



University
of Glasgow

Murdoch, Derek J. (2011) *Data analysis of retinal recordings from multi-electrode arrays under in situ electrical stimulation*. PhD thesis.

<http://theses.gla.ac.uk/2997/>

Copyright and moral rights for this thesis are retained by the Author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Data analysis of retinal recordings
from multi-electrode arrays under *in*
situ electrical stimulation

Derek Murdoch



University
of Glasgow

Department of Physics and Astronomy

College of Science and Engineering

University of Glasgow

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

April 2011

©D. Murdoch April 2011

Abstract

The development of retinal implants has become an important field of study in recent years, with increasing numbers of people falling victim to legal or physical blindness as a result of retinal damage. Important weaknesses in current retinal implants include a lack of the resolution necessary to give a patient a viable level of visual acuity, question marks over the amount of power and energy required to deliver adequate stimulation, and the removal of eye movements from the analysis of the visual scene. This thesis documents investigations by the author into a new CMOS stimulation and imaging chip with the potential to overcome these difficulties. An overview is given of the testing and characterisation of the components incorporated in the device to mimic the normal functioning of the human retina. Its application to *in situ* experimental studies of frog retina is also described, as well as how the data gathered from these experiments enables the optimisation of the geometry of the electrode array through which the device will interface with the retina. Such optimisation is important as the deposit of excess electrical charge and energy can lead to detrimental medical side effects. Avoidance of such side effects is crucial to the realisation of the next generation of retinal implants.

Declaration

Except where explicit reference is made to the work of others, this thesis is the result of my own work. None of this material has been submitted for any other degree at the University of Glasgow or any other institution.

Contents

1	Introduction	2
2	Background	12
2.1	Human visual system	12
2.1.1	Eye	12
2.1.2	Visual pathway	16
2.2	Mechanisms and physiology of the retina	18
2.3	Lateral inhibition in the retina	24
2.4	Diseases of the retina	25
2.5	Strategies for restoration of sight	28
2.6	Development of retinal implants	33
2.7	Group strategy	39
3	Multi-electrode arrays	43
3.1	Theory of key nanofabrication processes	43
3.1.1	Optical lithography [63]	43
3.1.2	Electron beam evaporation [64]	45
3.1.3	Etching techniques and liftoff	45
3.1.4	Polyimide substrate	47
3.2	Mask design and manufacture	47
3.3	Fabrication	52
3.4	Electrical testing and characterisation	58
4	CMOS imaging and stimulation chip	63
4.1	Introduction	63

4.2	Overview of IPIX II components	66
4.2.1	Pixel array	66
4.2.2	Voltage controlled oscillator (VCO)	68
4.2.3	Neural network	72
4.2.4	Biphasic output	76
4.2.5	Electrode interface with retina	78
4.3	Testing and results	85
4.3.1	Testing environment	85
4.3.2	Neural network linearity	87
4.3.3	VCO behaviour under test signal	93
4.3.4	Biphasic output	94
4.3.5	Comparison of electrode-electrolyte interface model with experiment	98
4.3.6	Extensions of model	99
5	Retinal experiments	108
5.1	Experimental setup	108
5.2	Java runtime environment and data selection	110
5.3	Analysis process and software	113
5.3.1	C/Bash analysis	114
5.3.2	Artifact subtraction	117
5.3.3	Scanning and spike detection	123
5.3.4	Histogramming and construction of strength-duration curve	124
5.3.5	Principal component analysis	127
5.3.6	Clustering	129

5.4	Testing and proof-of-concept analysis	130
5.4.1	Testing of artifact subtraction	130
5.4.2	Spike processing and proof-of-concept analysis	131
5.4.3	MATLAB analysis	136
5.4.4	Increase of critical mass of data	142
6	Final analysis	144
6.1	Analysis framework	144
6.2	Analysis and results	145
6.2.1	Experimental apparatus testing prior to experiment . .	145
6.2.2	Spike analysis and construction of strength-duration curve	148
6.2.3	Determination of rheobase and chronaxie	156
6.3	Discussion and implications for future work	160
7	Conclusions	164

List of Figures

1	Example of a prosthetic toe from the Egyptian Museum in Cairo. (Image taken from Ref. [4])	4
2	Schematic cross-section of the retina showing implantation sites for the two main categories of implant. (Image taken from Ref. [11])	5

3	Example of a strength-duration curve showing (1) the rheobase current, (2) the doubling of the rheobase, and (3) the chronaxie as the pulse duration corresponding to twice the rheobase current. Image taken from Ref. [12].	9
4	Sagittal horizontal section of the human eye. (Image taken from Ref. [25])	13
5	Schematic cross-section of the neural structure of the retina, showing the various nuclear and plexiform layers. (Image taken from Ref. [14])	15
6	High magnification EM images of cone and rod outer segments in the ground squirrel retina. (Image taken from Ref. [25])	18
7	Diagram of the organisation of centre-surround responses using horizontal cell circuitry to provide the antagonistic surround. (Image taken from Ref. [25])	22
8	Three types of horizontal cells in human retina (Image taken from Ref. [25])	23
9	Bipolar cell types in human retina (Image taken from Ref. [25]). In addition to those listed in the legend, also shown are flat (FMB) and invaginating (IMB) cone types.	24

10	Normal macula of an elderly patient. The asterisk represents the location of the fovea, which lies directly in the visual axis. The macula (boxed area), which is adapted for high acuity vision, is located temporal to the optic nerve (arrow). It is approximately 6mm in diameter and centered on the fovea. The vascular arcades are indicated by arrowheads. (Image taken from Ref. [25])	27
11	Colour fundus photograph from an individual with dry AMD. Numerous small and intermediate-sized drusen are visible in the macular region (oval). (Image taken from Ref. [25])	28
12	Diagram showing the basic operational principles of the Boston retinal implant. Image taken from Ref. [11].	35
13	A photograph of the prototype subretinal implant developed by the German SUB-RET project. Image source: www.retina-implant.de	37
14	Close-up image of the electrode sites of the 8-electrode array, as viewed from within <i>Tanner L-Edit</i>	48
15	Close-up image of the electrode sites of the 16-electrode array, as viewed from within <i>Tanner L-Edit</i>	48
16	Image of the wafer design showing the arrangements of the 8-electrode arrays (bottom row) and the 16-electrode arrays (top row), as viewed from within <i>Tanner L-Edit</i>	49
17	Close-up image of the electrode sites of the single-electrode stimulation array, as viewed from within <i>Tanner L-Edit</i>	50

18	Image of the wafer design showing the arrangement of the copies of the focussed stimulation array, as viewed from within <i>Tanner L-Edit</i>	51
19	Close-up image of the electrode sites of the correlated stimulation array, as viewed from within <i>Tanner L-Edit</i> . The larger circles correspond to the larger stimulation electrodes.	51
20	Graphical representation of the 4" silicon wafer prior to commencement of the fabrication process.	54
21	Graphical representation of the 4" silicon wafer after completion of the polyimide application process.	54
22	Graphical representation of the 4" silicon wafer after completion of the metal deposition process.	55
23	Graphical representation of the 4" silicon wafer after completion of the photoresist application process.	55
24	Graphical representation of the 4" silicon wafer after completion of the exposure and development processes.	55
25	Graphical representation of the 4" silicon wafer after completion of the wet etching process.	55
26	Graphical representation of the 4" silicon wafer after completion of the reactive ion etching process.	56
27	Graphical representation of the 4" silicon wafer after completion of the flood exposure process and application of the passivation layer.	56
28	Graphical representation of the 4" silicon wafer after completion of the via etching process.	56

29	Graphical representation of the 4" silicon wafer after completion of the electroplating process.	56
30	16-electrode array at 20 × magnification. Each wire track leads out to an individual bondpad (not shown).	59
31	16-electrode array at 50 × magnification, showing etch detail and via holes.	59
32	16-electrode array at 100 × magnification, showing measurements of electrode radii.	60
33	8-electrode array at 100 × magnification, showing measurements of electrode radii and wire track widths.	60
34	A photograph of the completed wafer (courtesy of I. MacVicar), showing the arrangement of the five copies of the focussed stimulation array.	60
35	Plot of the mean impedance with respect to frequency for twenty 5 μm-diameter electrodes.	61
36	Plot of the mean phase angle with respect to frequency for the same selection of electrodes.	62
37	Photograph of the architecture of the IPIX II chip, clearly showing the central array (hexagon), level hold/shift and VCO (inner square) and output driver circuitry (outer square). The array is 7211 × 7077 μm ² in size, with the pixels spaced at intervals of 107.5 μm. The bond pads are 50 × 180 μm ² and have a pitch of 230 μm.	64
38	Block diagram of the IPIX II chip, giving a more detailed overview of the components seen in the photograph.	66

39	Block diagram of an individual channel in the central array, showing the key components.	67
40	Diagram of the photosensor structure within a given pixel. . .	67
41	Diagram of the signals transmitted to the chip during photosensor operation.	69
42	Diagram of the signals transmitted to the VCO and timer circuitry during operation.	70
43	Diagram of the VCO circuit.	72
44	Diagram showing how the principal data signals are snaked through the array. The order of the bits transmitted to each pixel varies depending on the row addressed. The white squares denote photosensors, which are placed on alternate sides of pixels to produce a hexagonal-close-packed arrangement for greater resolution.	73
45	Diagram of the “configuration register load” signals transmitted to the chip prior to the commencement of operation. . . .	76
46	Diagram of a typical biphasic pulse as generated by the biphasic output circuit of the VCO.	78
47	Schematic representation of the simulated circuit applied by Borkholder [73] and Mathieson et al. [79] in recent studies. C_i denotes the interfacial capacitance, R_t the charge transfer resistance, C_w and R_w the capacitance and resistance associated with the Warburg impedance and R_s the spreading resistance. (Image taken from Ref. [79])	79

48	Diagram of the double layer capacitance created across the interface when an electrode is placed in an electrolyte solution.	80
49	Table of the key parameters in the PSPICE model with respect to electrode diameter.	85
50	The relationships of the impedances modelled in PSPICE to the stimulation frequency for a selection of electrode diameters.	85
51	Diagram of the LabVIEW VI used in the testing and characterisation of the IPIX II chip, showing all the key controls and readouts.	88
52	<i>Top left:</i> The behaviour of the VCO frequency vs. light intensity for selected combinations of neural network weights. <i>Top right:</i> Behaviour of the minimum frequency of the VCO vs. neural network weight for the selected combinations. <i>Bottom right:</i> Behaviour of the gradient of VCO frequency with respect to light intensity vs. neural network weight for the selected combinations. <i>Bottom left:</i> Graph of the correlation coefficient for the linear fits performed on the data, verifying reliability of fit.	91

53	<i>Top left:</i> The behaviour of the VCO frequency vs. light intensity for selected combinations of neural network weights in <i>C3Sel</i> mode. <i>Top right:</i> Behaviour of the minimum frequency of the VCO vs. neural network weight for the selected combinations. <i>Bottom right:</i> Behaviour of the gradient of VCO frequency with respect to light intensity vs. neural network weight for the selected combinations. <i>Bottom left:</i> Graph of the correlation coefficient for the linear fits performed on the data, verifying reliability of fit.	92
54	Table of the y-intercepts and gradients for linear behaviour of VCO frequency for selected neural network weights.	93
55	Graph of the behaviour of the output pulse frequency of the VCO with respect to test signal voltage.	94
56	Table of the ranges over which the current is expected to behave linearly in the two modes of the IPIX II.	96
57	Behaviour of the output current in <i>LowI</i> mode with respect to both positive and negative amplitudes.	97
58	Behaviour of the output current in <i>HighI</i> mode with respect to both positive and negative amplitudes.	98
59	Behaviour of the output current in <i>LowI</i> mode with respect to both positive and negative amplitudes, showing linear behaviour of the current in the specified range and the minimal charge imbalance.	99

60	Behaviour of the output current in <i>HighI</i> mode with respect to both positive and negative amplitudes, showing linear behaviour of the current in the specified range and the minimal charge imbalance.	100
61	Graph of a typical IPIX II stimulus waveform, with <i>ZeroSel</i> mode disabled (red trace). Note that the amplitude is not zeroed during the interphase delay as with <i>ZeroSel</i> mode enabled (black trace).	101
62	Graph of a typical IPIX II stimulus waveform, with <i>ZeroSel</i> mode enabled. Note that the amplitude is now zeroed during the interphase delay.	102
63	Comparison of modelled impedances with those of 7 μm -diameter electrodes from a different study.	103
64	Comparison of mean impedance recorded for a 16-electrode array with electrodes of 5 μm diameter.	104
65	The relationship of the electrode output voltage modelled in PSPICE to that of a real electrode.	105
66	The modified model circuit showing the new components. R_t , C_i and r_s are defined as before, R_{gen} and C_{gen} are the components associated with the pulse generator, R_{probe} represents the resistance of the probe, and R_{ret} represents the resistance of the retina.	105
67	Plot of the simulated variation of the input pulse voltage with respect to electrode diameter.	106

68	Plot of the simulated variation of charge imbalance with respect to frequency for selected electrode diameters.	107
69	Screenshot of the Java viewer showing a typical data set. . . .	111
70	Screen capture of the Java environment designed by A. R. Moodie for <i>in vivo</i> experiments in the IBLS.	112
71	Flow diagram showing the key steps of the data analysis process.	114
72	Example of spike well isolated by the averaging method compared to the interpolation method.	117
73	Example of spike poorly isolated by the averaging method compared to the interpolation method.	118
74	Graph showing a raw data set (black trace), the poorly interpolated artifact (red trace) and its consequent impact on the final data set (green trace).	121
75	Graph showing the refined interpolation of a data set (black trace), with the modelled artifact (red trace) adhering more closely to the raw data. Points in the final data set (green trace) prior to cessation of anodic saturation have been assigned zero values.	122
76	Typical profile of a spontaneous spike as seen in recordings of retinal signalling.	124
77	Example of a strength-duration curve showing (1) the rheobase current, (2) the doubling of the rheobase, and (3) the chronaxie as the pulse duration corresponding to twice the rheobase current. Image taken from Ref. [12].	126

78	Example of a 3D scatter plot showing the first three principal components of the data set classified into three clusters.	128
79	Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 0.5 ms stimulation at 100 mV. The trough at 75 ms is a noise event and has been correctly removed.	131
80	Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 0.5 ms stimulation at 250 mV.	132
81	Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 10 ms stimulation at 1000 mV.	133
82	Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 10 ms stimulation at 100 mV.	134
83	Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 0.1 ms stimulation at 20 mV.	135
84	Graph showing how the <i>#define</i> offset may be used to select the point at which to begin artifact interpolation on the data.	136

85	Histogram of spikes detected with pulses of strength 1600 mV.	137
86	Histogram of spikes detected with pulses of strength 1800 mV.	137
87	Histogram of spikes detected with pulses of duration 0.2 ms.	138
88	Histogram of spikes detected with pulses of duration 5 ms.	138
89	A PCA plot for the test data set.	139
90	A cluster plot for the data shown in Fig. 89, showing classification into four clusters.	140
91	Plot of the averaged spikes for each cluster of the data analysed in Figs. 89 and 90.	141
92	Example of the waveforms detected by the recording electrodes for a stimulus duration of 0.5 ms and strength of $0.1 \mu\text{A}$.	147
93	Example of the waveforms detected by the recording electrodes for a stimulus duration of 5 ms and strength of $10 \mu\text{A}$.	148
94	Gaussian fit and histogram of detected spikes for a pulse strength of $10 \mu\text{A}$ and duration of 1 ms. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.	149
95	Gaussian fit and histogram of detected spikes for a pulse strength of $20 \mu\text{A}$ and duration of 1 ms. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.	150
96	Histogram of detected spikes for a pulse strength of $20 \mu\text{A}$ and duration of 5 ms.	151
97	Separation of the spikes detected under a stimulus strength of $20 \mu\text{A}$ and duration of 5 ms into three clusters by means of PCA and k-medoid clustering analysis.	152

98	Separation of the spikes detected under a stimulus strength of $20 \mu\text{A}$ and duration of 5 ms into four clusters by means of PCA and k-medoid clustering analysis.	153
99	Cluster diagram of the principal components of the detected spikes shown in Fig. 97.	154
100	Cluster diagram of the principal components of the detected spikes shown in Fig. 98.	155
101	Gaussian fits and histograms of detected spikes for a pulse strength of $20 \mu\text{A}$ and duration of 5 ms, separated into three clusters by means of PCA and k-medoid clustering analysis. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.	156
102	Gaussian fits and histograms of detected spikes for a pulse strength of $20 \mu\text{A}$ and duration of 5 ms, separated into four clusters by means of PCA and k-medoid clustering analysis. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.	157
103	Graph of the numbers of spikes per trial generated on the eight most active electrodes under $10 \mu\text{A}$ stimulation.	158

1 Introduction

Blindness is a condition which continues to blight lives across the world. In 2002 the World Health Organisation [1] estimated that around 37 million people worldwide were blind, and another 124 million suffering from severe visual impairment. This can arise from a number of factors, including genetics or injury, however the vast majority of cases arise as a result of disease. Cataracts, which can be corrected by surgery, are responsible for nearly half of all cases of disease-related blindness. Much of the remainder of cases arise from degenerative conditions such as glaucoma, age-related macular degeneration (AMD) and diabetic retinopathy [1]. Rates for other diseases such as trachoma and onchocerciasis (more commonly known as “river blindness”) are higher than the world average in developing countries where communities may not have the same standard of sanitation. In the developed world, by contrast, incurable diseases such as age-related macular degeneration are more prevalent, although in some cases they can be managed effectively with treatments such as laser photocoagulation [2, 3]. However, these treatments only slow down the onset of blindness, and blindness brought on in this manner may only be avoidable by the use of some form of prosthesis.

The work undertaken in this thesis aims to make an important contribution to the development of a visual prosthesis, specifically a retinal prosthesis that bypasses those parts of the retina affected by diseases which affect that part of the eye. Such a prosthesis is thought to be the only way to restore sight to people affected by AMD, at a level of resolution that even closely approximates that of natural sight. It is unlikely, however, to be of any value to sufferers from other, somewhat rarer, degenerative conditions such as glau-

coma and diabetic retinopathy, as the more extensive damage inflicted by those conditions necessitates a prosthesis which acts on a more central stimulation site such as the optic nerve; however the prevalence of AMD (and the rarer, but similar condition retinitis pigmentosa) means that there will always be significant demand for such a specialised implant.

The field of prosthetics has been established since ancient times. The Egyptian Museum in Cairo contains an example of a wooden prosthetic toe, shown in Fig. 1, which would have been worn by an amputee in order to help maintain his balance while walking [4]. Nowadays, many types of prosthetic limbs and joints, and even artificial hearts and lungs, are available for amputees and sufferers from various forms of degenerative disease.

Later prosthetic research has focussed on the use of electricity to correct disabilities relating to the nervous system. The best known achievements in this field are the artificial heart pacemaker and the cochlear implant. The earliest realisation of the former can be traced to 1958 when one was successfully implanted in a patient in Sweden suffering from bradyarrhythmia, a disturbance of the rhythm of the heart which leads to an abnormally slow heart rate [5]. Although the pacemaker had to be replaced a substantial number of times over the course of the patient's lifetime, the concept was proven. Other developments followed, including RF control and power sources. Now the pacemaker is a mature technology which possesses the capability to mimic the heart's natural rhythm very closely [6].

The spur for the development of the cochlear implant came in 1957, when Djourno and Eyries placed an electrode on a deaf patient's auditory nerve



Figure 1: Example of a prosthetic toe from the Egyptian Museum in Cairo. (Image taken from Ref. [4])

which received its input by means of solenoidal coils. The patient reported being able to hear background noises [7]. Although this device did not function for very long, a second device was tested which aided deaf patients in lipreading by enabling them to recognise the rhythm of speech. Research in this field progressed rapidly and by 1972 a single electrode implant was commercially available. By 2005, development had reached a stage where the devices were totally implantable, without the need for external components [8].

The use of electricity for such therapeutic purposes is therefore now well established, and lends itself to the application of a retinal prosthesis. The main problem facing the numerous groups of scientists and engineers researching

such a device is how it is to be realised. No consensus has yet been reached on this question.

Current realisations of retinal prostheses fall into two categories, as shown

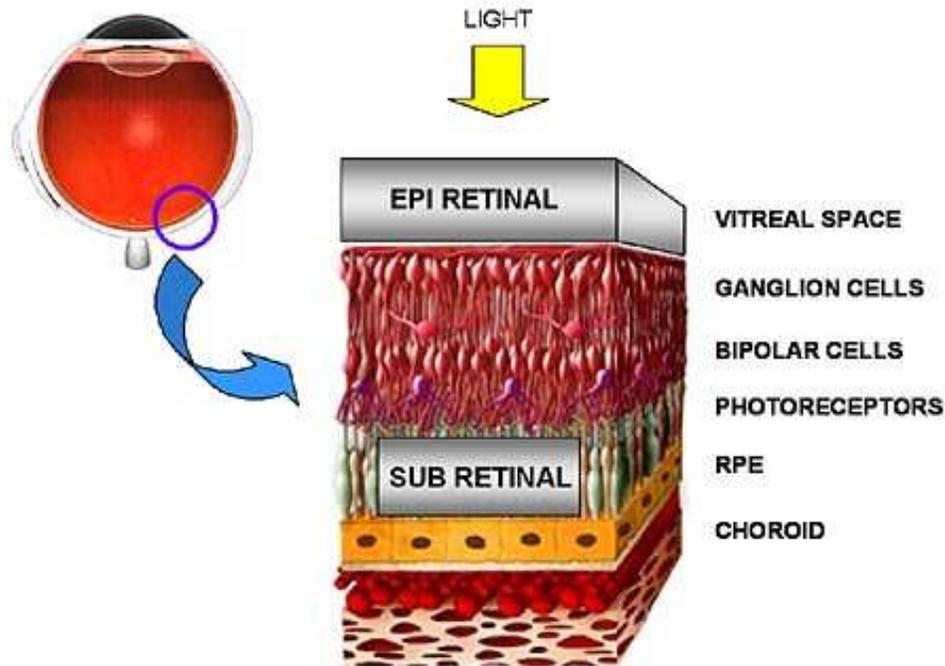


Figure 2: Schematic cross-section of the retina showing implantation sites for the two main categories of implant. (Image taken from Ref. [11])

in Fig. 2. The designs in each case are quite different. One common factor is that both are designed to treat a condition which predominantly destroys the rod and cone cells of the retina, hereafter referred to as photoreceptors, and consequently both take the place of these cells in some fashion. Subretinal prostheses are designed to function as direct replacements for the photoreceptors, and stimulate the retina through its existing neural structure. This is a more “organic” approach in the sense that it exploits the remaining healthy retinal tissue and largely avoids the use of complex electronics. However, the

requirement for such implants to be located in the subretinal space means highly invasive surgery is required for their chronic implantation [9]. In addition, the most prominent realisations of such a device require to make use of some form of amplification as incident light alone can generate only a fraction of the current necessary to stimulate the retina. Epiretinal implants are designed to sit on the inner surface of the retina and stimulate the ganglion cell layer, the layer of retinal cells which forms the point of departure for the optic nerve, directly. This avoids the more complex and invasive implantation procedures involved with subretinal implants, however, the added complication of replicating the behaviour of the intermediate components of the retina is introduced. Another disadvantage of this approach is that, in most realisations, eye movements are taken out of the equation, placing added pressure on the brain to adapt to the new forms of stimulus it is receiving from the prosthesis, as a result of dependence on an external camera to film visual input. While the brain is thought to be capable of such adaptations, it is desired that natural sight be restored to the greatest extent possible.

The most viable realisation of a retinal prosthesis will be one which combines the strengths of both approaches; the self-containment and accurate replication of retinal behaviour afforded by subretinal implants, and the ease of implantation and electronic sophistication of epiretinal implants. In order to realise such a device, a multi-disciplinary approach to the problem has been adopted, built around the Glasgow Retinal Group, which conducts its research and development work under the auspices of the Glasgow Experimental Particle Physics Group. The work undertaken at Glasgow on sensor development for particle physics experiments has enabled the Group to base

their proposed retinal implant on such technology. The sensors currently incorporated within the ATLAS detector at CERN [10] make use of silicon detectors which can provide high resolution tracking of microscopic interactions of particles at the various collision sites around the accelerator ring. Detection is accomplished by means of p-n junctions – a form of semiconductor detector of which photodiodes are a subset – which react to energy deposits from interaction products by forming electron-hole pairs which in turn produce a momentary pulse of current which, in the context of the experiments, is then detected as a signal by the built-in readout electronics and stored for data analysis. It is this operational principle which underpins the operation of the IPIX stimulation chip that forms the “front end” of the proposed device. Its analogue in the context of an implantable retinal prosthesis is the detection of incident photons as would be done by the photoreceptors of the healthy retina, followed by the processing of these visual stimuli by the subsequent components of its neural structure. The strength of this design is that it is able to incorporate much more intricate electronics than was previously possible. Where previous realisations of retinal implants have relied on external electronic processing, in the case of the most prominent realisations of epiretinal implants, or in some cases no such processing at all, the expertise gained from developments at CERN will enable such processing to be conducted on-chip, with particular circuits mimicing the functions of each layer of the retina. The most important of these is an on-chip neural network which also mimics certain functions of the horizontal cells, a group of cells which project between conduction pathways in the retina and subtract out steady shifts in cell membrane potential, thereby enabling continued function

under prolonged exposure to light.

The other component of the proposed device is an electrode array which delivers visual stimuli to the retina. This component of the device remains a work in progress. Earlier realisations of electrode arrays for implantation have tended to have a square geometry, however the current electrode array has a hexagonal-close-packed array of microelectrodes to better match the layout of the retinal cells. There remains, however, an open question over what the optimal size for the electrodes which make up the array should be. The work documented in this thesis aims to build on modelling work by Kovacs and Borkholder at Stanford University, in order to determine the optimal energy deposit into the retina, and the consequences of this for the electrode size. This is accomplished by the use of data from experiments on frog retina conducted within the Institute of Biomedical and Life Sciences (IBLS) at the University of Glasgow to build a strength-duration curve, as shown in Fig. 3, from which important parameters known as the rheobase and the chronaxie, which are discussed in more detail in the following chapter, may be determined. These in turn may be used to determine the threshold energy for stimulation of the retina and hence the optimal electrode size for delivery of such a level of energy.

Chapter 2 gives an overview of all the relevant background theory. It begins with a description of both the human eye and the wider visual system of which it is part. There then follows a more detailed description of the structure of the retina, and the role played by each of its principal layers in the processing of visual input. It is vitally important to understand these as the most viable realisation of a visual prosthesis will be one which replicates

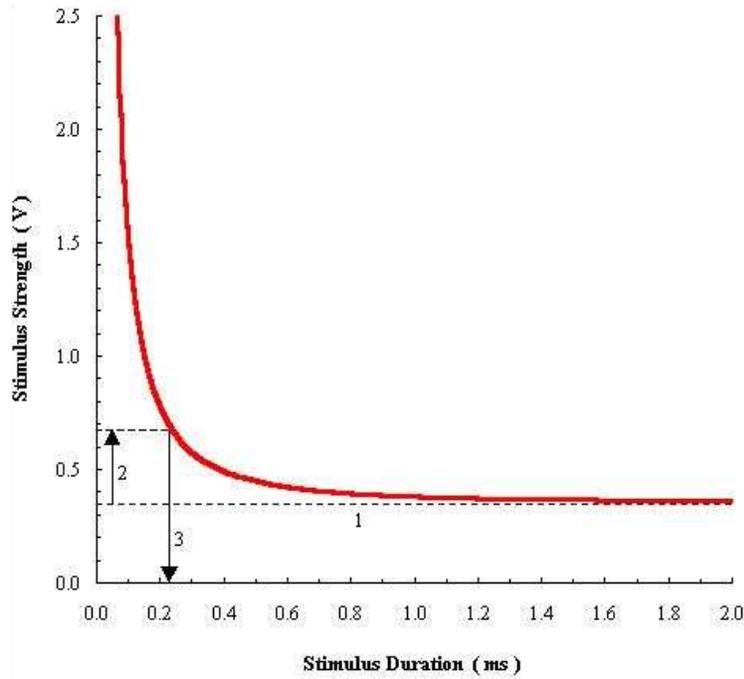


Figure 3: Example of a strength-duration curve showing (1) the rheobase current, (2) the doubling of the rheobase, and (3) the chronaxie as the pulse duration corresponding to twice the rheobase current. Image taken from Ref. [12].

them most closely. Lastly, overviews are presented in turn of the strategies pursued to date in the drive to restore sight to the blind, with emphasis not only on retinal implant development but that of cortical and optic nerve devices also; and of the strategy the Glasgow Retinal Group have undertaken to date in their efforts to develop a viable retinal implant.

Chapter 3 deals with the first component of the proposed retinal implant, the electrode array. In that context, it is to serve as the conduit through which electrical stimulation pulses generated by visual input are delivered to the appropriate position on the retina. The electrode arrays fabricated by

the author for the work documented in this thesis, however, have a different purpose – that of studying retinal signalling under *in situ* experimental conditions with a view to optimising the size of the electrodes that will eventually be included in the completed device. An overview is given of the designs to which they were fabricated, and the processes by which these were realised.

Chapter 4 covers the other component of the proposed retinal implant, the IPIX II stimulation chip. It is to form the “front end” of the device, and its function is to detect the incoming light and process it in a similar manner to that of the healthy retina. A description is given of the various components of the chip and how each of their functions mimics that of the corresponding layer of the retina. This is followed by descriptions of the testing and characterisation performed by the author on each of these components in turn and how the results of these tests compare with the specifications to which they were manufactured.

Retinal experiments are ongoing within the IBLS at the University of Glasgow, with whom the Retinal Group collaborate for these purposes. These, and the analyses of the data generated by them, are discussed in chapters 5 and 6. Chapter 5 falls into three clear sections. The first gives an overview of the experimental setup used in the IBLS, discussing in turn the experimental apparatus itself and the software governing its operation. This is followed by a description of the software written by the author specifically for the analysis of the large data sets generated by these experiments and the strategy undertaken for it. Finally, the results of a preliminary, proof-of-concept analysis, utilising and testing the software, are presented. In chapter 6 the results of the analysis performed by the author on an expanded data set are

presented and used to derive the principal result of this thesis, the optimal electrode diameter for the next generation of arrays to be included in the proposed implant. Discussion of the result and some of its consequences for the future of the project is also presented.

2 Background

This chapter gives an overview of the physiological background to the research undertaken in this thesis. In order to build a device which accurately simulates the operation of the retina, it is important to understand not only retinal processing but also how it contributes to the operation of the human visual system as a whole and the various degenerative medical conditions that affect it. The entire system is therefore described, from the entry of incident light into the eye, to its processing by the retina, to its transmission via the optic nerve to the visual cortex and its processing there. Also described is the manner in which the progress of the medical conditions by which the retina can be affected has inspired the development of various types of electronic visual prostheses. Particular emphasis is placed upon retinal prostheses as research in this area has informed much of the thrust of the work documented in this thesis.

2.1 Human visual system

2.1.1 Eye

The human visual system begins with the eye itself [13], a horizontal section of which is shown in Fig. 4. The outermost layers of the eyeball are made up of the sclera, cornea and choroid. Strands of the protein collagen contained in the cornea have a distinctive regular arrangement which renders it transparent. The sclera, on the other hand, is opaque.

The choroid is the layer inside the sclera, and is served by the most abundant blood supply per unit weight of any part of the human body. The ciliary

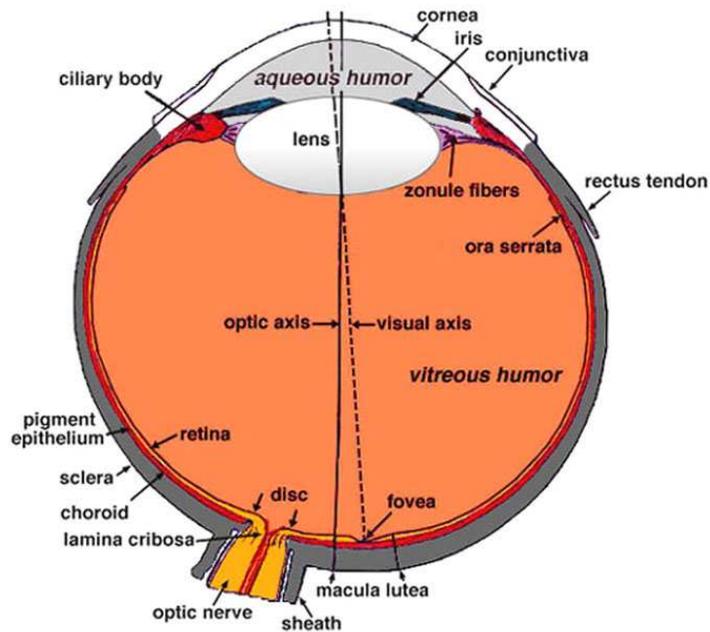


Figure 4: Sagittal horizontal section of the human eye. (Image taken from Ref. [25])

body, the ring of muscle to which the eye lens and the iris are attached, extends forward from here. The lens is biconvex, and has a similar protein structure to the cornea, albeit with the strands arranged concentrically in a similar fashion to the layers of an onion. The lens is made up of 66% water and forms the boundary between the anterior and posterior chambers of the eyeball.

The iris has two muscles which govern the size of the pupil: a circular one which induces constriction, and a radial one which induces dilation. Behind the iris, the epithelium of the ciliary body secretes the aqueous humour, a plasma-like fluid which drains through a network of canals at the junction of the cornea and sclera. In the posterior chamber of the eyeball, behind the

lens, is located the vitreous humour, which is a viscous gel composed of a network of collagen strands. This abuts directly onto the retina.

The retina is a layer of nervous tissue which lines the inner surface of this posterior chamber almost to the extent of the ciliary body, interrupted by the optic disc which serves as the point of departure for the axons of retinal ganglion cells which make up the optic nerve. The retina is structurally centred on a pit known as the fovea, located on the nasal side of the visual axis and characterised by a high density of retinal neurons displaced outward in a rim leaving only photoreceptors in the centre.

The eye is served by the blood supply from the carotid artery. Subsidiary to this are the retinal circulation drawn from the central artery of the optic nerve, and the choroidal circulation entering from the vicinity of the optic nerve and proceeding through a network of blood vessels radiating forward toward the ciliary body and iris. The outermost layers of the retina not served by the retinal circulation are served by this circulation. The cornea, aqueous humour, lens and vitreous humour are avascular (i.e. not served by blood circulation) and therefore depend on diffusion of nutrients and waste products to and from the choroid.

Eye movements are governed by three pairs of muscles inserted into the sclera from the orbit. The most important eye movements are those along the horizontal meridian. The visual field of the eye is theoretically 180° , although in practice it is limited by the nose, cheek bones, and eyebrows with only the temporal side unrestricted. There is a blind spot 5° in diameter at approximately 15° on the nasal side of the visual axis. A further property is

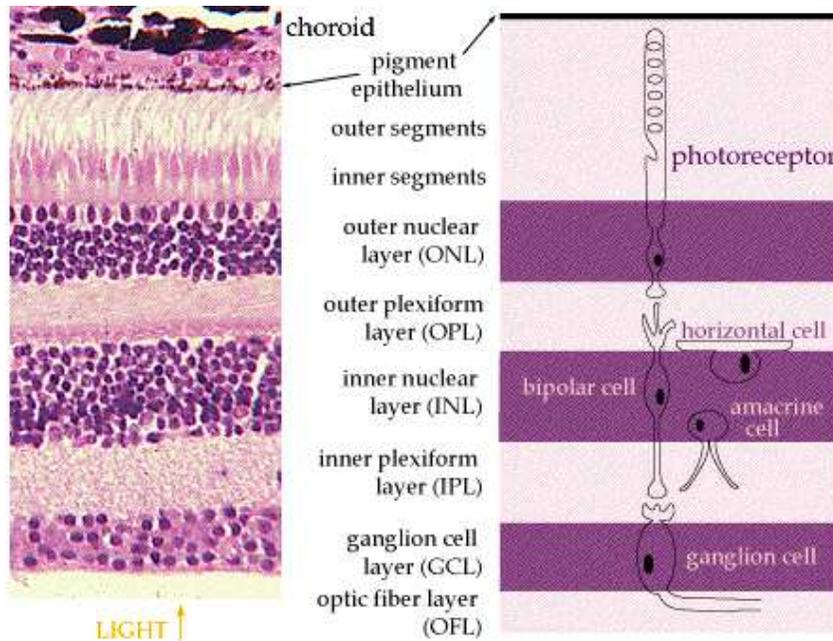


Figure 5: Schematic cross-section of the neural structure of the retina, showing the various nuclear and plexiform layers. (Image taken from Ref. [14])

that due to the nature of the optical projection of light, the image formed on the retina is inverted and reversed. 1° of the visual field corresponds to a distance of approximately $300 \mu\text{m}$ on the retinal surface of an average adult human eye.

Eye movements provide an important motor system as images must be focused on the fovea for proper viewing and also scanning is required to build up a proper picture of the visual scene. Eye movements fall into three main categories: saccadic movements, in which the eye moves to look at an object which appears in the periphery of the visual field; smooth pursuit movements, in which the eye moves to track an object moving across the visual field while keeping the head steady, with head movements required to keep

up with the object at higher velocities; and fixation movements which correct the position of the eyes to fix the image on the fovea.

In addition, the human eye possesses a property known as accommodation [16]. This enables an adjustment to occur in the refractive power of the lens when an object which is being examined is brought closer or moved further away. This declines with age as the lens loses its flexibility. Other important deficiencies which can occur in the human eye include spherical aberration, in which light rays passing through the periphery are refracted more sharply than those through the centre; astigmatism, in which a misshapen cornea causes rays of light through one meridian to be refracted more sharply, rendering it impossible to accommodate all the rays; and chromatic aberration, in which light of shorter wavelengths is refracted less than light of longer wavelength, leading to coloured blurring of vision.

2.1.2 Visual pathway

The visual pathway [17] begins with retinal ganglion cell axons combining in an orderly arrangement to form the optic nerves. The left and right optic nerves converge in such a way as to align the respective visual fields onto the appropriate side of the brain in preparation for stereoscopic vision. These pass onto the lateral geniculate nucleus (LGN), which serves as the main relay station for visual signals to the brain. The neurons in this area behave in a similar fashion to the ganglion cells of the retina. The next stage is the primary visual cortex which has similar mapping to the LGN and accounts for approximately half of the visual cortex. The prestriate cortex and higher cortical areas follow and have a more complex structure comprising various

visual areas.

In order for appropriate spatial form processing to take place, specifications of orientation, length and location of an object relative to other features of the image are required. These are encoded by particular neurons in the pre-striate visual cortex, the hierarchy of which is well understood [20]. Three dimensional interpretation of the visual scene is best achieved with binocular vision, however certain properties of an image can be used to make such interpretations monocularly. Absolute size can be used to judge distance; relative sizes can be used to judge separations of objects along the visual axis; shading and contours enable perception of the depth of the object in the visual field; perspective enables positioning of the object in three dimensions to be deduced from distortion of its known shape; and relative motion enables the distance of a moving object to be judged. Binocular vision allows instantaneous comparison of the different images seen by both eyes. When the eyes look at an object, images fall on the left and right halves of the fovea. The two streams are then combined in the visual cortex giving the perception of a single image [18].

The photoreceptors in the retina can be classified according to their response to colour. Young [19] classified cones into red, green and blue categories, corresponding to the three primary colours of light. Hering [21] classified retinal ganglion cells into three principal categories: red ON/green OFF, blue ON/yellow OFF, and black/white. These are now recognised as manifestations of different levels of the human visual system's response to colour. The red, green and blue cones exhibit responses at wavelengths which correspond to those colours' positions on the visible portion of the electromagnetic

spectrum. In the visual cortex the corresponding neurons are characterised by centres and surrounds of different colours. These are classified into the same categories as other cortical neurons according to the properties of their receptive fields. *Simple* cells respond to stimuli of particular length and orientation on a given part of the retinal surface, *complex* cells respond in addition to motion of such stimuli, while *hypercomplex* have the added property of terminating their responses once an edge becomes too long.

