current. However, the noise does not scale equally with the input current. E.g., the noise is not halved when the input current is halved. In order to achieve the required 8 bit resolution, the SNR of the bias generator must remain above

$$SNR_{bias,gen} = 20log_{10}(2^8) = 48.16dB \tag{3.7}$$

Fig. 3.11 shows the output noise current and SNR over different input bias currents. Conclusively, the input current must remain above 2.8 nA to achieve sufficient SNR. Since this calculation does not take into account the noise generated after the bias generation circuit, this number is only a theoretical maximum, and should in practice be higher.

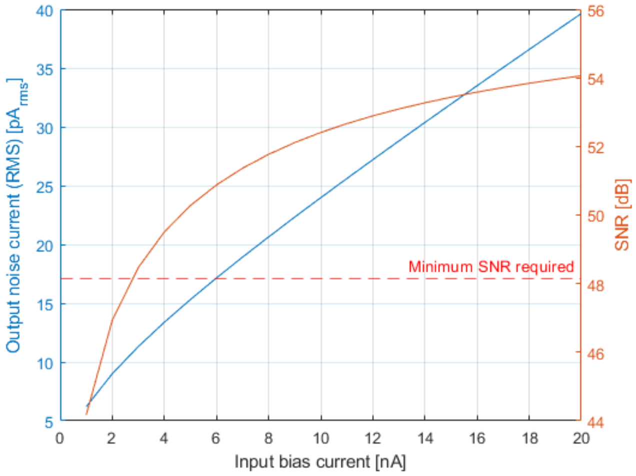
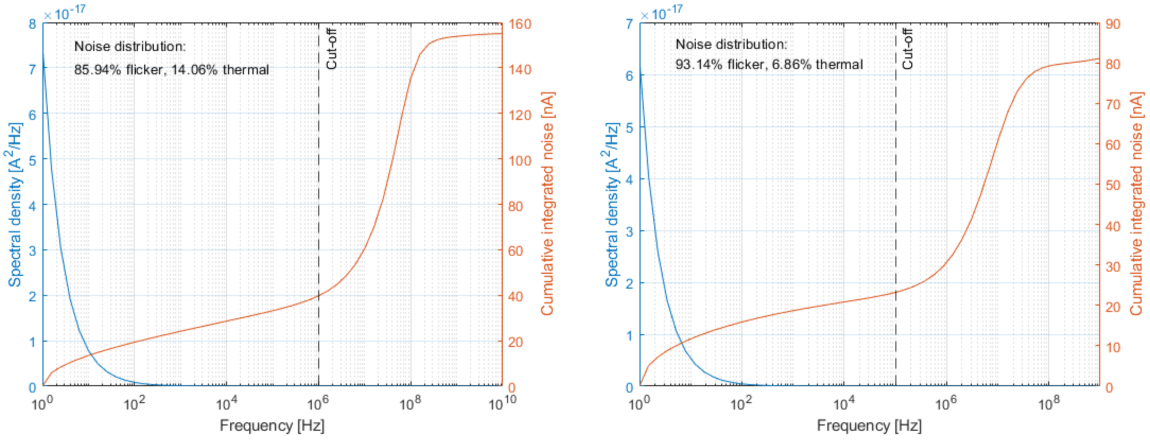


Figure 3.11: Output noise current (RMS) generated by the bias generator and SNR over different input bias currents.

Given that the input bias current was 80 nA (Sec. 2.2.1) with unity gain of the output driver, a gain of 8 was chosen. This reduces the input bias current to 10 nA, which leaves some margin for noise generated by the rest of the system.

Using minimum size transistors, the noise spectrum and integrated RMS noise is shown in Fig. 3.12a and 3.12b for the anodic stimulator and cathodic stimulator, respectively. The relative distribution of the noise origin up to the cut-off frequency is also shown.



(a)

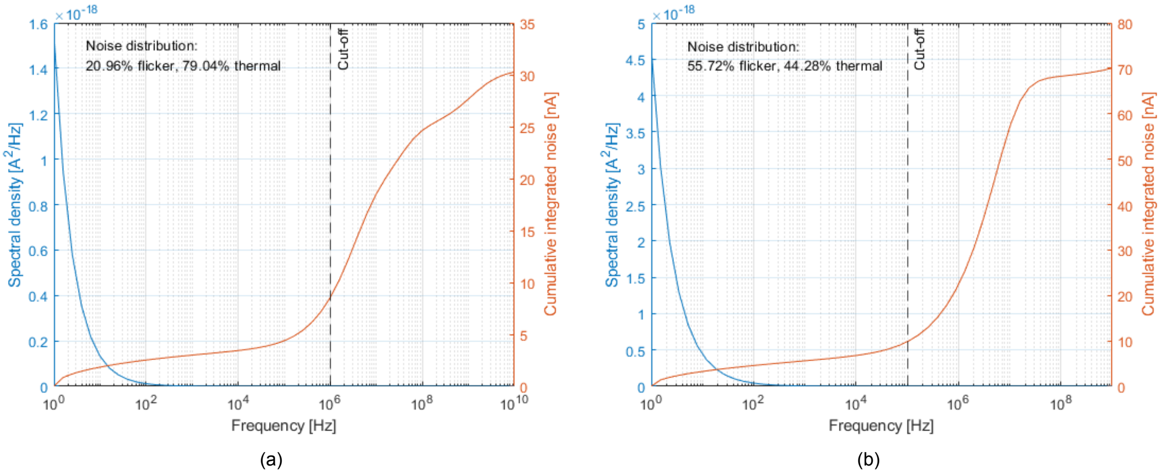
(b)

Figure 3.12: Total current noise spectral density and integrated current noise (RMS) for (a) the anodic stimulator, and (b) the cathodic stimulator, both with minimum size transistors.

The dominant noise source for both stimulators is flicker noise. In order to reduce the flicker noise present, the area of the transistors have to be increased, as flicker noise is inversely proportional to the area (Eq. 3.6). The final noise spectral densities and integrated noise are shown in Fig. 3.13a and 3.13b for the anodic stimulator and cathodic stimulator, respectively. The resulting RMS integrated output noise current is 8.6 nA for the anodic stimulator, and 9.8 nA for the cathodic stimulator. Both of which are below $0.5LSB = 11.7nA$, and thus sufficient.

3.4. Voltage compliance monitor

For the voltage compliance monitor (Fig. 3.14), the capacitor needs to be sized, and the comparator needs to be designed.



(a) (b)

Figure 3.13: Total current noise spectral density and integrated current noise (RMS) for (a) the final anodic stimulator, and (b) the final cathodic stimulator.

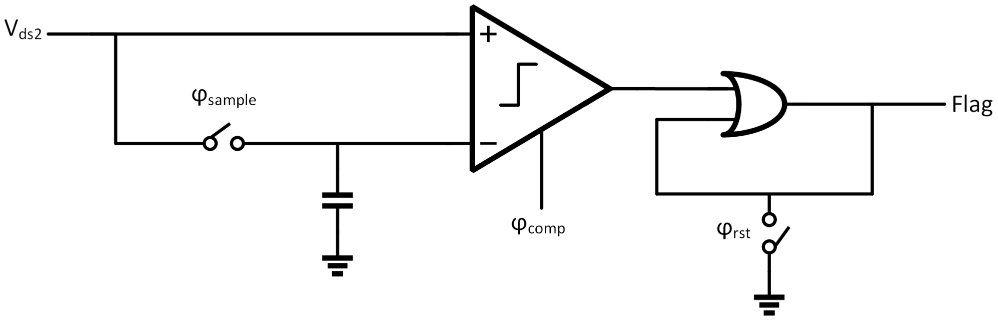


Figure 3.14: The proposed voltage compliance monitor.

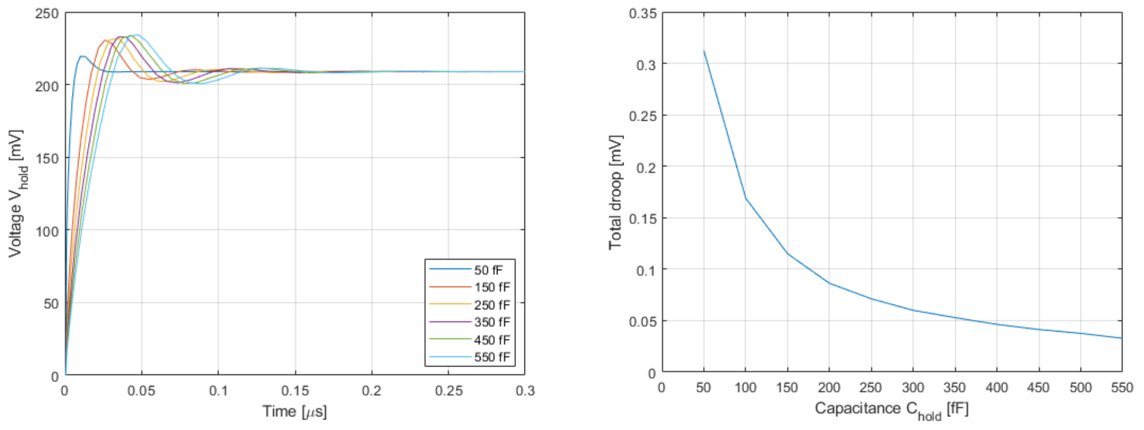
Ideally, for the capacitor, the acquisition time is small, and the total droop is minimal. The shortest time in which the capacitor needs to be charged is 1 µs, and the longest period the voltage has to be held before the comparator is clocked is 48 µs (longest cathodic pulse is 50 µs, and both the sampling and comparator clocking period are 1 µs each). There is an inherent trade-off between acquisition time and droop rate, so a balance needs to be found here. Furthermore, this capacitor will be implemented as a metal-oxide-metal capacitor (MOM cap), given that this saves area by allocating layers metal3 to metal5 (M3-M5) for the capacitor and having M1-M2 for the transistors and routing. As a consequence, this also means the area used to implement the capacitor is of no concern, as long as it remains within the boundaries of the used area underneath.

Regardless, an upper limit to the capacitance can be calculated. Since the current necessary to charge the capacitor is added to the total output current, more charge is injected into the load as a result. From previous calculations 117 fC is determined to be the maximum allowable charge error. The maximum value for the hold capacitor is then

$$C_{hold,max} = \frac{Q_{error}}{V_{hold}} = \frac{117fC}{220mV} = 531.82fF \tag{3.8}$$

The sampling transient for different hold capacitor values can be seen in Fig. 3.15a. The total droop over the period of 48 µs is plotted against the capacitances in Fig. 3.15b.

Since for all capacitance values the acquisition time is still well within the 1 µs range, and the total droop is still well below 5 mV, the largest possible capacitance can be taken. This has the additional benefit of limiting the effects of high frequency noise (i.e. reduces kT/C noise), allowing for a more accurate sampled voltage. As mentioned before, it should ideally be possible to implement this capacitor without incurring any extra area penalty, and the maximum capacitance possible (determined after layout) is 423.97 fF.



(a) (b)

Figure 3.15: The (a) sampling transient over time, and (b) total droop for different capacitance values for the hold capacitor.

3.4.1. Comparator

The topology used will be the modified StrongARM latch comparator [46] (Fig. 3.16), primarily for the reason that it is a simple, robust design with offset from one differential pair only [47]. Due to this, it becomes easier to tune the offset by introducing trimming transistors in parallel with the input differential pair, which is necessary as will be explained later. To ensure fast latching at $V_{CM} = 220mV$, a PMOS version is used.

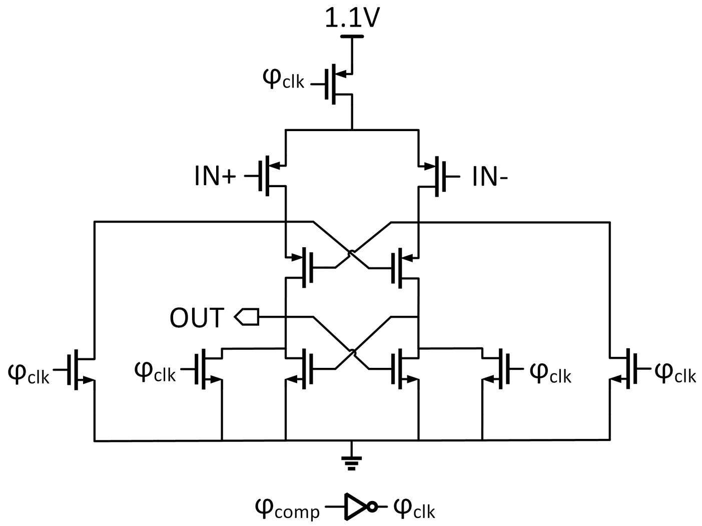


Figure 3.16: Transistor-level schematic of the modified StrongARM latch topology. The clock is inverted, as it was first designed for an NMOS version.

The histogram of the offset (N=200) of this comparator can be found in Fig. 3.17. Given that the required offset is at least -5 mV, trimming is necessary.

Trimming can, as mentioned earlier, easily be done by adding trimming transistors in parallel with the input differential pair (see Fig. 3.18).

The offset after trimming can be read in Tab. 3.2. Only trimming of the right side is shown, as the left side has the same offset but with reversed polarity. By using eight trimming transistors (for on each side of the input), the offset can consistently be trimmed to at least -5 mV.

An example of the entire voltage compliance monitor functioning can be seen in Fig. 3.19. The monitor

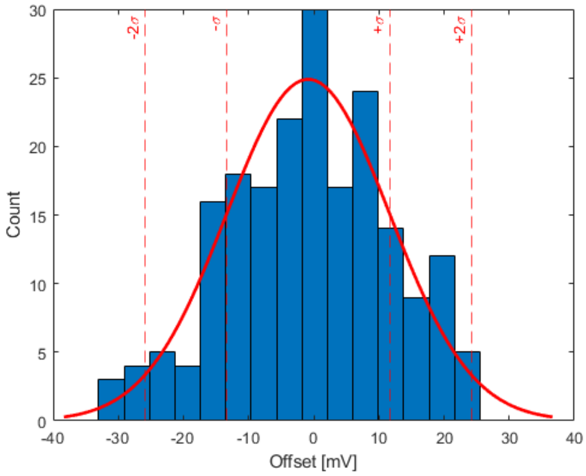


Figure 3.17: Histogram showing the offset distribution of the modified StrongARM latch comparator.

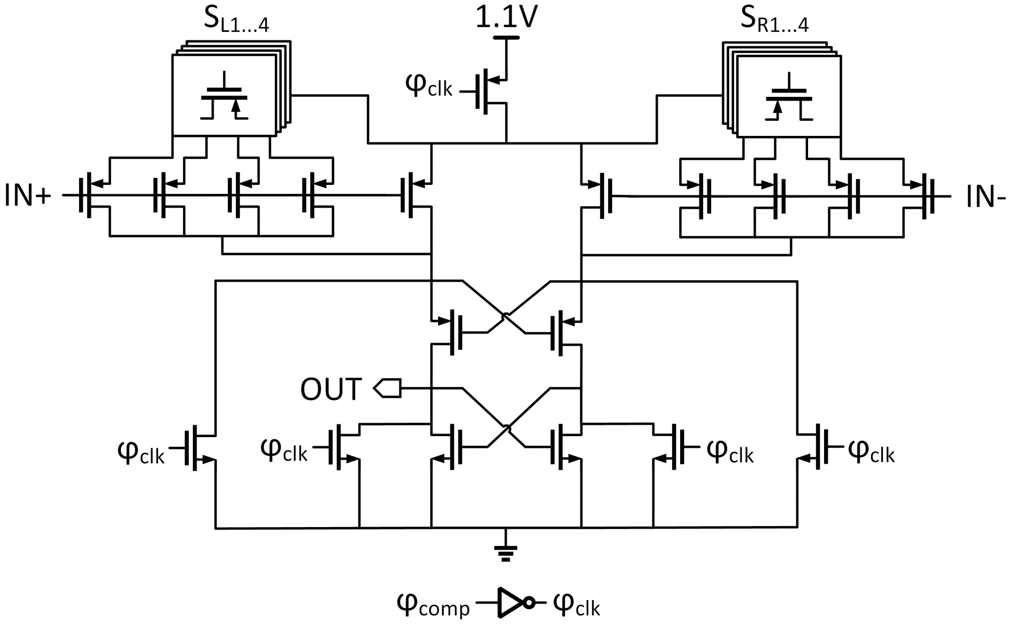


Figure 3.18: Transistor-level schematic of the modified StrongARM latch topology, with trimming transistors.

can be seen raising a flag in the second cycle (each cathodic cycle taking 10 µs in this case), as output current decreases due to decreasing voltage headroom for the cathodic driver.

