



**TÉCNICO**  
LISBOA

# **Computational Models, Neuronal Metrics and System Identification in Bioelectronic Vision**

**João Carlos da Silva Martins**

**Orientador: Doutor Leonel Augusto Seabra de Sousa**  
**Co-orientador: Doutor José Jasnau Caeiro**

**Tese especialmente elaborada para obtenção do Grau de Doutor em  
Engenharia Eletrotécnica e de Computadores**

**Tese Provisória**

**Novembro 2012**



## **Abstract**

The advances in signal processing technology and the development of biocompatible materials lead to the development of different types of new bioelectronic devices to circumvent human disabilities, opening the possibility to interface these prostheses directly with the nervous system. Among these devices are the bioelectronic vision systems which attempt to convey vision to profoundly blind people. Similarly to other neuroprostheses the development of bioelectronic vision systems imply the modeling of the signal processing occurring in the involved neuronal centers, namely the retina. Within this scope this thesis discusses and assesses several retina models, and proposes two additional models, representative of the most relevant models' classes, which further extends the available models and alternatives for retina modeling. A set of neuronal metrics, from different classes, are also discussed and applied to the tuning and assessment of neuronal models. The drawbacks of its application are discussed and several effective methods for the selection of the metric's parameters are provided. The last topic addressed in this thesis is system identification and parameter estimation of models, with the introduction of a new algorithm, based on multiple model adaptive estimators, that can be used in the identification and tuning of both linear and nonlinear systems, which are applied in the context of neuronal modeling.

## **Keywords**

Computational retina models, neural metrics, systems identification, bioelectronic vision systems.



## Resumo

Os avanços das técnicas de processamento de sinal e o desenvolvimento de novos materiais biocompatíveis permitem o desenvolvimento de diversos tipos de novos dispositivos bioeletrônicos para contornar diferentes deficiências humanas, abrindo a possibilidade de interface direta destes dispositivos com o sistema nervoso. Entre estes dispositivos estão os sistemas de visão bioeletrônicos que tentam restaurar algum tipo de visão a cegos profundos. No entanto, tal como é comum no desenvolvimento de neuropróteses, o desenvolvimento de sistemas de visão bioeletrônicos implica a modelação do processamento de sinal que ocorre nos centros neuronais envolvidos, sendo a retina de particular importância. Nesta tese discutem-se e avaliam-se diversos modelos de retina, sendo propostos dois modelos adicionais representativos das principais classes de modelos, conjuntamente com técnicas alternativas para a modelagem da retina. São igualmente analisadas e discutidas um conjunto de métricas neuronais, representativas das diferentes classes de métricas existentes, que são aplicadas na afinação e avaliação dos modelos neuronais. São apresentados os diversos inconvenientes da sua aplicação e são propostos métodos eficazes para a selecção dos parâmetros para estas métricas. Um último tópico abordado é a identificação de sistemas e a estimação de parâmetros de modelos, com a introdução de um novo algoritmo baseado na técnica de estimação adaptativa por múltiplos modelos. Esta técnica pode ser utilizada quer na identificação e ajuste de sistemas lineares como de sistemas não lineares, sendo aplicada no contexto da modelagem neuronal.

## Palavras-Chave

Modelos computacionais da retina, métricas para o código neuronal, identificação de sistemas biológicos, sistemas de visão bioeletrónica.



## Acknowledgments

I want to express my gratitude and love to my wife Elisabete and to our sons João Francisco and Ana Catarina, for their endless patience and comprehension. This time is for sure... it's almost done. I want also to thank, once again, my parents for their impact in my life, because they taught me what are the most important things in life, and for their example. A word of gratitude to my brothers Pedro and Rui for their friendship and support, it has been good to grow and learn many things with them.

A special thanks to my supervisor Prof. Leonel Sousa for the advice and guidance along my academic journey, and to Prof. José Caeiro who joined this journey lately but is also a longtime companion. Thanks also to the Signal Processing Systems (SiPS) research group at Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento em Lisboa (INESC-ID) for the friendly and challenging work environment.

A final word of appreciation to all my friends for being around and sharing their lives with me.



# Contents

<b>Contents</b>	<b>x</b>
<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xv</b>
<b>List of Acronyms</b>	<b>xix</b>
<b>Mathematical Notation</b>	<b>xxiii</b>
<b>1. Introduction</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.2. Main Objectives . . . . .	7
1.3. Main Contributions . . . . .	8
1.3.1. Development and Test of Retina Models . . . . .	8
1.3.2. Neural Metrics . . . . .	9
1.3.3. System Modeling and Identification . . . . .	10
1.4. Thesis Organization . . . . .	12
<b>2. The Human Visual System</b>	<b>15</b>
2.1. Introduction . . . . .	15
2.2. The Neuron . . . . .	15
2.2.1. Neuron Anatomy . . . . .	16
2.2.2. Neuron Dynamics . . . . .	17
2.2.3. The Neural Response Function . . . . .	21
2.3. The Human Visual System . . . . .	25
2.3.1. The Eye . . . . .	26
2.3.2. The Retina . . . . .	27
2.3.3. Retina Operation . . . . .	33
2.3.4. The Visual Pathway . . . . .	39
2.4. Conclusions . . . . .	42
<b>3. Neuron and Retina Models</b>	<b>45</b>
3.1. Introduction . . . . .	45

3.2.	The Retina Neural Code . . . . .	45
3.2.1.	Macroscopic Retinal Response Characteristics . . . . .	46
3.2.2.	Time Code or Rate Code . . . . .	48
3.2.3.	Classification of Retina Models . . . . .	49
3.2.4.	Retina Model's Assumptions . . . . .	50
3.3.	Structural Models Processing Blocks . . . . .	51
3.3.1.	The Hodgkin-Huxley Neuron Model . . . . .	52
3.3.2.	The Integrate and Fire Model . . . . .	59
3.3.3.	The Leaky Integrate-and-Fire Model . . . . .	63
3.3.4.	Stochastic Integrate-and-Fire Models . . . . .	67
3.4.	Modeling Input Stimuli . . . . .	70
3.4.1.	Synapses Modeling . . . . .	71
3.4.2.	Visual Stimuli Input Current . . . . .	73
3.5.	Retina Models . . . . .	73
3.5.1.	Deterministic Models . . . . .	74
3.5.2.	Stochastic Models . . . . .	79
3.5.3.	White Noise Model . . . . .	86
3.6.	Conclusions . . . . .	93
<b>4.</b>	<b>Modeling the Retina</b> . . . . .	<b>95</b>
4.1.	Introduction . . . . .	95
4.2.	A Static/Dynamic Retina Model . . . . .	95
4.2.1.	Model Structure . . . . .	96
4.2.2.	Model Implementation . . . . .	97
4.2.3.	Model Tuning . . . . .	100
4.2.4.	Experimental Results . . . . .	105
4.3.	A Multi Kernel LNP Model . . . . .	108
4.3.1.	Model Structure . . . . .	108
4.3.2.	Model Tuning . . . . .	112
4.3.3.	Experimental Results . . . . .	115
4.4.	Poisson Based Firing Rate Models versus SLIF Models . . . . .	117
4.5.	Conclusions . . . . .	119
<b>5.</b>	<b>Neural Activity Metrics</b> . . . . .	<b>121</b>
5.1.	Introduction . . . . .	121
5.2.	The Definition of Metric . . . . .	122
5.3.	Firing Rate Metrics . . . . .	123
5.3.1.	Mean Squared Error . . . . .	124
5.3.2.	Normalized Mean Squared Error . . . . .	125
5.3.3.	Percent Variance Accounted For . . . . .	125
5.3.4.	Analysis of the Firing Rate Metrics . . . . .	126
5.4.	Spike Train Metrics . . . . .	131
5.4.1.	Spike Time Metric . . . . .	132
5.4.2.	Interspike Interval Metric . . . . .	135

5.4.3.	Spike Train Distance Metric . . . . .	137
5.4.4.	Spike Train Metrics Analysis . . . . .	141
5.5.	Spike Events Metrics . . . . .	145
5.5.1.	Spike Events Metric Analysis . . . . .	152
5.6.	Tuning and Assessment of Retina Models . . . . .	157
5.7.	Conclusions . . . . .	164
<b>6.</b>	<b>System Identification and Parameters Estimation</b>	<b>167</b>
6.1.	Introduction . . . . .	167
6.2.	System Modeling and Identification . . . . .	167
6.2.1.	State Space Representation . . . . .	169
6.3.	Multiple Model Adaptive Estimators . . . . .	171
6.3.1.	MMAE Algorithm Setup and Properties . . . . .	171
6.3.2.	The Multiple Model Adaptive Estimator for Nonlinear Systems	173
6.4.	Constellation Based Multiple Model Adaptive Estimators . . . . .	175
6.4.1.	Design of the Models Constellation . . . . .	177
6.4.2.	System Identification and Parameters Estimation . . . . .	178
6.4.3.	Application to Nonstationary Systems . . . . .	180
6.5.	System State Estimators . . . . .	181
6.5.1.	The Kalman Filter . . . . .	181
6.5.2.	The Extended Kalman Filter . . . . .	183
6.5.3.	The Unscented Kalman Filter . . . . .	184
6.6.	Experimental Results . . . . .	186
6.6.1.	Linear System Identification . . . . .	186
6.6.2.	Nonlinear System Identification . . . . .	192
6.7.	Conclusions . . . . .	196
<b>7.</b>	<b>Conclusions</b>	<b>199</b>
	<b>Appendices</b>	<b>203</b>
<b>A.</b>	<b>Blindness In The World</b>	<b>205</b>
A.1.	Extent and Main Causes of Visual Impairment . . . . .	205
<b>B.</b>	<b>Characterization of the Neural Response</b>	<b>209</b>
B.1.	Spikes: The Essence of the Neural Code . . . . .	210
B.1.1.	Retina Stimulation and Responses Recording . . . . .	212
B.1.2.	Spike Trains and Firing Rates . . . . .	222
B.1.3.	Spike Triggered Average . . . . .	232
B.1.4.	Spike Train Autocorrelation Function . . . . .	237
B.1.5.	The Spike Triggered Covariance . . . . .	238
B.2.	Stimulus and Response Statistics, and Firing Probabilities . . . . .	239
B.2.1.	Spike Train Statistics . . . . .	240
B.2.2.	Homogeneous Poisson Model of Spike Trains . . . . .	242

B.2.3. Inhomogeneous Poisson Model of Spike Trains . . . . .	248
B.2.4. Firing Rates and Neuron Refraction . . . . .	251
B.3. Spiking Mechanisms . . . . .	254
B.3.1. Generation of Poisson Spike Trains . . . . .	254
B.3.2. Integrate-and-Fire Spike Generation . . . . .	255
B.4. Integration in the Leaky Integrate-and-Fire Model . . . . .	257
<b>C. Bioelectronic Vision Systems</b>	<b>261</b>
C.1. Bioelectronic Vision Labs . . . . .	261
C.2. Components of a Bioelectronic Vision System . . . . .	263
C.3. Types of Visual Prosthesis . . . . .	265
C.3.1. Retinal Neuroprosthesis . . . . .	266
C.3.2. Cortical Visual Neuroprosthesis . . . . .	268
C.4. Retinal Prostheses . . . . .	269
C.4.1. Epiretinal implants . . . . .	270
C.4.2. Subretinal implants . . . . .	272
C.5. Retinal Bioelectronic Vision System Design . . . . .	273
C.6. Cortical Visual Prostheses . . . . .	277
<b>Bibliography</b>	<b>285</b>

# List of Figures

1.1. Diagram of the human visual system. . . . .	4
1.2. Components of a visual neuroprosthesis. . . . .	4
1.3. Corticis cortical visual prosthesis concept. . . . .	7
2.1. Neuron anatomy. . . . .	17
2.2. Propagation of an action potential along a neuron's axon. . . . .	18
2.3. Neuron gap junction. . . . .	21
2.4. Representation of a spike train. . . . .	21
2.5. Spike waveform of a retinal ganglion cell. . . . .	22
2.6. Neural response function of a retinal ganglion cell. . . . .	22
2.7. Stimulus and spike train from an ON-type retinal ganglion cell. . . . .	23
2.8. Firing rate of an ON-type retinal ganglion cell. . . . .	24
2.9. Diagram section of the human eye. . . . .	26
2.10. Light micrograph of a vertical section through the retina. . . . .	28
2.11. Simplified schematic organization of the retina. . . . .	29
2.12. Spectral sensitivity of the photoreceptors. . . . .	30
2.13. Spatial distribution of photoreceptors. . . . .	31
2.14. Micrographs of the fovea cross section and of the foveal periphery. . . . .	32
2.15. Human retina photograph. . . . .	32
2.16. The receptive field of a cone. . . . .	34
2.17. Photoreceptor to bipolar cell connection. . . . .	35
2.18. Connections of a horizontal cell. . . . .	36
2.19. Connections between bipolar cells and ganglion cells. . . . .	37
2.20. AII amacrine cell connections. . . . .	38
2.21. Visual pathways . . . . .	40
2.22. The lateral geniculate nucleus (LGN) cell layers . . . . .	41
3.1. The squid giant neuron. . . . .	53
3.2. The Hodgkin-Huxley neuron model . . . . .	54
3.3. Hodgkin-Huxley steady-state conductances and decay time constants . . . . .	56
3.4. Action potential waveform generated by the Hodgkin-Huxley model. . . . .	57
3.5. Subthreshold action potential waveform of the Hodgkin-Huxley model. . . . .	58
3.6. Spike train and firing rate generated by the Hodgkin-Huxley model. . . . .	59
3.7. Gate dynamics in the Hodgkin-Huxley model. . . . .	60
3.8. The leaky integrate-and-fire (LI&F) model. . . . .	61

3.9.	Integrate-and-fire model response for constant stimulus current. . . .	62
3.10.	Leaky integrate-and-fire model response for constant stimulus current.	65
3.11.	Firing rate versus stimulus current for integrate-and-fire models. . . .	66
3.12.	General integrate-and-fire model of the retina. . . . .	73
3.13.	General LNP model of the retina. . . . .	74
3.14.	Block diagram of the deterministic model. . . . .	74
3.15.	Discrete spatial DoG. . . . .	77
3.16.	Block diagram of the pseudo-stochastic model. . . . .	80
3.17.	Distorted sinus base functions. . . . .	82
3.18.	Spike triggered average represented by basis functions. . . . .	83
3.19.	Generation of a noise sequence with a specific autocorrelation. . . . .	86
3.20.	The white noise model structure. . . . .	87
3.21.	White noise model of a salamander and of a rabbit RGC. . . . .	92
4.1.	Dynamic retina model. . . . .	96
4.2.	Spike response of the dynamic and static model of a rabbit RGC. . .	106
4.3.	Structure of Linear-Nonlinear-Poisson model. . . . .	108
4.4.	Eigenvalues ratio. . . . .	110
4.5.	Additive logistic retina model. . . . .	114
4.6.	NMSE for data sets for different polynomial orders. . . . .	116
4.7.	Firing rate comparison with the multi kernel model. . . . .	117
4.8.	Spike response for the additive logistic retina model. . . . .	118
5.1.	Comparison of two firing rates. . . . .	123
5.2.	Rabbit type-ON RGC responses for a ON-OFF full-field stimulus. . .	127
5.3.	Salamander RGC responses to white-noise full-field stimulus. . . . .	128
5.4.	Inter-trial evaluation for different smoothing parameters. . . . .	129
5.5.	Comparing smoothed and non-smoothed PSTHs. . . . .	131
5.6.	Comparison of two spike trains. . . . .	132
5.7.	Spike train transformation in the spike time metric. . . . .	133
5.8.	Spike train transformation path in the interspike interval metric. . . .	136
5.9.	Spike waveform shape comparison with the spike distance metric. . .	139
5.10.	Inter-trial evaluation for different temporal sensitivities. . . . .	143
5.11.	Inter-trial evaluation with the spike distance metric. . . . .	144
5.12.	Parsing a set of spike trains into firing events. . . . .	145
5.13.	Characterization of spike trains into firing events. . . . .	148
5.14.	Inter-trial evaluation of RGC responses for different smoothing values.	153
5.15.	Salamander RGC and SLIF model responses. . . . .	156
5.16.	Evaluation of model responses with spike time metric. . . . .	156
5.17.	Evaluation of model responses with the interspike interval metric. . .	157
5.18.	Evaluation of model responses with the spike events metric. . . . .	160
5.19.	Evaluation of model responses with the spike events metric. . . . .	160
5.20.	Evaluation of model responses with the spike train distance metric. .	161
5.21.	Evaluation of model responses with the spike train distance metric. .	161

5.22. Evaluation of model responses with NMSE metric . . . . .	162
5.23. Evaluation of model responses with the NMSE metric. . . . .	162
6.1. The MMAE algorithm block diagram. . . . .	172
6.2. Multiple model adaptive estimator and constellation adapter. . . . .	175
6.3. Models constellation topology and evolution. . . . .	177
6.4. Linear system dynamics in state-space. . . . .	182
6.5. Probabilities evolution of a set of models with the CBMMAE. . . . .	190
6.6. Time evolution of the system parameters' estimates for the linear system. . . . .	192
6.7. Estimation error of the linear system parameters. . . . .	193
6.8. Zooming of a parameter estimation and error. . . . .	194
6.9. Parameter estimation of a linear time-variant system linear system. . . . .	195
6.10. Estimates of the nonlinear system parameters. . . . .	196
6.11. Estimates errors for the nonlinear system parameters. . . . .	197
A.1. Global causes of blindness. . . . .	206
B.1. Representation of a spike train. . . . .	210
B.2. Spike waveform of a rabbit's retinal ganglion cell. . . . .	210
B.3. Neuronal response function of a retinal ganglion cell. . . . .	211
B.4. Cellular recording of neuronal signals. . . . .	212
B.5. Retinal ganglion cell responses for a ON-OFF full-field stimulus. . . . .	214
B.6. Retinal ganglion cell responses for white-noise full-field stimulus. . . . .	215
B.7. Spatially uniform visual stimuli. . . . .	216
B.8. Spatially non-uniform visual stimuli. . . . .	217
B.9. Spatially nonuniform Gabor functions. . . . .	218
B.10. Stimuli with spatial and temporal modulation. . . . .	218
B.11. Multi-dimensional Gaussian white noise stimulus. . . . .	219
B.12. Spatially non-uniform visual stimuli . . . . .	220
B.13. Microelectrode array. . . . .	220
B.14. Experimental apparatus for retina data acquisition and analysis. . . . .	221
B.15. Retina preparation for data acquisition. . . . .	222
B.16. The $\delta_{\Delta}$ function. . . . .	223
B.17. Stimulus and spike trains from a ON-type retinal ganglion cell. . . . .	225
B.18. Firing rate ON-type RGC. . . . .	226
B.19. The rectangular (boxcar) filter window. . . . .	229
B.20. The Gaussian filter window. . . . .	230
B.21. The $\alpha$ function filter. . . . .	231
B.22. Firing rate for neural responses filtered with different windows. . . . .	232
B.23. Procedure for the STA computation. . . . .	233
B.24. Time reversed spike triggered average . . . . .	234
B.25. Spike number probability density for a Poisson spike train. . . . .	243
B.26. Interspike time interval probability density of a Poisson spike train. . . . .	243
B.27. Integrate-and-fire spike generation from firing rate. . . . .	257

B.28. The leaky integrate-and-fire (LI&F) model. . . . .	257
C.1. World map of vision neuroprosthesis research groups. . . . .	262
C.2. Components of a visual neuroprosthesis. . . . .	263
C.3. Main components of a retinal neuroprosthesis . . . . .	267
C.4. Cortivis intracortical visual prosthesis. . . . .	270
C.5. IRP test device. . . . .	274
C.6. Layout of a retinal implant. . . . .	276
C.7. Electrode layout. . . . .	277
C.8. The Dobbelle visual prosthesis. . . . .	278
C.9. Microelectrode array. . . . .	279
C.10. The CORTIVIS bioelectronic vision system. . . . .	280

# List of Tables

3.1. Hodgkin-Huxley model parameters . . . . .	57
4.1. Error between static/dynamic model and RGC responses. . . . .	107
4.2. Error for the generalized additive retina model. . . . .	116
5.1. Values of firing rate metrics for RGC responses. . . . .	131
5.2. Limit values of spike train metrics for RGC responses. . . . .	141
5.3. Mean and limit values of spike events metric for RGC responses. . . . .	155
6.1. System modeling methodologies. . . . .	169
6.2. Neuron model parameters . . . . .	188
C.1. Main pros and cons of visual prostheses approaches. . . . .	284



# Abbreviations and Acronyms

<b>AGC</b>	automatic gain control
<b>AMD</b>	age-related macular degeneration
<b>CBMMAE</b>	constellation based multiple-model adaptive estimation
<b>CCD</b>	charged couple device
<b>CDF</b>	cumulative density function
<b>CGC</b>	contrast gain control
<b>CORTIVIS</b>	Cortical Visual Neuroprosthesis for the Blind
<b>CRT</b>	cathode ray tube
<b>DC</b>	direct current
<b>DoG</b>	difference of Gaussians
<b>EKF</b>	extended Kalman filter
<b>EM</b>	Expectation-Maximization
<b>EPSC</b>	excitatory post-synaptic current
<b>fps</b>	frames-per-second
<b>GCL</b>	ganglion cell layer
<b>GS</b>	Gaussian Sum filter
<b>HH</b>	Hodgkin-Huxley model
<b>I&amp;F</b>	integrate-and-fire
<b>IF</b>	integrate-and-fire model
<b>IIR</b>	infinite impulse response
<b>INL</b>	inner nuclear layer

**IPL** inner plexiform layer  
**IPSC** inhibitory post-synaptic current  
**IRP** Intraocular Retinal Prosthesis  
**ISI** interspike time interval  
**KF** Kalman filter  
**LCD** liquid crystal display  
**LGN** lateral geniculate nucleus  
**LIF** leaky integrate-and-fire model  
**LNP** linear-nonlinear-Poisson  
**LoG** Laplacian of Gaussian  
**MEA** microelectrode array  
**MEMS** microelectromechanical systems  
**ML** maximum likelihood  
**MMAE** multiple-model adaptive estimation  
**MSE** mean squared error  
**NIR** near-infrared  
**NMSE** normalized mean squared error  
**ONL** outer nuclear layer  
**OPL** outer plexiform layer  
**PCA** principal component analysis  
**pdf** probability density function  
**pixel** picture element  
**pmf** probability mass function  
**PSTH** peri-stimulus time histogram  
**RF** receptive field  
**RGC** retinal ganglion cell

**RNC** Retina Neural Code  
**RP** retinitis pigmentosa  
**SLIF** stochastic leaky integrate-and-fire model  
**STA** spike triggered average  
**STC** spike triggered covariance  
**TFT** thin-film transistor  
**UKF** unscented Kalman filter  
**V1** cortex visual area 1  
**%VAF** percent-Variance-Accounted-For  
**WHO** World Health Organization



# Mathematical Notation

Symbol	Meaning
$x, y$	real/complex variable
$\mathbf{x}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\phi}$	column vectors (bold lowercase Roman/Greek alphabet letters)
$\mathbf{A}, \mathbf{B}, \boldsymbol{\Theta}, \boldsymbol{\Phi}$	matrices (bold uppercase Roman/Greek alphabet letters)
$\emptyset$	empty set
$\in$	in the set
$\notin$	not in the set
$x(t)$	continuous signal denoted by the independent variable $t$ inside parentheses, $t \in \mathbb{R}$
$x[n]$ or $x[t]$	discrete signal/sequence denoted by the independent variable $n$ or $t$ inside brackets, $n, t \in \mathbb{Z}$
$\lceil x \rceil$	ceiling function. $\lceil x \rceil$ is the smallest integer not less than $x$
$\lfloor x \rfloor$	floor function. $\lfloor x \rfloor$ is the largest integer not greater than $x$
$\ln(x)$	the natural logarithm (Nepper base) of $x \in \mathbb{R}^+$ : $\ln(e^x) = x$ ;
$\delta(t)$	the generalized continuous Dirac delta function: $\delta(t) = 0, \quad t \neq 0; \quad \int_{-\infty}^{+\infty} \delta(\tau) d\tau = 1;$ properties: $\int_{-\infty}^{+\infty} f(\tau) \delta(x - \tau) d\tau = f(x);$ $\delta(t) = \frac{dH(t)}{dt}$
$H(t)$	the Heaviside step function: $H(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$ $H(t) = \int_{-\infty}^t \delta(\tau) d\tau$

$N(x; \mu, \sigma^2)$	<p>normal Gaussian function with mean <math>\mu</math> and variance <math>\sigma^2</math>:</p> $N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}$ $N(x; \mu, \sigma^2) = N((x - \mu)\sigma; 0, 1)$
$N(\mathbf{x}; \bar{\mathbf{x}}, \Sigma)$	<p>vectorial normal probability density function with mean vector <math>\bar{\mathbf{x}}</math> and covariance matrix <math>\Sigma</math>:</p> $N(\mathbf{x}; \bar{\mathbf{x}}, \Sigma) = \frac{1}{(2\pi)^{n/2} \sqrt{ \Sigma }} e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T \Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}})}, \quad \mathbf{x} \in \mathbb{R}^n$
$\text{erf}(x)$	<p>error function:</p> $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt; \quad x \in \mathbb{R}$
$\Phi(x)$	<p>standard normal cumulative density function:</p> $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt; \quad x \in \mathbb{R}$ $\Phi(x) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x}{\sqrt{2}} \right) \right]$
$\Phi(x; \mu, \sigma^2)$	<p>normal cumulative density function with mean <math>\mu</math> and variance <math>\sigma^2</math>:</p> $\Phi(x; \mu, \sigma^2) = \int_{-\infty}^x N(t; \mu, \sigma^2) dt$ $= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt, \quad x \in \mathbb{R}$ $= \Phi \left( \frac{x - \mu}{\sigma}; 0, 1 \right) = \Phi \left( \frac{x - \mu}{\sigma} \right);$ $\Phi(x; \mu, \sigma^2) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x - \mu}{\sqrt{2}\sigma} \right) \right]$
$\delta[n]$ or $\delta_{ij}$	<p>the discrete Kronecker delta function:</p> $\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \quad \text{or} \quad \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$ $\delta[n] = H[n] - H[n - 1]$
$H[n]$	<p>discrete unit step function:</p> $H[n] = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases},$ $H[n] = \sum_{k=-\infty}^n \delta[k]$
$y(t) = x(t) * h(t)$	<p>continuous convolution operation:</p> $y(t) = \int_{-\infty}^{+\infty} x(\tau) h(t - \tau) d\tau$

$y[t] = x[t] * h[t]$	discrete convolution operation: $y[t] = \sum_{\tau=-\infty}^{+\infty} x(\tau)h(t - \tau)$
$\mathbf{A}_{i*}$ and $\mathbf{A}_{*j}$	denotes the $i^{\text{th}}$ row and the $j^{\text{th}}$ column of matrix $\mathbf{A}$ , respectively
$\mathbf{x}_i$	denotes the $i^{\text{th}}$ element of the column vector $\mathbf{x}$
$\mathbf{I}$	identity matrix
$\mathbf{0}$	null matrix
$\mathbf{A}^T$	matrix transpose of $\mathbf{A}$
$\mathbf{A}^{-1}$	matrix inverse of $\mathbf{A}$
$\mathbf{A}_{n \times m}[n] * x[n]$	columnwise discrete convolution of matrix $\mathbf{A}_{n \times m}$ with the signal $x[n]$ $\mathbf{A}_{n \times m}[n] * x[n] = [\mathbf{A}_{*1}[n] * x[n] \cdots \mathbf{A}_{*m}[n] * x[n]]^T$
$\ \mathbf{x}\ $	norm of vector $\mathbf{x}$
$\mathbf{N}$	set of natural numbers
$\mathbf{Z}$	set of integer numbers
$\mathbf{R}$	set of real numbers
$\mathbf{C}$	set of complex numbers
$\mathbf{R}^n$	$n$ -dimensional space of real numbers
$A \times B$	cartesian product of $A$ and $B$
$f : A \mapsto B$	mapping of set $A$ to $B$ by the function $f$
$F[A]$	image of $A$
$P[x]$	probability of the event $x$
$P(x)$	discrete probability distribution (probability mass function) of the random variable $x$
$p(x)$	continuous probability density function of the random variable $x$
$\arg \max_x f(x)$	value of $x$ for which the function $f(x)$ takes its maximum
$\rho(t)$	continuous neural response function: $\rho(t) = \sum_i \delta(t - t_i)$
$\rho[n]$	discrete neural response function: $\rho[n] = \sum_{i=1}^n \delta[n - n_i]$
$r(t)$	firing rate [spikes/s] or [Hz] $r(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle \rho(\tau) \rangle d\tau$ $= \left\langle \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \rho(\tau) d\tau \right\rangle$



*Learning is the only thing the mind never exhausts,  
never fears, and never regrets.*

Leonardo da Vinci  
(Vinci, Italy, 1452 – Amboise, France, 1519)

*The more I learn, the more I realize how much I do  
not know.*

Socrates  
(Athens, Greece, 469 a.C. – 399 a.C.)

# 1

## Introduction

### 1.1. Motivation

**I**N recent years the development of prosthetic devices to circumvent diverse human impairments has experienced an increasing interest. Several research institutions, mostly universities, have been involved in the development of this area, and important professional organizations have promoted a series of relevant publications about these subjects (like [IEEE, 2012]), with an exponential growth in the number of projects and papers during the last years.

A particular area of this research is concerned with retina modeling and the development of visual prostheses in order to convey some kind of vision to visually impaired people. The expression "some kind of vision" is frequently employed by scientists when referring to this goal and reflects the huge extension of this task due to the complexity of the human visual system, ranging from the retina's neural circuitry to the involved deep brain processes.

Vision is a fundamental sense of paramount relevance in modern society. It is through vision that the human being perceives much of the surrounding information, feels diverse forms of art and communicates with other human beings. Nowadays, in modern society, much of the information is supported on images and video putting visually impaired people in great disadvantage. However, the greatest difficulties arise in individual terms.

For everyday life, vision is an indispensable resource. For performing even the most simple tasks vision is indispensable, therefore blindness represents a severe impairment. People who lost their sight usually become very dependent of others, which constitutes a social problem. For blind people it is very difficult to recognize other people, perceive landscapes, and discern objects of the daily life. In addition, the visually impaired usually have severe motion restrictions and depend on others

## 1. Introduction

to move safely. People who have been able to see for years lose a relevant part of their quality of life when they become blind. Loss of vision is not only an enormous psychological burden, but it also causes severe handicaps and tremendous difficulties in moving in strange, and even in formerly familiar, environments.

Human vision disabilities can be of different types and have different origins. Thus, depending on the nature of the disability, different approaches have been used to circumvent vision impairments. Impairments in the eye's optical system, which is responsible for transmitting and focusing light as a sharp image on the retina, usually are usually easily overcome with the use of an external corrective optical system. Optical lenses, like glasses or contact lenses, are a typical optical choice to remedy this problem. Another option is surgical intervention, as in the case of cataracts, where an eye lens transplant can be performed<sup>1</sup>.

The current challenge is to circumvent damages at the retina and superior vision center levels, that frequently lead to profound blindness. Besides traumatic accidents several diseases, like diabetic retinopathy or glaucoma, age-related macular degeneration (AMD) and retinitis pigmentosa (RP) can damage the retina and the optic nerve permanently. In the case of profound blindness the remaining hope is to combine the increasing knowledge about the biology and anatomy of the human visual system with the amazing advances in science and technology. The combination of knowledge from these domains towards the development of vision prosthesis established an emerging field designated by bioelectronic vision.

Several projects involving multidisciplinary research groups have been promoted to develop and demonstrate the feasibility of artificial vision systems. The huge challenge of artificially restoring vision to the blind poses engineering and biological problems hard to overcome and requiring at the end the meandering process, in personal, legal and ethical terms, of clinical human testing. Due to the initial cost of such prostheses, mainly people living in industrialized countries are expected to be the first beneficiaries from such devices, when bioelectronic vision becomes possible through the design of visual neuroprostheses. Therefore, blind people affected by diabetes or age-related macular degeneration (AMD) and retinitis pigmentosa (RP) should be the ones to first take advantage of visual neuroprostheses, retinal or cortical, while people suffering from glaucoma could only take advantage of cortical neuroprosthesis due to the generalized damage of the retina neural layer and optic nerve.

Retinal neuroprostheses require the presence of viable cells in the inner retina. Therefore, diseases limited primarily to the outer retina are potentially treatable with a retinal neuroprosthesis. The references [Margalit et al., 2002] and [Weiland et al., 2005] present an extended overview of retinal neuroprostheses. A retinal neuroprosthesis example, entitled the "Bionic Eye", uses a new ceramic material to substitute the retina's photoreceptors, that acts as an optic detector transducing light into electrical impulses, by means of the photo-ferroelectric effect [Wu, 2006]. An-

---

<sup>1</sup>Appendix A lists the most common causes of blindness with a description of the diseases associated and its prevalence.

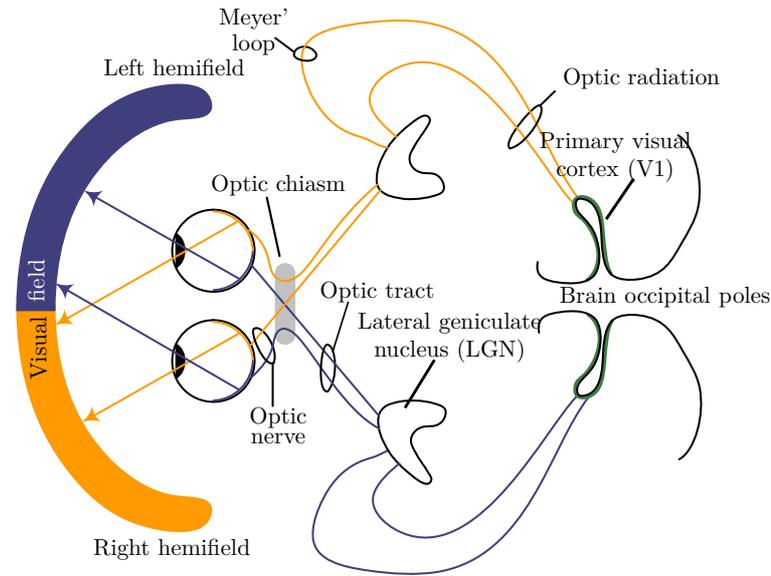
other example of a retinal neuroprosthesis is the artificial silicon retina (ASR) microchip [Chow et al., 2004], which uses the well known and developed silicon technology to design a device that tries to mimic the retina [Optobionics Corporation, 2012].

For profoundly blind people, where the earlier visual centers like the retina and optic nerve are irreversibly damaged, the remaining solution to provide a vision sense is to stimulate directly the superior vision centers in the brain. The efforts for conveying some kind of vision to profoundly blind people already have some history [Rizzo III and Wyatt, 1997]. The first permanent device developed and applied for chronic stimulation of the neural tissue was developed in 1968 [Brindley and Lewin, 1968]. Although it was observed that the electrical stimulation of the occipital lobe of the human cortex causes a subject to perceive phosphenes, that are on the base of the visual sensations, it was also concluded that the indiscriminate injection of electrical current into the brain can produce short and long term complications [Agnew and McCreery, 1990]. The Dobbelle Institute, founded by the pioneer in artificial vision William Dobbelle, conducted a series of experiments earlier in 1974 [Dobbelle and Mladejovsky, 1974; Dobbelle et al., 1974] where blind patients have been implanted with cortical neuroprostheses. Despite a relative success, it also shown that focal epileptic activity can be induced by electrical stimulation.

Therefore, in what concerns brain stimulation the solution points to the use of microelectrode arrays made available by the semiconductor technology manufacturing, such as the microelectrode array, developed in the Utah University and known as the Utah Electrode Array, to stimulate the visual neuronal structures [Maynard, 2001]. These microelectrodes can be deeply inserted in the virtual cortex by using a pneumatic insertion technique [Maynard et al., 1997; Normann et al., 1999], for intracortical stimulation with very low and controlled electrical currents without provoking major injuries. Presently, silicon micromachining and micromanufacturing technologies allow the fabrication of very small arrays with a large number of microelectrodes capable of stimulating only the neurons nearest to the electrode and with a small amount of current (in the order of dozens of micro ampere). The reference [Donoghue, 2002] provides a general perspective of cortex electronic interfaces.

Figure 1.1 shows a schematic overview of the entire visual system, from the eye to the brain. The foremost component of the visual system is the eye. The eye is responsible for gathering and transducing light energy— photons— to neural electrical signals, in the form of electrical impulses, that are successively transformed and posteriorly transmitted to the brain for further information extraction. Roughly speaking, the eye is composed by an optical system that focuses light onto the retina, a neuronal tissue. In the retina the light patterns are encoded into electrical signals where the neuronal processing starts, so that the retina can be seen as an extension of the brain, the only brain’s accessible component outside the skull. A visual neuroprosthesis must start by mimicking the optical eye system, which does not pose any major technological difficulties nowadays, and mimic the neuronal processing occurring in the retina, where the true challenge lies. Therefore, a significant stage of a vision prosthesis design comprises the choice and evaluation of a retina neural

## 1. Introduction



**Figure 1.1.** Diagram of the human visual system.

processing model.