2.2 Mechanisms and physiology of the retina

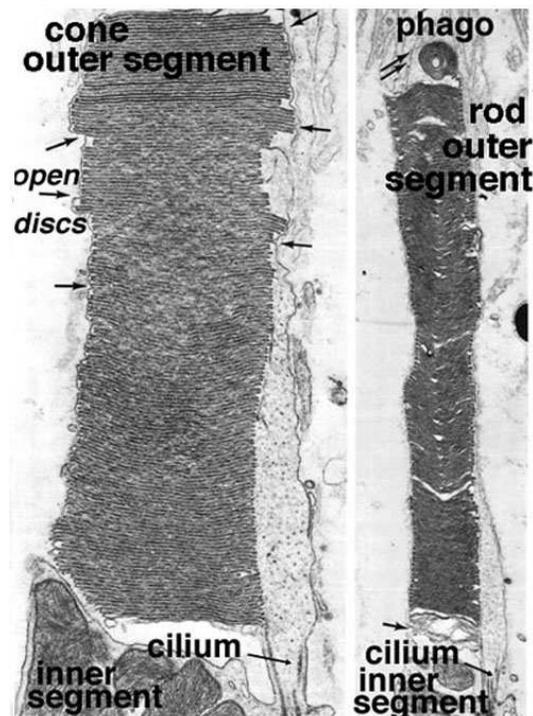


Figure 6: High magnification EM images of cone and rod outer segments in the ground squirrel retina. (Image taken from Ref. [25])

The retina [22] lines the choroidal surface of the eyeball and takes the

form of a complex neural network with multiple layers. Schematic cross-sections of these are shown in Fig. 5. The first (outermost) of these layers is the *pigment epithelium*, which consists of a network of hexagonal cells fused together to form a barrier through which nutrients and waste products diffuse to and from the choroidal circulatory system. The very outermost layer of this is known as *Bruch's membrane*. Connected to this are the *photoreceptor layer* and the *outer nuclear layer*. These contain the outer segments and nuclei respectively of the photoreceptor cells — the *rods* and *cones*. Rods are extremely sensitive to small numbers of photons and as a consequence have a key role to play in night vision. Their sensitivity is such that they saturate and deactivate under daylight conditions. Conversely, cones have a much higher activation threshold and only activate under daylight conditions. The next layers are the *outer plexiform layer* and *inner nuclear layer*, followed by the *inner plexiform layer* and *ganglion cell layer*. The inner nuclear layer contains the nuclei of the next major groups of cells, the *bipolar*, *horizontal*, and *amacrine cells*. The bipolar cells serve as the main conduction pathway between the outer nuclear layer and the ganglion cell layer, although some connect via the amacrine cells. The horizontal cells project between photoreceptors and oppose their polarisation state, allowing them to continue to respond to illumination after the initial onset. The ganglion cell layer contains the nuclei of the ganglion cells, of which there are many subsets; the two most abundant types are ON and OFF, which can be distinguished by their responses to illumination; ON ganglion cells increase their firing rate under illumination, while OFF ganglion cells increase their firing rate when the illumination is reduced. Physiologically, these cells connect to different

regions of the inner nuclear layer; the ON cells connect to the inner two-thirds while the OFF cells connect to the remaining outer portion.

The fovea is a pit in the retina 5° in diameter, located on the nasal side of the visual axis as noted above. It is characterised by a high density of retinal neurons, arranged such that the ganglion cells are displaced outward in a rim leaving only photoreceptors in the centre. This region is important to the functions of visual acuity and colour perception. The photoreceptors and higher retinal neurons are not evenly distributed. There are estimated to be around 5-7 million cones, 75-150 million rods and around 1 million ganglion cells contained in the retina. Cones have a peak density at the centre of the fovea, while rods are not present in the central 2° but reach a peak density at around 20° , gradually becoming more sparsely distributed at wider angles. Retinal ganglion cells and cones roughly mirror each other in their distribution, with around half of cones being located in the central 18° and half of retinal ganglion cells located in the central 13° .

The cell bodies of the photoreceptors contain the nuclei and sub-cellular structures - these form the inner segments. On the outer side, extending towards the pigment epithelium, cilia expand into tube-like structures, each one containing a stack of around a thousand discs. Photoreceptors renew themselves by creating new discs from infolding of their plasma membranes. The newest discs are those closest to the incoming light, while those furthest away have had substantial exposure to incident light. The older discs are shed at the outer end and ingested by the pigment epithelium. Complete turnover of discs in a photoreceptor typically takes between 9 and 12 days. In the case of rods, recovery processes take place during periods of inactivity in bright

daylight [24]. On the inner side the axon connects to the bipolar cell layer of the retina. Classification of photoreceptors is based on the shapes of their outer segments - rods have more tube-like outer segments than cones, though some cones in the central portion have more rod-like shapes to account for their close packing. Another important difference is that in rods the axons form single channels called spherules, whereas in cones they form three-way connections known as pedicles. Cone pedicles form three-way synapses called triads, where the central processing is performed by bipolar cells and the lateral processing is performed by horizontal cells. Rod spherules form similar connections however in this case there is more than one central process and there are several lateral processes.

Transduction of light into electrical signals occurs in the outer segments of the photoreceptors. 80% of the protein contained in rods consists of a visual pigment known as rhodopsin. The peak absorbance for this occurs at 498 nm, which is in the green portion of the visual spectrum. In cones the chemistry is similar, however in this case there are three different visual pigments, whose peak absorbances occur at 420 nm (blue), 534 nm (green), and 564 nm (yellow/green). Another important difference is that the effect of the incidence of single photons is greatly reduced [23].

There are three types of post-receptor neuron, as indicated above: horizontal cells, bipolar cells and amacrine cells. Horizontal cells possess only dendrites and connect across photoreceptor terminals to form a lateral pathway. They depolarise in darkness and hyperpolarise in light. The neurotransmitters they release have an inhibitory effect, and as a consequence of this they act to oppose the polarisation state of the photoreceptors to which

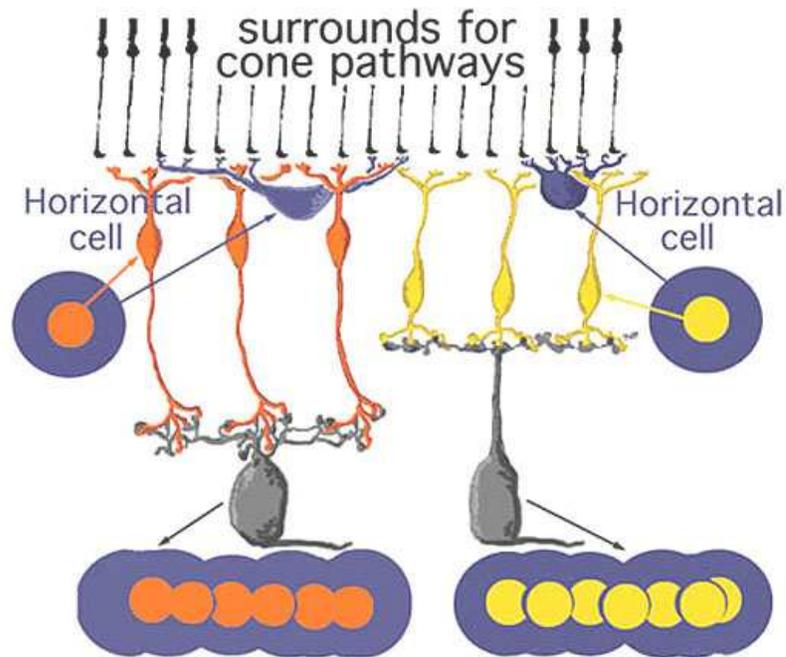


Figure 7: Diagram of the organisation of centre-surround responses using horizontal cell circuitry to provide the antagonistic surround. (Image taken from Ref. [25])

they connect, and subtract out steady shifts in their membrane potentials. This creates a mechanism of post-receptoral adaptation which enables the photoreceptors to continue to respond to illumination for a longer time than would otherwise be possible.

There are three types of bipolar cells. *Invaginating cone* bipolar cells form the central components of cone triads, and cause sign inversions in the signals they receive, as the effect of incident light is to cause a depolarisation followed by a hyperpolarising overshoot in the membrane potential on recovery. *Flat cone* bipolar cells hyperpolarise in response to light in the same fashion as cones. *Rod* bipolar cells connect to rods and also depolarise in response

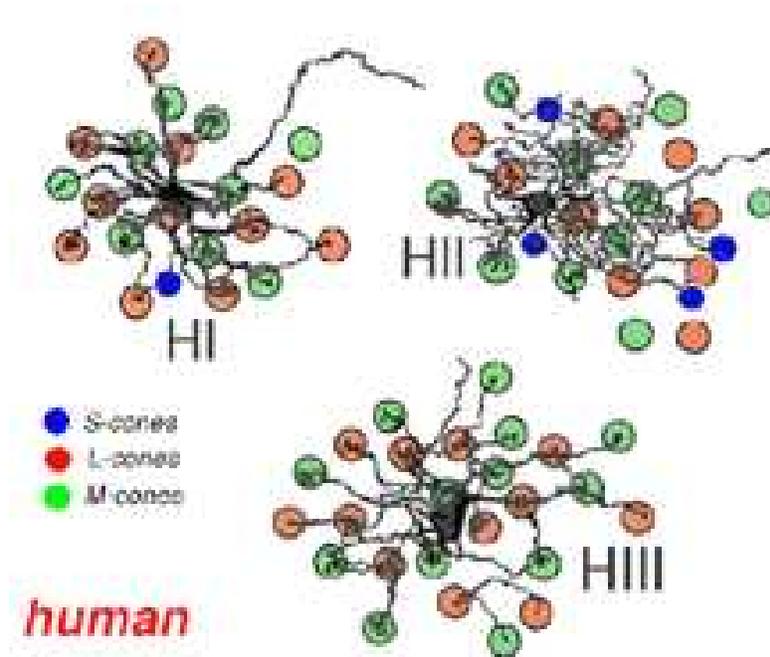


Figure 8: Three types of horizontal cells in human retina (Image taken from Ref. [25])

to light but have a substantially slower response time than flat cone cells.

The response of ganglion cells is similar to that of bipolar cells, except in this case it is an action potential that is generated. This can be viewed as an analogue-to-event conversion. Ganglion cells in the inner two-thirds of the inner plexiform layer are driven by invaginating cone bipolar cells, and increase their firing rate in response to illumination: these cells are therefore known as ON cells. Those in the outer third are driven by flat cone bipolar cells and increase their firing rate in response to termination of illumination: these are known as OFF cells. Rod bipolar cells connect to the ganglion cell layer indirectly through amacrine cells, which can be viewed as ganglion cells that have failed to develop a viable axon, and release inhibitory neurotrans-

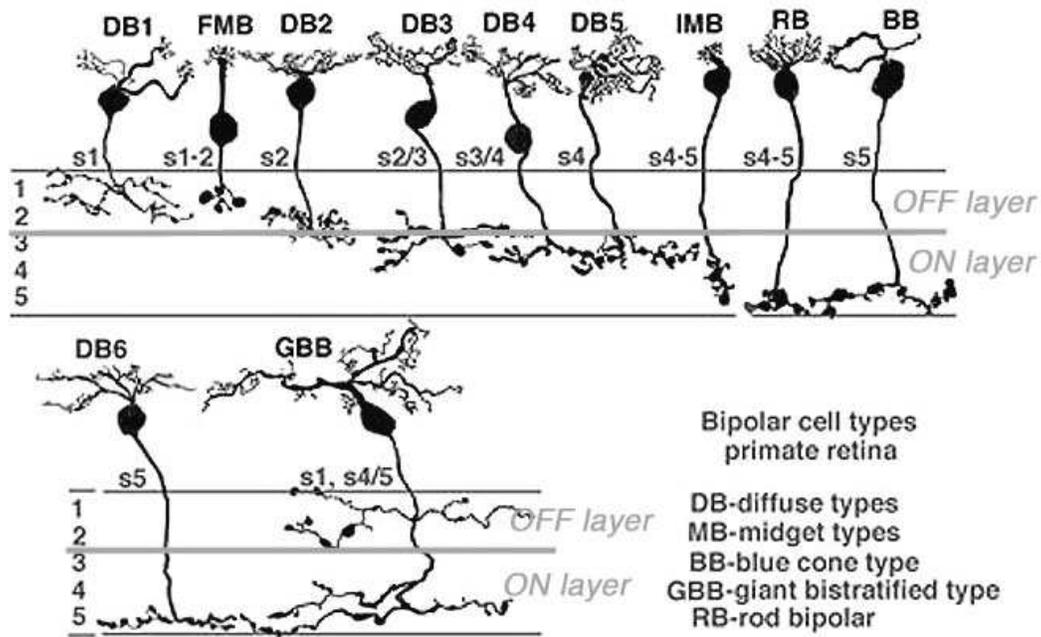


Figure 9: Bipolar cell types in human retina (Image taken from Ref. [25]). In addition to those listed in the legend, also shown are flat (FMB) and invaginating (IMB) cone types.

mitters on depolarisation. They also project to ganglion cells driven by cone bipolar cells. Their effect is to generate an ON response.

2.3 Lateral inhibition in the retina

Among the more important mechanisms within the retina is the lateral inhibition provided by horizontal cells. Horizontal cells are similar to photoreceptors in that they hyperpolarise in response to light and depolarise in darkness. This arises from the release of glutamate, a neurotransmitter which is secreted by photoreceptors under conditions of darkness. This release ceases under light conditions [15]. Horizontal cells secrete gamma

amino butyric acid (GABA), a neurotransmitter which acts to hyperpolarise the photoreceptors to which the cells project. As this release of GABA is reduced under hyperpolarisation of the horizontal cells, the inhibition of the photoreceptors is also reduced and they therefore depolarise, cancelling out the initial hyperpolarisation induced by the incident light. The overall result of this action of horizontal cells therefore is to compensate for steady shifts in the membrane potentials of photoreceptors.

It is thought that this mechanism has another important property, namely the provision of centre-surround organisation within the retina. Information is “summed” from a sizeable network of horizontal cells, which influences the photoreceptors and their downstream bipolar cells to include responses from the surrounding areas in the retina in addition to their own. This leads to a centre-surround organisation in retinal signalling, similar to that found in the neurons of the visual cortex [25].

2.4 Diseases of the retina

There are a number of diseases which affect the retina, however the current thrust of retinal implant research has focused on the effects of two in particular – age-related macular degeneration (AMD) and retinitis pigmentosa (RP). The reason for this is that these attack the photoreceptors on a large scale while leaving the remainder of the neural structure of the retina largely intact.

Age-related macular degeneration predominantly affects the elderly, although it can also come about as a side-effect of hypertension or regular smoking

[26], and arises from failure of the digestive mechanism of the pigment epithelium that forms part of the photoreceptor renewal cycle. Material which should be digested by the pigment epithelium instead accrues between it and the photoreceptor layer, creating a layer of deposits which disrupts the diffusion of nutrients and waste products to and from the choroidal blood supply. Another contributory factor is thought to be thickening of the basal lamina of the pigment epithelium with age. This leads to degradation and eventual failure of the photoreceptors in the fovea and wider macular region, causing loss of central vision and serious implications for the viability of the patient's sight. This is known as the "dry" form of the condition and is the more common of the two forms. In the "wet" or "exudative" form, Bruch's membrane is broken. This leads to the formation of redundant new blood vessels which invade the pigment epithelium and subretinal space. Blood and other fluids may then leak in from these, causing the pigment epithelium to develop severe fibrovascular scarring and the photoreceptors to become misaligned and deform, which in turn leads to failure of those photoreceptors and ultimately the affected area of the retina itself [27].

The term retinitis pigmentosa actually refers to a number of inherited diseases (such as Usher syndrome) in which rods in the outer regions of the retina progressively fail, gradually reducing the angular spread of the visual field until only the central field is left [28]. As with AMD, retinitis pigmentosa arises as a result of a disruption of the composition of the pigment epithelium by the accumulation of undigested material from the photoreceptors, in this case affecting the rods rather than the cones. There is also a corresponding reduction in visual acuity in the latter stages of the disease's progress as cone

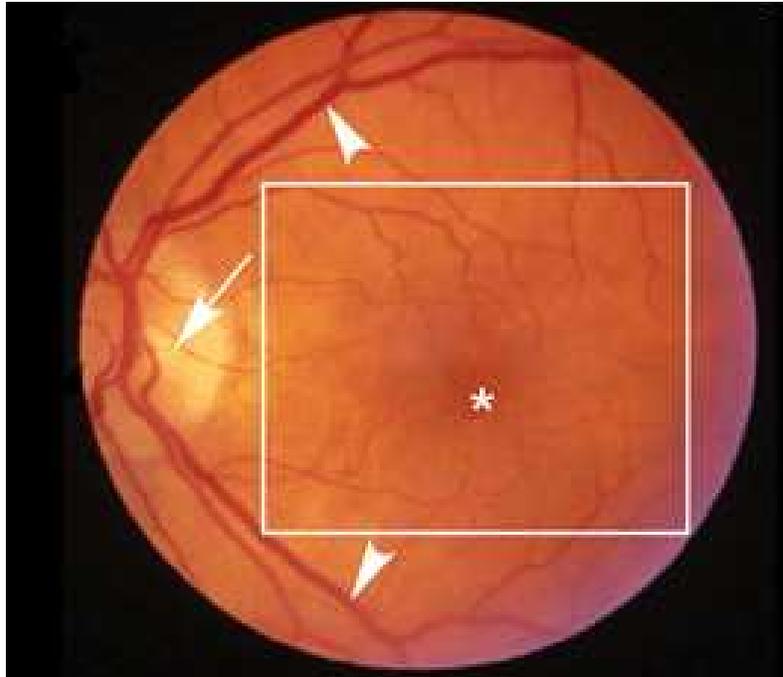


Figure 10: Normal macula of an elderly patient. The asterisk represents the location of the fovea, which lies directly in the visual axis. The macula (boxed area), which is adapted for high acuity vision, is located temporal to the optic nerve (arrow). It is approximately 6mm in diameter and centered on the fovea. The vascular arcades are indicated by arrowheads. (Image taken from Ref. [25])

photoreceptors in the central macular region come under attack [29].

Various treatments for these diseases have been studied to date. For early stage sufferers the possibility exists of manipulating the molecular environment of the region, in order to suppress the chemical growth factors involved in the blood vessel formation that characterises the progress of wet AMD. Another growth factor of interest is that derived within the pigment epithelium itself, which prevents angiogenesis and may also improve the general

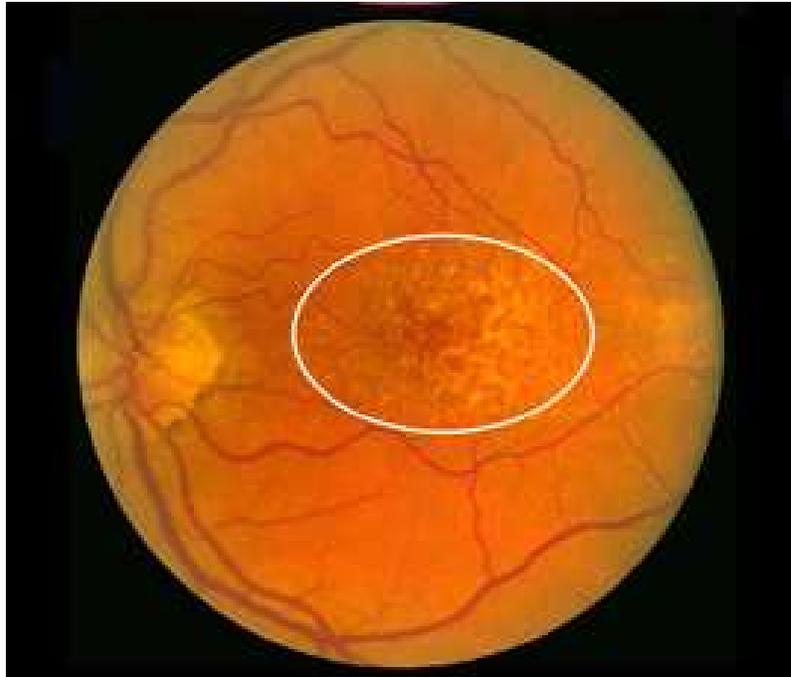


Figure 11: Colour fundus photograph from an individual with dry AMD. Numerous small and intermediate-sized drusen are visible in the macular region (oval). (Image taken from Ref. [25])

health of the pigment epithelium, paving the way for restoration of the barrier between retina and choroid [30]. Trials are also underway on a variety of anti-angiogenic agents and laser treatments [2, 3, 31]. There are currently no treatments for advanced AMD or RP [27, 28], and these conditions require a restoration of vision by artificial means such as those outlined below.

2.5 Strategies for restoration of sight

While the concept of a visual prosthesis is a 20th century one, the first of what would come to be known as phosphenes were actually observed in 1755, when in the course of an investigation into the therapeutic use of electricity,

LeRoy [32] passed electricity through the eye of a test subject who reported a sensation of light [32]. The term phosphene was first coined however in 1929, when Foerster used an implanted array of electrodes to stimulate the occipital portion of the visual cortex [33].

The most important question in visual prosthesis research centres around the positioning of the device, namely whether it should be positioned in the visual cortex, on the optic nerve or on the retina. Early attention focused on direct stimulation of the visual cortex, but more recent work has tended to concentrate on less central stimulation sites. A particularly important reason for this is that implantation in such sites involves less invasive surgical procedures than is the case for the visual cortex. In addition, the retinotopic mapping is much more complex in the visual cortex than in the retina. Paving the way for a retinal implant is the fact that inner retinal cells can survive degeneration of the outer layer in cases of diseases like age-related macular degeneration and retinitis pigmentosa, as noted in the previous section, however there is a risk of counterproductive reorganisation of the retinal structure as a result of this.

The principal attempt at a surface cortical stimulator was made by the Dobbelle group, who implanted arrays of sixty-four electrodes at the cortical surface of several patients which were found to supply phosphene vision [36, 37], and enable the patients to recognise 15 cm tall characters from 1.5 m away. However the charge injections required were found to be much higher than those generally considered acceptable for artificial stimulation in terms of eliciting visual responses. In light of this and the fact that such high charge, and by extension energy deposits, have adverse effects on human tissue, it is

thought that the electrode densities necessary to elicit visual responses are not possible in cortical surface stimulation. Additional difficulties the group encountered include controlling the number of phosphenes elicited by each electrode, interactions between elicited phosphenes, and pain and epileptic activity induced in the patients by the large electrodes. The patients also reported inability to distinguish individual phosphenes, instead reporting groups of such phosphenes surrounded by “halos”.

Intracortical stimulation moved a step forward when Brindley and Lewin developed a helmet equipped with eighty transmitters which each connected to an electrode implanted at the cortical surface. The subject, a female glaucoma sufferer, reported thirty-five phosphenes, sixteen of which were within the central 10° of the subject’s visual field. Another important finding from the study was that the subject reported that the phosphenes moved with eye movements, and with further training the subject was able to successfully learn to read “visual” Braille. These findings proved that visual prostheses were a fertile avenue for further research [38]. Other recent efforts in this direction include the Illinois Intracortical Visual Prosthesis, and the Utah Electrode Array. The former consisted of 152 intracortical electrodes which were implanted in a male macaque monkey. Their experiments demonstrated that mapping of the receptive field went hand in hand with eye tracking by training the animal to perform to perform basic memory saccade tasks. This was an important development in this field, as other studies [39, 40, 41] had shown that simple information about the position of the object in the visual field was not enough to enable a reliable picture of the visual scene to be built up. The Utah Array consists of multiple silicon spikes with platinum

electrode tips arranged in a square grid of side length 4.2 mm. It is designed as a more general prosthesis for patients with extreme muscular impairment but has been designed to be adaptable to most such cases including visual impairment.

Optic nerve stimulators offer significant advantages over both cortical and retinal stimulators from the surgical point of view – the surgical procedure to implant an optic nerve stimulator is substantially less invasive than those for cortical or subretinal and even epiretinal implants. In the same manner as the idea of electrically stimulating the retina to restore sight developed from the use of cochlear implants to restore hearing, so did the idea of electrically stimulating the optic nerve develop from the use of “cuff” electrodes to stimulate specific nerve fibres within muscles in order to promote regeneration and reacquisition of their use. A collaboration headed up by Veraart et al [35] in Belgium developed what is termed a “self sizing spiral cuff” electrode with four contacts placed at 0° , 90° , 180° and 270° on the optic nerve and undertook a series of testing and modelling studies. It was found to have similar retinotopic mapping problems to a cortical implant owing to the small number of electrodes requiring to process a wide range of inputs (the optic nerve comprises approximately 1.2 million axons within a 2 mm cylindrical structure), but also to promote optic nerve regeneration, and with tuning of the various stimulation parameters was able to induce phosphene vision in the patient, who with training was able to learn to recognise basic patterns.

The first retinal stimulator was patented in 1953 by Australian radio engineer Graham Tassicker and was based on the Becquerel effect, a voltage produced between two electrolyte-dwelling electrodes under unequal illumination. It

was implanted in a patient, between the choroid and sclera of the eye. The subject reported white light in the vicinity of the implant [34]. While the retina lends itself easily to stimulation by implanted arrays of microelectrodes, stimulation by this means is only feasible in sufferers from diseases such as age-related macular degeneration and retinitis pigmentosa, which attack the photoreceptor layer of the retina while leaving the remainder of its neural structure largely intact. In the case of conditions such as glaucoma, which also attack the higher neural layers such as the ganglion cell layer, stimulus requires to be delivered further upstream, to the optic nerve or the visual cortex.

Retinal implants currently in development fall into two categories — *epiretinal* and *subretinal*. Epiretinal implants are placed on the inner surface of the retina and stimulate the ganglion cells directly, whereas subretinal implants are placed inside the photoreceptor layer of the retina, as shown in Fig. 2. The Retinal Group, a multi-disciplinary collaboration bringing together scientists in the University of Glasgow’s Department of Physics & Astronomy and Institute of Biomedical & Life Sciences (IBLS) as well as microelectronics experts at the STFC Rutherford Appleton Laboratory (RAL), are working toward a device in the former category, as were the Boston Retinal Implant Project which brings together technology specialists at the MIT Research Laboratory of Electronics and medical scientists at Harvard Medical School [11], however they have now adopted a subretinal approach. The other most prominent collaborations working toward devices in the latter category are a US Department of Energy backed project bringing together retinal specialists at the University of Southern California’s Doheny Eye Institute, technology

specialists at various facilities within the National Laboratory system — including Los Alamos and Lawrence Livermore — and the private company Second Sight Medical Products Inc. in California [42, 43, 44, 45]; the private company OptoBionics Inc. based in Illinois [49]; the German SUB-RET project, which brings together electronics specialists based in Stuttgart and the specialist Eye Hospitals of the universities of Regensburg and Tübingen [81]; and a collaboration between Stanford University’s Departments of Neurobiology and Ophthalmology and Hansen Experimental Physics Laboratory [82].

2.6 Development of retinal implants

As noted above, a common factor in the progress of age-related macular degeneration and retinitis pigmentosa is the fact that they both destroy the photoreceptors of the retina while leaving the remaining neural components of the retina, most importantly the ganglion cells, intact in relatively large proportions [58]. Retinal implants currently in development are being designed to exploit this. Unfortunately this means that they are unlikely to be of benefit in the treatment of retinal diseases that also attack the ganglion cells, such as glaucoma and diabetic retinopathy.

The Boston group [11] have developed an implant which transfers information about the visual scene and power signals to deliver the electricity to the electrodes. Because of the necessity for the device to be implantable chronically and biocompatibility concerns, these are transmitted wirelessly by means of RF transmission. A transmitting apparatus is mounted on the arm of a specially-designed pair of glasses and the implant itself is equipped with a

receiving coil. The required information about the visual scene is captured by a camera, also mounted on the glasses, and then analysed to generate the appropriate pattern of stimulus. The power signals are transmitted separately. Electrical currents are passed through electrodes at the appropriate locations on the retina to deliver the stimulus. Six experiments have been performed on humans to date. These were relatively brief, generally lasting only for hours. The electrode arrays were placed in the subretinal space and small electrical currents delivered. The test subjects reported small spots of light and basic motion detection [21].

Another epiretinal implant was one developed by Schanze et al at Philipps University Marburg in Germany [56]. This one operated on the same basic principle as those documented above, with the important difference that the apparatus for reception and transmission of power and signals were contained in an intraocular lens which was to be implanted in place of the original eye lens. The capture of the visual scene was handled as before by a camera mounted in a pair of glasses, also including the necessary data processing circuits and transmitter coils. The electrode array was to be mounted on the inner retinal surface in a similar manner to other epiretinal implants, and the two connected by wire tracks and encapsulated in parylene.

This implant was fabricated with a smaller number of electrodes than planned for testing purposes and implanted in three adult cats. One implant failed to function due to an error in the fabrication process, while one of the other two failed to elicit a response from the retina. This was found to be due to detachment of the array from the retinal surface. Responses obtained were verified by infrared stimulation of the eyes, which evoked no response, as would be

expected. A further study was conducted, also with three cats implanted with the device, which demonstrated that the device was capable of generating selective responses in particular areas of the retina and consequently visual cortex [57]. However, the design was found to require modification in order to be scalable, and the parylene encapsulation remained to be proved stable on the scale of chronic implantation.

The USC group have developed two devices so far in collaboration with Sec-

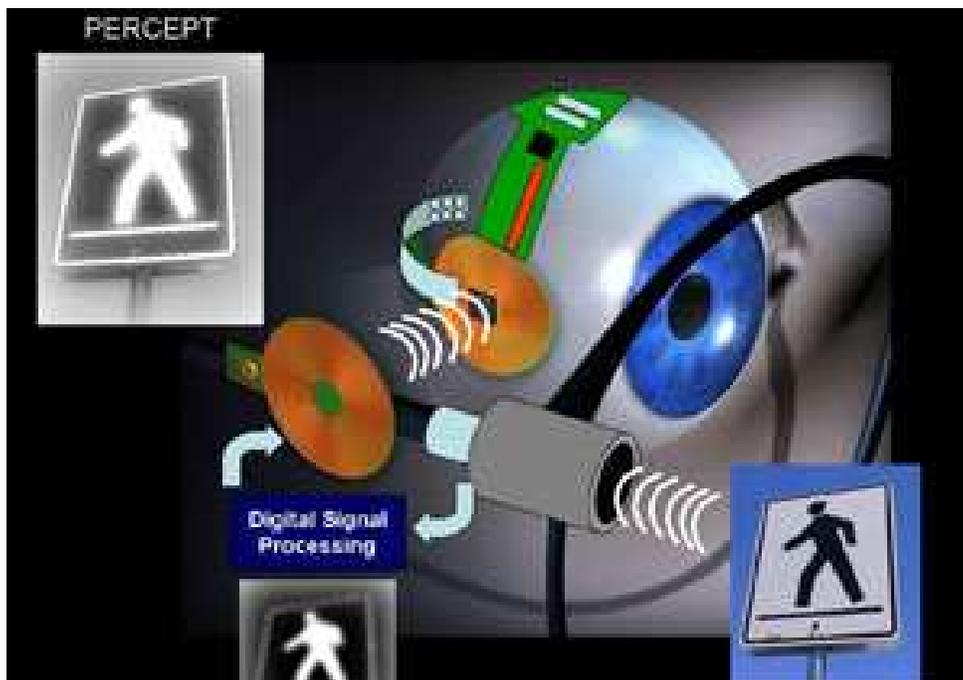


Figure 12: Diagram showing the basic operational principles of the Boston retinal implant. Image taken from Ref. [11].

ond Sight, with a third currently in development. The first, known as *Argus I*, was implanted in six patients between 2002 and 2004, all of whom were in various stages of retinitis pigmentosa. It consists of an array of sixteen electrodes in a 1-inch square package incorporating similar wireless transmis-

sion mechanisms to the Boston device. Power is supplied to the device by means of a belt-worn battery pack. The patients were able to detect when lights were switched on, describe motion and count items. Five of these devices have since approved for long-term home use [46]. The second, known as *Argus II*, is currently undergoing clinical trials. It is smaller, 5 mm by 5 mm, has more electrodes (sixty), and can be implanted in a much shorter time than its predecessor [80]. The third is currently undergoing design and fabrication studies in participating Department of Energy laboratories. It will have two hundred or more electrodes and be made of flexible material in order to enable it to conform more closely to the curvature of the eyeball. It will also be substantially smaller again [80]. The ultimate goal of this project is to produce a lasting device enabling the recipient to recognise faces and read large print.

A last epiretinal type of implant was prototyped by a group at the University of Nagoya in Japan which would have been a “hybrid” implant combining electronics with tissue grafting. The motivation behind this was to attempt to provide therapeutic regeneration of the surviving ganglion cell layer and ensure a strong connection to the brain by cultivating a layer of neural tissue on top of the electrode array whose axons would connect into the optic nerve. However, a study of the device in adult cats failed to demonstrate the formation of a viable axonal connection [47].

The Zrenner group [81] have developed a device based on a microchip of 3mm diameter and 50 μm thickness, containing an array of 1,500 70 μm by 70 μm pixels. Each of these is equipped with a photodiode, an amplification circuit and a stimulating electrode. The photodiodes absorb the incident light and

convert it into electrical energy. The amplification is necessary in order for the electrical pulses generated by the photodiodes to reach the thresholds necessary for stimulus, as the incident light alone cannot generate the energy necessary. All the components which make up the device are attached to a polyimide ribbon, and all but the stimulation electrodes embedded in silicon. The device has a total length of 100 mm, a total width of 3 mm and a total thickness of 0.1 mm. One end of the device is implanted in the retina itself, while the other is sewn onto the eyeball and enclosed within the conjunctiva. Clinical trials have been undertaken for this device — in the first instance with energy transmitted from an external source, and in a later series with high frequency bursts of light. The ultimate aim is to achieve spatial orientation, a visual field of 12° , and to give patients the ability to count items and recognise faces, read lettering without visual aids and see in surround brightness over a range of 10 to 100,000 Lux [48].

Problems with photodiode transduction of light into electrical signals have



Figure 13: A photograph of the prototype subretinal implant developed by the German SUB-RET project. Image source: www.retina-implant.de.

also affected the devices developed by OptoBionics, Inc. in Illinois [49]. As found by Zrenner et al, elicitation of visual responses is not possible with photodiodes without amplification in place. In their device, 5mm by 5mm photodiodes were coupled to electrodes which in early experiments were exposed to a flash stimulator [50]. Action potentials were observed but issues

arose with the scalability of the device. The scaled version [51] was found to remain stable and functional when chronically implanted in the subretinal space, which was an important result [52]. However, after removal, the electrodes were found to have corroded and there was also found to be a decline in the number of photoreceptors in the region. Their goal is now therefore to work around this degeneration of the retina by etching holes into the device to allow the flow of nutrients, while also improving the ratio of photodiode area to electrode area in order to boost charge injection [53]. Ten such devices have been chronically implanted so far. These have diameters of 2mm and thicknesses of $25\mu\text{m}$ and incorporate 3500 photodiodes each. Subjects reported improved vision across the entire field for 18 months after implantation. However this is thought to be due more to side-effects from the surgery rather than the presence of the device itself.

Each category of retinal implant has its respective advantages and disadvantages. The most important advantages of epiretinal implants are the adaptability of their charge injection mechanisms and their applicability to both high and low light intensities. The main disadvantages are that they require external cameras to film their visual input and have yet to attain the resolution necessary for a viable degree of visual acuity; in implants assembled so far, small numbers of electrodes are arranged over a wide area, placing a limit on the amount of detail patients can see. Additionally, they put the patient at risk of proliferative vitreous reaction. Subretinal implants such as those developed by OptoBionics Inc. are at a more advanced stage in terms of chronic implantation — six patients have had devices implanted to date — and have the benefit of not requiring external cameras. However,

they have no external power supply and therefore need high brightness levels to work — they have shown no evidence so far that they can reach the power levels necessary to elicit responses from the retina. Furthermore, the implantation procedure for such implants can be more complex than for epiretinal implants [81].

2.7 Group strategy

A common factor in epiretinal implants developed to date is a dependence on an external camera for visual input. Subretinal implants have the related problem of requiring high brightness levels to work. Neither of these can be regarded as satisfactory solutions. Use of an external camera mounted on a pair of glasses has the crucial disadvantage of removing eye movements from the process of interpretation of the visual scene. This is important because when the brain transmits a nervous signal to move the eyes, it anticipates a corresponding change in the image transmitted back to it by them. This will not occur with implants which depend on external cameras for visual input as these are necessarily restricted to head movements only. Sensitivity to eye movements can only be achieved by development of a device which incorporates both acquisition and processing of visual input, while being small enough to be implanted in the eye as a single unit.

The Glasgow particle physics group, of which the Retinal Group is a subsidiary, is an established member of the ATLAS and LHCb collaborations at CERN. As part of its contribution to these, the group were tasked with assisting in the development of CMOS silicon strip pixel detectors for the inner detector mechanism of the ATLAS experiment and similar detectors for

the vertex locator mechanism of the LHCb experiment. This work led to the idea of employing such pixel sensors as the front end of an epiretinal implant in conjunction with a suitable electrode array. These sensors possess certain properties which lend merit to the idea of adapting their principles for the purpose of developing a fully implantable retinal prosthesis. Both systems make use of semiconductors in their detectors, and require high resolution, low power imaging devices, with on-pixel processing desirable. These design requirements resulted in the development of the IPIX programmable stimulation chip.

There have been two major strands to the Group's electrode array development work. The first, which is beyond the scope of the work documented in this thesis, has been undertaken in collaboration with the Santa Cruz Institute for Particle Physics (SCIPP) and the Salk Institute in San Diego. This work looks into how the retina, as an extension of the brain, interprets the visual scene. To this end, arrays of 519 electrodes at $30\ \mu\text{m}$ pitch were manufactured from indium tin oxide (ITO) on a glass substrate [59] for use in conjunction with the Retinal Readout System developed by SCIPP for the purpose of conducting *in vitro* studies of retinal signalling [60]. Tests on guinea pig retina using these arrays demonstrated good detection of retinal action potentials, and showed that such arrays, appropriately adapted, could form a viable conduit for artificial stimulation of the retina.

The second is the development of flexible arrays for *in vivo* testing, and ultimately implantation. Flexible arrays of 8 electrodes were fabricated for use in retinal experiments, with one of 61 electrodes also fabricated in order to demonstrate the scalability of the technology [61]. These arrays had elec-

trodes of diameter $50\ \mu\text{m}$, and as a consequence demonstrated the ability to detect action potentials from several adjacent ganglion cells at once under *in vivo* experimental testing. Smaller electrodes were therefore seen as necessary in order to improve the capability of isolating individual ganglion cells, an important requirement for electrode arrays intended for *in vivo* implantation.

The work documented in this thesis has as one of its primary goals the optimisation of the size of electrode required for the retinal implant envisaged by the Group. Electrode arrays of 16 electrodes have been fabricated for use in retinal experiments conducted in the Institute for Biomedical and Life Sciences (IBLS) at the University of Glasgow. The results of these experiments will be used to determine a number of important parameters, key among these being the *rheobase*, or minimum current, required for stimulation, and the *chronaxie*, which is the minimum stimulation time required at twice the rheobase current. These are in turn related to the minimum charge, current and energy deposits into the retina, which can be used to optimise the electrode diameter. This result will then inform the development of the next generation of electrode arrays for the retinal implant.

These principal components of the retinal implant device then require to be joined together in order to ensure that the stimulus is directed to the appropriate position on the retina. Off-chip connections are sufficient for the purposes of experimental testing and characterisation, however a device to be implanted in the eye will require to be self-contained in order to serve any useful purpose. To this end, a programmable on-chip neural network was incorporated into the next IPIX chip, known as the IPIX II. This was also

tested and characterised as part of the work documented in this thesis.