Table 3.2: Offset after trimming of the right input for the comparator.

Trimming configuration (left on - right on)	Offset [mV]
3-4	-9.2 mV
2-3	-10.5 mV
1-2	-12.3 mV
0-1	-15.2 mV
2-4	-20.0 mV
1-3	-22.7 mV
0-2	-27.7 mV
1-4	-33.0 mV
0-3	-38.4 mV
0-4	-48.9 mV

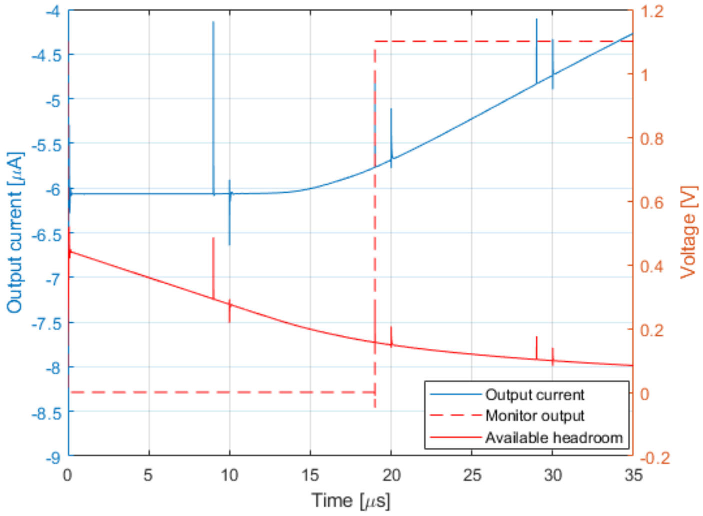


Figure 3.19: Example waveforms showing the output current and the voltage headroom available, with the resulting voltage compliance monitor output.

3.5. Demultiplexer

The demultiplexer is in nominal operating conditions attached to an output of $V_{CM} = 1.65V$, and thus requires HV devices. Furthermore, the clock signal driving these switches need to be 0–3.3 V as a result. Due to the digital circuits using standard 1.1 V devices, a level-shifter is needed. A simple latch level-shifter is used, and can be seen in Fig. 3.20, using HV transistors. Since the anodic side of the demultiplexer is switched with PMOS and the cathodic side is switched with NMOS, the same level-shifter can be used for both but with different output points (i.e. OUT and inverse OUT). In this way, all digital high signals can be considered conducting, regardless of phase.

The output of the level-shifter can be seen in Fig. 3.21. As briefly mentioned before, the output (blue line) is connected to the cathodic NMOS switches, and the inversed output (red line) is connected to the anodic PMOS switches.

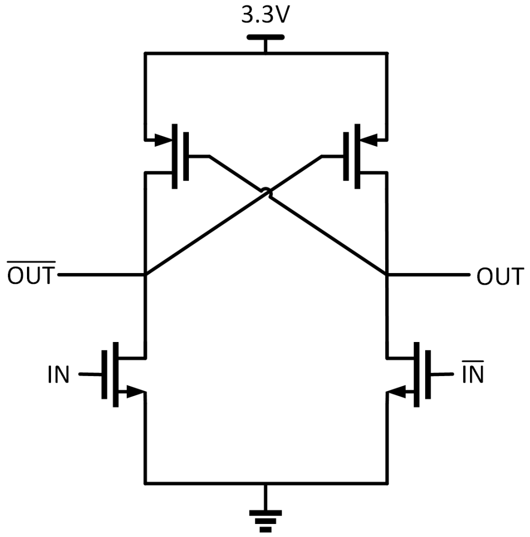


Figure 3.20: Transistor-level schematic of the latch level-shifter.

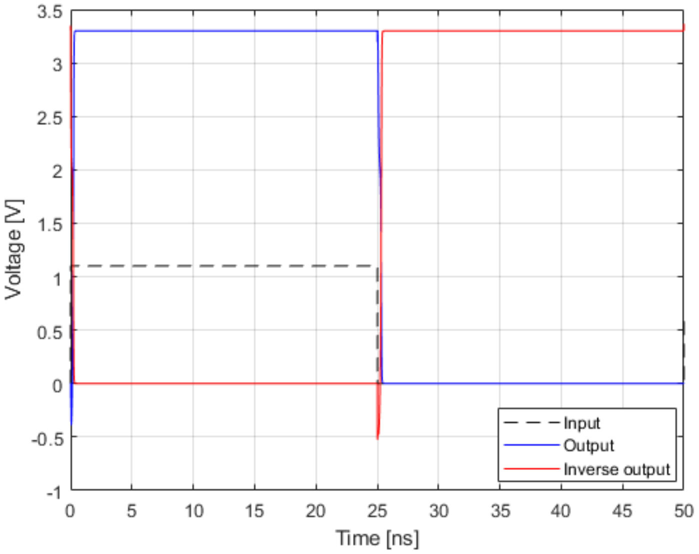


Figure 3.21: Example output of the level-shifter.

3.6. Layout

The layout of the full system is seen in Fig. 3.22. The same layout, but without the sampling capacitor visible, with annotations marking each sub-system can be seen in Fig. 3.23. The total area is 25.3 µm by 13.0 µm.

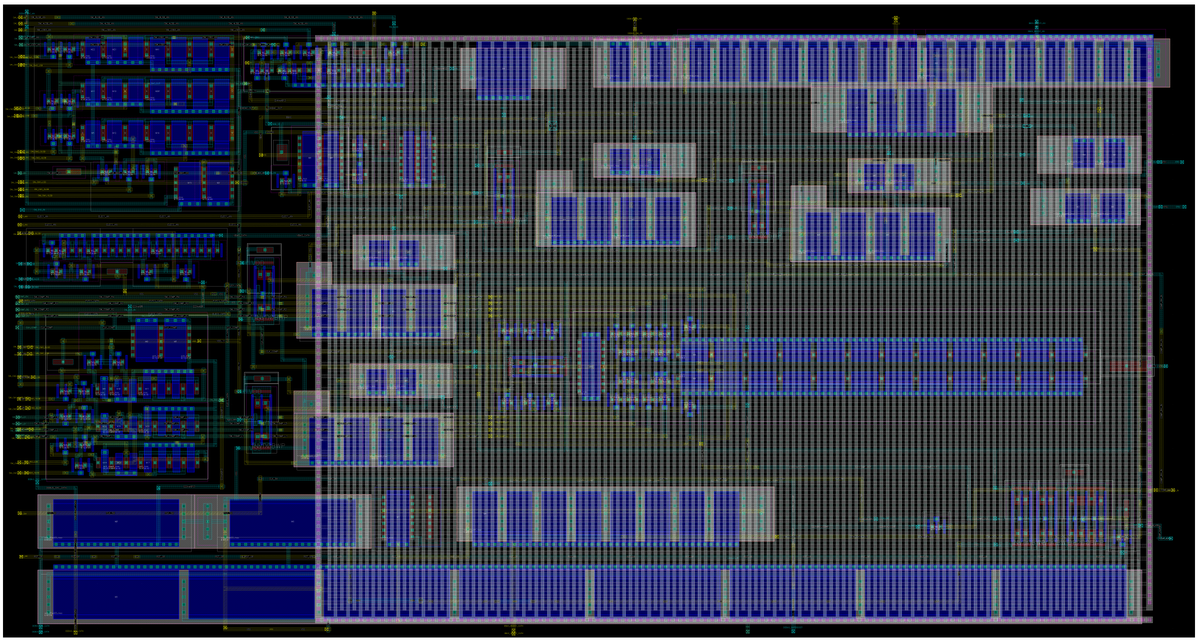


Figure 3.22: Layout of the full design.

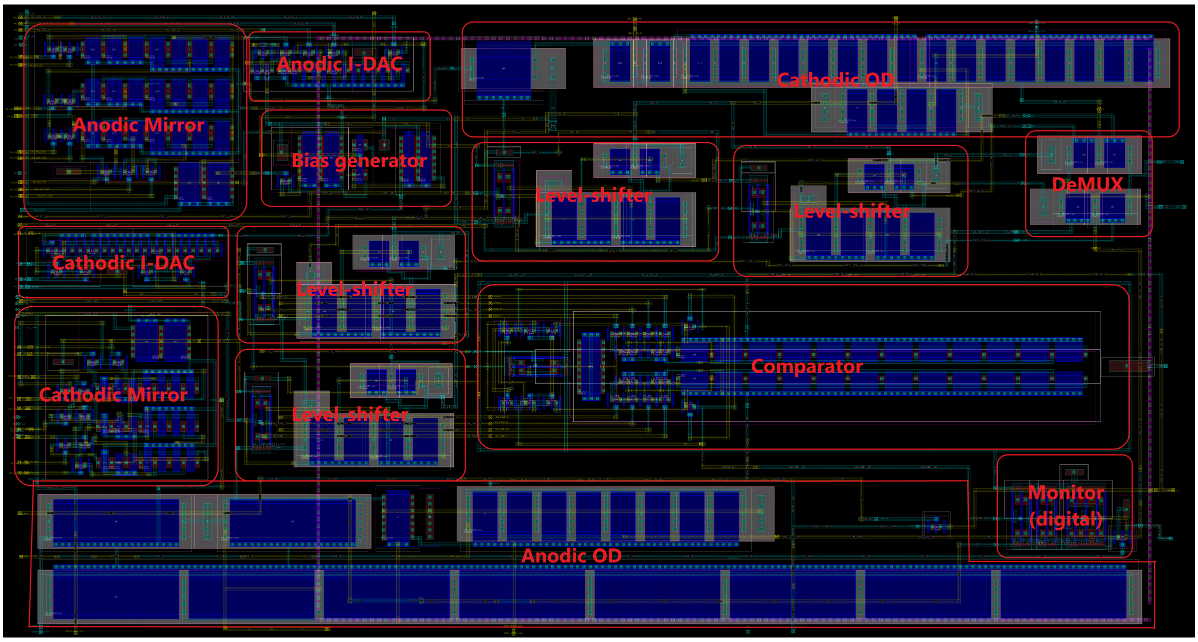


Figure 3.23: Layout of the full design with annotation, without sampling capacitor.

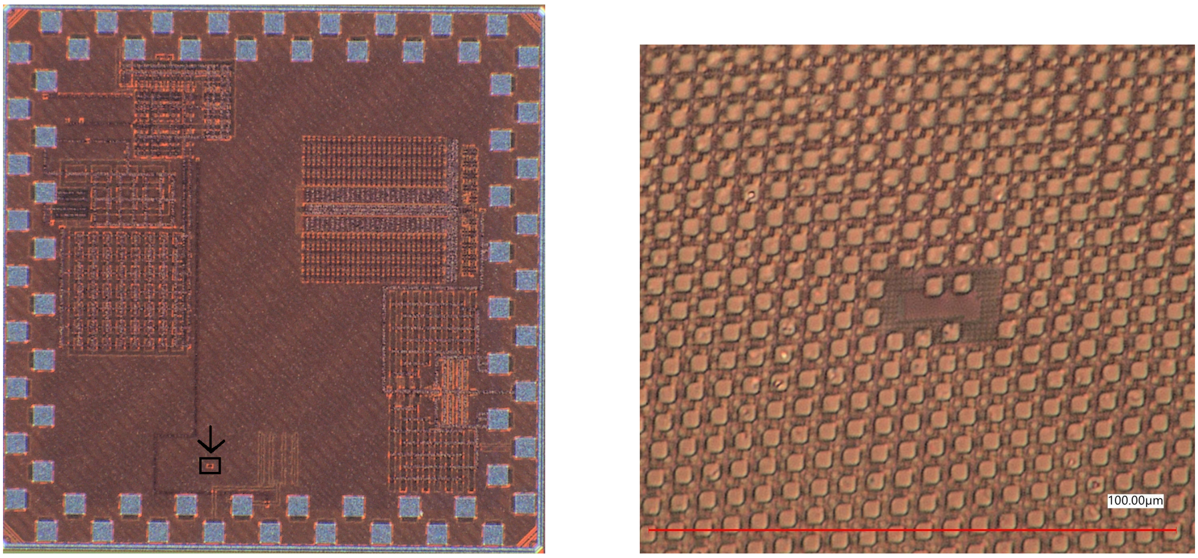
4

Measurements

In this Chapter the measurement set-up is detailed. Additionally, the measurement results are shown and, if necessary, compared to post-layout simulations for context.

4.1. Measurement set-up

A micrograph of the fabricated chip in CMOS 40-nm technology can be seen in Fig. 4.1a and 4.1b.



(a)

(b)

Figure 4.1: Micrograph of (a) the total fabricated chip with a black box around the neurostimulator, and (b) a zoomed in version to only the neurostimulator.

Since the fabricated chip also contains the designs of four other members of the research group, it was decided to make a daughterboard PCB to which the chip is wirebonded. This daughterboard can then be inserted (via pin headers) into the main PCB of each person. This allows each person to test the same chip, without having to directly wirebond or solder it. The two PCBs can be found in Appendix F.1.

On the main PCB, the most important blocks are the microcontroller (MCU) and the current generation circuitry.

4.1.1. Microcontroller

The used MCU is the PIC32MK1024GPK064. This MCU was primarily chosen for its SRAM size: it ideally needs to be able to store all the output current calibration data in its entirety. Since the ADC of the MCU has a resolution of 12-bits, each ADC reading requires at least two bytes of SRAM. However, since the MCU operates with a 32-bit CPU, it is easiest to store the ADC results in 32-bit (4 byte) format. To improve the resolution of the ADC, oversampling is applied. The ADC samples 256x per input code, and then takes the average of this and stores the result in SRAM. Given that there are 40960 total input codes, the entire output calibration requires 163.84 KB. This fits into the 256 KB SRAM of the chosen MCU.

In addition, the MCU is capable of UART serial communication, which is also what the digital logic on the chip (that controls the switches and frequency of stimulation) uses. For more information about the UART communication protocol, see Appendix G.

4.1.2. Current generation circuitry

The neurostimulator requires three bias currents: I_{stim} (input current of the bias generator), $I_{OD,an}$ (bias current for the anodic output driver), and $I_{OD,cath}$ (bias current for the cathodic output driver). These three currents are generated using the circuit of Fig. 4.2 [48].