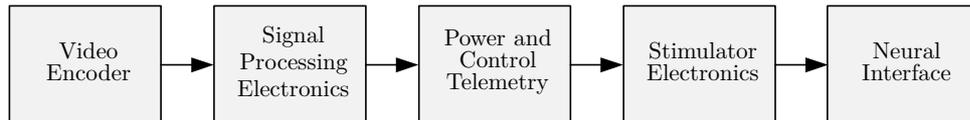
Bioelectronic vision systems can be classified in two main classes of visual neuroprostheses: *i*) retinal neuroprostheses, and *ii*) cortical neuroprostheses.<sup>2</sup> Retina neuroprostheses are suitable only when the front end of the retina is functioning properly. When the retina is not functioning, including the optic nerve, and only the brain visual centers remain intact the implantation of a cortical neuroprosthesis is the last hope. For profoundly blind people, whose optical neurotransmitters are irreversibly damaged, cortical neuroprostheses are interfaced directly with the visual processing center in the brain, known as the visual cortex area V1, circumventing the earlier vision processing centers.

The concept and components of a bioelectronic vision system supported on a complete visual neuroprosthesis that interfaces directly with the brain is depicted in Fig. 1.3. Figure 1.2 depicts the main components of a bioelectronic vision system, and it includes a set of components which, depending on the type of visual neuroprosthesis, can be the biological structure itself or its electronic circuit counterpart.<sup>3</sup>

Image acquisition can be achieved by a generic small and full functional digital camera that is suited for a vision prosthesis, both for cortical and optic nerve neu-

<sup>2</sup>Section C.3 presents a more detailed description of the various types of visual prosthesis.

<sup>3</sup>Section C.2 describes the components of a bioelectronic vision system in more detail.



**Figure 1.2.** Components of a visual neuroprosthesis.

roprostheses, in terms of dynamic range, sensitivity and depth of field and, as also important, in terms of aesthetic. For retinal neuroprostheses the image encoder can be integrated into the neural interface, and lay at the plane of the retina, with the advantage that the eye optics can be used to project the image in the encoder.

The digital signal processing system has to transform the visual space image into a set of discrete signals, modeling the neuronal signal processing occurring in the retina, and should take into account the visuotopic organization of the target structure (the retina or the cerebral cortex, depending on the prosthesis type). Then a module is used to transmit power, control signals, and the encoded visual stimulus to the implanted electronics to induce the perception of phosphenes – the entoptic phenomenon characterized by the sensation of seeing light.

In the signal processing block the more difficult challenge is the mapping of the visual space into the visuotopic organization of the target structure, particularly the visual cortex. This is an even somewhat complicated task due to the uniqueness of this map among individuals, and because it is conformal only at the low resolution level; for high spatial resolutions this mapping seems to be locally random. Therefore, parameterizable models have to be developed for implementing this module and properly stimulating individuals. This is a somewhat more complicated task due to the plasticity of the visual pathways and the different possible combinations between electrodes and phosphenes elicited. Based on the developed models, this module transforms the image into a discrete set of signals that drive the stimulators. To adapt the intensity of the incoming light signals into the range level of the neurons being stimulated an automatic gain control (AGC) can perform similarly to what is done by the photoreceptors. The first three components of the visual neuroprosthesis in Fig. 1.2 are susceptible of being included into a single device attached to a set of eyeglasses. The remaining blocks of the visual neuroprosthesis are likely to be located inside the patient.

The information can be transmitted to the neural interface using a percutaneous connector or using receptive field (RF) telemetry. The neural stimulator receives power, encoded data, and control signals from the connection. It must be capable of exciting multiple electrodes at the same time to evoke consistent phosphenes, and must be capable of controlling the amount of power delivered to avoid the damage of the surrounding tissues, and should also be capable to disable nonfunctional electrodes.

The last element in a vision prosthesis is the interface with the nervous system. This interface makes the bridge between the nervous system and the signal processing electronics. It mediates the transduction between the electrical currents generated by the electronic device into ionic currents that flow inside the human body. The materials employed in the fabrication of retinal neuroprosthesis are silicon chips and specific ceramic materials. For cortical interfaces the oxidized iridium is a candidate material since it has shown good biocompatibility and a good electronic to ionic current transducer. Although, there are many compatibility issues that should be taken into account.

Resuming, when the visual signal reaches the optic nerve or the visual cortex in

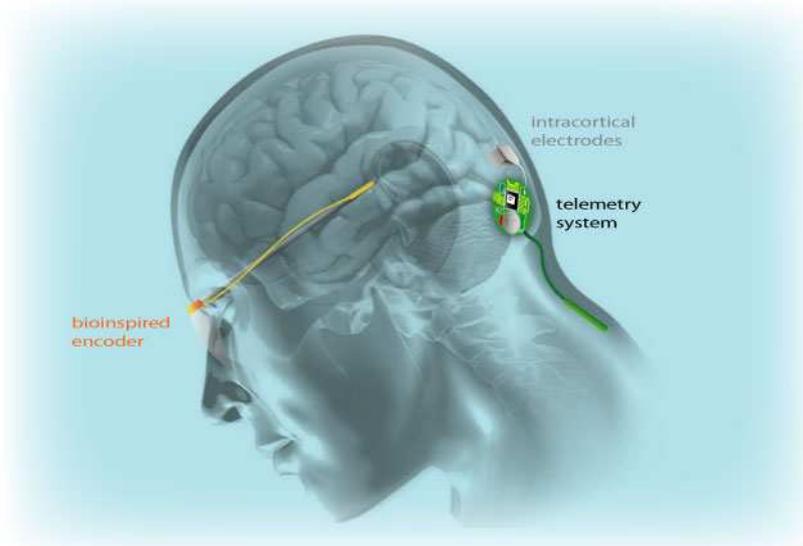
## 1. Introduction

a healthy person, it has been subjected to a series of neural processing stages. A vision prosthesis must mimic these processing stages. When the interface to the visual system is made at the level of the retina's ganglion cell layer, as for the case of an epiretinal neuroprosthesis, the information at the output must be identical to the one produced by a healthy retina; the transformation of the visual space to the retinotopic space is done by modeling the neural processing of the retina. For the case of a cortical neuroprosthesis the signal processing occurring along the visual pathway should be adequate to generate the proper stimulus for the neural interface.

The discovery that by direct electrical stimulation of the visual centers, that range from the eye to the brain visual cortex, it was possible to elicit visual sensations - the phosphenes [Bair and O'Keefe, 1998], triggered the research and development of devices capable of providing some kind of vision. This endeavor is being recently boosted by the advances in electronics and materials science. Research is being done to design and develop cortical visual neuroprostheses through intracortical stimulation, but none of these prostheses has been permanently applied for chronic stimulation [Warren and Normann, 2003].

The European project "Cortical Visual Neuroprosthesis for the Blind (**CORTIVIS**)" has been conducted over the last few years to design and develop a complete visual neuroprosthesis designed to restore useful vision to profoundly blind people [Project CORTIVIS, 2006]. Namely, the discrimination of shape and location of objects, allowing the navigation in a familiar environment, and to read enlarged text, resulting in a substantial improvement in the standard of living of blind and visually impaired persons. In this project intracortical microstimulation through one or more Utah Electrode Arrays implanted into the primary visual cortex is performed. The system is composed of a primary unit located outside the body and a secondary unit, implanted inside the body, that communicate with each other using wireless communication technology. A prototype of the proposed system including all components is described in detail in [Martins and Sousa, 2009]. A device preview is shown in Fig. 1.3, where the interface with the visual system is made at the cortical level, by intracortical microstimulation using the Utah Microelectrode Array and the power and control signals are transmitted by a radio frequency link. In the paper by [Piedade et al., 2005] the different electronic modules developed for the prosthesis are described. A relevant part of the project was the evaluation and choice of the retina's processing model to be implemented in the signal processing module, to convert the image of the visual field to the set of discrete signals for the neural stimulation at the visual cortex area.

Motivated by the open questions left by the **CORTIVIS** project, a second project termed "Retina Neural Code (**RNC**)", funded by the Portuguese Foundation for Science and Technology (FCT), was developed in the meanwhile [Project RNC, 2008]. The main goal of the **RNC** project was to further unveil the coding mechanisms of the retina by investigating more accurate models and to study the metrics employed in the evaluation of such models.



**Figure 1.3.** The Cortivis cortical visual prosthesis concept [Project CORTIVIS, 2006].

## 1.2. Main Objectives

A first goal of this thesis is to investigate and analyze retina models from the most representative classes suitable to be used in a bioelectronic vision system. After this first step a subsequent goal is to develop and propose improvements into the models. These improvements follow two directions: the proposal of alternative structures for the models such that the response accuracy to visual stimuli is improved, and afterwards to research and apply alternative tools, that have shown good results in other scientific areas, to the estimation and tuning of retina models.

Another main objective is centered in the analysis of neuronal metrics. Among the neuroscientists' community there is not a consensus about the coding and decoding of the neural code performed by the nervous centers. The studies on this subject did not come to a final conclusion so far of what are the relevant features of the neuronal code used by the neuronal centers to encode and extract information, since the neuronal signals are composed by a temporal series of stereotyped events, called action potentials or spikes, which are modeled as a Dirac delta function occurring at a precise time instant. In the scientific literature on neural centers models, and vision centers in particular, several metrics are employed that explore a particular characteristic of the neuronal response and are applied in a somewhat *ad-hoc* manner. The goal within this context is to gather the neuronal metrics appearing in the literature, and study its applicability in the tuning and assessment of retina models. An important issue is the devising of methods for the selection of meaningful values for the metrics' parameters instead of the *ad-hoc* procedure followed until now.

Another objective, which is closely related to the first one, was to apply and develop alternative methods to the estimation of system models, and neuronal retina models in particular. There is a huge number of methods and algorithms currently available

## 1. Introduction

that are used in engineering to model and estimate the most diverse systems. Among these algorithms are adaptive techniques based on linear and nonlinear estimates of the system's state to infer its parameters. These algorithms are normally based on the Kalman filter and on its several flavors. This thesis proposes and describes a new method, based on the multiple model adaptive estimators, to estimate the neuron state and training its parameters. Starting from the classical perspective, and using a linear system model, this method is extended to nonlinear systems by applying nonlinear estimators, like the extended Kalman filter [Gelb, 1974] and the unscented Kalman filter [Julier and Uhlmann, 1997].

### 1.3. Main Contributions

As stated before, the work presented in this thesis is focused in three main areas related to the field of artificial vision. First, in the modeling of visual systems, particularly the neural processing occurring in the retina. The assessment of the retina models leads to another important topic that is the organization and analysis of metrics to evaluate the neuronal activity, which is the second main topic developed. A third topic pursued is the identification and tuning of systems parameters, with a special emphasis on neuronal models, with the development of a new method to identify system models and estimate its parameters.

#### 1.3.1. Development and Test of Retina Models

The accurate modeling of the retina behavior is of paramount importance for the development of visual prosthesis. Several types of approaches and models have been proposed by scientists. The majority of the neuronal models, including the retina, follow one of two principal approaches: a (stochastic) integrate and fire mechanism or a nonlinear Poisson process.

However, most of the models are concerned with the description of a particular phenomena and are not meant to be used in a prosthesis. In a preliminary stage a series of representative retina models were collected and assessed using different types of stimuli. In a second stage these two types of models were further investigated. For the stochastic integrate and fire model it was demonstrated that if certain noise conditions are met, namely if the noise process is of Gaussian nature (which is a mild condition due to the central limit theorem) it is equivalent to a Poisson model. Within this framework a dynamic model for the retina is proposed that besides the usual feedforward and feedback mechanisms also includes a dynamic feedforward and feedback path, both dependent on input stimulus and on spike history, where the amplitude is modulated by a linear function of the stimulus and spike history. This model is shown to have a better response, in terms of several error metrics, against the typical static model by increasing the spiking time precision. These results were published in [Tomás et al., 2008].

White noise analysis techniques were also applied to the specific modeling of the

retina. This time a canonical linear-nonlinear-Poisson (LNP) model was chosen and white-noise analysis was applied to obtain the linear and non-linear filter kernels. This model departs from the classical LNP model, which uses a single linear component stimulus (see [Chichilnisky, 2001]), by including several linear components of the stimulus that are individually combined by nonlinear kernels to generate the firing potential. To avoid the need of a huge amount of neuronal data, usually needed to populate an  $n$ -dimensional histogram, it was proposed the use of generalized additive models to model the nonlinear functionals that transform the linear kernels. These nonlinear functionals are estimated with the *weighted backfitting algorithm*. This approach has the advantage that permits to plainly include both excitatory and suppressive kernels in the model, making the retina's model response more accurate. This model, and the obtained results, are described in [Martins et al., 2007].

Resuming, the first stage of this work allowed to investigate and further develop retina models by proposing two additional modeling approaches within the two main frameworks of retina models.

### 1.3.2. Neural Metrics

Although not completely independent of the previous topic, a second important stage of this work was the organization and experimental analysis of neural metrics and selection of appropriate values for its parameters.

In general, in engineering to know and measure the accuracy of how a model mimics a given system is of paramount importance. The evaluation of retina models raises a few questions. What metric should be used to evaluate a retina model? Are these metrics similar to the common metrics employed in other engineering fields? How the results from system modeling and identification methods used in engineering can be used in the development of retina models? The answers to these questions lead to the organization and experimental analysis of a set of neural metrics and methods for the appropriate selection of its parameters' values.

Many of the neural metrics proposed in the literature have been developed for the purpose of studying particular aspects of the neural code or to compare the responses of specific neural systems, so that their application is not general. Furthermore, no quantitative measurements or methods are provided to decide which metric should be used in a given context. These metrics normally depend on one or more free parameters and no clues or methods existed to choose these parameters properly. The followed approach was to use these metrics for directly comparing neural responses and study their sensitivity to the different parameters, and applying the results to the tuning and assessment of retina models, where some metrics have shown to be more appropriate than others.

Since the neuronal metrics are targeted to measure different characteristics of the neuronal response, they can be classified into one of the classes: firing rate metrics, spike train metrics, and firing event metrics.

The firing rate metrics measure the reliability of the neural code in terms of the firing rate by comparing their mean firing rate. These metrics are close to the ones

## 1. Introduction

found in the engineering fields like the mean-squared error. This class comprises three metrics, namely the mean squared error (**MSE**), the normalized mean squared error (**NMSE**) and the percent-Variance-Accounted-For (**%VAF**). The main problem with this type of metrics relies in the calculation of the mean firing rate, since the neural response is impulsive (series of impulse Dirac delta functions) it becomes very dependent on the size of the time bin used. To avoid these problems the firing rates are smoothed before being compared, a process that should also be done with care. Namely, for the smoothing of the firing rate a Gaussian function is used, with zero mean and standard deviation  $\sigma$ . In order not to remove the higher frequency components of the firing rate,  $\sigma$  must be chosen has the minimum value that minimizes the inter-trial comparison, which has the advantage that as the number of trials used in the ensemble average increases the chosen value for  $\sigma$  decreases and the accuracy of the firing rate estimate increases. The analysis of the **%VAF** neuronal metric resulted in the conclusion that it should not be used since it can give a deceiving value for the error.

The spike train metrics measure the time precision in the occurrence of individual spikes, so that they compare directly the spike trains or an adequate altered version of them. This class of metrics include the *spike time metric*, the *spike interval metric* and the *spike distance metric*. The most surprising fact about the first two metrics is that dependent on the value of its free parameter, the error (metric distance) between a true spike train and between a null spike train (train with no spikes) can be smaller than between two trains from the same neuron generated with the same stimuli. To chose the parameter once again the intertrial error must be obtained first and then the value of the parameter must be such that the it is smaller than the comparison with the null train. Concerning the spike distance metric its parameter must take the value which gives the best temporal precision from a prior intertrial error measure.

Finally a spike events metric is also analyzed. This metric measures the distance between bursts of spikes from sets of spike trains. Its has several sensible steps ranging from the obtention of the peri-stimulus time histogram, with the implicit choice of the standard deviation for the smoothing function, to the delimitation of the events (a bunch of spikes with particular characteristics), which are interconnected and are very sensible to each other.

The main conclusions to be drawn from the work presented in this thesis are that the optimum parameters for the metrics' parameters depend on the data sets, however a general set of rules was obtained for selecting its value. These results are presented in [Martins et al., 2009] with subsequent applications to the tuning and evaluation of a series of representative retina neural models.

### 1.3.3. System Modeling and Identification

In the topic of system modeling and identification, this thesis proposes a system identification and a parameter tuning technique. In the effort to develop retina models, a method from the engineering field is applied to a state-space description of a neuron model whose parameters are estimated using multiple-model adaptive

estimation, and by employing a developed method for the selection and refinement of the models' parameters.

The multiple-model adaptive estimation (**MMAE**) have been used as a standard technique in system identification and state estimation. In its standard utilization the **MMAE** selects the system model with the highest probability by following a Bayesian criteria from a fixed set of pre-established models.

In the proposed approach a set of models is established with very few restrictive conditions. For  $N$  unknown parameters the number of models in the set must be  $2^N + 1$ , independently of the different values that the parameters can take. In the standard application of the **MMAE** for  $N$  parameters that can take  $M$  different values independently, the number of models to be used is  $M^N$ , so that the proposed approach is a minimum value for this expression. The constructed model set constitutes what is called a constellation of models. Relying in the properties of the **MMAE**, the model's constellation is shifted in the unknown parameters' space, a process designated by tracking, until the unknown parameters point is encircled by the constellation, a process termed bracketing, afterwards the constellation volume is reduced - shrinking process - and the parameters of the model are then refined to an arbitrary precision. This method allows naturally the identification of time-variable systems provided that the parameters change infrequently than the convergence rate of the **MMAE**.

This method was applied with good results to the identification and tuning of a simple stochastic integrate and fire neuron model [Martins et al., 2011a]

This technique was first mentioned, despite briefly and for the case of an unidimensional unknown parameter space, in [Martins, 2006]. The identification method and the results obtained and were published in [Martins et al., 2011a,b].

Organized list of publications:

- Review of retina models; test and evaluation of representative models of different classes:
  - Tomás, P., Martins, J. C., and Sousa, L. A. (2008). Towards a unified model for the retina: Static vs dynamic integrate and fire models. In *International Conference on Bio-Inspired Systems and Signal Processing (BIOSIGNALS 2008)*, volume 2, pages 528–533, Funchal, Madeira, Portugal. [Tomás et al., 2008]
  - Martins, S. F., Sousa, L. A., and Martins, J. C. (2007). Additive logistic regression applied to retina modelling. *IEEE International Conference on Image Processing, 2007. ICIP2007.*, 3:309–312. [Martins et al., 2007]
  - Martins, J. C. and Sousa, L. A. (2005). Comparison of computational retina models. In *5th IASTED Conference on Visualization, Imaging, and Image Processing.*, pages 156–161, Benidorm, Spain. [Martins and Sousa, 2005]

## 1. Introduction

- Organization and experimental analysis of neuronal metrics and selection of appropriate parameters values:
  - Martins, J., Tomás, P., and Sousa, L. (2009). Neural code metrics: Analysis and application to the assessment of neural models. *Neurocomputing*, 72(10 – 12):2337 – 2350. [Martins et al., 2009]
- Parameters estimation and system identification with a new method for the adaptive selection of models' parameters:
  - Martins, J. C., Caeiro, J. J., and Sousa, L. A. (2013). Constellation based multiple model adaptive estimators for system identification and parameters tuning. *Digital Signal Processing, Elsevier*. submitted. [Martins et al., 2013]
  - Martins, J. C., Caeiro, J. J., and Sousa, L. A. (2011b). On the use of adaptive model constellations in multiple model adaptive estimators. In *Proc. of RecPad 2011 - 17th Portuguese Conference on Pattern Recognition*, Porto, Portugal. [Martins et al., 2011b]
  - Martins, J. C., Caeiro, J. J., and Sousa, L. A. (2011a). A new approach to system identification and parameter tuning with multiple model adaptive estimators. In *7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011)*, pages 72 –77. [Martins et al., 2011a]
- Many of the background material that supports this work was published as the book:
  - Martins, J. C. and Sousa, L. A. (2009). *Bioelectronic Vision: Retina Models, Evaluation Metrics, and System Design*, volume 3 of *Series on Bioengineering & Biomedical Engineering*. World Scientific, Singapore. [Martins and Sousa, 2009]

### 1.4. Thesis Organization

This document is composed of 7 chapters and several appendices. It starts with a general introduction to the human vision system in Chap. 2, posing a particular emphasis on their anatomical and neurophysiological characteristics relevant for modeling. Chapter 3 presents the common approaches followed in the neuronal model and in the retina modeling, describing examples of various types of models employed to describe the retina. Chapter 4 describes and analyzes the retina models proposed and developed that are representative of the two main classes of neuronal models. Chapter 5 gathers the neuronal metrics employed in neuroscience, addresses their usefulness in evaluating neuronal responses and its applicability in the training and evaluation of neuronal models by proposing methods to wisely pick up its parameters. In Chap. 6 the system identification method proposed in this thesis is detailed with the results applied to the identification of neuron models, both a linear model and

a non-linear model. Finally, this thesis ends with Chap. 7 where the main conclusions are stated and possible future research directions are pointed. This thesis is complemented with several appendices included to increase the document readability and completeness by including additional information about the subject. Appx. A describes the main causes of blindness and its prevalence, Appx. B presents several methods and techniques used to study and characterize the retina and the neuronal response, and finally Appx. C describes the several approaches followed in the development of bioelectronic vision systems by briefly presenting several representative examples of each type of prosthesis.



*A good theoretical model of a complex system should be like a good caricature: it should emphasize those features which are most important and should downplay the inessential details. Now the only snag with this advice is that one does not really know which are the inessential details until one has understood the phenomena under study.*

Frenkel's guidance cited in Fisher, M. E. (1983). Scaling, universality and renormalization group theory. In Hahne, F. J., editor, *Lecture Notes in Physics*, volume 186, pages 1–139. Springer

# 2

## The Human Visual System

### 2.1. Introduction

THE visual system in humans, and in mammals in general, is very complex, so a substantial amount of research effort in the last decades has been directed at understanding the various aspects of the system, ranging from the physical to the biological and psychological processes involved.

After a short description of the eye composition, a somewhat detailed description of the anatomy and physiology of the human retina is made in the sequel. It serves as a basis for understanding the different processes occurring in the retina, to provide a functional evaluation of the retina models, and to give an insight into the challenges encountered in modeling such a complex and intricate network of neurons. It also helps to devise methods and simplifications that can be applied for deriving computational retina models. A summarized description of the visual pathway is given with a presentation of the principal visual processing centers, ending in the visual cortex.

Finally, some relevant issues related to retinal modeling are discussed. A brief overview of the basic processing blocks commonly used in retina modeling, as well as the taxonomy usually employed to classify those models is described. As the retina is a neural circuit, therefore it starts with a general description of the neuron's anatomy and dynamics.

### 2.2. The Neuron

To have a better understanding of the vision system, and of several phenomena related to the retinal neural cells' responses, it is important to have a general overview

## 2. The Human Visual System

of the neuron anatomy and the mechanisms involved in information encoding and communication.

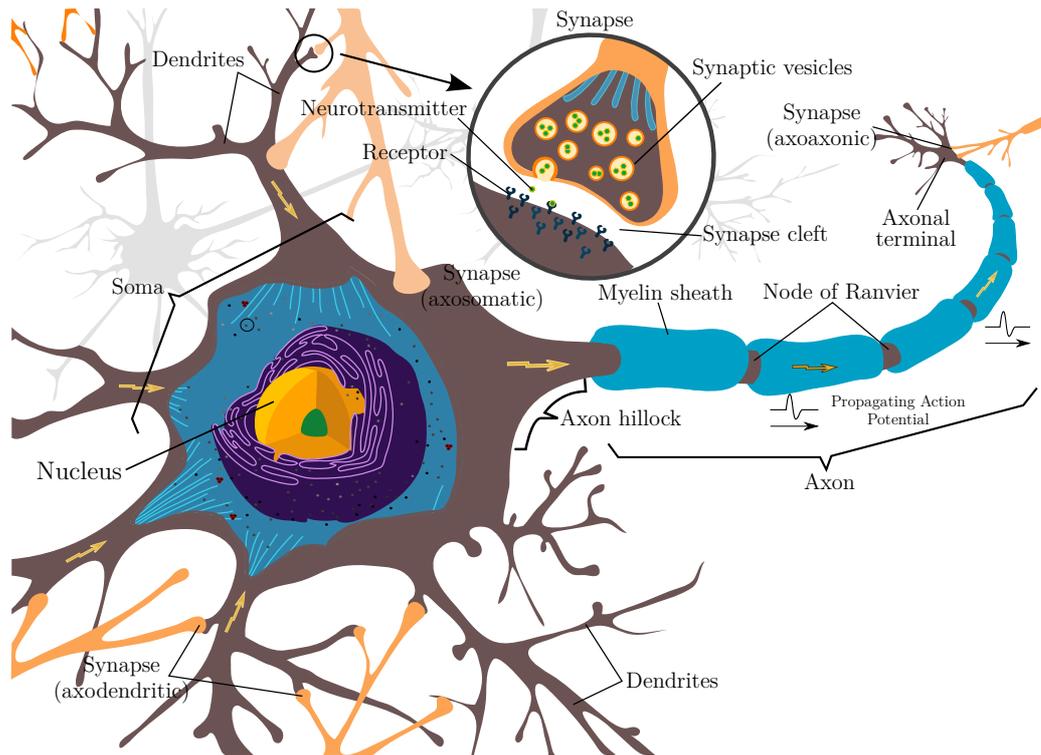
### 2.2.1. Neuron Anatomy

The neuron is the basic unit of information processing and is the building block of neural circuits, such as the retina and brain visual centers. The neuron is made up of three basic components, as shown in Fig. 2.1: the cell body, or *soma*; an extension, called the *axon*; and the *dendrites*. Dendrites look like the branches of a tree; they receive messages from other cells and communicate it to the soma. A single neuron can have more than 2000 dendritic branches, establishing connections between tens of thousands of other cells. Within the cell body (soma) is the nucleus, which contains the genetic material. The main role of the soma is to process all the information collected by the dendrites; if the sum of the electrical signals collected by the dendrites is strong enough, the neuron will fire an action potential. The axon transmits messages from the cell body to other neurons. The axon looks like a long tail, and can extend farther than 1 m (the largest axon in humans runs from the base of the spine to the big toe of the foot), and can be as wide as 1 mm. At its end, the axon divides into fine branches – the *axonal terminals* or *presynaptic terminals* – that make contacts with neighboring neurons. In many neurons, portions of the axon are covered by a myelin sheath. Myelin is a fatty substance whose role is to increase the speed and strength of the signal that travels down the axon, in addition to protecting the axon from external assaults.

The point of contact between two neurons is called the synapse (see detail in Fig. 2.1). A synapse is composed of a narrow space, called the *synaptic cleft*, between the axon ramification ends of the neuron that transmit a signal, the *pre-synaptic neuron*, and the dendrite of another neuron – the *post-synaptic neuron*. Most pre-synaptic terminals end on the dendrites of the post-synaptic neuron (axodendritic synapse), but the terminals can also target the cell body (axosomatic), and less frequently, the beginning or end of the axon of the receiving cell (axoaxonic synapse) (see Fig. 2.1). Synapses can be also found between neurons and other cells, such as muscle cells and gland cells, with appropriate receptors.

There are a number of different types of synapses found in the retina that can be classified into two main categories: chemical synapses and electrical synapses. Synapses which communicate through a *transmitter substance*, or *neurotransmitter*, are termed *chemical synapses*, and those in which two cells are electronically coupled together are termed *electrical synapses*. The specific sites where the neuron membranes connect in the electrical synapse are termed gap junctions.

A single neuron can receive synapses from a large number of other neural cells and its response can vary significantly, depending on which source cell (or set of cells) stimulated it. Moreover, the temporal history of stimulations can change the synaptic strength. This mechanism, typically referred to as *synaptic plasticity*, is an essential process that, in addition to roles in learning, is crucial for the physical building of the brain during its development and throughout an organism's life-



**Figure 2.1.** Anatomy of a neuron.

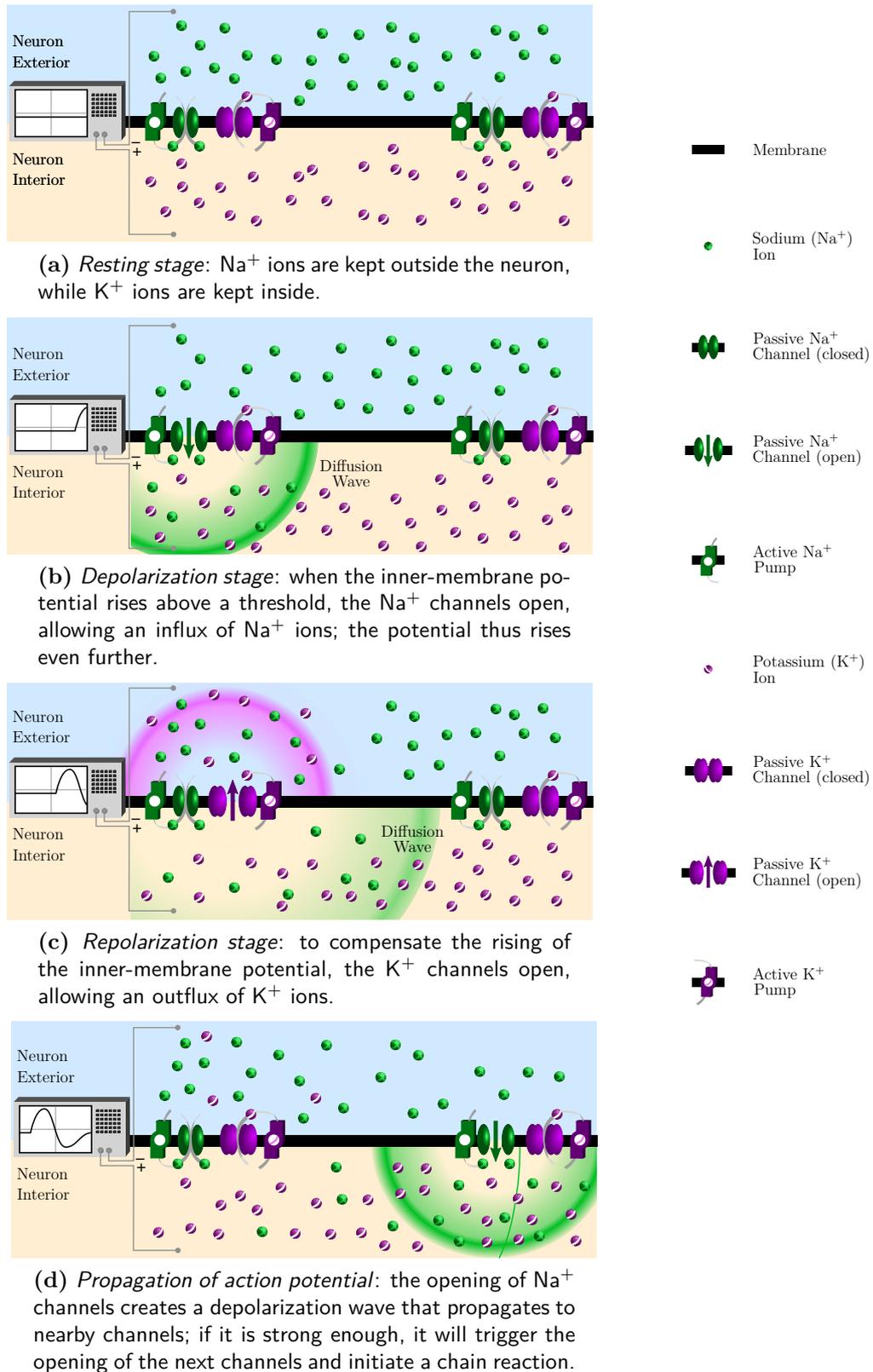
time [Dayan and Abbot, 2001; Gerstner and Kistler, 2002].

Besides the nerve cells, or neurons, the nervous system is also composed of other types of cells, like glial cells that outnumber the neurons. Glial cells are substantially smaller and provide support and protection for neurons. They are active in surrounding neurons and holding them in place; in supplying nutrients and oxygen to neurons; in insulating one neuron from another; and in destroying pathogens and removing dead neurons. Glia also have important developmental roles, such as guiding migration of neurons in early development and producing molecules that modify the growth of axons and dendrites [Purves et al., 2007].

### 2.2.2. Neuron Dynamics

When the integration of the signals gathered by the dendrites and delivered to the cell body surpasses a given threshold, an action potential is generated. An action potential is an electric pulse that is generated at the axon hillock, where the axon emerges from the cell body (see Fig. 2.1). This electrical signal travels along the axon to the axonal terminals and is refreshed along the way to prevent signal decay. This transmission and signal refreshing involves the movement of charged particles – namely ions – across the neuron membrane, and terminates with the release of a chemical substance into the synapse. The synaptically transmitted messages can be either excitatory or inhibitory.

## 2. The Human Visual System



**Figure 2.2.** Propagation of an action potential along a neuron's axon.

There are several types of ions involved in the transmission of the action potential along a neuron's axon. The propagation of electric signals due to excitatory stimulation is achieved mainly by means of an exchange of sodium ( $\text{Na}^+$ ) and potassium ( $\text{K}^+$ ) (positive ions), between the inside and outside of the neuronal membrane. Other ions involved in the process are chloride ( $\text{Cl}^-$ ) and organic ions ( $\text{A}^-$ ), such as amino acids, proteins, and nucleotides, in smaller concentrations. Along the axon, the neuron possess ionic pumps that move the  $\text{Na}^+$  ions to the outside of the cell (see Fig. 2.2a), while the  $\text{K}^+$  ions are moved inside; however, a small outward leak of  $\text{K}^+$  ions, larger than the inward leak of  $\text{Na}^+$  ions, keeps the potential difference between the inside and the outside negative. It is thus said that the membrane is hyperpolarized. When an action potential is delivered by the soma, the  $\text{Na}^+$  channel opens in the axon, as shown in Fig. 2.2b, so that there is a influx of sodium ions into the cell; this raises the potential inside the cell membrane, making it depolarized. To compensate, the nearby  $\text{K}^+$  ion channels open, leading to an efflux of  $\text{K}^+$  ions (Fig. 2.2c), repolarizing the membrane. Eventually, both channels close (first the sodium, then the potassium) and the pumps re-establish normal conditions. All this is achieved in less than 2 ms. Furthermore, as illustrated in Fig. 2.2b, once the  $\text{Na}^+$  channel is opened, diffusion waves of  $\text{Na}^+$  ions are generated; this depolarization wave then propagates to another nearby sodium channel, slightly raising the potential inside the cell. As a result, that nearby  $\text{Na}^+$  channel opens, and the cycle repeats itself. This process therefore triggers a chain of events that leads to the propagation of the electric signal down the axon to the axonal terminals.

The ion channels also have a refractory period, which prevents them from reopening in a short amount of time. This process is very important because it ensures that the cycle does not become unstable. Without this feature, once the channels were activated, they would enter a cycle of opening and closing; moreover, this property guarantees that the action potential always propagates forward.

The speed of action potential propagation is usually directly related to the size of the axon. Big axons generally have fast transmission speeds as increasing the size of the axon allows more of the sodium ions that form the internal depolarization wave to enter and remain inside the axon. To overcome the transmission speed limitations due to the limited size of the axons, myelin (above mentioned) is wrapped around many neurons' axons (see Fig. 2.1). The myelin sheath prevents the dissipation of the depolarization wave by inhibiting ion leakage, thus speeding up the transmission.

While in an excitatory synapse, the opening of  $\text{Na}^+$  channels leads to the excitation of the cell and the initiation of an action potential, the inhibitory synapse works by preventing this initiation. To accomplish this, the pre-synaptic neuron releases a packet of neurotransmitters which activate chloride ( $\text{Cl}^-$ ) channels. Once the channels are opened, the  $\text{Cl}^-$  ions flow into the neuron by diffusion; this lowers the potential inside the neuron and prevents the initiation of the action potential. The behavior of neurons is also influenced by other types of ions, such as calcium ( $\text{Ca}^{2+}$ ) and magnesium ( $\text{Mg}^{2+}$ ) [Gerstner and Kistler, 2002].

### Neural Communication through Chemical Synapses

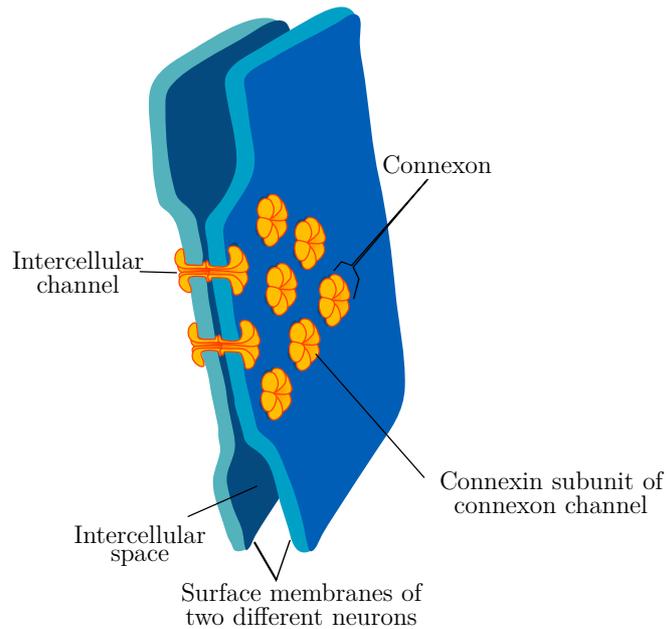
When the action potential reaches a chemical synapse at the axonal terminal, the pre-synaptic neuron communicates the information to the next neuron, the post-synaptic neuron. The process of information communication in the synapse is achieved as follows: *i*) the action potential triggers the opening of calcium channels, allowing for an influx of calcium ( $\text{Ca}^{2+}$ ) ions into the pre-synaptic neuron; *ii*) the excess of  $\text{Ca}^{2+}$  ions causes the pre-synaptic neuron to release a packet of organic molecules, referred to as *chemical messengers* or *neurotransmitters*, into the *synapse cleft*, a small gap between the two neurons. These neurotransmitters are stored in synaptic vesicles (see Fig. 2.1), and serve as the output signal, translating the neuron's electrical signal into a chemical signal; *iii*) when released by the pre-synaptic neuron, the neurotransmitters traverse the synaptic cleft and bind to special proteins in the post-synaptic neuron that produce a local electrical signal called the *synaptic potential*. Unlike the action potential, the synaptic potential is not propagated; instead, it triggers the opening of special ion channels - sodium ( $\text{Na}^+$ ) for excitatory synapses, or chloride ( $\text{Cl}^-$ ) for inhibitory synapses; *iv*) these ions then enter the post-synaptic neuron, generating an electric signal. In an excitatory synapse, this signal propagates to the post-synaptic neuron body, eventually resulting in the initiation of another action potential; in an inhibitory synapse, the  $\text{Cl}^-$  ions work to prevent the generation of action potentials by inhibiting excitatory electronic signals. This synaptic communication can be mediated by different types of neurotransmitters [Gerstner and Kistler, 2002]. In the retina, the neurotransmitter passing through the vertical pathways, which run from photoreceptors to bipolar cells to ganglion cells, is glutamate, while the horizontal and amacrine cells send signals using various excitatory and inhibitory amino acids, including catecholamines, peptides and nitric oxide [Kolb, 2003].