3 Multi-electrode arrays

An overview will now be given of the manufacturing process for the micro-electrode arrays, from the initial design stages to the electrical testing in preparation for use under experimental conditions. Microelectrode arrays are one of the principal components of any proposed retinal implant. The IPIX II, a CMOS imaging and stimulation chip, the features of which will be explored in detail in the next chapter, forms the front end of the device and is responsible for detecting incident light, converting that light into appropriate electrical signals and directing these signals to the appropriate positions on the retina. The electrode array forms the delivery mechanism for such signals. The fabrication work undertaken for the studies documented in this thesis was undertaken in the James Watt Nanofabrication Centre (JWNC) at the University of Glasgow. This facility enables a wide range of nanofabrication techniques to be carried out, and the most important to these studies are outlined below. After this, the mask designs for the arrays fabricated over the course of the current project are reviewed, and the fabrication processes for each described in detail.

3.1 Theory of key nanofabrication processes

3.1.1 Optical lithography [63]

This is the first stage of this fabrication process. The design of the device is divided into as many different layers as is needed and a photomask is produced for each of these layers. These are designed using a custom software utility, in this case *Tanner L-Edit*. Masks are fabricated from quartz or so-

dalime glass, depending upon the required resolution. Some mask aligning equipment, such as optical steppers, also allow for pattern reduction by the use of optical methods. Masks require to be produced with a high level of optical transparency at the desired exposure wavelength, and equally importantly a surface which minimises light scattering. They must also transfer a pattern to the device surface, and an opaque layer of chromium is used and placed in intimate contact with the sample during the exposure process.

Prior to this process, the surface of the wafer is coated with a photosensitive substance known as *photoresist* [62]. The mask is aligned to the sample using a mask aligner. In the JWNC, a *contact mask aligner* (Karl Suss MA6) is used for this purpose. Contact mask aligners operate by pressing the mask against the coated wafer during the process. The mask and the wafer are manually loaded into the appropriate areas of the device, and once alignment is complete, they are exposed to light. The optical source projects the image of the mask onto the wafer, with the chromium pattern masking the defined regions.

Various types of radiation are used for such processes, the most common being visible light (covering wavelengths of approximately 400-750 nm), ultra violet (300-400 nm), deep ultra violet (120-300 nm) and extreme ultra violet (10-120 nm), depending on the type of photoresist used and the level of resolution required. For the purposes of the work documented in this thesis, ultra violet light was used.

3.1.2 Electron beam evaporation [64]

Thin-film evaporators such as those used in the JWNC operate on a crucible heating system. In this system, wafers are loaded into a high-vacuum chamber and pumped with a diffusion pump. The material to be deposited – known as the “charge” – is located in a crucible. There are three distinct types of crucible heating systems, known as *resistive*, *inductive* and *electron beam*, distinguished by the method used to heat the charge. The tools used in the JWNC operate under the electron beam system, in which an electron gun located underneath the crucible fires a high-energy beam which is incident upon the charge. This beam is then “rastered” across the surface of the charge, melting a substantial portion of it. The melted portion of the charge is given off as a vapour which, because of the low pressure involved, simply radiates outward in straight lines toward the wafers. Up to 24 wafers may be loaded into some evaporation chambers, in which cases the deposition rate for a given wafer depends on its position in relation to the crucible. The devices used in the JWNC, however, operate on individual wafers.

3.1.3 Etching techniques and liftoff

The exposure process involves the formation of a photoresist image on the surface of the wafer. This image is then transferred to the layer below by means of etching. Two etching techniques are of importance in the studies documented in this thesis: *wet etching* and *reactive ion etching* (RIE). A related technique, known as *liftoff*, is also sometimes used.

In *wet* etching, three separate processes are involved, all of which must be completed if an etch is to be successful. The etchant in the solution must

first diffuse to the etch site. It must then react with the exposed surface of the wafer at the etch site. And finally the reaction products must diffuse away from the etch site. The slowest of these three processes determines the duration of the etch. The solution is normally agitated in some way, usually by means of ultrasound, in order to ensure that the etch proceeds in a uniform manner. This etching process is only used for tasks in which anisotropy is unimportant, as wet etching tends to undercut the material that is being etched.

Reactive ion etching (RIE) is a dry, anisotropic etching process and takes the place of wet etching for tasks which require a high degree of anisotropy. In this system the wafer is placed in an evacuated chamber and electrically isolated, before the gas to be used for the etch is introduced. An RF power signal is then applied to the wafer, resulting in an oscillating electric field which ionises the gas, creating a plasma. The free electrons are then either absorbed into the chamber walls or the wafer itself. The latter process results in a large negative voltage building up on the wafer, which results in the positive ions being drawn downward onto the sample and reacting both chemically and physically with it [66]. As the plasma must initially be in contact with the chamber walls, RIE is conducted at low pressure [66].

Ion bombardment from RIE can result in damage to the sample substrate, and so *liftoff* is used for more delicate tasks unsuited to straightforward wet etching. This is a related but somewhat different process from the straightforward wet etch process outlined above. In the case of liftoff the layer of resist is spun onto the wafer and patterned before the metal is deposited. The sample is then immersed in an acetone solution, in which the resist is

soluble. The metal deposited on top of the resist is “lifted off” by the dissolution of the resist while that deposited directly onto the sample remains behind [67].

3.1.4 Polyimide substrate

A device which is intended for *in vivo* implantation requires to be both biocompatible and flexible. Polyimides are increasingly becoming the material of choice for the housing and insulation of such devices. Richardson et al. showed them to be viable for use in neural implants [70], while Akin et al. demonstrated them to be non-toxic in *in vivo* tests on guinea pig [69]. Stieglitz et al. [71] compared polyimide to silicon, silicon dioxide and silicon nitride over a range of parameters and found it to offer in particular improved insulation in addition to lower density and greater flexibility. The polyimide favoured in these studies was tested on rabbit by Seo et al. [72] and found to be biocompatible. Polyimides can be used in standard cleanroom processing and in a wide variety of biomedical applications. It is for this reason that polyimide is used in the fabrication of the electrode arrays for these studies.

3.2 Mask design and manufacture

The first step towards the realisation of the 8- and 16-electrode arrays which were used in the retinal experiments was the design and manufacture of the masks. Electron beam lithography was used to make these masks. This is a process in which a beam of electrons is emitted in a patterned fashion over a surface coated with photoresist, in similar fashion to the radiation used

in photolithography. The purpose behind this is to create small structures on the resist which are transferred to the substrate, again as occurs in photolithography. It is used to write masks as it enables features to be written to nanometre precision, which is important as variations in feature size are an important source of systematic error and require to be minimised as much as possible.

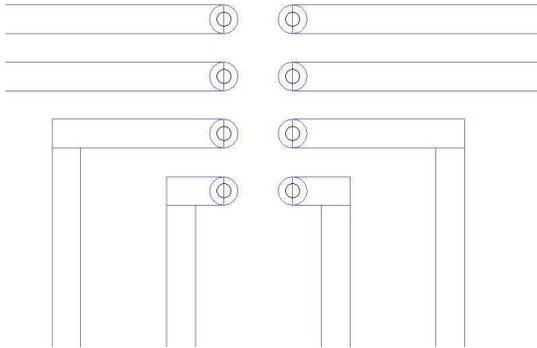


Figure 14: Close-up image of the electrode sites of the 8-electrode array, as viewed from within *Tanner L-Edit*.

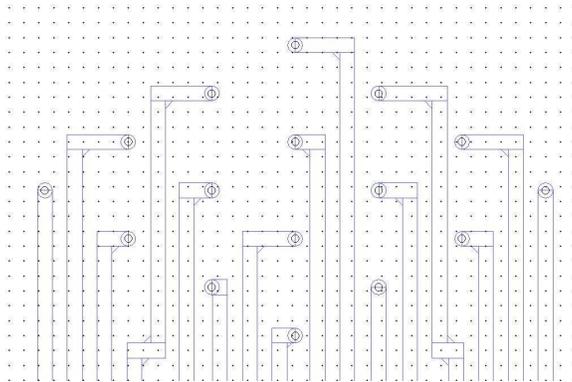


Figure 15: Close-up image of the electrode sites of the 16-electrode array, as viewed from within *Tanner L-Edit*.

The designs for the arrays were produced using the *Tanner L-Edit* software

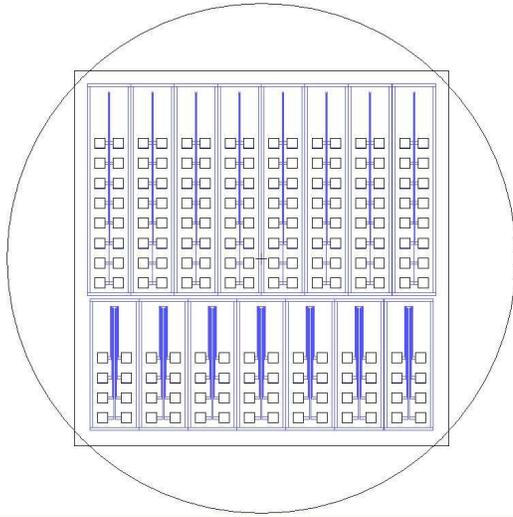


Figure 16: Image of the wafer design showing the arrangements of the 8-electrode arrays (bottom row) and the 16-electrode arrays (top row), as viewed from within *Tanner L-Edit*.

utility which allows mask layouts to be designed to particular specifications. The designs for the 8-electrode arrays used for proof of concept and 16-electrode arrays are shown in Figs. 14 and 15. Three layers from the *L-Edit* palette were used. The first, and lowest, layer was used for base alignment marks – these require to be placed on the mask for every layer in order to ensure features on different layers are in the correct positions in relation to one another. The second layer was used for the wire tracks which are the conducting components of the arrays. The third layer was used for the vias layer to reflect the fact that the vias for the electrodes are etched into the passivation layer of polyimide that is applied to the device after the first etch stage. Once the designs were finalised, they were arranged to make maximum use of the area of a 4” silicon wafer. The optimal arrangement turned out to be eight copies of the 16-electrode array and seven copies of the 8-electrode array design, as shown in Fig. 16. After the design was completed, it was then submitted to the electron beam lithography system to enable the mask to be manufactured.

As can be seen in Figs. 14 and 15, the initial electrode array designs had a rectangular arrangement of electrodes in the case of the 8-electrode design and a hexagonal arrangement in the case of the 16-electrode design. The 8-electrode design had $30\ \mu\text{m}$ vias, $50\ \mu\text{m}$ wire tracks and $50\ \mu\text{m}$ pitch, while in the case of the 16-electrode design the vias were $5\ \mu\text{m}$ in diameter, the wire tracks $10\ \mu\text{m}$ in diameter and the pitch was $60\ \mu\text{m}$ [68]. The latter design was employed in retinal experiments conducted within the IBLS under the direction of Dr. J. D. Morrison. The results of these experiments informed the development of the software used for their analysis as documented later in this thesis, and also enabled proof of the electrode-electrolyte interface model as a viable basis for the prediction of the properties of electrodes of a particular diameter, as documented in the next chapter.

Two further configurations were designed for the purpose of enabling the

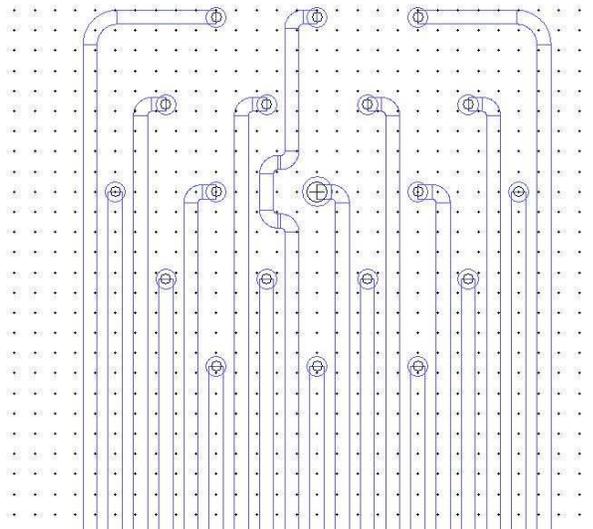


Figure 17: Close-up image of the electrode sites of the single-electrode stimulation array, as viewed from within *Tanner L-Edit*.

use of *single-electrode stimulation* and *correlated stimulation* in such studies,

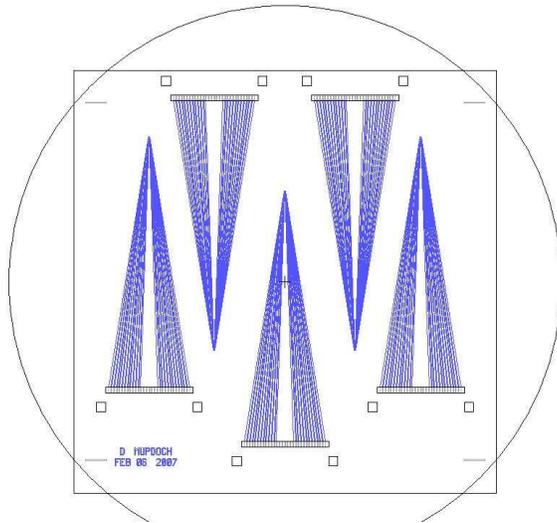


Figure 18: Image of the wafer design showing the arrangement of the copies of the focussed stimulation array, as viewed from within *Tanner L-Edit*.

and enabling clearer understanding to be achieved of the signalling of retinal neurons.

Single-electrode stimulation involves the use of a single stimulation electrode

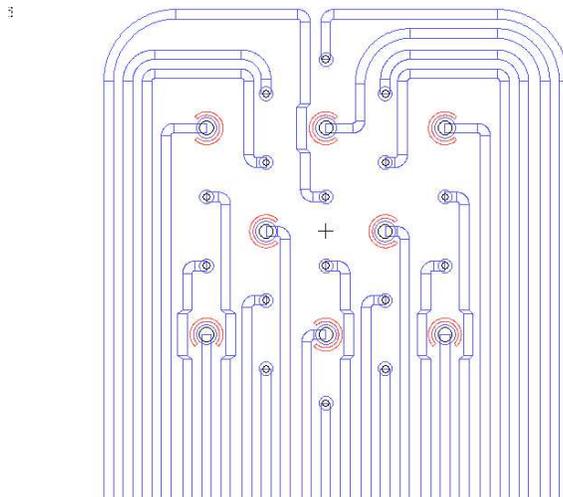


Figure 19: Close-up image of the electrode sites of the correlated stimulation array, as viewed from within *Tanner L-Edit*. The larger circles correspond to the larger stimulation electrodes.

to deliver the maximum possible stimulation to a particular area of the retinal

sample, surrounded by smaller recording electrodes to detect the response. *Correlated stimulation*, on the other hand, uses several stimulation electrodes in addition to the range of recording electrodes to enable simultaneous signalling of several neurons to be studied and correlated, in order to build up a more complete picture of the behaviour of the tissue sample.

As before [68], the first step towards the realisation of these new designs was the design and manufacture of the masks by optical lithography. The designs for the arrays were once again produced using the *L-Edit* software. Once the designs were finalised, they were arranged to make maximum use of the area of a 4" silicon wafer as before. In the case of both, five copies of the array were arranged as shown in Fig. 18. After these designs were completed, the single-electrode stimulation array was then submitted to the electron beam lithography system in the James Watt Nanofabrication Centre to enable the masks to be manufactured. Two masks were manufactured; one for the wire tracks and another for the vias.

3.3 Fabrication

The fabrication procedure for the arrays designed followed a number of distinct steps which are outlined below.

The first step was to coat the 4" silicon wafer (Fig. 20) with PI-2545 polyimide compound (Fig. 21). This substance forms the substrate of the arrays, and has the two useful properties of being flexible in its solid state — a requirement for any array being developed for *in vivo* implantation — and being an insulator, which therefore eliminates the unwanted resistive effects

that arise from a conducting or semi-conducting substrate. The polyimide is decanted on to the wafer surface, with care being taken to avoid the edge of the wafer. Once this is done, it is evened out over the surface of the wafer by means of spin coating. The appropriate thickness for a single polyimide layer, for the purposes of this project, is approximately 2–3 μm . The wafer is slowly ramped to a final speed of 2000 rpm for 30 s. After spinning, the wafer is placed on a hotplate for 4 minutes at 140°C, in order to set the polyimide for deposition of further layers. Five layers of polyimide are applied to the wafer prior to the *metal deposition* stage.

In the metal deposition stage, an electron-beam evaporator was used to deposit a 50 nm layer of titanium followed by a 150 nm layer of gold onto the polyimide surface of the wafer (Fig. 22). The gold ensures that the thin and narrow wire tracks have the required conductivity; the titanium is applied simply to aid the adhesion of the gold, as it does not adhere readily to polyimide.

Next is the application of *primer* and *photoresist* (Fig. 23). Photoresist is a substance which forms a protective coating on the wafer but alters its chemical structure upon exposure to ultra-violet light. It is important in the following stage of the procedure, in which the designed mask containing the array layouts is used to pattern the wafer. The primer, in this case hexamethyldisilazane (HMDS), is applied simply to facilitate application of the photoresist. The photoresist used is of type S1818 and is applied by a filtered syringe and spun at 4000 rpm for 30 s. The wafer is then baked in a 90°C oven for 30 minutes.

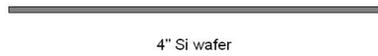


Figure 20: Graphical representation of the 4" silicon wafer prior to commencement of the fabrication process.

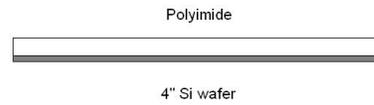


Figure 21: Graphical representation of the 4" silicon wafer after completion of the polyimide application process.

The *exposure and development* stage comes next (Fig. 24). The designed mask is placed in the mask aligner, directly above the wafer, and the sections of the wafer defined by the mask are exposed to ultra-violet light for approximately 4–5 s. The wafer is then placed in a 1:1 solution of developer concentrate and water for 75 s. The success of this process can be adversely affected by any contaminants present on the mask. In this event, the situation is rectified by cleaning the mask, flood exposing the wafer (essentially the same process as detailed above but without a mask in the aligner), re-applying the resist and repeating the procedure.

The next stage of the procedure is the *etch*, where the unwanted metal is removed from the wafer. The gold is wet-etched with potassium iodide solution (Fig. 25), for which the typical etch rate is approximately $2\text{--}5\text{ nms}^{-1}$ — the gold is therefore etched through after around 45–60 s. The titanium is dry-etched by means of RIE (Fig. 26); this is necessary because wet etching

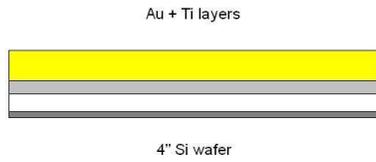


Figure 22: Graphical representation of the 4" silicon wafer after completion of the metal deposition process.

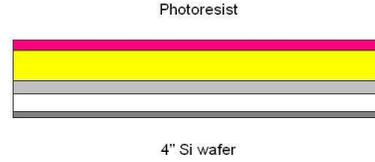


Figure 23: Graphical representation of the 4" silicon wafer after completion of the photoresist application process.

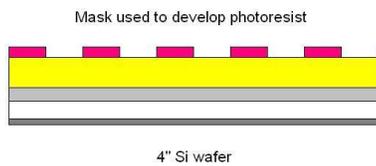


Figure 24: Graphical representation of the 4" silicon wafer after completion of the exposure and development processes.

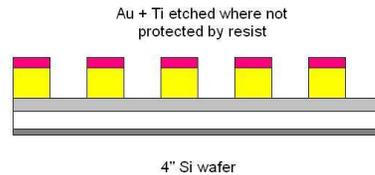


Figure 25: Graphical representation of the 4" silicon wafer after completion of the wet etching process.

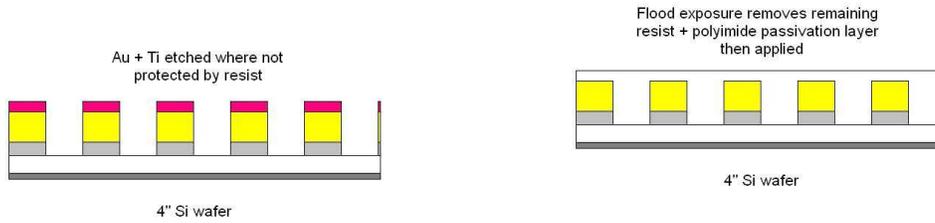


Figure 26: Graphical representation of the 4" silicon wafer after completion of the reactive ion etching process.

Figure 27: Graphical representation of the 4" silicon wafer after completion of the flood exposure process and application of the passivation layer.

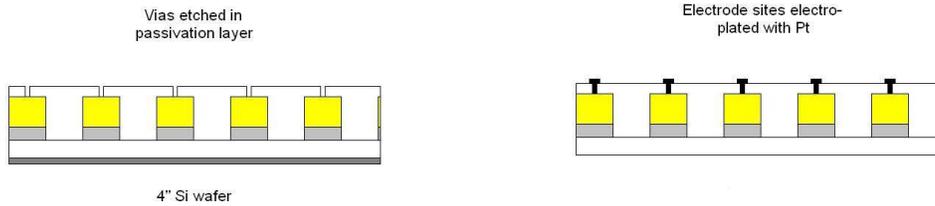


Figure 28: Graphical representation of the 4" silicon wafer after completion of the via etching process.

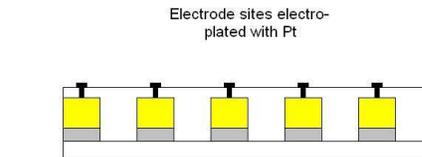


Figure 29: Graphical representation of the 4" silicon wafer after completion of the electroplating process.

is an isotropic process and wet-etching this layer in addition to the gold will create an undercut by etching the side walls of the metal, causing the metal tracks to peel off from the substrate. RIE avoids this as it is an anisotropic process. A 4-minute etch completely removes the titanium layer where it is not protected by the resist. This is conducted with $SiCl_4$ at a flow rate of 18 sccm, under a pressure of 9 mTorr and an RF power of 250 W.

After this process has been completed, the remaining photoresist is removed by flood exposing the wafer and developing as described above. An insulating top layer of polyimide, known as the *passivation layer*, is then applied, using the spin-coating method outlined above, and holes known as *vias* dry-etched through the polyimide (Figs. 27 and 28). This process is conducted in an Oxford Plasma Technology BP80 RIE system using an 80:20 mixture of C_2F_6 and O_2 and takes 18 minutes. These vias are to enable electrical connections to be made to the bondpads and the electrodes themselves.

The last stage of the fabrication before the arrays may be removed from the wafer is the *cure*, where the wafer is placed in an oven at an initial temperature of 20°C, and ramped up to 200°C at a rate of 4 °Cmin⁻¹. It is held at this temperature for 30 minutes before being ramped to 350 °C, again at a rate of 4 °Cmin⁻¹, and held at that temperature for 60 minutes. After this the temperature is ramped back down to 20 °C at 4 °Cmin⁻¹ in order to enable removal of the wafer from the oven.

After the arrays are removed from the wafer, electrical connection to the bondpads is achieved by soldering the individual wires of a length of 16-channel cable to each; the opposite ends are soldered to the 16 pins of a

standard connector which enables the device to be connected to the low-noise readout electronics.

The final step in the procedure is to electroplate the electrode sites with a rough-surfaced platinum deposit known as *platinum black* (Fig. 29). This has the effect of increasing their surface area and therefore greatly reducing their impedance. An increase in surface area of around two orders of magnitude is usually achievable in this manner [79]. The process involves immersing the tip of the array containing the electrode sites in chloroplatinic acid and connecting a current source to the bondpads and a platinum wire return electrode. Thus, when current is passed, the electrode sites are driven to negative potential and positively-charged platinum ions from the acid are drawn into the vias. A current density of $4 \text{ nA}\mu\text{m}^2$ is used to electroplate the electrodes, meaning that around 80 nA is usually required for a $5 \mu\text{m}$ diameter electrode. The process is complete when the voltage driving the current decreases and reaches a minimum (usually around 0.5 V). The sites are platinised one at a time in order to ensure a better quality of platinisation. To facilitate activation of individual channels, a switching device was used to isolate the individual sites. Once the completed devices are platinised to a fairly consistent impedance (around $200 \text{ k}\Omega$ at 1 kHz frequency for $5 \mu\text{m}$ diameter electrode sites [79]) they are ready to proceed to the testing stage.

3.4 Electrical testing and characterisation

Once the arrays have been fabricated, they must be electrically tested to ensure that they have the required electrical properties. A Keithley LCR

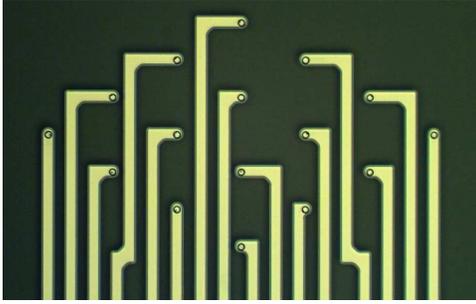


Figure 30: 16-electrode array at 20 \times magnification. Each wire track leads out to an individual bondpad (not shown).

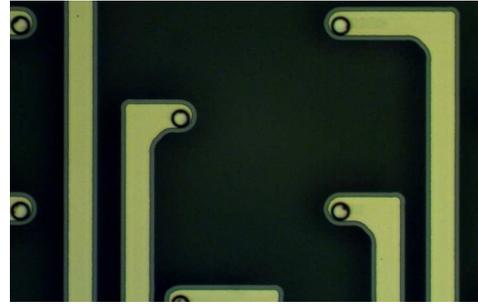


Figure 31: 16-electrode array at 50 \times magnification, showing etch detail and via holes.

meter and Cascade semiconductor probe station were used to measure the resistances of their bondpads and the capacitances between the individual channels. Arrays which possess reasonably consistent values for both across all measurement sites are retained for use in retinal experiments. The inter-channel impedances are also usually measured — these are typically in the $M\Omega$ – $G\Omega$ range.

An additional check that is performed on the arrays is to measure the impedances of the electrodes under experimental conditions. This is accomplished by immersing an array in physiological saline and connecting it to an LCR meter using a probe station. A platinum reference wire is immersed in the saline along with the array to facilitate measurement and each channel is selected individually for testing with the switching device. The probe station is connected to a PC running a LabVIEW environment which contains VIs allowing the performance of a variety of electrical tests. For the purposes of this study, measurements of impedance and phase angle vs. frequency were

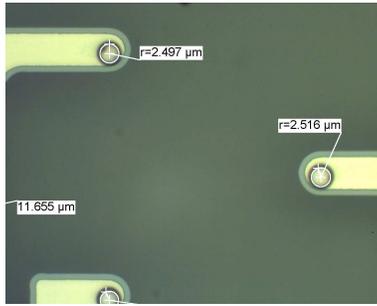


Figure 32: 16-electrode array at $100\times$ magnification, showing measurements of electrode radii.

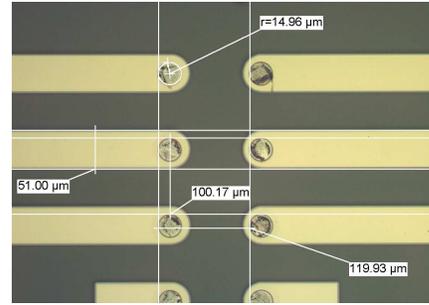


Figure 33: 8-electrode array at $100\times$ magnification, showing measurements of electrode radii and wire track widths.



Figure 34: A photograph of the completed wafer (courtesy of I. MacVicar), showing the arrangement of the five copies of the focussed stimulation array.

made. This was accomplished by applying a 0.1 V sinusoidal waveform to the system, increasing its frequency in a stepwise fashion from a minimum of

120 Hz to a maximum of 100 kHz, and recording the impedance and phase angle at each frequency in between. The impedance curve should have an approximate “backward s” shape. The phase angle curve should have an approximate “forward s” shape from large negative angle (strongly capacitive behaviour) to small negative angle (strongly resistive behaviour). Examples of such curves are shown in Figs. 35 and 36.

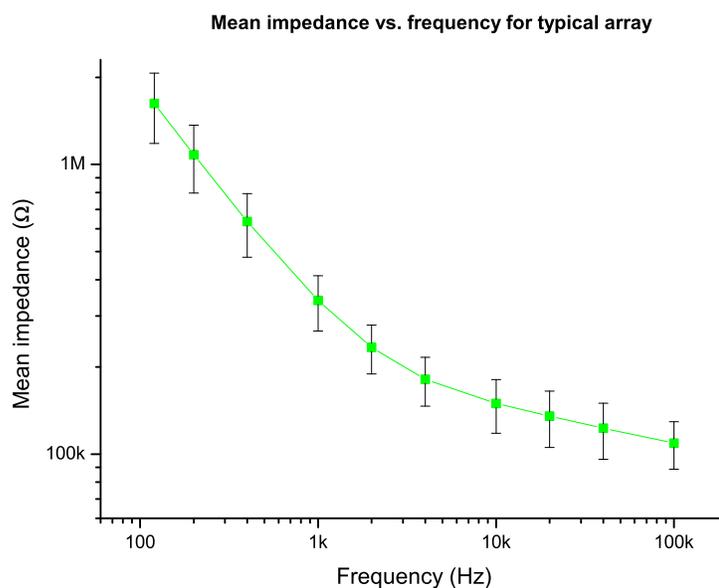


Figure 35: Plot of the mean impedance with respect to frequency for twenty $5 \mu\text{m}$ -diameter electrodes.

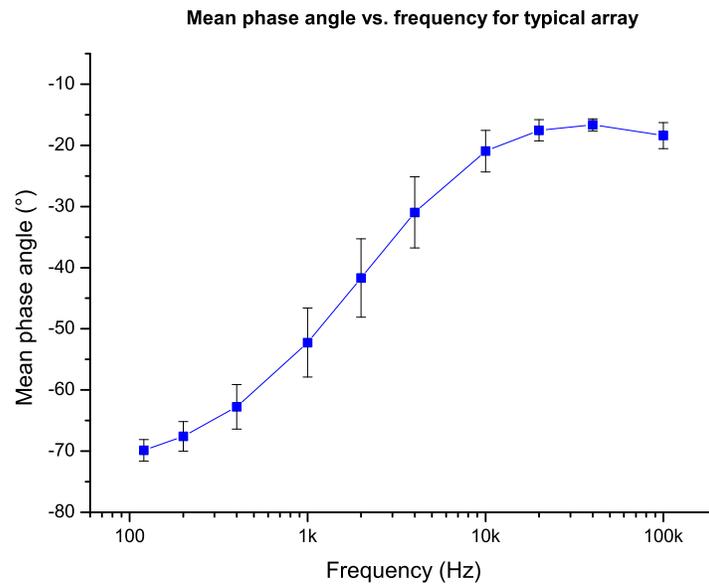


Figure 36: Plot of the mean phase angle with respect to frequency for the same selection of electrodes.

4 CMOS imaging and stimulation chip

The electrode arrays documented in the previous chapter form the delivery mechanism for the stimulation of the retina by the implant. The IPIX II, a CMOS chip incorporating the capability to both detect photons and direct stimulus to the correct area of the retina, forms the other principal component. An overview follows of how the device operates, from the initial detection of incident photons to the final output of the stimulus pulse, and how it builds on the functionality of its predecessor, the IPIX. There then follows an overview of the electrode model which forms part of the foundation of these studies, its relationship to experimental data, and how it has influenced the development of the IPIX II.

4.1 Introduction

The IPIX series of programmable retinal stimulation chips is the other major component of the retinal implants envisaged by the Glasgow group. Two such devices have been developed so far under the direction of M. Prydderch at the STFC Rutherford Appleton Laboratory. The second of these, known as the IPIX II, is currently undergoing characterisation and refinement work, and several experimental studies have been carried out of its operation. These are documented later in this chapter.

The original IPIX design consisted of a 10×10 array of $100 \mu\text{m}$ pixels. A photodiode inside each pixel detected light from an image and converted this into an electrical signal. These were read out by a shift register with a pointer which addressed each row in turn and directed the control signals to the ap-

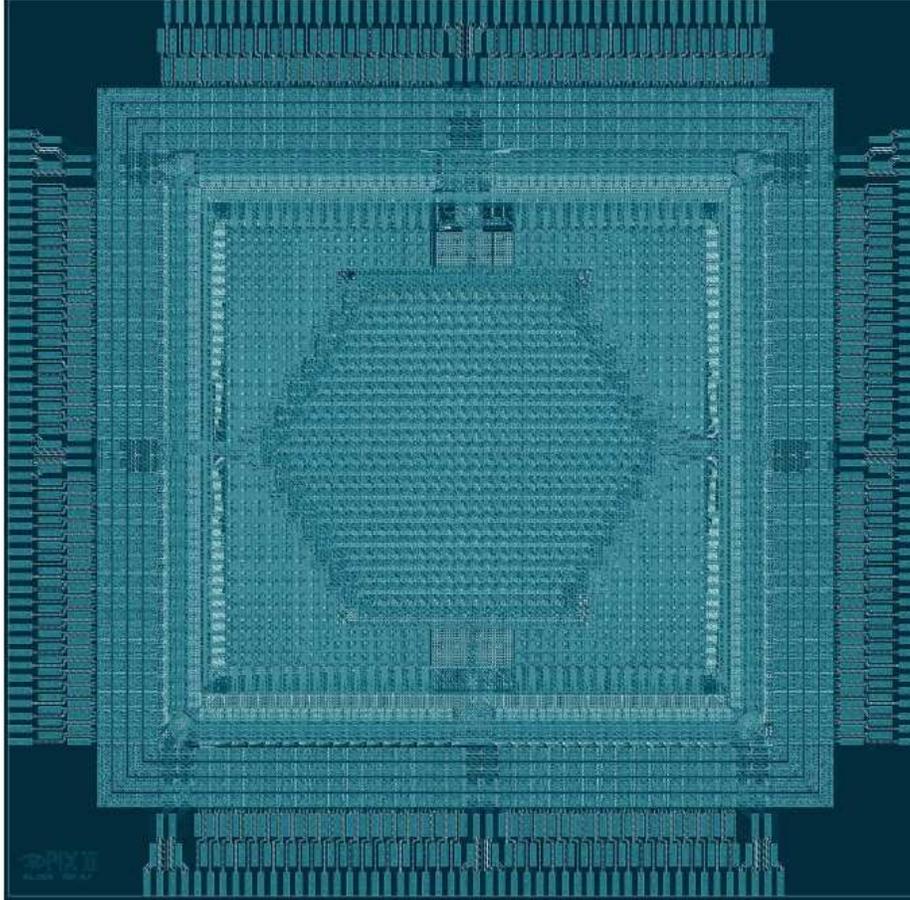


Figure 37: Photograph of the architecture of the IPIX II chip, clearly showing the central array (hexagon), level hold/shift and VCO (inner square) and output driver circuitry (outer square). The array is $7211 \times 7077 \mu\text{m}^2$ in size, with the pixels spaced at intervals of $107.5 \mu\text{m}$. The bond pads are $50 \times 180 \mu\text{m}^2$ and have a pitch of $230 \mu\text{m}$.

appropriate pixels, a forerunner to the inclusion of local on-chip neural network capability in the IPIX II. The time taken to revisit each row determined the integration time. The output of each pixel on a row would be driven onto the appropriate column readout line by an NMOS source follower. The

chip could be either be configured to operate from the photodiodes (internal source) or from an external source. The central nine pixels were connected to bondpads for the purpose of interfacing with an electrode array.

A photograph of the architecture of the IPIX II chip is shown in Fig. 37, and the corresponding block diagram in Fig. 38. The design falls into three clear sections. The central, hexagonal section contains the pixel array itself, as well as the associated shift register and local on-chip neural network components. The innermost of the two square sections contains the level hold, level shift and voltage-controlled oscillator (VCO) circuitry, while the outermost square contains the biphasic timing circuitry and output drivers for the VCOs. These are routed out to bondpads, which connect to the microelectrodes. One pixel is designated as a test pixel for the purposes of experimental testing. This pixel has two bondpads dedicated to it, one of which connects to the neural network output, bypassing the VCO and subsequent circuit components in order to facilitate independent testing of the local on-chip neural network. A block diagram depicting the layout of all these components is shown in Fig. 39.

The IPIX II chip operates in a similar manner to its predecessor [78]. Incident photons are detected by the pixel array, and photocurrents generated by the photodiodes are converted to voltages which in turn drive VCOs. These in turn generate biphasic stimulus pulses which are passed to the electrode array to be transmitted to what will ultimately be a live retina under *in vivo* conditions.

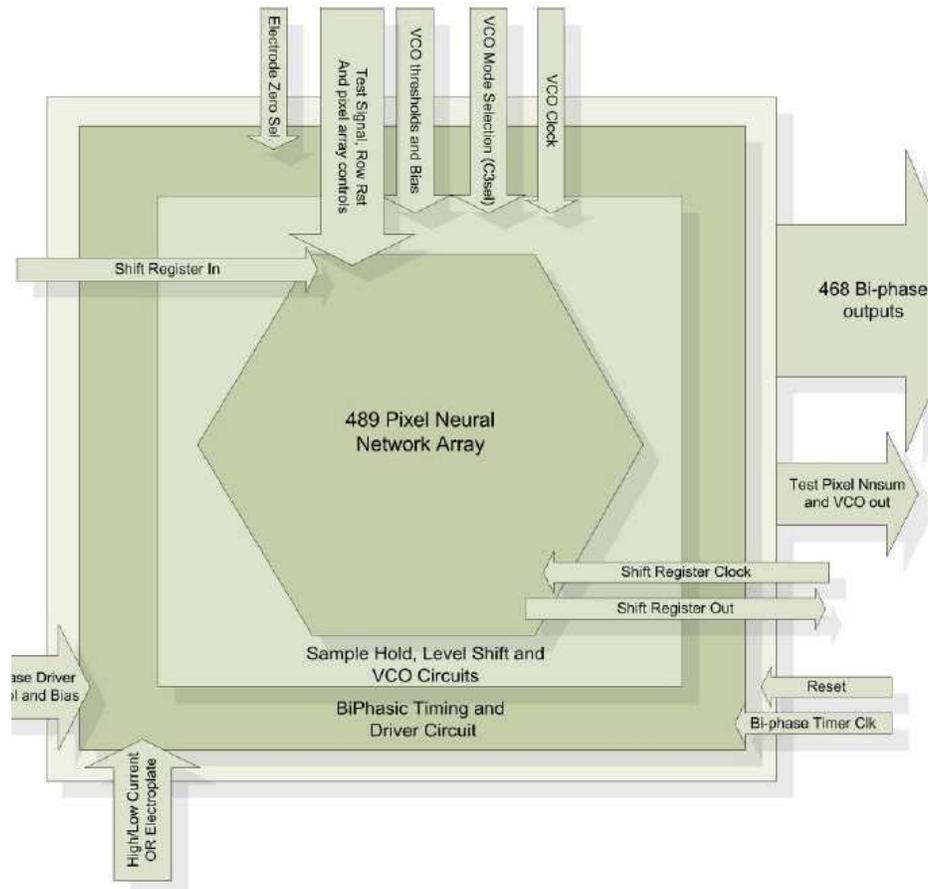


Figure 38: Block diagram of the IPIX II chip, giving a more detailed overview of the components seen in the photograph.