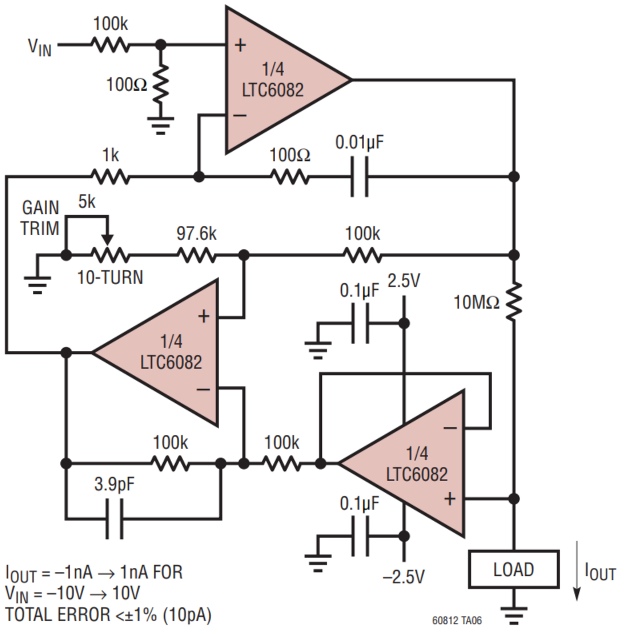


Figure 4.2: Current generation circuitry. The $10\,\mathrm{M\Omega}$ resistor is replaced with a $100\,\mathrm{k\Omega}$ resistor, to achieve a -100–100 nA output range. Taken from [48].

4.2. Results

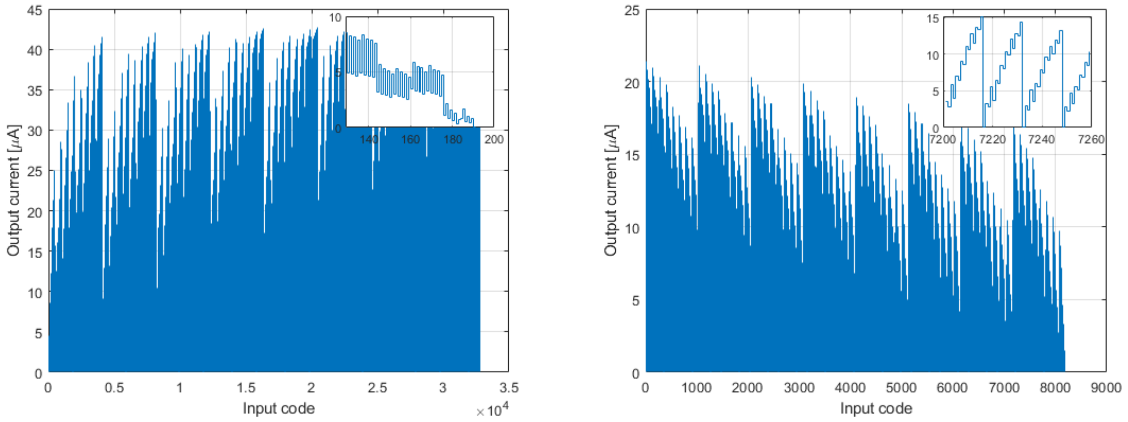
4.2.1. Output current calibration

The output current of each input code is calibrated using an automatic script written for the MCU. This was done for three fabricated chips. Fig. 4.3a and 4.3b show the obtained input-output waveform for the anodic and cathodic stimulator respectively, of chip 3.

The redundancy at each output current (for an 8-bit resolution) for all three chips can be seen in Fig. 4.4a and 4.4b for the anodic and cathodic stimulator, respectively. Finally, Tab. 4.1 shows the output range availability, i.e. how much of the full output range has a corresponding input code, for all three chips.

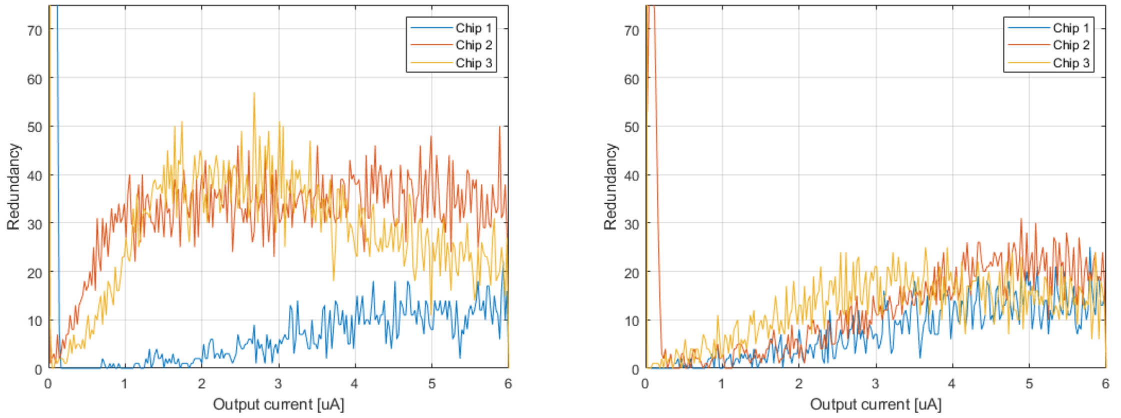
Table 4.1: Measured output range availability of three chips.

	Chip 1	**Chip 2**	**Chip 3**
Anodic	0.8320	1.000	0.9922
Cathodic	0.8633	0.9375	0.9727



(a) (b)

Figure 4.3: The measured output current for each input code for chip 3 of (a) the anodic stimulator, and (b) the cathodic stimulator. In the top right corner is a zoomed in version of the graphs.



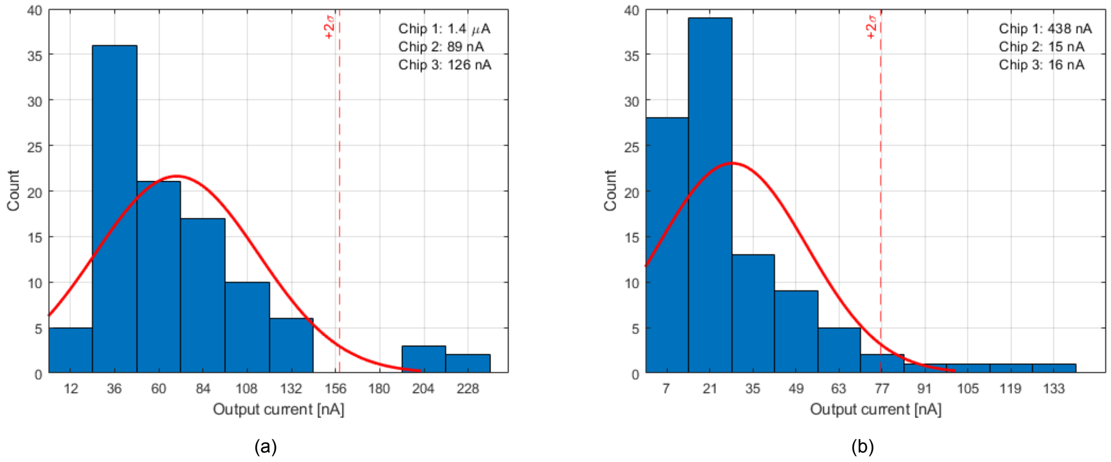
(a) (b)

Figure 4.4: The measured redundancy over the full output range of (a) the anodic stimulator, and (b) the cathodic stimulator.

Chip 1 seems to underperform significantly, based on the other two chips and (pre-layout) simulations[1] (see also Appendix C). To understand why this is the case, post-layout MC simulations of a single input code are performed, and then checked with the measured performance. Since the greatest variation in output current comes from the bias generation and the DAC, the input code that will be checked is when the gain nominally is unity. Fig. 4.5a and 4.5b show the post-layout MC simulations and the measured output currents of each chip, for the anodic and cathodic stimulator, respectively. This shows that the reason for the poor performance of chip 1 is the incredibly high initial input current, as the increase is seen in both the anodic and cathodic case. The exact reason for this increased input current for chip 1 only is unfortunately at the time of writing unknown.

---

[1] Post-layout MC simulations of all input-codes in one run were not performed, as this would require weeks of simulation.



(a) (b)

Figure 4.5: Histogram of post-layout simulation of a single unity gain configuration for (a) the anodic stimulator, and (b) the cathodic stimulator.

Finally, a calibration script was run that generates a LUT (to 8-bit resolution). The resulting input-output relation of the anodic and cathodic stimulator are plotted in Fig. 4.6. The corresponding INL and DNL are shown in Fig. 4.7a, 4.7b, 4.8a, and 4.8b.

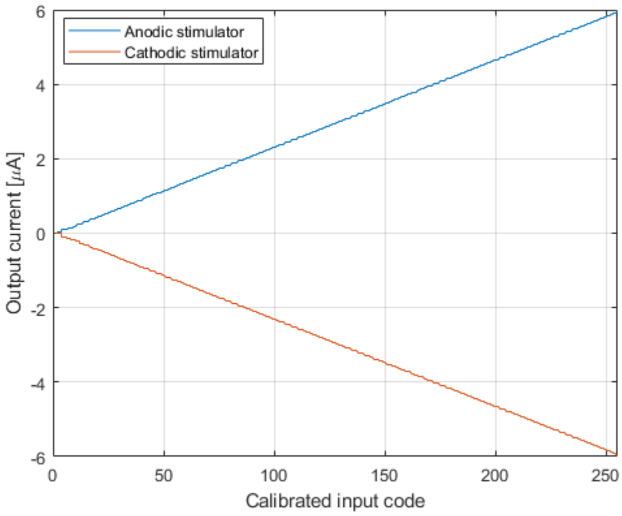
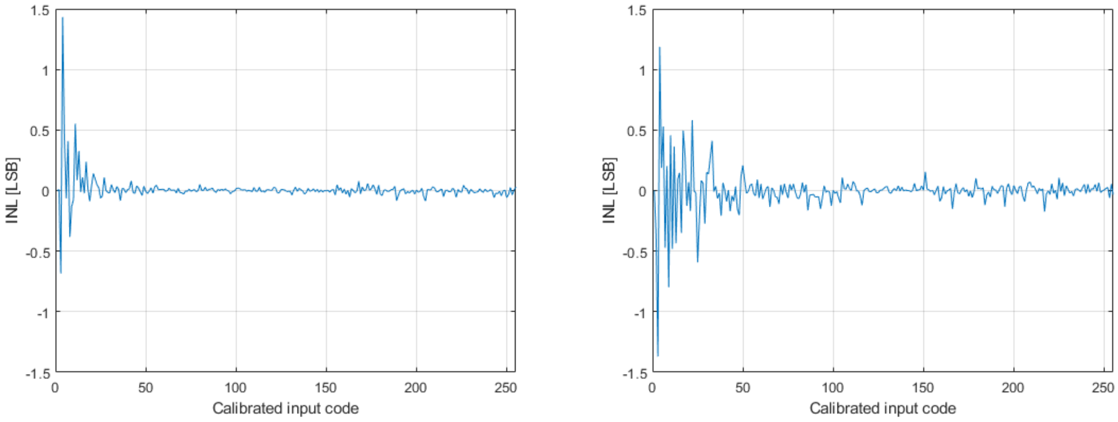
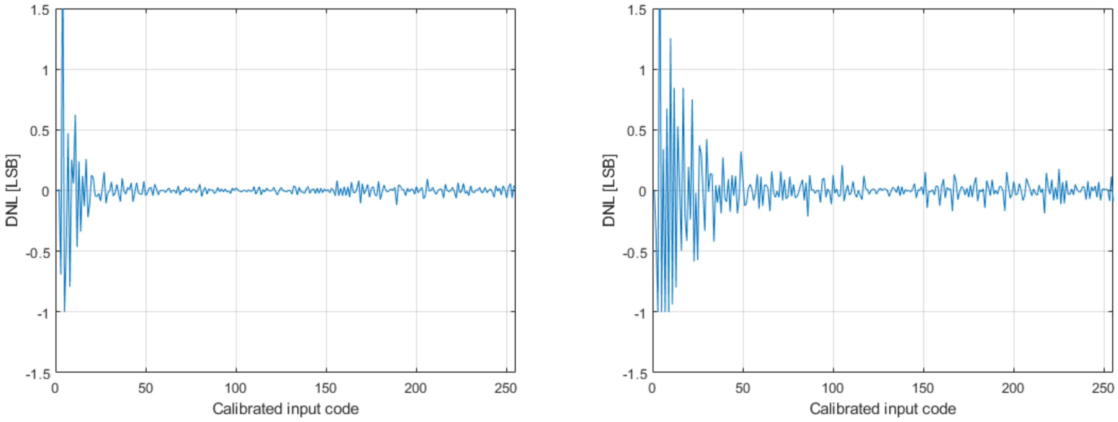


Figure 4.6: Output current of the configurations stored in the LUT for both the anodic and cathodic stimulator.



(a) (b)

Figure 4.7: INL of (a) the calibrated anodic stimulator, and (b) the calibrated cathodic stimulator.



(a) (b)

Figure 4.8: DNL of (a) the calibrated anodic stimulator, and (b) the calibrated cathodic stimulator.

4.2.2. Output driver

The V_{DS} of the bottom output mirror transistor of the cathodic driver can be measured. For the anodic driver, there unfortunately was no pad available to do so. Tab. 4.2 shows the measured values of V_{DS}. The measured voltages align with Fig. 3.7b.

Table 4.2: Measured V_{DS} of the bottom output mirror transistor in the cathodic driver for various bias currents.

Bias current	**Voltage**
10 nA	171 mV
30 nA	196 mV
50 nA	210 mV
60 nA	216 mV
70 nA	221 mV
80 nA	225 mV
90 nA	229 mV
100 nA	233 mV

The output impedance and voltage compliance were simulated in post-layout for both the anodic and cathodic output driver. Tab. 4.3 show these results.

Table 4.3: Output impedance and voltage compliance for the anodic and cathodic output driver.

	Output impedance	Voltage compliance
Anodic	82.04 MΩ	472.9 mV
Cathodic	137.6 MΩ	296.8 mV

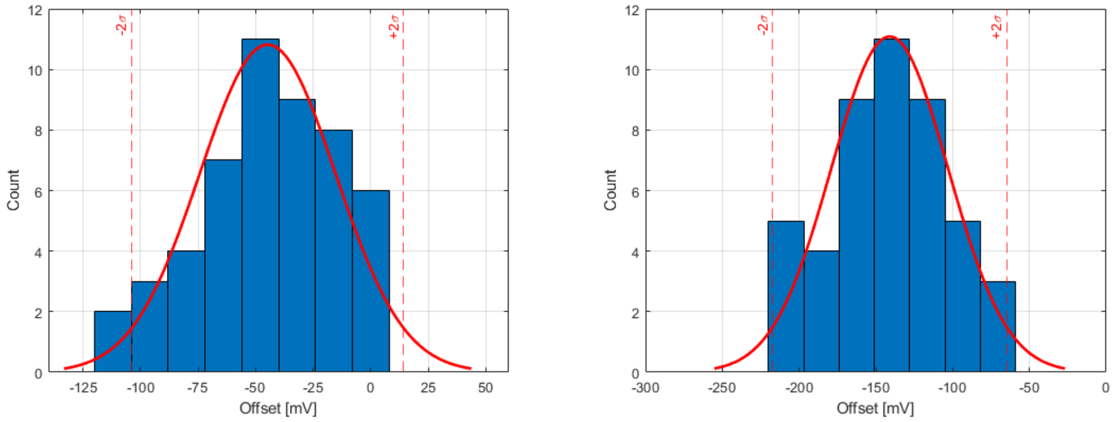
4.2.3. Voltage compliance monitor

The offset of the comparator in the voltage compliance monitor is measured as follows: the cathodic driver is disconnected from any load, and all switches in the current mirror in front of the driver are disconnected. Then, 220 mV is directly placed at the inverting input of the comparator, and a voltage ramp is placed on the non-inverting input. The moment the flag of the monitor is raised is then taken as the comparator switching point, and from this the offset is determined. This is then via a script repeated for all 256 comparator configurations.