The synaptic potential is not stereotyped, like the action potential; instead, its amplitude depends on the stimulus strength. The potential can be more positive than the neuron resting potential (depolarizing), enhancing the neuron's ability to fire an action potential (rendering them excitatory), or they can be more negative than the resting potential (hyperpolarizing), making the neuron less likely to fire an action potential (rendering them inhibitory).

### Neural Communication through Electrical Synapses

In an electrical synapse the two neurons are electrically coupled at specific sites of the membrane surface – the gap junctions. A gap junction consists of several channels that allow ions, and small molecules, to pass from one cell to the other cell. Each channel consists of two subchannels, termed *connexons*, one from each cell. Each connexon is composed by six protein subunits termed connexin (see Fig. 2.3). Gap junctions can differ in the number of channels and in the types of connexons that compose them, however they always have a depolarizing effect.

Signaling by electrical synapses is very fast because the action potential pass directly from one neuron membrane to the other, through the gap junction, which



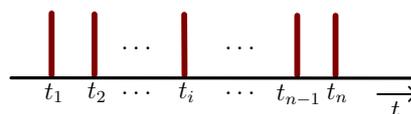
**Figure 2.3.** Gap junction between two neurons.

allows the direct exchange of ions. The electric current is proportional to the potential difference between the presynaptic and the postsynaptic membranes.

### 2.2.3. The Neural Response Function

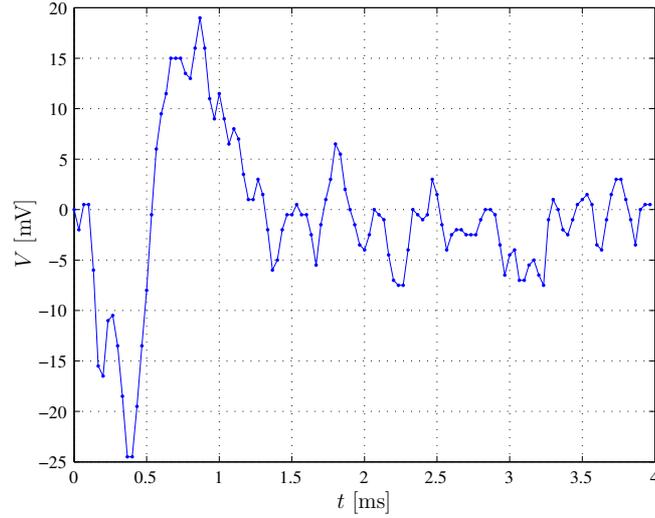
The neuron output is composed by a discrete sequence of voltage pulses, also called evoked potentials, or action potentials, and briefly termed by *spikes*. These are positioned at time instants  $\{t_i\}$ , with  $i = 1, \dots, n$ , like represented in Fig. 2.4, and is called a spike train. The waveform of the evoked potentials has a stereotyped shape for a given class of neural cells, and is used to classify the ganglion cells [Wandell, 1995]. The waveform in Figure 2.5 corresponds to an evoked potential of a rabbit OFF-type retinal ganglion cell (RGC).

Because of the stereotypical form of the spikes, and although their time lengths, amplitudes, and shapes show slight variations, the information carried to the brain is encoded in the spikes' temporal occurrence instants. Due to this fact, a spike train can be represented by a time series of equal amplitude bars, with a bar positioned at every instant where a spike occurs. Figure 2.6 shows a graphical representation of a spike train segment from the response of a rabbit transient brisk OFF-type ganglion cell when excited with a Gaussian random stimulus [Keat et al., 2001].



**Figure 2.4.** Representation of a spike train.

## 2. The Human Visual System



**Figure 2.5.** Spike waveform of a ganglion cell from a rabbit's retina obtained with a sampling frequency  $f_s = 30 \text{ kHz}$ .

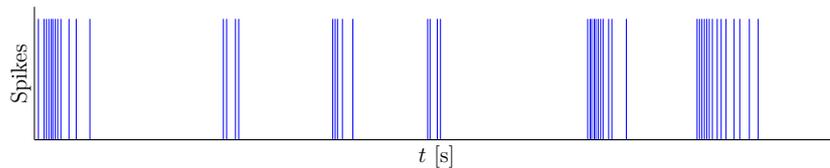
A spike train can be represented mathematically by a series of Dirac delta functions positioned at the time occurrence instants  $t_i$  like:

$$\rho(t) = \sum_{i=1}^n \delta(t - t_i), \quad (2.1)$$

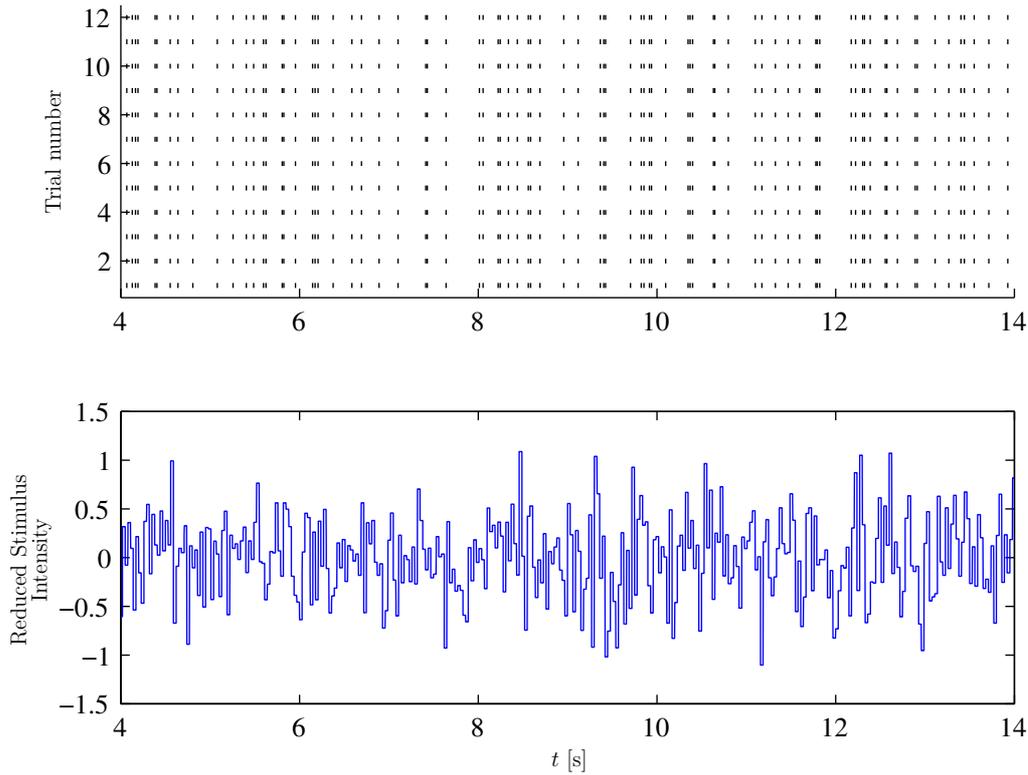
where  $n$  is the total number of spikes in the train,  $\rho(t)$  is the *neural response function*. The neural response function disregards the height and shape of the action potentials, so that all information is contained in the time instants arrivals of the spikes; thus, the spike train is considered to be a point process [Brenner et al., 2002].

Integrating the neural response function in Eq. (2.1) the number of spikes,  $n$ , present in the recording of a neuron response within a total elapsed time  $T$  is obtained by:

$$\begin{aligned} n &= \int_0^T \rho(\tau) d\tau \\ &= \sum_{i=1}^n \int_0^T \delta(\tau - t_i) d\tau, \end{aligned} \quad (2.2)$$



**Figure 2.6.** Neural response function of a retinal ganglion cell.



**Figure 2.7.** Neural spike trains from a Salamander ON-type retinal ganglion cell (*Top*) when driven by the uniform white noise stimulus obtained from sampling a Gaussian distribution (*Bottom*), (data from [Keat et al., 2001]).

where the integral in the last term evaluates to one, as a Dirac delta function.

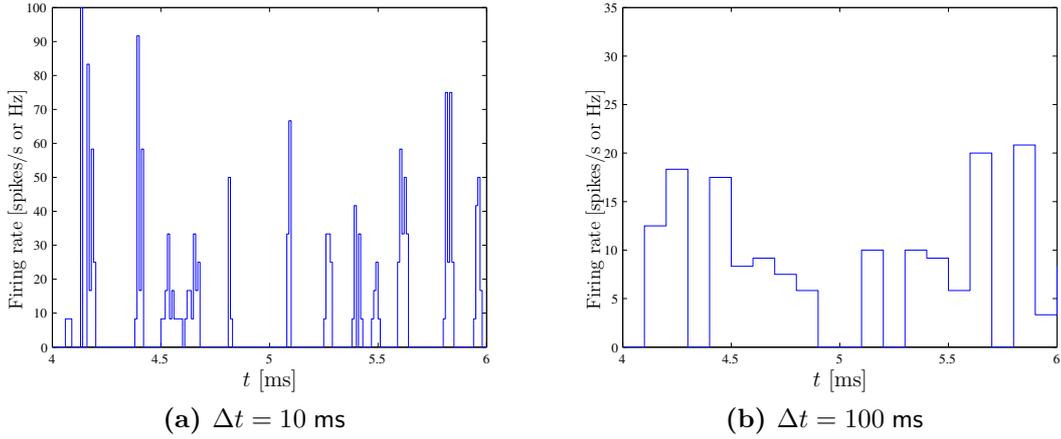
The neural response can be characterized by several other quantities obtained from the neural function such as the *spike-count rate*. The spike-count rate,  $r$ , is the total number of spikes,  $n$ , in a trial, divided by its total time duration,  $T$ :

$$r = \frac{n}{T} = \frac{1}{T} \int_0^T \rho(\tau) d\tau . \quad (2.3)$$

The spike-count rate is the time average of the neural response over a particular trial, and does not contain any temporal information about the neural response.

The average neural response function over many experimental trials for the same stimulus, like the trials displayed in Fig. 2.7, is called the *average neural response*. This average, represented by  $\langle \rho(t) \rangle$ , is computed by summing all the individual neural response functions for each trial and dividing the result by the number of trials  $M$ . If the neural response function for the trial  $j$ , where  $j = 1, \dots, M$ , is represented by

## 2. The Human Visual System



**Figure 2.8.** Retinal ganglion cell ON-type firing rate using different time bin widths for the neuron response.

$\rho_j(t)$ , the average neural response is:

$$\langle \rho(t) \rangle = \frac{1}{M} \sum_{j=1}^M \rho_j(t) . \quad (2.4)$$

If the time instant of a spike occurrence at time  $t_i$ , with  $0 \leq t_i \leq T$ , in the trial  $j$ , for  $j = 1, \dots, M$ , is represented by  $t_{ij}$  then the neural response average can be written as:

$$\langle \rho(t) \rangle = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{n_j} \delta(t - t_{ij}) , \quad (2.5)$$

where  $n_j$  represents the total number of spikes in the trial  $j$ , which is usually different from trial to trial.

A *time-dependent firing rate* can be obtained by counting spikes over short intervals of time from a large number of trials where the neuron is repeatedly stimulated with the same stimulus. The firing rate at time  $t$  can be computed by counting the number of spikes that occur within the time interval  $[t, t + \Delta t]$  and by dividing the result by  $\Delta t$ ; thus, the firing rate is measured in spikes/s or Hz. The precision of the firing rate increases by using a small time bin width  $\Delta t$ ; as a result, a higher temporal resolution can be achieved. If only a single spike train is used with a narrow time bin  $\Delta t$ , the firing rate will be either zero or one, and therefore an average over multiple trials should be made. The time-dependent firing rate is defined as the average number of spikes over several trials appearing in an interval between the time instants  $t$  and  $t + \Delta t$ , divided by the interval length  $\Delta t$ . Figure 2.8 displays the firing rates calculated from trials of a retinal type ON-type ganglion cell with different time bins.

The time-dependent firing rate is defined as:

$$\begin{aligned} r(t) &= \frac{1}{M} \sum_{j=1}^M \frac{1}{\Delta t} \int_t^{t+\Delta t} \rho_j(\tau) d\tau \\ &= \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle \rho(\tau) \rangle d\tau, \end{aligned} \tag{2.6}$$

where the integral in the first equality computes the number of spikes between the times  $t$  and  $t + \Delta t$  in the  $j$ th trial, while the integral in the second equality captures the number of spikes in the time range between  $t$  to  $t + \Delta t$  of the neural response average.

Formally, the time interval length of Eq. (2.6) can be made to go to zero. Thus, by taking the limit we have

$$\begin{aligned} r(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle \rho(\tau) \rangle d\tau \\ &= \left\langle \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \rho(\tau) d\tau \right\rangle. \end{aligned} \tag{2.7}$$

Inside the angle brackets in the last expression is the derivative of the integral of  $\rho(t)$ . Therefore, the formal definition of the firing rate can be written as:

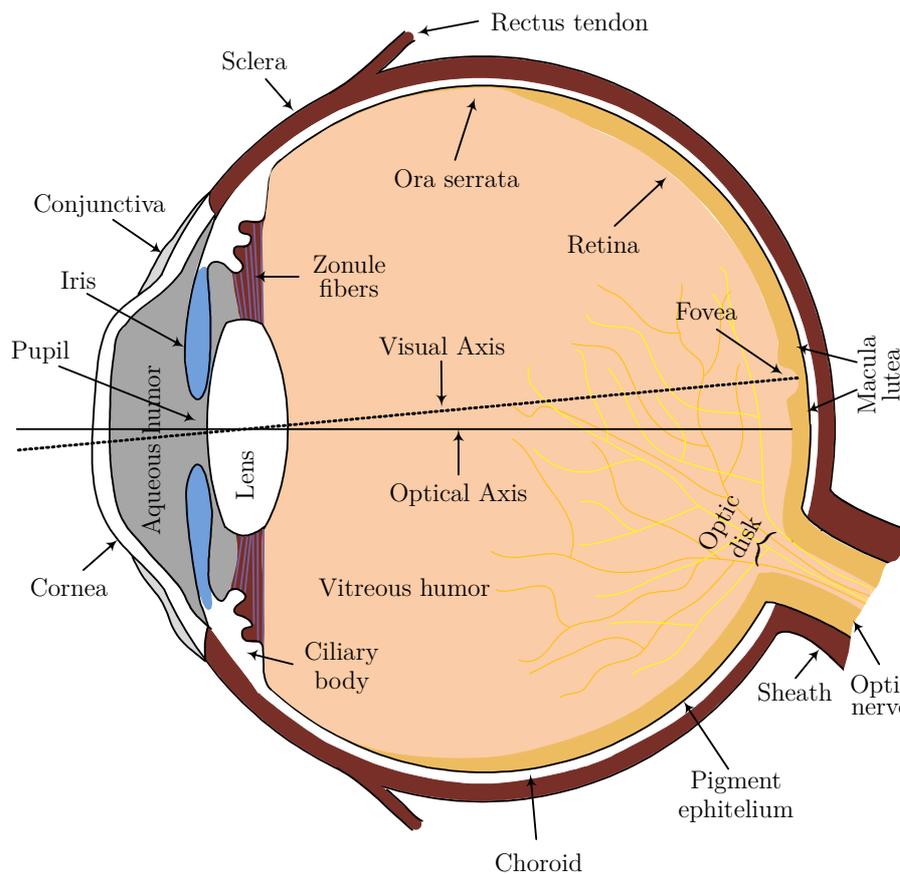
$$r(t) = \langle \rho(t) \rangle, \tag{2.8}$$

that states that the firing rate is the average of the neural function over many trials. Several other quantities can be computed from the neural response. In Appendix B various other quantities related to the neural response are presented and the way of calculating them in practice is discussed.

## 2.3. The Human Visual System

The ancient Greeks thought that it was the light rays emitted by the eyes that touch the objects, making them visible. This idea was related to the fact that, despite that objects are far away, they can still be sensed. It was only in the seventeenth century (1625) that the German Jesuit priest and physicist Christoph Scheiner (1573 – 1650) showed that it is the light entering the eye that produces the image. Since then, remarkable advances have been made in the knowledge and understanding of this marvelous sense, although there are still many unanswered questions. The sections that follow resumes what is currently known about the eye anatomy and physiology [Wandell, 1995; Rodieck, 1998].

## 2. The Human Visual System



**Figure 2.9.** Schematic section of the human eye.

### 2.3.1. The Eye

From a functional point of view, the eye is an optical system that gathers light and focuses it on its rear surface where the retina lies. Fig. 2.9 shows a schematic representation of the horizontal cross section of the human eye with the different parts labeled.

Looking into someone's eyes, it is possible to immediately identify several components. One is the pupil, an aperture in the center of the eye that appears dark due to the light absorbing pigments of the retina on the back. Around the pupil is the iris, a colored muscle that confers the color to the eyes, but whose color does not have any functional relevance. The iris muscle controls the size of the pupil according to the light conditions. The pupil diameter can range from 1.5 mm to 8 mm, becoming smaller in brighter light conditions.

Covering the external surface of the eye is the *cornea*. The cornea is a tough transparent membrane that acts as the first lens encountered by the light entering the eye. It has a round surface, with a refraction index of 1.38, and acts as a convex lens that is primarily responsible for bending light into the eye. Around the iris is an external white surface, called the *sclera*, that makes part of the wall that supports

the eyeball and is continuous with the *dura mater* – the tough membrane that covers the central nervous system.

Partially covered, and behind the iris, is the lens, a transparent tissue consisting of many fibers, which are also called *crystalline*. The crystalline lens has a bi-convex shape with a refraction index of 1.4, which is higher than any other eye component. However, because it is surrounded by media with a similar refraction index, light bends less as it passes through the lens than it does when passing through the cornea. It is the lens, in conjunction with the cornea, that allows the formation and focus of a sharp image on the back of the eye. The lens is connected to the *ciliary body* by several ligaments, called *zonule fibers*, depicted in Fig. 2.9. The contraction or relaxation of the zonule fibers, through the action of the ciliary body muscle, changes the shape of the lens and mediates the constant focusing of the image on the retina. This process is called accommodation and constitutes the zonule fibers' most important function.

In the eyeball, there are three chambers of fluid. Between the cornea and the iris lies the *anterior chamber*, and between the iris and the zonule fibers lies the *posterior chamber*. These two chambers are filled with the *aqueous humour*. A third chamber, called the *vitreous chamber*, is located behind the lens and is filled with the *vitreous humour* that occupies the entire space between the lens and the retina, representing two-thirds of the eye's volume. The vitreous humour is a viscous fluid whose refraction index is identical to the eye's optical system, so that it does not bend light. The vitreous chamber is also responsible for the round shape of the eye.

Behind the vitreous humour is the retina, which covers 65% of the inside eyeball and is one of the most important components of the eye. This is where the image is focused and transduced to neural signals, to be posteriorly conducted to the brain by the optic nerve. The previously presented components of the eye have the unique function of focusing the image on the retina, playing a role similar to that of an optical system, while neural processing actually begins at the retina. Figure 2.9 displays the path followed by a light ray entering the eye's optics and hitting the retina, showing that the visual axis differs from the optical axis. The visual axis ends in a special point in the retina, termed fovea, that is the point of highest visual acuity and is always directed at the object that one is paying attention to at the moment.

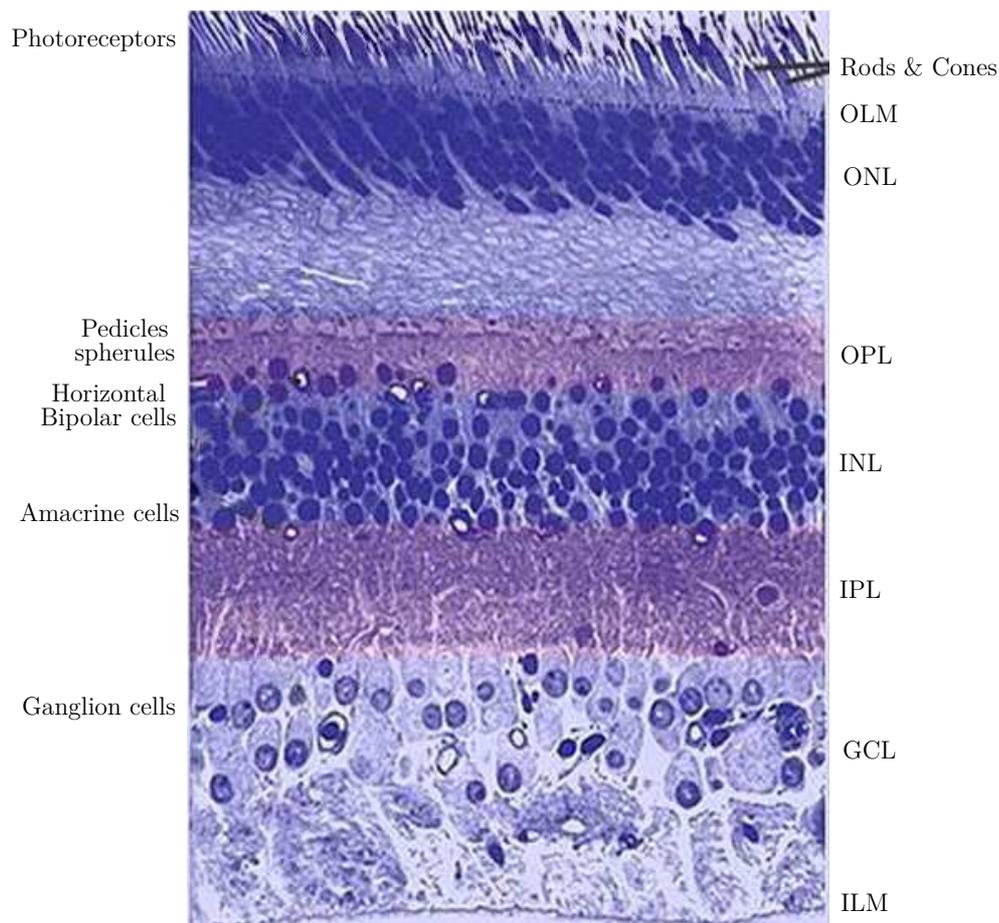
The eyeballs are held in their respective optical cavity by various ligaments, muscles and fascial expansions. Figure 2.9 shows the rectus tendon which is connected to one of the two pairs of muscles running to the skull called the rectus muscles. An additional pair of muscles, called oblique muscles, are responsible for rotating the eyeball in the orbit.

### 2.3.2. The Retina

The retina can be seen as an extension of the brain, it is responsible for transducing light into electrical nervous pulses, and for the early processing stages of neural visual signals. To understand the retina functioning it is essential first to know its anatomy [Dowling, 1987; Kolb, 2003].

The retina is a circular disc with a diameter of approximately 4.2 mm and a

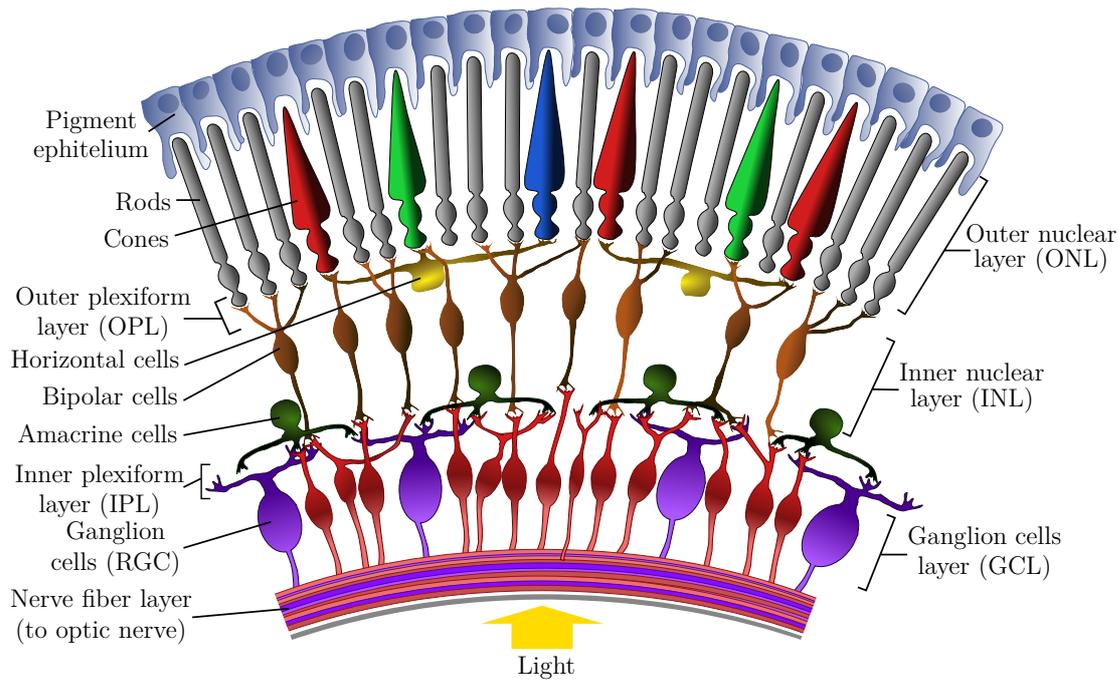
## 2. The Human Visual System



**Figure 2.10.** Light micrograph of a vertical section through the retina [Kolb et al., 2012].

thickness of 0.5 mm. It is composed of several layers of neurons that can be easily distinguished in a micrograph, like the one reproduced in Fig. 2.10. The darker layers, called the *nuclear* layers, contain the neuron cell bodies, while the pale layers, called the *plexiform* layers, contain their dendrites and axons. Figure 2.11 portrays a simplified schematic of the organization human retina, with the neurons composing each layer labeled.

The human retina has two types of photoreceptors: rods and cones. The rods are long and thin, and are about 120 million in number (about 94.5% of all photoreceptors). The rods are very sensitive to light (capable of perceiving even a single photon), and enable *scotopic vision* – the visual response at lower orders of illuminance magnitude. The cones are less numerous than the rods (totaling about 7 million), have a shorter and thicker shape, and are less sensitive to light. The cones provide the eye's *photopic vision* – the visual response at 5 to 6 orders of illuminance magnitude – and are responsible for color perception. There are three different types of cones in the human retina, the blue, green and red cones, corresponding to the visible light wavelength to which they are most sensitive. Figure 2.12 shows the spectral



**Figure 2.11.** Simplified schematic organization of the retina.

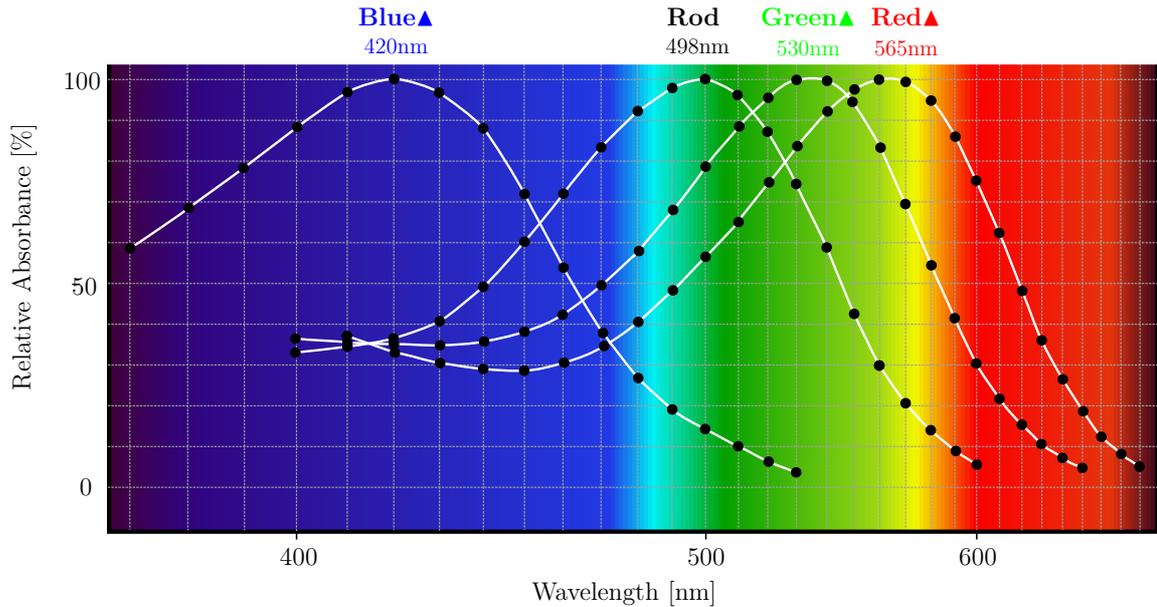
responses of the different types of photoreceptors present in the human retina. At intermediate levels of illuminance, both rods and cones are active, enabling *mesopic vision*.

The photoreceptors are not distributed uniformly in the retina. While the cones are almost exclusively concentrated in the fovea, where there are no rods, the rest of the retina is populated predominantly by rods. Figure 2.13 displays a graphic with the photoreceptor distribution throughout the retina. Foveal cones are densely packed hexagonally, as is shown in the photograph of the cross section of the human fovea in Fig. 2.14a. As distance from the fovea increases, the cones become larger and are packed less densely, like the photograph of the foveal periphery in Fig. 2.14b shows; the spaces between cones are filled by rod photoreceptors.

The fovea appears as a small dimple aligned with the visual axis (see Fig. 2.9), defining the operational center of the retina in bright light. Since in dim light an object focused on the fovea is not visible, at night we have to look to objects slightly sideways and is difficult to perceive color. The circular area around the fovea, with a diameter of approximately 6 mm, is the central retina; this region extends to the peripheral retina and further extends to the *ora serrata*, 21 mm from the center of the optic disc (see Fig. 2.9).

The light has to traverse all retinal layers to be sensed by the photoreceptors, as shown in Fig. 2.11, because photoreceptors must be in contact with the pigment epithelium (see Fig. 2.9). The exception to this rule occurs in the fovea, where nerves are pushed away so that the cones are directly exposed to light. Nevertheless, the photoreceptors are always in direct contact with the pigment epithelium because the

## 2. The Human Visual System

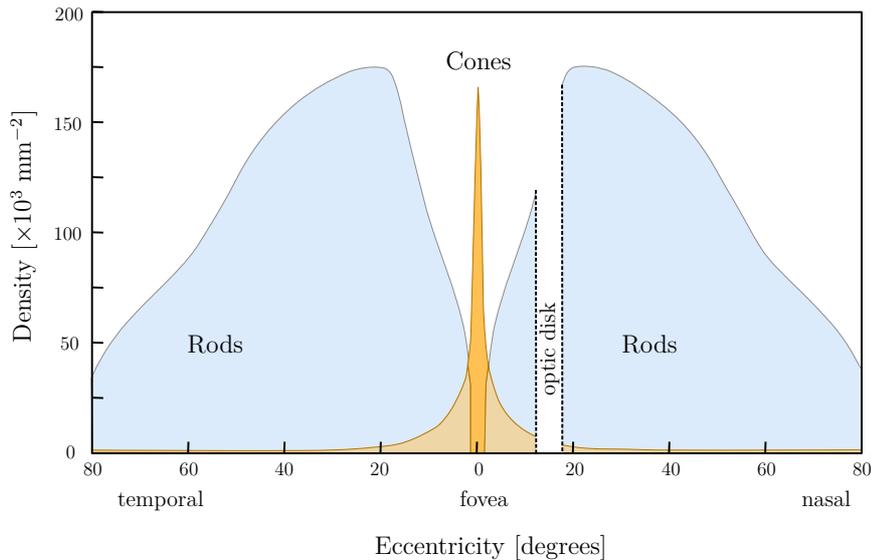


**Figure 2.12.** Spectral sensitivity of the photoreceptors.

pigment-bearing membranes of the photoreceptors have to be in contact with the epithelial layer, which provides a steady stream of several retinal molecules used to transduce light, like vitamin A and its aldehydes. These molecules sense light by changing their conformation in response to photons, and are then recycled back in the pigment epithelium [Rodieck, 1998].

The bottom layer of the retina in Fig. 2.10 is where the retina ganglion cells lie, and it is called the ganglion cell layer (**GCL**). The axons coming from the **RGCs** converge to one point where they leave the eye together and run to the brain, forming the optic nerve that also contains the blood vessels necessary for vascularization of the retina. This point is also called the blind spot, or optic disk, since there are no light receptive cells in this region (see Fig. 2.9). The existence of the blind spot is not perceived, because the blind spot of one eye is compensated by the other eye, and vice-versa.

The intermediate cell layers, namely the bipolar, horizontal and amacrine cells, are responsible for processing the electrical stimuli coming from the different photoreceptors and enhancing the features relevant for the brain to extract information. These cells are organized in layers, as shown in the schematic of Fig. 2.11 and in Fig. 2.10, where the layers are labeled with the corresponding cell types. The photoreceptor cells' bodies are located in the outer nuclear layer (**ONL**). The inner nuclear layer (**INL**) contains the cell bodies of the bipolar, horizontal and amacrine cells. The outer plexiform layer (**OPL**) is located between these two neuron layers and contains the synaptic connections between the photoreceptors, bipolar, and horizontal cells. The **GCL** contains the ganglion cells, whose axons link to the brain. Between this layer and the inner nuclear layer is the inner plexiform layer (**IPL**), containing sev-



**Figure 2.13.** Spatial distribution of photoreceptors along the retina.

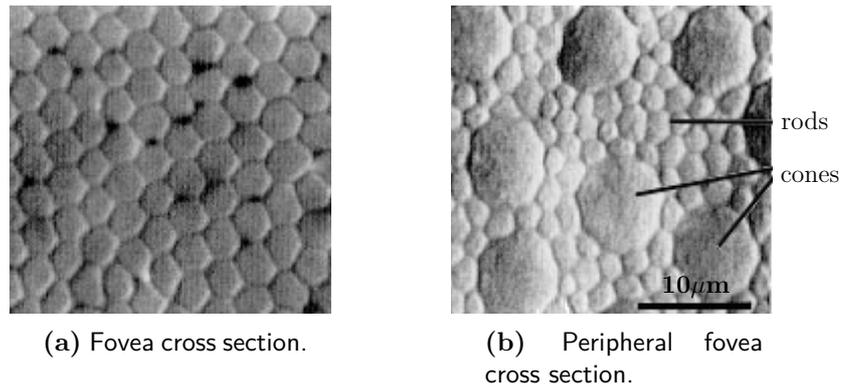
eral kinds of synaptic contacts between the ganglion cells and the horizontally and vertically directed amacrine cells. Despite its complexity, we can obtain a general overview of the structure of the human retina by considering a simplified organization such as the one depicted in Fig. 2.11, which is a very basic view of the structure of the retina. In reality, the connections between the different types of cells are very intricate, as can be previewed from Fig. 2.10.

The central retina is thicker than the peripheral retina for two main reasons: the high density of photoreceptors present, mainly cones, and a greater density of cones connecting second order neurons located in the inner nuclear layer, called *cone bipolar cells*. There is also a greater number of ganglion cells in the ganglion cell layer of the central retina, which implies a greater number of synaptic interactions.

At the middle point of the central retina is the fovea, whose constitution is significantly different from the central and peripheral retina. In the foveal pit, the neuron layers are radially displaced and cones are compactly aligned in a hexagonal structure and exposed directly to light (see Fig. 2.14a). Around the foveal pit is the foveal rim, or *parfovea*. It is the thickest zone of the retina due to its six layers of ganglion cells that connect the central cones to the optic nerve. The foveal area, including the rim around the parfovea, termed the *perifovea*, constitutes the *macula lutea*, which can be distinguished from the rest of the retina due to its yellow pigmentation (see Fig. 2.9). Figure 2.15 shows a photograph of the human retina taken with an ophthalmoscope, in which the fovea and the optic nerve are pointed out, and in which the blood vessels that vascularize the eye are visible.

Axons from both types of photoreceptors end in the **OPL**, where they establish visual pathways with neurons of the subsequent layers. To establish synaptic contacts with second order neurons, the cones end in a terminal called the *pedicle*, and the rods end in a terminal called the *spherule*, where the connections are established.