4.2 Overview of IPIX II components

4.2.1 Pixel array

The first component is the pixel array, which simulates the response of the rods and cones of the retina to incident photons. In the original IPIX this was a 10×10 array of $100 \mu\text{m}$ pixels. In the IPIX II design a hexagonal arrangement has been adopted. The current array is 13 pixels along each side and incorporates 469 pixels in total. The architecture of the pixels is broadly

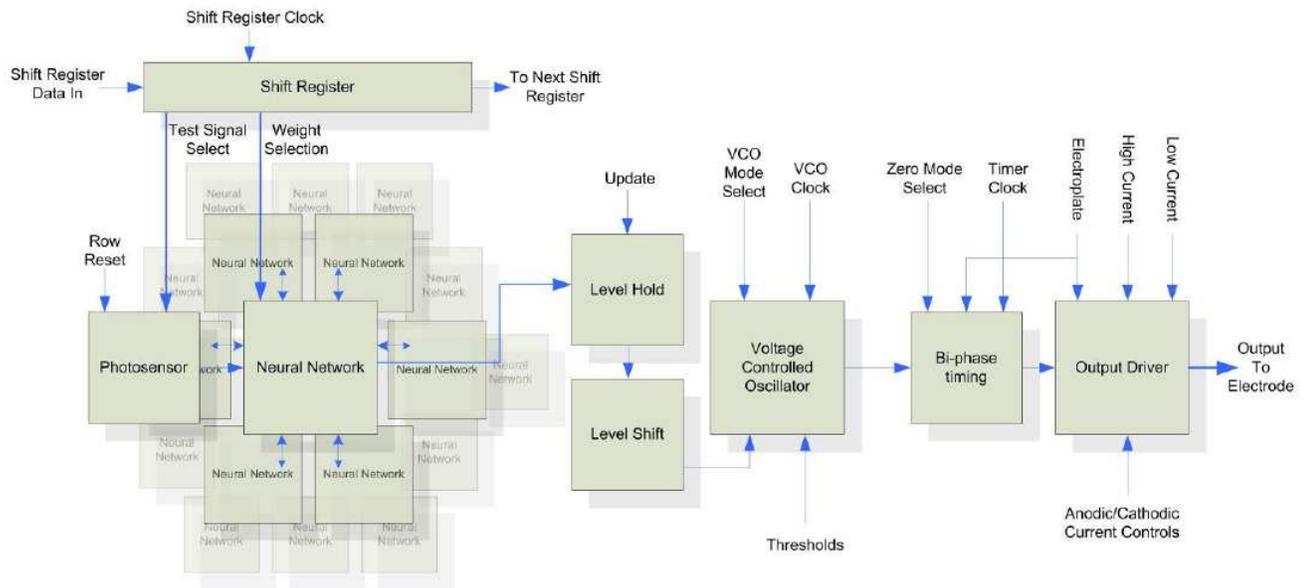


Figure 39: Block diagram of an individual channel in the central array, showing the key components.

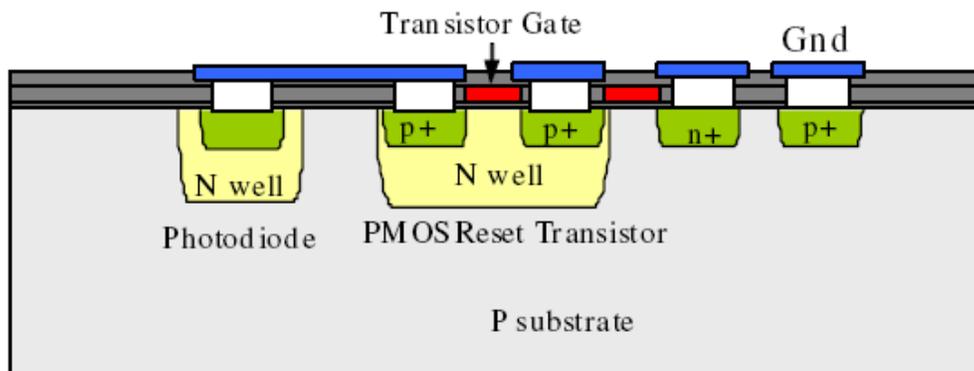


Figure 40: Diagram of the photosensor structure within a given pixel.

similar. The photodiodes in the photosensors take the form of n -type wells implanted into p -type substrates as shown in Fig. 40, small enough to keep

their capacitances and leakage currents within limits conducive to a good signal-to-noise ratio. Their positions alternate depending on their position within the array to give a hexagonal-close-packed arrangement, as can be seen in Fig. 44, in order to give better resolution than is possible with a square array. The capacitance of the photodiode acts as an integrator of the generated photocurrent, and a NMOS source follower then converts this into a voltage. The power supply rail keeps the pixel array at 3.3 V under darkness. This voltage is reduced under exposure to photons, with the saturation point at around 1 V, which corresponds to the maximum frequency output of the VCO.

Three digital signals are transmitted to the pixel array at various times in the course of photosensor operation, as shown in Fig. 41: *RstB*, *RowRst* and *Update*. *RstB* is a reset signal which acts on the VCO and biphasic timing logic circuits. *RowRst* is brought low (logic zero) in order to reset the photodiodes and held for 10 μs before being set high (logic one) again. The *Update* signal is brought high after the chosen integration time has elapsed and held, again for 10 μs before being switched to low again. *RowRst* may then be applied and a new cycle begun.

4.2.2 Voltage controlled oscillator (VCO)

The next stage of the system is the VCOs. Their behaviour is designed to be analogous to that of the bipolar and amacrine cell layers of the retina. As previously described, these are the cells which generate slow-wave potentials in response to the activation of the photoreceptors by incident photons. This process is simulated by variation of the voltage supplied to the VCO, which

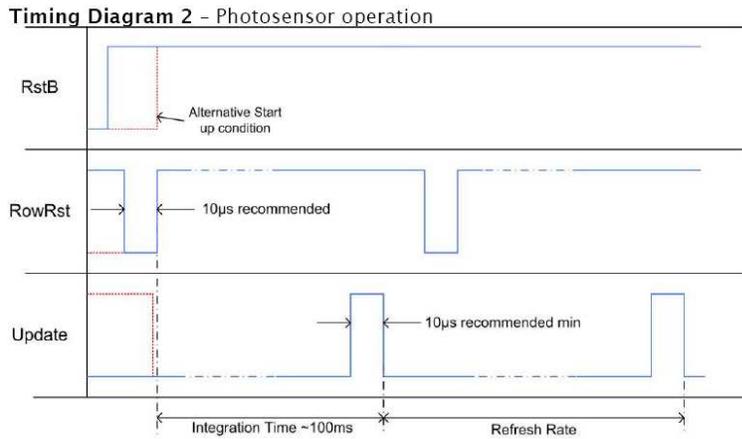


Figure 41: Diagram of the signals transmitted to the chip during photosensor operation.

results in a signal whose frequency is in inverse proportion to this voltage. The operation of the VCO is governed by four signals; the VCO clock, the ϕ_1 and ϕ_2 clock signals and the timer clock, as shown in Fig. 42. The VCO clock is used to generate the ϕ_1 and ϕ_2 signals which govern the operation of the VCO circuit. ϕ_1 has the same frequency and phase as the VCO clock, whereas ϕ_2 , while also having the same frequency, is in antiphase. The frequency is designed to be set at 1.97 MHz as this ensures a 400 μs pulse width and 100 μs interphase delay, however this can be varied. The VCO clock is independent of the timer clock which is transmitted through a separate connection.

The circuitry for the VCO, as seen in Fig. 43, has remained largely the same as that of the IPIX. The main advance in the IPIX II is the incorporation of

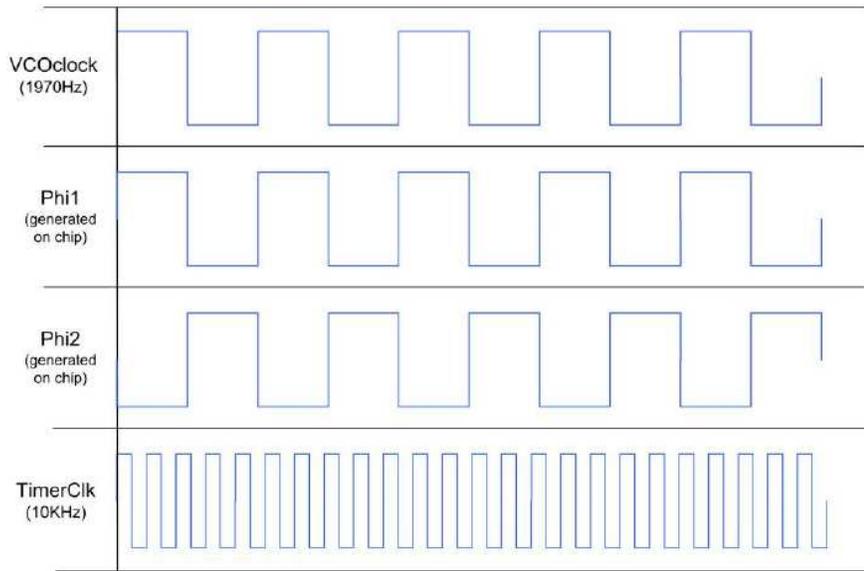
Timing Diagram 3 – VCO and Timer circuit clocks.

Figure 42: Diagram of the signals transmitted to the VCO and timer circuitry during operation.

two operating modes covering different ranges of frequency. While the VCO of the original IPIX covered only a frequency range of 10-50 Hz, the VCO of the IPIX II can be set to cover either 0-50 or 0-100 Hz, by setting the *C3Sel* digital switch to high or low. The 0-50 Hz range, corresponding to charging of both C3 capacitors, is the default.

The circuit is driven by the two distinct clock signals, ϕ_1 and ϕ_2 . These govern the sequence of charge and discharge of the capacitors C1, C2 and C3. During the ϕ_1 stage, C2 charges from the reference voltage, A_{ref} , to the comparator voltage and C1 charges by the difference between the reference and control voltages. These charges then transfer to C3 during the ϕ_2 stage. C1 and C2 are small in relation to C3 (26 fF and 10.15 fF respectively

compared with 250 fF or 500 fF for C3 depending on the selected mode of operation) so the output of the amplifier only increases by a fraction of the two voltages. The amplifier is not reset, so these cycles continue until the amplifier output crosses the upper comparator threshold, V_{th} , at which point the comparator output switches, as does the polarity of the charge of C3. The sequence of the clock signals also swaps over. The effect is that the amplifier output begins to step back until it crosses the lower comparator threshold, V_{tl} . In this way the VCO produces successive anodic and cathodic square-wave pulses similar in profile to the type required for *in vivo* stimulation. The centre frequency of the circuit is fixed by the ratio of C2 and C3, the two clock frequencies and the settings of the two comparator thresholds. The comparator thresholds are generally set to provide for a swing of around 2 V at the output of the amplifier. As only two possible capacitor ratios are built into the circuitry, the only other flexibility lies in the choice of frequencies for ϕ_1 and ϕ_2 .

The frequency of oscillation is determined by the following equation:

$$f = \frac{1}{2\left(\frac{\Delta V_{thresh}}{\Delta V_{step}}\right) \times \Phi_{PERIOD}}$$

in which ΔV_{thresh} is the difference between V_{th} and V_{tl} , and Φ_{PERIOD} is the combined period of the two clock signals. ΔV_{step} is the step size for the output voltage of the amplifier, which remains the same for both directions and is given by:

$$\Delta V_{step} = \frac{C2(V_{cp} - V_{ref}) + C1(V_{ref} - V_{cntl})}{C3}$$

in which V_{cp} is the positive output voltage of the comparator, V_{cntl} is the input control voltage and V_{ref} the VCO reference voltage. This equation

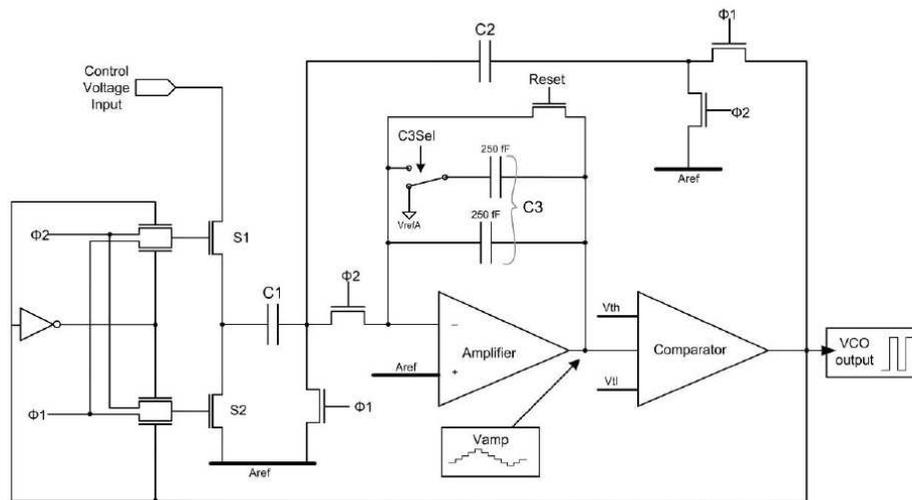


Figure 43: Diagram of the VCO circuit.

shows intuitively how selection of the value of $C3$ alters the behaviour of the system.

4.2.3 Neural network

In the original IPIX design, the output from the pixels was read out along the columns before being fed back to the VCOs. This was always intended as a forerunner to the inclusion of a local on-chip neural network to serve as the intermediate stage between the pixel sensors and the electrode array, which has been included in the IPIX II design. This is necessary in order to simulate the functioning of the horizontal cells of the retina which project to the photoreceptors. As noted previously, the horizontal cells perform two principal functions – opposition of the polarisation states of the photoreceptors to which they project, and subtraction of steady shifts in membrane potential which can result from prolonged exposure to incident light – both

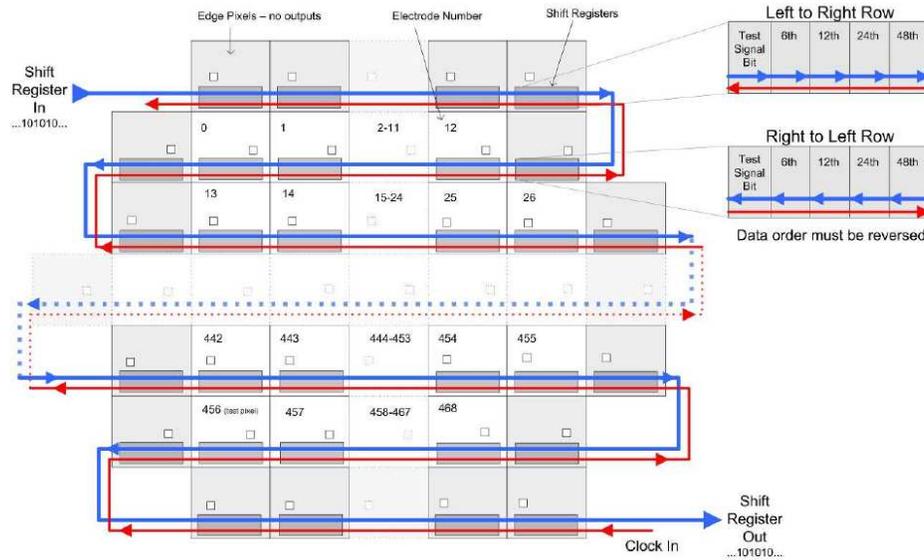


Figure 44: Diagram showing how the principal data signals are snaked through the array. The order of the bits transmitted to each pixel varies depending on the row addressed. The white squares denote photosensors, which are placed on alternate sides of pixels to produce a hexagonal-close-packed arrangement for greater resolution.

for the purpose of facilitating continued response to incident light. The former process is simulated by the subtraction of a weighted sum of the output currents of the six nearest neighbours from the output current of each pixel, thereby providing a zero response under uniform light and edge or boundary detection under variations in incident light intensity. This process may under certain conditions lead to a negative result being returned by the summation. The level hold is programmed to reject these, thereby ensuring that the currents supplied to the VCO remain within the appropriate ranges, in analogy with the latter process.

The readout and data transfer within the chip has also been significantly altered owing to the new design of the pixel array. Rather than data signals being transmitted via columns, which is impractical with a hexagonal array design, the data signals are snaked through the array from the top left, with the ordering of the data bits dependent on the direction of data flow in a given row of the array, as shown in Fig. 44. The two principal signals distributed in this way are the shift register and the clock, which snakes through from the bottom right. The shift register is important in governing the use of the on-chip neural network.

The shift register consists of five bits for each pixel, and these are set to low whenever the register is reset. The ordering of these bits alternates depending on which row of the array is being addressed at a given time; however, taking the first row of the array as an example, the order of the bits from left to right (looking down on the chip from above in its optimal orientation) is: test signal, and $\frac{1}{6}$, $\frac{1}{12}$, $\frac{1}{24}$ and $\frac{1}{48}$ weights of the neural network. The test signal is connected to the VCO input and when activated (set to high) supplies a PC-controlled voltage to the VCO rather than a voltage created by exposure to incident light. The remainder of the bits activate particular combinations of the weights of the neural network. The purpose of these weights is to vary the sharpness of the detail of the detected image. If a pixel has the $\frac{1}{6}$ weight activated and is exposed to uniform light, then the output of the neural network for that pixel will be zero. If there is a variance in the current of one or more of the neighbouring pixels, the effect will be one of edge (or boundary) detection. The greater the weight, the sharper the resolution of these edges and boundaries.

A number of signals are transmitted to the network prior to the beginning of operation as shown in Fig. 45. A reset signal, $SRrst$ is applied and then brought high in order to remove the reset conditions on the chip. Its effect is to connect all the photosensors for full light sensitivity (by setting the test signal selector in the shift registers to low) and deselect all the neural network weights (by setting their shift register bits to low). $SRClk$ is the clock signal for the loading of the shift register and is transmitted in the opposite direction to the shift register configuration data. It can be set for any frequency up to 50 MHz. $SRin$ is the signal which sets up the array for operation under experimental conditions. Values for each of the bits in the shift register (for all pixels in the array) are stored on disk in an ASCII file which can be edited by the user in order to fix the settings of the pixel array or individual pixels – the latter being desirable for the experimental work documented in the next section.

The neural network operates by means of current sources and sinks and is designed to mimic, as closely as possible, the lateral inhibition functions performed by the horizontal cells of the retina. For a given pixel, an electrical signal is generated by the photons incident upon it, the voltage corresponding to which is driven into the neural network by the NMOS source follower. The neural network generates a current from this, which it mirrors to the pixel's six nearest neighbours, scaled by the selected weight. These weighted currents are in turn mirrored into a summing node and a weighted sum of these subtracted from the non-weighted current in the pixel. Mathematically the calculation performed corresponds to:

$$Output = LightIntensity - \sum_{i=1}^6 (W_i \times NeighbourIntensity_i)$$

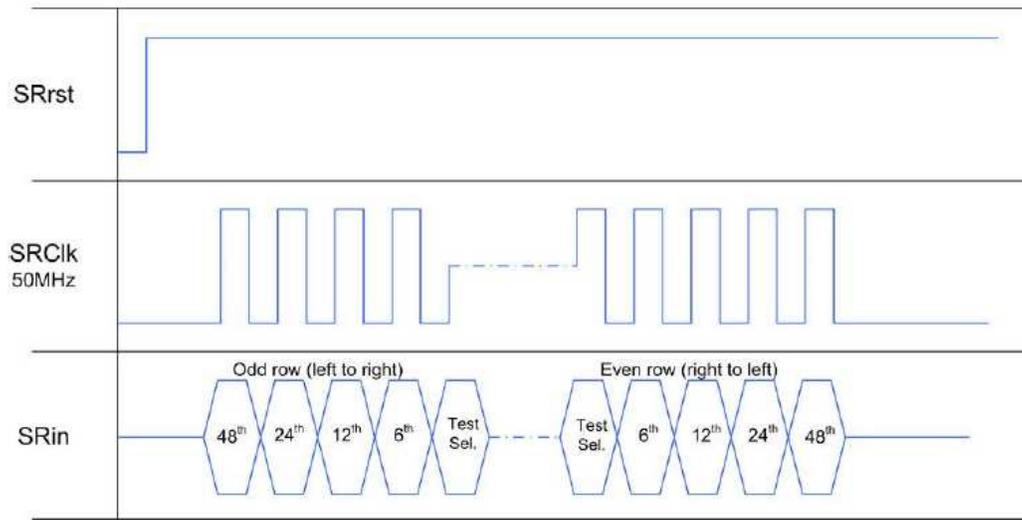
Timing Diagram1 - Configuration Register load

Figure 45: Diagram of the “configuration register load” signals transmitted to the chip prior to the commencement of operation.

The result of this sum is then transmitted out of the central array to the level hold storage capacitor. This is refreshed regularly by use of the *Update* signal described previously. The values stored are then buffered to the VCO via the level shift stage. This is an important component as it enables the suppression and subsequent discard of negative-valued neural network sums, which can be generated by large differences in current between pixels. The VCO alters its frequency output in response to these inputs.

4.2.4 Biphasic output

The last components of the chip are the biphasic timing logic and output circuits. The purpose of the logic circuits is to drive the biphasic output of the system according to the voltages governing the VCO. The timing circuit

consists of a series of NAND, NOR and NOT gates. The biphasic pulses produced by the circuit are defined by three parameters: a pulse width t_1 , an interphase delay t_2 which separates the cathodic and anodic phases of the pulse in time, and the period of the waveform, T . These are determined by the circuit architecture.

The biphasic output drivers follow on from this. They generate the current waveforms to be transmitted to the bond pads which connect to the electrode array. They are the final stage of the system prior to the electrode array and are designed to mimic the analogue-to-digital conversion performed by the ganglion cells of the retina on the slow-wave potentials generated by the bipolar and amacrine cells. They operate in a similar fashion to the neural network, mirroring reference currents into switchable sources and sinks, which are connected to bond pads leading to the electrode array. The switching is achieved by directing the output of the mirroring node onto the appropriate rail of the power supply. The timing circuits control this process, creating a biphasic pulse similar in profile to Fig. 46.

A facility has been built into the circuitry to enable independent control of both the positive and negative current amplitudes in order to ensure exact charge balancing of the delivered pulse. This is controlled by the *PveControlV* and *NveControlV* faders on the LabVIEW instrument documented in the next section (see Fig. 51), and sets the base current from which the other mirrors work, of which there are two, one for high current (*HighI*) and one for low current (*LowI*). Another constraint on the current stems from the RC behaviour of the electrodes. It is most pronounced in *HighI* mode as the voltage across the impedance of the electrode/electrolyte interface may not

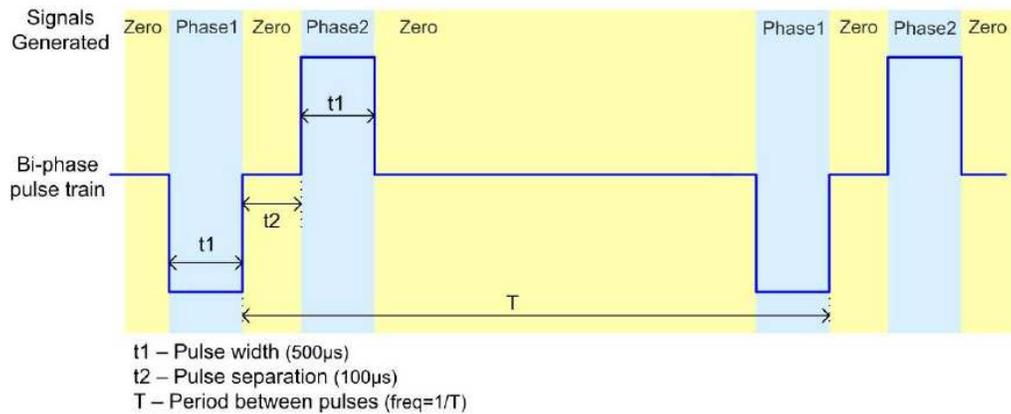


Figure 46: Diagram of a typical biphasic pulse as generated by the biphasic output circuit of the VCO.

exceed the supply voltage.

Under *in vivo* conditions, the bond pads would at this point be connected to a specially fabricated electrode array. For the purpose of the *in situ* experiments documented in this chapter, the chip was mounted on a printed circuit board (PCB) whose circuits mimic the presence of an electrode/electrolyte interface in the manner documented in the next section.

4.2.5 Electrode interface with retina

The electrode model used is based on the circuit diagram shown in Fig. 47, as used by Borkholder [73] and Mathieson et al.[79] in recent electrode studies, and has a number of antecedents [75, 76, 77]. Modelling plays an important part in any development process as it enables the properties of the system under consideration to be determined for any set of parameters, and facilitates consequent optimisation of these parameters. This model will enable the optimal electrode size for an implantable array to be determined.

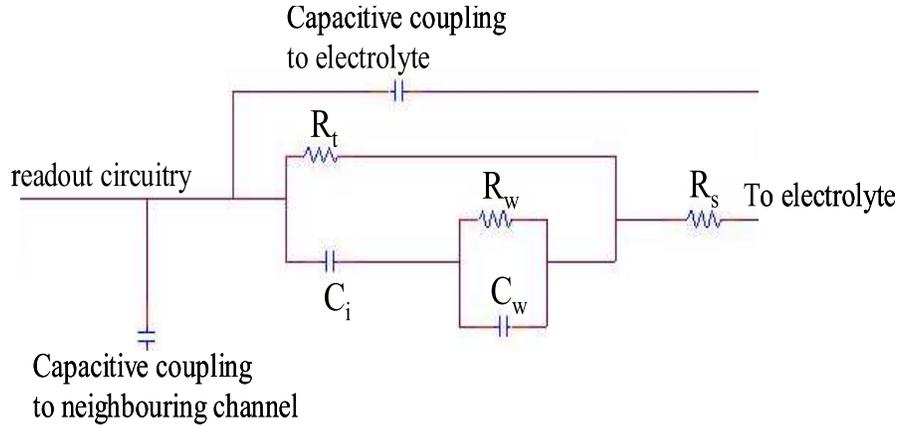


Figure 47: Schematic representation of the simulated circuit applied by Borkholder [73] and Mathieson et al. [79] in recent studies. C_i denotes the interfacial capacitance, R_t the charge transfer resistance, C_w and R_w the capacitance and resistance associated with the Warburg impedance and R_s the spreading resistance. (Image taken from Ref. [79])

The model proceeds from the description of a metal-electrolyte interface. An electrolyte is essentially any solution in which ions, rather than electrons in the case of a solid, are the principal conduit for charge transport. When a metal is introduced into such a medium, an electrical interface is created, and chemical reduction reactions begin. These in turn lead to development of an electric field across the interface, which drives opposite oxidation reactions and inhibits reduction reactions. This process continues until an equilibrium state arises. The electrolyte is affected by the field in a number of ways. Water dipoles from the solution form a *hydration sheath* at the metal surface. Solute ions then collect on top of this forming what is known as the *outer Helmholtz plane*. Some such ions may even adsorb onto the metal surface – this collection of such ions is known as the *inner Helmholtz plane*. In

this manner, an electrical double layer is formed at the interface. This is the

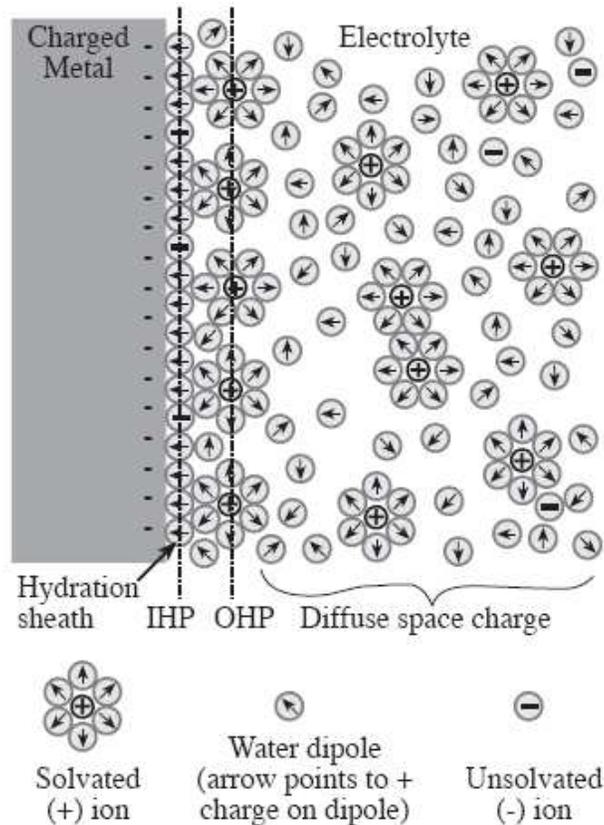


Figure 48: Diagram of the double layer capacitance created across the interface when an electrode is placed in an electrolyte solution.

foundation upon which Kovacs and Borkholder proceeded to build the model used for the electrode/electrolyte interface tested in this thesis.

The model circuit is driven as an AC circuit, usually at a particular frequency to simulate certain experimental conditions. However, for the purposes of gauging the accuracy of the model the frequency has been varied over a range from 100Hz–100kHz; this covers the range of interest for retinal studies (usually around 1 kHz). There are two main branches to the circuit.

The *interfacial capacitance* represents a layer of positive ions surrounding a negatively-charged metal interface, and is in inverse proportion to the surface area of the electrode. It was first modelled by Helmholtz in 1879 [83], who assumed that the metal surface in combination with the outer Helmholtz plane formed a simple capacitor with the water dipoles serving as the dielectric. Therefore the interfacial capacitance was governed by the relation:

$$C_H = \frac{\epsilon_0 \epsilon_r}{d_{OHP}}$$

with ϵ_0 denoting the permittivity of free space, ϵ_r the relative permittivity of the electrolyte, and d_{OHP} the distance of the outer Helmholtz plane from the electrode surface. The latter two are assumed to take the values of 78.5 Fm^{-2} and 0.5 nm respectively in physiological saline at 25°C . This however failed due to its assumption that charged solute ions were confined to the outer Helmholtz plane. Guoy and Chapman [84] accounted for this by defining an ion cloud where most of the charge was concentrated on the outer Helmholtz plane but decayed exponentially into the bulk of the solution. Thus the relation governing the capacitance became:

$$C_G = \frac{\epsilon_0 \epsilon_r}{L_D} \cosh\left(\frac{zV_0}{2V_t}\right)$$

with L_D denoting the Debye length, which is approximately 0.78 nm in an electrolyte of concentration 300 mMl^{-1} [76]. The hyperbolic cosine term compensates for the presence of mobile charges. Here z denotes the valence of the electrolyte ions, which can be taken to be 1 for physiological saline, V_0 denotes the potential at the electrode, assumed by Borkholder to be 50 mV , and V_t denotes the characteristic thermal voltage. This described the relationship better but tended to overestimate the value. The current

relationship was established when Stern [85] opted to combine the two approaches, both taking the outer Helmholtz plane to be a bound layer of ions and defining an ion cloud decaying into the bulk. In terms of the circuit, this is the equivalent of the two antecedent capacitances being added in series:

$$\frac{1}{C_i} = \frac{1}{C_H} + \frac{1}{C_G}$$

Resonant circuit theory requires that this have an impedance in inverse proportion to the signal frequency as well as the surface area of the electrode.

In parallel to this is a DC branch with impedance given by the *charge transfer resistance*. It is necessary for a resistive path to exist, as a DC potential applied across the interface can disturb the equilibrium and enable the flow of current. This potential difference is known as the *overpotential*. There are a number of contributions to this, each arising from a different process. The most important of these is the contribution from charge transfer, which gives rise to the charge transfer current density and associated resistance. This resistance is also in inverse proportion to the surface area of the electrode, and is generally measured experimentally in order to determine the current density. The relation governing the charge transfer resistance is expressed mathematically as:

$$R_t = \frac{V_t}{J_0 z}$$

where J_0 is the exchange current density, and V_t and z are defined as before. J_0 takes an approximate value of $8 \times 10^{-4} \text{ Acm}^{-2}$ from experimental studies of platinum hydrogen reactions [77]. Another important contribution arises from diffusion – if the current density becomes sufficiently large, reactants are no longer able to diffuse to the interface in time to participate in the

associated reduction and oxidation reactions. The associated impedance is known as the *Warburg impedance*. This represents a situation occurring at very high stimulation frequencies which fall outwith the frequency range used in these studies and so this impedance can therefore be neglected. In series with these is the *spreading resistance*, which is related to the spreading of current from the confines of the electrical circuitry into an extensive conducting solution. Unlike the two impedances detailed above, this depends not on the surface area but the geometric area of the electrode. The other series elements that have been incorporated into the model are the resistance of the physiological saline in which the electrode is immersed under experimental conditions, and the resistance of the retinal tissue sample itself. Capacitive couplings between channels have also been included.

Two types of interface can be defined from the exchange current density – *ideally non-polarisable* interfaces where the current density required to disturb equilibrium tends to infinity, and *ideally polarisable* ones where the current density required tends toward zero. Naturally, the studies documented in this thesis are concerned with those of the latter type.

The majority of parameters in the model are proportional in some manner to the surface or geometric area of the electrode, and consequently can be defined with respect to an *area-independent constant*, essentially the proportionality constant of the relationship. The surface area is defined as that of the electrode site after platinisation is complete and is usually around two orders of magnitude greater than the geometric area. The interfacial capacitance, C_i , taking the values of all the parameters to be as indicated above, can be expressed as:

$$C_i = 0.7 \text{ pF} \mu\text{m}^{-2} \times A_s$$

and the charge transfer resistance, R_t , as:

$$R_t = 3.3 \text{ G}\Omega \mu\text{m}^2/A_s$$

where A_s denotes the surface area of the electrode. The spreading resistance, R_s , is inversely proportional to the *geometric* area, A_g , of the electrode:

$$R_s = \frac{\rho \times \sqrt{\pi}}{4 \times \sqrt{A_g}}$$

where ρ denotes the resistivity of the electrode material, taken to be $72 \text{ } \Omega \text{ cm}$ for platinised electrodes. Using the above relations, the overall impedance may be modelled for electrodes of any size, as can be seen in Figs. 49 and 50. The general trend, however, is the same. Initially, at zero frequency, a state essentially corresponding to a DC input signal, the interfacial capacitor has, from resonant circuit theory, infinite impedance. The current must then necessarily flow through the DC branch containing the charge transfer resistor. As a result, the initial impedance is given by the sum of the impedances of the charge transfer resistor and the spreading and other resistors. As the frequency of the input signal is increased, the impedance of the interfacial capacitor is reduced, and more of the current begins to flow along the AC branch of the circuit, until a situation is reached where it is more energetically favourable for the current to flow along this route and so no further current flows through the DC branch. At high frequencies, the interfacial capacitor has negligible impedance and as a consequence the effective impedance is equivalent to that of the spreading and other resistors.

Diam. (μm)	C_i (nF)	R_t (k Ω)	R_s (k Ω)
5	1.4	1.65k	72
10	5.6	412	36
15	12.6	183	24
20	22.4	103	18
25	35.1	65.9	15
30	50.5	45.8	12
35	68.7	33.6	10
40	89.7	25.8	9
45	113	20.3	8
50	140	16.5	7.2

Figure 49: Table of the key parameters in the PSPICE model with respect to electrode diameter.

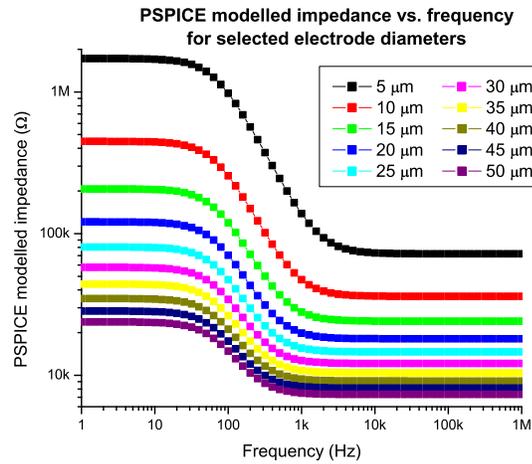


Figure 50: The relationships of the impedances modelled in PSPICE to the stimulation frequency for a selection of electrode diameters.

4.3 Testing and results

4.3.1 Testing environment

Before the IPIX II can be used in experiments, it must be properly characterised. The characterisation work documented in this thesis centred around two main components of the system: the neural network and the VCO, with the latter operated in *C3Sel*, *HighI* and *LowI* modes.

The IPIX II is controlled by a LabVIEW VI, designed by Dr. Keith Math-

ieson, with the layout shown in Fig. 51. There are three main blocks of controls which are of interest in the testing and characterisation of the chip, which can be seen above the readout screen. The first block, on the far left, contains the controls for the analogue input lines, each of which are controlled by faders. The first two on the left are *PveCntrlV* and *NveCntrlV*, which control the independent variation of the positive and negative amplitudes of the pulses generated by the VCO. The next two are *Hth* and *Lth*, which correspond to the upper and lower thresholds on the window comparator of the VCO circuit. These were set to 2.5 V and 0.55 V and were not altered at any stage of the testing. The last fader on the right of the panel governs the amplitude of the test signal when this function is activated. In addition to the faders, there are also digital input fields for each parameter to enable accurate fine tuning.

The next block contains an array of switches which control the writing of the shift register, an operation which takes place at the activation of the chip. The bottom row is an array of five switches, each of which corresponds to one of the five bits of the shift register, the test signal and four neural network weights. Above these is another switch which enables or disables the shift register writing operation. When this is on, a new ASCII file is written which forms the input to the shift register. It is uploaded to the chip when the *SRin* signal is applied. Ones and zeroes are written to the appropriate sections of the file, depending on whether the test signal or particular combinations of neural network weights are activated. In situations where it is desired to alter the state of individual pixels, the switch is left off, and the text file may instead be manually edited by entering ones and zeroes in the appropriate

parts of the file.

The third block, on the right, deals with digital control settings. It is here that modes such as *C3Sel*, *HighI/LowI*, and *ZeroSel* may be activated. There is also an electroplate mode to allow renewal of the electrodes during extended experimental studies.

The readout screen below the controls shows the two output signals from pixel 456, which is the pixel left as a test pixel and has both of these outputs connected to the system (NI device).

The PCB on which the chip is mounted was enclosed in a light-tight box with a camera lens directly above the chip as the only available entrance for light. The light source is a standard desk lamp. The position of this, as well as the settings of the camera lens, must be adjusted prior to every experimental session involving light sensitivity in order to ensure the system is calibrated properly.

4.3.2 Neural network linearity

The on-chip, local neural network operates by subtracting a weighted sum of the currents generated by six neighbouring pixels from that generated in a given pixel and outputting this sum to the VCO circuit. This mimics the lateral inhibition operation performed by the horizontal cells within the retina. This can be visualised by considering the behaviour of the array under uniform light with the $\frac{1}{6}$ weight enabled. Under such circumstances, the currents generated by the six neighbouring pixels are the same, to within a reasonable margin, as that generated by the centre pixel. As a result, be-

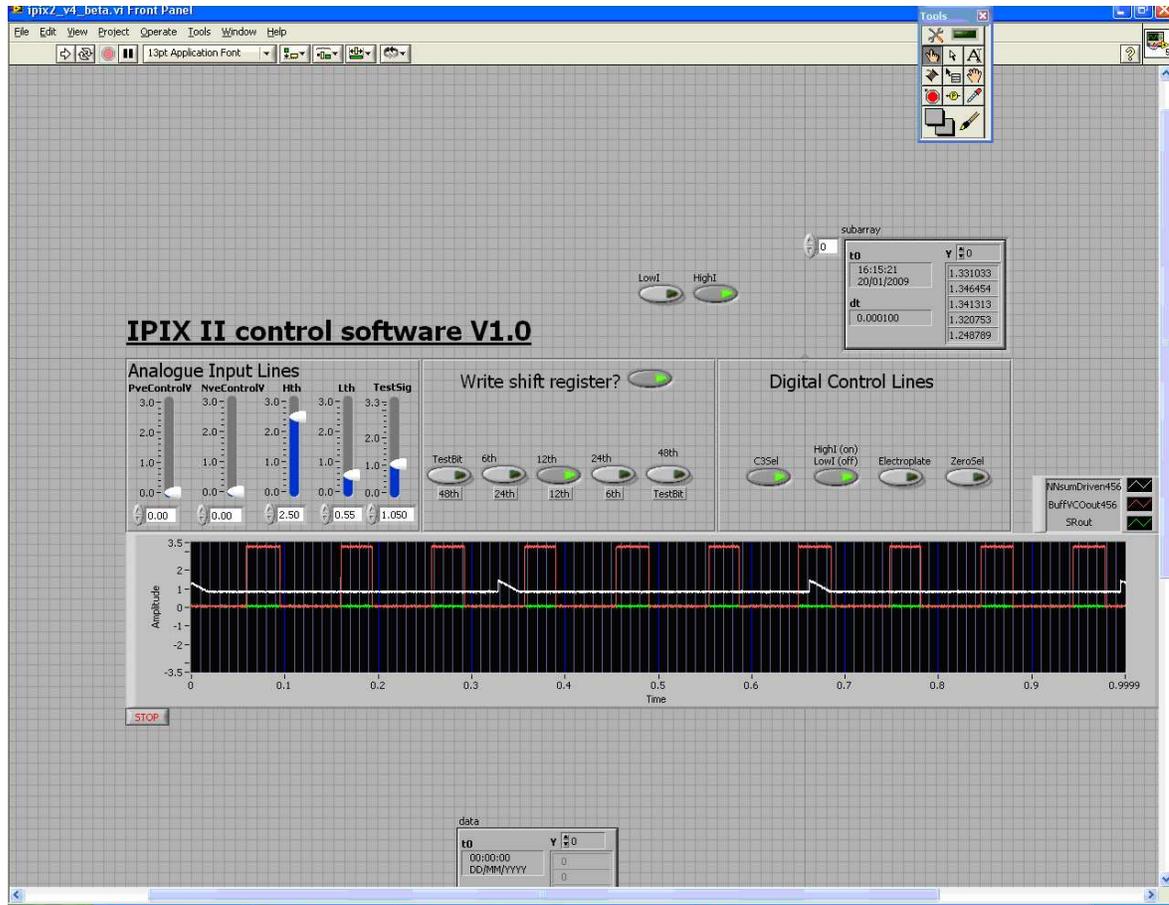


Figure 51: Diagram of the LabVIEW VI used in the testing and characterisation of the IPIX II chip, showing all the key controls and readouts.

cause the neighbouring values are being summed in sixths, they sum to zero, mirroring the behaviour of the horizontal cells in the retina under similar conditions. If any are different, however, the sum will not be zero and an edge or boundary effect will be created. The weight applied to the neural network determines how sharp the edge or boundary is. With the $\frac{1}{48}$ weight enabled, for example, the summation will produce non-zero results and a much softer boundary will result.