Tab. 4.4 shows the measured offset of two chips when untrimmed and when the inverting input is fully trimmed (i.e. all switches at the inverting input are on). The trimming shows a clear decrease in offset, however the initial untrimmed offset is incredibly high. Fig. 4.9a and 4.9b show the histogram of post-layout offset in the untrimmed case and the fully trimmed case. The measured offset difference between fully trimmed and untrimmed aligns with the post-layout simulations (about 97 mV). The measured offset on the other hand does not coincide at all with the post-layout simulations.

Table 4.4: Measured offset of the comparator for chip 1 and chip 3, at untrimmed (config. 255) and fully trimmed at the inverting input (config. 15).

	Chip 1	Chip 3
Config. 255	145 mV	148.3 mV
Config. 15	52 mV	30.2 mV



(a) (b)

Figure 4.9: Histogram of post-layout simulation of the offset of the comparator when (a) untrimmed, and (b) fully trimmed at the inverting input.

4.2.4. Real-time stimulation waveforms

In this Section, the current output will be directly measured. First, the measured output will be compared to the calibrated output and then the multiplexing will be shown.

4.2.4.1. Calibration vs measured output

The output current is checked to ensure that the calibrated LUT is indeed correct. This is done for 6 μA, 3 μA, 1 μA, 138 nA[2], and 23 nA. Since the output current contains high frequency noise (which

---

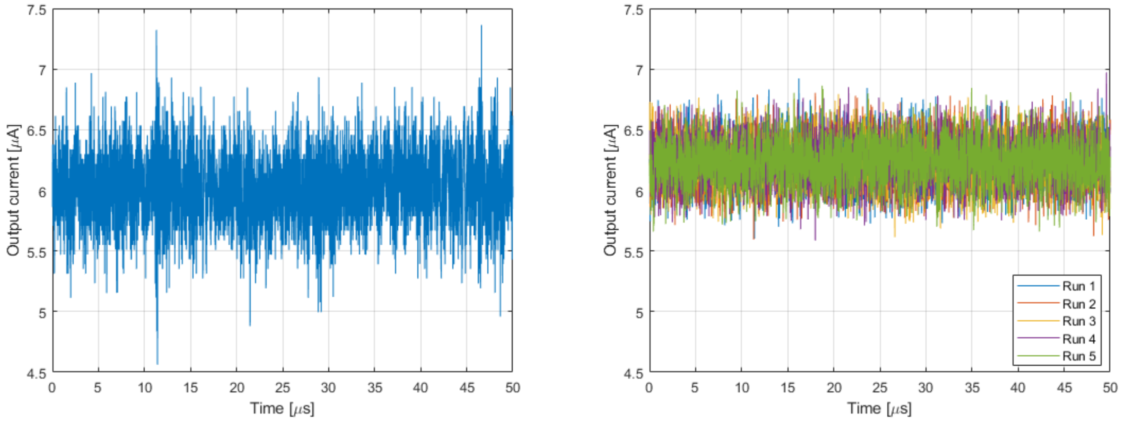
[2] There was no output available at 100 nA, so the closest possible value was taken.

is expected, see Appendix E), the average injected charge is determined and from this, the current. Tab. 4.5 shows these results.

Table 4.5: Calibrated current and measured current for chip 3 (anodic, cathodic) format. For currents below 1 µA only the anodic stimulator is shown, as availability is not important at low currents for the cathodic stimulator.

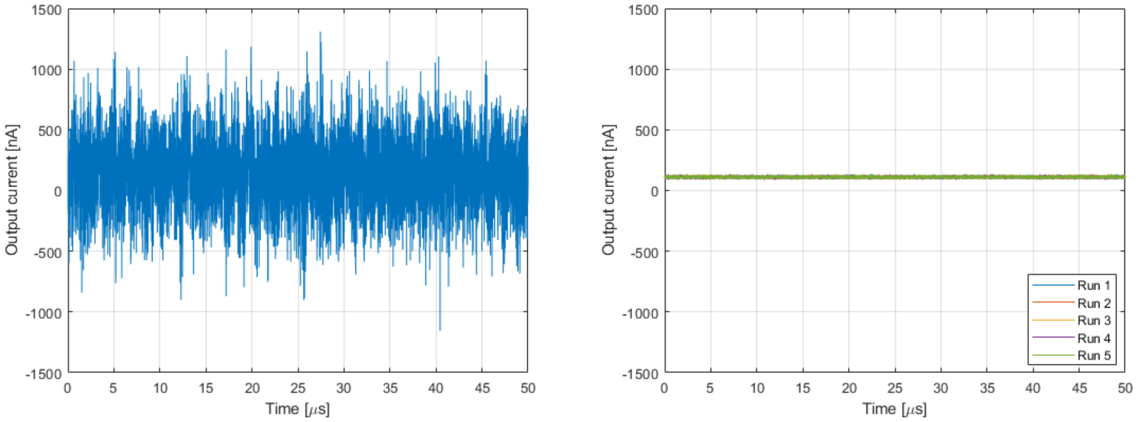
	Calibrated	Measured
6 µA	(6 µA, 6 µA)	(6.02 µA, 6.03 µA)
3 µA	(3 µA, 2.999 µA)	(2.89 µA, 2.92 µA)
1 µA	(1.003 µA, 0.999 µA)	(0.97 µA, 0.94 µA)
138 nA	136.2 nA	138.0 nA
23 nA	24.1 nA	29.4 nA

For the larger currents (>1 µA), the error is larger than an LSB (i.e. 23 nA). Furthermore, Fig. 4.10a and 4.10b show the output current that is measured and obtained from post-layout, respectively at high output. Fig. 4.11a and 4.11b show the same but at low currents.



(a) (b)

Figure 4.10: Transient output current at high output as (a) measured, and (b) obtained from post-layout.



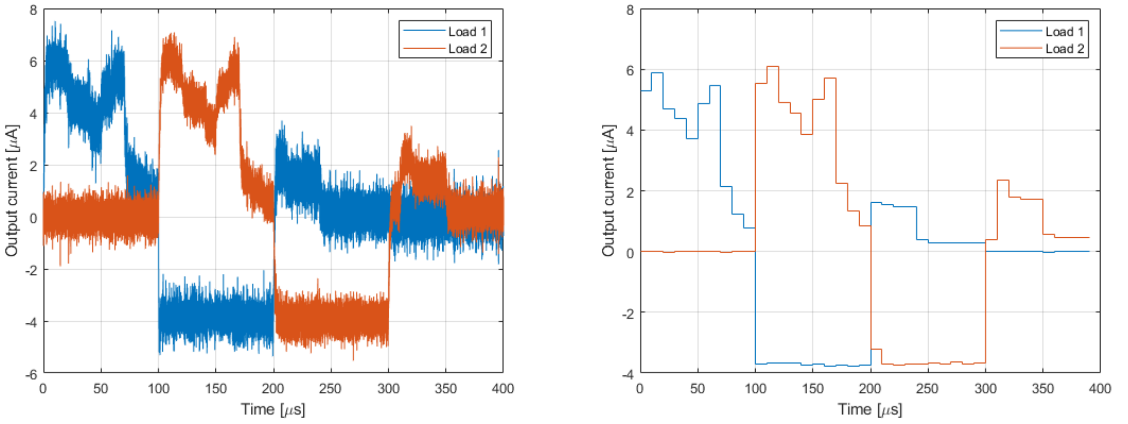
(a) (b)

Figure 4.11: Transient output current at low output as (a) measured, and (b) obtained from post-layout.

While the noise at high output current seems to coincide relatively well between measured and simulated, it becomes clear at low output that this may not be the case, as the same noise is present when it should not be there. This measured noise does not significantly affect the average injected charge, but can be the cause for the error between calibrated output and measured output.

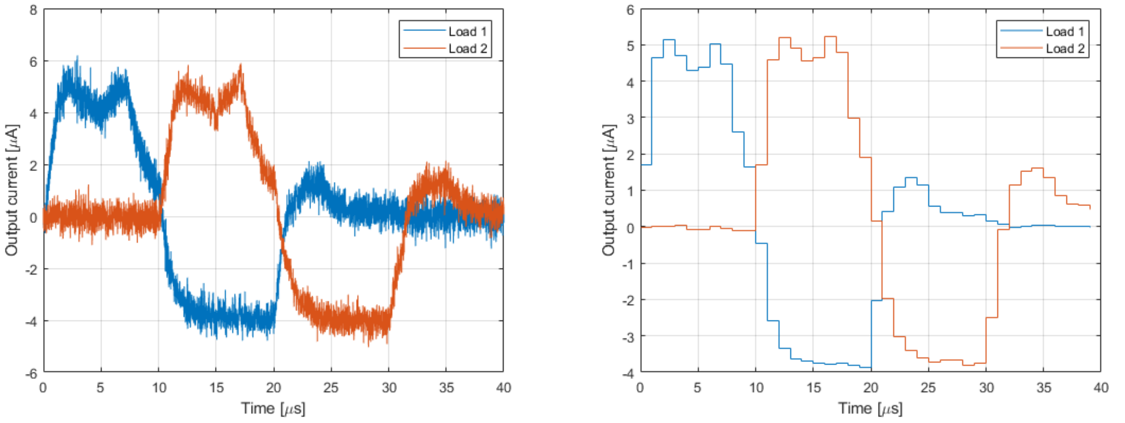
4.2.4.2. Waveform multiplexing

To test the multiplexing and waveform interleaving, an example waveform is generated. Fig. 4.12a and 4.13a show the measured output current with a step duration of 10 µs and 1 µs. The average current based on the amount of injected charge per step (i.e. 10 µs or 1 µs) is plotted in Fig. 4.13a and 4.13b.



(a) (b)

Figure 4.12: Example of a triphasic waveform with interleaving at a step duration of 10 µs, with (a) the measured output current, and (b) the step-averaged output current.



(a) (b)

Figure 4.13: Example of a triphasic waveform with interleaving at a step duration of 1 µs, with (a) the measured output current, and (b) the step-averaged output current.

From the measured output waveforms it is evident that there are slewing issues. This slewing does not change when changing the bias currents of the output drivers, which indicates that the issue should not be due to fabrication variation. Regardless, a post-layout MC simulation with example triphasic waveforms is done to verify this (Fig. 4.14). As there is no significant slewing present in these simulations, the issue has to originate from elsewhere.

The issue is most likely the current generation circuitry on the PCB: while the output current has been measured and is indeed found to be accurate, the settling time due to changing output voltage is significantly longer than the stimulation step duration, causing inaccurate (and insufficient) biasing while

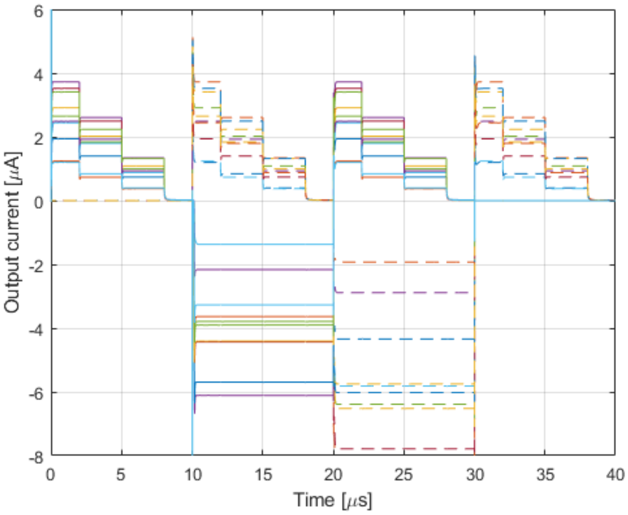
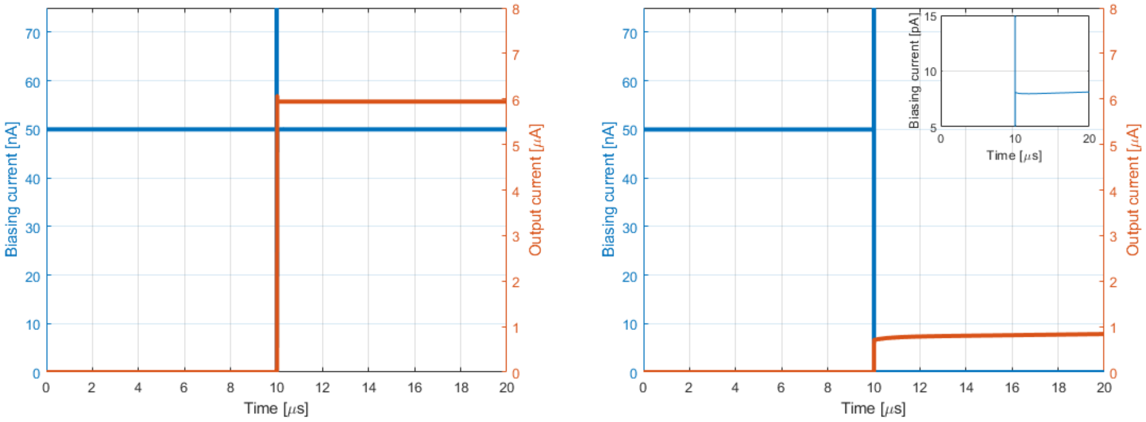


Figure 4.14: Post-layout MC simulation of example interleaved triphasic waveforms. The solid line represents load 1, and the dashed line represents load 2. Each colour is a different MC run.

settling. Fig. 4.15a and 4.15b show the difference between biasing current after changing the output current to 6 µA with an ideal current source and with the current generation circuitry. The latter drops its biasing current to a couple of picoamperes as soon as the output current changes, and then slowly rises back up to 50 nA over 1 ms, which is clearly too slow for stimulation purposes.



(a)

(b)

Figure 4.15: Bias current generation when applying a changing output current. The nominal bias current is 50 nA. (a) contains an ideal biasing current source, and (b) contains the used current generation circuitry. Note: absolute values are not representative, as the correct transistor model is not available in LTspice.

4.2.5. Power consumption

The power consumption is measured in the following conditions: the anodic stimulator is stimulating at 6 μA to load 1, the cathodic stimulator is stimulating at 6 μA to load 2, and both loads are set to a $V_{CM} = 1.65$ V. Then, the current was measured of the 1.1 V, 1.65 V, and 3.3 V rails. Only the current that is used by the neurostimulator is measured, so the power consumption of the digital part of the chip is not considered. Fig. 4.16 shows the distribution of the power consumption. As the output driver has a nominal gain of eight, and the 1.1 V rail is used only before the output driver, it is expected that most of the power is consumed by the output currents (i.e. 1.65 V and 3.3 V rails). The resulting peak power efficiency is, from these results, calculated to be 62.12%.

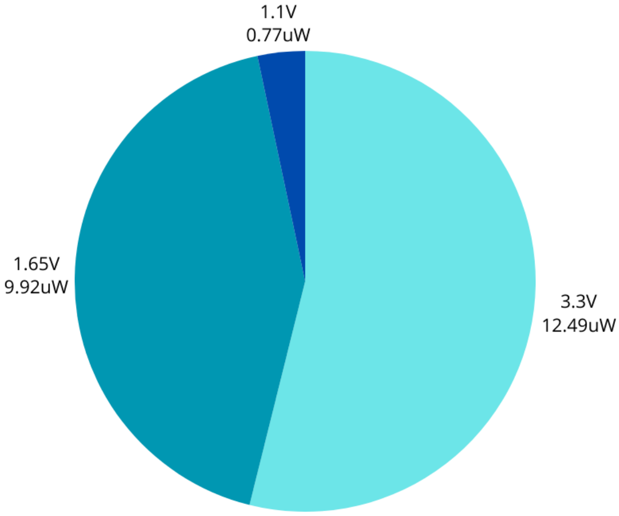


Figure 4.16: Measured distribution of the power consumption.