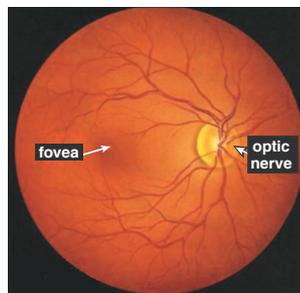
## 2. The Human Visual System



**Figure 2.14.** Micrograph of the cross section of the fovea and of the foveal periphery: (a) cones densely packed in fovea, and (b) fovea periphery with bigger cone photoreceptors interleaved with rods [Curcio et al., 1990].

The cone pedicle ends in approximately 30 extensions associated with 30 triads of neurons. Each triad is composed of a central element, which is a dendritic terminal of a bipolar cell, and two lateral elements which are dendritic terminals of horizontal cells. There are also what are called *basal junctions*, consisting of different varieties of bipolar cell dendrites that form synaptic contacts under the surface of the cone pedicle. The rod spherule has two extensions; each one is associated with four, second order neurons. This group of four neurons consists of two horizontal cell axon terminals and two central rod bipolar cell dendrites.

In addition to the pathways between photoreceptors and second order neurons, there are also contacts between photoreceptors. These contacts are established between cones and cones and between cones and rods. The cone pedicles have small extensions from their bases that establish contacts with neighboring pedicles and spherules. These contacts are called *telodendria*. A rod can have 3 to 5 telodendrial contacts from neighboring cones, and cones can have as many as 10 telodendrial contacts to their neighboring rods. The exceptions are the cones that are sensitive to blue light, also called *S-cones*, which do not have many telodendria and have a second order specific kind of bipolar cell. These direct interactions between photoreceptors appear to degrade spatial resolution and color perception by mixing signals coming



**Figure 2.15.** Photograph of a human retina from an ophthalmoscope [Kolb et al., 2012].

from different photoreceptors at different locations. One justification for telodendrial contacts is to allow rods to use neural pathways, devoted to both rods and cones, that are more rapid than the specific rod pathways.

### 2.3.3. Retina Operation

Having a grasp of the composition and organization of the retina we will focus now on its operation [Kolb, 2003; Kolb et al., 2012]. As mentioned earlier, the light must traverse all retinal layers before reaching the photoreceptors that lay at the back of the retina. The photoreceptors absorb photons that, by means of a biochemical reaction, are converted to an electrical signal capable of stimulating the forward neuronal layers of the retina. Electrical stimuli arrive at the ganglion cell layer and are sent to the brain through the optic nerve as a sequence of stereotyped voltage pulses, called *action potentials*, evoked potentials, *spikes*. (Figure 2.5 displays a typical spike waveform from a rabbit RGC).

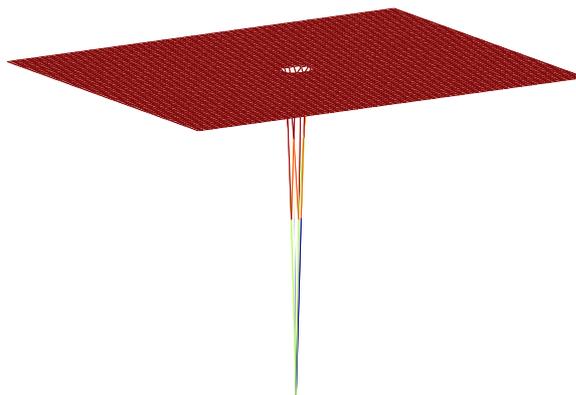
The optic nerve contains about  $10^6$  optic nerve fibers, corresponding to extensions of the ganglion cell axons. Three different types of horizontal cells, 11 types of bipolar cells, at least 25 different types of amacrine cells have been morphologically distinguished and 18 types of morphologically different ganglion cells in the human retina have been identified [Kolb et al., 2012].

When excited by light, the photoreceptors send a neurotransmitter through the vertical pathways of the retina. The horizontal and amacrine cells introduce excitatory or inhibitory signals in the retina's neural network, depending on their nature. A central concept in neuroscience is the notion of receptive field (RF). In the case of the retina, the RF can be characterized by the area on which light influences the neural response. Rods and cones react to light directly falling over them, so they have narrow receptive fields. Figure 2.16 represents a photoreceptor's receptive field. The rods can detect dim light and respond to relatively slow changes in luminance, while cones deal with bright light signals and can detect rapid light fluctuations. The process of image decomposition begins at the photoreceptor layer and continues in the first synapses of the visual pathway, existent between the photoreceptors and the bipolar cells.

Glutamate is a neurotransmitter that enables the transmission of the electrical neural signals from the photoreceptors through the bipolar cells to the ganglion cells. This neurotransmitter allows the establishment of electrical conductive channels between the axon of the presynaptic neuron and the dendrite of the postsynaptic neuron. Depending on the type of ion flow, the neurotransmitter effect can be excitatory or inhibitory through depolarization or hyperpolarization of the postsynaptic neuron, respectively. The horizontal and amacrine cells can send excitatory or inhibitory signals using several types of substances [Kolb et al., 2012].

While in the dark the rods and cones constantly release neurotransmitter and cease its releasing when excited by light. For example, when a green cone is in the dark, its internal potential is at rest, but when is illuminated by green light it becomes hyperpolarized – the electrical potential of its membrane gets more negative, stopping

## 2. The Human Visual System



**Figure 2.16.** Graphical representation of the narrow receptive field of a cone.

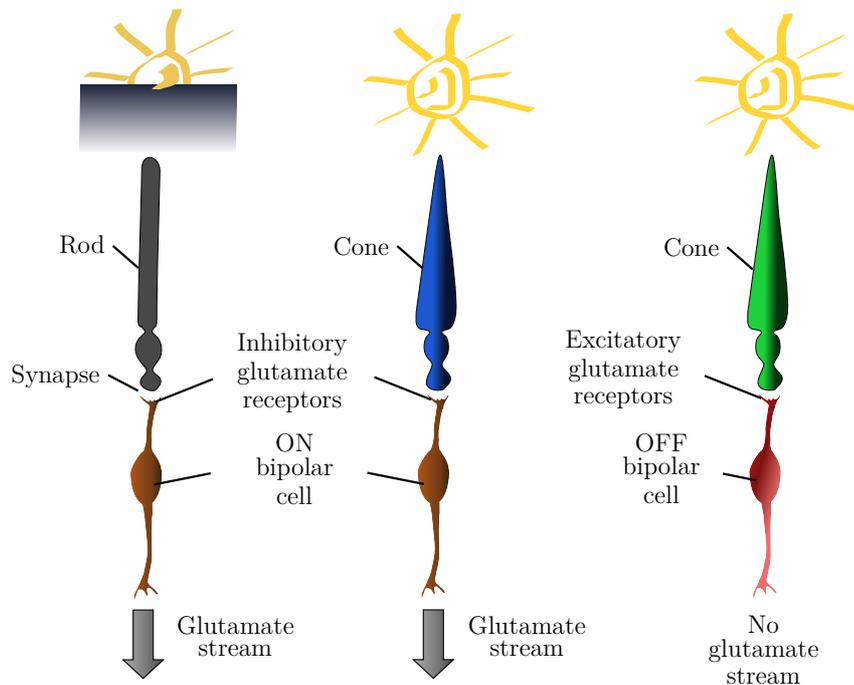
the neurotransmitter release over the duration of the light flash.

The distinct types of bipolar cells react differently to the neurotransmitter. Some bipolar cells can re-sensitize their glutamate receptors quickly and react to rapid changes in the visual signal by firing at a relatively high frequency, while others take more time to re-sensitize their glutamate receptors, so that they respond relatively slowly to the same amount of stimulus. On the other hand, some bipolar cells' receptors, which possess a hyperpolarizing receptive field, respond to glutamate by activating an OFF pathway, so that they detect dark images against a lighter background. Other bipolar cells possess an inhibitory glutamate receptor that prevents the cell from firing when it receives the neurotransmitter; thus, the glutamate activates an ON pathway, and the cell detects a light image against a darker background and possesses a depolarizing receptive field. Figure 2.17 illustrates the different types of connections between photoreceptors and bipolar cells and the streams of information controlled by the glutamate molecule. A single bipolar cell receives input from a small number of cones and has a medium sized receptive field.

The parallel sets of visual channels of ON-type (detecting light areas on dark backgrounds) and of OFF-type (detecting dark areas on light backgrounds) are fundamental to sight, as vision depends on perceiving the contrast between an object and its background. Connections between ON-type bipolar cells and ON-type ganglion cells and between OFF-type bipolar cells and OFF-type ganglion cells occur in specific regions of the inner plexiform layer.

If the images were transmitted to the brain via the bipolar and ganglion cells alone, they would be grainy and blurry. The role of horizontal cells is to define the edges and enable the perception of fine details in an image. Each horizontal cell receives its input from several cones, and so its receptive field is large. The receptive field becomes even broader because the plasma membranes fuse with those of neighboring horizontal cells at gap junctions (see Fig. 2.18).

A single bipolar cell, with its ON or OFF light response, would carry a fairly blurry response to its ganglion cell. Horizontal cells add an opponent signal that is spatially



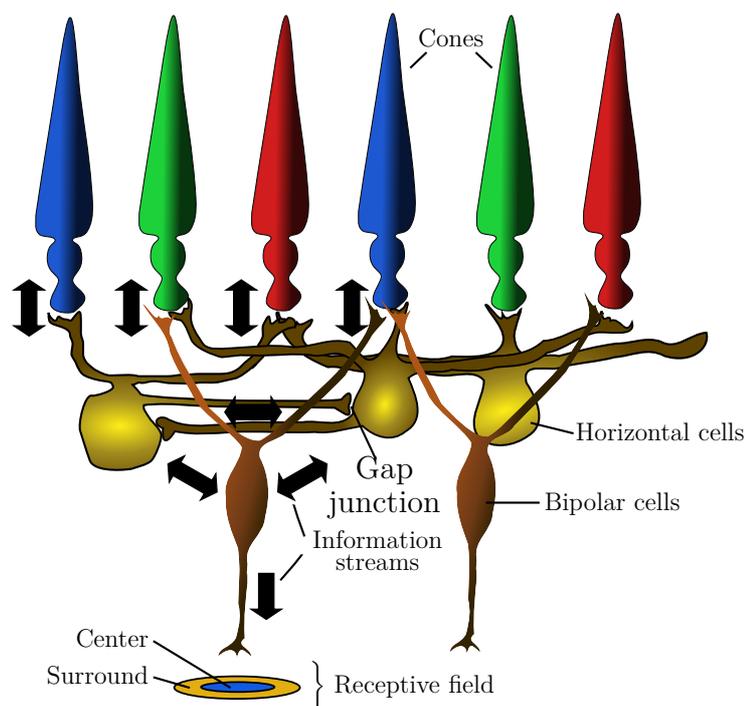
**Figure 2.17.** Photoreceptor to bipolar cell connections. The ON-type bipolar cells (a,b); OFF-type bipolar cell (c).

constructive, giving the bipolar cell what is known as a *center surround organization* (see Fig. 2.18). The bipolar center signals can be either ON or OFF, and the horizontal cells add an OFF or ON surround signal through two different mechanisms - either directly or by sending feedback information to the cone photoreceptors, which then feed forward again to the adjacent connected bipolar cells. Figure 2.18 illustrates the information streams performed by the horizontal cells in conjunction with the bipolar cells.

Horizontal cells also receive feedback signals, from the inner plexiform layer, that influence its activity. The result is that horizontal cells modulate the photoreceptors' signals under different lighting conditions, allowing signaling to become less sensitive in bright light and more sensitive in dim light, as well as shaping the receptive field of the bipolar cells. Horizontal cells also make the bipolar cells' response color-coded through feedback circuits to cones.

Ganglion cells have a receptive field that is also organized in concentric circles. The amacrine-cell circuitry in the inner plexiform layer conveys additional information to the ganglion cells, thus sharpening the boundary between the center and surround areas in their receptive field even further than the horizontal cell input alone. There are two main types of ganglion cells in the human retina with ON centers and OFF centers that form the major output from the retina to the subsequent visual centers in the brain. The ON-center ganglion cells become active when a spot of light falls in the center of their receptive field and are inactivated when light falls on the field periphery. The OFF-center ganglion cells act in the opposite way; their activity

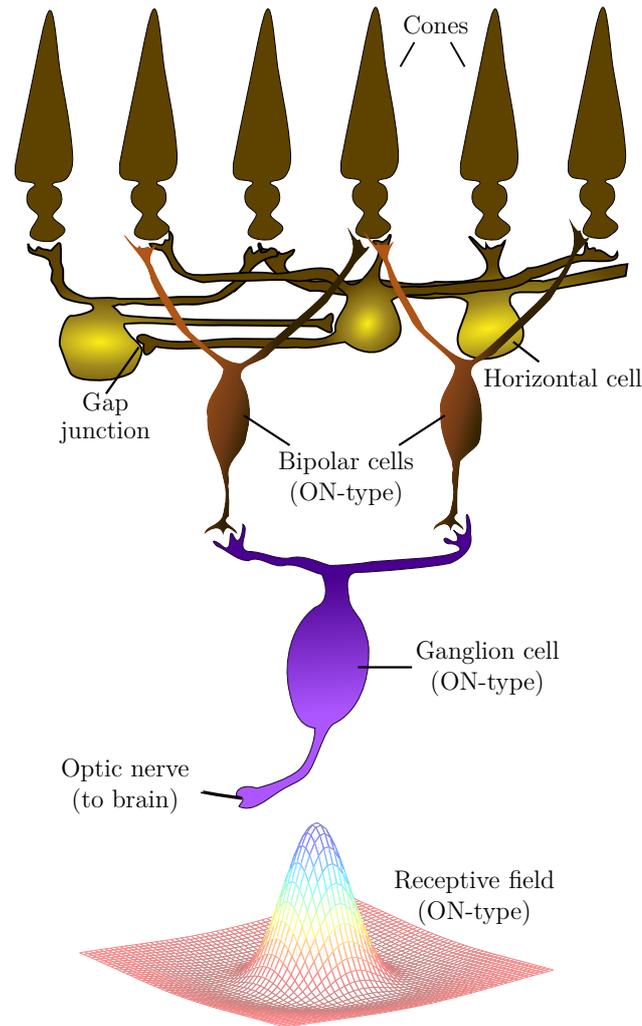
## 2. The Human Visual System



**Figure 2.18.** Connections between photoreceptors, horizontal cells and bipolar cells, with the information streams indicated.

increases when the periphery of their receptive field is lit and decreases when light falls on the center of the field. The horizontal cells convey antagonistic surround signals to bipolar cells, and consequently to the ganglion cells. This kind of processing sharpens the boundaries of the images. The receptive fields of the ganglion cells can be modeled as the difference between two Gaussian functions, giving them a so called Mexican-hat shape. Figure 2.19 presents a sketch of the connections of an ON-type ganglion cell with its receptive field modeled as a difference between two Gaussian functions, represented at the bottom of the figure. An OFF-type ganglion cell is connected to OFF-type bipolar cells, and its receptive field would have a symmetric shape relative to the one in Fig. 2.19.

In contrast to the rest of the retina, the organization of the retinal cells in the fovea region contains midget ganglion cells, which possess tiny dendritic trees that are connected in a one-to-one ratio with midget bipolar cells. The channel from midget bipolar to midget ganglion cells carries the information from a single cone, thus relaying a point-to-point image from the fovea to the brain. Each red or green cone in the central fovea connects to two midget ganglion cells, so at any time each cone can either transmit a dark-on-light (OFF) signal or a light-on-dark (ON) message. The message that is sent to the brain carries both spatial and spectral information at the highest spatial resolution. As a result of being connected to only one cone, the receptive fields of the midget ganglion cells are much narrower than their counterparts on the rest of the retina. Blue cones are organized in a different way: they transmit information through a specific blue bipolar cell to a different type of ganglion cell,

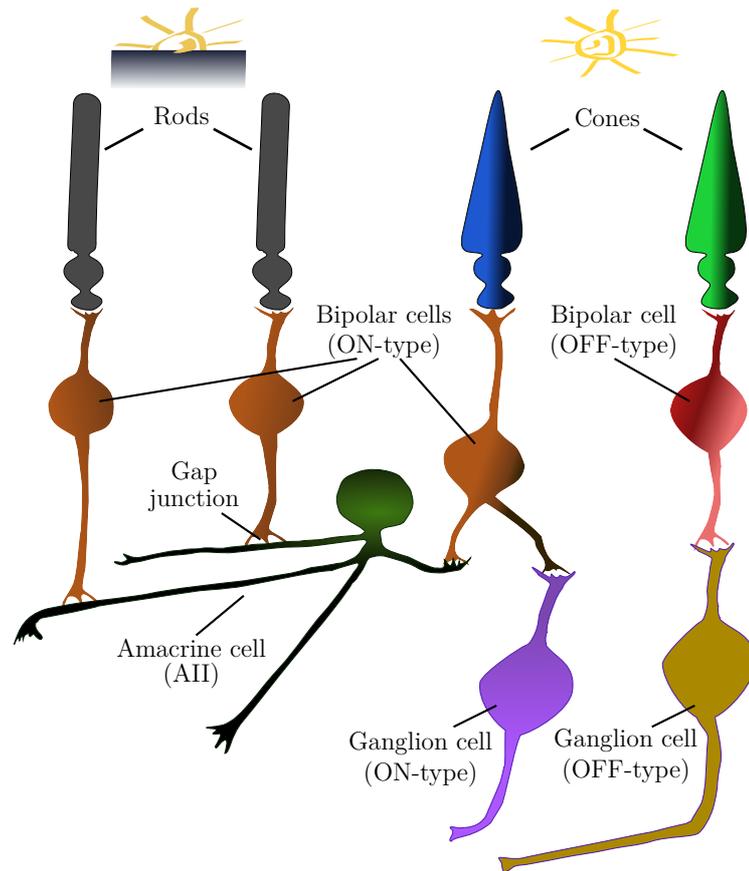


**Figure 2.19.** Connections between bipolar cells and ganglion cells.

which can carry both a blue ON and a yellow OFF response.

At the bottom of the inner plexiform layer, just above the ganglion cell layer, are the amacrine cells. Among the several types of amacrine cells, there is a special type denominated AII. These cells receive information from the bipolar cells and transmit information to ganglion cells and bipolar and amacrine cells. They also provide interconnections between ON and OFF systems of bipolar and ganglion cells. Another important type of amacrine cell is designated A17. These amacrine cells are crucial in the pathways starting in rods. Whereas the cones connect directly to bipolar cells, and these to ganglion cells, the bipolar cells that receive their input from rods do not synapse directly to the ganglion cells. All bipolar cells connecting to rods are of the ON-type (see Figure 2.17), and use the AII and A17 amacrine cells to send signals to the ganglion cells. A single AII amacrine cell can be connected to as many as 30 rod bipolar cells and can transmit a depolarization signal both to ON cone bipolar cells and their ON ganglion cells, and to OFF cone bipolar cells and

## 2. The Human Visual System



**Figure 2.20.** All amacrine cell connections.

their OFF ganglion cells. Therefore, AII amacrine cells make it possible for rods to use the faster cone pathways. Figure 2.20 illustrates the connections made by an AII amacrine cell. The A17 amacrine cells also collect the signals from thousands of rod bipolar cells that are modulated, amplified, and transmitted to AII amacrine cells. Signal modulation and amplification allow the perception of very weak light signals, and consequently enable night vision.

There are several other types of amacrine cells that spread horizontally, interacting with hundreds of bipolar cells and many ganglion cells. They can even connect to neighboring amacrine cells through gap junctions, increasing their action radius and the speed with which signals can be transmitted across large areas of the retina. Another role played by the amacrine cells is in modulating the retinal response for different illuminance conditions. By liberating several kinds of neurotransmitters, they inhibit or reinforce the synaptic connections between the neuronal layers in the retina.

All these singularities of light processing in the retina suggest that a significant part of the construction of the visual images occurs in the retina. A quantitative analysis of the responses of each retina cell type permits the evaluation of each cell type's contribution to the ganglion cell response. The analysis performed in [Meister and Berry II,

1999] revealed that the photoreceptors, horizontal cells, and bipolar cells produce responses to light that are basically linear. On the contrary, under the same light stimulus conditions, the amacrine cells showed strong nonlinear distortions, which made it possible to distinguish between *sustained* and *transient* amacrine cells [Victor, 1999]. This kind of experiment allows the functional differences between classes of cells in the retina to be distinguished and the development of models in a cascading manner, giving biological meaning to each block in the signal processing pathway.

### 2.3.4. The Visual Pathway

In addition to the retina, the spatiotemporal processing of the illuminance pattern gathered by the eye continues to take place all along the visual pathway to the cerebral cortex. Figure 2.21 shows a schematic view of the visual pathways, from the retina to the cerebral cortex. A visual prosthesis can potentially be interfaced with the nervous system in different places along the visual pathway, with inherent advantages and disadvantages [Warren and Normann, 2003].

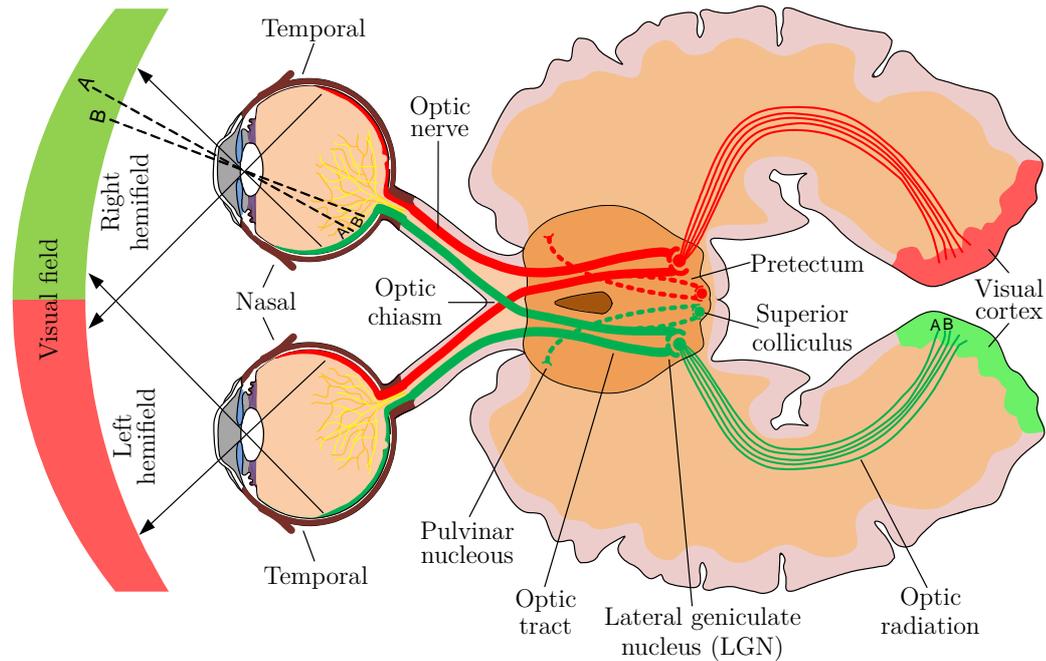
All the neurons along the visual pathway possess a receptive field. The receptive field construct can be expanded to include other characteristics of the visual stimuli, such as the shape, size, intensity, color, and location in the visual space, that drive the neuron to respond optimally. Another important issue is the mapping of the visual space to the neural space. This mapping is visuotopic, meaning that the neurons along the visual pathway are arranged such that their receptive fields form an organized and approximately linear map of the visual space. As a consequence, objects that are close together in the visual space evoke neural activity in nearby neurons in the brain. This implies that a rectangle in the visual space will result in neural activity in a similarly shaped arrangement of neurons in the visual centers, although this arrangement may be stretched along each axis, rotated, and/or warped. The external electrical stimulation of this ensemble of neurons would result in the perception of the outline of a rectangle. Figure 2.21 illustrates the visuotopic organization of the visual pathway, with points A and B in the retina mapping to points A and B in the visual cortex.

The visual pathway is a highly parallel signal processing system. This parallelization occurs along two principal parallel pathways: the M pathway (M for *magno*, or large) and the P pathway (P for *parvo*, or small). These distinct parallel pathways begin in the retina and extend through the visual pathway. The M and P pathways represent different features of an object placed on the visual space, such as where the object is located and what object it is, respectively [Warren and Normann, 2003].

### Optic Nerve and Tract

The optic nerve contains the nerves running from the retina to the optic chiasm – the location where the nerves coming from the left and right eyes intersect (see Fig. 2.21). The nerves that go from the optic chiasm to the subcortical tracts compose the optic tract. The optic nerve has a length of approximately 50 mm and the optic tract is

## 2. The Human Visual System



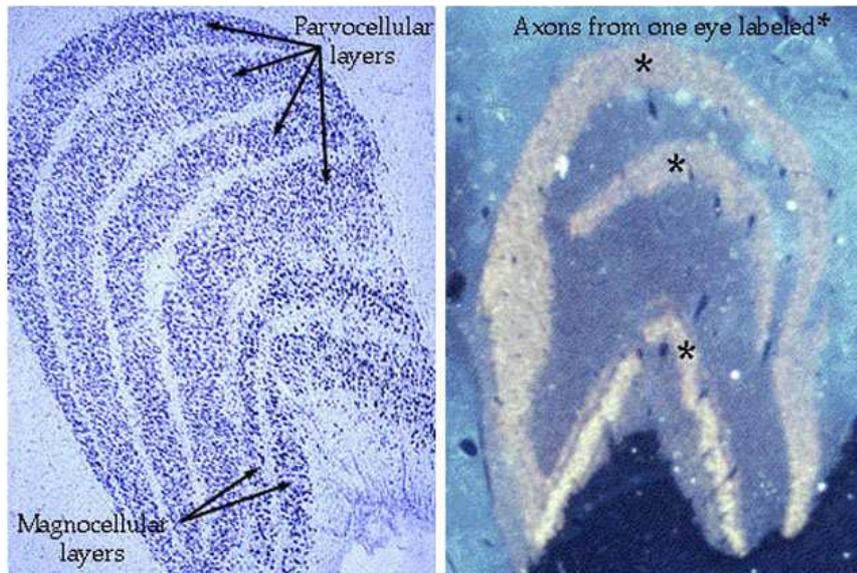
**Figure 2.21.** Schematic cross section of the human brain with the visual pathways depicted.

about 30 mm in length. It is uniquely composed of the axons of  $1.2 \times 10^6$  retinal ganglion cells, and do not include neuron cell bodies. The optic nerve contains the axons from the nasal visual field and from the temporal to the fovea visual field of a single eye. In the optic chiasm, the optic nerve axons are reorganized so that the optic tract almost exclusively contains the axons representing the contralateral visual hemifield. Besides the retinal ganglion cell axons, the optic nerve also contains an artery and vein, which irrigate the retina. Figure 2.21 shows a schematic cross section of the human brain and illustrates in a more detailed manner the different visual fields and pathways.

The fibers of the optic nerve are also visuotopically organized, with the upper retina represented along its dorsal side, the central retina along the lateral side, and nasal visual field along the medial side. However, this visuotopic organization changes along the nerve [Warren and Normann, 2003]. Because the optic nerve and the optic tract are extensions of the retinal ganglion cell axons, their receptive fields have the same structure.

### Subcortical structures

The axons coming from the retinal ganglion cells target three subcortical structures: the *superior colliculus*, the *pretectum*, and the **LGN** (see Fig. 2.21). The superior colliculus and the pretectum are localized on the top of the midbrain, and are associated with the saccadic eye movements and pupillary reflexes, respectively. Neither



**Figure 2.22.** LGN cell layers [Kolb et al., 2012].

of these structures is suited for a visual prosthesis implant or bioelectronic vision system.

The **LGN** is located on the ventral side of the *thalamus*, and its neurons are considered relay neurons: they receive the input from the retinal ganglion cells and pass it to the cortex. The volume of the **LGN** is small,  $7 \times 7 \times 2$  mm, with six functionally independent cells *laminae*, stacked in a form of a distorted 'U' (see Fig. 2.22). Each lamina receives input from only one P or M pathway and from only one of the contralateral or ipsilateral eyes.

Figure 2.22 displays a microscopic cross section of the **LGN**, in which the projections of the small P cells and large M cells from the two eyes to *parvocellular* and *magnocellular* layers of the **LGN** are indicated. Each eye projects its **RGC** axons to alternating layers, as indicated on the right side of the picture shown in Fig. 2.22.

The **LGN** laminae are visuotopically organized, and the visuotopic maps are registered between lamina. Half of the neurons of the **LGN**, representing half of its area, have receptive fields in the fovea and surrounding region [Warren and Normann, 2003]. Despite its difficult access, the neural organization of the **LGN** make it eligible for a neuroprosthesis implant and there is some experiments to demonstrate its feasibility [Pezaris and Reid, 2007].

## Visual Cortex

The visual cortex is the final stage in the visual pathway. Even in the visual cortex, the M and P pathways continue to follow distinct processing paths. The majority of the **LGN** axons run to a cortex region, designated as the primary visual cortex because it is the first visual processing region at the cortical level, or cortex visual area 1 (**V1**), with dimensions within the range of  $2500 - 3200$  mm<sup>2</sup>. This brain region

## 2. The Human Visual System

is also known as the striate cortex. Visual processing takes place in subsequent areas of the cortex referred to as 2, 3, 4 and 5 (V2, V3, V4 and V5), reflecting their hierarchical relationships.

All these areas have a laminar structure containing six layers with a total thickness of approximately 2 mm from the *pia mater* (the more external region of the brain) to the white matter (the internal region of the brain). Processing of the visual information is performed in an upright manner, with information passing between neurons in a column from pia mater to white matter, and in an horizontal fashion, in which information is integrated across a number of columns. Layer 4 in area **V1** is further subdivided into four layers: 4A, 4B, 4C $\alpha$  and 4C $\beta$ . Layers 4C $\alpha$  and 4C $\beta$  receive input from the M and P pathways, respectively, and layer 6 in **V1** sends feedback to LGN [Grill-Spector and Malach, 2004].

Every area in the visual cortex is visuotopically organized. The visuotopic maps are registered between lamina so that all neurons in a column, from pia mater to white matter, have receptive fields in the same region of visual space. Half of the neurons in the visual area, representing half of the region size, have receptive fields in the fovea or surround region. The visual space corresponding to the foveal region is represented at the posterior part of **V1**.

The receptive fields in the visual cortex, with the exception of layer 4C of **V1**, which has receptive fields identical to the LGN, are more complex than those of previous centers. They still have subregions of preferred ON-type illuminance, lacking OFF-type illuminance, but their shape is no longer annular. They are somewhat elongated, giving preference to visual stimulus with a bar form oriented in a given direction. The visuotopic organization of the retinal input, preserved through all stages along the visual pathway, becomes less distinct in the cortical areas. A neuron can be driven by the contralateral eye, by the ipsilateral eye, or by both eyes. Nevertheless, nearby neurons tend to prefer the same orientation and receive information from the same eye. After the **V1** area, the organization of the visual pathway becomes more complex and the optimal visual stimulus becomes less evident.

Only a small fraction of all the stages that exist along the visual pathway are suitable for the implantation of a neural interface to the nervous system. The main issues hindering implantation are related with retinotopic organization and with surgical access for interface implantation. The place for implanting the prosthesis is also related to the damage present in prior stages of the visual pathway. The area **V1** is the most adequate for implanting a visual prosthesis at the visual cortex level [Warren and Normann, 2003].

## 2.4. Conclusions

This chapter provided a general overview of the human visual system with the goal to provide a perspective of the issues and challenges involved in the development of bioelectronic prostheses. The human visual system is very complex, involving different areas of the nervous system, and comprising several distinct processes. Two main

references related to the subject treated in this chapter are [Kolb, 2003; Kolb et al., 2012]. In particular [Kolb et al., 2012] is a major source of information about the human vision system, with the additional value that it is periodically updated and further complemented with comprehensive material from different researchers. It contains up-to-date information about the human visual system, that is only superficially touched on this introductory overview. The review in [Grill-Spector and Malach, 2004] contains relevant information about the organization of the human visual system in general, and about the visual cortex in particular, addressing the retinotopic organization of the visual system. The understanding of retinal vision processes is far from complete. Every day brings to light additional findings about the vision sense, where the division between the physiological and psychological phenomena is sometimes hard to establish [Werblin and Roska, 2007].

Despite this thesis deals mainly with retina models, there are several studies about the hypothesis to interface visual prosthesis in different structures along the visual pathway. However, besides prosthesis interface at the retinal and optic nerve level, the most promissory high level visual structure for an interface is at the brain cortex level.



*Given how much is now known, it might be fair to ask, are we finished with the retina, or are there more surprises on the horizon? ... Given such unexpected findings, it appears that there may still be much more to learn about how the retina works.*

Kolb, H. (2003). How the retina works. *Scientific American*, 91:28–35

# 3

## Neuron and Retina Models

### 3.1. Introduction

A major goal in neuroscience is the modeling of neural processing centers. The modeling of the neural processing centers is a major goal in neuroscience. Additional knowledge about how the brain and the central nervous system work is provided thus improving the prediction about the behavior of the nervous system. In the context of the retina, these models are essential for the design of signal processing blocks necessary for the development of bioelectronic vision systems. In the context of other neural processing centers the models are also required for the development of prosthesis to overcome diverse types of impairments or even to extend the human natural senses.

In the previous chapter a description of the main biological processes occurring in the retina that go from the transduction of light into the evoked potentials or spikes, that are sent to the brain, and from which all visual information is extracted was provided. This chapter is dedicated to outline several representative spiking neural mechanisms used in different signal processing modules employed in retina models, and to present some state of the art retina models belonging to different classes, both in its continuous and discrete-time forms, providing the necessary steps for its computational implementation targeting a bioelectronic vision device.

### 3.2. The Retina Neural Code

A central question in computational neuroscience is how the evoked potentials encode information, and the dual problem, given a spike train, how the stimulus that

### 3. Neuron and Retina Models

originated it can be reconstructed. These two processes are termed coding and decoding, respectively, and are intimately related [Rieke et al., 1997]. In this context, the question is how the visual image is encoded into the spike trains that are sent to the brain, in a parallel fashion, through the optic nerve?

The establishment of a neural model for the retina involves the implicit answers to the following questions: "What code is used by the retina to codify visual information?" and secondly "How is that code generated?". The answer to the former question would be like having a dictionary for this neural language, where we search the translation for a given visual stimuli. The answer to the latter question can be seen as the construction of the dictionary for this particular neural language - the retina neural code.

The brain does not interpret directly the light intensity pattern falling on the retina that composes the image, but instead, it extracts information from the spike pattern carried by the optic nerve. The brain starts by processing these incoming spike sequences, and it sends information to motor neurons in the form of another sequence of spikes. The language of the brain is composed of sequences of spikes: the brain listens to spikes, uses spikes in its internal processing, and communicates with the external world using spikes [Rieke et al., 1997].

The neural code is normally understood by neuroscientists from two perspectives: with the rate code perspective, one assumes that the brain encodes information only in the mean firing rate of neurons [Bialek et al., 1991; Eggermont, 1998]; whereas the time code perspective considers that the individual spike occurrence times play a significant role in the encoding of information [Berry II and Meister, 1998; Reinagel and Reid, 2002]. The rate-code perspective is supported by the fact that the neurons responses can be so variable that the brain can only extract information from the mean firing rate. An example of such behavior is the excitation of the retina by moving bars: while the firing rate of ganglion cell's responses increases considerably as the stimuli enters its receptive field, the trial-to-trial variability of spiking instants can be quite large [Passaglia et al., 2001]. Additionally, even in the absence of stimuli, the retinal cells still show spontaneous activity. The time-code perspective states that the output from single neurons is reproducible from trial to trial [Reinagel and Reid, 2002]. Thus, the precise timing of spikes is important in the coding process [Reinagel, 2001; Uzzell and Chichilnisky, 2004]. A key factor in these issues is the temporal precision of the individual action potentials, which carries important information.

#### 3.2.1. Macroscopic Retinal Response Characteristics

Several interesting approaches have been proposed to modulate various features of the retinal processing, particularly at the macroscopic level. Some relevant examples are the general Weber-Fechner law, [Wandell, 1995], which interprets the light adaptation that occurs in the retina, or the edge detector proposed by Marr, known as a Laplacian of Gaussian (LoG) [Marr and Hildreth, 1980; Lim, 1990], inspired in the ability of the retina to devise contours in images. While these models provide

some insight into the processes occurring in the visual system, they are not able to mimic the fine details of retinal processing and do not provide a sequence of evoked potentials, as desired.

The conversion of light into a spike pattern occurring in the retina, has some distinctive characteristics that deserve to be pointed out. We will focus on the neural code employed by the ganglion cells of the retina for conveying visual information to the brain.

A remarkable characteristic of the retina is the amount of information compression that it performs since the light is collected by 120 millions photoreceptors whereas the optic nerve has only about 1.25 millions fibers, corresponding to the retinal ganglion cells' axons, a relation of about 100:1. Therefore, on average, there are 100 photoreceptors for 1 ganglion cell, although this relation varies systematically along the retina. Another important feature is that the response of the retina depends on the temporal characteristics of the stimulus. For stationary stimuli the response vanishes, as well as above a certain frequency. Only within a given bandwidth does the retina fire spikes, behaving as a temporal bandpass filter. Moreover, the retina is not sensitive to the absolute illuminance, but instead is sensitive to its variation: the contrast.

Regarding the spatial composition of the stimuli, the retina output is directly related to the notion of receptive field (**RF**) – its response is maximal if the stimulus resembles the spatial form of the **RF**. This raised the existence of different types of retinal cells with ON or OFF centers, and OFF or ON surrounds, respectively, a retinal process known as lateral inhibition [Meister and Berry II, 1999].