This behaviour also requires to be investigated with respect to the intensity of the incident light. To this end, a set of neutral density (ND) filters were employed in various combinations to produce step changes in the intensity of the light delivered to the chip. The filters used had optical density factors of 0.26, 0.6 and 0.9. The light intensity produced by each ND filter or combination thereof was then calculated as a fraction of the unfiltered light intensity. For the purpose of these experiments the chip was placed in a light-tight box with a camera lens mounted on top through which light may be transmitted. The setup was then calibrated by exposure to the ambient light of the laboratory in order to simulate high light conditions and give a 100 Hz stimulation rate, the maximum under the IPIX II specifications. The ND filters were then introduced in order to attenuate the light and thereby simulate the full range of light conditions within which the eye operates.

The behaviour that should be expected is that as the incident light intensity decreases, the frequency of oscillation of the VCO should decrease in a corresponding fashion, with the two parameters linked in a roughly linear relationship. Taking into consideration the neural network weights, the $\frac{1}{6}$ should give the lowest frequency output as it represents the largest amount being subtracted off the pixel output. Other combinations of weights subtract smaller fractions of the neighbouring pixel currents and thus the output of the VCO should be larger. In these experiments, the test signal is set low in order that the light response of the centre pixel dominates. This is to ensure clear distinction between the two modes of operation and ease of identification of the trends exhibited.

Series of measurements were undertaken for the combinations of ND filters

listed above, as well as for the following combinations of neural network weighting factors: $\frac{1}{24} + \frac{1}{12}$, $\frac{1}{48} + \frac{1}{12}$, $\frac{1}{12}$, $\frac{1}{48} + \frac{1}{24}$, $\frac{1}{24}$, and $\frac{1}{48}$. The behaviour of the parameters can be observed in Fig. 52. Other series of measurements were taken for the same parameters but with the *C3Sel* mode activated on the chip. The behaviour of the parameters for this can be observed in Fig. 53. One hundred measurements were taken for each brightness level using a digital oscilloscope that calculates the mean and standard deviation to a high accuracy. The results are as shown in Fig. 54. What can be observed in Fig. 52 is that as different weights or combinations of weights are activated the response of the VCO is indeed reduced to different degrees, and the reduction is linearly dependent on the weight applied. This is clear when the variation of the intercepts and gradients are subjected to a linear fit. There is a clear linear trend in both parameters. The intercept represents the minimum frequency of the VCO under a particular weight of the neural network. This decreases as the weight is increased, for instance with a $\frac{1}{6}$ enabled on all six neighbouring pixels. For ease of inspection of the behaviour of the chip, all the pixels on the chip were set for the same neural network weighting factor with only the test pixel being set for light sensitivity. The minimum frequency is as low as it gets as this represents the largest amount being subtracted from the signal. With $\frac{1}{48}$, the lowest available weight, enabled, the minimum frequency is a maximum under weight. The only higher maximum is that under complete disablement of the neural network with nothing being subtracted. The gradient of the linear fit also follows a linear trend. Again, this trend is a decay. What this means is that as a larger weight is enabled and, as noted above, more is being subtracted from the output of the centre

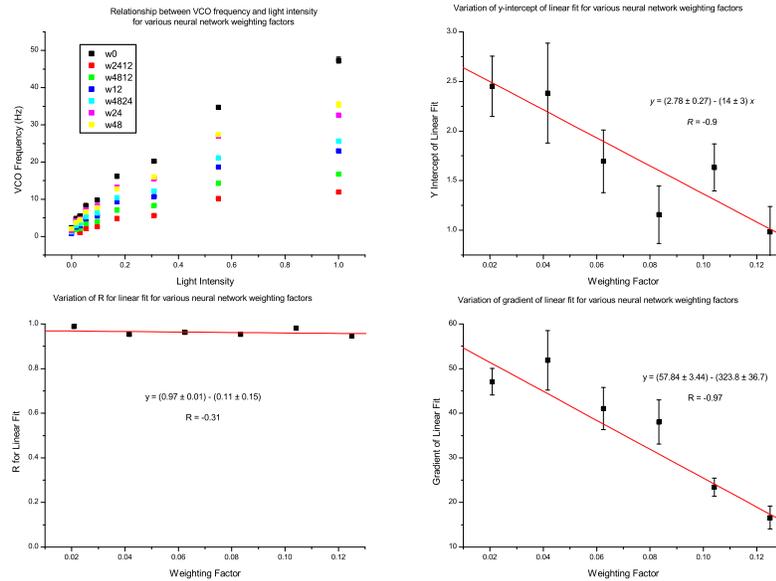


Figure 52: *Top left:* The behaviour of the VCO frequency vs. light intensity for selected combinations of neural network weights. *Top right:* Behaviour of the minimum frequency of the VCO vs. neural network weight for the selected combinations. *Bottom right:* Behaviour of the gradient of VCO frequency with respect to light intensity vs. neural network weight for the selected combinations. *Bottom left:* Graph of the correlation coefficient for the linear fits performed on the data, verifying reliability of fit.

pixel, the effect created tends more towards a shallowing of the increase of frequency with respect to light intensity. So the lowest gradient corresponds to the activation of the $\frac{1}{6}$, while the highest, other than when the neural network is not in operation, is that corresponding to the activation of the $\frac{1}{48}$.

Also presented is the behaviour of the correlation coefficient, which is used to gauge how well the fit describes the data. As can be noted, this is relatively

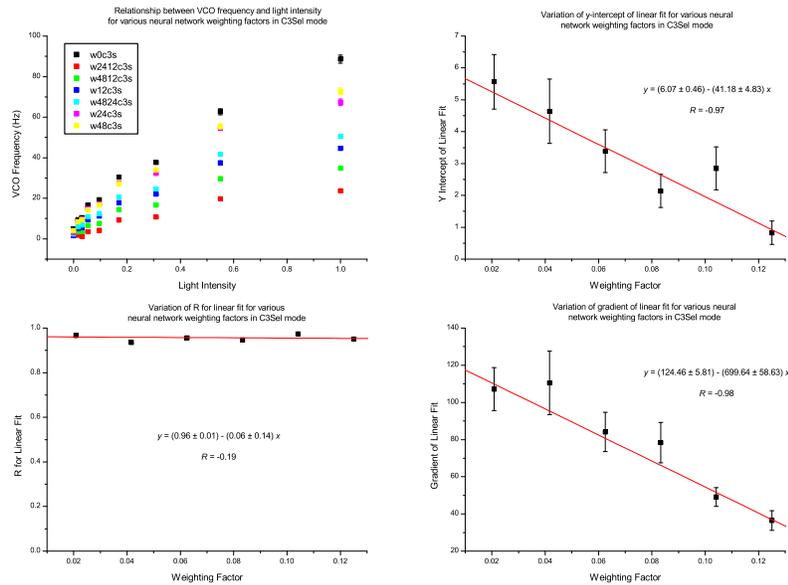


Figure 53: *Top left:* The behaviour of the VCO frequency vs. light intensity for selected combinations of neural network weights in *C3Sel* mode. *Top right:* Behaviour of the minimum frequency of the VCO vs. neural network weight for the selected combinations. *Bottom right:* Behaviour of the gradient of VCO frequency with respect to light intensity vs. neural network weight for the selected combinations. *Bottom left:* Graph of the correlation coefficient for the linear fits performed on the data, verifying reliability of fit.

constant for all weight modes of the neural network, suggesting that the fits give a reliable description of the behaviour of the system.

Finally, the same is presented with *C3Sel* mode activated in Fig. 53. It is clear what should be expected with *C3Sel* mode activated. In this state, the C3 capacitance is doubled, thereby halving the step size as defined in the previous section, and doubling the maximum and intermediate frequencies

Weight	y -intercept	Gradient	R
0	3.14 ± 0.47	64.42 ± 7.43	0.96
$\frac{1}{24} + \frac{1}{12}$	0.98 ± 0.26	16.63 ± 2.55	0.95
$\frac{1}{48} + \frac{1}{12}$	1.63 ± 0.24	23.43 ± 2.02	0.98
$\frac{1}{12}$	1.16 ± 0.29	38.03 ± 4.94	0.95
$\frac{1}{48} + \frac{1}{24}$	1.70 ± 0.32	41.06 ± 4.73	0.96
$\frac{1}{24}$	2.38 ± 0.50	51.84 ± 6.68	0.95
$\frac{1}{48}$	2.45 ± 0.31	47.07 ± 2.98	0.99

Figure 54: Table of the y -intercepts and gradients for linear behaviour of VCO frequency for selected neural network weights.

compared with the corresponding behaviour in *C3Sel*-inactive mode.

4.3.3 VCO behaviour under test signal

The next component to be tested was the VCO, which was characterised by means of varying the test signal – controlled by the *TestSig* fader on the LabVIEW VI (see Fig. 51). It is known that the voltage across the pixel varies from 3.3 V under conditions of darkness to a saturation point of around 1 V and so the test signal amplitude was varied over this range, both in *C3Sel* mode and otherwise. The behaviour of the frequency output of the VCO is as shown in Fig. 55. As expected, the frequency of oscillation is a maximum for the saturation voltage of 1 V, and linearly decays up to the power supply rail voltage of 3.3 V.

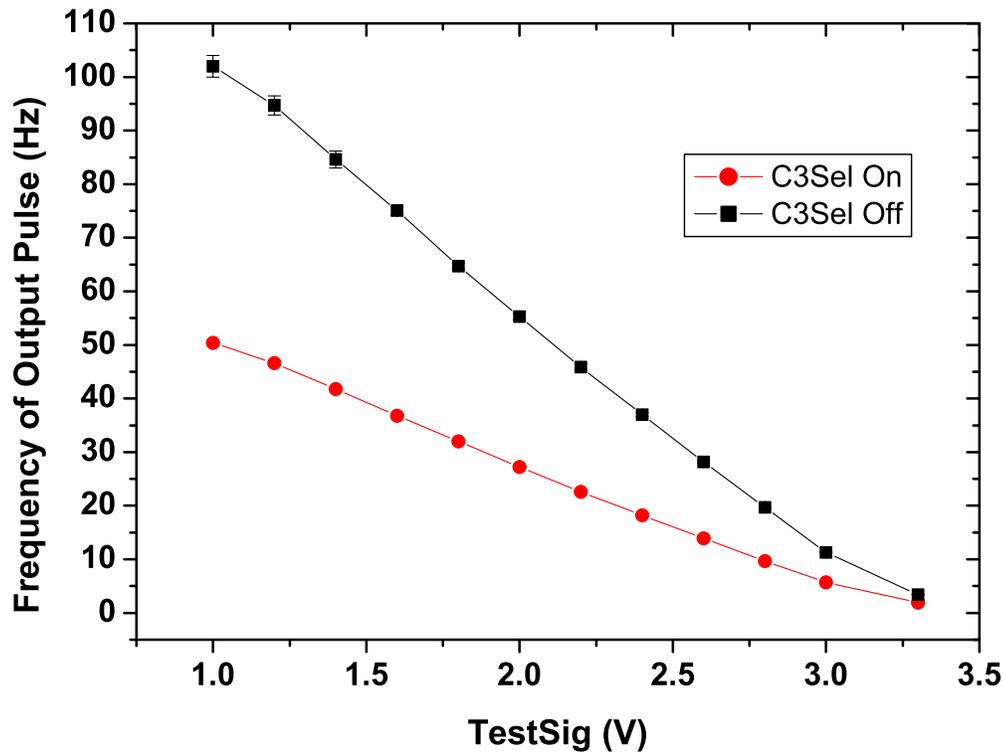


Figure 55: Graph of the behaviour of the output pulse frequency of the VCO with respect to test signal voltage.

4.3.4 Biphasic output

The IPIX II output is designed to take the form of a biphasic pulse, with an interphasic delay of $100 \mu\text{s}$ separating the cathodic and anodic phases. Due to the potential medical issues that can arise from depositing net charges into the retina, as previously indicated, balancing of the charges in the stimulus pulse is very important. So, as noted above, the IPIX II design incorporates the ability to adjust both the positive and negative amplitudes of the pulse independently in order that such balance may be achieved under *in situ*

experimental conditions.

The two modes in which the chip can operate are directly related to the two current mirrors built into the VCO output circuitry. Independent control of $PveCntrlV$ and $NveCntrlV$ enables the user to fix the base current used by these mirrors. The current is reduced to 5% of its previously generated value in *LowI* mode. As noted in the previous section, the output current is limited by the RC behaviour of the electrodes, particularly in *HighI* mode when the voltage across the resistance of the interface becomes limited by that of the power supply. The behaviour of the output current requires to be gauged for the full range of applied control voltages. Both voltages can be varied from 0 V to a maximum of 3.3 V and the output current measured. The resistance of the interface is simulated by a circuit on the PCB that mirrors the model circuit described in the previous section, with an interfacial capacitance of 1 pF, a charge transfer resistance of 1.6 M Ω and a spreading resistance of 70 k Ω . A resistor is then connected across the output of the circuit. For the purposes of the experiment different values of resistor were connected depending on the mode in which the chip was being operated - a 100 k Ω resistor was used for *HighI* mode and a 1 M Ω resistor for *LowI* mode. By measuring the voltage drop across this resistor the current being pushed through the circuit may be determined by the application of Ohm's Law.

The expected behaviour is that at low voltages current should be limited (in either mode, though less so in *LowI* mode) and should then exhibit a linear decay across a particular range of voltages fixed by the resistance of the interface. Plateaus should also be observed at low voltages as another basic manifestation of Ohm's Law – the resistance of the interface remains

Control Voltage	HighI Output (simulated)	LowI Output (simulated)	HighI Output (measured)	LowI Output (measured)
1 V	30 μA	1.5 μA	14.5 μA	1.25 μA
2.5 V	1 μA	100 nA	1.89 μA	121 nA

Figure 56: Table of the ranges over which the current is expected to behave linearly in the two modes of the IPIX II.

constant, therefore the current driven through it is large at low voltages, and the circuit should be unable to deliver this current. In theory, if the resistance of the interface were less, then the linear decays seen in the figures would begin from substantially larger maxima than are observed. At higher voltages, the current becomes negligible. Again, this is consistent with a high resistance.

Indications of operational range for the IPIX II are given in Fig. 56. It should be noted that those given for *HighI* in the table correspond to the use of a 47 k Ω resistor. These represent the ranges over which the current is expected to behave in a linear fashion.

The behaviour exhibited is a linear decay in both directions over a range of about 1 to 2.5 V as can be seen in Figs. 57 and 58. It can be seen by applying linear fits in both directions and summing the two waveforms together that the charge balance was kept very much in place during the experimental run, with only a slight variation in gradient in both directions, pointing to a slight imbalance on the order of less than 100 nA, as shown in Figs. 59 and 60.

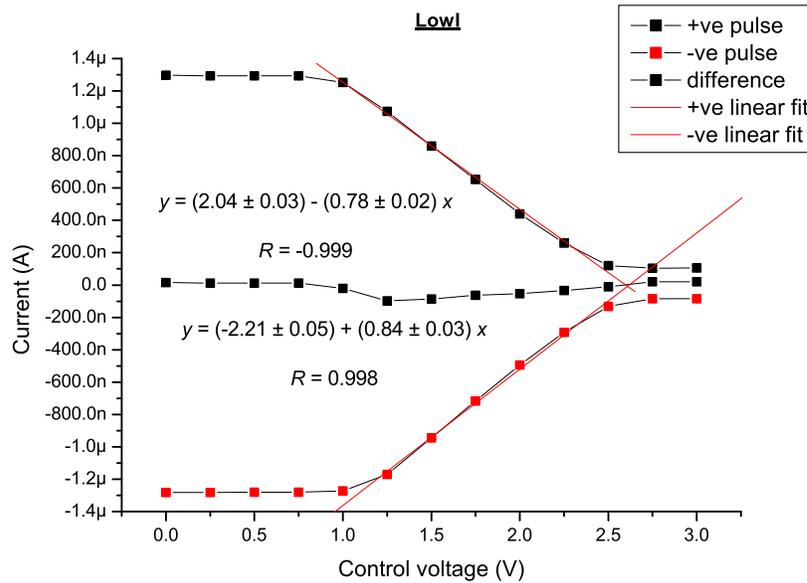


Figure 57: Behaviour of the output current in *LowI* mode with respect to both positive and negative amplitudes.

An important strength of the IPIX II's design is the incorporation of “zero selection” in the biphasic output circuitry. The importance of this stems from the RC characteristics of the electrode simulation circuit. If after the cathodic portion of the stimulus pulse is transmitted there is no “zero selection” then upon cessation of the phase, and onset of the interphase delay, the delay period is taken up by the RC decay of the amplitude. This also occurs after the anodic phase.

This is not an acceptable situation as it casts doubt on the symmetry of the pulse and leads to greater charge and energy deposit than necessary. With zero selection in place, there is an active switch from the desired amplitude to ground, providing a falling edge to the pulse that matches the rising one at the beginning. The *ZeroSel* mode incorporated into the chip is designed

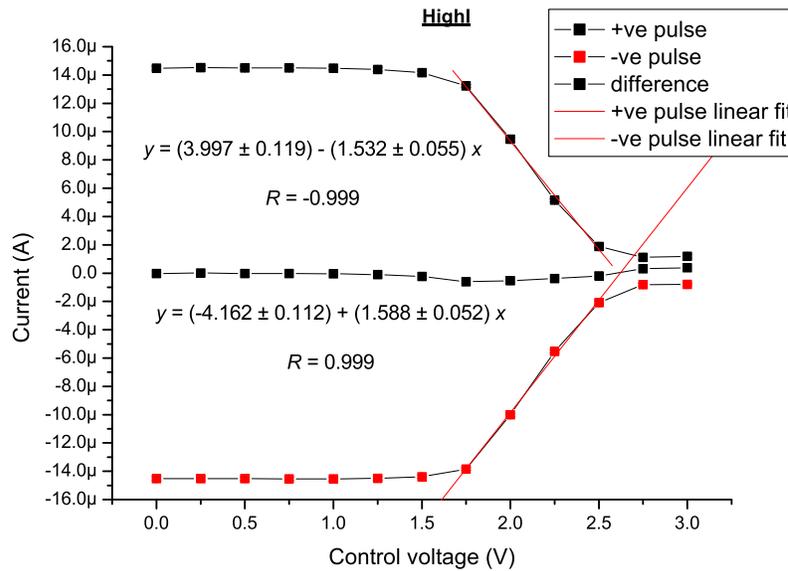


Figure 58: Behaviour of the output current in *HighI* mode with respect to both positive and negative amplitudes.

to provide such a mechanism. Figs. 61 and 62 show the effect of operating the chip in this mode, with Fig. 62 demonstrating the improvement it makes to the quality of the stimulus pulses generated by comparing the output with that corresponding to the absence of zero selection.

4.3.5 Comparison of electrode-electrolyte interface model with experiment

The validity of the model was verified by comparison with measured data. Two such comparisons were made. The first was made with impedance data for 7 μm -diameter electrodes from an array fabricated for a different study. As Fig. 63 shows, the experimental and modelled data are in good agreement. The second, and more important, comparison was with the data for the

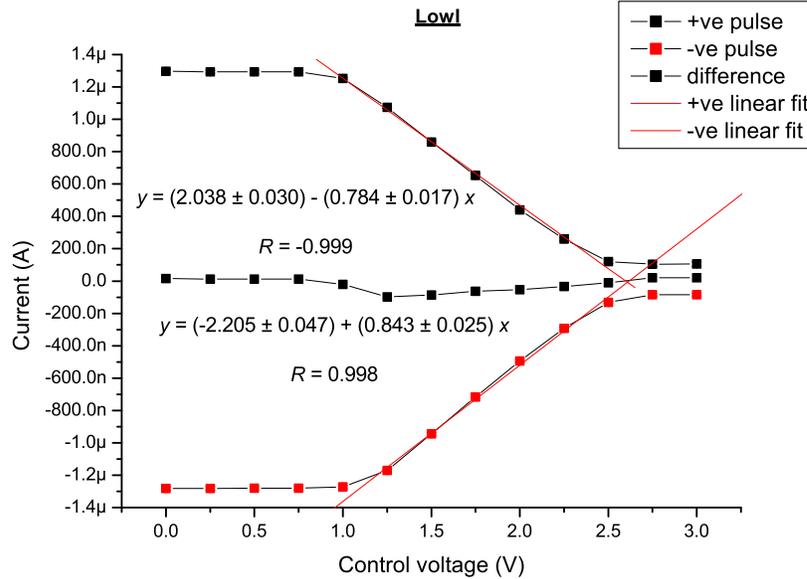


Figure 59: Behaviour of the output current in *LowI* mode with respect to both positive and negative amplitudes, showing linear behaviour of the current in the specified range and the minimal charge imbalance.

electrode arrays prepared for this study. A 16-electrode array from one of the wafers was tested using the LCR meter as before, to gauge the impedance of its electrodes over frequencies from 100 Hz–100 kHz. All electrodes on the array were tested. The agreement of the data with the model is shown in Fig. 64. The decrease of the impedance below that predicted by the model at higher frequencies is possibly due to variations in the actual diameters of the vias.

4.3.6 Extensions of model

The circuit shown in Fig. 47 was modified in an attempt to ensure its output adhered more closely to the experimental data. Most important among

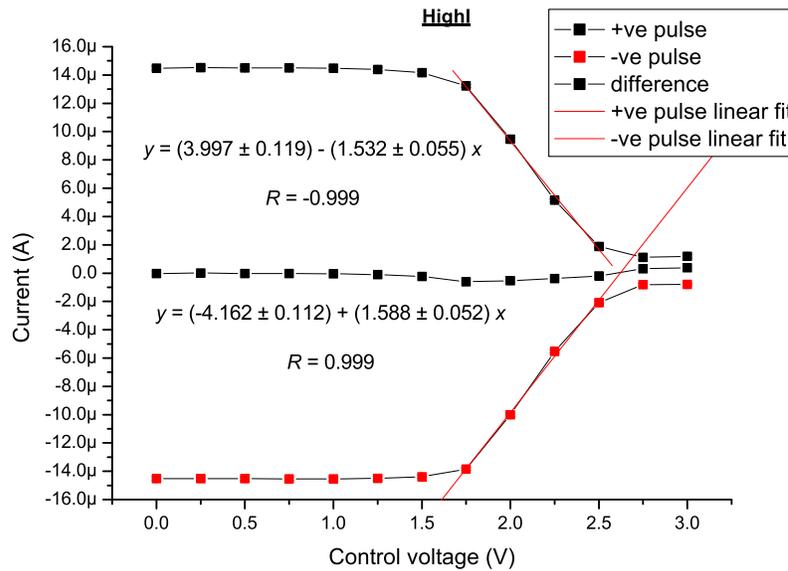


Figure 60: Behaviour of the output current in *HighI* mode with respect to both positive and negative amplitudes, showing linear behaviour of the current in the specified range and the minimal charge imbalance.

the modifications are the inclusion of capacitive and resistive components for the pulse generator representing the IPIX and a circuit element representing the probe in series with the electrode, with parameters taken from the probe specifications. Other components, such as the interfacial capacitance and the charge transfer resistance were varied slightly but kept in proportion as much as possible. The latest graph is shown in Fig. 65, and the modified circuit diagram shown in Fig. 66. The shape of this graph is now very similar to that of the actual electrode output.

The model was applied to lay the foundations for experimental work in the latter stages of the investigation documented in this thesis. Two important theoretical relationships were deduced which will require to be tested exper-

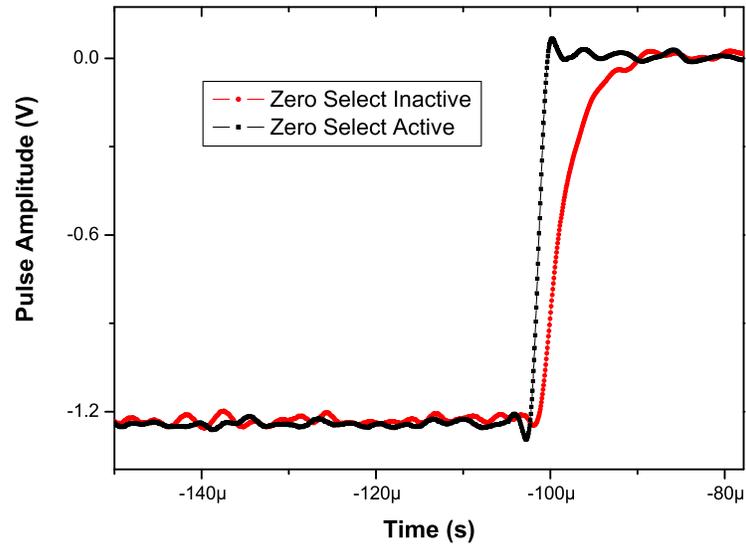


Figure 61: Graph of a typical IPIX II stimulus waveform, with *ZeroSel* mode disabled (red trace). Note that the amplitude is not zeroed during the inter-phase delay as with *ZeroSel* mode enabled (black trace).

imentally in future studies.

The first relationship was that of the input pulse voltage to the electrode diameter, as shown in Fig. 67. The input pulse voltage is the amplitude of the pulse that is initially fed into the circuit, and as the threshold charge density was determined to be around 2.25 nC per phase by Palanker et al [82], the analysis was aimed at investigating the voltages required to produce such a charge density. As expected, since impedance scales inversely with electrode surface area, the simulated voltage increased significantly for smaller electrodes, reaching a maximum of around 4 V. There was less significant variation for larger electrodes, from around 750 mV for a 20 μ m-diameter electrode to around 500 mV for a 50 μ m-diameter electrode.

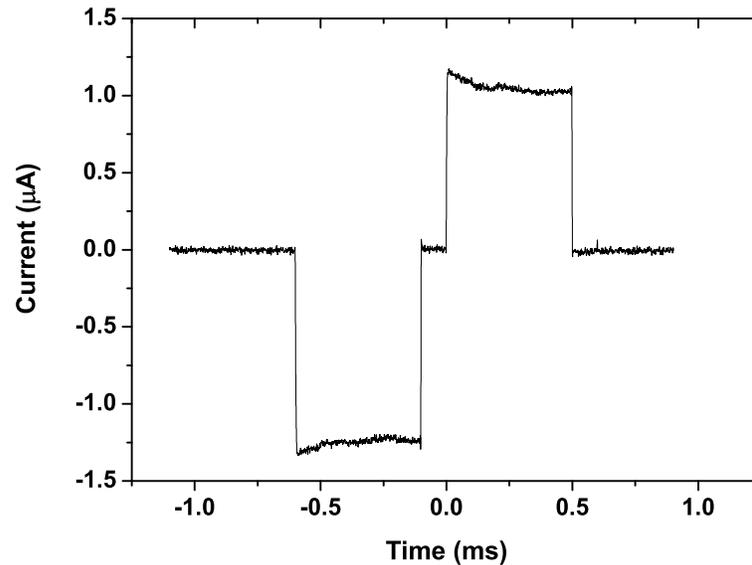


Figure 62: Graph of a typical IPIX II stimulus waveform, with *ZeroSel* mode enabled. Note that the amplitude is now zeroed during the interphase delay.

Another important phenomenon that manifests itself is that of *charge imbalance* — this occurs as a result of the time-dependent behaviour introduced by the parallel connection of the interfacial capacitor and the charge transfer resistor. It would not be expected under normal circumstances for there to be any imbalance because the modelled circuit is effectively a *differentiator* and should produce an output which integrates to zero for a square wave input. However, the exponential dependence in time introduced by the RC time-constant of the simulated circuit (which is around 2.3 ms for all configurations of the system) suggests that an imbalance is introduced for input signals whose periods are not significantly longer than this. In fact, physically, the nature of such behaviour means that in order to minimise such imbalance, at least five such periods should be allowed to lapse before the

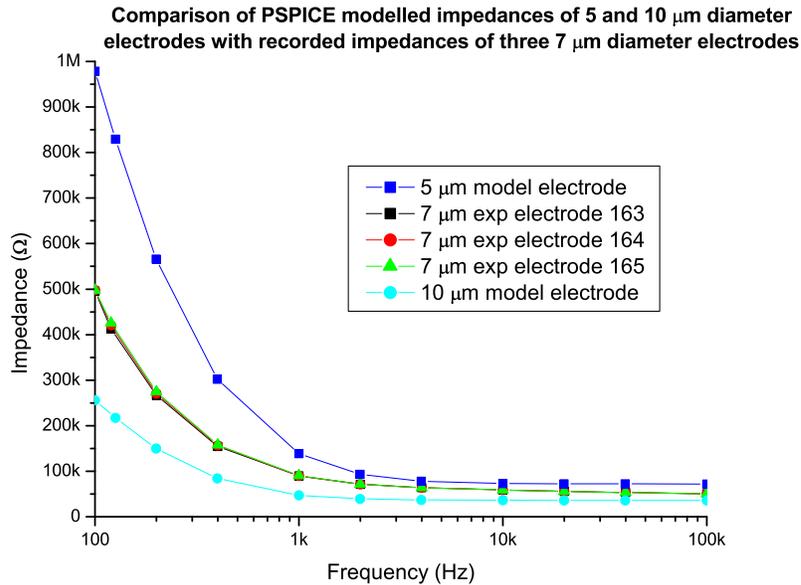


Figure 63: Comparison of modelled impedances with those of 7 μm -diameter electrodes from a different study.

transmission of another pulse, which would place an intrinsic limit on the range of frequencies over which trains of input pulses can be transmitted. Another factor which limits the frequency of stimulation is the refractory period of retinal ganglion cells, which generally ranges from 3 to 10 ms, thereby limiting the range of effective stimulation frequencies to a maximum of 100 Hz.

Fig. 68 shows the precise variation of the charge imbalance as a deposit per cycle. As the frequency increases, the system is being given progressively less time to return to its zero potential, resulting in a corresponding increase in charge imbalance per cycle. The imbalance levels off as the frequency approaches 10 kHz because the pulses used in the simulations had a duration of 0.1 ms, so for frequencies of 10 kHz and above the input is simply a DC

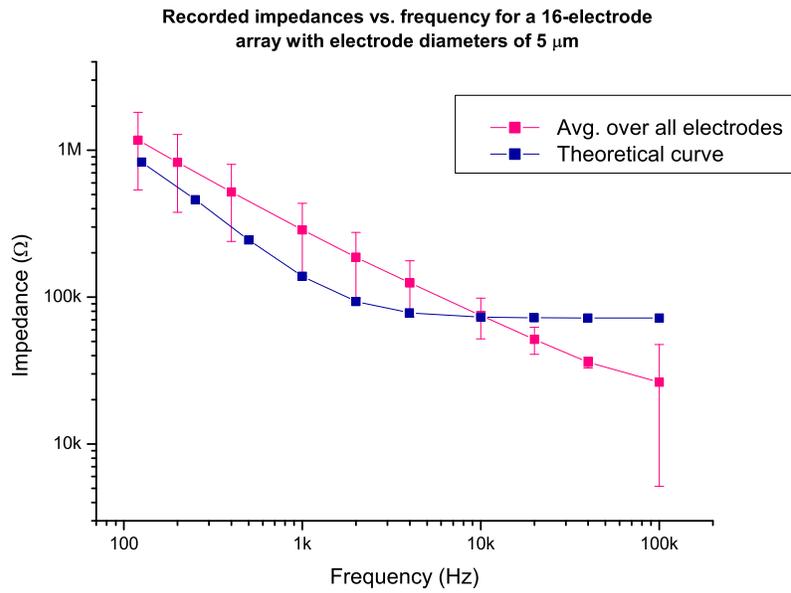


Figure 64: Comparison of mean impedance recorded for a 16-electrode array with electrodes of 5 μm diameter.

offset, and the charge imbalance must necessarily therefore be constant. It should also be noted that the charge imbalance introduced increases with the diameter of the electrode. It varies from 23 pC for a 5 μm electrode to approximately 1.2 nC for a 50 μm electrode at a frequency of 125 Hz, and from 266 pC for a 5 μm electrode to around 2.2 nC for a 50 μm electrode at 8 kHz, which was the highest frequency tested.

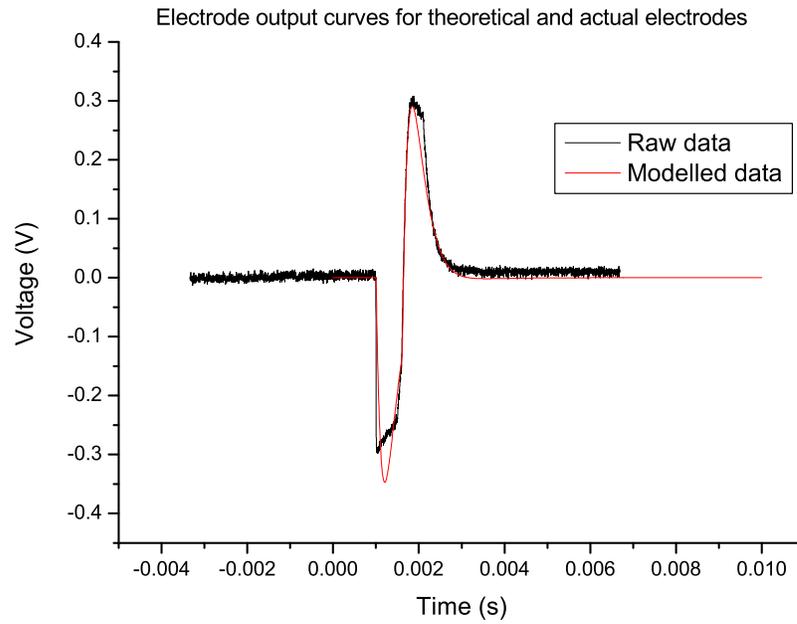


Figure 65: The relationship of the electrode output voltage modelled in PSPICE to that of a real electrode.

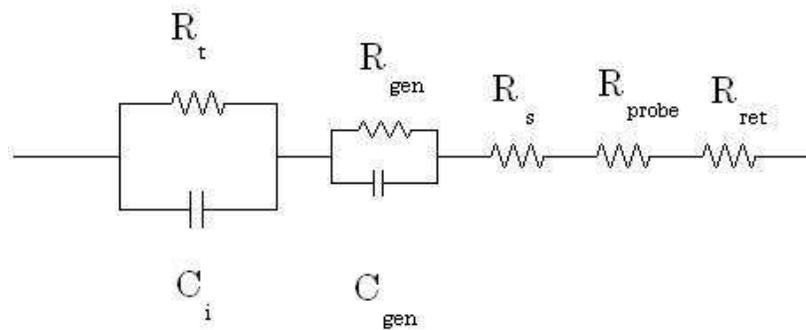


Figure 66: The modified model circuit showing the new components. R_t , C_i and r_s are defined as before, R_{gen} and C_{gen} are the components associated with the pulse generator, R_{probe} represents the resistance of the probe, and R_{ret} represents the resistance of the retina.

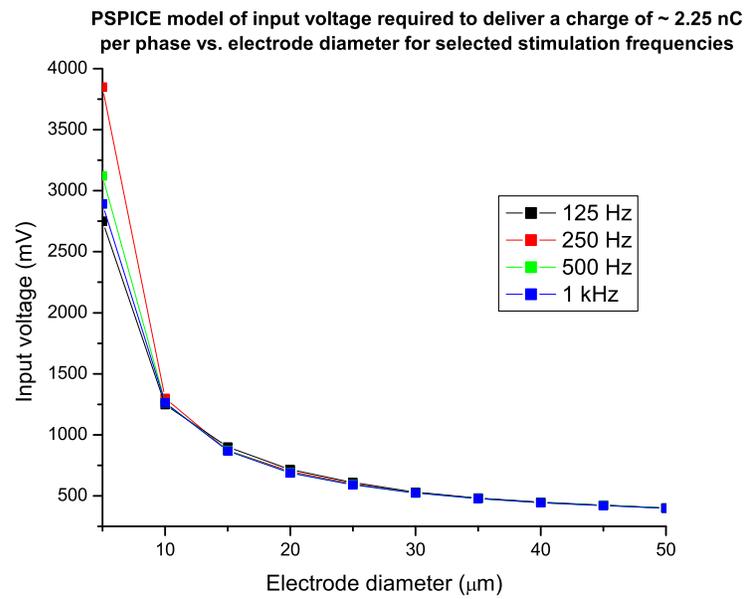


Figure 67: Plot of the simulated variation of the input pulse voltage with respect to electrode diameter.

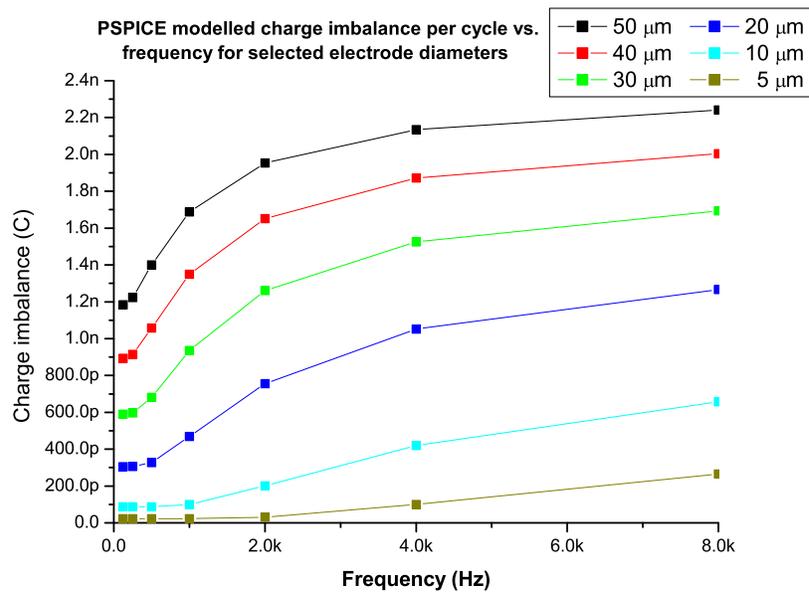


Figure 68: Plot of the simulated variation of charge imbalance with respect to frequency for selected electrode diameters.

5 Retinal experiments

In order for an implantable retinal prosthesis to be properly realised, the design must be optimised in order to ensure that its operation does not cause undue harm to the recipient. As indicated in the opening chapter, the deposition of excessive amounts of energy into the retina can cause unwanted heating effects and potential medical side-effects, making it essential that electrode geometry and consequently energy deposition are optimised such that no more is deposited than is necessary. It is to this end that the data analyses documented in the remainder of this thesis are directed. An overview is given first of the software created for the purpose of these analyses, followed by a description of the tests, including a proof-of-concept analysis, to which they were subjected in order to guarantee their functionality.