5

Conclusion and Recommendations

5.1. Conclusion

The objective of this thesis was to design a neurostimulator that is capable of artifact-free stimulation, and is small enough to be able to be scaled up to 10^4 electrodes. The proposed system makes use of mismatch to achieve high effective resolution, and is further optimized to the required specifications of the epiretinal implant (i.e. 0–6 µA output range and 8-bit resolution). To decrease the amount of stimulation units necessary, the anodic and cathodic stimulators are separated and individually multiplexed. As a result, a stimulation unit can provide stimulation to 50% more electrodes. Furthermore, a voltage compliance monitor is implemented in the cathodic stimulator to guarantee safe operation.

The designed system occupies a total area of 25.3 µm by 13 µm and was fabricated. The system performance was verified on a testbench, and it met the 8-bit resolution requirement over the 0–6 µA output range, showing a 0.9922 full output range coverage for the anodic stimulator, and 0.9727 full output range coverage for the cathodic stimulator. Additionally, the waveform interleaving was shown to function. However, the improper current biasing resulted in more slewing than expected: the fully functioning interleaving process was shown to work with no slewing in post-layout simulations.

5.2. Recommendations

5.2.1. System-level

The current system, when scaled up to 10^4 electrodes, would encompass an area of 1.6445 mm^2, which is more than the requirement of 1 mm^2. However, the current system only has a 1:2 demultiplexer as a proof of concept. This demultiplexer can easily be changed to control more electrodes without affecting stimulation functionality (as per Section 1.1.5). The exact demultiplexing ratio should be determined such that cross-talk is negligible and the elicited visual fidelity is unaffected. This requires further testing on either RGC models, or in-vitro RGCs.

5.2.2. Bias current generation

The current design uses external current sources as biasing. The current generation circuitry on the PCB did not work as initially expected, and caused significant slewing. Either a different external biasing current source needs to be implemented on the PCB, or the entire current biasing could be moved on-chip in a future design. An advantage of having it be on-chip is that a feedback loop could be implemented that pushes the output mirror transistor in the output drivers as close as possible to saturation automatically. In this way, the required voltage headroom could be minimized regardless of process variations.

5.2.3. Comparator offset

The offset of the comparator when untrimmed was determined to be in the order of -50 mV when performing post-layout simulations. The comparator itself is both circuit symmetric and operating point symmetric, and thus has no offset when in isolation. However, since the output node of the comparator is connected to, in this case, an inverter, the capacitance seen by the output node is now different than the capacitance seen by the inverted output node. This in turn explains the negative offset: a larger output node capacitance requires a larger current to charge to the same voltage. In the future, this should be addressed, either by connecting a dummy inverter to the inverting output, or through any other method that equalizes the output node capacitances.

Finally, the measured untrimmed on-chip offset was in the range of $150\,\text{mV}$ on multiple chips, which does not coincide with the post-layout results at all. The reason for this should be investigated. Since the chosen comparator generally can have offset due to unbalanced capacitive loading of either the output nodes or the precharged nodes (between the input transistors and the first latched transistors), any potential capacitive loading at these nodes due to the PCB should be looked into first. Simulations show that a capacitance of roughly $6\,\text{fF}$ at the unused output node can lead to the measured offset.

5.2.4. Measured noise

The origin of the noise that is present in all output current measurements should be analyzed. Especially since the noise is even present, with equal amplitude, when all the switches in the neurostimulator are disconnected and there therefore is no output current. This would indicate that the noise source is either due to not simulated parts of the chip, such as the pads or surrounding routing of other designs, or due to the PCB / measurement set-up. It is unlikely to be interference due to ground loops or nearby stray fields, as the frequency analysis of this noise showed no specific peaks.

Bibliography

[1] Stanford Medicine. Research - Stanford Artificial Retina Project. URL `https://med.stanford.edu/artificial-retina/research.html`. Accessed: 06-10-2022.

[2] Joshua R. Sanes and Richard H. Masland. The Types of Retinal Ganglion Cells: Current Status and Implications for Neuronal Classification. *http://dx.doi.org/10.1146/annurev-neuro-071714-034120*, 38:221–246, 7 2015. ISSN 15454126. doi:10.1146/ANNUREV-NEURO-071714-034120. URL `https://www.annualreviews.org/doi/abs/10.1146/annurev-neuro-071714-034120`.

[3] Kuanfu Chen, Yi Kai Lo, and Wentai Liu. A 37.6mm2 1024-channel high-compliance-voltage SoC for epiretinal prostheses. *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, 56:294–295, 2013. ISSN 01936530. doi:10.1109/ISSCC.2013.6487741.

[4] Dante G Muratore, E J Chichilnisky, D G Muratore, and Wu Tsai. Artificial Retina: A Future Cellular-Resolution Brain-Machine Interface. pages 443–465, 2020. doi:10.1007/978-3-030-18338-7_24. URL `https://link.springer.com/chapter/10.1007/978-3-030-18338-7_24`.

[5] Chun Wang, Emma Brunton, Saman Haghgooie, Kahli Cassells, Arthur Lowery, and Ramesh Rajan. Characteristics of electrode impedance and stimulation efficacy of a chronic cortical implant using novel annulus electrodes in rat motor cortex. *Journal of neural engineering*, 10(4), 8 2013. ISSN 1741-2552. doi:10.1088/1741-2560/10/4/046010. URL `https://pubmed.ncbi.nlm.nih.gov/23819958/`.

[6] Hyunjun Jung, Jintae Kim, and Yoonkey Nam. Recovery of early neural spikes from stimulation electrodes using a DC-coupled low gain high resolution data acquisition system. *Journal of Neuroscience Methods*, 304:118–125, 7 2018. ISSN 0165-0270. doi:10.1016/J.JNEUMETH.2018.04.014.

[7] Gonzalo E. Mena, Lauren E. Grosberg, Sasidhar Madugula, Paweł Hottowy, Alan Litke, John Cunningham, E. J. Chichilnisky, and Liam Paninski. Electrical stimulus artifact cancellation and neural spike detection on large multi-electrode arrays. *PLoS computational biology*, 13(11), 11 2017. ISSN 1553-7358. doi:10.1371/JOURNAL.PCBI.1005842. URL `https://pubmed.ncbi.nlm.nih.gov/29131818/`.

[8] Rohan Brash. *Artifact-Free Stimulation for Next-Generation Single-Cell Resolution Epiretinal Implants*. MSc Thesis, TU Delft, 2021. URL `https://repository.tudelft.nl/islandora/object/uuid%3A5d776dcf-95c6-4dce-a425-560595cf7d81`.

[9] Jay T. Rubinstein, Charles A. Miller, Hiroyuki Mino, and Paul J. Abbas. Analysis of monophasic and biphasic electrical stimulation of nerve. *IEEE Transactions on Biomedical Engineering*, 48 (10):1065–1070, 2001. ISSN 00189294. doi:10.1109/10.951508.

[10] Lvpiao Zheng, Zhouyan Feng, Yipeng Xu, Yue Yuan, and Yifan Hu. An Anodic Phase Can Facilitate Rather Than Weaken a Cathodic Phase to Activate Neurons in Biphasic-Pulse Axonal Stimulations. *Frontiers in Neuroscience*, 16:823423, 3 2022. ISSN 1662453X. doi:10.3389/FNINS.2022.823423/BIBTEX.

[11] Yu Wu, Dai Jiang, and Andreas Demosthenous. A Multi-Channel Stimulator with High-Resolution Time-to-Current Conversion for Vagal-Cardiac Neuromodulation. *IEEE Transactions on Biomedical Circuits and Systems*, 15(6):1186–1195, 12 2021. ISSN 19409990. doi:10.1109/TBCAS.2021.3139996.

[12] G. Luis E. Rueda, Marco Ballini, Nick Van Hellepute, and Srinjoy Mitra. Analysis of passive charge balancing for safe current-mode neural stimulation. *Proceedings - IEEE International Symposium on Circuits and Systems*, 9 2017. ISSN 02714310. doi:10.1109/ISCAS.2017.8050621.

[13] Chun-Yu Lin and Ming-Dou Ker. Overview of on-Chip Stimulator Designs for Biomedical Applications. *undefined*, 1(2):204–212, 4 2012. ISSN 21682011. doi:10.1166/JNSNE.2012.1023.

[14] Victoria H. Fan, Lauren E. Grosberg, Sasidhar S. Madugula, Pawel Hottowy, Wladyslaw Dabrowski, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. Epiretinal stimulation with local returns enhances selectivity at cellular resolution. *Journal of neural engineering*, 16(2):025001, 2019. ISSN 17412552. doi:10.1088/1741-2552/AAEEF1. URL `/pmc/articles/PMC6416068//pmc/articles/PMC6416068/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6416068/`.

[15] Elliot Greenwald, Christoph Maier, Qihong Wang, Robert Beaulieu, Ralph Etienne-Cummings, Gert Cauwenberghs, and Nitish Thakor. A CMOS Current Steering Neurostimulation Array with Integrated DAC Calibration and Charge Balancing. *IEEE Transactions on Biomedical Circuits and Systems*, 11(2):324–335, 4 2017. ISSN 19324545. doi:10.1109/TBCAS.2016.2609854.

[16] Sajjad Shahsavari and Mehdi Saberi. A Highly Linear 8-Bit M-2M Digital-to-Analog Converter for Neurostimulators. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 67(6):989–993, 6 2020. ISSN 15583791. doi:10.1109/TCSII.2019.2928548.

[17] Mario A. Meza-Cuevas, Dietmar Schroeder, and Wolfgang H. Krautschneider. A scalable 64 channel neurostimulator based on a hybrid architecture of current steering DAC. *Middle East Conference on Biomedical Engineering, MECBME*, pages 111–114, 2014. ISSN 21654255. doi:10.1109/MECBME.2014.6783218.

[18] Chia Chi Hsieh and Ming Dou Ker. Monopolar Biphasic Stimulator with Discharge Function and Negative Level Shifter for Neuromodulation SoC Integration in Low-Voltage CMOS Process. *IEEE Transactions on Biomedical Circuits and Systems*, 15(3):568–579, 6 2021. ISSN 19409990. doi:10.1109/TBCAS.2021.3087036.

[19] Paulo Cesar C. De Aguirre, Fabio D. Da Fonseca, and Alessandro G. Girardi. An 8-Bit 4x4 Segmented Current-Steering DAC for Bipolar Biphasic Stimulators. *Midwest Symposium on Circuits and Systems*, 2022-August, 2022. ISSN 15483746. doi:10.1109/MWSCAS54063.2022.9859492.

[20] Mikhail S. Yenuchenko. Alternative structures of a segmented current-steering DAC. *2018 International Symposium on Consumer Technologies, ISCT 2018*, pages 14–17, 7 2018. doi:10.1109/ISCE.2018.8408905.

[21] Anh Tuan Nguyen, Jian Xu, Diu Khue Luu, Cynthia K. Overstreet, Jonathan Cheng, Edward W. Keefer, and Zhi Yang. Redundant Crossfire: A Technique to Achieve Super-Resolution in Neurostimulator Design by Exploiting Transistor Mismatch. *IEEE Journal of Solid-State Circuits*, 56 (8):2452–2465, 8 2021. ISSN 1558173X. doi:10.1109/JSSC.2021.3057041.

[22] Jong Hyeon Chang, Jungil Park, Youngmi Kim Pak, and James Jungho Pak. Fitting improvement using a new electrical circuit model for the electrode-electrolyte interface. *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, pages 572–574, 2007. doi:10.1109/CNE.2007.369737.

[23] M. Sawan, Y. Laaziri, F. Mounaim, E. Elzayat, J. Corcos, and M. M. Elhilali. Electrode–tissues interface: modeling and experimental validation. *Biomedical Materials*, 2(1):S7, 3 2007. ISSN 1748-605X. doi:10.1088/1748-6041/2/1/S02. URL `https://iopscience.iop.org/article/10.1088/1748-6041/2/1/S02https://iopscience.iop.org/article/10.1088/1748-6041/2/1/S02/meta`.

[24] Bence Csomós, Dénes Fodor, and István Vajda. Estimation of Battery Separator Area, Cell Thickness and Diffusion Coefficient Based on Non-Ideal Liquid-Phase Diffusion Modeling. *Energies 2020, Vol. 13, Page 6238*, 13(23):6238, 11 2020. ISSN 1996-1073. doi:10.3390/EN13236238. URL `https://www.mdpi.com/1996-1073/13/23/6238/htmhttps://www.mdpi.com/1996-1073/13/23/6238`.

[25] Karolina Kolodziej, Malgorzata Szypulska, Wladyslaw Dabrowski, and Pawel Hottowy. Modelling and Cancellation of the Stimulation Artifact for ASIC-based Bidirectional Neural Interface. *Proceedings of 25th International Conference Mixed Design of Integrated Circuits and Systems, MIXDES 2018*, pages 449–453, 8 2018. doi:10.23919/MIXDES.2018.8436947.

[26] Kriangkrai Sooksood, Thomas Stieglitz, and Maurits Ortmanns. An active approach for charge balancing in functional electrical stimulation. *IEEE Transactions on Biomedical Circuits and Systems*, 4(3):162–170, 6 2010. ISSN 19324545. doi:10.1109/TBCAS.2010.2040277.

[27] Daniel R. Merrill, Marom Bikson, and John G.R. Jefferys. Electrical stimulation of excitable tissue: design of efficacious and safe protocols. *Journal of Neuroscience Methods*, 141(2):171–198, 2 2005. ISSN 0165-0270. doi:10.1016/J.JNEUMETH.2004.10.020.

[28] A. T. Gardner, Hunter J. Strathman, and Ross M. Walker. A multiplexed electrochemical measurement system for characterization of implanted electrodes. *Proceedings - IEEE International Symposium on Circuits and Systems*, 2020-October, 2020. ISSN 02714310. doi:10.1109/ISCAS45731.2020.9180496/VIDEO. URL `https://ieeexplore-ieee-org.tudelft.idm.oclc.org/document/9180496`.

[29] Hui Xie, Yi Wang, Zixin Ye, Shiyi Fang, Zhen Xu, Tianzhun Wu, and Leanne Lai Hang Chan. Monitoring cortical response and electrode-retina impedance under epiretinal stimulation in rats. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:1178–1187, 2021. ISSN 15580210. doi:10.1109/TNSRE.2021.3090904.

[30] Wendy Franks, Iwan Schenker, Patrik Schmutz, and Andreas Hierlemann. Impedance characterization and modeling of electrodes for biomedical applications. *IEEE Transactions on Biomedical Engineering*, 52(7):1295–1302, 7 2005. ISSN 00189294. doi:10.1109/TBME.2005.847523.

[31] Prateek Vajpayee, A. Srivastava, S. S. Rajput, and G. K. Sharma. Low voltage regulated cascode current mirrors suitable for sub-1V operation. *IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings, APCCAS*, pages 584–587, 2010. doi:10.1109/APCCAS.2010.5774891.

[32] Kevin Meng, Andreas Fellner, Frank Rattay, Diego Ghezzi, Hamish Meffin, Michael R. Ibbotson, and Tatiana Kameneva. Upper stimulation threshold for retinal ganglion cell activation. *Journal of neural engineering*, 15(4), 5 2018. ISSN 1741-2552. doi:10.1088/1741-2552/AABB7D. URL `https://pubmed.ncbi.nlm.nih.gov/29616983/`.

[33] Tianruo Guo, David Tsai, Chih Yu Yang, Amr Al Abed, Perry Twyford, Shelley I. Fried, John W. Morley, Gregg J. Suaning, Socrates Dokos, and Nigel H. Lovell. Mediating retinal ganglion cell spike rates using high-frequency electrical stimulation. *Frontiers in Neuroscience*, 13(APR):413, 4 2019. ISSN 1662453X. doi:10.3389/FNINS.2019.00413/BIBTEX.