Another important issue related to modeling is the linearity of the retinal response. The retina shows a linear correspondence between the input stimulus and the firing rate for only some ganglion cells and under restricted conditions – the range of variation of the light intensity must be small relative to its mean and cannot change too much over time. To model the nonlinearities of the retina response, a processing block denominated by contrast gain control (**CGC**) is frequently used [Meister and Berry II, 1999].

The retinal response changes depending on the illumination conditions. A consequence of this phenomenon, known as light adaptation, is that the capacity for distinguishing two different levels of illuminance depends of the mean illuminance level. The perceived contrast  $\Delta C$  depends on the absolute luminance  $I$ , following the Weber-Fechner law [Wandell, 1995] which can be expressed as:

$$\Delta C \propto \frac{\Delta I}{I}, \quad (3.1)$$

where  $\Delta I$  is the variation in the luminance intensity. This is an important characteristic, since the illumination of the natural world changes several orders of magnitude along the day. Some ganglion cells transmit information about the absolute light illumination level to control the pupil and eye aperture, and the visual response is somewhat immune to the mean intensity level, coding only the contrast changes.

### 3. Neuron and Retina Models

The spatial resolution is also affected by the light intensity conditions. In dim light conditions, the time response slows down and the ganglion cell integrates the incoming illuminance over a longer period of time. The receptive field gradually loses its antagonistic characteristics, reducing the surrounding area. This adaptation permits vision in dim light conditions, but at the cost of reduced temporal and spatial resolution. This adjustment of the stimulus-response relationship, dependent of the mean light level, takes place at different locations along the retina using distinct mechanisms [Meister and Berry II, 1999].

The retina not only adapts to the mean light intensity level, but also to its variance, or contrast. This much more finely tuned retinal response, denominated by contrast adaptation, represents a slow adjustment of the retinal code to the changes of the statistics of the visual image. The time course of contrast adaptation is different for a contrast increase compared to a contrast decrease, and it is supposed that contrast adaptation is not performed by individual photoreceptors, but that it is instead a collective process.

#### 3.2.2. Time Code or Rate Code

An old discussion, related not only to the retina neural code, but to the nervous system in general, relates to whether the neural code is a rate code or a time code [Eggermont, 1998; Nirenberg and Latham, 2003]. The spike trains are so variable for different trials with the same stimulus, that some scientists claim that the brain retrieves the information about the stimulus from a spike train through the neuron firing rate  $r(t)$ . This firing rate is obtained by averaging the responses from the fires of many identical ganglion cells [Meister and Berry II, 1999], and composes a rate code. On the other extreme are the scientists who advocate that the precise time occurrence of a spike, like the time interval between spikes convey relevant information about a given stimulus; all the time characteristics associated with the neural function  $\rho(t)$  are important [Victor, 1999] and compose the time code<sup>1</sup>. This point of view is reinforced by the fact that analyzing a trial response to a given stimulus involves discrete events in time, and not a continuous firing, and that each event is well described by the time of the first spike and the total number of spikes in the event [Berry II and Meister, 1998; Uzzell and Chichilnisky, 2004]. A somewhat intermediate view does not consider isolated spikes as the neural code nor the firing rate, but bunches of spikes, called events, that convey the information, reflected in the spike events metric [Keat et al., 2001], discussed in Sec. 5.5.

Another question is whether the neural code of the retina can be viewed as an individual process, where each individual neuron cell acts independently, or as a population process. Putting it another way, whether the firing of a neuron is dependent on the response of its neighboring neurons, which would imply that correlations between firing patterns generated by different neurons convey information [Eggermont, 1998].

---

<sup>1</sup>The firing rate  $r(t)$  and neural function  $\rho(t)$  are two distinct ways of characterizing a neuron response, and are defined in Appendix B.

For the case of retinal ganglion cells, several studies suggest that the neural coding at the retinal level is essentially an individual process, in which the population coding is responsible for carrying only a minor part of the information transmitted, and when the correlation between spike trains is not taken into account, more than 90% of the information about the stimulus can be retrieved [Nirenberg et al., 2001].

### 3.2.3. Classification of Retina Models

As presented in the previous chapter, the retina is organized in several layers by discrete processing units, such as photoreceptors, bipolar and horizontal cells, amacrine cells, and ganglion cells. Despite this fact, many retina models are developed using a continuous formalism, bypassing processes such as the discrete-space sampling done by the photoreceptors mosaic, but providing a mathematical description in a closed form.

Looking for a taxonomy to classify retina models, one suggestion is to separate the models into two groups: the functional models and the structural models [Wulf, 2001]. A finer possible classification is to order the neural models in terms of how detailed the neuron structure is modeled [Herz et al., 2006]. At the top of the scale are the detailed compartmental models, that describe with detail each dendrite taking into account its spatial structure, and each neuron interaction, it can be considered a very detailed structural model. At the bottom of the scale are the black box models. This type of models ignore the neuron's biophysical mechanisms, and only map the input/output relation, being equivalent to functional models.

In general, neuron models can also be classified as rate-code or time-code models, depending on whether the output is the firing rate or the discrete sequence of firing events, where the time occurrence of each spike is taken into account. In the case of the retina, the output of a rate-code model is just the mean firing rate of the RGC given the light stimulus, while in a time-code model its output corresponds to the spike train relative to the encoding of the gathered image. The time-code models are also commonly referred also as spiking neuron models, and are typically based on a spiking mechanism like integrate-and-fire (IF) model, while a rate-code model needs an additional firing mechanism that receives the firing rate and generates the spikes, like a Poisson spike generator.

A functional model attempts to mimic the functions of the retina as a black box, mapping its input into its output. These models describe the spatiotemporal receptive fields of the ganglion cell by a set of equations. Typically, they involve one input layer, where light enters, and one output layer, that furnishes the model output; often the input and output layers are seen as the photoreceptor and ganglion cell layers, respectively, with the other interneurons connections not taken explicitly into account [Wulf, 2001]. Functional models can be composed of different blocks, performing different types of processing. A functional model can have only a spatial block, where the temporal processing of the retina is set aside, or can be complemented by a temporal block that models the temporal processing of the retina. It can also be composed of only a temporal block, where the spatial processing is set aside.

### 3. Neuron and Retina Models

A functional model can also be composed of a single spatiotemporal block, where the separation between the spatial and temporal processing cannot be made. In this class of models the retina is seen as a black box, and the model tries to resemble its functioning, disregarding its internal structure, and only the information from the input stimuli and output response is used to obtain the model.

Within the class of functional models several subclasses can be distinguished. An evident division is between deterministic and stochastic models. Deterministic models, normally of the rate-code type, always produce the same output rate-code for a given input stimulus, where the variability is introduced only at the level of the spike generation. On the other side, stochastic models try to mimic directly the intrinsic variability of the retinal response by introducing noise sources into the model's internal structure, and may directly produce the spike trains. Amid these, there is the important class of white noise models. White noise models are obtained by applying white noise analysis techniques to the retina's responses.

In the structural models, the cell functional structures are modeled layer by layer, normally by a set of differential equations [Hennig and Funke, 2001]. Another equivalent approach is to model the neuron in terms of electric components that simulate the ions' channels and ions' currents flowing through the cell membrane [Wulf, 2001]. This class includes models of only a given retinal cell type, and models of an interconnected subset of retinal cells, resembling the anatomical structure of the retina. In a structural retina model each retina cell's type is modeled individually according to its layer, and then, by cascading these individual structures, a complete retina model is obtained. Frequently, the model for an individual cell is based on the Hodgkin-Huxley model of the neuron [Dayan and Abbot, 2001].

Many retina models are a mixing of these two structural and functional approaches where some retina structures are modeled individually and then connected to other blocks of the black box type to provide a complete retina model. An example of this approach is the pervasive use of the integrate-and-fire spike generation mechanism, that is based on the neuron dynamics, in many retina models.

There are other characteristics that can be used to classify the models, such as: the type of mathematical description used for the model (differential equations, discrete equations), relation grade with the functional anatomy of the retina, and extent of reproduction of experimental data (e.g. experimental data that cannot be reproduced) [Wulf, 2001].

#### 3.2.4. Retina Model's Assumptions

Generally, neural models rely on the assumption that neurons generate spikes independently of each other. In the case of retina ganglion cells, experimental evidence shows that they fire synchronously, even in darkness or under constant illumination [Meister et al., 1995; Warland et al., 1997]. However, comparing the stimulus reconstruction considering independent encoding of the stimulus by the retinal ganglion cells with the reconstruction taking the correlations into account give equally good results [Nirenberg and Latham, 1998]. Another study showed that the firing of the

first spike from a population encoding stimulus accounts for the discrimination of the stimulus from a population [Fernández et al., 2000].

Another common assumption is that the neuron’s firing probability varies with time and is exclusively a function of the input stimulus. These assumptions are not entirely true, as the neurons compose an intricate network with multiple feedforward and feedback paths, even in the primary neuronal layers of the retina. Moreover, a neuron cannot fire subsequent spikes with an arbitrarily small interspike interval because their internal and external ionic charge levels must be recovered in order to be able to fire again, which leads to the concepts of absolute and partial refractory periods [Berry II and Meister, 1998]. Several models take these phenomena into account by modifying their structure temporally.

A complete model of the retina attempts to map the incident light pattern on the photoreceptors’ layer to the spike train that comes out from the ganglion cells’ layer. This light pattern, focused on the retina by the eye’s optical system, can be generally described as a function of three variables, corresponding to: *i*) the bi-dimensional space vector,  $\mathbf{r} = [x\ y]^T$ , *ii*) as a function of time,  $t$ , that models the variation of environment components, by itself or due to the eyes’ movements, and *iii*) the light wavelength,  $\lambda$ , since the retina has different sensibilities according to the wavelength. The third spatial dimension, depth, that provides the 3D perception, is perceived by the brain through the angular difference between the two-dimensional images gathered individually by each eye. Hence, the visual stimulus signal can be described mathematically by the function  $s(\mathbf{r}, t, \lambda)$ . Normally, retina models consider the stimulus only as a function of space and time, so that its dependence on wavelength is deferred, or treated separately for each given wavelength and photoreceptor type. The light intensity patterns can also be continuous or discrete, both in space and/or in time (further details are given in Appendix B).

Although the surrounding visual stimulus are spatiotemporal, the retina response is frequently only modeled in time, and even in the spatiotemporal models it is common to consider the spatial and the temporal processing separable, and therefore are handled independently [Wandell, 1995].

In the following sections, it will be given representative examples of retina models and processing blocks from each model’s classes that have been thoroughly used and investigated.

### 3.3. Structural Models Processing Blocks

The class of structural neural models is broad. Nevertheless, the different structural models are further adaptations of the seminal Hodgkin-Huxley model. The majority of the models derived from the Hodgkin-Huxley (HH) model go in the direction of reducing its complexity, like the FitzHugh-Nagumo model [FitzHugh, 1961; Nagumo et al., 1962; Gerstner and Kistler, 2002], however a few models correspond to the inclusion of more terms in the HH model’s equations resulting in extremely complex models, difficult to analyze and to provide further insights into the neuron

### 3. Neuron and Retina Models

dynamics. On the other extreme is the Hopfield neuron model, which is a binary neuron model whose output is -1 or 1 depending on the linear combination of its input synapses, which are also -1 or 1. The Hopfield neuron model is used particularly in learning rules in associative memory [Dayan and Abbot, 2001], and are the basis of Hopfield's artificial neural networks.

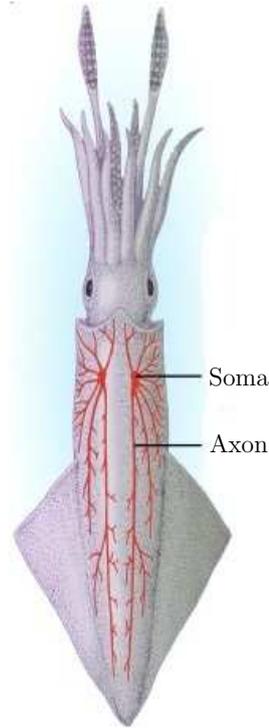
A set of neuron models with wide application in the modeling of different neural processing centers, including the retina's neuron cells, are the **IF** models [Reich et al., 1997; Gestri et al., 1980; Hérault and Durette, 2007]. These models are based on the characteristics of the neuronal membrane, whose properties are modeled as electrical components, being a simplification of the Hodgkin-Huxley nonlinear set of differential equations [Gerstner and Kistler, 2002]. Despite being around for a long time, [Hodgkin and Huxley, 1952; Brunel and van Mark C. W. Rossum, 2008], they still are the state of the art neuron models [Kistler et al., 1997; Jolivet et al., 2004; Burkitt, 2006; Mihalas and Niebur, 2009].

The article by Izhikevich [2004] provides a review of 20 different prominent spiking features, like different types of bursting, and classifies the most common spiking models according to their ability to display such features, and according to the number of floating point operations involved. Remarkably, the **HH** model is able to model all features observed in the neural response, however its computational complexity lead to the development of more simple spiking models.

#### 3.3.1. The Hodgkin-Huxley Neuron Model

The Hodgkin-Huxley model was developed with the aim to describe the flow of electric current through the surface membrane of the giant nerve fibre of the squid [Hodgkin and Huxley, 1952]. Fortunately nature provided the squid with a giant axon, up to 1 mm in diameter, amenable to be studied and analyzed even with the electronic devices available in the mid XX century. Additionally, comparing to other excitable nerve membrane, the squid axon was an ideal model system. It is a simple neuron, with essentially only two types of voltage-dependent conductances, and presented a suitably generic and tractable problem whose solution raised new and powerful techniques and the discovery of fundamental concepts [Nelson and Rinzel, 1995]. The giant axon extends from the head to the tail of the squid, and it innervates the squid's mantle muscle used to propel the squid through the water (see Fig. 3.1). The axon can be up to 1 mm in diameter, 100 to 1000 times larger than a mammalian axon, visible with naked eye.

By performing several experiments, Hodgkin and Huxley observed that the axon membrane potential  $V_m$  was related to the influx and efflux of ions through the axon's membrane of the giant squid ganglion neuron, and identified that these currents are due mainly to  $\text{Na}^+$  and  $\text{K}^-$  ions flux. By manipulating experimentally the ionic concentrations in the axon they were able to find the contributions of the different ionic conductances to the establishment and evolution of the membrane potential. The **HH** model for the neuron membrane is represented in Fig. 3.2.



**Figure 3.1.** The squid giant neuron (from [Randall et al., 2002])

The balance of charge in the electrical circuit of Fig. 3.2 is described by the differential equation:

$$C_m \frac{dV_m(t)}{dt} + I_{ion}(t) = I_s(t) \quad (3.2)$$

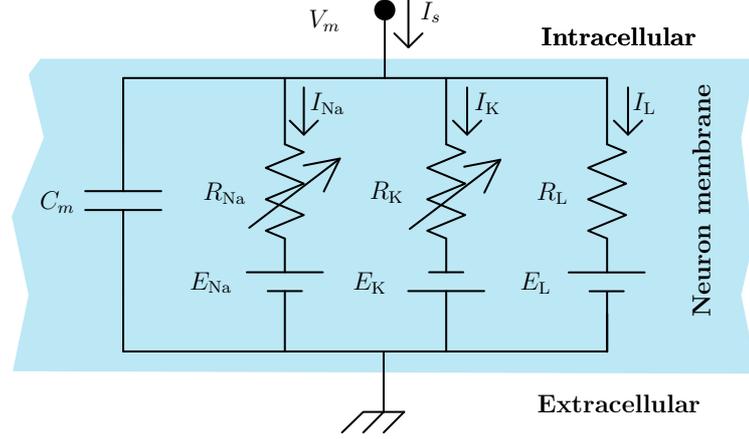
where  $C_m$  is the membrane capacitance,  $V_m$  is the intracellular potential relatively to the outside, and  $I_{ion}$  is the ionic current. The current  $I_s$  represents several types of external currents, like the synaptic current ( $I_{syn}$ ), and/or the current injected into the neuron directly by an intracellular electrode ( $I_{inj}$ ) (see Fig. B.4), and can also represent the current directly generated by a (visual) stimulus ( $I_{stim}$ ), each per unit area of cell membrane.

The current flow across the membrane has two major components. One associated with the charging of the membrane capacity  $C_m$ , and another major component is due to flux of ions across the membrane. The ionic current can be subdivided into current carried by sodium and potassium ions,  $I_{Na}$  and  $I_K$ , and a small leakage current  $I_L$  made up by the flux of chloride  $Cl^-$  and other minor ions types, such that

$$I_{ion} = I_{Na} + I_K + I_L. \quad (3.3)$$

The ionic permeability of the membrane is expressed in terms of the ionic conductances  $g_{Na}$ ,  $g_K$  and  $\bar{g}_L$ . Each ionic current is given by a conductance times the membrane potential,  $V_m$ , minus its respective reversal potential (Nernst equilibrium

### 3. Neuron and Retina Models



**Figure 3.2.** The Hodgkin-Huxley model of a neuron's membrane. The resistances  $R_{Na} = 1/g_{Na}$  and  $R_K = 1/g_K$  vary with time and with the membrane potential, while  $R_L = \bar{g}_L$  and other components are constant.

potential). The individual ionic currents are given by:

$$I_{Na} = g_{Na}(V_m - E_{Na}), \quad (3.4)$$

$$I_K = g_K(V_m - E_K), \quad (3.5)$$

$$I_L = g_L(V_m - E_L). \quad (3.6)$$

The conductance associated with each ionic current is a function of the membrane potential which can be seen as a result of the combining effects of a large number of microscopic ion channels embedded in the neuron's membrane that can be on the permissive or non-permissive state. When the physical gates of the ion channels are open ions can cross the membrane, while in the non-permissive state the channels are closed preventing the ions from crossing the membrane.

The conductances  $g_i$  in Eq. (3.4)-Eq. (3.6) depend if the states of the respective gates are on the permissive or non-permissive state. Hodgkin and Huxley modeled the ionic conductances using three types of empirical gating variables:  $m$ ,  $h$  and  $n$ ; and established empirically that the membrane conductance for each ion type depends on the gating variables as:

$$g_{Na} = \bar{g}_{Na}m^3h, \quad (3.7)$$

$$g_K = \bar{g}_Kn^4, \quad (3.8)$$

$$g_L = \bar{g}_L, \quad (3.9)$$

such that the ionic current is

$$\begin{aligned} I_{ion} &= \sum_k g_k(V_m - E_k) \\ &= \bar{g}_{Na}m^3h(V_m - E_{Na}) + \bar{g}_Kn^4(V_m - E_K) + \bar{g}_L(V_m - E_L), \end{aligned} \quad (3.10)$$

where  $m$  and  $h$  are the sodium gates, and  $n$  is the potassium gate. The conductances  $\bar{g}_{\text{Na}}$ ,  $\bar{g}_{\text{K}}$  and  $\bar{g}_{\text{L}}$  are the maximum values for the conductances for each ion type given in Table 3.1. The overall HH model becomes:

$$C_m \frac{dV_m(t)}{dt} = -\bar{g}_{\text{Na}} m^3 h (V_m(t) - E_{\text{Na}}) - \bar{g}_{\text{K}} n^4 (V_m(t) - E_{\text{K}}) - \bar{g}_{\text{L}} (V_m(t) - E_{\text{L}}) + I_s(t). \quad (3.11)$$

The gating variables  $m$ ,  $h$ , and  $n$  can be on the permissive (conducting) or non-permissive (non-conducting) state according to the time differential equations:

$$\frac{dm}{dt} = \alpha_m(V)(1 - m) - \beta_m(V)m, \quad (3.12)$$

$$\frac{dn}{dt} = \alpha_n(V)(1 - n) - \beta_n(V)n, \quad (3.13)$$

$$\frac{dh}{dt} = \alpha_h(V)(1 - h) - \beta_h(V)h. \quad (3.14)$$

The values of  $m$ ,  $h$  and  $n$  take values between 0 and 1 and give the fraction of gates for each ion type that are on the permissive state. Although  $m$ ,  $h$  and  $n$  do not depend directly on the membrane potential, the six rate constants  $\alpha_i$  and  $\beta_i$  in Eq. (3.12)–Eq. (3.13) depend on the membrane voltage. By adjusting a curve to the data of the giant axon of the squid, Hodgkin and Huxley found the empirical functions for  $\alpha_i$  and  $\beta_i$  as functions of the membrane potential. The rate constants per time (in ms) depend on the membrane potential (in mV) as:

$$\alpha_n(V) = 0.01(10 - V) \left[ \exp\left(\frac{10 - V}{10}\right) - 1 \right]^{-1} \quad (3.15)$$

$$\beta_n(V) = 0.125 \exp\left(\frac{-V}{80}\right) \quad (3.16)$$

$$\alpha_m(V) = 0.1(25 - V) \left[ \exp\left(\frac{25 - V}{10}\right) - 1 \right]^{-1} \quad (3.17)$$

$$\beta_m(V) = 4 \exp\left(\frac{-V}{18}\right) \quad (3.18)$$

$$\alpha_h(V) = 0.07 \exp\left(\frac{-V}{20}\right) \quad (3.19)$$

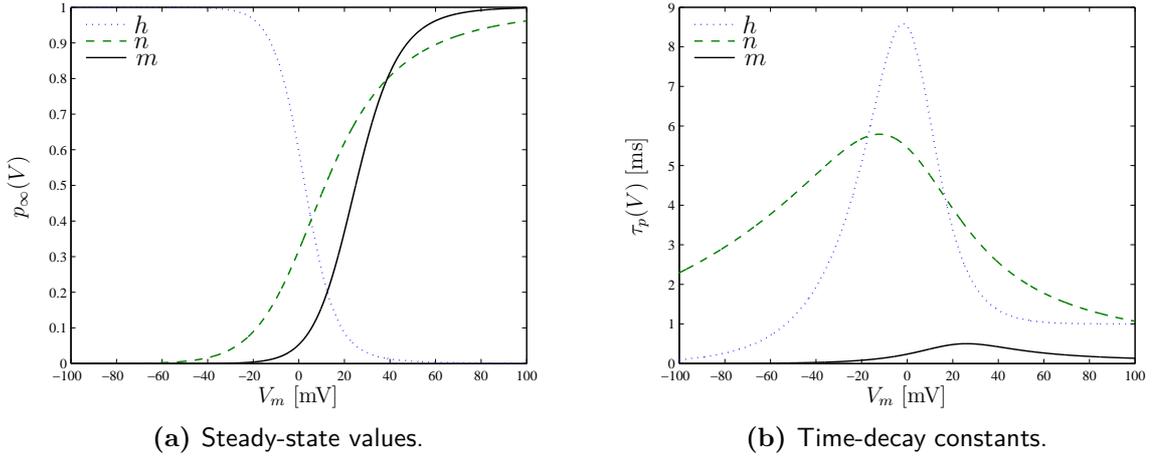
$$\beta_h(V) = \left[ \exp\left(\frac{30 - V}{10}\right) + 1 \right]^{-1}. \quad (3.20)$$

Keeping the membrane voltage  $V_m$  constant (clamped at a fixed value  $V$  in the physiological jargon) the fraction of gates in the permissive state will eventually reach a steady-state value:  $dp/dt = 0$ ,  $p \in \{m, h, n\}$  as  $t \rightarrow \infty$ , given by:

$$p_\infty(V) = \frac{\alpha_p(V)}{\alpha_p(V) + \beta_p(V)} \quad p \in \{m, n, h\}. \quad (3.21)$$

By changing the membrane voltage from  $V_m = V_0$  to  $V_m = V$  the rate constants change instantly according to Eq. (3.15)–Eq. (3.20), but the gating variables  $m$ ,  $h$ , and

### 3. Neuron and Retina Models



**Figure 3.3.** Hodgkin-Huxley steady-state  $m$ ,  $h$ ,  $n$  values (a) and decay time constants model's responses for a constant input stimulus current (b).

$n$  change according to Eq. (3.12)- Eq. (3.14). The solution of Eq. (3.12)- Eq. (3.14) when the membrane voltage changes from  $V_m = V_0$ , at  $t = 0$ , where the stationary value for  $p$ , for  $p \in m, n, h$ , is  $p_\infty(V_0) = \alpha_p(V_0)/[\alpha_p(V_0) + \beta_p(V_0)]$ , to  $V_m = V$  is given by:

$$p(t) = p_\infty(V) - [p_\infty(V) - p_\infty(V_0)] e^{-t/\tau_p}, \quad p \in \{m, h, n\} \quad (3.22)$$

where the decay time constants are given by:

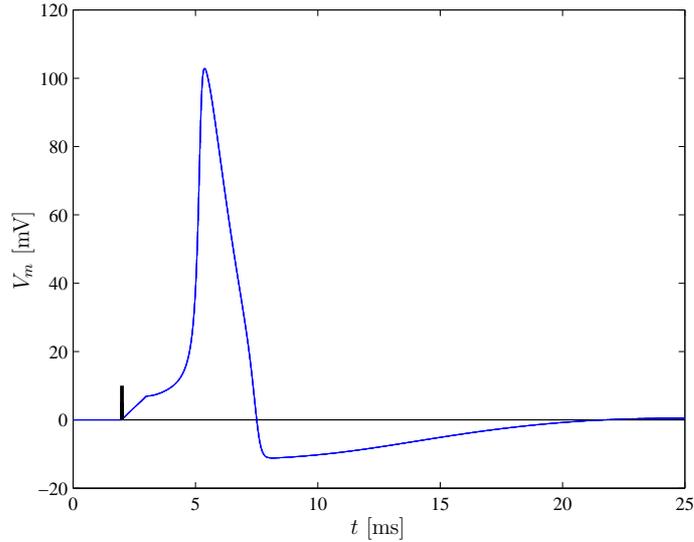
$$\tau_p(V) = \frac{1}{\alpha_p(V) + \beta_p(V)}, \quad (3.23)$$

and  $p_\infty(V) = \alpha_p(V)/[\alpha_p(V) + \beta_p(V)]$  is the stationary value for gating variable  $p \in \{m, h, n\}$ . The differential equations in Eq. (3.12)-Eq. (3.13) can be written as:

$$\frac{dp}{dt} = -\frac{1}{\tau_p(V)} [p - p_\infty(V)] \quad p \in m, h, n \quad (3.24)$$

indicating that  $p$  approaches the value  $p_\infty(V)$  with a time constant  $\tau_p(V)$ . Figure 3.3a displays the steady-state values for  $m$ ,  $h$  and  $n$  using Eq. (3.21) and Eq. (3.15)-Eq. (3.19), and Fig. 3.3b displays the gating time-constants.

To obtain a better insight on the dynamics of the HH model the equations Eq. (3.11) and Eq. (3.15)-Eq. (3.15) can be integrated numerically, using the parameters obtained by Hodgkin and Huxley for the squid axon given in Table 3.1. By injecting a short external current pulse with the form:  $I_s = I_{max}[\text{H}(t - t_0) - \text{H}(t - t_1)]$  with sufficient amplitude and time duration we can make the neuron to fire a spike. Figure 3.4 displays the waveform of an action potential, with an amplitude around 100mV, generated by the HH model for an input pulse of current with  $I_{max} = 8\mu\text{A}/\text{cm}^2$  with  $t_1 - t_0 = 1$  ms located at time  $t_0 = 2$  ms. In the action potential the spike itself can be identified, corresponding to the positive peak, followed by a refractory period where



**Figure 3.4.** Action potential waveform generated by the **HH** model for an external current impulse with  $8\mu\text{A}/\text{cm}^2$  of amplitude and duration of 1 ms (**I** – impulse current onset) for a resting potential  $E_r = 0\text{mV}$ .

the action potential goes below the resting potential ( $E_r = 0\text{V}$ ), where the neuron is unable to fire again, until the ion levels are recovered and the neuron membrane returns to the resting potential.

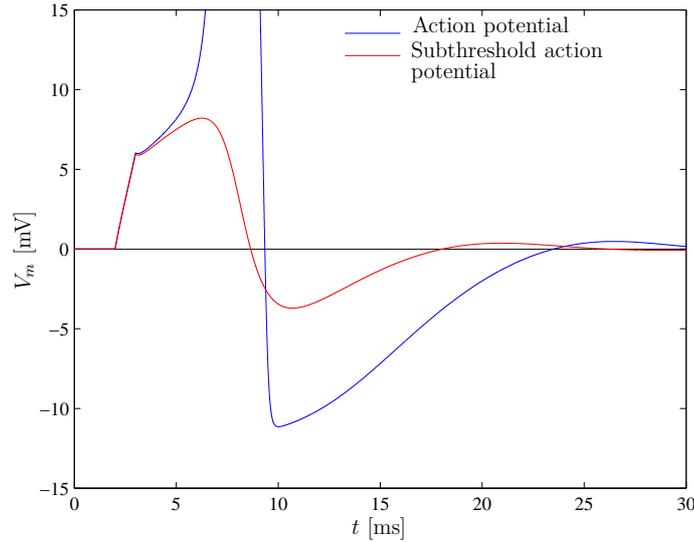
If the amplitude of the input current pulse is lower than a given value at around  $I_\theta = 6.9\mu\text{A}/\text{cm}^2$  the neuron membrane does not reach the threshold potential  $V_\theta$  and the neuron does not fire a spike. Figure 3.5 shows the waveform of a subthreshold action potential. The membrane potential for a the subthreshold potential stays below  $V_m = 10\text{mV}$ .

For a constant input current  $I_s = I_{max} H(t - t_0)$  the **HH** model fires spikes at a constant rate. Figure 3.5 displays a spike train generated by the **HH** model for  $I_s = 7 H(t - 2 [\text{ms}]) [\mu\text{A}/\text{cm}^2]$ . For a more intense stimulus (external current) the frequency of the spike train increases. The firing frequency raises more slowly as

Parameter	Value
$E_{\text{Na}}$	115 mV
$E_{\text{K}}$	-12 mV
$E_{\text{L}}$	10.6 mV
$C_m$	$1.0 \mu\text{F}/\text{cm}^2$
$\bar{g}_{\text{Na}}$	$120 \text{mS}/\text{cm}^2$
$\bar{g}_{\text{K}}$	$36 \text{mS}/\text{cm}^2$
$\bar{g}_{\text{L}}$	$0.3 \text{mS}/\text{cm}^2$

**Table 3.1.** Hodgkin-Huxley model parameters

### 3. Neuron and Retina Models



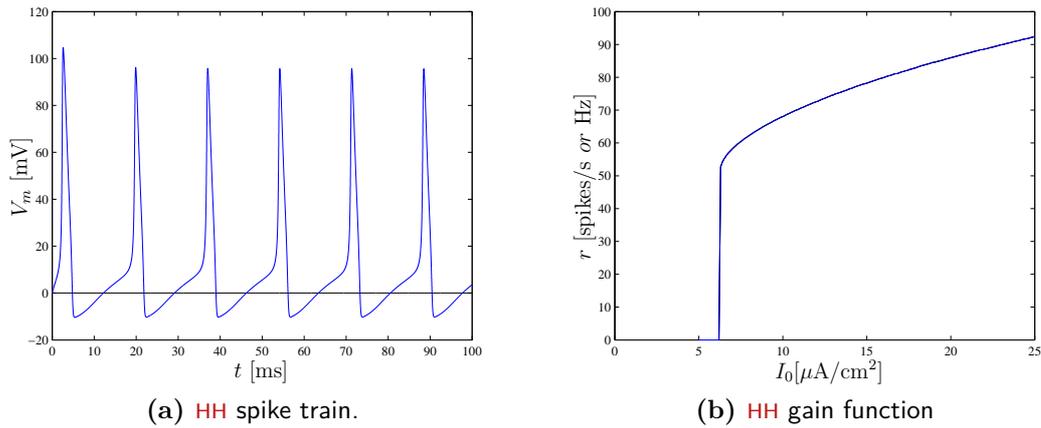
**Figure 3.5.** Subthreshold action potential waveform generated by the **HH** model using an external current impulse with  $6.9\mu\text{A}$  of amplitude and duration of 1 ms.

the input current amplitude increases, as depicted in Fig. 3.6b. The difficulty of the neuron to generate subsequent action potentials, even for a strong stimulus current, is due firstly to the fact that the membrane is hyperpolarized after a spike and the voltage needs more time to raise. Secondly, after the firing of a spike a large number of ionic channels are open and the resistance of the membrane is reduced, unlike at resting potential. Therefore, the depolarizing effect of the external current is smaller after an action potential. These effects account for the refractory behavior of every neuron.

The **HH** model mimics the neuron behavior in some important aspects. It reproduces accurately the time course of the action potential, describes naturally the refractory behavior of the neural system due to the slow recovery of sodium and potassium conductance after an action potential, and integrates the incoming synaptic current capacitively, essential characteristics of a neuron model dynamics [Abbott and Kepler, 1990; Izhikevich, 2004].

#### Extension of the Hodgkin-Huxley Model to Other Types of Neurons

The **HH** model describes accurately the dynamics of the squid giant axon whose membrane potential is mainly regulated by sodium and potassium ions channels. The membrane potential of the squid giant axon is controlled by an influx of sodium ions, that depolarize the membrane and raises an action potential, followed by an efflux of potassium ions that hyperpolarize the membrane again. Figure 3.7 illustrates the flow of sodium and potassium ions along the action potential. Nowadays, the accepted resting potential is  $-65\text{ mV}$  so that the equations Eq. (3.15)–Eq. (3.19) must be shifted, like the reversal potentials of Table 3.1 [Abbott and Kepler, 1990].



**Figure 3.6.** (a) Spike train generated by the Hodgkin-Huxley model in response to a step current input  $I_s = 7\text{H}(t)\ \mu\text{A}/\text{cm}^2$ , and (b) firing rate produced by the Hodgkin-Huxley model in response to a step current input  $I_s = I_0\text{H}(t)$ .

For other organisms, and other different types of neurons, the electric regulation of the membrane properties follows the mechanisms disclosed by Hodgkin and Huxley, however other types of ions join the sodium and potassium ions to model the neuron electrophysiological properties, including "persisting" sodium channels and more than one type of potassium channels, with various subdivisions.

The neuron membrane potential follows the equation Eq. (3.2), however the ionic current, described in the HH model by (3.10), is generalized to

$$I_{ion} = \sum_k I_k, \quad (3.25)$$

where each ionic current  $I_k$  is modeled by

$$I_k = \bar{g}_k m^{p_k} h^{q_k} h(V_m - E_k), \quad (3.26)$$

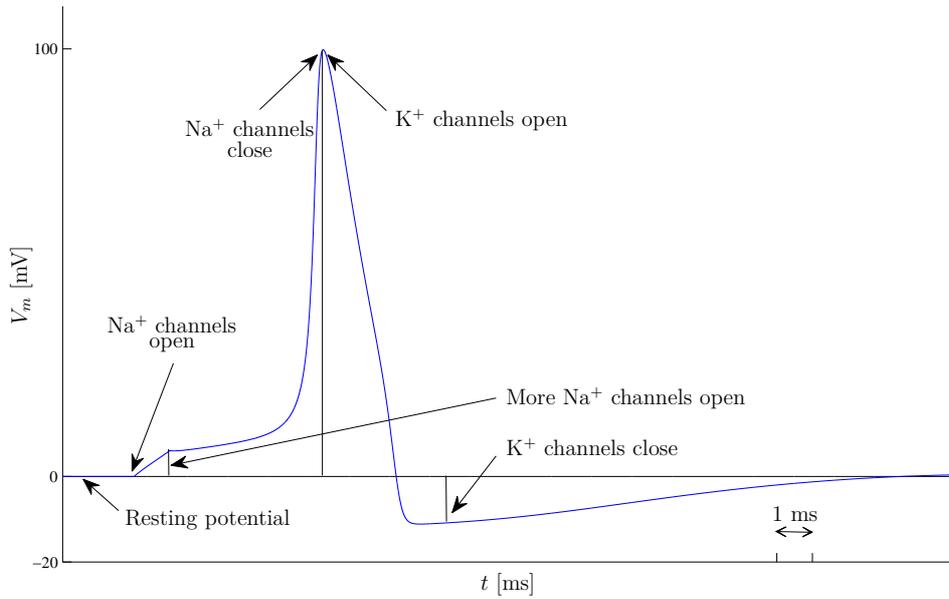
where  $\bar{g}_k$  is the maximum nominal conductance of the ion channel  $k$ ,  $E_k$  is the reversal potential, and  $m$  and  $h$  are the activation and inactivation variables. The parameters  $p_k$  and  $q_k$  depend on the ion type and take the values  $p_{\text{Na}} = 3$ ,  $q_{\text{Na}} = 1$  for sodium channels, and  $p_{\text{K}} = 4$ ,  $q_{\text{K}} = 0$  for potassium channels in the HH model. The HH model can even be further extended to include other ion channels that do not follow Eq. (3.25), like calcium channels [Gerstner and Kistler, 2002].

Despite its fidelity in describing the dynamics of the neuron's membrane potential, the HH model is computationally heavy [Izhikevich, 2004], and several simplification models are used that resembles the HH model.