5.1 Experimental setup

The experiments outlined here were carried out in the Faculty of Biomedical and Life Sciences (FBLS) under the supervision of Dr. J. D. Morrison. The experimental subject is a terminally anaesthetised frog — frogs are used because they provide a good model in terms of the size of their neurons and so the electrical thresholds involved in their stimulation should be similar. After anaesthesia, the cornea is removed and the aqueous humour absorbed, and then the lens removed to expose the retina. The electrode array is then advanced into the eyecup until it is in contact with the ganglion cell layer, a status verified by light stimulation [103]. In experiments using the 16-electrode arrays, stimulation pulses were fed into the system via one electrode on the array using a precision pulse generator to induce the retinal cells

to send signals up the optic nerve. These responses were then recorded by the remaining electrodes and traced graphically by a Java program produced specifically for the purposes of the experiments [104]. It was these traces which were fed into the C program detailed in the previous chapter for analysis. In the experiment documented in this chapter, a new arrangement was employed in order to more closely mimic in experiments the behaviour of the proposed retinal implant under *in vivo* conditions. Within this arrangement, a 61-electrode flexible array was used, with a selection of electrodes being used as stimulus electrodes, and a further selection used as recording electrodes. Eight electrodes were chosen to serve as stimulus electrodes, and these were connected to the biphasic output of the IPIX II. A further sixteen electrodes were chosen to serve as recording electrodes. This was done because the apparatus housed within the IBLs laboratory only incorporates that number of pre-amplifiers. These electrodes were connected to the data acquisition system.

The experimental setup falls into three clear sections. There is the experimental apparatus itself, which consists of a cradle in which the frog is placed during the course of the experiment, as well as a tungsten wire which acts as a independent recording which enables the performance of the system to be verified, and an additional wire which acts as a reference electrode – this was connected into the earth of the IPIX II in order to earth the entire system. These are both also advanced into the eyecup prior to experimental procedures. The tungsten wire usually forms the first channel in data files generated by the Java suite detailed in the previous chapter.

The other two stages are the electronic hardware and the software which

governs its operation. The experimental apparatus is controlled by two PCs. The first of these is to govern the operation of the IPIX II chip which, for the purposes of these experiments, was used to externally trigger the data acquisition routines of the Java suite, and to deliver stimulus pulses of the appropriate profile to the retinal sample via the VCO and biphasic output. The return electrode of the experimental apparatus is also, as indicated above, connected to the ground of the IPIX II, thereby earthing the entire system. The external trigger for the data acquisition system is provided by means of the signal *Update*. The biphasic output of the IPIX II is connected to those electrodes on the electrode array which have been designated as stimulus electrodes – these in turn deliver the stimulus pulses to the frog retina.

Once the external trigger is received by the data acquisition system, the system begins recording the data transmitted through the recording electrodes. The data is then logged in the Java suite as documented in the previous chapter, and initially stored in the PC running the suite. From there it is exported to the other PC running the IPIX and on to the Group's storage within the file servers of the Department of Physics & Astronomy.

5.2 Java runtime environment and data selection

The data sets used in the analyses documented in the ensuing chapters were collected by means of a Java software suite written by A. R. Moodie. It combines facilities for both the recording and viewing of data, and allows data from retinal experiments to be easily exported for more rigorous analysis.

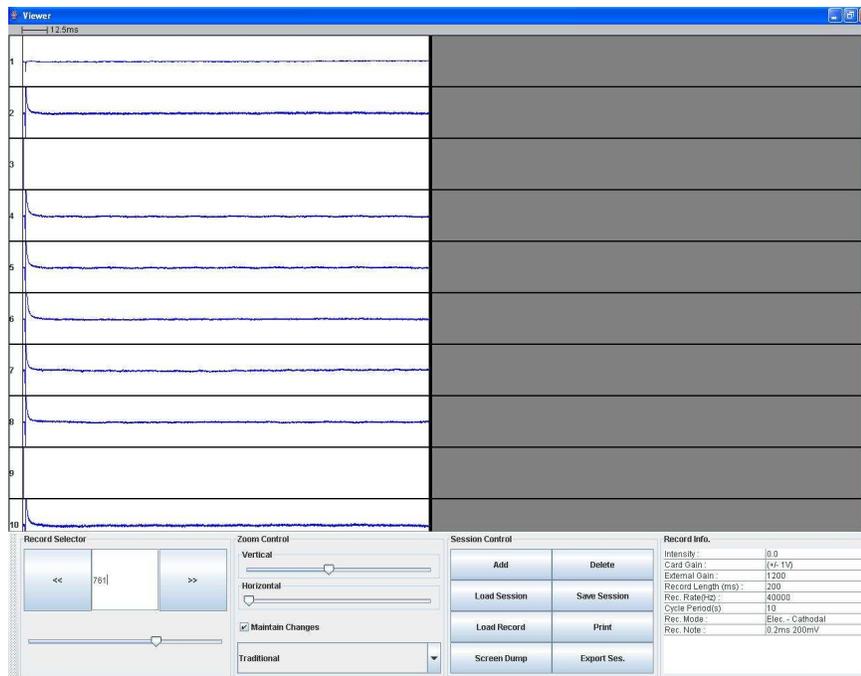


Figure 69: Screenshot of the Java viewer showing a typical data set.

A record in the data set as processed by the Java environment takes the form shown in Fig. 69. When setting up a recording run, a number of pieces of information initially need to be supplied in a dialog box. Chief among these are the number of channels, the sampling rate, the peak-to-peak voltage and the sample length. The number of channels is normally set to seventeen, to allow for the presence of a sixteen-electrode array and a reference electrode within the experimental setup. The purpose of the reference electrode is to provide an independent control sample against which the behaviour of the system may be verified. The sampling rate was set at 40 kHz for the proof-of-concept analysis documented in this chapter, with a record length of 200 ms. Once these parameters have been specified, a data acquisition screen is reached, as shown in Fig. 70. On this screen, the individual data channels are shown stacked vertically.

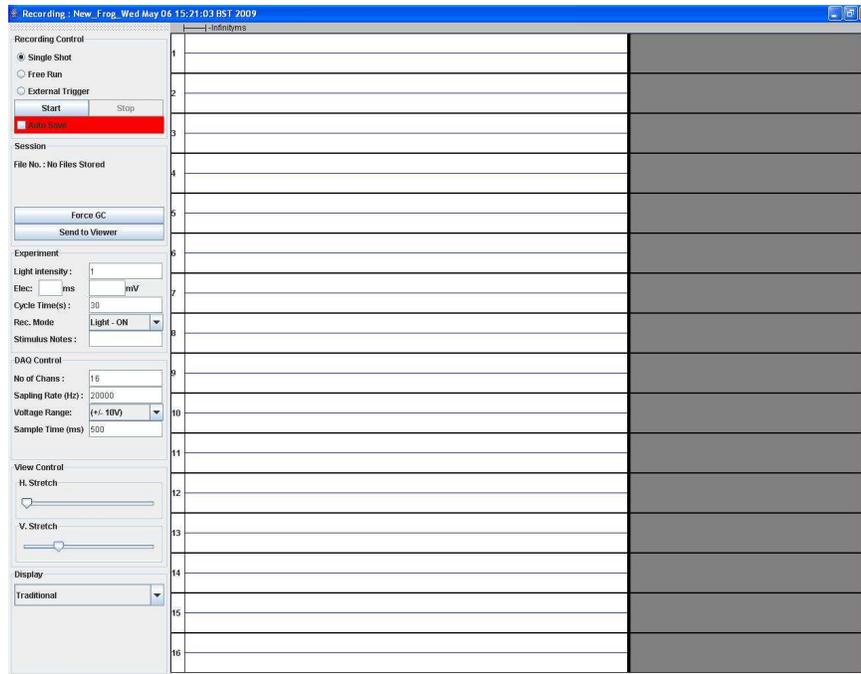


Figure 70: Screen capture of the Java environment designed by A. R. Moodie for *in vivo* experiments in the IBLS.

Each panel in Fig. 69 represents one of the sixteen electrodes of the array, with the exception of the very top panel, which represents the reference electrode. Individual panels may be isolated and telescoped using the vertical and horizontal adjustment controls along the bottom of the screen in order to examine their finer detail, particularly in the region of any spiking events. This may be used to conduct intuitive inspections of the data for spikes prior to more rigorous analysis, however this is only practical for relatively small data sets. For the purpose of the analyses documented in this thesis, no rigorous intuitive inspections were made of the data, save to visually account for the presence of spikes during the experimental run, as even the test data sets were relatively large. Rigorous analysis was conducted offline as detailed later in the chapter. In the bottom right corner of the screen a display is

situated containing the key parameters of the stimulation, in particular the pulse strength and duration, and it is these two key parameters with respect to which the data sets are divided up for analysis. Also indicated is whether the pulse is anodic or cathodic (i.e. whether the first phase of the pulse is in the positive or negative direction respectively). Cathodic pulses are more widely used in epiretinal implants as they are more effective at depolarising the retinal ganglion cell membranes than anodic pulses [87]. This is because these membranes generally have negative resting voltages and consequently the voltage increase supplied by a cathodic pulse has the effect of facilitating the onset of an action potential, as they produce a straightforward depolarising effect. When anodic pulses are used, on the other hand, the effect takes the form of a rebound, and any action potentials which may occur – which in turn lead to spiking events – do so on recovery. These are more commonly used in subretinal implants due to the different behaviour of the cells stimulated.

5.3 Analysis process and software

The analysis of spike data followed a number of distinct steps, as illustrated in Fig. 71. The first was to run the data through a “suite” of C programs under the control of a Bash command line shell script. The latter is designed to loop over the channels of a given recording from the physiological experiments, invoking the programs for each data trace in turn. The output files from this software suite inform the next steps of the analysis. A histogram file is used to generate histograms of incidence of spikes for each different pulse strength and duration, for the purpose of gauging threshold pulse strengths for each,

from which follow the principal results of this thesis. In addition, a routine written in the matrix manipulation language, MATLAB, incorporating a combination of principal component analysis and k -medoid clustering enables the identification of particular spike profiles, and consequently particular neurons within the retina.

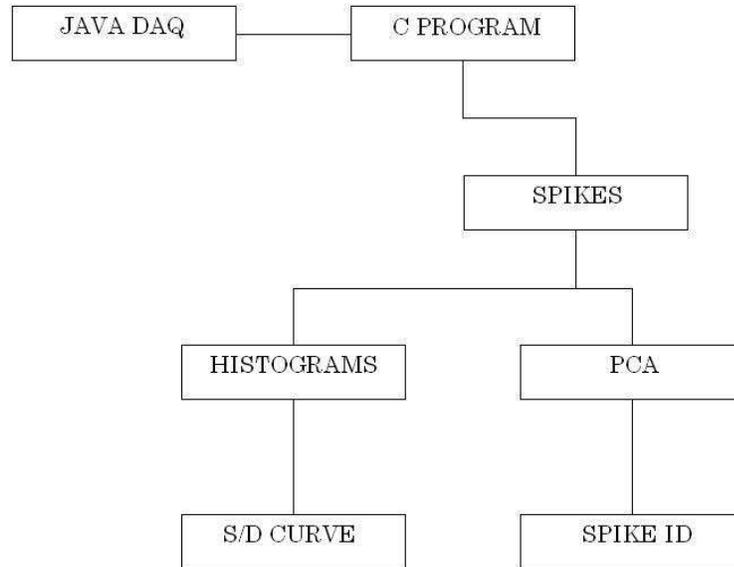


Figure 71: Flow diagram showing the key steps of the data analysis process.

5.3.1 C/Bash analysis

In order to reduce the amount of time required for data analysis, two C data processing programs were written, as well as a Bash command line shell script.

The first program was designed to enable automated classification of spikes in the data. Overall, a minimum of two output files are generated by the

main program. The first is a histogram file, in which the detected spikes are binned at 1 ms intervals across the duration of the data recording. This is important in enabling a strength-duration curve to be built up, as the number of spikes generated for each pulse strength and duration provide an important clue as to the threshold pulse strength for each duration. The second is a data file which enables the raw, interpolated and final data sets to be compared. This is done as a means of assuring the quality of operation of the artifact subtraction routine incorporated into the program to counteract the phenomenon of stimulus artifacts left on the data recordings. This arises as a result of the pre-amplifiers of the stimulation system becoming saturated by the large voltages involved, and skews the data in the immediate wake of the stimulus pulse, leading to spiking events at these positions in the recordings potentially being overlooked during the course of data analysis. The algorithms underpinning this process are described below. All other output files generated consist of “snapshots” of each spike detected, covering intervals from 1 ms prior to their onset to 2 ms after. These are used in the subsequent MATLAB analysis.

When the presence of sixteen electrodes is taken into account, along with the fact that in the experiments data is sampled at 40 kHz, over intervals as long as 2-3 s per trial, the volume of data recorded can quickly become large and time-consuming to analyse. A further saving in analysis time was therefore made by writing a routine in the shell-scripting language, Bash, to spawn jobs for multiple files. Many hundreds of files may be analysed in this manner with a single command. Each such job file generated by the shell script corresponds to one recording generated in the Java environment, and

loops over all sixteen electrodes.

As a consequence of the above, in order to allow the data to be collated more efficiently, the first program was modified to output a short job file consisting of a single command line. This command line invokes a second program written for the single purpose of collating all the spike data gathered over the sixteen electrodes of a record. The individual files containing data for single spikes are read in by this program, which gathers them together in a two-dimensional array that is output as a new, separate file. This is done on a piecemeal basis, with the file appended by the program as each channel in turn is analysed by the program within the shell. When the first channel is analysed, the name of the program is output to the file prior to the names of any files containing spike data in order to begin building the command line. Correspondingly, when the sixteenth and last channel is analysed, the name of the destination file is output to the file. When this command line invokes the program, the result is that all the snapshots of spike data from the main program are collated into a single file which defines a matrix. This is important for the other principal analysis performed on the spike data, involving principal component and k -medoid clustering analysis, which is performed in the MATLAB environment and has the purpose of enabling the detected spikes to be traced to particular neurons within the frog retina operated on during the experiments.

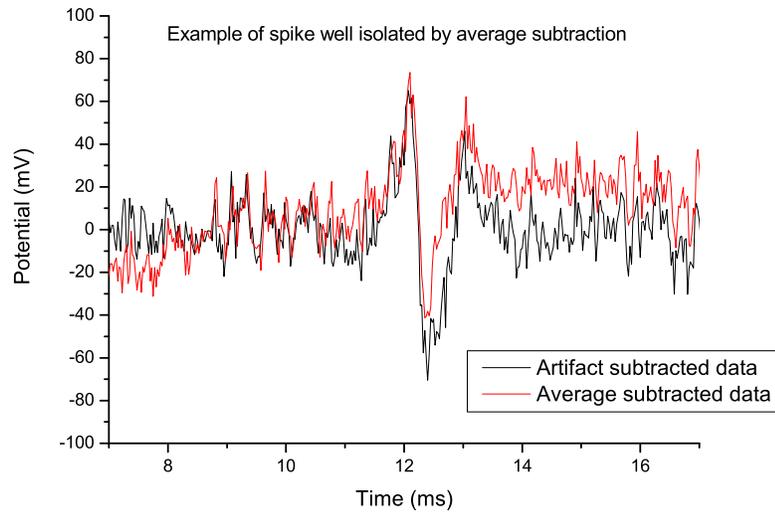


Figure 72: Example of spike well isolated by the averaging method compared to the interpolation method.

5.3.2 Artifact subtraction

The first major subroutine of the program is the mechanism by which the artifacts left by the stimulation pulses on the recordings are interpolated and subsequently removed by subtraction from each data channel in turn. It is important to remove these artifacts before reliable gauging of spiking activity is possible.

Two preliminary studies were carried out with the program: processing of the data by means of interpolation as implemented below, and comparison of this method with the subtraction of averaged spike-free channels. The latter method has been favoured in other fields involving the study of neural activity [105, 106, 107], and potentially lent itself to the retinal studies doc-

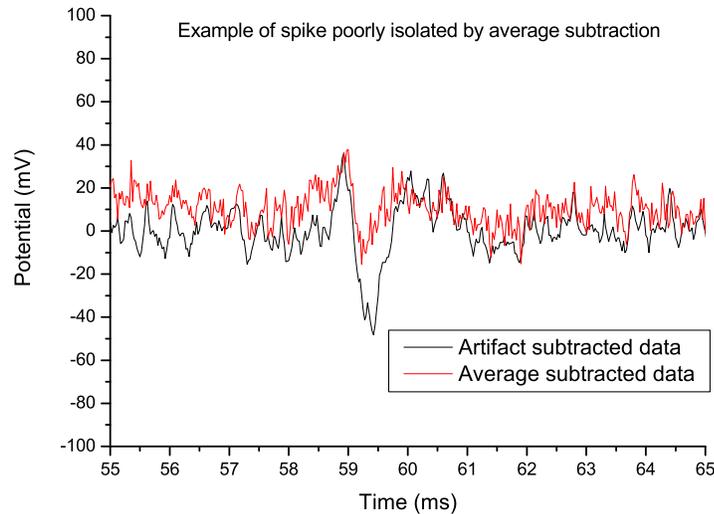


Figure 73: Example of spike poorly isolated by the averaging method compared to the interpolation method.

umented in this thesis because stimulation of a given area of retina does not always induce a spiking event, as the membranes of retinal ganglion cells have particular thresholds below which stimulation does not elicit such activity, therefore making it possible to accomplish artifact removal by averaging over those records not containing spiking events and subtracting the result from the record in which the spiking event is located.

Graphs of the traces from both studies are shown in Figs. 72 and 73. It can be seen from these that while the interpolation method generally yields statistically flat traces, the averaging method can lead to a much noisier profile with less clear isolation of the spiking event or events due to “additive” effects between the statistics of the channels over which the average is calcu-

lated. This suggests that interpolation is a more reliable method for isolation of neural spikes from experimental data than averaging.

It was decided to accomplish artifact subtraction by employing a process of spline interpolation to model the shape of the artifact whilst leaving the remainder of the data set intact, after which the modelled artifact can then simply be subtracted from the raw data set, providing a “statistically flat” trace from which spiking events may be more easily isolated. Initial attempts in this direction were made using linear interpolation. This is the simplest possible implementation of a spline interpolation process, whereby the data are simply averaged over intervals of a pre-determined size and the means of the data over those intervals assigned to the midpoints. These midpoints are then connected linearly by appropriate array element assignments.

This however failed to model the shape of the artifact as effectively as desired, so the decision was made to make use of *cubic* spline interpolation, which is generally far more accurate than linear spline interpolation over intervals of the size under consideration. This programming task made use of pre-written algorithms from Ref. [108]. Two functional routines were adopted - *spline()*, which obtains a second derivative for the data set under consideration (due to the discrete nature of the data it was decided to make use of the averaged data points from the earlier linear interpolation routine for this purpose), and *splint()*, to use the derivative along with the original data set to obtain values for the *Lagrange interpolation polynomial*:

$$y = Ay_j + By_{j+1} + Cy_j'' + Dy_{j+1}''$$

where:

$$A = \frac{x_{j+1}-x}{x_{j+1}-x_j}$$

$$B = 1 - A$$

$$C = \frac{1}{6}(A^3 - A)(x_{j+1} - x_j)^2$$

$$D = \frac{1}{6}(B^3 - B)(x_{j+1} - x_j)^2.$$

These equations gave the interpolated values of the data over the intervals between two successive averages.

The presence of the artifact at a non-zero position in the data set, combined with its generally biphasic form, posed a serious problem for both methods of interpolation used, as these factors tended to prevent the interpolation curves adhering closely enough to the shape of the artifact. This led to erratic behaviour of the artifact-subtracted data during this time interval, as shown in Fig. 74. Further sophistication therefore had to be built in. Following the lead of [92], it was noted that the artifact generally includes periods of saturation, first in the cathodic and then in the anodic direction. It was therefore decided that the starting points of the interpolated data sets should be fixed to the first points after the saturation in the anodic direction. This ensured that they adhered faithfully to the remainder of the data set, as shown in Fig. 75. Data points prior to the cessation of anodic saturation were simply assigned zero values.

A similar software package, known as SALPA (Stimulus Artifact Local Polynomial Approximation), has been independently developed by Wagenaar and Potter [93], which performs the same functions as outlined above, but importantly in real time as opposed to offline after the experiment in the case of

the algorithms described here, and in a manner which can be “tuned” for optimal performance as the experiment progresses. Also of note is SARGE (Stimulus Artifact Removal Graphical Environment), developed by Erez et al [94], which also performs artifact removal in real time and in a graphical environment for ease of adjustment of parameters. These systems both move the emphasis away from the data and algorithms towards a “workflow-centred” approach, enabling the artifact removal to be conducted with much greater efficiency. It may be of advantage to make use of one of these or implement a similar tailored system for future studies in this area.

During the course of the testing, it became clear that some records were be-

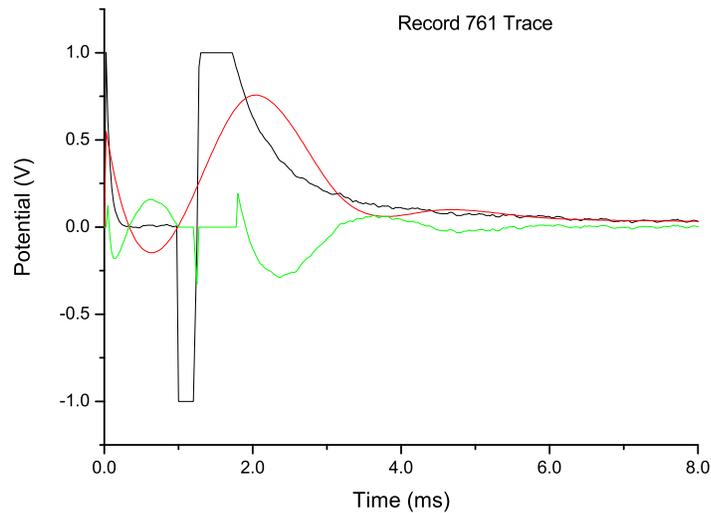


Figure 74: Graph showing a raw data set (black trace), the poorly interpolated artifact (red trace) and its consequent impact on the final data set (green trace).

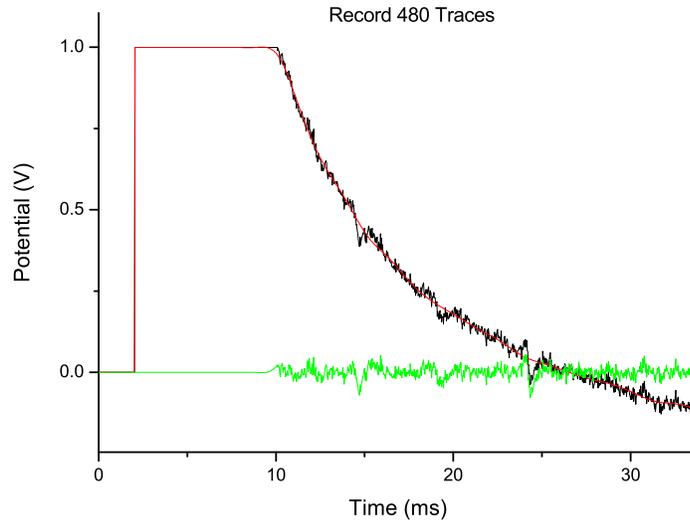


Figure 75: Graph showing the refined interpolation of a data set (black trace), with the modelled artifact (red trace) adhering more closely to the raw data. Points in the final data set (green trace) prior to cessation of anodic saturation have been assigned zero values.

ing “skipped” by the program during runtime because the anodic maximum of the stimulus artifact does not always form an exact plateau, leading it to fail the conditional test set out in the program. This problem was solved by altering the conditional test such that instead of successive array elements being required to be exactly equal, they were simply required to vary from the previous element by no more than a certain amount. This however also added the complication that particular shapes of stimulus artifact caused the conditional test to be satisfied by the cathodic phase, leading to the interpolation being conducted over an interval which includes the anodic phase of the artifact. A related problem was that on a significant number of traces,

the conditional test was not satisfied until a substantial time after the cessation of saturation, which could lead to spiking events being overlooked in the course of data analysis. These were avoided by bypassing the conditional test and instead *#defin*-ing an offset constant at the top of the program. This has an effect approximating that of a trigger, and can be adjusted depending on the width of stimulus pulse used in a given record.

5.3.3 Scanning and spike detection

Once the artifact has been subtracted from the data set, the data set is then ready to be scanned for spikes. The data file is read into a two-dimensional array at the beginning of the program, and a selected individual column from this, fixed by the command line passed from the Bash shell, is passed to the spike identification subroutine within the program. The function loops over all the elements of the column and spike detection is deemed to have been achieved when the amplitude of the data trace remains above the threshold – in either the positive or negative direction – for a minimum of four bins, corresponding to an interval of 100 μ s.

This threshold is calculated by means of a calibration subroutine which operates by calculating the signal background of the data set from the statistics of the last 100 data points of each trace – these are used because there are unlikely to be any time-locked spikes in this region under normal circumstances as elicited retinal signalling generally takes place at much shorter latencies. This signal background can then be used to calculate the total noise of the system. The threshold is then taken for convenience to be an

integer multiple (in this case, by a factor of five) of the total noise. Since the total noise is mathematically equivalent to the standard deviation of the signal background, a subroutine was incorporated into the program to calculate its value in the usual manner:

$$\sigma_{TOTAL} = \sqrt{\frac{1}{N-1} \sum_{i=N-100}^N (v_i - \bar{v})^2}. \quad [86]$$

where v_i represents the i th point in the data set under consideration, and \bar{v} represents the mean over all such points.

Once a spike has been detected, two things subsequently happen. First, the appropriate element of the histogram array – which define the bins of the histogram file documented above – is incremented; and second, a subroutine is invoked which generates an output file containing a snapshot of the spike for further study.

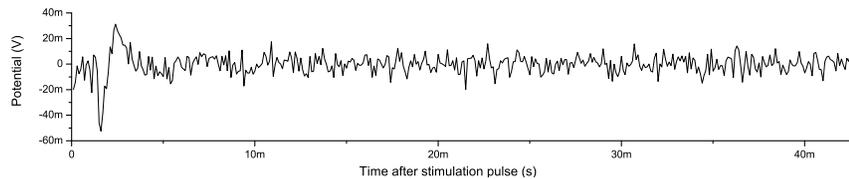


Figure 76: Typical profile of a spontaneous spike as seen in recordings of retinal signalling.

5.3.4 Histogramming and construction of strength-duration curve

This is the most important stage of the analysis. Once all the spike data for the full range of pulse strengths and durations has been collected, the data

is histogrammed, either for a given pulse strength or a given pulse duration. Both are favoured in this thesis. The main program is designed to output a histogram data file which contains a count of the number of spikes detected over the course of a given recording, usually binned in intervals of 1 ms. Thus there are two hundred bins for the test data sets in the ensuing proof-of-concept analysis, though the interval over which data is gathered may be altered. It is also instructive to bin the spikes according to pulse strength for a given duration, as taken together the two histograms can provide important clues as to where the threshold pulse strength lies with respect to a given pulse duration. In the case of these latter histograms there are twelve bins at intervals of 200 mV.

The importance of the histograms lies in their use to gauge a threshold voltage above which there is a statistical likelihood of spikes being detected for a given pulse duration. When this is done for the full range of pulse strengths, a clear strength-duration curve similar in profile to Fig. 77 should emerge. This will enable the determination of the principal result of this thesis, the minimum energy deposit into the retina. In order for any retinal implant to function safely, the energy delivered to the functional tissue must be minimised in order to avoid unwanted heating effects in the tissue. It is therefore important for this parameter to be known in order to design the system correctly.

This minimum energy deposit is generally derived mathematically from the strength-duration curve, by determining two important parameters known as the *rheobase* and *chronaxie*. The rheobase is the minimum current required for stimulation, below which no responses may be elicited from the tissue.

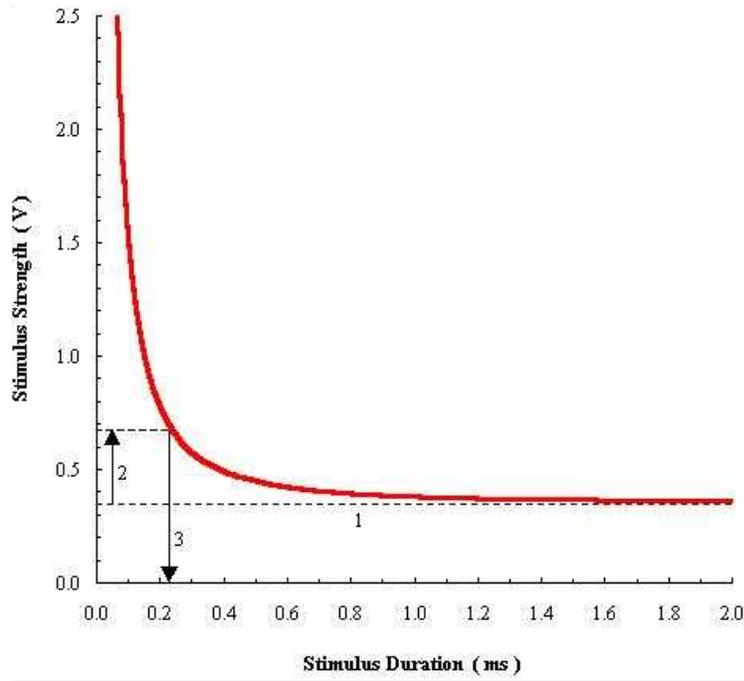


Figure 77: Example of a strength-duration curve showing (1) the rheobase current, (2) the doubling of the rheobase, and (3) the chronaxie as the pulse duration corresponding to twice the rheobase current. Image taken from Ref. [12].

This manifests itself as an asymptotic minimum on the strength-duration curve and may be determined by application of a linear fit to that portion of the curve. Related to this is the chronaxie, which is defined as the stimulation time required to elicit a response at a current of twice the rheobase. Three important relationships then follow from these [98]:

$$I_{th} = I_{rh} \times \left(1 + \frac{t_{ch}}{t}\right)$$

$$Q_{th} = I_{rh} \times t \times \left(1 + \frac{t_{ch}}{t}\right)$$

$$E_{th} = I_{rh}^2 \times r \times t \times \left(1 + \frac{t_{ch}}{t}\right)$$

where t is the stimulus duration and r the resistance of the path through which the current flows. The path resistance may be taken to be that simulated by the electrode-electrolyte interface model described in the previous chapter for the purpose of a proof-of-concept analysis such as that documented in this chapter, although in practice there are also less dominant contributions from the resistance of the retina itself. Alternatively, if impedance data is available for the electrode array used the impedance of the site used as the stimulus electrode may be used instead. Such data is routinely recorded prior to the use of electrode arrays in experiments.

5.3.5 Principal component analysis

The other main analysis to be performed on the spike data is principal component analysis. This has a different motivation from the above, namely that of enabling the tracing of particular spikes to different neurons within the retinal sample under analysis. Each type of neuron has its own characteristic spike signature.

Principal component analysis is a means of reducing the dimensionality of a data set by projecting it onto an orthonormal basis defined by the most significant principal components. The most significant principal components are those that capture most of the variability of the data, and in this case no more than the first three are generally needed. Principal component analysis can be conducted within the MATLAB programming environment by means of the single command *princomp()* [99]. Data is fed to this routine in the form of a matrix, as indicated in the previous section, with the spike data forming the rows and the columns corresponding to individual obser-

vations of the action potential. The routine then generates two matrices – a square matrix, C , which defines a transformation between the two bases, and a matrix S of the same dimensions as the data matrix which defines the representation of the data in the new basis. Once the principal components have been identified they must then be plotted in a 3D scatter plot – each cluster corresponds to a particular neuron in the vicinity of the electrode. The components generated are ranked according to their significance, so it is a simple matter of plotting the first three as shown in Fig. 78. A MATLAB script was written in order to accomplish this.

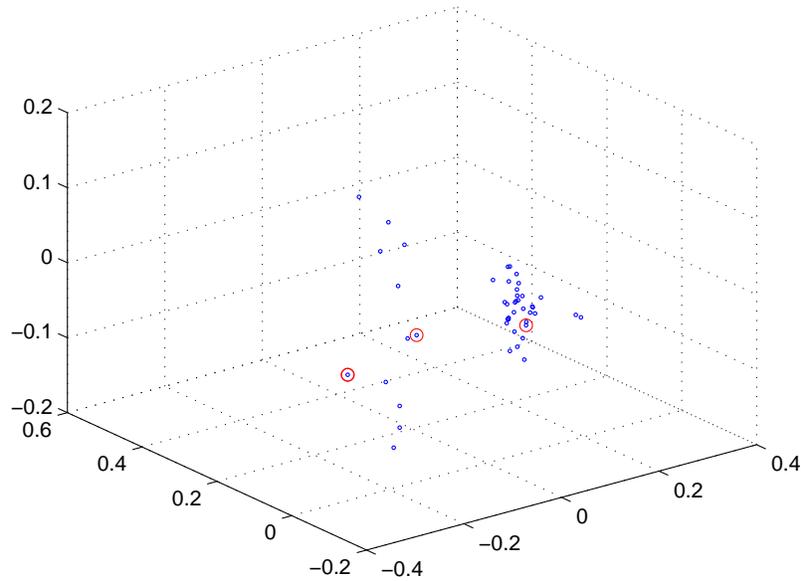


Figure 78: Example of a 3D scatter plot showing the first three principal components of the data set classified into three clusters.

5.3.6 Clustering

Once the principal component analysis is complete and the scores of the principal components plotted on an appropriate scatter plot, the plotted data must be sorted into clusters, which in physical terms correspond to the signalling of different neurons. These can be intuitively assigned by inspection of the plot, however it makes sense to adopt a more rigorous approach and build on the software-driven analysis. It was decided to make use of a hard partitioning technique known as *k-medoid clustering*, which has been proved to be effectively equivalent to principal component analysis [101]. An additional MATLAB script was written in order to enable automation of this process. A subroutine performs the partitioning procedure for up to a maximum of five clusters and assigns colours to each cluster in order that they are clearly delineated for easy inspection. This number was chosen because a given electrode is highly unlikely to detect signalling from more than five different neurons. The spikes corresponding to each cluster can also be plotted and averaged. This is accomplished by computing the inverse of the matrix C and multiplying it on the right by the matrix S , thereby reversing the transformation – there is a direct correspondence between the rows of S and the rows of the data matrix, which correspond to spikes. The averaging of the spikes for a given cluster is accomplished in MATLAB with the *mean()* function. In this manner, the detected spikes may be traced to particular neurons within the retina and classified accordingly.

5.4 Testing and proof-of-concept analysis

5.4.1 Testing of artifact subtraction

The robustness of the artifact subtraction routine was tested by running experimental data through the program. The data used was a series of recordings taken of frog retina and optic tectum using pulse widths ranging from 0.05 ms to 10 ms, and pulse voltages ranging from 20 mV up to 3500 mV. The figures shown indicate that the artifact subtraction routine is able to produce statistically flat traces which may then be scanned for spikes. Fig. 79 shows a recording taken under 0.5 ms stimulation with a 100 mV pulse. This offers a good demonstration of the capabilities of the routine, given the pronounced unevenness of the trace. The trough at 75 ms is much too broad to constitute a spiking event and has correctly been removed. Fig. 80 shows a recording under 0.5 ms of stimulation with a voltage of 250 mV, again demonstrating a statistically flat trace and also that the downturn from saturation has been correctly detected. Figs. 81 and 82 show examples of the routine operating on recordings under 10 ms stimulation pulses, with voltages of 1000 mV and 100 mV respectively. Fig. 83 shows the effect on a recording at 20 mV with a 0.1 ms stimulus pulse, and demonstrates that the routine is robust even at the lowest pulse strengths, where saturation may not occur. An example of the *#define* trigger at work is shown in Fig. 84.

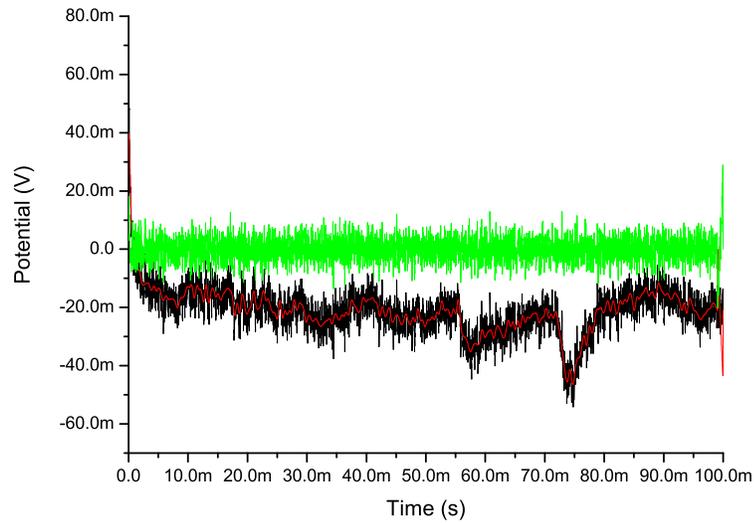


Figure 79: Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 0.5 ms stimulation at 100 mV. The trough at 75 ms is a noise event and has been correctly removed.

5.4.2 Spike processing and proof-of-concept analysis

Once all the spike data has been gathered for the full range of voltages and pulse durations under consideration, the timings of the spikes detected can then be plotted in a histogram for each combination of pulse duration and stimulus voltage. Inspection of these histograms reveals a threshold voltage for each pulse duration, below which the histograms will show a statistical unlikelihood of spiking activity and above which clear statistical likelihoods of spiking activity are seen. These threshold voltages are then converted into currents and plotted in a strength-duration curve.

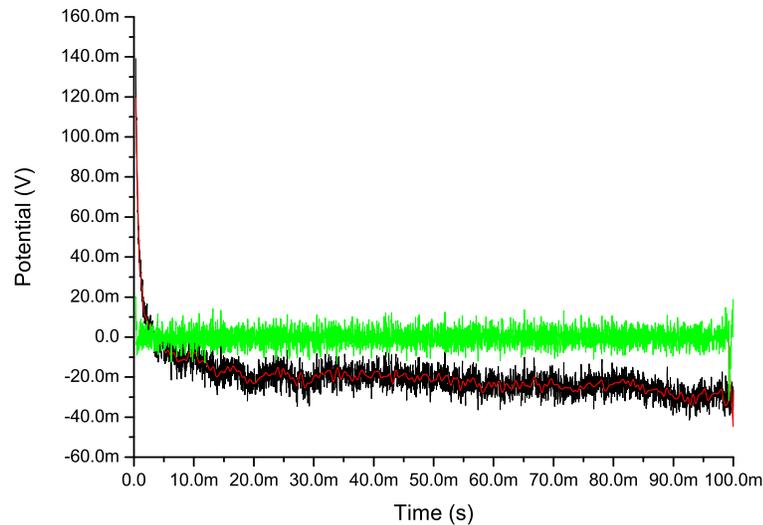


Figure 80: Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 0.5 ms stimulation at 250 mV.

In order to prove the robustness of the processing software designed for the work documented in this thesis, a test data set selected from the above recorded on frog retina was subjected to analysis. From this data set, records were selected that utilised cathodic stimulus pulses, of strengths ranging from 200 mV to 2400 mV in increments of 200 mV, and durations of 0.2 ms, 0.5 ms, 1 ms, 2 ms and 5 ms. The numbers of records varied for each combination of parameters, with a minimum of only four devoted to some, and a few having as many as sixteen, though most had in the region of six to eight records devoted to them. A total of 356 records were analysed.