[34] Chris Sekirnjak, Pawel Hottowy, Alexander Sher, Wladyslaw Dabrowski, A. M. Litke, and E. J. Chichilnisky. Electrical stimulation of mammalian retinal ganglion cells with multielectrode arrays. *Journal of Neurophysiology*, 95(6):3311–3327, 6 2006. ISSN 00223077. doi:10.1152/JN.01168.2005/ASSET/IMAGES/LARGE/Z9K0050673980012.JPEG. URL `https://journals.physiology.org/doi/10.1152/jn.01168.2005`.

[35] Ralph J. Jensen, Ofer R. Ziv, and Joseph F. Rizzo. Thresholds for activation of rabbit retinal ganglion cells with relatively large, extracellular microelectrodes. *Investigative ophthalmology & visual science*, 46(4):1486–1496, 4 2005. ISSN 0146-0404. doi:10.1167/IOVS.04-1018. URL `https://pubmed.ncbi.nlm.nih.gov/15790920/`.

[36] David Boinagrov, Jim Loudin, and Daniel Palanker. Strength–Duration Relationship for Extracellular Neural Stimulation: Numerical and Analytical Models. *Journal of Neurophysiology*, 104(4):2236, 10 2010. ISSN 00223077. doi:10.1152/JN.00343.2010. URL `/pmc/articles/PMC2957463//pmc/articles/PMC2957463/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2957463/`.

[37] Manjunatha Mahadevappa, James D. Weiland, Douglas Yanai, Ione Fine, Robert J. Greenberg, and Mark S. Humayun. Perceptual thresholds and electrode impedance in three retinal prosthesis subjects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(2):201–206, 6 2005. ISSN 15344320. doi:10.1109/TNSRE.2005.848687.

[38] Nishal P. Shah and E. J. Chichilnisky. Computational challenges and opportunities for a bi-directional artificial retina. *Journal of neural engineering*, 17(5), 10 2020. ISSN 1741-2552. doi:10.1088/1741-2552/ABA8B1. URL `https://pubmed-ncbi-nlm-nih-gov.tudelft.idm.oclc.org/33089827/`.

[39] M. Reza Pazhouhandeh, Michael Chang, Taufik A. Valiante, and Roman Genov. Track-and-zoom neural analog-to-digital converter with blind stimulation artifact rejection. *IEEE Journal of Solid-State Circuits*, 55(7):1984–1997, 7 2020. ISSN 1558173X. doi:10.1109/JSSC.2020.2991526.

[40] Jay Shah, Christopher Quinkert, Brett Collar, Michael Williams, Ethan Biggs, and Pedro Irazoqui. A Highly Miniaturized, Chronically Implanted ASIC for Electrical Nerve Stimulation. *IEEE Transactions on Biomedical Circuits and Systems*, 16(2):233–243, 4 2022. ISSN 19409990. doi:10.1109/TBCAS.2022.3153282.

[41] Pascal Raffelberg, Roman Burkard, Reinhard Viga, Wilfried Mokwa, Peter Walter, Anton Grabmaier, and Rainer Kokozinski. Current Controlled CMOS Stimulator with Programmable Pulse Pattern for a Retina Implant. *PRIME 2018 - 14th Conference on Ph.D. Research in Microelectronics and Electronics*, pages 253–256, 8 2018. doi:10.1109/PRIME.2018.8430332.

[42] Hossein Kassiri, Gairik Dutta, Nima Soltani, Chang Liu, Yu Hu, and Roman Genov. An impedance-tracking battery-less arbitrary-waveform neurostimulator with load-adaptive 20V voltage compliance. *European Solid-State Circuits Conference*, 2016-October:225–228, 10 2016. ISSN 19308833. doi:10.1109/ESSCIRC.2016.7598283.

[43] Nishal P. Shah, Sasidhar Madugula, Lauren Grosberg, Gonzalo Mena, Pulkit Tandon, Pawel Hottowy, Alexander Sher, Alan Litke, Subhasish Mitra, and E. J. Chichilnisky. Optimization of Electrical Stimulation for a High-Fidelity Artificial Retina. *International IEEE/EMBS Conference on Neural Engineering, NER*, 2019-March:714–718, 5 2019. ISSN 19483554. doi:10.1109/NER.2019.8716987.

[44] Lauren E. Grosberg, Karthik Ganesan, Georges A. Goetz, Sasidhar S. Madugula, Nandita Bhaskhar, Victoria Fan, Peter Li, Pawel Hottowy, Wladyslaw Dabrowski, Alexander Sher, Alan M. Litke, Subhasish Mitra, and E. J. Chichilnisky. Activation of ganglion cells and axon bundles using epiretinal electrical stimulation. *Journal of neurophysiology*, 118(3):1457–1471, 9 2017. ISSN 1522-1598. doi:10.1152/JN.00750.2016. URL `https://pubmed.ncbi.nlm.nih.gov/28566464/`.

[45] Erich W. Schmid, Wolfgang Fink, and Robert Wilke. Simultaneous vs. sequential and unipolar vs. multipolar stimulation in retinal prostheses. *International IEEE/EMBS Conference on Neural Engineering, NER*, pages 190–193, 2013. ISSN 19483546. doi:10.1109/NER.2013.6695904.

[46] Yun Ti Wang and Behzad Razavi. 8-Bit 150-MHz CMOS A/D converter. *IEEE Journal of Solid-State Circuits*, 35(3):308–317, 2000. ISSN 00189200. doi:10.1109/4.826812.

[47] Behzad Razavi. The StrongARM latch [A Circuit for All Seasons]. *IEEE Solid-State Circuits Magazine*, 7(2):12–17, 3 2015. ISSN 19430582. doi:10.1109/MSSC.2015.2418155.

[48] Analog Devices. LTC6081/LTC6082 - Precision Dual/Quad CMOS Rail-to-Rail Input/Output Amplifiers datasheet. URL `https://www.analog.com/media/en/technical-documentation/data-sheets/60812fd.pdf`. Accessed: 18-12-2023.

[49] Teresa Serrano-Gotarredona and Bernabé Linares-Barranco. Systematic width-and-length dependent CMOS transistor mismatch characterization and simulation. *Analog Integrated Circuits and Signal Processing*, 21(3):271–296, 1999. ISSN 09251030. doi:10.1023/A:1008330121404/METRICS. URL `https://link-springer-com.tudelft.idm.oclc.org/article/10.1023/A:1008330121404`.

[50] Marcel J.M. Pelgrom, Aad C.J. Duinmaijer, and Anton P.G. Welbers. Matching Properties of MOS Transistors. *IEEE Journal of Solid-State Circuits*, 24(5):1433–1439, 1989. ISSN 1558173X. doi:10.1109/JSSC.1989.572629.

A

Code

This chapter contains the relevant written software used during the thesis project. Code written for the MCU is not included here.

A.1. RXF Performance Characterization

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%   RXF Performance Characterization     %
%   Author: Hsukang Chen                 %
%   Last modification date: 17-Oct-2023  %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

close all;
clear all;

load dataset_name.mat           % Load your dataset
MCTest = workspace_name;        % Change second part of this to name
                                % of dataset as seen in workspace

max_current = 6e-6;             % Maximum current relevant
RXF_resolution = 8;             % Resolution to which results
                                % are characterized
steps = 2^RXF_resolution - 1;   % Calculate amount of discrete steps
oversampling_factor = 5;        % OSF for calibration
runs = size(MCTest,2) / 2;      % Find out how many MC runs were done

MC_absolute = zeros(1, size(MCTest, 1)); % Preallocate variable
odd_vector = 1:2:2*runs;                 % Create vector of odd numbers

for i = 1:runs
    odd = odd_vector(i);      % The odd part contains timestamp of data
    even = odd_vector(i) + 1; % The even part contains output current data

    MC_absolute(i,:) = transpose(MCTest(:,even)); % Make matrix with
                                                  % only output currents
end

MC_abs_sorted = sort(MC_absolute, 2); % Put outputs in ascending order

% Sort up to max_current plus 10% (for some leeway)
```

```
% Technically not necessary, but speeds up simulation by discarding
% the outputs that fall outside our relevant range
indices_max = find(MC_abs_sorted>(max_current*1.1));
MC_abs_sorted(indices_max) = NaN;

MC_abs_diff = diff(MC_abs_sorted, 1, 2);      % Create difference matrix

diff_min = min(MC_abs_diff, [], 2);           % Find minimum difference
diff_max = max(MC_abs_diff, [], 2);           % Find maximum difference
diff_mean = mean(MC_abs_diff, 2, 'omitnan'); % Find mean difference
% NOTE: omitnan is necessary as it otherwise considers NaNs a numerical 0

current_steps = 2^RXF_resolution; % Technically same as 'steps' + 1
% Artifact from since then removed code (too much hassle to change all
% other code to 'steps')

% Split max_current into current_steps of bins
current_step_array = linspace(0, max_current, current_steps);

% Sort analog output current into digital (uA -> LSBs)
discrete_MC_abs_sorted = discretize(MC_abs_sorted, current_step_array);

% If running slowly, INCREASE THIS to preallocate sufficiently!
sens_array = NaN(100000, current_steps);

for j = 1:current_steps      % For all steps
    % Find the indeces of all points that are equal to the step
    idx = find(discrete_MC_abs_sorted == j);
    for k = 1:size(idx, 1)
        % Disregard last column, as diff cannot be calculated,
        % since there is no next column.
        if idx(k) < ((runs-1) * size(discrete_MC_abs_sorted, 2))
            % Put the difference into the array
            sens_array(k,j) = MC_abs_diff(idx(k));
        end
    end
end

sens_min = min(sens_array, [], 1); % Minimum difference
sens_max = max(sens_array, [], 1); % Maximum difference
sens_mean = mean(sens_array, 1, 'omitnan'); % Mean difference
sens_95 = prctile(sens_array , 95, 1); % 95th percentile
sens_5 = prctile(sens_array , 5, 1); % 5th percentile

figure % Plots the difference at 5th, 50th (median), and 95th percentile
plot(current_step_array, sens_mean.*10^9, 'k');
hold on
plot(current_step_array, sens_95.*10^9, 'b');
hold on
plot(current_step_array, sens_5.*10^9, 'r');
legend("50th %", "95th %", "5th %");
title("Smallest possible amplitude step at each output current");
xlabel("Output current (A)");
ylabel("Smallest possible step (nA)");

% Make a matrix that contains whether or not there is AT LEAST 1 output
```

```
% possible for each output, for each MC run.
occurence = zeros(runs, current_steps);
occurence_boolean = zeros(runs, current_steps);
for j = 1:runs
    for i = 1:current_steps
        occurence(j,i) = sum(discrete_MC_abs_sorted(j,:) == i);
        if occurence(j,i) > 0
            occurence_boolean(j,i) = 1;
        end
    end
end

% Convert to a ratio of occurrence
occurence_ratio = zeros(1,current_steps);
for k = 1:current_steps
    occurence_ratio(k) = sum(occurence_boolean(:,k)) / runs;
end

% Plots the ratio of occurrence over the full range
figure
plot(current_step_array, occurence_ratio);
ylabel(”Ratio of occurrence”);
xlabel(”Output current (A)”);
title(”Occurrence of output step over full range”);

% Checks availability over full range
hits_fullrange_per_run = zeros(runs,1);
for x = 1:runs      % For each MC run
    nr_hits = 0;    % Reset hit count
    for z = 1:steps % For each discrete output step
        if occurence_boolean(x,z) == 1 % Check if run x has step z at
                                       % least once
            nr_hits = nr_hits + 1;
        end
    end
    hits_fullrange_per_run(x) = nr_hits;
end

% Divide amount of hits by total steps for percentage
pct_fullrange_per_run = hits_fullrange_per_run./steps.*100;

% Plots availability of the output range per run
figure
histogram(pct_fullrange_per_run, ’BinWidth’, 0.25);
% Use bin widths of 0.25% (can be adjusted to whatever binwidth)
title(”Distribution of availability of full range”);
xlabel(”Availability of full range (%)”);
ylabel(”Number of MC runs”);

% Checks how many runs have the FULL output range available
boolean_fullrange_possible = zeros(runs,1);
for k = 1:runs
    if pct_fullrange_per_run(k) == 100 % If run k has full coverage
                                       % of entire range
      boolean_fullrange_possible(k) = 1;
    end
```

```
end

% Ratio of runs that cover the entire range
ratio_fullrange_possible = sum(boolean_fullrange_possible) / runs;
```

A.2. Chip Calibration and Characterization

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%   Chip calibration and characterization   %
%   Author: Hsukang Chen                    %
%   Last modification date: 20-Dec-2023     %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear all;

load dataset.mat                % Load your dataset
decData = hex2dec(fileName);    % Change to your file name
decData = decData ./ 100;       % Same as mult_factor in MCU IDE
ADC_supply = 3.3;
ADC_resolution = 12;            % Resolution of the MCU ADC
ADC_LSB = ADC_supply / 2^ADC_resolution;
resistance = 54.6e3;            % Change to measured resistance value
Anodic_Cathodic_Cutoff = 32769; % Change to first cathodic index (32769)

for i = 1:(Anodic_Cathodic_Cutoff - 1)  % Creates configuration arrays
    AnodicConfig(i) = i-1;
end

for i = 1:(size(decData, 1) - Anodic_Cathodic_Cutoff + 1)
    CathodicConfig(i) = i-1;
end

for i = 1:size(decData, 1)      % Calculate current from the measured voltage
    if (i < Anodic_Cathodic_Cutoff)
        fullAnodicData(i, 1) = AnodicConfig(i);
        fullAnodicData(i, 2) = (decData(i) * ADC_LSB) / resistance ;
    else
        fullCathodicData(i-Anodic_Cathodic_Cutoff + 1, 1) = CathodicConfig(i-Anodic_Cathodic_Cutoff + 1);
        fullCathodicData(i-Anodic_Cathodic_Cutoff + 1, 2) = (ADC_supply - decData(i) * ADC_LSB) / resistance;
    end
end

sorted_Anodic = sortrows(fullAnodicData, 2);        % Sort in ascending order
sorted_Cathodic = sortrows(fullCathodicData, 2);

%% Anodic find - Finds the closest configuration for your output current
wantedValue = value;         % Change value to output current you want
dist    = abs(sorted_Anodic(:,2) - wantedValue);
minDist = min(dist);
logicalIdx  = (dist == minDist);
minVal   = sorted_Anodic(logicalIdx, 2)
minIdx = find(sorted_Anodic(:,2) == minVal);
```

```
minConfig_Anodic = sorted_Anodic(minIdx, 1)

%% Cathodic find - Finds the closest configuration for your output current
wantedValue = value; % Change value to output current you want
dist    = abs(sorted_Cathodic(:,2) - wantedValue);
minDist = min(dist);
logicalIdx  = (dist == minDist);
minVal  = sorted_Cathodic(logicalIdx, 2)
minIdx = find(sorted_Cathodic(:,2) == minVal);
minConfig_Cathodic = sorted_Cathodic(minIdx, 1)

%% Generate full LUT - Loop through all LSBs and find closest output
   current
resolution = 8;
max_output = 6e-6;
LSB = max_output / 2^resolution;
integerArray = 0:(2^resolution - 1);
outputValues_wanted = integerArray * LSB;
output_Anodic = zeros(256, 1);
output_Cathodic = zeros(256,1);

for i=1:2^resolution
    wantedValue = outputValues_wanted(i);
    dist    = abs(sorted_Anodic(:,2) - wantedValue);
    minDist = min(dist);
    logicalIdx  = (dist == minDist);
    minVal  = sorted_Anodic(logicalIdx, 2);
    output_Anodic(i) = minVal(1);

    dist    = abs(sorted_Cathodic(:,2) - wantedValue);
    minDist = min(dist);
    logicalIdx  = (dist == minDist);
    minVal  = sorted_Cathodic(logicalIdx, 2);
    output_Cathodic(i) = minVal(1);
end

%% INL/DNL on LUT - Calculate INL and DNL
for i=1:2^8
    INL_anodic(i) = (output_Anodic(i) - outputValues_wanted(i))/LSB;
    INL_cathodic(i) = (output_Cathodic(i) - outputValues_wanted(i))/LSB;
end

for i=1:2^8
    if (i==1)
        DNL_anodic(i) = output_Anodic(i)/LSB;
        DNL_cathodic(i) = output_Cathodic(i)/LSB;
    else
        DNL_anodic(i) = ((output_Anodic(i) - output_Anodic(i-1))/LSB) - 1;
        DNL_cathodic(i) = ((output_Cathodic(i) - output_Cathodic(i-1))/LSB
            ) - 1;
    end
end
```

B

Mismatch

While it is fully possible (and perhaps even quicker) to simply change parameters in Cadence Virtuoso, run a simulation, and check its output many times over to find an optimal value, it would be prudent to first discuss the nature of mismatch and its consequences to reduce the amount of ”Spice monkeying” (or ”Cadence monkeying”).