### 3.3.2. The Integrate and Fire Model

The integrate-and-fire model can be seen as a simplification of the HH model. The behavior of the neuron was described in terms of its electrical capacitance and resistance properties earlier than the HH model [Brunel and van Mark C. W. Rossum,

### 3. Neuron and Retina Models

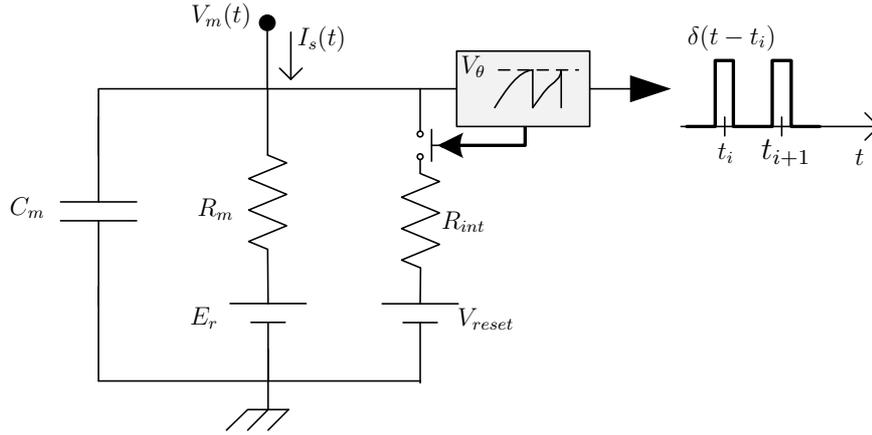


**Figure 3.7.** Gating dynamics across the generation of the neuron action potential.

2008], however the IF model was introduced later [Brunel and van Rossum, 2007]. It is a widespread model used for the modeling of many different types of neurons, including the modeling of the retina [Hérault and Durette, 2007] and other visual centers [Casti et al., 2008].

The most simple form for the IF model considers that the ionic current term is zero in Eq. (3.2), so that there is no leak current. This model was first analyzed in [Gerstein and Mandelbrot, 1964]. The subthreshold dynamics of the neuron is modeled by considering that it just integrates the input current. The firing mechanism is modeled considering that a spike is fired whenever the membrane voltage,  $V_m(t)$ , surpasses a threshold voltage,  $V_\theta$ , from below, and whenever a spike is fired the membrane potential is set to its reset value,  $V_{reset}$ . The general circuit for the (leaky) integrate-and-fire model is depicted in Fig. 3.8 where  $E_r$  is the rest reverse potential. The resting potential of the membrane takes values between -70mV to -65mV, and the threshold potential takes values around -55mV to -50mV.

The non leaky integrate-and-fire model considers that the neuronal membrane behaves like an ideal capacitor that is charged by the input stimulus current until it reaches a limit value,  $V_\theta$ , and discharges to a reset value,  $V_{reset}$ . This behavior is equivalent to considering a very high membrane resistance ( $R_m \rightarrow \infty$ ) in Fig. 3.8, which is equivalent to an open circuit, and by regarding the switch internal resistance as very small ( $R_{int} \rightarrow 0$ ), which is equivalent to a short circuit. The relation between the voltage,  $V_m(t)$ , across the neuron's membrane capacity,  $C_m$ , and the



**Figure 3.8.** The integrate-and-fire (I&F) model.

input stimulus current,  $I_s(t)$ , is given by the simple differential equation:

$$C_m \frac{dV_m(t)}{dt} = I_s(t) . \quad (3.27)$$

Equation (3.27) states that the input current  $I_s(t)$  is integrated to yield the membrane voltage  $V_m(t)$ . Whenever  $V_m(t)$ , crosses the threshold potential  $V_\theta$  a spike is fired and the switch is closed in Fig. 3.8, resetting the membrane potential to the reset value,  $V_{reset}$ .

The non leaky integrate-and-fire model is characterized by possessing an infinite memory, so that even a very small current applied for a sufficiently long period will eventually trigger a spike, and a constant firing rate is produced for every nonzero constant input current.

The membrane potential,  $V_m(t)$ , can be obtained by integrating the differential equation in Eq. (3.27) between the time instants  $t_i$ , and  $t$ , which gives

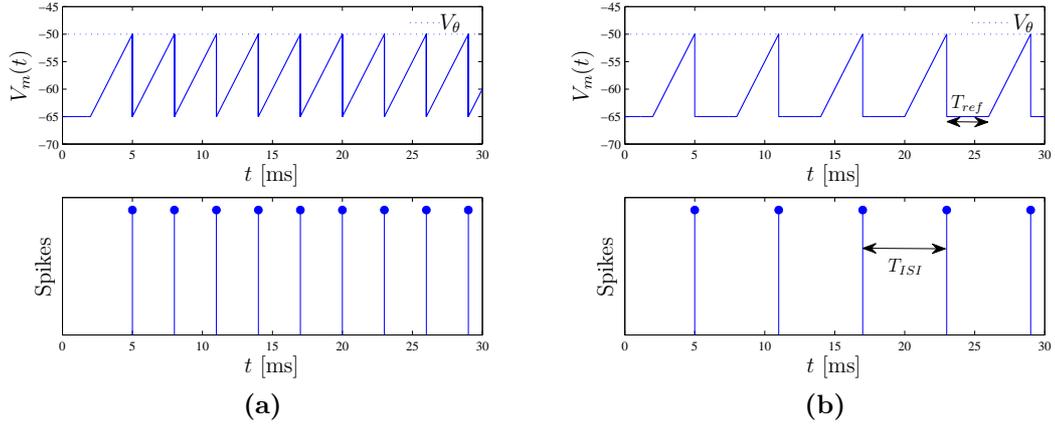
$$V_m(t) = V_m(t_i) + \frac{1}{C_m} \int_{t_i}^t I_s(t) dt, \quad t_i \leq t < t_{i+1} . \quad (3.28)$$

If it is stated that the neuron fired a spike at the time instant  $t_i$ , its membrane is at the reset potential for  $t = t_i$ ,  $V_m(t_i) = V_{reset}$ , from Eq. (3.28), it follows that the membrane potential is given by

$$V_m(t) = V_{reset} + \frac{1}{C_m} \int_{t_i}^t I_s(t) dt, \quad t_i \leq t < t_{i+1} , \quad (3.29)$$

between two successive spikes. Equation (3.29) represents the membrane potential excursion for  $t_i \leq t < t_{i+1}$ , where  $t_{i+1}$  is the time instant of the next generated spike after the spike at  $t_i$ .

### 3. Neuron and Retina Models



**Figure 3.9.** Integrate-and-fire model responses for a constant input stimulus current. *Top:* Membrane potential. *Bottom:* Generated spike sequence. (a) without refractory period (b) and with a refractory period. ( $I_s = 5\text{H}(t - 2)$   $\mu\text{A}$ ,  $C_m = 1$   $\mu\text{F}$ , and  $T_{ref} = 5$  ms; ( $V_{reset} = -65$  mV,  $V_\theta - V_{reset} = 15$  mV)).

For a constant input current,  $I_s(t) = I_s$ , from Eq. (3.29), the membrane potential is given by

$$V_m(t) = V_{reset} + \frac{1}{C_m} I_s (t_{i+1} - t_i), \quad t_i \leq t \leq t_{i+1}. \quad (3.30)$$

This expression is valid between the time instant that the neuron fired a spike,  $t = t_i$ , until the firing of the next spike,  $t = t_{i+1}$ , when the membrane is set to the reset potential and its voltage follows the expression of Eq. (3.30) once again. Since a neuron fires a spike whenever  $V_m(t)$  surpasses  $V_\theta$  from below, the time between two consecutive spikes,  $T_{isi}$ , can be obtained from Eq. (3.30) by knowing that  $V_m(t_{i+1}) = V_\theta$ , which gives

$$T_{isi} = t_{i+1} - t_i = C_m \frac{V_\theta - V_{reset}}{I_s}. \quad (3.31)$$

From the time between spikes, the firing rate,  $r = 1/T_{isi}$ , becomes

$$r = \frac{1}{C_m} \frac{I_s}{V_\theta - V_{reset}}, \quad (3.32)$$

which is constant since the time between spikes is also constant. Figure 3.9 displays the time evolution of the membrane voltage for a constant input current, and displays the corresponding spike sequence generated by the model. By adjusting the threshold potential to  $V_\theta - V_{reset}$ , the reset potential in Fig. 3.8 can be set to zero ( $V_{reset} = 0$ ).

Equation (3.32) states that the neuron can fire at an arbitrarily high frequency for a high input stimulus current,  $I_s$ . However, in real neurons this situation cannot occur due to the dynamics of the ion channels across the membrane, which introduces a refractory period after every firing. A neuron that has fired a spike is unable to produce another for a period of time  $T_{ref}$ , regardless of the strength of the stimulus.

After this dead time, in the absence of relative refractoriness, its firing ability returns immediately to its prior steady state. The integrate-and-fire model can be extended to take into account the refractory period by stating that after a fire, the neuron is inactive during a given time  $T_{ref} \neq 0$ , and only after this period it will be able to fire again. Referring to Fig. 3.8, the refractory period can be modeled by stating that the switch is closed during a time equal to  $T_{ref}$ , and only after this period it is open again and the capacitor can start charging to fire a new spike ( $R_m \rightarrow \infty, R_{int} \rightarrow 0$  for the non leaky IF model).

In the case of a constant input stimulus current, taking into account the refractory period  $T_{ref}$ , the time between two consecutive spikes is

$$T_{isi} = T_{ref} + \frac{C_m V_\theta}{I_s}, \quad (3.33)$$

which gives to the firing rate:

$$r = \frac{I_s}{C_m V_\theta + T_{ref} I_s}. \quad (3.34)$$

From Eq. (3.34), we observe that for high stimuli currents (meaning that the neuron is heavily excited) the maximum firing rate is

$$r_{max} = \lim_{I_s \rightarrow \infty} \frac{I_s}{C_m V_\theta + T_{ref} I_s} = \frac{1}{T_{ref}}, \quad (3.35)$$

which is limited by  $T_{ref}$ . Figure 3.9b displays the time evolution of the membrane voltage for the integrate-and-fire model considering a refractory period after the firing of each spike.

### 3.3.3. The Leaky Integrate-and-Fire Model

The leaky integrate-and-fire (LIF) model is closer to the HH model. It assumes that the ionic conductances are constant, so that the membrane conductance is only modeled by a passive leakage term:  $I_{ion} = (V_m(t) - E_r)/R_m$ . The assumption that the conductances are constant is a good approximation for small fluctuations of the membrane potential around its resting potential. The neuron models of the IF type only model the subthreshold dynamics of the membrane potential. Relatively to the simple IF model, the LIF model takes into account the fact that the neuron's memory is finite by introducing a leakage term so that past input stimuli events are discarded as time goes by.

The leakage term models the current drain through the cell's membrane, and is included in the differential equation Eq. (3.27), thus turning into:

$$C_m \frac{dV_m(t)}{dt} = -\frac{V_m(t) - E_r}{R_m} + I_s(t) \quad (3.36)$$

which is similar to the HH model. A single current, similar to the leak current  $I_{leak}(t)$  in the HH model, is maintained proportional to the membrane voltage, through the

### 3. Neuron and Retina Models

membrane resistance  $R_m$  following Ohm's law. The leaky integrate-and-fire model is represented by the electric circuit in Fig. 3.8, where the resistance  $R_m$  is connected in parallel with the membrane capacity  $C_m$ . Frequently the rest potential is considered to be zero:  $E_r = 0$  V. The conductance  $1/R_m$  encompasses the different ionic and leak conductances of the membrane that are assumed to be constant and independent of the membrane potential.

By introducing the time constant  $\tau_m = R_m C_m$ , the differential equation for the leaky integrate-and-fire model takes the form:

$$\tau_m \frac{dV_m(t)}{dt} + V_m(t) - E_r = R_m I_s(t), \quad (3.37)$$

where the time constant  $\tau_m$  characterizes the membrane dynamics. If the input current compensates the leak current, the membrane potential keeps rising until it reaches the firing threshold,  $V_\theta$  and a spike is fired. After the firing of a spike the membrane potential is set to its reset value,  $V_{reset}$ . In the absence of a stimulus current the membrane potential decays exponentially to its rest potential,  $E_r$ .

For the case of a constant input stimulus current,  $I_s$  must be strictly bigger than  $(V_\theta - E_r)/R_m$ , so that it can compensate the maximum drain current,  $I_\theta = (V_\theta - E_r)/R_m$ , in order for the neuron to fire. For a constant input current  $I_s(t) = I_s$ , therefore Eq. (3.37) becomes

$$C_m \frac{dV_m(t)}{dt} + \frac{V_m(t) - E_r}{R_m} = I_s. \quad (3.38)$$

Considering the initial condition that at the time instant  $t_i$ , the neuron just fired a spike so that its membrane potential is at the reset value  $V_m(t_i) = V_{reset}$ , the solution of Eq. (3.38) is:

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-(t-t_i)/\tau_m} + R_m I_s \left(1 - e^{-(t-t_i)/\tau_m}\right). \quad (3.39)$$

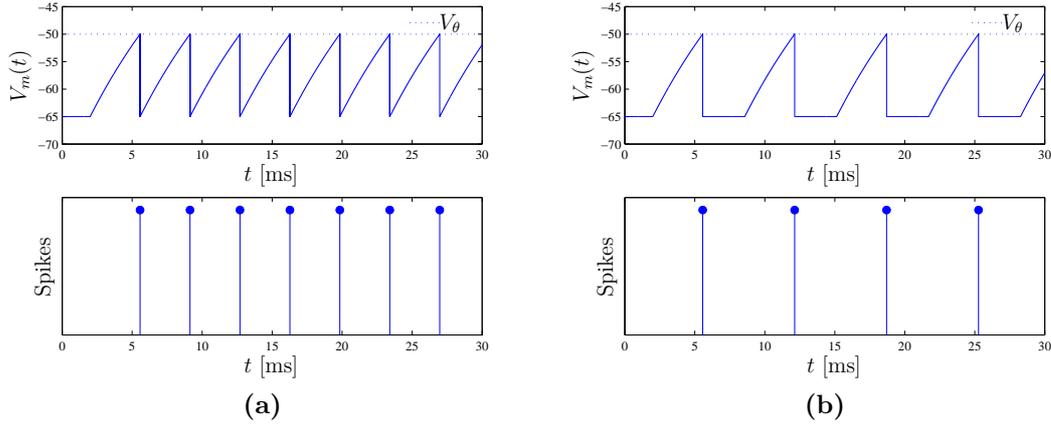
Knowing that a spike occurs whenever  $V_m(t) = V_\theta$ , the time of the next spike occurrence for a constant input current can be obtained from Eq. (3.39). Making  $V_m(t_{i+1}) = V_\theta$  in Eq. (3.39) the interspike time interval (ISI) is

$$\Delta t = t_{i+1} - t_i = \tau_m \ln \left( \frac{R_m I_s - (V_{reset} - E_r)}{R_m I_s - (V_\theta - E_r)} \right). \quad (3.40)$$

By introducing an absolute refractory period of  $T_{ref}$ , the ISI becomes  $T = T_{ref} + (t_{i+1} - t_i)$ , so that the firing rate for the leaky integrate-and-fire model for a constant input current is

$$r = \begin{cases} 0, & I_s \leq I_\theta \\ \left[ T_{ref} + \tau_m \ln \left( \frac{R_m I_s - (V_{reset} - E_r)}{R_m I_s - (V_\theta - E_r)} \right) \right]^{-1}, & I_s > I_\theta \end{cases}. \quad (3.41)$$

Figure 3.10b displays the model's membrane potential for the leaky integrate-and-fire model stimulated by a constant input current with an absolute refractory period.



**Figure 3.10.** Leaky integrate-and-fire model's response for a constant input stimulus current. *Top:* Membrane potential. *Bottom:* Generated spike sequence. (a) without refractory period (b) and with a refractory period. ( $I_s = 5H(t - 2 \text{ ms}) \mu\text{A}$ ,  $C_m = 1 \mu\text{F}$ ,  $\tau_m = 10 \text{ ms}$ , and  $T_{ref} = 3 \text{ ms}$ ; ( $E_r = V_{reset} = -65 \text{ mV}$ ,  $V_\theta = -50 \text{ mV}$ )).

The general solution of the differential equation for the LIF model can be found for a general input current stimulus. From the differential equations theory (or the Laplace transform theory) the solution of Eq. (3.36) is

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) + \frac{1}{C_m} \int_{t_i}^t e^{-\frac{t-\tau}{\tau_m}} I_s(\tau) d\tau, \quad t \geq t_i. \quad (3.42)$$

Another convenient way of representing Eq. (3.42) is

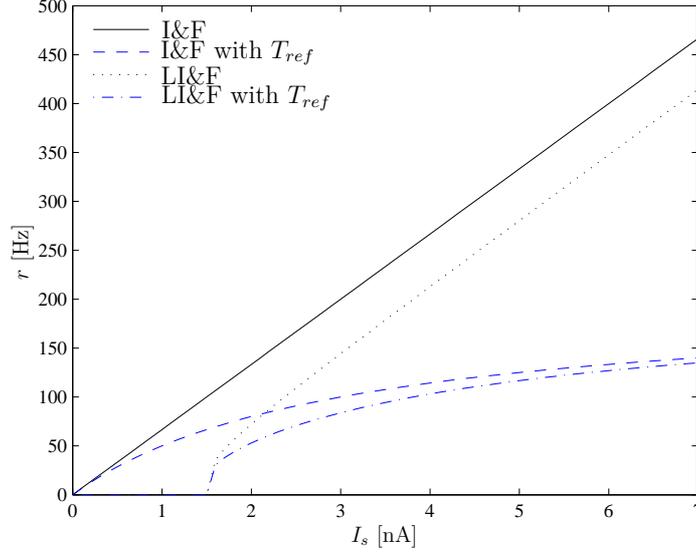
$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{(t-t_i)}{\tau_m}} + \frac{1}{C_m} \int_0^{t-t_i} e^{-x/\tau_m} I_s(t-x) dx, \quad t \geq t_i, \quad (3.43)$$

where the integration variable is the time elapsed since the last spike and  $V_m(t)$  is the membrane potential given that it was occurred a spike at instant  $t_i$ . By a change of variables the integral in Eq. (3.42) can be transformed into a convolution integral. Recurring to the convolution operator Eq. (3.42) can be written as:

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{(t-t_i)}{\tau_m}} H(t - t_i) + \frac{1}{C_m} \left[ e^{-\frac{(t-t_i)}{\tau_m}} H(t - t_i) \right] * [I_s(t + t_i) H(t)]. \quad (3.44)$$

Equation (3.44) expresses that after the firing of a spike at time  $t_i$ , the membrane potential is reset to potential  $V_{reset}$ , that decays exponentially to the rest potential,  $E_r$ , to which is added the convolution of the input stimulus current since the firing of the last spike, that occurred at  $t_i$ , by the low-pass filter with impulse response:  $h(t) = e^{-t/\tau_m} H(t)$ . In many situations the rest potential is made to be zero:  $E_r = 0$ . Equation (3.44) is valid from the firing of a spike (time instant  $t_i$ ) until the membrane

### 3. Neuron and Retina Models



**Figure 3.11.** Firing rate versus stimulus current for the integrate-and-fire model. ( $C_m = 1 \mu\text{F}$ ,  $\tau_m = 10 \text{ ms}$ , and  $T_{ref} = 5 \text{ ms}$ ,  $E_r = V_{reset} = -65 \text{ mV}$ ,  $V_\theta = -50 \text{ mV}$ ).

potential reaches the threshold value  $V_\theta$ , when a new spike is fired (time instant  $t_{i+1}$ ), and the potential is reset  $V_m(t_{i+1}) = V_{reset}$  and integration restarts. Equation (3.44) is used repeatedly to obtain the next firing instant by considering the previous time instant  $t_{i+1}$  as the new  $t_i$ .

To implement the LIF model in a digital computer the discrete counterpart of the differential equation in Eq. (3.37) must be obtained. In the discrete implementation, we want to find the membrane voltage,  $V_m(t)$  for the time instants  $t = nT_s$ , where  $n = 0, 1, 2, \dots$ , and where  $T_s$  is the sampling period, corresponding to find  $V_m[0], V_m[T_s], V_m[2T_s], \dots, V_m[nT_s], V_m[(n+1)T_s], \dots$ . The simplest way to obtain the discrete form of Eq. (3.37) is to approximate the continuous derivative by the forward discrete approximation, also known as the Euler approximation, given by:

$$\frac{dy(nT_s)}{dt} = \frac{y[(n+1)T_s] - y[nT_s]}{T_s}. \quad (3.45)$$

This approximation improves as the sampling period  $T_s$  tends to zero. Applying this approximation to the differential equation in Eq. (3.37), and dropping the sampling period in the arguments, it becomes

$$\tau_m \frac{V_m[(n+1)] - V_m[n]}{T_s} + V_m[n] - E_r = R_m I_s[n], \quad (3.46)$$

in discrete-time. Applying basic mathematical manipulations Eq. (3.46) becomes:

$$V_m[n+1] = \left(1 - \frac{T_s}{\tau_m}\right) V_m[n] + \frac{T_s}{\tau_m} E_r + \frac{T_s}{C_m} I_s[n]. \quad (3.47)$$

The membrane potential should be initialized with the rest potential,  $V_m[0] = E_r$ , and whenever it crosses the threshold potential, when  $V_m[n] \geq V_\theta$ , a spike is fired and its value is set to the reset potential,  $V_m[n] = V_{reset}$ . Another transformation from the continuous to discrete domain, frequently used in signal processing [Oppenheim et al., 1999a], is the bilinear transform which gives:

$$V_m[n+1] = \frac{2 - T_s/\tau_m}{2 + T_s/\tau_m} V_m[n] + \frac{2T_s/\tau_m}{2 + T_s/\tau_m} E_r + \frac{R_m T_s/\tau_m}{C_m(2 + T_s/\tau_m)} [I_s[n] + I_s[n+1]] , \quad (3.48)$$

for the difference equation of the LIF model.

### 3.3.4. Stochastic Integrate-and-Fire Models

*In vivo* output of a neuron shows a certain degree of irregularity even for stable input stimuli [Reich et al., 1997]. This behavior is pervasive to all nervous structures to a lesser or greater extent, as in the cortex structures [Kuriscak et al., 2012] and motion sensitive neurons [Lewen et al., 1997]. This is a neural phenomena whose origin is not completely understood and matter of current research [Kuriscak et al., 2012].

Like other neural systems, the response of the retina also shows a certain degree of variability even when stimulated by the same visual stimulus [Berry et al., 1997; Berry II and Meister, 1998; Uzzell and Chichilnisky, 2004]. Even for a stationary external injected current the response of a neuron is not perfectly periodic unlike predicted by the LIF model and is depicted in Fig. 3.10. This same behavior is obtained by the HH model as Fig. 3.6b shows. To mimic the variations in the neural response the LIF model is extended by adding a noise source to its dynamics.

Mainly two methods are used to include a degree of variability in the neuron models' response by noise addition: by adding an escape function to the firing threshold [Jolivet and Gerstner, 2004] or by including diffusive noise into the model input [Plesser and Gerstner, 2000].

#### Escape Neuron Models

In the escape models the firing threshold is considered to be variable, or noisy, so that the neuron can fire even when the formal firing threshold  $V_\theta$  has not been surpassed. This is modeled by introducing a firing intensity or escape rate which depends on the actual state of the neuron. In the standard LIF model the neuron fires a spike whenever  $V_m$  reaches the threshold potential  $V_\theta$ . In the escape model the neuron can fire with a given probability as a function of the distance between the actual membrane and the threshold, so that spikes can occur with a probability density given by

$$r(t) = f(V_m(t) - V_\theta), \quad (3.49)$$

where the probability increases as the distance  $|V_m(t) - V_\theta|$  gets smaller. In Eq. (3.49)  $r(t)$  is called the stochastic intensity in the context of mathematical theory of point

### 3. Neuron and Retina Models

processes [Papoulis and Pillai, 2002], and firing rate intensity in the context of neuron models (see Sec. B.2.1). In more general and elaborated models the functional in Eq. (3.49) can also depend on the variation of the membrane potential and on time directly, so that Eq. (3.49) becomes

$$r(t) = f[dV_m(t)/dt, V_m(t), t]. \quad (3.50)$$

The escape function in Eq. (3.50) has the restriction that  $f \rightarrow 0$  for  $V_m \rightarrow -\infty$ . For the escape function:

$$f[V_m(t) - V_\theta] = \begin{cases} 0, & V_m(t) < V_\theta \\ 1/\Delta, & V_m(t) \geq V_\theta \end{cases}, \quad (3.51)$$

where  $\Delta$  is the mean escape rate, the neuron never fires for  $V_m(t) < V_\theta$ , and for  $\Delta \rightarrow 0$  it fires immediately whenever  $V_m(t)$  crosses the threshold  $V_\theta$ , resembling the deterministic LIF model. Other smooth escape functions are [Gerstner and Kistler, 2002]:

$$f[V_m(t) - V_\theta] = \frac{1}{\tau_0} e^{\beta[V_m(t) - V_\theta]}, \quad (3.52)$$

with an exponentially growing escape rate and parameters  $\tau_0$  and  $\beta$ , that for  $\beta \rightarrow \infty$  resumes to the noiseless model Eq. (3.51). The escape rate can also be approximated by a piecewise linear function like:

$$\begin{aligned} f[V_m(t) - V_\theta] &= \begin{cases} 0, & V_m(t) < V_\theta \\ \beta[V_m(t) - V_\theta], & V_m(t) \geq V_\theta \end{cases} \\ &= \beta[V_m(t) - V_\theta] \text{H}(V_m(t) - V_\theta), \end{aligned} \quad (3.53)$$

where  $\beta$  is the slope of the escape function for  $V_m(t) \geq V_\theta$ . Another typical escape rate is given by the sigmoidal function:

$$\begin{aligned} f[V_m(t) - V_\theta] &= \frac{1}{2\Delta} \left[ 1 + \text{erf} \left( \frac{V_m(t) - V_\theta}{\sqrt{2}\sigma} \right) \right] \\ &= \frac{1}{\Delta} \Phi \left( \frac{V_m(t) - V_\theta}{\sigma} \right) \\ &= \frac{1}{\Delta} \Phi(V_m(t); V_\theta, \sigma^2), \end{aligned} \quad (3.54)$$

where  $\Delta$  is the time constant and  $\sigma$  the threshold noise intensity ( $\text{erf}(x)$  is the error function and  $\Phi(x)$  is the normal cumulative density function, see p. [xxi](#)).

Departing from a continuous escape model, the probability of firing during the time bin  $\Delta t$  of a neuron that emitted a spike at  $t_i$  is

$$P[\text{spike in } [t, t + \Delta t] | V_m(t), t_i] = \int_t^{t+\Delta t} r(\tau) d\tau \approx r(t) \Delta t. \quad (3.55)$$

Since the firing rate  $r(t) = f[V_m(t) - V_\theta]$  can be very high for  $V_m(t) > V_\theta$ , as given by Eq. (3.52) for example, then  $\Delta t$  must be very small so that  $r(t)\Delta t \leq 1$ , so that the probability axioms are not violated.

The discrete approximation for the probability that a spike occurs in the time interval  $[t, t + \Delta t]$  can be improved by calculating the probability that a spike does not occur in the interval  $[t, t + \Delta t]$ , which can be approximated by:

$$P[\text{no spike in } [t, t + \Delta t] | V_m(t), t_i] = \exp\left(-\int_t^{t+\Delta t} \underbrace{f[V_m(\tau) - V_\theta]}_{r(\tau)} d\tau\right) \quad (3.56)$$

$$\approx \exp(-f[V_m(t) - V_\theta]\Delta t)$$

so that Eq. (3.55) can be approximated by

$$P[\text{spike in } [t, t + \Delta t] | V_m(t), t_i] = 1 - \exp\{-f[V_m(t) - V_\theta]\Delta t\}. \quad (3.57)$$

which has the advantage that even if the escape function, defined by:  $r(t) = f[V_m(t) - V_\theta]$ , is unbounded the spiking probability is bounded:

$$0 \leq P[\text{spike in } [t, t + \Delta t] | V_m(t), t_i] \leq 1. \quad (3.58)$$

### Stochastic Integrate-and-Fire Model

A simple way to introduce variability in the output of a noise model is to add a noise component directly to the membrane potential, which constitutes the essence of diffusive neuron models. The LIF model differential equation of Eq. (3.37) can be extended by adding a noise term  $\xi(t)$  to the input current that becomes:

$$\tau_m \frac{dV_m(t)}{dt} = -V_m(t) + R_m I_s(t) + \sigma \xi(t), \quad (3.59)$$

by assuming a null resting potential:  $E_r = 0$ , and where  $\tau_m = R_m C_m$ , and  $\sigma$  gives the noise power. This equation corresponds to the Langevin equation for the LIF model. The term  $\xi(t)$  in Eq. (3.59) is a Gaussian white noise stochastic process with mean and autocorrelation function:

$$E\{\xi(t)\} = 0; \quad E\{\xi(t)\xi(t')\} = \delta(t - t'). \quad (3.60)$$

The noise term  $\sigma\xi(t)$  can be viewed as the influence of the synaptic current on the membrane potential resulting from the stochastic spike arrival from background activity in the neural network. For  $\sigma = 0$  this influence is null and the model resumes to the LIF model of Eq. (3.37), for an increasing  $\sigma$  in Eq. (3.59) the influence of the background noise raises. The neuron fires a spike whenever the membrane potential  $V_m(t)$  reaches the threshold potential  $V_\theta$ . This model is known as the stochastic

### 3. Neuron and Retina Models

leaky integrate-and-fire (**SLIF**) model, or diffusive neuron model, and is frequently used in several retina models [Keat et al., 2001; Pillow et al., 2005; Paninsky, 2006; Capela et al., 2007].

Due to the linearity of the differential equation Eq. (3.59) its solution follows Eq. (3.43) with the noise term added, becoming:

$$V_m(t) = V_{reset} e^{-\frac{(t-t_i)}{\tau_m}} + \frac{R_m}{\tau_m} \int_0^{t-t_i} e^{-x/\tau_m} I_s(t-x) dx + \frac{\sigma}{\tau_m} \int_0^{t-t_i} e^{-x/\tau_m} \xi(t-x) dx. \quad (3.61)$$

Introducing the continuous convolution operator Eq. (3.61) can be written as

$$V_m(t) = h_{LIF}(t) * [\tau_m V_{reset} \delta(t-t_i) + R_m I_s(t) H(t-t_i) + \sigma \xi(t) H(t-t_i)], \quad (3.62)$$

where

$$h_{LIF}(t) = \frac{1}{\tau_m} e^{-t/\tau_m} H(t), \quad (3.63)$$

is the **LIF** low-pass filter kernel with a pole at  $s_p = 1/\tau_m$ , where  $\tau_m = R_m C_m$ . Due to the variability of noise term it is impossible to obtain an analytic solution for Eq. (3.61). However, considering that the noise mean is zero, the expected trajectory of the membrane potential after the emission of a spike at  $t_i$ , where the membrane potential is reset to  $V_m(t_i) = V_{reset}$ , is given by

$$\mu_{V_m}(t) = E\{V_m(t)\} = V_{reset} e^{-(t-t_i)/\tau_m} + \frac{R}{\tau_m} \int_0^{t-t_i} e^{-x/\tau_m} I_s(t-x) dx. \quad (3.64)$$

Taking into account the variance of the noise process in Eq. (3.60), the variance of the membrane trajectory is given by:

$$\sigma_{V_m}^2(t) = E\{(V_m(t) - \mu_{V_m}(t))^2\} = \frac{\sigma^2}{2\tau_m} (1 - e^{-2(t-t_i)/\tau_m}), \quad (3.65)$$

showing that the noisy membrane trajectory drifts away from the noiseless membrane potential. The standard deviation from the noiseless trajectory is

$$\lim_{t \rightarrow +\infty} \sigma_{V_m}(t) = \sigma_{V_m(\infty)} = \frac{1}{\sqrt{2\tau_m}} \sigma, \quad (3.66)$$

which is approached with a time constant of  $\tau_m/2$  as indicated by Eq. (3.65).

## 3.4. Modeling Input Stimuli

The input stimuli in the models described before was described as a general input stimulus current,  $I_s(t)$ . This current accounts for different ways to stimulate the neuron: it can be a directly injected current into the neuron by an electrical electrode, like in experimental setups; can be induced by the pre-synaptic neurons to which the neuron's dendrites contact with; or, in retina models, this stimulus current can be obtained directly from the visual stimulus that impinge the retina's photoreceptors.

### 3.4.1. Synapses Modeling

In Sec. 2.2.2 the communication between neurons was briefly described quantitatively. After traveling along the axon, the action potential arrives at the neuron synapse where it triggers the opening of calcium channels provoking an efflux of calcium ions that causes the pre-synaptic neuron to release neurotransmitters into the synaptic cleft (see Fig. 2.1). The neurotransmitter molecules diffuse along the synaptic cleft and reach the postsynaptic neuron where they activate the receptors present at the synapses of the membrane, leading to the opening of specific ion channels which results on an excitatory post-synaptic current (EPSC) or inhibitory post-synaptic current (IPSC). The synaptic input can be modeled in two ways: as current synapses or conductance synapses [Burkitt, 2006].

#### Conductance Synapses

The EPSC and IPSC can be modeled recurring to a time dependent conductivity  $g_{syn}(t)$  by:

$$I_{syn}(t) = g_{syn}(t)(V_m - E_{syn}), \quad (3.67)$$

where  $E_{syn}$  is the reversal potential of the synapse, and  $V_m$  is the membrane potential. Typically, for inhibitory synapses  $E_{syn} \approx -75\text{mV}$  whereas for excitatory sinapses  $E_{syn} \approx 0\text{mV}$ , considering a resting potential of  $-65\text{mV}$ .

The synaptic conductance  $g_{syn}$  is frequently modeled as a combination of exponentials. In the case of excitatory synapses, that rely on the glutamate neurotransmitter, the main receptors are the AMPA and NMDA. The AMPA current activates and deactivates rapidly, while the NMDA receptor is slower [Dayan and Abbot, 2001]. The conductance of the postsynaptic membrane for the AMPA receptor, for a action potential occurring at  $t = t_0$ , is modeled by:

$$g_{syn} = \bar{g}_{syn} B \left( e^{-(t-t_0)/\tau_d} - e^{-(t-t_0)/\tau_r} \right) \text{H}(t - t_0), \quad (3.68)$$

where  $\tau_r$  and  $\tau_d$  is the rise and decay time, respectively, with  $\tau_r < \tau_d$ , and  $B$  is a normalization constant. The synaptic conductance is also frequently described by an  $\alpha$ -function model [Bernard et al., 1994] that has the form:

$$g_{syn} = \frac{\bar{g}_{syn}(t - t_0)}{\tau_r} e^{(1-(t-t_0))/\tau_r} \text{H}(t - t_0), \quad (3.69)$$

and for a spike occurring at  $t = t_0$  it grows from 0 and reaches its peak value  $\bar{g}_{syn}$  for  $t - t_0 = \tau_r$ .

#### Current Synapses

In a current synapse the stimulus current is independent of the membrane potential and is modeled by

$$I_{syn} = C_m \left[ \sum_{k=1}^{N_E} a_{E,k} \rho_{E,k}(t) + \sum_{k=1}^{N_I} a_{I,k} \rho_{I,k}(t) \right], \quad (3.70)$$

### 3. Neuron and Retina Models

where  $a_{E,k} > 0$  and  $a_{I,k} < 0$  give the variations in the potential due to a single synaptic action potential for the excitatory and inhibitory synapses, respectively.  $N_E$  and  $N_I$  are the total number of excitatory and inhibitory synapses, respectively. The associated charge delivered to the neuron by the excitatory synapse is  $C_m a_{E,k}$  and  $C_m a_{I,k}$  for the inhibitory synapse. The expressions:

$$\rho_{E,k}(t) = \sum_{t_{E,k}} \delta(t - t_{E,k}) \quad \text{and} \quad \rho_{I,k}(t) = \sum_{t_{I,k}} \delta(t - t_{I,k}), \quad (3.71)$$

represent excitatory and inhibitory sets of spikes (e.g. [Jolivet and Gerstner, 2004]), respectively.

For a spike train with a single excitatory spike,  $\rho(t) = \delta(t - t_0)$ , the contribution of the excitatory current given by Eq. (3.70) to the membrane potential in the LIF model is given by:

$$V_m(t) = a_E e^{-(t-t_0)/\tau_m} H(t - t_0), \quad (3.72)$$

where  $\tau_m = C_m R_m$  is the membrane time constant.

To consider a synaptic current with a finite time duration, and not only infinitesimal duration, the synaptic current can also be modeled as:

$$I_{syn}(t) = C_m a_E e^{-t/\tau_s} H(t), \quad (3.73)$$

where  $\tau_s$  is the synaptic time constant. This postsynaptic current can be modeled from an action potential with the form of Eq. (3.71) as the solution of the differential equation:

$$\frac{dI_{syn}(t)}{dt} + \frac{1}{\tau_s} I_{syn}(t) = C_m a_E \delta(t). \quad (3.74)$$

The postsynaptic membrane potential for a synaptic current with the form of Eq. (3.73) has the form:

$$V_m(t) = a_E \Delta \left[ e^{-t/\tau_m} - e^{-t/\tau_s} \right] H(t), \quad (3.75)$$

with  $1/\Delta = 1/\tau_s - 1/\tau_m$ .