The pulses were sorted according to combinations of their parameters as

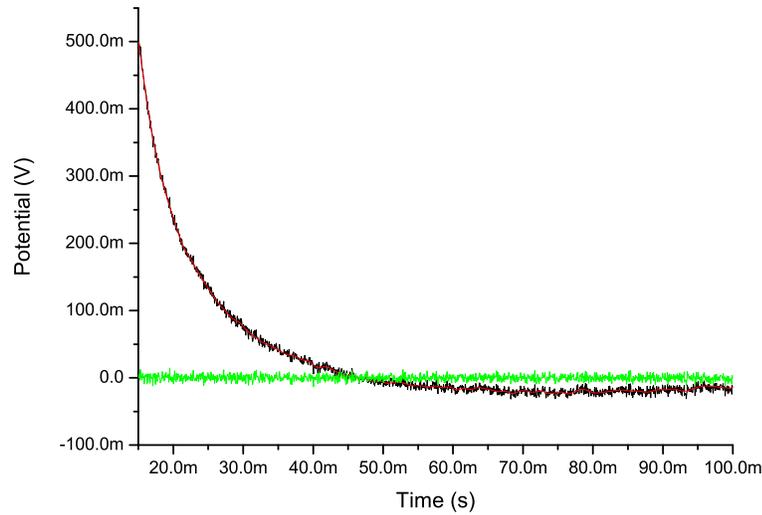


Figure 81: Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 10 ms stimulation at 1000 mV.

documented previously, and groups of records corresponding to particular combinations were analysed in turn, usually in ascending order of first pulse strength and then duration. After C/Bash processing each group was subjected to scrutiny of the detected spikes and then histogrammed with respect to time for each pulse duration. A total of 46 spike traces were detected within the 356 records processed, and histograms of some of these can be viewed in Figs. 85 and 86. It is important to note that each detected spike trace is not necessarily an individual spike as several electrodes may detect a spike from a range of distances, so the total number of spikes detected is likely to be much less than this.

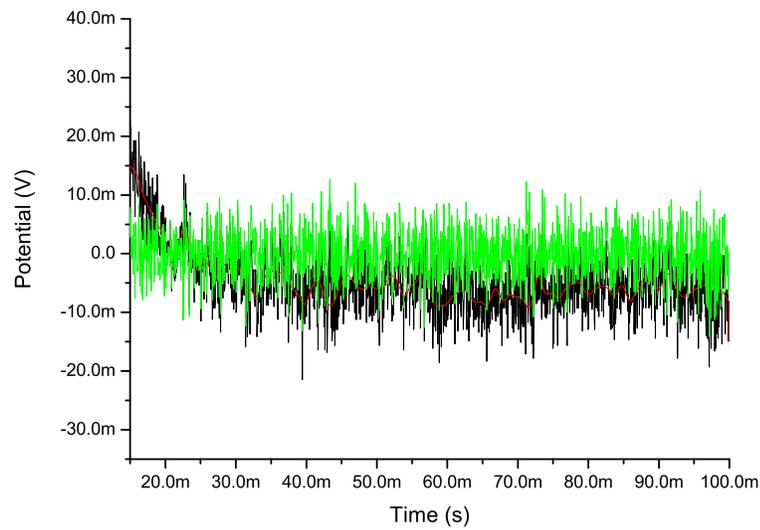


Figure 82: Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 10 ms stimulation at 100 mV.

It can clearly be seen on both histograms that spikes occur at both short latency – in the immediate wake of the stimulus – and longer latency – on the order of 10-100 ms.

As an aid to gauging the threshold pulse strengths for each pulse duration, with a view to constructing a strength-duration curve, it is also instructive to construct a second set of histograms, each corresponding to a particular pulse duration, with the binning conducted with respect to pulse strength. The two most active histograms are shown in Figs. 87 and 88.

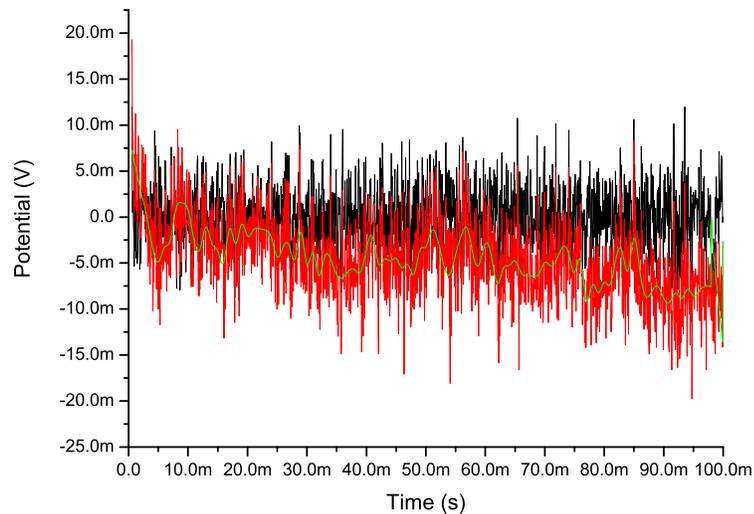


Figure 83: Graph showing a raw data set (black trace), the interpolated artifact (red trace) and its consequent impact on the final data set (green trace) for a recording made under 0.1 ms stimulation at 20 mV.

While these histograms demonstrate the robustness of the processing software, it is clear that the data set does not possess the critical mass of statistics necessary to enable the determination of the spiking threshold to the degree of precision necessary for the construction of a strength-duration curve or the determination of rheobase and chronaxie with respect to the curve. While Fig. 87 seems to suggest a relatively high threshold as would be expected with a pulse duration of 0.2 ms, and Fig. 88, with its high peak at 1600 mV seems to suggest a much lower threshold, again as would be expected at a pulse duration of 5 ms, the intermediate histograms, not shown, do not contain enough data to support any meaningful conclusions. The only conclusions that may be drawn from these histograms is that with more data a

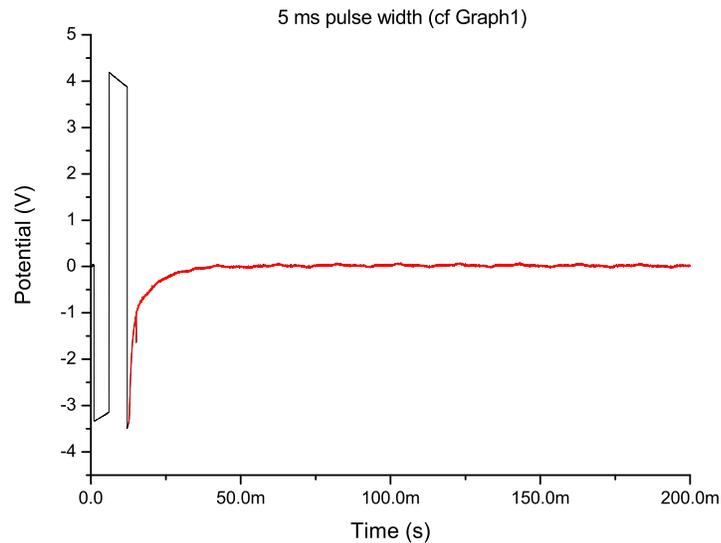


Figure 84: Graph showing how the *#define* offset may be used to select the point at which to begin artifact interpolation on the data.

more meaningful pattern, in line with the expected behaviour, may emerge.

5.4.3 MATLAB analysis

The robustness of the MATLAB scripts for the principal component analysis was tested in much the same way using the 46 spikes detected during the histogramming analysis. Here there is no need to keep the spikes detected in the analysis of different combinations of parameters apart, as was initially done in the previous analysis; what is of interest here are the profiles of the various spikes. So the spike data files generated by the C/Bash analysis process were collated into a single file and read in for analysis. There are four windows in the resulting plot, shown in Fig. 89: the top left shows an overlay plot of all 46 spikes, in the same format as generated by the C/Bash software;

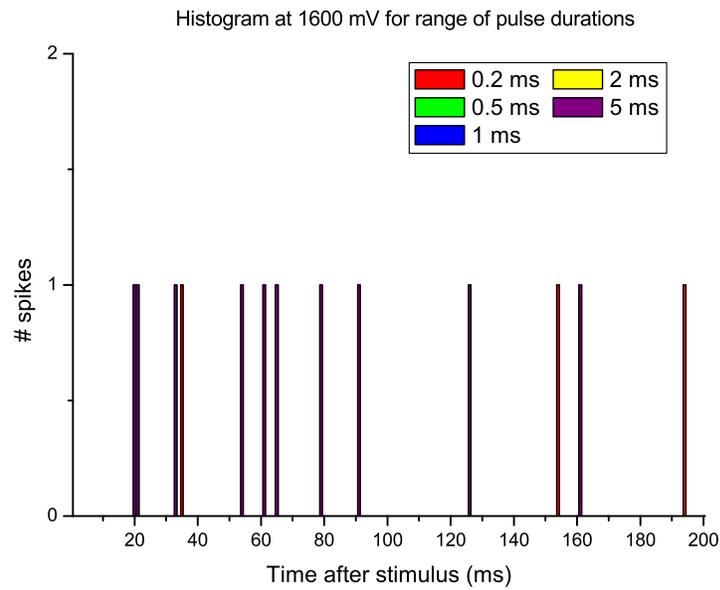


Figure 85: Histogram of spikes detected with pulses of strength 1600 mV.

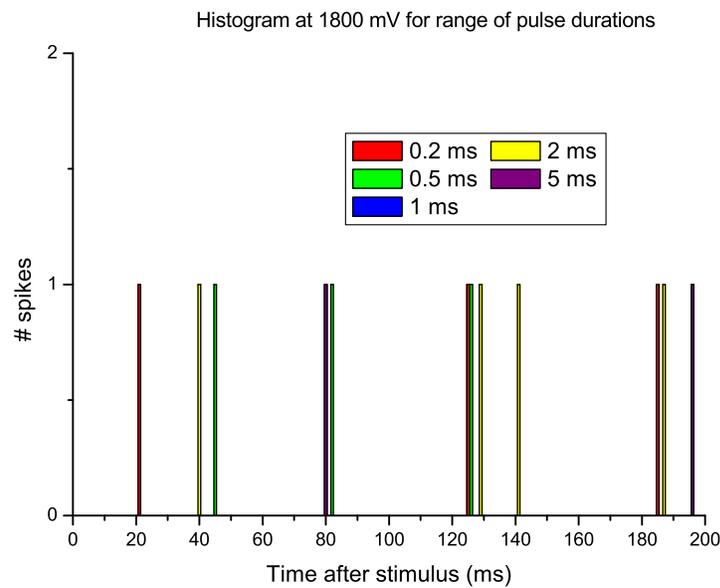


Figure 86: Histogram of spikes detected with pulses of strength 1800 mV.

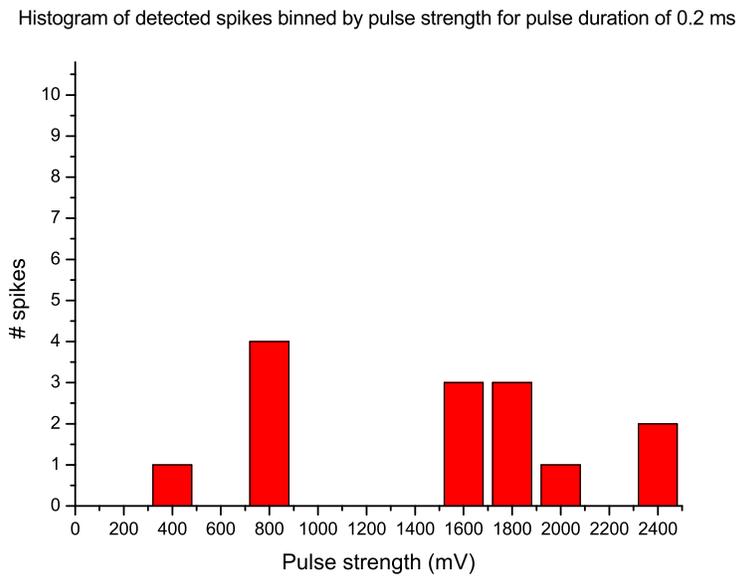


Figure 87: Histogram of spikes detected with pulses of duration 0.2 ms.

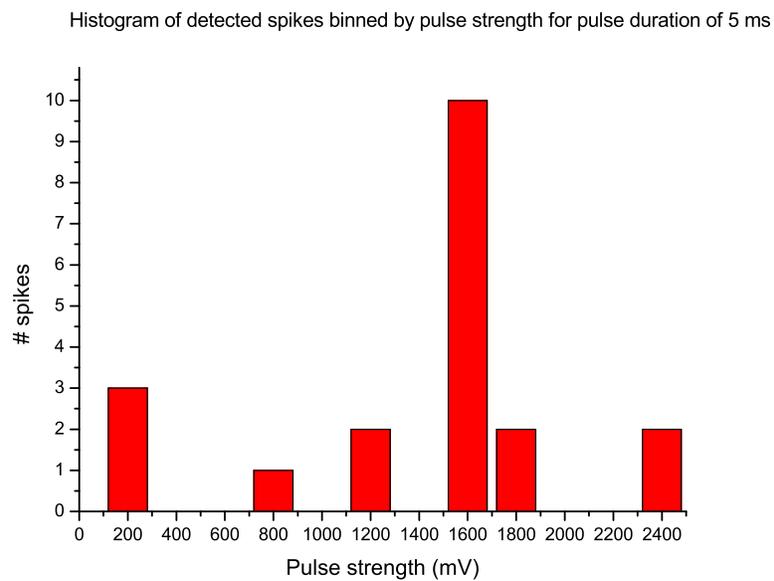


Figure 88: Histogram of spikes detected with pulses of duration 5 ms.

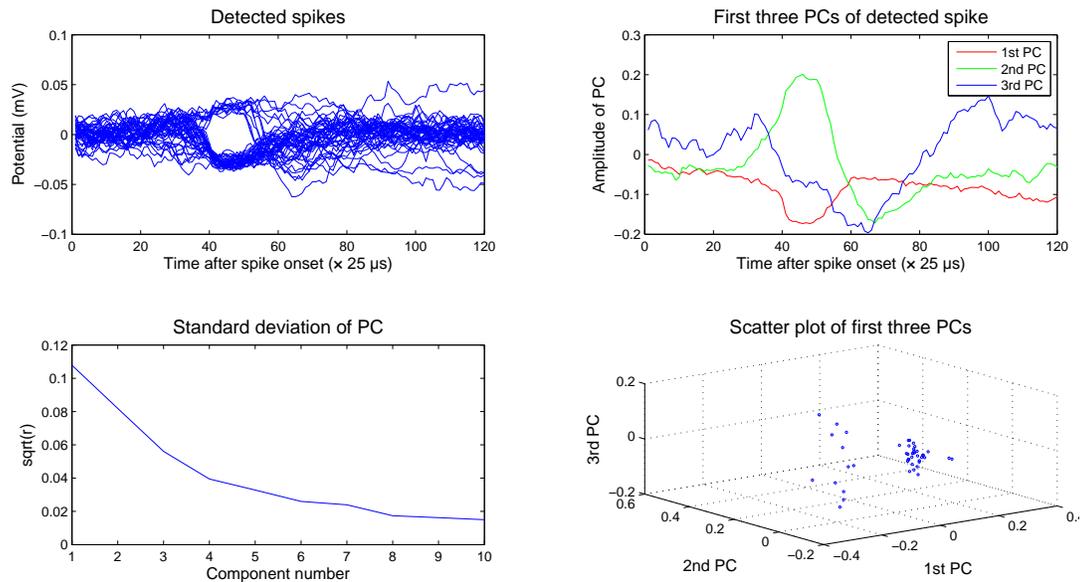


Figure 89: A PCA plot for the test data set.

the top right shows an overlay plot of the three most significant principal components of the data, while the bottom left shows the variances of the ten most significant principal components; and the bottom right contains the scatter plot of the scores of the three most significant principal components of the data. The results of the analysis for four clusters are shown in Fig. 90, and the averaged spikes corresponding to each cluster are shown in Fig. 91.

It is clear that statistics are the only barrier to a clear description of the results shown. The overlay plot of spikes in Fig. 89 shows a clear majority of one particular type of spike, and this is borne out by the presence of a very clear cluster on the scatter plot. However, there is also a significant minority of at least one other type of spike, and this appears to be borne out by the profiles shown on the plot of spike averages in Fig. 91; each

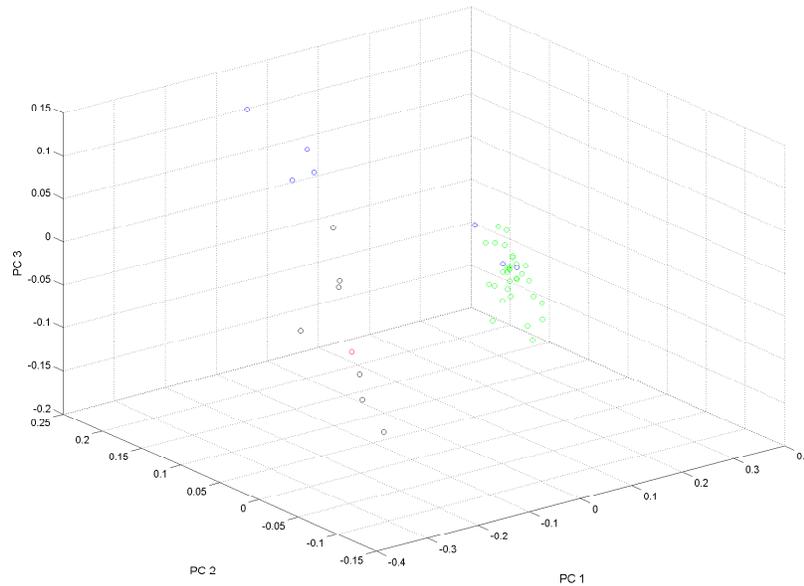


Figure 90: A cluster plot for the data shown in Fig. 89, showing classification into four clusters.

colour corresponds to the cluster of that colour on the scatter plot in Fig. 90. However there is not enough data present to verify this rigorously. The presence of a significant number of apparent outliers on the scatter plot to the left of the main cluster suggest that with a greater critical mass of statistics these could in fact be proven to be clusters, and possibly spikes, in their own right.

A greater critical mass of statistics also opens up opportunities for more sophisticated analysis techniques to be applied. *Bayesian clustering* is the next step from principal component analysis in terms of rigour, and involves modelling each cluster as a multivariate Gaussian and assigning spikes to each cluster based on its likelihood of falling within that cluster [99]. However an

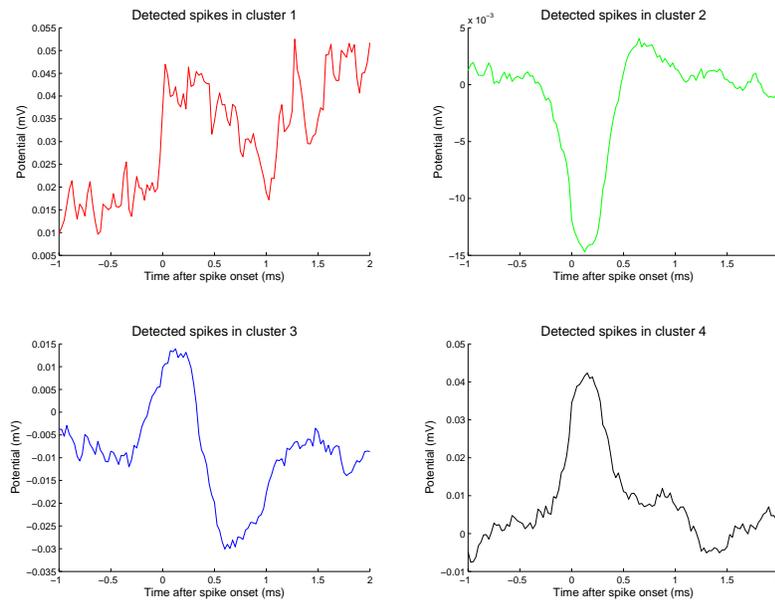


Figure 91: Plot of the averaged spikes for each cluster of the data analysed in Figs. 89 and 90.

important drawback to modelling clusters as Gaussians is that a number of effects can occur which lead to non-Gaussian cluster shapes. Chief among these are electrode drifts during recording, bursting leading to variations in spike shape, overlapping spikes and non-stationary background noise.

Two methods have been developed which can overcome this difficulty. One is to adopt a hierarchical method as applied by Fee et al [95], in which the data is initially sorted into a large number of clusters which are then progressively merged according to common elements within them. Another, more recently adopted for the purpose of spike sorting is *superparamagnetic clustering*, which has been applied to the problem by Quiroga et al [96]. In this method, spikes are grouped into clusters as a function of a single

parameter known as the temperature. The number of clusters increases with temperature as in statistical mechanics, however an optimal temperature or range of temperatures can be found which corresponds to the optimal solution.

The use of this, or any other more sophisticated method, however, depends on having more statistics than are currently available.

5.4.4 Increase of critical mass of data

In light of the findings of the above, a new data framework had to be designed in order to ensure that not only were the necessary range of pulse strengths and durations covered by the new data set, but also that the necessary critical mass of statistics was available in order to enable meaningful conclusions to be drawn from the study. It was therefore decided to make some alterations to the ranges of parameters tested. The range of pulse durations would change from the previous arrangement to 0.1 ms, 0.2 ms, 0.5 ms, 1 ms, 2 ms and 5 ms. This change was made because the expected tendency of the strength-duration curve to plateau at longer pulse durations meant that use of overly long pulse durations simply added to the necessary recording time for no benefit, and the other selected values were more likely to yield points of structural importance on the curve. It was also decided to make use of a range of pulse currents instead of voltages – the levels to be used were 200 nA, 500 nA, 1 μ A, 2 μ A, 5 μ A, 10 μ A and 20 μ A. Currents are preferable to voltages as they provide much greater control. The opportunity was also taken to halve the recording rate to 20 kHz – as spiking events occur over sufficiently long time scales for this change not to have any adverse effects

on the analysis – and reduce the recording lengths to just 50 ms and 100 ms from the original 200 ms, as spikes at longer latency are more likely to have arisen as a result of a network response by the inner retina rather than direct activation of the ganglion cell layer by the stimulation pulse.

6 Final analysis

The analysis documented in the previous chapter demonstrated the robustness of the developed software suite for both the detection of spiking events from retinal data and the localisation of such events to particular neurons within the retinal tissue sample. The goal of the analysis documented in this chapter was to build on this by replicating the previous results with a data set with substantially improved statistics. With this larger critical mass of statistics, data from the analysis documented in the previous chapter suggested that the histograms would show a substantially greater number of detected spikes, thereby enabling a strength-duration curve to be constructed with far greater certainty. In addition, the PCA plot generated in the MATLAB environment would contain more, and larger, clusters than the one generated in the previous chapter, providing further confirmation of the detection of spikes from different neurons.

6.1 Analysis framework

In light of the findings of the previous chapter, a new data framework was designed in order to ensure that not only were the necessary range of pulse strengths and durations covered by the new data set, but also that the necessary critical mass of statistics was available in order to enable meaningful conclusions to be drawn from the study. It was therefore decided to make some alterations to the ranges of parameters tested. The range of pulse durations would change from the previous arrangement to 0.1 ms, 0.2 ms, 0.5 ms, 1 ms, 2 ms and 5 ms. This change was made because the expected tendency of the strength-duration curve to plateau at longer pulse durations

meant that use of overly long pulse durations simply added to the necessary recording time for no benefit, and the other selected values were more likely to yield points of structural importance on the curve. It was also decided to make use of a range of pulse currents instead of voltages – the levels to be used were 200 nA, 500 nA, 1 μ A, 2 μ A, 5 μ A, 10 μ A and 20 μ A. Currents are preferable to voltages as they provide much greater control. The opportunity was also taken to halve the recording rate to 20 kHz – as spiking events occur over sufficiently long time scales for this change not to have any adverse effects on the analysis – and reduce the recording lengths to just 50 ms and 100 ms from the original 200 ms, as spikes at longer latency are more likely to have arisen as a result of a network response by the inner retina rather than direct activation of the ganglion cell layer by the stimulation pulse.

6.2 Analysis and results

6.2.1 Experimental apparatus testing prior to experiment

Prior to embarking upon the experiment itself the performance of the experimental apparatus was verified. This is done in order to ensure that the signals involved in the experiment are being correctly picked up by the system. To this end, a data recording process was undertaken in which pulses of varying strength and duration generated by the IPIX II were passed through the stimulus electrodes and recorded by the recording electrodes. The data was logged in the usual manner. The electrode array was immersed in water rather than physiological saline in order to minimise the attenuation of

the pulse that would result from immersion in the latter, by increasing the impedance of the interface and therefore the voltage drop across the electrode.

The levels of current used in the tests documented were $0.1 \mu\text{A}$, $0.2 \mu\text{A}$, $0.5 \mu\text{A}$, $1 \mu\text{A}$, $2 \mu\text{A}$, $5 \mu\text{A}$, $10 \mu\text{A}$, $15 \mu\text{A}$ and $20 \mu\text{A}$, and the pulse durations used were $100 \mu\text{s}$, $200 \mu\text{s}$, $500 \mu\text{s}$, 1 ms , 2 ms and 5 ms . The full range of currents were tested only for $500 \mu\text{s}$. The rest of the pulse durations were tested only on currents of $0.1 \mu\text{A}$, $1 \mu\text{A}$ and $10 \mu\text{A}$. There were also tests conducted on other stimulus electrodes for pulses of $500 \mu\text{s}$ duration in order to guarantee their operation.

The impedances of the electrodes were found to vary across the array. Of the eight stimulus electrodes, numbers 3 and 7 were found to be open circuits while number 4 was a short circuit. This left five usable stimulus electrodes, of which number 2 was found to have the lowest impedance. Consequently it was the one chosen to be used in the experiment. An additional factor which must be taken into account in the interpretation of the data is the positioning of the recording electrodes at varying distances from the stimulus electrode. The further away the recording electrode from the stimulus electrode, the smaller the amplitude of the waveform recorded by it. However, the basic profile of the waveform should be similar across all data records and channels. The system saturates at a level of approximately 2.5 V due to the architecture of the biphasic output circuitry of the IPIX II.

Examples of the waveforms detected by the electrodes are shown in Figs. 92 and 93. The graphs are arranged in numerical order from 1 to 16 with each

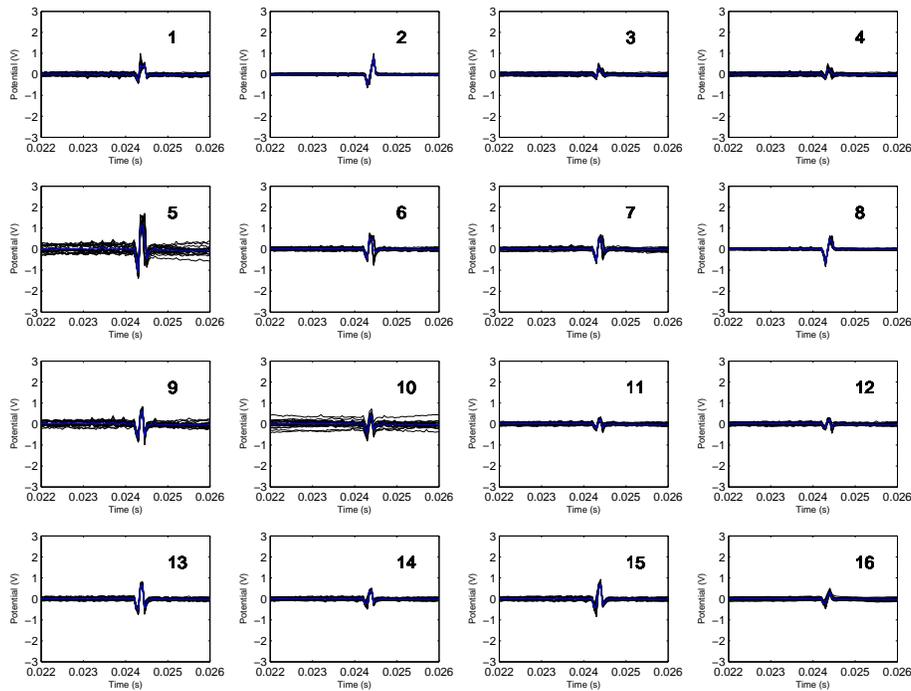


Figure 92: Example of the waveforms detected by the recording electrodes for a stimulus duration of 0.5 ms and strength of $0.1 \mu\text{A}$.

number corresponding to a channel of the recording. Twenty distinct recordings were made for each pulse strength and duration. These were plotted using an adaptation of the MATLAB routine used to generate the overlay plots from the PCA analysis and then averaged. The black traces correspond to the overlay plots of the individual records, while the blue traces correspond to the averaged waveforms.

As indicated above, the waveforms detected by the recording electrodes may be classified according to the distance of each electrode from the stimulus electrode. In the case of electrodes which are relatively close to the stimulus electrode, the channel will be saturated at high currents for all but the very

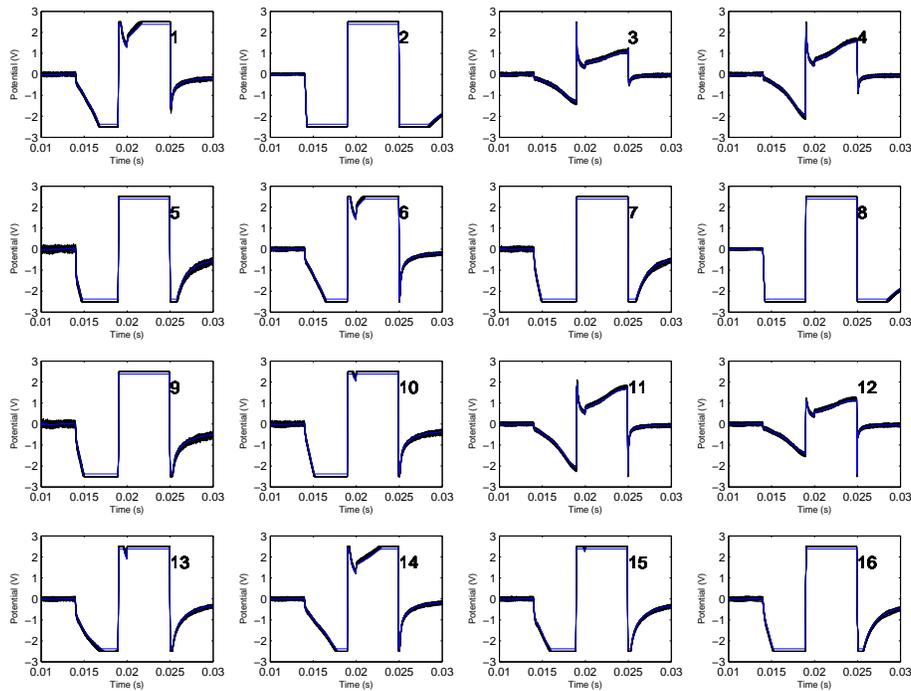


Figure 93: Example of the waveforms detected by the recording electrodes for a stimulus duration of 5 ms and strength of $10 \mu\text{A}$.

shortest pulse durations. This can be seen in the plot corresponding to electrode 5 in Fig. 93. At the other end of the spectrum, if a recording electrode is further away from the stimulus electrode, the detected waveform will have a much smaller amplitude, for example, as with recording electrode 3 in Fig. 92.

6.2.2 Spike analysis and construction of strength-duration curve

The analysis followed the same steps as the one documented in the previous chapter. The pulses were sorted according to combinations of their parameters as before, and groups of records corresponding to particular combinations

were once again analysed in ascending order of first pulse strength and then duration. After C/Bash processing each group was histogrammed with respect to time for each pulse duration.

It can be seen from some of the histograms shown, as documented below, that a much clearer pattern can be observed with the critical mass of statistics now available in these particular cases. Fewer spikes were detected for the majority of other pulse strengths and durations, however those for greater pulse strengths demonstrate much clearer evidence of activity than was previously the case.

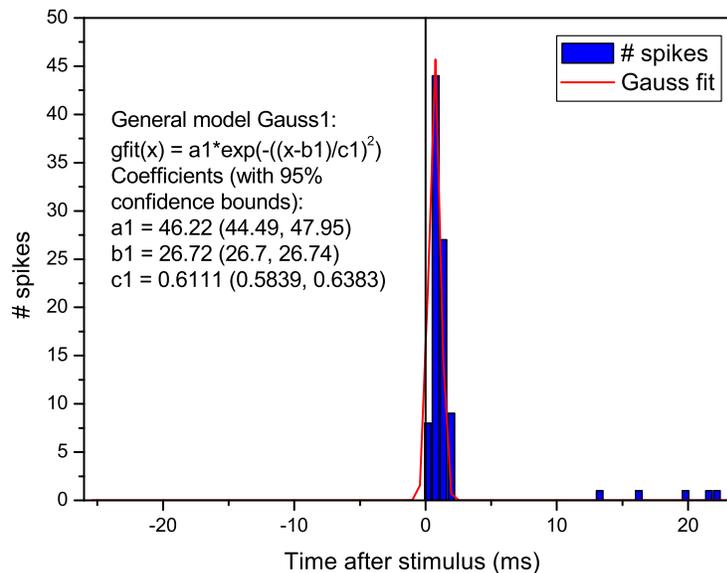


Figure 94: Gaussian fit and histogram of detected spikes for a pulse strength of $10 \mu\text{A}$ and duration of 1 ms. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.

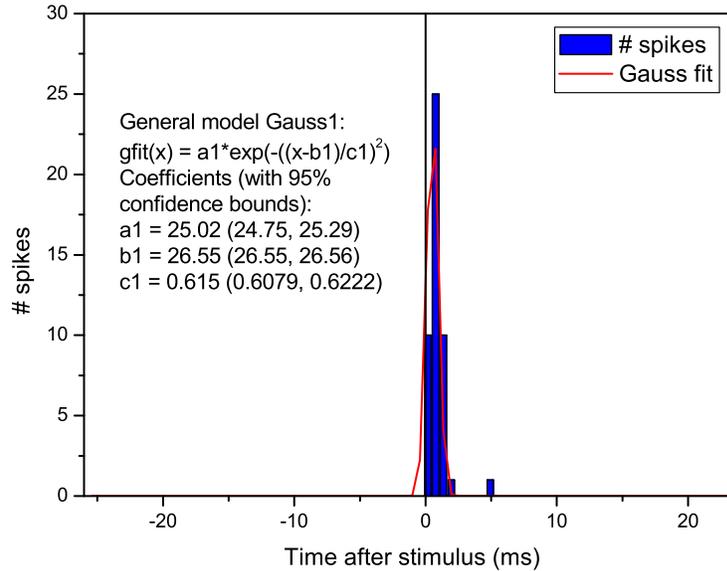


Figure 95: Gaussian fit and histogram of detected spikes for a pulse strength of $20 \mu\text{A}$ and duration of 1 ms. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.

The histograms for 1 ms pulses at pulse strengths of $10 \mu\text{A}$ and $20 \mu\text{A}$ are shown in Figs. 94 and 95, overlaid with the corresponding Gaussian fits to the data generated in MATLAB. Both show a pronounced peak at approximately 30 ms on the timebase of the graph, which corresponds to an interval of approximately 4 ms after the cessation of the stimulation pulse. This represents clear evidence that a spike has been elicited from the retina by the application of the pulse. These are likely to be from the bipolar cell layer rather than the ganglion cell layer as spikes from the latter tend to occur at much shorter latency and are likely to have gone undetected due to the saturation of the amplifiers from the stimulus pulse.

The evidence for this behaviour is strengthened somewhat by examination of

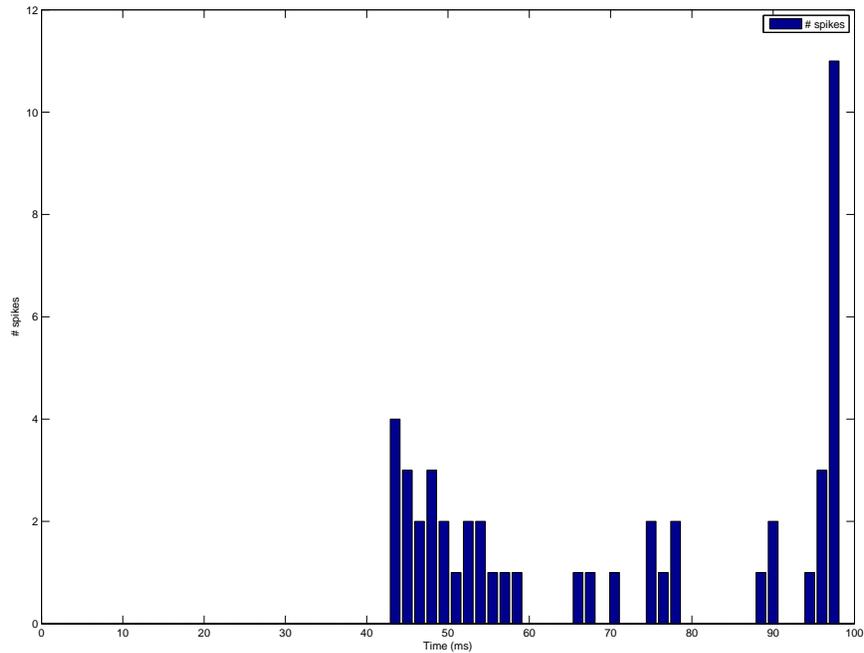


Figure 96: Histogram of detected spikes for a pulse strength of $20 \mu\text{A}$ and duration of 5 ms.

the histograms for 5 ms pulses of similar strength. The data gathered for $15 \mu\text{A}$ pulses, not shown, did not have the statistical weight to support a Gaussian fit, however viewed in conjunction with the histograms and Gaussian fits generated for $20 \mu\text{A}$ pulses, shown in Fig. 96, it reveals evidence that a train of spikes has been elicited from the retina under stimulation by pulses of such strength and duration. The spike distribution in this case may no longer be Gaussian, and instead may point to a network response from the inner retina rather than stimulation of individual bipolar or ganglion cells.

In the previous chapter it was documented that principal component analysis (PCA) provided a robust means of determining the type of neuron from

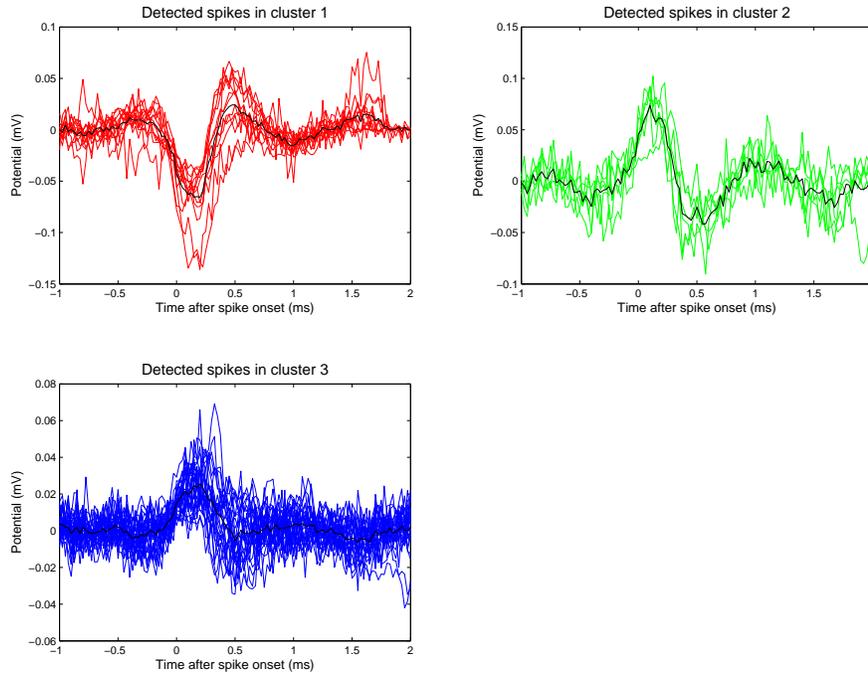


Figure 97: Separation of the spikes detected under a stimulus strength of $20 \mu\text{A}$ and duration of 5 ms into three clusters by means of PCA and k-medoid clustering analysis.

which a spike has been elicited, as it could discriminate easily between the different signatures associated with each. The technique was applied to the current data set and the resulting graphs are shown in Figs. 97 and 98. The clustering analysis technique was then applied in order to separate the spikes into groups corresponding to each type of neuron. The resulting cluster plots can be seen in Figs. 99 and 100, and the separate histograms and Gaussian fits for each cluster in Figs. 101 and 102.