There are several known contributors that can each cause the output current to deviate, which can approximately be grouped into two categories: (1) deviations caused by the inherent inaccuracy of chip production (e.g. varying doping concentrations, material thickness variations, etc.), and (2) deviations caused by systematic gradients [49]. Since systematic gradients often occur over larger dies and are gradual changes, these will not be taken into consideration for now. The random nature of mismatch is primarily attributed to the first category, i.e. even if two transistors are right next to each other, they will exhibit mismatch.

A well-known and commonly used mismatch model is the Pelgrom model [50] wherein parameter deviation between two transistors follow the equation

$$\sigma^2 = \frac{A^2}{WL} \tag{B.1}$$

with σ as the standard deviation, A as a constant dependent on the used technology, and W and L being the transistors widths and lengths. Most of the time, the dominant transistor parameter affecting output current is the threshold voltage V_{TH}. Since any variation in V_{TH} is directly seen in the output current (as the overdrive voltage V_{OV} changes) by a factor set by its transconductance, the variation in output current caused by mismatch can also be related to the inverse of the gate area via

$$\sigma_I^2 = \frac{4A_{VT}^2}{WL \cdot V_{OV}^2} \tag{B.2}$$

and can thus be tuned by increasing or decreasing the gate area. Notably, increasing the overdrive voltage also reduces mismatch, but often is set by the required current. Ideally, minimum size transistors (i.e. 120 nm by 40 nm) would correspond to the optimal mismatch needed.

Since A_{VT} (in Eq. B.2) is effectively an unknown parameter, simulation data will have to be used to find σ_I. Fig. B.1a and B.1b contain the proportional standard deviation of the output current of a standard current cell obtained via simulation compared against the theoretical curve, for NMOS and PMOS devices respectively, over different amount of fingers.

The authors of [21] pose that a standard deviation of around 0.1 to 0.2 is optimal as higher standard deviations does not increase the effective resolution, and instead decreases yields at the higher end of the output range as seen in Fig. B.2.

This estimation of optimal mismatch is done under the assumption that the mismatch follows a Gaussian distribution. This assumption is verified via simulation, and an example of the histograms with its corresponding distribution can be seen for both NMOS and PMOS in Fig. B.3a and B.3b.

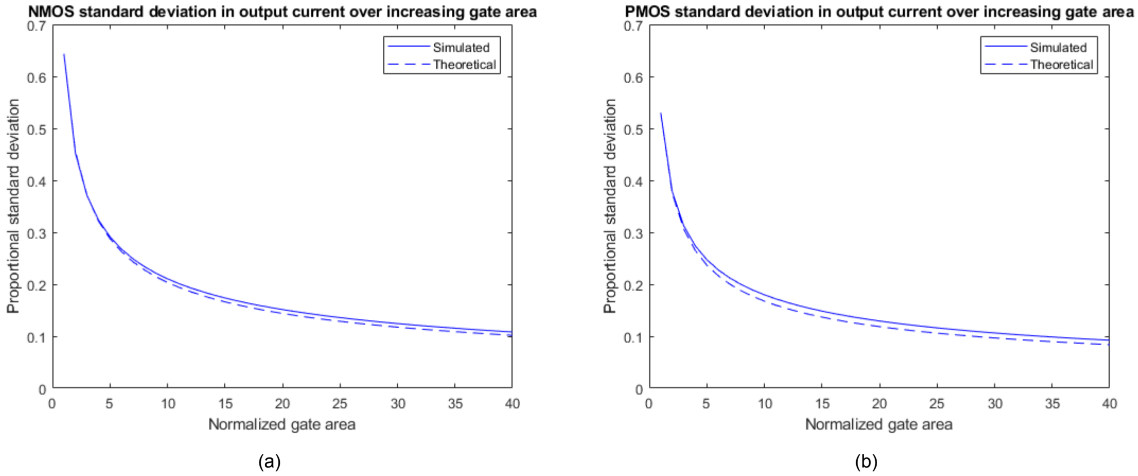


Figure B.1: Proportional standard deviation in output current against transistor gate area (with 200 Monte Carlo runs), with both simulated (via Cadence Virtuoso) and theoretical values for (a) NMOS devices, and (b) PMOS devices.

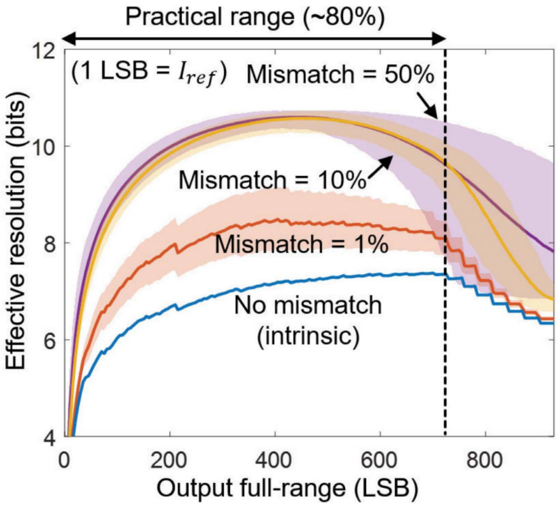
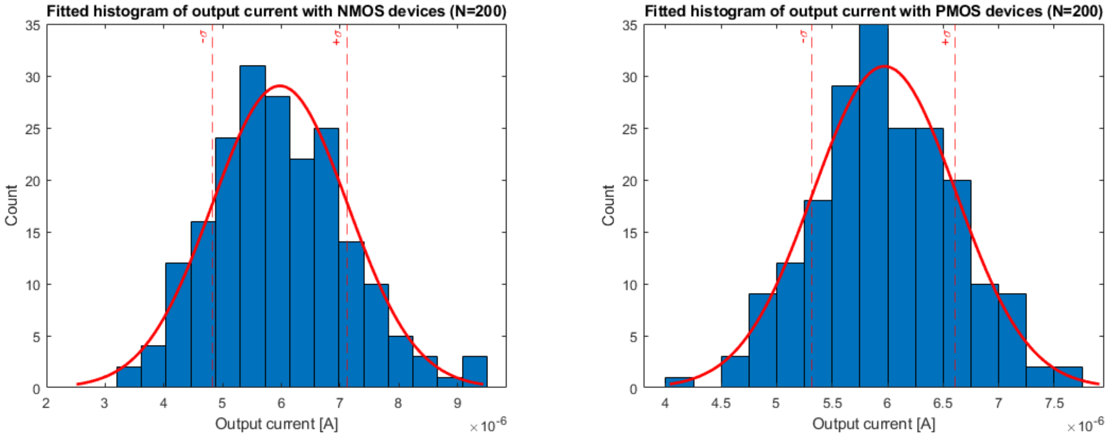


Figure B.2: Simulated resolution of a standard RXF set-up with varying mismatch, with the shaded regions indicating the 90% confidence interval. At higher output LSBs the confidence interval widens for higher mismatch, without the benefit of achieving higher effective resolution mid-range. Taken from [21].

Coming back to the posed optimal mismatch of 0.1 to 0.2, in the technology used this would require transistors that are about 10x the size of a minimum size transistor.



(a)

(b)

Figure B.3: Output current distribution due to mismatch and fitting for (a) NMOS devices, and (b) PMOS devices.

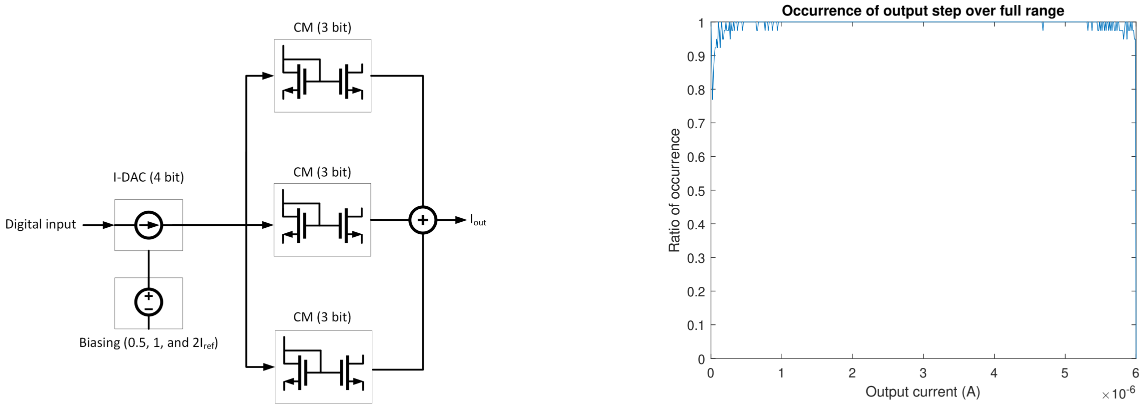
C

RXF Testing Designs

In order to be able to determine which RXF topologies were the most suitable for this project, many different topologies were tested. The topologies with major, but in the end unused changes not shown in the main work are shown in section C.1. Furthermore, more detailed results of the topologies shown in the main work are presented in section C.2.

C.1. Tested topologies

The topology with 4b I-DAC and three 3b current mirrors was also tested with triple biasing. This biasing resulted in an I-DAC LSB of not only the normal 80 nA, but also half and double. This was attempted so as to increase the RoO at higher output currents, similar to how the double-biasing topology boosted RoO at lower output currents. The topology and its corresponding RoO is shown in Fig. C.1a and C.1b, respectively. Ultimately, this topology was discarded as a possibility due to still not having the desired RoO at low output currents. While using a double-bias with $1I_{LSB}$ and $2I_{LSB}$ could have been an option to increase RoO at high output currents only, the decision was made to not do so in this way as the results are very similar to the pseudo-segmented topology, but with one less bit necessary in the calibrated LUT.

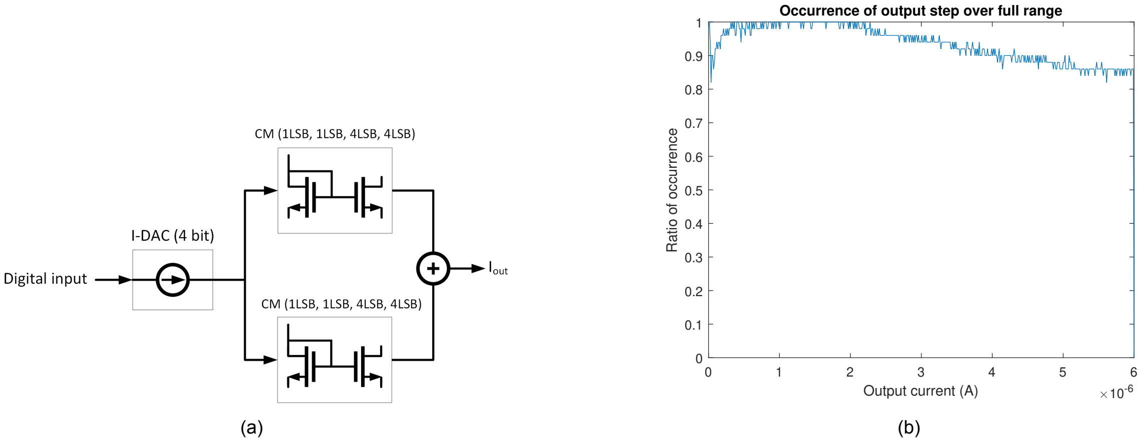


(a)

(b)

Figure C.1: (a) Topology with a triple-biased I-DAC, and (b) its corresponding RoO.

Another topology tested was with single-biasing, but with pseudo-segmented current mirrors instead of binary current mirrors. The idea behind this topology was to simply create extra redundancy at the tails of the output range by including more LSB and MSB branches. The resulting topology was with a 4b IDAC, and two current mirrors, each with gains of: 1LSB, 1LSB, 4LSB, 4LSB. This topology and its resulting RoO is shown in Fig. C.2a and C.2b, respectively. The reason this topology was not chosen is evident: the RoO overall simply is not sufficient. The RoO at the two tails is increased compared to the first standard of 4b DAC and two 3b current mirrors, but not enough. Segmenting the current mirrors does not help as much as segmenting the IDAC, as the ”earlier” the segmentation takes place, the more of an effect it has on the positions of redundancy in the final output. This can be conceptually explained by the fact that any mismatch present in the IDAC will be amplified by the following current mirrors, resulting in a larger spread of final output, which in turn helps distribute the redundancy over a larger portion of the output range. For these reasons, it was decided to only segment the IDAC.

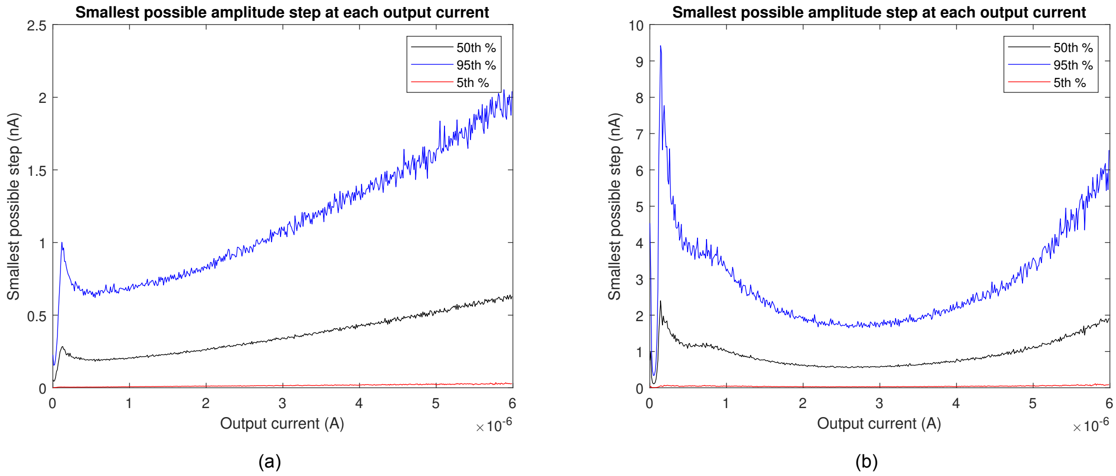


(a) (b)

Figure C.2: (a) Topology with pseudo-segmented current mirrors, and (b) its corresponding RoO.

C.2. Extra results of chosen topologies

Although the ratio of occurrence is a strong indicator for performance, there are other metrics that can help characterize the DAC. One such parameter is the smallest possible amplitude step size at each point of the full output range. This effectively indicates how densely populated the non-calibrated output is. Fig. C.3a and C.3b show this parameter at the 5th percentile, the 50th percentile (the median), and the 95th percentile for the RXF topology for the anodic and cathodic phase, respectively. The percentiles are calculated after taking the entire output of all MC runs as one big unit. One should note that even if the smallest possible amplitude step theoretically allows for a higher ”resolution” stimulator, it does not guarantee that any given MC run will produce outputs that have that amplitude step over the entire range.