Another typical model for the synapse current is given by the  $\alpha$ -function with the form:

$$I_{syn}(t) = C_m a_E t e^{-t/\tau_\alpha} H(t), \quad (3.76)$$

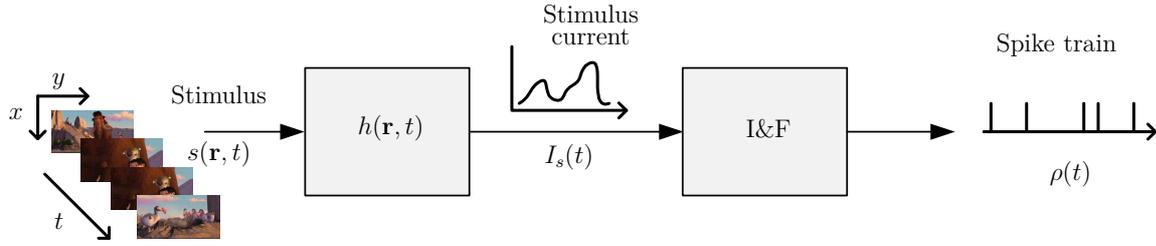
which can be obtained as the solution of the differential equation:

$$\frac{d^2 I_{syn}(t)}{dt^2} + \frac{2}{\tau_\alpha} \frac{dI_{syn}(t)}{dt} + \frac{1}{\tau_\alpha^2} I_{syn}(t) = C_m a_E \rho(t), \quad (3.77)$$

with  $\rho(t) = \delta(t)$  for a single spike. For the input corresponding to a for a barrier of spikes, given by  $\rho(t) = \sum_{t_i} \delta(t - t_i)$ , the solution of Eq. (3.77) gives the superposition of a series of postsynaptic currents. The membrane potential for the LIF model due to a single input postsynaptic current with the shape of an  $\alpha$ -function given by Eq. (3.77) is:

$$V_m(t) = a_E \Delta^2 \left( e^{-t/\tau_m} - e^{-t/\tau_\alpha} \left( 1 + \frac{t}{\Delta} \right) \right) H(t), \quad (3.78)$$

where  $1/\Delta = 1/\tau_\alpha - 1/\tau_m$ , and  $\tau_m$  is the membrane time constant.



**Figure 3.12.** Block diagram of a general integrate-and-fire model of the retina.

### 3.4.2. Visual Stimuli Input Current

In the first layer of the retina, the stimulus current comes from the photoreceptors' perception of the input light pattern, and can be obtained by the convolution of the visual stimuli with a filter or with a bank of filters, resembling the neuron receptive field. If one is interested in modeling the **RGC**, then the shape of the filter resembles the temporal and spatial receptive field of the neuron; it can be, for example, the spike triggered average (**STA**) (see Section **B.1.3**) [Chichilnisky, 2001; Pillow and Simoncelli, 2003]. It is common to use a set of base functions (e.g., Laguerre base functions [Tomás and Sousa, 2007], or distorted sine base functions [Keat et al., 2001] as the ones represented in Fig. 3.17) to compose the input filter. This filter can be interpreted as the spatio-temporal receptive field of the neuron that selects the relevant spatial and temporal features from the stimulus to fire a spike. Figure 3.12 shows a block diagram of a simple integrate-and-fire model of the retina with the input stimulus current generator block.

For a general visual input stimulus with spatial and temporal dependency, the input stimulus current is obtained through the expression:

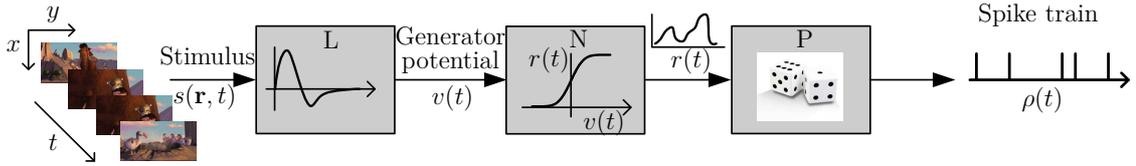
$$I_s(t) = \underbrace{\left( \sum_{i=0}^N c_i h_i(x, y, t) \right)}_{h(x, y, t)} * s(x, y, t), \quad (3.79)$$

where  $*$  represents the convolution operation that is accomplished in space and in time. The sum index  $N$  is the number of basis functions composing the impulse response of the input filter. In general the basis functions of the input filter,  $h_i(x, y, t)$ , can be nonlinear.

## 3.5. Retina Models

The following sections are devoted to the description of several functional models of the retina. These models are classified according to the type of processing they use to estimate the neural code of the **RGCs**. The functional models presented in the sequel are tuned and assessed in Chap. 5, in the context of metric analysis, to compare its performance.

### 3. Neuron and Retina Models



**Figure 3.13.** Block diagram of a general Linear-Nonlinear-Poisson (LNP) retina model.

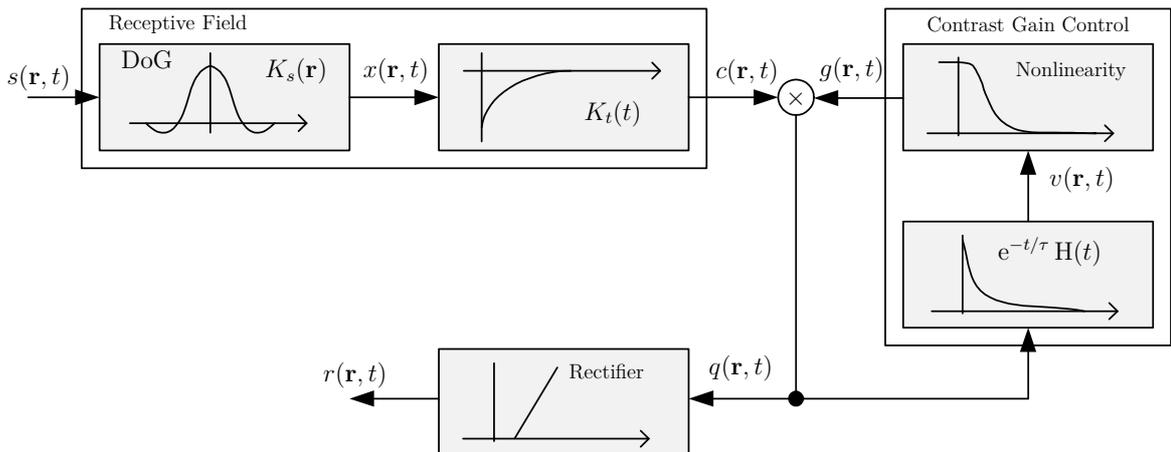
#### 3.5.1. Deterministic Models

The deterministic models do not take into account the variations in the retina response to a given stimulus. Using the same stimulus as the input for a deterministic model will always produce the same response. The variability of the neuron response is introduced by the spike generation mechanism (normally a non-homogeneous Poisson process).

The deterministic retina model described next was reported in [Wilke et al., 2001; Thiel et al., 2003], and outputs the firing rate. It was developed to model the temporal and spatial response of ON and of ON-OFF-type ganglion cells of a turtle retina to the movement of a white bar crossing its visual field. It was observed that the temporal behaviors of these two kinds of neurons are similar [Wilke et al., 2001]. The difference is mainly noted in their spatial response, where the ON-type ganglion cells respond only to the onset of the bar, whereas the ON-OFF-type cells respond to the onset and to the offset of the light bar.

#### Model Description

The block diagram of this model is depicted in Fig. 3.14. The signal  $s(\mathbf{r}, t)$  represents the light stimulus pattern that hits the retina as a function of time  $t$  and space  $\mathbf{r}$ , where the vector  $\mathbf{r}$  represents the stimulus spatial dependence,  $\mathbf{r} = [x \ y]^T$ . The first block of the model resembles the RF of the RGC. The spatiotemporal stimulus pattern is convolved with a kernel,  $K(\mathbf{r}, t)$ , resulting in the activation signal of the ganglion



**Figure 3.14.** Block diagram of the deterministic model.

cell,  $c(\mathbf{r}, t)$ . This operation is described by the convolution

$$c(\mathbf{r}, t) = K(\mathbf{r}, t) * s(\mathbf{r}, t) . \quad (3.80)$$

In a real retina, the spatiotemporal filter  $K(\mathbf{r}, t)$  may have a time delay:  $K(\mathbf{r}, t) = \tilde{K}(\mathbf{r}, t + \delta t)$ . For typical **RFs** of **RGCs** this kernel can be factorized, within a good approximation, in a time and a space kernel [Wandell, 1995], such that

$$K(\mathbf{r}, t) = K_s(\mathbf{r})K_t(t) . \quad (3.81)$$

To model the spatial **RF** of a **RGC**, a difference of Gaussians (**DoG**) [Wandell, 1995] is frequently used to simulate it [Rodieck, 1965]. A **DoG** has the mathematical description:

$$K_s(\mathbf{r}) = \frac{A_C}{2\pi\sigma_C^2} \exp\left(-\frac{\mathbf{r}^2}{2\sigma_C^2}\right) - \frac{A_S}{2\pi\sigma_S^2} \exp\left(-\frac{\mathbf{r}^2}{2\sigma_S^2}\right) ; \quad (3.82)$$

where the parameters  $A_C$  and  $A_S$  give the weight of the center of the **RF** relative to its surroundings. The parameters  $\sigma_C^2$  and  $\sigma_S^2$  ( $\sigma_C^2 < \sigma_S^2$ ) control the diameter of the center and of the outer Gaussian functions, respectively. The spatial behavior of the **RF** of the ON and OFF ganglion cells can be modeled by carefully choosing these weights. The effect of a **DoG** is a band-pass filtering of the spatial information of the image.

The temporal kernel corresponds to a high-pass filter, typically expressed by:

$$K_t(t) = \delta(t) - \alpha e^{-\alpha t} H(t) , \quad (3.83)$$

where  $1/\alpha$  is the decay rate of the filter response and  $H(t)$  is the continuous Heaviside unit step function.

The feedback loop of the model consists of a contrast gain control block inserted to capture the characteristics of the retinal response [Berry II et al., 1999]. The signal  $c(\mathbf{r}, t)$  is modulated by  $g(\mathbf{r}, t)$ , giving

$$q(\mathbf{r}, t) = c(\mathbf{r}, t)g(\mathbf{r}, t) . \quad (3.84)$$

The neuron activation signal,  $q(\mathbf{r}, t)$ , is obtained through the relation

$$q(\mathbf{r}, t) = g(\mathbf{r}, t) (K(\mathbf{r}, t) * s(\mathbf{r}, t)) . \quad (3.85)$$

The **CGC** loop includes a low-pass temporal filter that integrates the neuron activation signal, which usually has the following impulse response

$$v(\mathbf{r}, t) = B q(\mathbf{r}, t) * \left(e^{\frac{-t}{\tau}} H(t)\right) , \quad (3.86)$$

where parameter  $B$  controls the amplitude and  $\tau$  the time duration of the integration. Finally, the signal  $q(\mathbf{r}, t)$  passes through a static nonlinear function, resulting in a factor that modulates the **RF** output. The nonlinear function has the form

$$g(\mathbf{r}, t) = \frac{1}{1 + ([v(\mathbf{r}, t)]_+)^4} , \quad (3.87)$$

### 3. Neuron and Retina Models

where  $[x]_+ = x \text{H}(x)$  is the rectification operator.

At the output of the model, the activation signal is rectified to obtain the instantaneous firing rate of the **RGC**,  $r(\mathbf{r}, t)$ . The rectification function has the form

$$r(\mathbf{r}, t) = \tilde{\alpha}[q(\mathbf{r}, t) + \Theta]_+, \quad (3.88)$$

where  $\tilde{\alpha}$  establishes the scale and  $\Theta$  the baseline for the firing rate. When the stimulus  $s(\mathbf{r}, t)$  is spatially uniform and temporally constant, the activity signal  $q(\mathbf{r}, t)$  is zero. Thus, the neuron response is just the baseline neural activity, which is equal to the firing rate,  $r(\mathbf{r}, t) = \tilde{\alpha}\Theta$ .

#### Computational Implementation

The spatial processing carried out by the **RF**, corresponding to the first block of the model depicted in Fig. 3.14 inside the **RF** outer block, corresponds to the difference of two-dimensional Gaussian functions of Eq. (3.82). This kernel can be obtained as the product of two one-dimensional Gaussian bell-shaped curves for the center, one Gaussian for each direction, summed with the product of two other one-dimensional Gaussian curves for the outer sub-kernel. Taking the spatial center sub-kernel,  $K_{s_C}(x, y)$  as an example, this can be written as

$$K_{s_C}(x, y) = \frac{1}{2\pi\sigma_{C_x}\sigma_{C_y}} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_{C_x}^2} + \frac{y^2}{\sigma_{C_y}^2}\right)} = \frac{1}{2\pi\sigma_C^2} e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma_C^2}}, \quad (3.89)$$

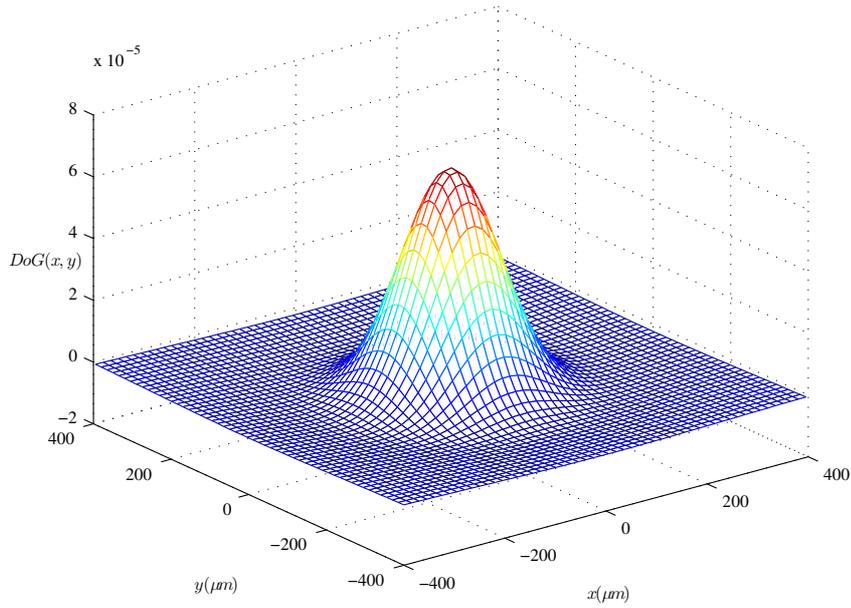
where  $x, y \in \mathbb{R}$ . The last equality in Eq. (3.89) results from the fact that the **RF** has no preferred direction, so  $\sigma_{C_x} = \sigma_{C_y} = \sigma_C$ . The outer sub-kernel has an identical form, and applying different weights to the center and outer sub-kernels, one obtains the expression in Eq. (3.82), where we used the fact that the squared magnitude of the position vector,  $\mathbf{r} = [x \ y]^T$ , is given by  $\mathbf{r}^2 = \mathbf{r}^T \mathbf{r} = x^2 + y^2$ .

In order to implement the analog filter given by Eq. (3.82) in a digital computer, it must first be converted into the discrete form. To discretize Eq. (3.82), we start by making  $x = n_1\Delta x$  and  $y = n_2\Delta y$ , where  $\Delta x$  and  $\Delta y$  are the spatial sampling interval length in the  $xx$  and  $yy$  directions, respectively, and  $n_1, n_2 \in \mathbb{Z}$ . Usually the sampling grid is equally spaced in the  $xx$  and  $yy$  directions, so that  $\Delta x = \Delta y = \Delta r$ . Starting by substituting the expression for sampling  $x$  and  $y$  in Eq. (3.89), we get

$$K_{s_C}(n_1\Delta r, n_2\Delta r) = C e^{-\frac{\Delta r^2}{2\sigma_C^2}(n_1^2+n_2^2)}, \quad (3.90)$$

where  $C$  is a constant to be determined. To obtain the correct value for the constant  $C$ , one must remember that the multiplicative factors in Eq. (3.89) were included in order to normalize the integral value of the Gaussian function to one. After sampling the Gaussian function to get a total sum equal to one, it must be normalized by the width of the sampling interval, so that Eq. (3.90) becomes:

$$K_{s_C}(n_1\Delta r, n_2\Delta r) = \frac{\Delta r^2}{2\pi\sigma_C^2} e^{-\frac{\Delta r^2}{2\sigma_C^2}(n_1^2+n_2^2)}. \quad (3.91)$$



**Figure 3.15.** Discrete spatial DoG.

Finally, introducing two terms like the one in Eq. (3.91) for the center and the other for the outer components of the **RF**, with weights  $A_C$  and  $A_S$ , respectively, and by omitting the argument dependence on  $\Delta r$  so that  $K_s(n_1\Delta r, n_2\Delta r) = K_s[n_1, n_2]$ , the discrete equation for the spatial **DoG** kernel becomes

$$K_s[n_1, n_2] = \frac{A_C}{2\pi(\sigma_C/\Delta r)^2} e^{-\frac{\Delta r^2}{2\sigma_C^2}(n_1^2+n_2^2)} - \frac{A_S}{2\pi(\sigma_S/\Delta r)^2} e^{-\frac{\Delta r^2}{2\sigma_S^2}(n_1^2+n_2^2)}. \quad (3.92)$$

The operation of the spatial kernel on the stimulus corresponds to its convolution with a matrix for the two-dimensional filter for each frame obtained at each instant of time sampling [Lim, 1990]. The shape of this spatial kernel is similar to the one of the **LoG** two-dimensional function, proposed for detecting intensity changes or contours in an image [Lim, 1990; Wulf, 2001]. Figure 3.15 shows a contour plot of a **DoG**, from Eq. (3.92), with  $\sigma_C = 80\mu m$ ,  $\sigma_S = 3\sigma_C$ ,  $A_C = 3$ ,  $A_S = 0.8A_C$ , and a spatial sampling period of  $\Delta r = 1\mu m$ .

The second block of the **RF**, corresponding to the temporal kernel, with expression Eq. (3.83) in continuous time, has the Laplace transform [Oppenheim et al., 1999b]

$$K_t(s) = \frac{s}{\alpha + s}, \quad (3.93)$$

where  $s$  is the Laplace complex variable. The transfer function of Eq. (3.93) has a pole at  $s_p = -\alpha$  and a zero at  $s_z = 0$ , corresponding to a high-pass temporal filter.

To achieve the discretization of this high-pass filter, we can map the Laplace frequency domain into the discrete  $z$ -transform domain by means of the bilinear transform. The bilinear transform is an adequate frequency mapping used to transform

### 3. Neuron and Retina Models

a continuous-time system representation, in the  $s$ -domain into the discrete-time  $z$ -domain [Oppenheim et al., 1999a] which as the expression

$$s = \frac{2}{T_s} \frac{1 - z^{-1}}{1 + z^{-1}}, \quad (3.94)$$

where  $T_s$  is the time sampling period.

By applying the bilinear transform to Eq. (3.93) we get

$$\begin{aligned} K_t(z) &= K_t(s) \Big|_{s=\frac{2}{T_s} \frac{1-z^{-1}}{1+z^{-1}}} \\ &= \frac{2}{2 + \alpha T_s} \frac{1 - z^{-1}}{1 - \left(\frac{2 - \alpha T_s}{2 + \alpha T_s}\right) z^{-1}}. \end{aligned} \quad (3.95)$$

This digital filter has a pole at  $z_p = (2 - \alpha T_s)/(2 + \alpha T_s)$ , inside the unit circle.

Applying the properties of the  $z$ -transform [Ziemer et al., 1998], the difference equation relating the input and the output of the temporal kernel, designated by  $x[n]$  and  $c[n]$  in Fig. 3.14, respectively, can be written in a suitable form to be implemented in a computer as:

$$c[n] = \frac{2 - \alpha T_s}{2 + \alpha T_s} c[n - 1] + \frac{2}{2 + \alpha T_s} (x[n] - x[n - 1]). \quad (3.96)$$

The CGC is composed of two blocks: a low-pass filter and a nonlinear function. The low-pass filter of the CGC can be converted into an equivalent digital filter following the same procedure used before for the temporal kernel.

The Laplace transform of the CGC low-pass filter takes the form:

$$h(t) = B H(t) \exp(-t/\tau) \xrightarrow{\mathcal{L}} H(s) = \frac{B}{1/\tau + s}, \quad (3.97)$$

and by applying the bilinear transform, the expression for its transfer function is

$$\begin{aligned} H(z) &= H(s) \Big|_{s=\frac{2}{T_s} \frac{1-z^{-1}}{1+z^{-1}}} \\ &= \frac{B\tau T_s}{2\tau + T_s} \frac{1 + z^{-1}}{1 - \left(\frac{2\tau - T_s}{2\tau + T_s}\right) z^{-1}}. \end{aligned} \quad (3.98)$$

By using the  $z$ -transform properties, we get the difference equation relating the output signal  $v[n]$  with the input signal  $q[n]$  of the CGC low-pass filter:

$$v[n] = \frac{2\tau - T_s}{2\tau + T_s} v[n - 1] + \frac{B\tau T_s}{2\tau + T_s} (q[n] + q[n - 1]). \quad (3.99)$$

The nonlinearity of the model described by Eq. (3.87) can be directly implemented in the digital domain with no difficulty. Finally, the last block of the model, denoted by the rectifier block in the model structure of Fig. 3.14, corresponds to expression Eq. (3.88), whose computational implementation is also straightforward.

This model produces at its output the RGC firing rate. Thus, to obtain the neural response, a spike generator that generates spikes according to the input firing rate is required (see Sec. B.3).

### 3.5.2. Stochastic Models

The stochastic retina model discussed in this section was proposed by [Keat et al., 2001]. It produces as its output the spike train itself, avoiding the need to use an external spike generator. The produced spike trains change from trial to trial for the same stimulus due to the intrinsic stochastic behavior of the model. It attempts to directly model the variability present in the responses of the RGCs.

While other models take into account that these variabilities are part of a stochastic process adopted to generate the spike trains (the process is usually considered to be a Poisson process), this model outputs the spike train, directly modeling its variability by several means along the stimulus processing. The effort in the development of the model to include the variability of the spike trains, and not leave it for the spike generator, is justified by its proponents in consideration of the fact that the variability of the spike trains are kept within a limited interval, smaller than that of a Poisson process, and not accurately accounted for by an external spike generator. The variability exhibited by the spike trains is modeled by the inclusion of two Gaussian noise sources within the model.

This model is only concerned with the temporal dependencies of the retina response with the stimulus. However, it is suitable to be extended in order to include a spatial treatment of the stimulus. In particular, if it is assumed that the temporal and spatial processing of the stimulus can be separated, as discussed for the deterministic model, the spatial processing can be modeled by an adequate function [Dayan and Abbot, 2001], such as a DoG [Wandell, 1995].

This model was originally applied to predict the response of the retina to Gaussian random flicker stimuli for several different types of RGCs from different vertebrates, namely rabbit, salamander, and cat. It was also applied to model LGN neuron cells in the case of the cat [Keat et al., 2001].

#### Model Description

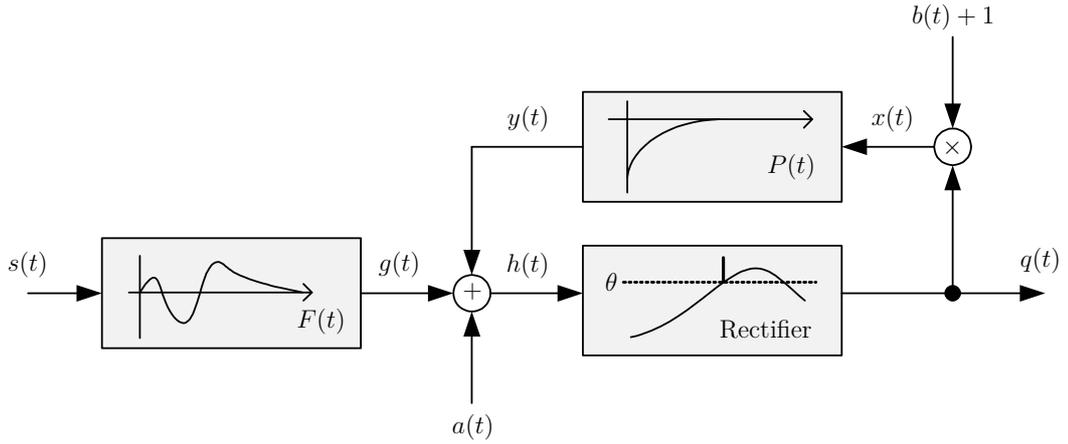
The structure of the stochastic model is represented in the block diagram of Fig. 3.16. The temporal stimulus  $s(t)$  is filtered by a linear filter, with impulse response  $F(t)$ , producing the generator potential  $g(t)$ . The filter function  $F(t)$  is synthesized using a linear combination of orthonormal functions,  $f_j$ , weighted by  $k_j$ :

$$F(t) = \sum_{j=1}^N k_j f_j(t) . \quad (3.100)$$

The basis functions are distorted sinusoidal functions with expression:

$$f_j(t) = \begin{cases} \sin \left( \pi j \left( 2 \frac{t}{\tau_F} - \left( \frac{t}{\tau_F} \right)^2 \right) \right), & \text{if } 0 \leq t \leq \tau_F \\ 0, & \text{otherwise} \end{cases} . \quad (3.101)$$

These base functions are preferred because of the lower number of parameters needed to reproduce the filter waveform accurately regarding, for example, the traditional sine functions, as in a Fourier series. This filter waveform, which is similar to the STA,



**Figure 3.16.** Block diagram of the pseudo-stochastic model.

has an impulse response with a high amplitude near the origin that decays rapidly over longer times. The parameter  $\tau_F$  controls the time duration of the filter response.

The signal  $g(t)$  in Fig. 3.16 results from the convolution of the stimulus  $s(t)$  with this linear filter  $F(t)$ :

$$g(t) = \int_{-\infty}^t s(\tau) F(t - \tau) d\tau . \quad (3.102)$$

Fifteen different components with the form Eq. (3.101) are used to synthesize  $F(t)$ , which corresponds to making  $N = 15$  in Eq. (3.100). This filter selects the stimuli patterns to which the model will fire events because the signal  $g(t)$  will be strongest when the visual stimulus follows a pattern similar to this filter response. Most of the model's parameters are used to appropriately adjust this filter function.

Posteriorly, the signal  $g(t)$  is summed with a noise component  $a(t)$ , and with a feedback signal coming from the feedback block, resulting in the signal  $h(t)$ , which is then compared with a threshold. The threshold block is composed of three main parts: *i*) the signal  $h(t)$  is compared with a threshold level  $\theta$ , so that it has a term of the form  $\delta(h(t) - \theta)$ , corresponding to the firing of a spike; *ii*) it has a term of the form  $H(\dot{h}(t))$  in order to guarantee the model only fires one spike when the signal crosses the threshold in the upward direction; and finally, *iii*) these two terms are multiplied by the derivative of  $h(t)$  with respect to time, denoted by  $\dot{h}(t)$ . Thus, only when the signal crosses the threshold from below a spike is fired, while when the signal  $h(t)$  crosses the threshold from above nothing happens, and the firing intensity is proportional to the intensity increase of  $h(t)$ . The output signal of the threshold block is:

$$q(t) = \delta(h(t) - \theta) \frac{dh(t)}{dt} H\left(\frac{dh(t)}{dt}\right) . \quad (3.103)$$

These first two blocks, the filter  $F(t)$  and the threshold block, are intended to predict the time occurrence of the firing events, correspondingly making  $a(t) = b(t) = P(t) = 0$  in the model structure of Fig. 3.16. In order to predict the correct number

of firing events, a feedback block is introduced in the model to take into account the refractory period of the ganglion cells after a firing event. Each fired spike triggers a negative after-potential  $P(t)$  that is added to the generator potential  $g(t)$ , lowering the signal  $h(t)$  immediately after a firing. The after-potential function  $P(t)$  has the form

$$P(t) = B e^{-t/\tau_p} H(t) , \quad (3.104)$$

where  $B$  and  $\tau_p$  are two other parameters of the model that define the characteristics of the low-pass filter. The parameter  $B$  controls the amplitude of the after potential, while  $\tau_p$  gives the time decay of the after potential.

The after-potential makes the signal  $h(t)$  drop below the threshold after the firing of a spike. However, if  $g(t)$  continues to rise in such a way that it compensates for the negative potential  $P(t)$ , the model will fire again, such that large increments of  $g(t)$  lead to a train of several spikes. After a firing event, the signal  $h(t)$  is lower than  $g(t)$  as a result of the accumulated after potentials, and the probability of subsequent firing events is reduced until the after-potential decays. This negative feedback loop simulates both repetitive firing within a firing event and a refractory behavior after the firing of a spike.

The input signal of the feedback block can be written as:

$$q(t) = \delta(h(t) - \theta) \dot{h}(t) H(\dot{h}(t)) , \quad (3.105)$$

where  $\dot{h}(t)$  is included in the threshold function such that the potential block input in the feedback loop is proportional to the slope, or instantaneous increase, of  $h(t)$ . With the addition of the feedback loop, the generator potential becomes

$$h(t) = g(t) + a(t) + \int_{-\infty}^t q(\tau)(1 + b(\tau))P(t - \tau)d\tau . \quad (3.106)$$

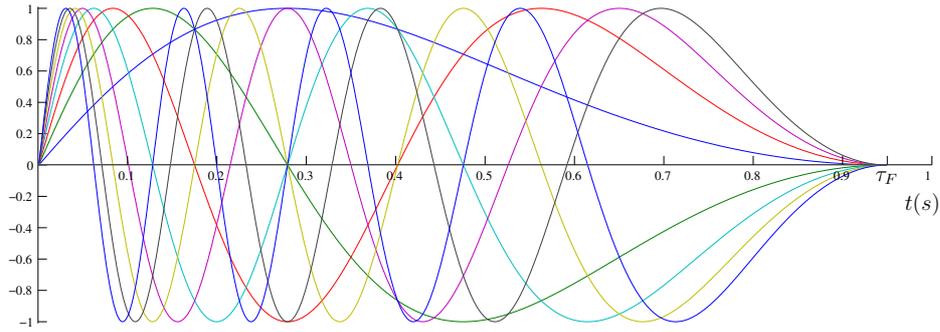
The output spike train is a series of delta functions, as described by Eq. (B.7), occurring at time instants  $t_i$ , whenever the generator potential  $h(t)$  crosses the threshold  $\theta$  from below. Its expression is equal to Eq. (3.105) without the  $h(t)$  derivative term, and is written as

$$\hat{\rho}(t) = \delta(h(t) - \theta) H(\dot{h}(t)) . \quad (3.107)$$

As mentioned previously, this model also attempts to model the variability of the neural response from trial to trial. This variability comprises the variation in the total number of spikes and the variation in the time instants that spikes occur. This is modeled by including two Gaussian noise sources:  $a(t)$  and  $b(t)$ . The amplitudes of the noise signals change over time, and consequently, the spike trains change between trials.

The random signal  $a(t)$  is added to the generator potential  $g(t)$  before the threshold block, introducing a random variability into the exact time of occurrence of the threshold crossing. This noise source has a Gaussian distribution with a zero mean, a standard deviation of  $\sigma_a$ , and an exponentially decaying autocorrelation function with

### 3. Neuron and Retina Models



**Figure 3.17.** Distorted sinus base functions ( $\tau_F = 0.95s$ ).

time constant  $\tau_a$ . The variability introduced by  $a(t)$  was not sufficient to reproduce the neural response variability, despite the fact that the spike occurrence variability was well modeled. The variability in the spike number was lower than the observed results from real neurons, and therefore the noise source  $b(t)$  was also included.

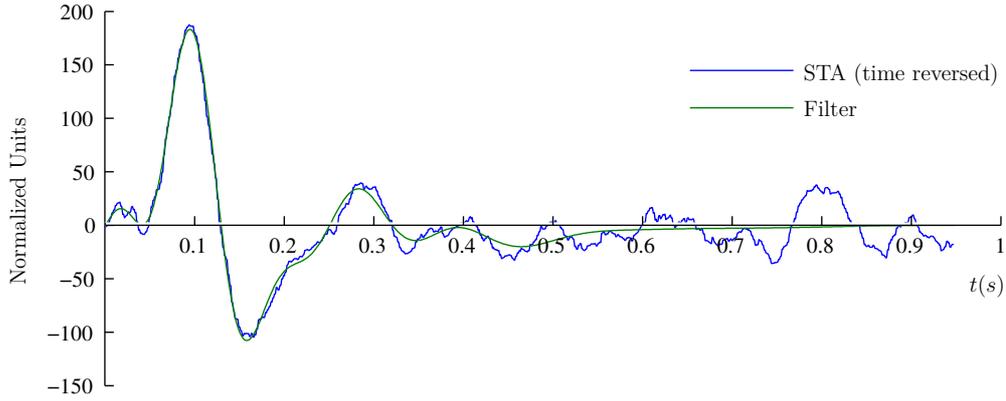
The noise source  $b(t)$  drives, in conjunction with the output of the threshold block, the negative after-potential generator by randomly modulating its amplitude after each spike. This noise source has a Gaussian distribution with a zero mean and a standard deviation of  $\sigma_b$ .

#### Computational Implementation

The stochastic retina model described in the previous section considers continuous signals and systems, and must be discretized in order to be implemented in a digital computer. The steps to implement this retina model in a digital computer are described.

The impulse response  $F(t)$  of the filter, represented by the first block in Fig. 3.16, was decomposed as a linear combination of basis functions consisting of distorted sinus functions described by (Eq. (3.101)). These functions were discretized using a sampling period  $T_s$ , for a filter length equal to  $\tau_F$ , for  $j = 1, \dots, N$ . After being sampled, the resulting  $N$  vectors were orthonormalized using the Gram-Schmidt procedure [Arfken and Weber, 2005], leading to the vectors  $\mathbf{f}_j$ , where  $j = 1, \dots, N$ . Figure 3.17 shows a plot of the first eight of these distorted sinus functions, with  $j = 1, \dots, 8$ .

The filter impulse response  $F(t)$  was initialized with a shape equal to the time reverse of the STA. The STA was defined in Section B.1.3 in its continuous and discrete form, and it is denoted in its discrete-time form by the vector  $\mathbf{s}_{spk}$ . The time reverse of the STA was decomposed as a linear combination of the basis functions, as Eq. (3.100) indicates. The discrete orthonormal vectors  $\mathbf{f}_j$ , with  $j = 1, \dots, N$ ,



**Figure 3.18.** Time reversal of the spike triggered average and its representation by basis functions ( $\tau_F = 0.95$  s;  $T_s = 1$  ms).

were brought together to compose the matrix  $\mathbf{F}_\perp$ , which is written as

$$\mathbf{F}_\perp = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{f}_1 & \mathbf{f}_2 & \cdots & \mathbf{f}_N \\ | & | & \cdots & | \end{bmatrix}. \quad (3.108)$$

The coefficients  $k_j$ ,  $j = 1, \dots, N$  in Eq. (3.100) that are components of the vector  $\mathbf{k}$  were initialized with the values resulting from the matrix-vector product:

$$\mathbf{k} = \mathbf{F}_\perp^T \tilde{\mathbf{s}}_{spk}, \quad (3.109)$$

where  $\tilde{\mathbf{s}}_{spk}$  is the vectorial representation of the time reverse of the STA.

Figure 3.18 shows the time reverse of the STA superimposed with its reconstruction using the base functions of Fig. 3.17 after orthonormalization. The filter impulse response, represented by the vector  $\mathbf{F}$ , is obtained through the matrix product

$$\mathbf{F} = \mathbf{F}_\perp \mathbf{k}, \quad (3.110)$$

and filters the stimulus entering the model.

The threshold block in Fig. 3.16, corresponding to Eq. (3.105), has a direct implementation. If the discrete input signal  $h[n-1]$  is smaller than the threshold  $\theta$ , but  $h[n]$  is bigger than  $\theta$  (meaning that there is a threshold crossing in the upward direction), a spike is fired and the output signal has the value  $q[n] = h[n] - h[n-1]$ . If this situation does not occur, then the output signal is made equal to  $q[n] = 0$ . It is worth noting that the signal  $q[n]$  is proportional to the slope of  $h[n]$  whenever it crosses the threshold from below, which means that the feedback potential block is excited by a signal proportional to the slope of the generator potential. The signal  $q[n]$  can be computed by using the discrete counterpart of Eq. (3.105),

$$q[n] = \delta [h[n] - \theta] \mathbf{H} [h[n] - h[n-1]] (h[n] - h[n-1]). \quad (3.111)$$

### 3. Neuron and Retina Models

As a matter of fact, in order to compute the equation, the term  $\delta[h[n]-\theta]$  in Eq. (3.111) must be made equal to one whenever  $h[n]$  is in the vicinity of  $\theta$ , that is when  $|h[n] - \theta| < \gamma$ , since the signal  $h[n]$  would be exactly equal to  $\theta$  only by chance.

The signal  $q[n]$  is convolved with the negative potential that has the continuous expression Eq. (3.104). This potential corresponds to a low-pass continuous filter, as can be seen from its Laplace transform in Eq. (3.112). The discrete counterpart of the after-potential can be obtained by applying the bilinear transform to map the Laplace frequency domain into the  $z$ -transform domain.

The Laplace transform of the negative potential is

$$P(t) = Be^{-t/\tau_p} \text{H}(t) \xrightarrow{\mathcal{L}} P(s) = \frac{B}{1/\tau_p + s}, \quad (3.112)$$

which has a unique pole at  $s_p = -1/\tau_p$  corresponding to a stable filter. The application of the bilinear transform to Eq. (3.112) leads to the same equivalent discrete filter in the  $z$ -transform domain as obtained in Eq. (3.99) with  $\tau = \tau_p$ :

$$\begin{aligned} P(z) &= P(s) \Big|_{s=\frac{2}{T_s} \frac{1-z^{-1}}{1+z^{-1}}} \\ &= \frac{B\tau_p T_s}{2\tau_p + T_s} \frac{1 + z^{-1}}{1 - \left(\frac{2\tau_p - T_s}{2\tau_p + T_s}\right)z^{-1}}. \end{aligned} \quad (3.113)$$

Referring to the properties of the  $z$ -transform [Ifeakor and Jervis, 2002], the difference equation for the filter corresponding to the negative feedback potential obtained from Eq. (3.113) is

$$y[n] = \frac{2\tau_p - T_s}{2\tau_p + T_s} y[n-1] + \frac{B\tau_p T_s}{2\tau_p + T_s} (x[n] + x[n-1]), \quad (3.114)$$

which is used in the computational implementation of the model.