The spike plots shown in Figs. 97 and 98 appear to show the presence of at least two unique types of neuron. This is clear when the two figures are compared. Taking Fig. 97 for three clusters first, the top left plot shows a

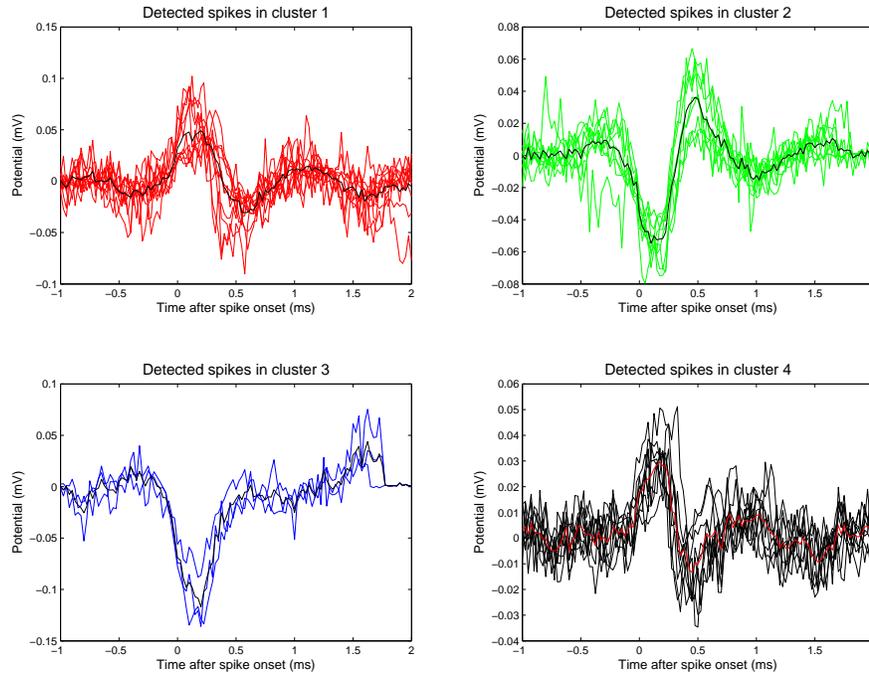


Figure 98: Separation of the spikes detected under a stimulus strength of $20 \mu\text{A}$ and duration of 5 ms into four clusters by means of PCA and k-medoid clustering analysis.

spike with an initial trough in the cathodal direction with a minimal overshoot in the anodal direction. Spikes of this type make up around a third of the events detected. The top right plot shows what could be described as an inverted version of the first type of spike, with a pronounced anodal phase and minimal cathodal overshoot on recovery. These account for only a small number of the detected events. The final plot in the bottom left of the figure shows a substantial number of spikes, with similar profile to the second type but smaller amplitudes. This could indicate that this group of spikes is in fact of the same type as the second group, however more detected spikes are needed to allow clearer clustering and improve the accuracy of the neuron identification.

It may however be deduced from Fig. 98 that there are no more than three unique types of neuron signalling under the applied stimulus. The top left plot shows a category of spikes which corresponds to the top right category in Fig. 97, while the bottom right plot clearly corresponds to those spikes assigned to the bottom left plot in Fig. 97. The remaining two plots in the figure broadly correspond to the top left plot in Fig. 97. This is clear when the cluster plot for four clusters, shown in Fig. 100, is inspected; two clusters can be seen overlapping, which suggests the spikes in these clusters all in fact belong to the same category.

The threshold for activation of a given neuron can be determined by plotting

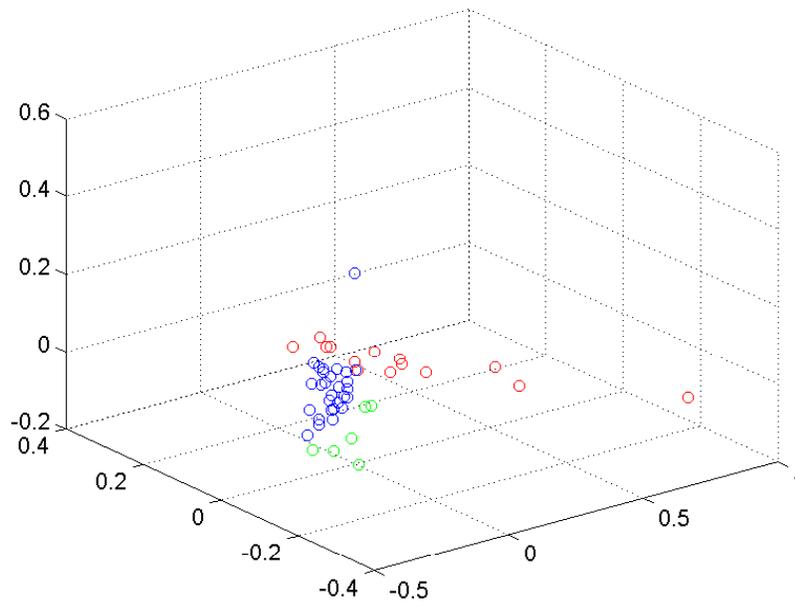


Figure 99: Cluster diagram of the principal components of the detected spikes shown in Fig. 97.

the efficiency of stimulation for the retinal sample as a whole with respect to

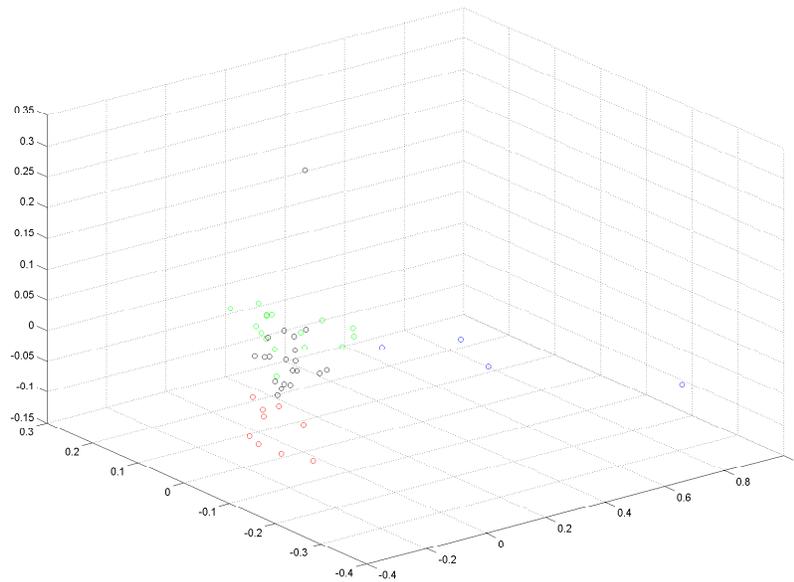


Figure 100: Cluster diagram of the principal components of the detected spikes shown in Fig. 98.

the stimulus current. This is determined by dividing the number of spikes generated by the number of trials conducted. A sigmoidal fit may then be applied to the resulting plot, with the level corresponding to 0.5 spikes per trial corresponding to the threshold for activation of the neuron under consideration.

Once this process has been completed for all the available data, the determined thresholds may then in theory be plotted with respect to stimulus duration in order to construct the strength-duration curve. The strength-duration curve requires to have a very specific form which is shown in the previous chapter. Unfortunately it remained the case that insufficient stimulation had been achieved in order to enable this. Fig. 103 shows the spikes

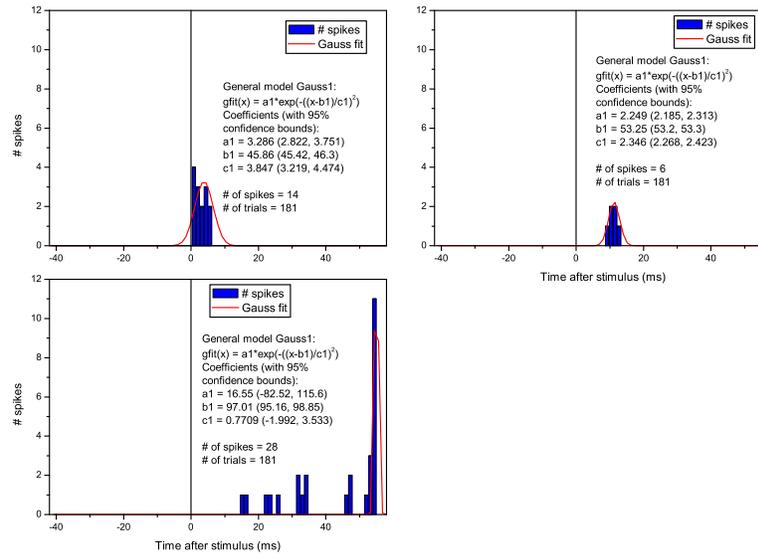


Figure 101: Gaussian fits and histograms of detected spikes for a pulse strength of $20 \mu\text{A}$ and duration of 5 ms, separated into three clusters by means of PCA and k-medoid clustering analysis. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.

per trial statistics for the most active stimulation data, that gathered for pulses of strength $10 \mu\text{A}$. The most active channels in those trials yielded a maximum of 1.3 spikes per trial.

6.2.3 Determination of rheobase and chronaxie

As noted in the previous chapter, two important parameters follow intuitively from the strength-duration curve: the *rheobase* and the *chronaxie*.

The rheobase is the asymptotic plateau towards which the strength-duration

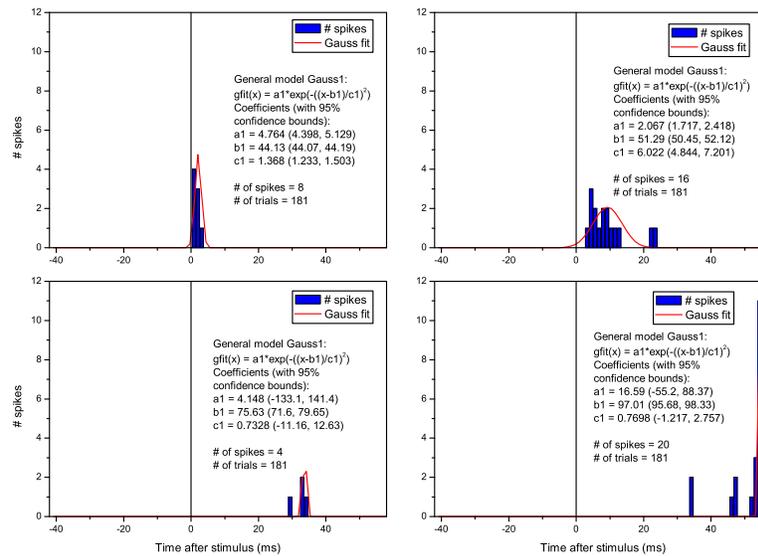


Figure 102: Gaussian fits and histograms of detected spikes for a pulse strength of 20 μA and duration of 5 ms, separated into four clusters by means of PCA and k-medoid clustering analysis. 0 ms on the x -axis represents the end of saturation from the stimulus pulse.

curve tends. It can be estimated by applying a linear fit to the flat portion of the graph. The equation of the trendline obtained should show a negligible gradient and a non-zero y -intercept; this latter number represents the rheobase. The chronaxie is then the pulse duration which corresponds to a current twice as large. As noted above, it was not possible to construct a suitably robust strength-duration curve using the data available. However, it was possible to use the data to apply an exponential fit to those thresholds determined and use the resulting equation to determine approximate values for the rheobase and chronaxie:

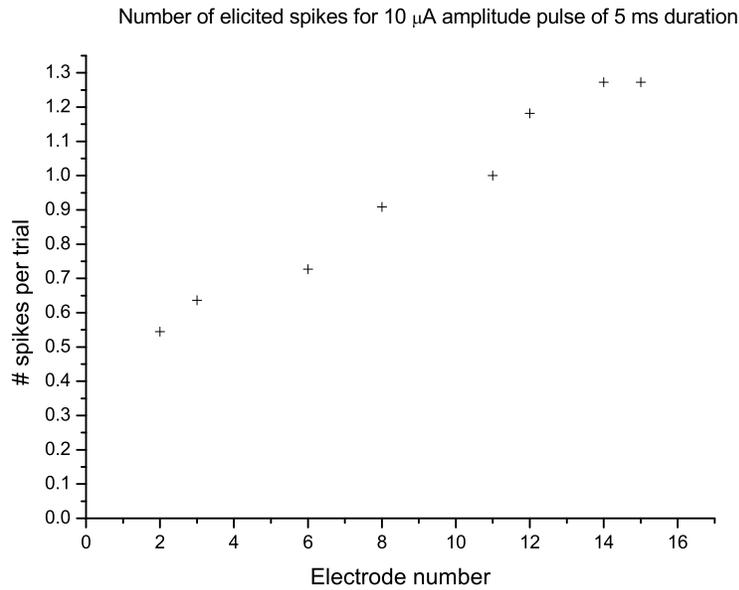


Figure 103: Graph of the numbers of spikes per trial generated on the eight most active electrodes under 10 μA stimulation.

$$I_{rh} = 0.83 \mu\text{A}$$

$$t_{ch} = 0.88 \text{ ms}$$

These values may then be substituted into the equations from the previous chapter in order to obtain a threshold energy for stimulation of the retina. The impedance used is that recorded for the stimulus electrode used in the experiment.

It turns out that these equations take on much simpler forms when the chronaxie is taken as the stimulus duration:

$$I_{th} = I_{rh} \times \left(1 + \frac{t_{ch}}{t}\right) = 2 \times I_{rh} \text{ (for } t = t_{ch}\text{)} = 1.66 \mu\text{A}$$

Similarly, since the threshold charge corresponds to the passing of the threshold current for a period of time equal to the chronaxie, its equation also reduces to a form similar to the above:

$$Q_{th} = I_{rh} \times t \times \left(1 + \frac{t_{ch}}{t}\right) = 2 \times I_{rh} \times t \text{ (for } t = t_{ch}) = 1.46 \text{ nC}$$

The threshold energy derives from the standard equation for the total energy stored in a capacitor, as the system is modelled as such:

$$E_{th} = \frac{1}{2} \times Q_{th} \times V_{th}$$

The threshold charge, Q_{th} , is defined as above. The threshold voltage, V_{th} , is taken to be as defined by Ohm's Law, with the current taken to be the threshold current as defined above, and the impedance to be that recorded for the stimulus electrode. It should be recalled however that this form only predicts the behaviour for shorter pulse durations.

$$E_{th} = 2 \times [I_{rh} \times t] \times [I_{rh} \times r]$$

The threshold energy is therefore now expressed in its other common form with respect to electric power and time:

$$E_{th} = 2 \times I_{rh}^2 \times r \times t$$

The factor of 2 is once again a result of the chronaxie being taken as the stimulus duration. Values have been determined for all of the parameters in the above, and if an interfacial impedance of 500 k Ω is assumed these may

then be substituted in the equation with the result:

$$E_{th} = 2 \times I_{rh}^2 \times r \times t = 606.2 \text{ pJ}$$

This is an important result as it enables the electrode array parameters to be optimised in order to deliver the minimum energy necessary for stimulation of the retina. However, further experiments will require to be carried out in order to obtain the critical mass of statistics necessary to verify this result. As noted previously, the deposit of excess energy into the retina can produce potentially damaging heating effects which required to be avoided if the proposed implant is to be used for an extended period of time.

6.3 Discussion and implications for future work

The minimisation of the energy deposit into the retina documented above is important as it enables unwanted heating effects to be avoided. This is necessary because these can potentially lead to medical side effects.

The case has been made for keeping the electrical components of any retinal implant, whether epi- or subretinal, clear of the retina itself to the fullest extent possible, as studies of dog retina under the influence of intraocular heaters have shown that distant heat sources within the eye cause minimal damage to the retina even under chronic exposure [109]. However, those same studies also showed there to be a limit on the amount of heat retinal tissue could safely absorb at close proximity, placing it at 50 mW over an area of 1.4 mm² on the retinal surface, or a power density of 35.7 mWmm⁻². It is

of vital importance that the energy delivered to the retina by the stimulus electrodes of the implant does not exceed this if the implant is to be viable for chronic implantation.

The threshold energy determined in the previous section is in the pJ range, subject to verification by further experimental study, however this applies to a single stimulus. The repetition rate of the implant circuitry must be taken into account, in addition to the incidence of photons entering the eye, as the implant is intended to mimic the natural functioning of the retina. Thousands of photons per second are likely to be detected by the pixel array at peak intensity and these two factors will combine to drive the energy deposit up from the relatively modest value obtained from the data analysis. However it must also be noted that the refractory period of a typical retinal ganglion cell is around 3 ms, therefore making repetition rates of more than 300 Hz impractical. In practice, the repetition rate of the device is more likely to be in the region of 100 Hz. Therefore this may be taken as a sensible upper limit for the repetition rate.

So if a repetition rate of 100 Hz and stimulation at the threshold energy with pulse duration set at the chronaxie are assumed, the theoretical power of the device can be determined to be:

$$P_{cycle} = E_{th} \times 100 \times \frac{6 \times 10^{-10}}{0.88 \times 10^{-3}} = 68.2 \mu W$$

Once the 519 electrodes of the array are taken into account, the overall theoretical power of the device is:

$$P_{overall} = 68.2 \mu W \times 519 = 35.4 mW$$

It should be noted that this assumes all electrodes on the array being activated at the same time, which would be highly unlikely under *in vivo* implanted conditions.

To determine the implications of this result, the area over which this is distributed must be calculated. The currently proposed electrode array for the Group's retinal implant has 519 electrodes arranged in a hexagonal close packed formation with a pitch of $60\ \mu\text{m}$. With fifteen electrodes along one side of the hexagon the array covers an area of approximately $2.1\ \text{mm}^2$, which is one and a half times that for which the experimental limit of $50\ \text{mW}$ was determined. This implies that the theoretical limit for the implantable array under a power density of $35.7\ \text{mWmm}^{-2}$ is $75\ \text{mW}$. As can be noted from the above, the power of the device using the above parameters, which equates to a power density of $16.8\ \text{mWmm}^{-2}$, is well within the limit and therefore there is no obstacle to the continued use of $5\ \mu\text{m}$ diameter electrodes for the implantable array. Indeed if the impedance constraints on such small electrodes were to prove prohibitive to the safe operation of the device under future testing there is ample scope for increasing the diameter of the electrodes to a level which corresponds to a more manageable impedance across the interface.

As an example, consider a doubling of the electrode diameter to $10\ \mu\text{m}$. This would reduce the impedance by a factor of approximately 4, and if the stimulus current is to be maintained at the same level as before, this will entail a quartering of the power from the previous level when the square dependence of the current is taken into account. And if the pitch of the electrode array were increased to $100\ \mu\text{m}$, thereby mimicing the geometry of an array fabri-

cated by Borkholder et al [111] while maintaining the hexagonal close packed structure of the array, this would have the effect of increasing the area covered by the array by a factor of approximately 3, to 5.8 mm^2 , thereby leaving the power of the device within safe levels.

Another additional consideration when fixing the diameter of the electrodes of the array is whether or not such a diameter allows the device to deliver the resolution necessary to enable a reasonable degree of visual acuity to be restored. One series of tests which remain to be conducted with the IPIX II device and in conjunction with the array (and possibly other arrays with a range of other diameters) is to test the response of the device under patterns of incident light rather than a uniform flow, as was done for the neural network testing in chapter 4, with each combination of neural network weights activated in turn. These will test how well the implant will perform in terms of resolving certain images and therefore how well it lends itself to chronic implantation with a view to restoring a patient's sight, an important property to guarantee before the device can proceed to clinical trials.

7 Conclusions

The work documented in this thesis has aimed to make an important contribution toward the realisation of a fully functional and viable retinal prosthesis. The work has taken in two broad threads. The first involved testing and characterisation of the IPIX II, the CMOS pixel detector and stimulation chip which builds on the functionality of its predecessor, the IPIX, in a number of important ways and will form the “front end” of the completed device. Allied to this was the fine-tuning of an electrode model originated by Kovacs and Borkholder [73], with the result that the behaviour of the overall system of the proposed retinal implant could be predicted for the first time. The behaviour of each component of the system was tested in order to verify that it was in line with that predicted. The second thread involved the use of electrode arrays in conjunction with the aforementioned electrode model to optimise the key dimensions of the electrode array that will form the basis of the implant. This was accomplished by using a combination of stimulus and recording electrodes in experiments on frog retina to create recordings of retinal activity for a number of different pulse strengths and durations. These recordings were analysed and the threshold for elicitation of an action potential determined for each pulse duration. These were then to be plotted in a strength-duration curve, from which two important properties, the *rheobase* and the *chronaxie* were determined, the values of which were used to optimise the charge and energy deposit of the electrode array and relate these properties to the geometry of the electrode array.

Each component of the IPIX II has been designed to mimic the behaviour of a particular component of the retina, as documented in chapter 2. The

first major component tested was the on-chip local neural network, which is designed to mimic the behaviour of the horizontal cells which form the first layer of neural processing within the retina after the activation of the photoreceptors by incident light. The VCOs mimic the bipolar and amacrine cells which generate the action potentials which underpin retinal signalling, and, along with the VCOs, mimic the analogue-to-digital conversion which takes place within the ganglion cell layer prior to the transmission of signals along the optic nerve to the visual cortex. All the major features of the IPIX II have been verified as operating in the required manner, and so it can be deemed to form a viable front end for the retinal implant proposed by the Group.

This enables the implant to build on the functionality of its predecessors to date in two important ways. First, it successfully bypasses the photoreceptor layer and other processing layers of the retina by introducing comparable processing between light detection and ganglion cell layer activation, thereby giving similar benefits to those gained from the use of subretinal implants without the need for the complex invasive surgery involved in their implantation. Second, the pixel array enables the implant to reside entirely in the back of the eye and therefore incorporate eye movements into the analysis of the visual scene, which is not possible in currently available implants which make use of external cameras.

The second major strand of the work involved the use of the IPIX II in experiments involving frog retina in order to verify its operation under experimental conditions but also primarily to gather spike data from the resulting data set and use this data to construct a strength-duration curve. The purpose of this

was to derive important parameters of the frog retina such as optimal current deposit, optimal charge deposit, and optimal energy deposit. It is this latter parameter that can be related to the geometry of the electrode array through the electrode model. Unfortunately, while the IPIX II did successfully stimulate frog retina under experimental conditions, not enough retinal spikes were observed across a wide enough range of stimulation parameters to enable a robust strength-duration curve to be constructed.

Applying an exponential fit to the available strength-duration data and substituting in the equations documented in chapter 6, approximate values of $0.83 \mu\text{A}$ for the rheobase and 0.88 ms for the chronaxie were obtained.

The optimal current, charge and energy deposits in relation to these parameters were found to be $1.66 \mu\text{A}$, 1.46 nC and 606.2 pJ respectively. When these values were applied to the electrode model, taking into account other factors such as level of safe exposure per unit area and number of electrodes on the array, the theoretical power of the device was found to be well within the safe limit of exposure of 50 mW over an area of 1.4 mm^2 , equivalent to a power density of 35.7 mWmm^{-2} . This indicates that there are no obstacles to continued use of $5 \mu\text{m}$ diameter electrodes in the implantable array, and indeed that there is substantial scope for increasing the size of the electrodes to $10 \mu\text{m}$ diameter in order to provide a more manageable level of impedance across the interface. These results remain to be verified with the appropriate critical mass of data in order to lend them the desired statistical weight.

References

- [1] S. Resnikoff, D. Pascolini, D. Etya'ale, I. Kocur, R. Pararajasegaram, G. P. Pokharel, S. P. Marriotti, *Bulletin of the World Health Organisation* **82** (2004) pp. 844-851.
- [2] J. Moisseiev, A. Alhalel, R. Mosuri, G. Treister, *Arch Ophthalmol* **113** (1995) pp. 185-189.
- [3] N. M. Bressler, J. P. Gills, *British Medical Journal* **321** (2000) pp. 1425-1427.
- [4] http://www.egyptarchive.co.uk/html/hidden_treasures/hidden_treasures_31.html.
- [5] R. Elmqvist, A. Senning, *Medical Electronics: Proceedings of the Second International Conference on Medical Electronics, Paris, 24-27 June 1959* (1960) pp. 253-254.
- [6] S. A. P. Haddad, R. P. M. Houben, W. A. Serdijin, *Engineering in Medicine and Biology Magazine IEEE* **25** (2006) pp. 38-48.
- [7] A. Djourno, C. Eyries, B. Vallencien, *C. R. Seances Soc Biol Fil* **151** (1957) pp. 423-425.
- [8] R. J. Briggs, H. C. Eder, P. M. Seligman, R. S. Cowan, K. L. Plant, J. Dalton, D. K. Money, J. F. Patrick, *Otology and Neurotology* **29** (2008) pp. 1141-119.
- [9] A. Y. Chow, M., T. Pardue, J. I. Perlman, S. C. Ball, V. Y. Chow, J. R. Hetling, G. A. Peyman, C. Liang, E. B. Stubbs, Jr., N. S. Peachey, *Journal of Rehabilitation Research and Development* **39** (2002) pp. 313-332.

- [10] S. D'Auria, C. del Papa, E. Charles, D. Fasching, K. H. Becks, G. Lenzen, C. Linder, *Nucl. Inst. and Meth.* **456** (2001) pp. 217-232.
- [11] The Boston Retinal Implant Project,
<http://www.bostonretinalimplant.org>.
- [12] http://www.unm.edu/~toolson/rheobase_chronaxie_determination.htm.
- [13] H. Davson, *Physiology of the Eye*, 5th Edition, MacMillan (London), 1990.
- [14] Neuroscience Tutorial: Eye and Retina, Washington University School of Medicine, <http://thalamus.wustl.edu/course/eyeret.html>.
- [15] S. C. Massey, *Progress in Retinal Research* **9** (1990) pp. 399-425.
- [16] R. T. Hennessey, T. Idia, K. Shina, H. W. Liebowitz, *Vision Res* **16** (1976) pp. 587-589.
- [17] D. H. Hubel, *Eye, Brain and Vision*, Scientific American Library (New York), 1989.
- [18] R. L. De Valois, K. K. De Valois, *Spatial Vision*, Oxford University Press, 1988.
- [19] R. W. Young, *J Ultrastruct Res* **61** (1977) 172-185.
- [20] D. H. Hubel, T. N. Wiesel, *Proc Roy Soc Serb* **198** (1977) 1-59.
- [21] J. Rizzo, J. L. Wyatt, J. Loewenstein, S. Kelly, D. Shire, *Invest Ophthalmol Vis Sci* **44** (2003) pp. 5362-5369.
- [22] J. E. Dowling, *The Retina*, Harvard (Cambridge, MA), 1987.

- [23] H. Kolb, *Invest Ophthalmol Vis Sci* **35** (1994) 2385-2404.
- [24] R. W. Young, *Invest Ophthalmol* **15** (1976) 700-725.
- [25] H. Kolb, E. Fernandez, R. Nelson, *Webvision: The Organisation of the Retina and the Visual System*, <http://webvision.med.utah.edu>.
- [26] J. R. Evans, *Prog Retin Eye Res* **20** (2001) pp. 227-253.
- [27] A. Chopdar, U. Chakravarthy, D. Verma, *British Medical Journal* **326** (2003) pp. 485-488.
- [28] H. G. Wood-Gush, *Journal of the Royal Society of Medicine* **82** (1989) pp. 355-358.
- [29] R. B. Szamier, E. L. Berson, *Invest Ophthalmol Vis Sci* **16** (1977) pp. 947-962.
- [30] H. Rasmussen, K. W. Chu, P. Campochiaro, P. L. Gehlbach, J. A. Haller, J. T. Handa et al., *Hum Gene Ther* **12** (2001) pp. 2029-2032.
- [31] T. Yasukawa, H. Kimura, Y. Tabata, H. Miyamoto, Y. Honda, Y. Ikada, Y. Ogura, *Invest Ophthalmol Vis Sci* **40** (1999) pp.2690-2696.
- [32] C. LeRoy, *Mem. Acad. Roy. Sci. Paris* (1755) pp. 60-98.
- [33] O. Foerster, *J Psychol Neurol, Lpz* **39** (1929) pp. 463-485.
- [34] G. E. Tassicker, *British Journal of Physiological Optics* **13** (1956) pp. 102-105.
- [35] C. Veraart, M. Wanet-Defalque, B. Grard, A. Vanlierde, J. Debelke *Artificial Organs* **27** (1996) pp. 996-1004.

- [36] W. H. Dobbelle, M. G. Mladejovsky, J. P. Girvin, *Science* **183** (1974) pp. 440-444.
- [37] W. H. Dobbelle, *American Society for Artificial Internal Organs* **46** (2000) pp. 3-9.
- [38] G. S. Brindley, W. S. Lewin, *J Physiol (Lond)* **196** (1968) pp. 479-493.
- [39] J. R. Bartlett, R. W. Doty, *Acta Neurobiol Exp (Warsz)* **40** (1980) pp. 713-728.
- [40] M. Bak, J. P. Girvin, F. T. Hambrecht, C. V. Kufta, G. E. Loeb, E. M. Schmidt, *Med Biol Eng Comput* **28** (1990) pp. 257-259.
- [41] E. M. Schmidt, M. J. Bak, F. T. Hambrecht, C. V. Kufta, D. K. O'Rourke, P. Vallabhanath, *Brain* **119** (1996) pp. 507-522.
- [42] R. R. Lakhanpal, D. Yanai, J. D. Weiland, G. Y. Fujii, S. Caffey, R. J. Greenberg, E. de Juan, Jr., M. S. Humayun, *Curr Opin Ophthalmol* **14** (2003) pp. 122-127.
- [43] J. Weiland, M. S. Humayun, *Artificial Organs* **27** (2003) pp. 961-962.
- [44] E. Margalit, M. Maia, R. J. Greenberg, G. Y. Fujii, G. Torres, D. V. Piyathaisere, T. M. O'Hearne, W. Liu, G. Lazzi, G. Dagnelie, D. A. Scribner, E. de Juan, Jr., M. S. Humayun, *Surv. Ophthalmol* **47** (2002) pp. 335-356.
- [45] M. S. Humayun, *Transactions of the American Ophthalmological Society* **99** (2001) pp. 271-300.

- [46] M. S. Humayun, E. de Juan, Jr., G. Dagnelie, R. J. Greenberg, R. H. Propst, D. H. Phillips, *Arch Ophthalmol* **114** (1996) pp. 40-46.
- [47] T. Yagi, Y. Ito, H. Kanda, S. Tanaka, M. Watanabe, Y. Uchikawa, *Sys Man and Cyb* **4** (1999) pp. 382-385.
- [48] F. Gekeler, P. Szurman, S. Grisanti, U. Weiler, R. Claus, T. O. Greiner, M. Volker, K. Kohler, E. Zrenner, K. U. Bartz-Schmidt, *Graefes Arch. Clin. Exp. Ophthalmol.* (2006).
- [49] OptoBionics, Inc., <http://www.optobionics.com>.
- [50] A. Y. Chow, V. Y. Chow, *Neuroscience Letters* **225** (1997) pp. 13-16.
- [51] N. S. Peachey, A. Y. Chow, M. T. Pardue, J. I. Perlman, V. Y. Chow, *Retinal Degenerative Diseases and Experimental Therapy* (1999) pp. 471-478.
- [52] A. Y. Chow, V. Y. Chow, K. H. Packo, J. S. Pollack, G. A. Peyman, R. Schuchard, *Arch Ophthalmol* **122** (2004) pp. 460-469.
- [53] A. Y. Chow, M. T. Pardue, V. Y. Chow, G. A. Peyman, C. Liang, J. I. Perlman, N. S. Peachey, *IEEE Transactions on Rehabilitation Engineering* **9** (2001) pp. 86-95.
- [54] E. Zrenner, *Science* **295** (2002) pp. 1022-1025.
- [55] D. Palanker et al., *J. Neural Eng.* **2** (2005) S105-S120.
- [56] T. Schanze, L. Hesse, C. Lau, N. Greve, W. Haberer, S. Kammer, T. Dörge, A. Rentzos, T. Stieglitz, *IEEE Transactions on Biomedical Engineering* **54** (2007) pp. 983-992.

- [57] P. Walter, Z. F. Kisvarday, M. Gortz, N. Alteheld, G. Rossler, T. Stieglitz, U. T. Eysel, *Invest Ophthalmol Vis Sci* **46** (2005) pp. 1780-1785.
- [58] N. E. Medeiros, C. A. Curcio, *Invest Ophthalmol Vis Sci* **42** (2001) pp. 795-803.
- [59] K. Mathieson, W. Cunningham, J. Marchal, J. Melone, M. Horn, D. Gunning, R. Tang, C. Wilkinson, V. O'Shea, K. M. Smith, A. Litke, E. J. Chichilnisky, M. Rahman, *Nucl. Inst. and Meth. A* **513** (2003) pp. 51-56.
- [60] A. M. Litke, N. Bezayiff, E. J. Chichilnisky, W. Cunningham, W. Dabrowski, A. A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, R. S. Kalmar, K. Mathieson, D. Petrusca, M. Rahman, A. Sher, *Nuclear Science Symposium Conference Record* **2** (2003) pp. 951-955.
- [61] C. Adams, K. Mathieson, D. Gunning, W. Cunningham, M. Rahman, J. D. Morrison, M. L. Prydderch, *Nucl. Inst. and Meth. A* **546** (2005) pp. 154-159.
- [62] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication* (2nd Edition), 183-200, Oxford University Press, 2001.
- [63] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication* (2nd Edition), 152-155, Oxford University Press, 2001.
- [64] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication* (2nd Edition), 208-215, Oxford University Press, 2001.

- [65] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication* (2nd Edition), 262-264, Oxford University Press, 2001.
- [66] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication* (2nd Edition), 277-281, Oxford University Press, 2001.
- [67] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication* (2nd Edition), 283-285, Oxford University Press, 2001.
- [68] D. J. Murdoch, "Data analysis of retinal recordings from multi-electrode arrays," 1st year Ph.D. report, University of Glasgow, Glasgow, UK, 2006.
- [69] T. Akin, K. Najafi, R. H. Smoke, R. M. Bradley, *IEEE Trans Biomed Eng* **41** (1994) 305-313.
- [70] R. R. Richardson, J. A. Miller, W. M. Reichert, *J Biomat* **14** (1993) 627-635.
- [71] T. Stieglitz, H. Beutel, M. Schuettler, J.-U. Meyer, *Biomed Microdevices* **2** (2000) 283-294.
- [72] J. M. Seo et al., *Mater. Sci. Eng. C* **24** (2004) p. 185.
- [73] D. A. Borkholder, "Cell based biosensors using microelectrodes," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1998.
- [74] K. Mathieson, S. Kachiguine et al., *IEEE Trans. Nucl. Sci.* **51** 5 (2004) 2027-2031.
- [75] J. Newman, *J. Electrochem. Soc.* **113** 5 (1966) 501-502.

- [76] A. J. Bard, L. R. Faulkner, *Electrochemical methods*, John Wiley & Sons Inc., New York, 1980.
- [77] G. T. A. Kovacs, In *Enabling technologies for cultured neural networks* (D. A. Stenger, T. M. McKenna, eds.), 121-165, Academic Press, New York, 1994.
- [78] M. L. Prydderch, M. J. French, K. Mathieson et al. *Sensors, Cameras, and Systems for Scientific/Industrial Applications VII* **6068** (2006) 6803.
- [79] K. Mathieson, S. Kachiguine et al., *IEEE Trans. Nucl. Sci.* **51** 5 (2004) 2027-2031.
- [80] ABT Artificial Retina Project, <http://doemedicalsciences.org/abt/retina/retinas.shtml>.
- [81] Retina Implant GmbH, <http://www.retina-implant.de>.
- [82] D. Palanker et al., *J. Neural Eng.* **2** (2005) S105-S120.
- [83] H. L. Helmholtz, *Ann Phys Chem* **7** (1879) 377-382.
- [84] M. Guoy, *J Phys* **9** (1910) 457-468.
- [85] O. Stern, *Z. Electrochem* **30** (1924) 508-516.
- [86] G. Woan, *The Cambridge Handbook of Physics Formulas*, Cambridge University Press, 2000.
- [87] F. Rattay, *Neuroscience* **89** (1999) 335-346.
- [88] W. Nakamura, K. Anami, T. Mori, O. Saitoh, A. Cichocki, S. Amari, *Trans Biomed Eng* **53** (2006) 1294-1308.

- [89] L. Lemieux, A. Salek-Haddadi, O. Josephs, P. Allen, N. Toms, C. Scott, K. Krakow, R. Turner, D. R. Fish, *NeuroImage* **14** (2001) 780-787.
- [90] P. J. Allen, G. Polizzi, K. Krakow, D. R. Fish, L. Lemieux, *NeuroImage* **8** (1998) 229-239.
- [91] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in C* (2nd Edition), Cambridge University Press, 1992.
- [92] E. B. Montgomery et al., *Journal of Neuroscience Methods* Volume 144, Issue 1, 15 May 2005, Pages 107-125.
- [93] D. A. Wagenaar, S. M. Potter, *Journal of Neuroscience Methods* **120** (2002) 113-120.
- [94] Y. Erez, H. Tischler, A. Moran, I. Bar-Gad, *Journal of Neuroscience Methods* **191** (2010) 45-59.
- [95] M. S. Fee, P. P. Mitra, D. Kleinfeld, *Journal of Neuroscience Methods* **69** (1996) 175-188.
- [96] R. Q. Quiroga, Z. Nadasdy, Y. Ben-Shaul, *Neural Comp* **16** (2004) 1661-1687.
- [97] http://www.unm.edu/~toolson/rheobase_chronaxie_determination.htm.
- [98] D. Prutchi, M. Norris, *Design and Development of Medical Electronic Instrumentation*, Wiley, 2005.
- [99] M. S. Lewicki, *Comput. Neural Syst.* **9** (1998) R53-R78.
- [100] B. Balasko, J. Abonyi, B. Feil, *Fuzzy Clustering and Data Analysis Toolbox*.

- [101] C. Ding, X. He, *Proc. of Int. Conf. Machine Learning* (2004) 225-232.
- [102] J. I. Loewenstein, S. R. Montezuma, J. F. Rizzo, Jr., *Arch Ophthalmol* **122** (2004) 587-596.
- [103] C. Adams et al., *Nucl. Instr. and Meth. A* **546** (2005) 154-159.
- [104] A. R. Moodie, "A data acquisition and analysis system for use with multi-electrode arrays in the development of a retinal prosthesis for degenerative retinal diseases," 2009, Ph.D. thesis, University of Glasgow, Glasgow, 2009.
- [105] W. Nakamura, K. Anami, T. Mori, O. Saitoh, A. Cichocki, S. Amari, *Trans Biomed Eng* **53** (2006) 1294-1308.
- [106] L. Lemieux, A. Salek-Haddadi, O. Josephs, P. Allen, N. Toms, C. Scott, K. Krakow, R. Turner, D. R. Fish, *NeuroImage* **14** (2001) 780-787.
- [107] P. J. Allen, G. Polizzi, K. Krakow, D. R. Fish, L. Lemieux, *NeuroImage* **8** (1998) 229-239.
- [108] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in C* (2nd Edition), Cambridge University Press, 1992.
- [109] J. I. Loewenstein, S. R. Montezuma, J. F. Rizzo, Jr., *Arch Ophthalmol* **122** (2004) 587-596.
- [110] E. Margalit, M. Maia, J. D. Weiland, R. J. Greenberg, G. Y. Fujii, G. Torres, D. V. Piyathaisere, T. M. O'Hearn, W. Liu, G. Lazzi, G. Dagnelie, D. A. Scribner, E. de Juan, Jr., M. S. Humayun *Surv Ophthalmol* **47** (2002) 335-356.

- [111] D. A. Borkholder, J. Bao, N. I. Malufa, E. R. Perl, and G. T. A. Kovacs,
J. Neuro.Sci. Meth. **77** (1997) 61-66.