(a) (b)

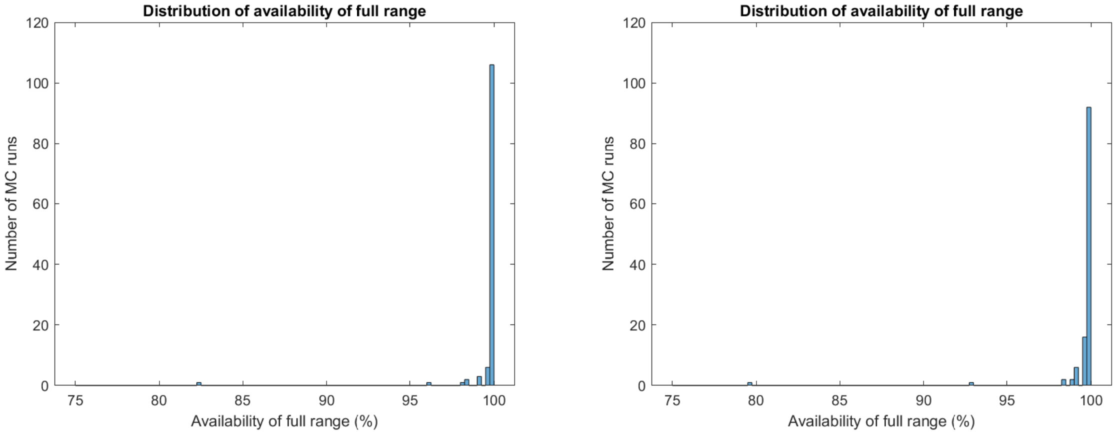
Figure C.3: Smallest possible amplitude step over full range of (a) the topology for the anodic phase, and (b) the topology for the cathodic phase.

Table C.1: Full range coverage of the stimulator.

	Full range coverage
Anodic phase	0.8833
Cathodic phase	0.7667

Another interesting parameter is how many of the MC runs have outputs available over the entire output range at the desired resolution of 8 bits. In other words, the ratio of full range coverage. The values for this can be found in Table C.1.

Since these values are not both 1.00, it is interesting to see what the distribution is of how much percent of the output range is covered by any given MC run. In other words, how much coverage there is of the full output range. Fig. C.4a and C.4b show this for the anodic and cathodic phase, respectively.



(a)

(b)

Figure C.4: Histogram showing the percentage of full range availability of 120 MC runs with (a) the topology for the anodic phase, and (b) the topology for the cathodic phase. Bin width for both histograms is 0.25%.

D

Oversampling the DAC

This chapter will go into more detail about oversampling and show a comparison of the performance with and without it.

D.1. Oversampling details

As briefly explained in the main body of this work, oversampling is a way to increase the control over the injected charge. This is beneficial since it allows for flexibility in stimulation time, without having to take an area penalty of the stimulator itself. As an example, suppose that (unrealistically) the output of the stimulator is [1; 2; 3; 4; 5] A. If this were to be oversampled by a factor 2, for instance to adjust the duration of each phase of the triphasic waveform from 10 μs to 20 μs, the following combinations would arise:

[[1,2]; [1,3]; [1,4]; [1,5]; [2,3]; [2,4]; [2,5]; [3,4]; [3,5]; [4,5]]

This is not including repeating the same output twice, which of course still is possible. With these combinations, the averaged output would be:

[1.5; 2; 2.5; 3; 2.5; 3; 3.5; 3.5; 4; 4.5]

When taking into account mismatch, the average of e.g. [1,5] and [2,4] would not be both exactly the same, indicating the importance of redundancy here again. When applying this to realistic output values, the RoO can be plotted with and without oversampling. Fig. D.1a and D.1b show the RoO without and with oversampling, respectively. It should be noted that the topology used is not the same as in the final design and the output current here goes up to 6.5 μA instead of 6 μA. Nevertheless, it functions as a good indication of how oversampling affects RoO.

D.2. Computational requirements

It is worth mentioning that applying and calculating all the possible combinations and their averaged output is quite computationally expensive. Especially when taking into account that the actual DAC topologies have 32768 and 8192 configurations for the anodic phase and cathodic phase, respectively, and that an oversampling factor (OSF) up to 5 can be expected. The approximate amount of possible combinations is

$$\text{Combinations} = \text{Configurations}^{\text{OSF}} \tag{D.1}$$

In case of an OSF of 5, this leads to a staggering $3.8 * 10^{22}$ combinations for the anodic phase, and $3.7 * 10^{19}$ combinations for the cathodic phase. Even though this only would have to be done once (or very infrequently at least), this is quite a computationally expensive task. Adjustments could be made to only allow combinations of configurations that are within e.g. 20 % of each other in output. This would significantly reduce the amount of possible combinations, while theoretically not affecting RoO

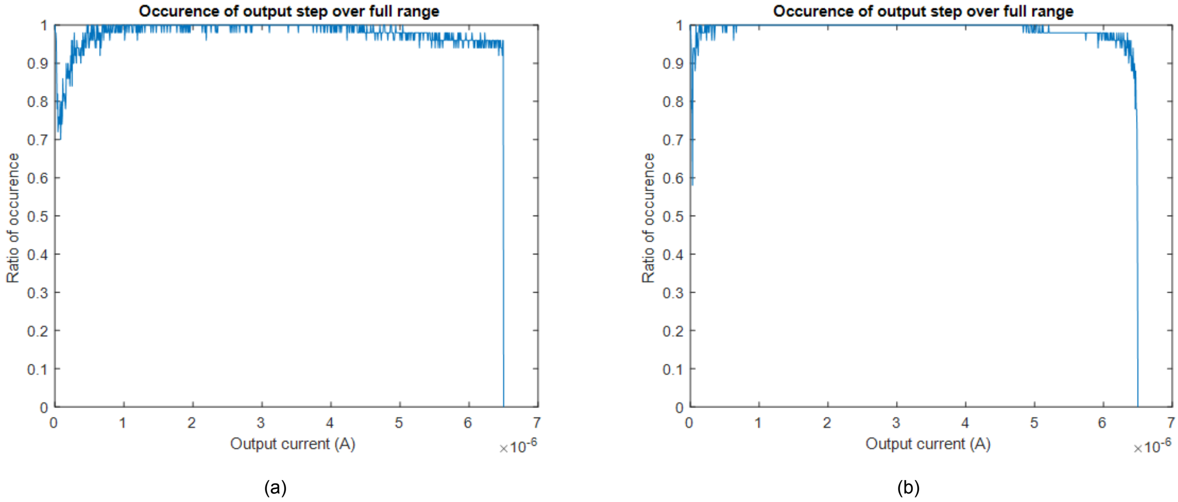


Figure D.1: Example of ratio of occurrence when scaling output bins to 9 bits when (a) not applying oversampling, and (b) oversampling by two. Note that the topology used for both these cases are not the one used in the final design, and are only used to visualize the result of the concept.

that much. Unfortunately, optimizing the oversampling algorithm is outside the scope of this thesis, and is thus not included here.

E

Noise

The mechanism of action for stimulation, as mentioned in Section 1.1.4, is the injected charge and its corresponding electrical field. For the cathodic working phase, this means that the total injected charge is the important parameter that needs to be accurate. For the anodic corrective phases, it is also the total injected charge that determines the residual artifact duration, as the residual voltage (Fig. 1.3) is determined by the residual charge.

Due to this, any noise that has a frequency above the period of the stimulation pulse can be disregarded, as it will average out to zero, and thus also have no effect on the total injected charge. To confirm this, simulations that included noise up to a certain frequency were performed, and the total injected charge was derived for each of these simulations. This charge was then compared to the total injected charge without any noise present. The delta is then

$$\Delta Q_f = Q_{noise,f} - Q_{base} \tag{E.1}$$

where Q_{noise_f} is the total injected charge in presence of noise up to frequency f, and Q_{base} is the total injected charge without any noise. This is then performed multiple times, and the RMS of ΔQ_f is calculated. Tab. E.1 and E.2 show the resulting $Q_{diff,RMS}$ for 25 runs per frequency, at 1 μs and 10 μs pulse widths. As expected, the longer the pulse width, the lower the cut-off frequency is.

$F_{max,noise}$	$Q_{diff,RMS}$
10 kHz	0.296 fC
100 kHz	1.716 fC
1 MHz	7.088 fC
5 MHz	8.085 fC
10 MHz	9.378 fC
100 MHz	8.119 fC
1 GHz	8.669 fC

Table E.1: Total injected charge difference for 1 μs pulses.

$F_{max,noise}$	$Q_{diff,RMS}$
10 kHz	9.48 fC
100 kHz	50.8 fC
1 MHz	58.34 fC
10 MHz	54.7 fC
100 MHz	64.63 fC
1 GHz	60.63 fC

Table E.2: Total injected charge difference for 10 μs pulses.

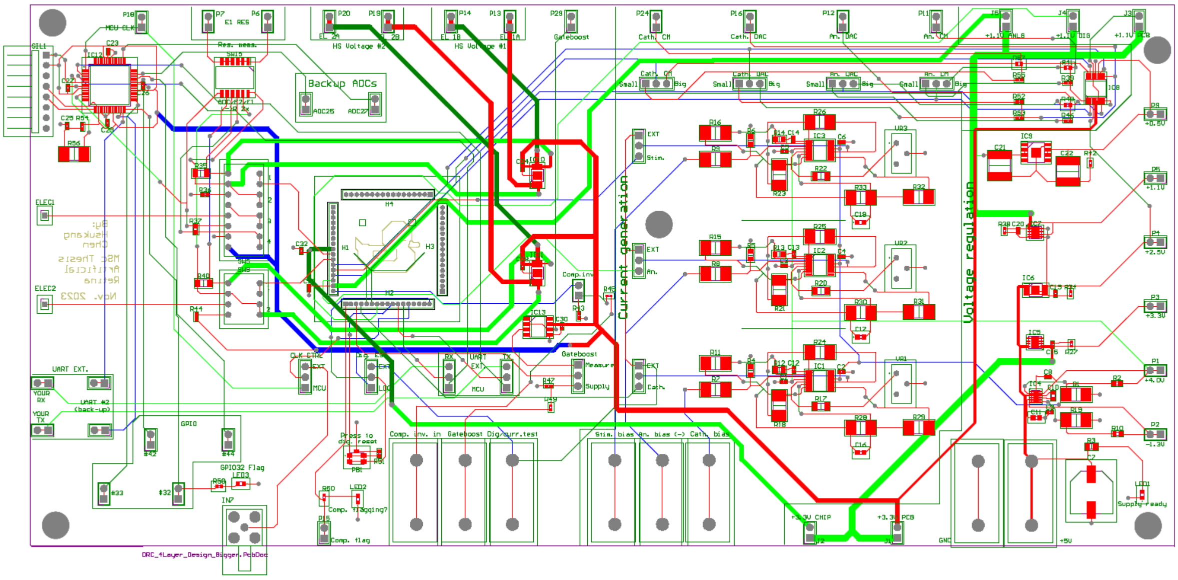
Since the anodic phase has a minimum step duration of 1 μs, and the cathodic phase has a minimum step duration of 10 μs, the relevant noise cut-off frequencies are 1 MHz and 100 kHz respectively.

F

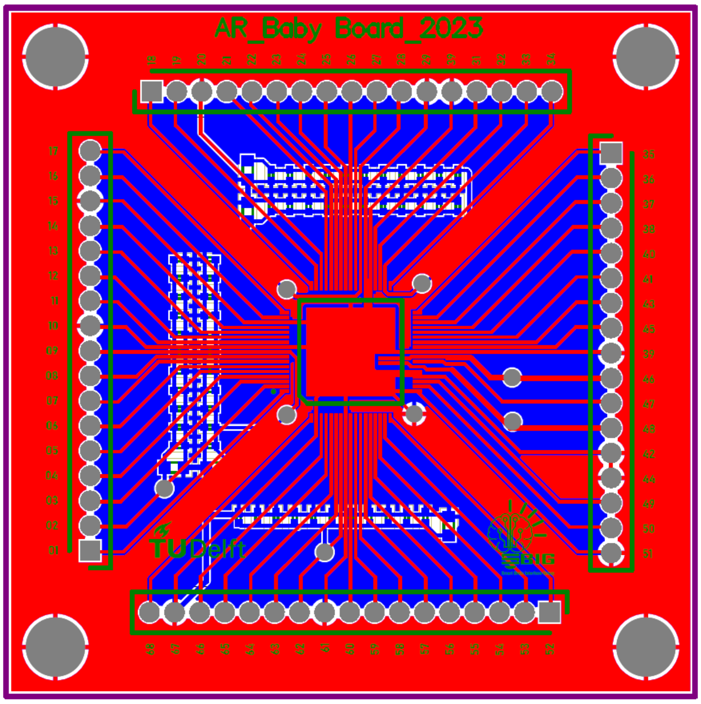
PCB design

F.1. PCB view

F.1.1. Main PCB



F.1.2. Daughterboard PCB



G

UART serial communication

The communication between MCU and the chip is done via serial UART. This Appendix will elaborate on the protocol used and on the registers on chip.

The used UART protocol operates at a baud rate of 115200, has no parity bit and uses a single stop bit. Each packet written from MCU to the chip consists out of five bytes (Fig. G.1). Byte #1 and #2 are preamble bytes, byte #3 is used to specify write (WR = 0) or read (WR = 1) from the register and the corresponding register address. Byte #4 and #5 contain the data that needs to be written to the register.

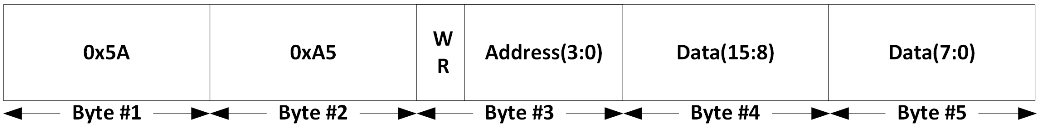


Figure G.1: Packet format to write from MCU to the chip.

This protocol is shared with another design, so only the relevant registers will be discussed here. There are two types of registers important: (1) the static register, and (2) the dynamic registers. The static registers are updated once and remain constant until updated again. The dynamic registers contain the control for the demultiplexer, the reset switch of the voltage compliance monitor and the input code for the AWG.

Tab. G.1 shows the static registers, their addresses, and their data formats. Writing ”00” to the ”Read back Register” requests the chip to write to the MCU with the current content of a specific register. The CLK periods are based off of the 40 MHz clock, e.g. a stimulation period of 40 would lead to a stimulation frequency of 1 MHz.

Table G.1: Addresses and data format for relevant static registers.

Description	Address	Data
Dynamic Register	0	Data(0)
Read back Register	0	Data(2:1)
CLK comparator period	1	Data(15:0)
CLK comparator on-time	2	Data(7:0)
CLK S/H on-time	2	Data(15:8)
CLK stimulation period	3	Data(15:0)
Comparator config.	4	Data(7:0)

Tab. G.2 contains the dynamic register addresses and their data format. To write to these registers, the static register ”Dynamic Register” needs to be set to ”0”, and then the settings for the dynamic register are to be sent 60 times. After this, a command can be sent to either execute the dynamic register only once and then stop, or to continuously execute its content.

Table G.2: Addresses and data format for the dynamic registers.

Description	**Address**	**Data**
Anode Input Code	8	Data(14:0)
Reset switch of monitor	8	Data(15)
Cathode Input Code	9	Data(12:0)
Demultiplexer Control	9	Data(15:13)