The last items to include in the discrete implementation of the stochastic model are the Gaussian noise sources:  $a(t)$  and  $b(t)$ . Concerning  $b(t)$ , it is enough to generate a random sequence with a Gaussian distributed amplitude with a zero mean and unit variance and multiply each new generated sample by the desired standard deviation  $\sigma_b$ . In the case of  $a(t)$ , in addition to the variance  $\sigma_a^2$ , this noise sequence has an autocorrelation function with an exponential form with a constant decay rate equal to  $\tau_a$ .

To generate a discrete noise sequence equivalent to  $a(t)$ , a linear system driven by a discrete white noise sequence with a zero mean and unit variance can be used. The transfer function of this system must be determined such that its output is a random sequence with the desired statistical characteristics. This process is called prewhitening, as it is the reverse of whitening [Orfanidis, 1990]. The required continuous noise has an autocorrelation function with the form

$$R_a(\tau) = \sigma_a^2 e^{-|\tau|/\tau_a}, \quad (3.115)$$

that takes the value  $R_a(0) = \sigma_a^2$ , for a lag  $\tau = 0$ , corresponding to the variance of the continuous random noise. The discrete noise sequence should have an autocorrelation sequence equal to Eq. (3.115) at the sampling points. Using a sampling period equal to  $T_s$ , the autocorrelation sequence becomes

$$\begin{aligned} R_a[l] &= \sigma_a^2 e^{-|l|T_s/\tau_a} = \sigma_a^2 \underbrace{(e^{-T_s/\tau_a})}_{\varrho}^{|l|} \\ &= \sigma_a^2 \varrho^{|l|}. \end{aligned} \quad (3.116)$$

By considering a linear shift invariant system with impulse response  $h[n]$  subjected to the input  $x[n]$  that produces the output signal  $y[n] = h[n] * x[n]$ , if the autocorrelation of the input signal is  $R_x[l]$  and the output autocorrelation is  $R_y[l]$ , the relationship between these two discrete autocorrelation functions is [Therrien, 1992]

$$R_y[l] = h[l] * h[-l] * R_x[l]. \quad (3.117)$$

By taking the  $z$ -transform of this expression, we get

$$S_y[z] = H[z] H[z^{-1}] S_x[z], \quad (3.118)$$

where  $S_x[z]$  and  $S_y[z]$  are the  $z$ -transforms of the input and output signals, respectively, corresponding to their spectral densities.

A discrete pure white noise sequence,  $w[n]$ , with a zero mean and a variance  $\sigma_w^2$  has the autocorrelation function

$$R_w[l] = \sigma_w^2 \delta[l], \quad (3.119)$$

and for the particular case of a white noise sequence with unit variance, the spectral density is:

$$R_w[l] = \delta[l] \xrightarrow{\mathcal{Z}} S_w(z) = 1. \quad (3.120)$$

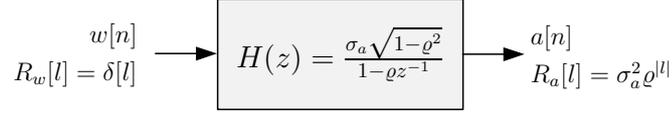
If a discrete white noise sequence with zero mean and unit variance is used as the input for the prewhitening filter, the system transfer function  $H(z)$  must be found that transforms this input white noise sequence into a discrete output random sequence with the spectral density

$$R_a[l] = \sigma_a^2 \varrho^{|l|} \xrightarrow{\mathcal{Z}} S_a(z) = \frac{\sigma_a^2(1 - \varrho^2)}{(1 - \varrho z)(1 - \varrho z^{-1})}, \quad |\varrho| < |z| < 1/|\varrho|. \quad (3.121)$$

Plugging the input and output spectral densities,  $S_w(z)$  and  $S_a(z)$ , respectively, into Eq. (3.118), the equation becomes

$$\begin{aligned} H(z)H(z^{-1}) &= \frac{\sigma_a^2(1 - \varrho^2)}{(1 - \varrho z)(1 - \varrho z^{-1})} \\ &= \frac{\sigma_a \sqrt{1 - \varrho^2}}{1 - \varrho z} \frac{\sigma_a \sqrt{1 - \varrho^2}}{1 - \varrho z^{-1}} \quad |\varrho| < |z| < 1/|\varrho|. \end{aligned} \quad (3.122)$$

### 3. Neuron and Retina Models



**Figure 3.19.** System function to generate the required noise sequence.

From the factorization in Eq. (3.122), it is possible to identify

$$H(z) = \frac{\sigma_a \sqrt{1 - \rho^2}}{1 - \rho z^{-1}}, \quad (3.123)$$

which corresponds to a causal, stable, and minimum-phase system [Therrien, 1992]. Calculating the inverse  $z$ -transform for this transfer function, we get the impulse response

$$h[n] = \sigma_a \sqrt{1 - \rho^2} \rho^n H[n], \quad (3.124)$$

where  $H[n]$  is the discrete Heaviside unit step function. By applying the  $z$ -transform properties to Eq. (3.123), we obtain the following difference equation for the filter:

$$a[n] = \rho a[n - 1] + \sigma_a \sqrt{1 - \rho^2} w[n]. \quad (3.125)$$

From Eq. (3.116), we have that  $\rho = e^{-T_s/\tau_a}$  in the last expressions. Figure 3.19 is a sketch of a first order linear system that generates the desired noise sequence when driven by white noise.

The estimation of the discrete neuronal function  $\hat{\rho}[n]$  is computed from the input signal of the threshold block  $h[n]$  by

$$\hat{\rho}[n] = \delta[h[n] - \theta] H[h[n] - h[n - 1]], \quad (3.126)$$

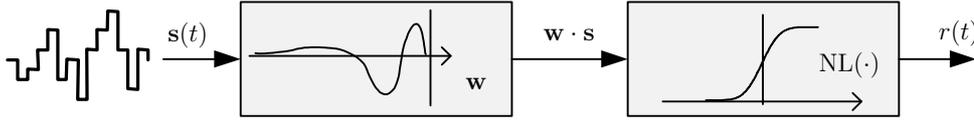
which is similar to the expression for the signal  $q[n]$  in Eq. (3.111) except for the amplitude. Therefore, we can compute the output of the threshold block from  $q[n]$  as:

$$\hat{\rho}[n] = \begin{cases} 1, & q[n] > 0 \\ 0, & q[n] = 0 \end{cases}, \quad (3.127)$$

that can be written simply as  $\hat{\rho}[n] = 1 - \delta[q[n]]$ . It should be noted from Eq. (3.111) that  $q[n] \geq 0$ .

#### 3.5.3. White Noise Model

A simple model based on a basic form of white noise analysis of the RGCs' responses to random light patterns was proposed in [Chichilnisky, 2001] and [Baccus and Meister, 2002]. The use of white-noise analysis of system model started with the seminal works of Norbert Wiener.



**Figure 3.20.** The white noise model structure.

White noise analysis has some relevant features; for example, it can generate a quantitative model for the spatial, temporal and spectral responses of the visual system neurons by taking into account its nonlinearities [Schwartz et al., 2002; Chichilnisky, 2001; Rust et al., 2004; Simoncelli et al., 2004; Baccus and Meister, 2002]. These models can be considered deterministic, since they do not include any intrinsic noise source, and the variability is introduced by the spike generator.

### Model Description

The white noise model is portrayed as a block diagram in Fig. 3.20. The first block is a linear filter with an impulse response equal to the time-reverse of vector  $\mathbf{w}$ , which is proportional to the STA if the stimulus space possesses certain characteristics. The second block is a nonlinear function that maps the generator potential signal, given by the inner vector product  $\mathbf{w} \cdot \mathbf{s} = \mathbf{w}^T \mathbf{s}$ , onto the retina firing rate. The filter impulse response is obtained by white noise analysis of the ganglion cell responses, and the nonlinear function on the second block is obtained by fitting a curve to the generator potential plotted against the firing rate. This retina model gives the neuronal firing rate as the output, assuming that it depends only on the generator signal. The generator signal is a linear combination of the visual stimulus that reaches the retina over a specific region and time period. In order to have a spike train, a spike generator must be used to convert the estimated firing rate to the neural function.

The time interval  $T$ , during which the retina is stimulated and the produced spike trains are recorded, is discretized into time bins of width  $\Delta t$ , resulting in a total recording number of time bins  $N = \lfloor T/\Delta t \rfloor$ . The number of spikes observed in the time bin  $n$  is represented by  $f_n$ , where  $1 \leq n \leq N$ . Furthermore, the number of spikes is related to the neural response function  $\rho(t)$  defined in Eq. (B.7) by

$$f_n = \int_{(n-1)\Delta t}^{n\Delta t} \rho(\tau) d\tau. \quad (3.128)$$

Each time bin  $n$  has a corresponding stimulus vector, which is represented by  $\mathbf{s}_n$ , with dimension  $k$ , whose elements are the stimulus intensities as a function of space and time, in the time bins immediately preceding the instant  $n\Delta t$ . That is,

$$\mathbf{s}_n = \begin{bmatrix} s(\mathbf{r}, n\Delta t) \\ s(\mathbf{r}, (n-1)\Delta t) \\ \vdots \\ s(\mathbf{r}, (n-k+1)\Delta t) \end{bmatrix}. \quad (3.129)$$

### 3. Neuron and Retina Models

The definition of  $\mathbf{s}_n$  in Eq. (3.129) considers spatial information, which is indicated by the displacement vector  $\mathbf{r}$  in its elements, since the stimulus can carry spatial information. For the case when the stimulus is spatially uniform, this dependency is dropped. The displacement vector  $\mathbf{r}$  spans the retinal area that influences the RGC under analysis, corresponding to their RF. In general, it is a matrix of positions for each value of the time index  $n$ .

The time duration of the stimulus vector  $\mathbf{s}_n$ , which is equal to  $k\Delta t$ , must be sufficiently long to exceed the ganglion cell memory. This corresponds to the period over which the stimulus can affect the cell response; thus, the neuron response  $f_n$  at time  $n\Delta t$  depends only on the stimulus  $\mathbf{s}_n$ .

Each stimulus vector  $\mathbf{s}_n$  can be viewed as a point in a  $k$ -dimensional stimulus space  $S$ , and it is assumed that it is drawn randomly from  $S$  with a probability distribution given by  $P(\mathbf{s})$ . This probability distribution is considered to be radially symmetric about the origin in stimulus space, and so any two stimulus vectors  $\mathbf{s}$  and  $\mathbf{s}^* \in S$  with equal vector length have equal probability of being drawn from the distribution, that is,

$$|\mathbf{s}| = |\mathbf{s}^*| \Rightarrow P(\mathbf{s}) = P(\mathbf{s}^*) . \quad (3.130)$$

The radial symmetry implies negative entries in some of the stimulus vectors, meaning that the stimulus entries specify the contrast, or deviation, from a mean intensity level. These stimulus vectors can be generated using a Gaussian white noise sequence, which corresponds to sampling a Gaussian density function with zero mean and a standard deviation equal to the desired contrast. In general, the stimulus intensity for every spatial location, time bin, and wavelength could be drawn from Gaussian noise sequences.

The modeled neural response,  $R(\mathbf{s})$ , is the average value of the spike count  $f$  in the time bin immediately following the stimulus  $\mathbf{s}$ . This can be written as

$$R(\mathbf{s}) = \langle f|\mathbf{s} \rangle = \sum_f f P(f|\mathbf{s}) , \quad (3.131)$$

which states that  $R(\mathbf{s})$  is proportional to the expected response given the stimulus. In terms of the cell firing rate  $r(t)$ , the relation is

$$R(\mathbf{s}_n) = r(n\Delta t) \cdot \Delta t . \quad (3.132)$$

The angle bracket notation in Eq. (3.131) represents the trial average across experiments for the same stimulus. The right hand equality in Eq. (3.131) comes from the definition of statistical expectation, where  $P(f|\mathbf{s})$  is the probability distribution of the number of spikes  $f$  given a certain stimulus  $\mathbf{s}$ . To be successful, the final model should predict the average number of spikes per time bin observed after the presentation of a given stimulus.

The white noise analysis estimates  $R(\mathbf{s})$ . For the model in analysis, it is assumed that  $R(\mathbf{s})$  is a static nonlinear functional of a real linear function of the stimulus:

$$R(\mathbf{s}) = \text{NL}(\mathbf{w} \cdot \mathbf{s}) , \quad (3.133)$$

where  $\mathbf{w}$  is a fixed weighting vector and  $NL(\cdot)$  is a real-valued nonlinear function of one variable. The vector  $\mathbf{w}$  weights the neuron stimulus intensities over space and time and defines the neuron stimulus selectivity. The dot product  $\mathbf{w} \cdot \mathbf{s} = \mathbf{w}^T \mathbf{s}$  is the generator signal that controls the firing rate through the nonlinear function  $NL(\cdot)$ . (See Fig. 3.20.)

Equation (3.133) shows that the model response does not depend on previous responses, but solely on the stimulus, meaning that the spikes are generated by a Poisson-like process with a rate parameter equal to the expected response  $R(\mathbf{s})$ . To obtain a spike train, a spike generator must be added at the output of the model in Fig. 3.20.

The weight vector  $\mathbf{w}$ , as in many retinal ganglion cells models, is related to the STA. For this particular setup, the STA (see Sec. B.1.3), that corresponds to the average stimulus preceding the spikes generated by the neuron, is given by

$$\mathbf{s}_{spk} = \frac{\sum_{n=1}^N \mathbf{s}_n f_n}{\sum_{n=1}^N f_n}, \quad (3.134)$$

where  $N$  denotes again the total number of time bins. Equation (3.134) represents the discrete spike triggered average  $\mathbf{s}_{spk}$ , which is now a vector of dimension  $k$  (the same dimension as  $\mathbf{s}_n$ ); the time dependency of the definition given by Eq. (B.39) is implicit.

Dividing the numerator and denominator of Eq. (3.134) by the total time duration  $T$  of the stimulus, it becomes

$$\mathbf{s}_{spk} = \frac{\frac{1}{T} \sum_{n=1}^N \mathbf{s}_n f_n}{\frac{1}{T} \sum_{n=1}^N f_n}. \quad (3.135)$$

If the response record lasts for a long time, such that  $T \rightarrow \infty$ , the denominator in Eq. (3.135) is the average firing rate  $\langle r \rangle$ , defined in Eq. (B.23). Also, as  $T \rightarrow \infty$  the numerator in Eq. (3.135) tends to the time average  $\langle \mathbf{s}f \rangle$ . Statistically, this expectation can be expressed as the sum of all stimulus response pairs weighted by the probability of observing that particular stimulus response pair:

$$\langle \mathbf{s}f \rangle = \sum_{\mathbf{s}} \sum_f \mathbf{s}f P(\mathbf{s}, f). \quad (3.136)$$

Using Bayes' rule, and by applying Eq. (3.131), the expression in Eq. (3.136) can be written as

$$\begin{aligned} \langle \mathbf{s}f \rangle &= \sum_{\mathbf{s}} \sum_f \mathbf{s}f P(\mathbf{s}) P(f|\mathbf{s}) \\ &= \sum_{\mathbf{s}} \mathbf{s} P(\mathbf{s}) \underbrace{\sum_f f P(f|\mathbf{s})}_{R(\mathbf{s})} \\ &= \sum_{\mathbf{s}} \mathbf{s} P(\mathbf{s}) R(\mathbf{s}) \end{aligned} \quad (3.137)$$

### 3. Neuron and Retina Models

Replacing the result of Eq. (3.137) in Eq. (3.135) and considering large values of  $T$ , the following expression is obtained for the STA:

$$\mathbf{s}_{spk} = \frac{1}{\langle r \rangle} \sum_{\mathbf{s}} \mathbf{s} P(\mathbf{s}) R(\mathbf{s}) . \quad (3.138)$$

The spike triggered average in Eq. (3.138) approaches a sum of stimulus vectors, where each vector is weighted by its probability of being drawn, times the average response it induces, that is normalized by the average firing rate.

The radial symmetry of the stimulus space means that there are two stimulus vectors,  $\mathbf{s}^* \in S$  and  $\mathbf{s} \in S$ , positioned symmetrically relative to the vector  $\mathbf{w}$ , that have equal probability of being drawn. This means that  $P(\mathbf{s}) = P(\mathbf{s}^*)$ . Introducing this result into Eq. (3.138), and using the equality in Eq. (3.133), we obtain

$$\begin{aligned} \mathbf{s}_{spk} &= \frac{1}{\langle r \rangle} \sum_{\mathbf{s}, \mathbf{s}^*} [\mathbf{s} P(\mathbf{s}) \text{NL}(\mathbf{w} \cdot \mathbf{s}) + \mathbf{s}^* P(\mathbf{s}^*) \text{NL}(\mathbf{w} \cdot \mathbf{s}^*)] \\ &= \frac{1}{\langle r \rangle} \sum_{\mathbf{s}, \mathbf{s}^*} (\mathbf{s} + \mathbf{s}^*) P(\mathbf{s}) \text{NL}(\mathbf{w} \cdot \mathbf{s}) , \end{aligned} \quad (3.139)$$

where the factorization in the last equality results from the fact that  $\mathbf{s}$  and  $\mathbf{s}^*$  have equal probability of being drawn and are symmetric around  $\mathbf{w}$ . This implies that  $\mathbf{w} \cdot \mathbf{s} = \mathbf{w} \cdot \mathbf{s}^*$ . The symmetry of  $\mathbf{s}$  and  $\mathbf{s}^*$  around the vector  $\mathbf{w}$  also implies that  $\mathbf{s} + \mathbf{s}^*$  is proportional to  $\mathbf{w}$ . Furthermore, since all other quantities in Eq. (3.139) are scalars, it can be concluded that the vector  $\mathbf{w}$  is proportional to the STA:

$$\mathbf{w} \propto \mathbf{s}_{spk} . \quad (3.140)$$

This reasoning leads to the conclusion that the linear part of the model in Fig. 3.20, corresponding to  $\mathbf{w}$ , is equal to the STA vector  $\mathbf{s}_{spk}$  multiplied by a gain factor.

The linear weighting vector  $\mathbf{w}$  expresses the way the neuron integrates the visual stimuli. The temporal structure of  $\mathbf{w}$  corresponds to the time-reverse of the neuron's impulse response, and the spatial structure of  $\mathbf{w}$  describes the neuron's RF. The neuron memory can also be obtained by examining the time duration of the impulse response, and the spectral response of the neuron can be characterized by analyzing the chromatic structure of  $\mathbf{w}$ .

The gain factor can be taken into account by including a scale factor to adjust the nonlinearity function. Thus, a proportionality constant equal to unity can be used in Eq. (3.140), such that  $\mathbf{w} = \mathbf{s}_{spk}$ .

Finally, the nonlinear function has to be estimated. Since  $\text{NL}(\cdot)$  is a real-valued, positive function of the generator potential  $g_n = \mathbf{w} \cdot \mathbf{s}_n$ , and the values of  $f_n$  are available, it is possible to plot the values of the generator potential  $g_n$  against the number of spikes in each bin  $f_n$ , for every  $0 < n \leq N$ , to estimate the form of  $\text{NL}(\cdot)$ . The nonlinearity can be estimated by plotting the generator potential against the number of spikes it produces and fitting the nonlinear function using a minimization error criterion. For example, the nonlinear mapping function has the expression [Chichilnisky, 2001]:

$$f(x) = r_{max} \Phi(\beta x + \gamma) , \quad (3.141)$$

where  $\Phi(x)$  represents the normal cumulative density function (**CDF**)– the integral of the standard Gaussian function [Abramowitz and Stegun, 1965] which is related to the error function,  $\text{erf}(x)$ , by:

$$\Phi(x) = \frac{1}{2} + \frac{1}{2} \text{erf}(x/\sqrt{2}). \quad (3.142)$$

The parameter  $r_{max}$  gives the maximum firing rate of the neuron,  $\beta$  is the sensitivity of the nonlinearity to the generator signal and  $\gamma$  gives the baseline firing rate, which defines the firing rate in the absence of any visual stimulation. The parameter  $\gamma$  can also be negative, meaning that the generator potential must overcome this value to fire a spike. The drawback of the normal **CDF** function, in terms of computational implementation, is that its expression,

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \quad (3.143)$$

does not have a closed mathematical form and must be calculated numerically. This imposes some constraints on its implementation in a signal processing system.

Another commonly used nonlinear mapping function is the sigmoidal function [Dayan and Abbot, 2001], with the expression

$$f(x) = \frac{r_{max}}{1 + e^{(g_{1/2}-x)/\Delta g}}, \quad (3.144)$$

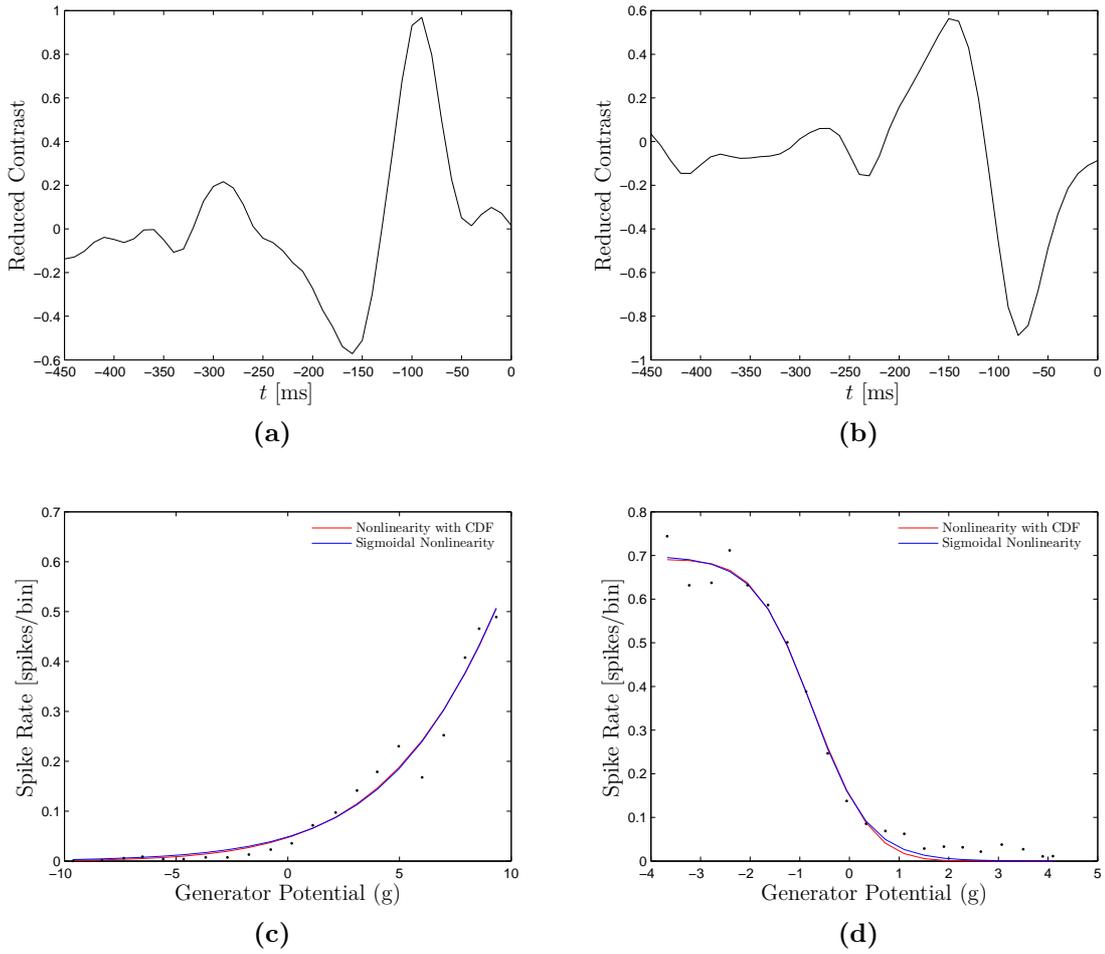
where the parameter  $r_{max}$  once again represents the neuron maximum firing rate,  $g_{1/2}$  is the generator potential value that produces a firing rate with half of its maximum value, and  $\Delta g$  controls how quickly the firing rate increases as a function of the generator potential  $g$ . For a negative  $\Delta g$ , the firing rate decreases monotonically with the generator potential. Plots for the **STA** and for the least squares adjustment of the nonlinear function for a salamander ON-type and for a rabbit OFF-type brisk **RGC**, corresponding to the expressions in Eq. (3.141) and Eq. (3.144), in Fig. 3.21 are similar in the interval of interest, which means that both nonlinearities can be used in the model without noticeable differences.

### Computational Implementation

This simple white noise model can be considered as fitting a curve to the number of fired spikes as a function of a stimulus characteristic, specifically the stimulus projection onto the **STA**, which is a common procedure in neuron modeling in theoretical neuroscience [Dayan and Abbot, 2001]. Its implementation for a computer is quite straightforward, since all expressions are already in discrete form.

First, the model parameters must be estimated (the **STA**) using experimental data; the example in Fig. 3.21 uses a salamander ON-type cell and a rabbit OFF-type brisk transient cell [Keat et al., 2001]. The vector  $\mathbf{w}$  is obtained from the spike triggered average stimulus, depicted in Fig. 3.21 for both types of cells used. The spike occurrence is considered to be located at the time instant 0, and so the negative time values are the time instants before the spike occurrence.

### 3. Neuron and Retina Models



**Figure 3.21.** White noise model characterization for a salamander ON-type cell and for a rabbit OFF-type transient brisk cell. The spike triggered average for the salamander cell (a) and the rabbit cell (b). Generator signal versus number of spikes per time bin for the salamander cell (c) and for the rabbit cell (d), with  $\Delta t = 10$  ms

To fit the nonlinear function a table with the points is built, corresponding to the generator potential  $g_n = \mathbf{w} \cdot \mathbf{s}_n$  and the number of fired spikes,  $f_n$ , for the respective time bins. The number of elements in this table depends on the time bin width. The used time bin,  $\Delta t = 10$  ms, was large when compared to the sampling period of  $T_s = 1$  ms. The number of spikes per time bin oscillates with the generator potential. The range of the generator potential values was divided into 20 equal subintervals, and the values of the generator potential and of the spikes per bin were averaged for each interval. The plots in Fig. 3.21c and Fig. 3.21d show the plots of the generator potential against the number of spikes per time bin for the analyzed RGC cells. We observe that the nonlinear function shows different behaviors from the plot in Fig. 3.21c, for the case of the salamander ON-type cell, and in Fig. 3.21d, for the case of the rabbit OFF-type transient brisk cell. These graphs also show that using the

sigmoidal nonlinearity of Eq. (3.144) leads to a curve fit similar to the one obtained with the normal CDF function within the spread of points in the plot. The sigmoidal nonlinearity is preferred when taking into account computational feasibility.

The inner product between the vector  $\mathbf{w}$  and the succeeding stimulus vectors  $\mathbf{s}_n$  can be viewed as a linearly filtered discrete stimulus signal  $s[n]$  with an impulse response equal to the time reverse of  $\mathbf{w}$  [Moon and Stirling, 2000]. This approach was followed in the model implementation, and the filter output was used as the argument to drive the nonlinear function. This, in turn, produces the firing rate  $r[n]$ . In Chap. 5, the results for this model are further described and evaluated.

## 3.6. Conclusions

This chapter presents the most relevant neural structures used in neural models and in retina models in particular. Departing from the Hodgkin-Huxley model the most relevant spiking structures are discussed and analyzed. The modeling of the synaptic input stimuli and visual stimuli was also discussed. These blocks are ubiquitous and form the basis of many retina models, and are used in the proposed retina models presented in the next chapter.

The second part of the chapter is devoted to providing an overview of the diversity of retina models by describing representative models from different classes. These models are further analyzed and assessed in Chap. 5 in the context of analysis and discussion of the application of neural activity metrics in the models' evaluation.



*The question of nature's susceptibility to mathematical description has some deep philosophical aspects, and in practical terms we have to take a more pragmatic view of models. Our acceptance of models should thus be guided by "usefulness" rather than "truth".*

Ljung, L. (1999). *System Identification: Theory for the User*. Pearson Education, New Jersey, USA

# 4

## Modeling the Retina

### 4.1. Introduction

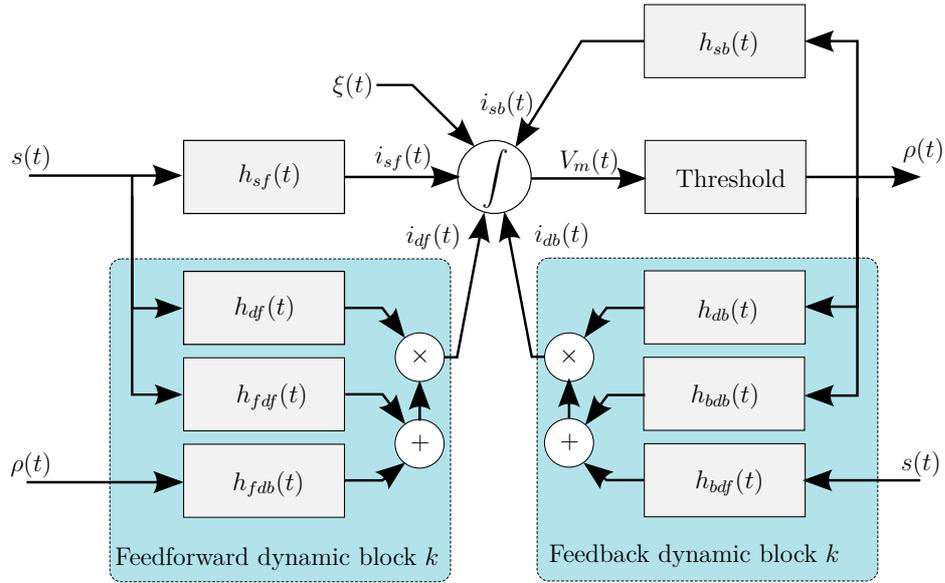
THIS chapter presents two distinct models for the retina. These models were developed based on the leaky integrate-and-fire model and on the linear-nonlinear-Poisson model, representative of two main classes of neuron models, namely the spiking retina models and firing rate retina models, respectively. The first model uses a set of static and dynamic filters and the leaky integrate-and-fire (LIF) mechanism to produce the spike train output, belonging to the spiking neuron models [Tomás et al., 2008]. The second proposed retina model is based on the white-noise analysis of the retina responses to obtain excitatory and inhibitory filters that are mapped through functional nonlinearities estimated to generate the firing rate. This model uses a Poisson firing rate mechanism to generate the spike trains, belonging to the class of firing rate models [Martins et al., 2007]. Both models take into account the temporal processing of the visual stimuli since the available retina data is only of the full-field type.

### 4.2. A Static/Dynamic Retina Model

In order to study a structural model a dynamic retina model is proposed and analyzed. This model is composed by distinct dynamic and static components so that it can be analyzed solely with its static components and with the static and the dynamic components. The variants of this retina model are implemented, tuned, and the results analyzed [Tomás et al., 2008].

The static structure is studied, where the neuron current has only two static components: one dependent on the input stimulus and the other on the spike history.

#### 4. Modeling the Retina



**Figure 4.1.** Dynamic retina model.

Afterwards, a dynamic structure is added where, besides the static components, two additional dynamic components are included, one dependent on the stimulus and the other on the spike history. Both dynamic currents are modulated by a signal which is a function of the stimulus and on the spike history. The complete structure of the model is depicted in Fig. 4.1.

The output feedback mechanism takes into account the effect that a spike emission has on the dynamics of the neuron, like the refraction behavior. The dynamic model modulates its output not only by the stimulus characteristics but also by the recent spike history. The model's parameters are initialized using spike-triggered analysis.

##### 4.2.1. Model Structure

The firing mechanism of this model is based on the **SLIF** model described in Sec. 3.3.4, where the current is integrated and a spike is fired whenever the threshold is surpassed from below. The current generation depends both on the stimulus and on the spike history. As described in Chap. 2 the horizontal and amacrine cells establish feedback paths to model the response of the retina to changes in the illuminance and contrast of the scene [Kolb, 1997; Baccus and Meister, 2002], despite these feedback mechanisms being minimal in the fovea, these adaptation processes also occur at the bipolar [Rieke, 2001], and ganglion cells [Kim and Rieke, 2001]. In a bioelectronic prosthesis the techniques used in video cameras for the illuminance and contrast adaptation mechanisms can be employed. However the spike history effects such as refractoriness and burstiness demands the inclusion of output feedback mechanisms.

The static current components in Fig. 4.1 are given by:

$$i_{sf}(t) = h_{sf}(t) * s(t), \quad (4.1a)$$

$$i_{sb}(t) = h_{sb}(t) * \rho(t), \quad (4.1b)$$

where  $h_{sf}(t)$  is the static forward filter, and  $h_{sb}(t)$  is the static backward filter. The signal  $s(t)$  is the input visual stimulus and  $\rho(t)$  is the output spike sequence.

The forward dynamic block is composed by a set of linear filters  $h_{df}^{(k)}(t)$  that filter the input stimuli, and whose amplitude is dynamically modulated by the linear function  $c_{df}^{(k)}(s(t), \rho(t))$  that depends both on stimulus and spike history. The dynamic forward current is given by:

$$i_{df}^{(k)}(t) = c_{df}^{(k)}(t)[h_{df}^{(k)}(t) * s(t)], \quad (4.2)$$

where

$$c_{df}^{(k)}(t) = h_{fdf}^{(k)}(t) * s(t) + h_{fdb}^{(k)}(t) * \rho(t). \quad (4.3)$$

In the backward dynamic block the output spike sequence is filtered by the set of linear filters  $h_{db}^{(k)}(t)$  whose amplitude is modulated by  $c_{db}^{(k)}(t)$  which depends also on the stimulus and on the spike history. The dynamic backward current is given by:

$$i_{db}^{(k)}(t) = c_{db}^{(k)}(t)[h_{db}^{(k)}(t) * \rho(t)], \quad (4.4)$$

where

$$c_{db}^{(k)}(t) = h_{bdf}^{(k)}(t) * s(t) + h_{bdb}^{(k)}(t) * \rho(t). \quad (4.5)$$

The overall input current is given by:

$$I_s(t) = i_{sf}(t) + i_{sb}(t) + \sum_k i_{df}^{(k)}(t) + \sum_k i_{db}^{(k)}(t), \quad (4.6)$$

and the membrane potential  $V_m(t)$  follows Eq. (3.62).

## 4.2.2. Model Implementation

The model is discretized with a sampling period  $T_s$ . The filters of the static components were discretized into the vectors:

$$\mathbf{h}_{sf} = [h_{sf}[1] \ h_{sf}[2] \ \cdots \ h_{sf}[M]]^T, \quad \mathbf{h}_{sb} = [h_{sb}[1] \ h_{sb}[2] \ \cdots \ h_{sb}[M]]^T, \quad (4.7)$$

where  $M$  is the filter's memory length. Each filter is described as a linear combination of basis functions, which allows to reduce the dimensionality of the model parameters and decrease the parameters overfitting to the data in the training phase.

Considering the discrete basis functions' vectors  $\mathbf{b}_k = [b_k[1] \ b_k[2] \ \cdots \ b_k[M]]^T$ ,  $k = 1, 2, \dots, B$ , organized in matrix form as:

$$\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_B], \quad (4.8)$$

#### 4. Modeling the Retina

the filters  $\mathbf{h}_x$  in Eq. (4.7), with  $x \in \{sf, sb\}$ , are synthesized by:

$$\mathbf{h}_x = \sum_{k=1}^B a_{xk} \mathbf{b}_k = \left( \mathbf{a}_x^T \mathbf{B}^T \right)^T = \mathbf{B} \mathbf{a}_x, \quad (4.9)$$

where  $\mathbf{a}_x = [a_{x1} \ a_{x2} \ \dots \ a_{xB}]^T$  are the synthesis coefficients of the filter impulse response in terms of the base functions  $\mathbf{b}_k$ . The impulse response of the filters in the dynamic components  $\mathbf{h}_x^{(k)} = [h_x^{(k)}[1] \ h_x^{(k)}[2] \ \dots \ h_x^{(k)}[M]]^T$ , with  $k = 1, 2, \dots, C$  and  $x \in \{df, db, fdf, fdb, bdf, bdb\}$ , can be organized in the matrix:  $\mathbf{H}_x = [\mathbf{h}_x^{(1)} \ \mathbf{h}_x^{(2)} \ \dots \ \mathbf{h}_x^{(C)}]$  which can be written in terms of the basis functions matrix  $\mathbf{B}$  as:

$$\mathbf{H}_x = [\mathbf{B} \mathbf{a}_x^{(1)} \ \mathbf{B} \mathbf{a}_x^{(2)} \ \dots \ \mathbf{B} \mathbf{a}_x^{(C)}] = \mathbf{B} \mathbf{A}_x, \quad (4.10)$$

where the matrix  $\mathbf{A}$  is defined by:

$$\mathbf{A}_x = [\mathbf{a}_x^{(1)} \ \mathbf{a}_x^{(2)} \ \dots \ \mathbf{a}_x^{(C)}], \quad (4.11)$$

whose columns are the synthesis coefficients of the dynamic filters.

From Eq. (4.1) the discrete static current components are given by:

$$i_{sf}[n] = (\mathbf{B} \mathbf{a}_{sf})[n] * s[n], \quad (4.12a)$$

$$i_{sb}[n] = (\mathbf{B} \mathbf{a}_{sb})[n] * \rho[n-1]. \quad (4.12b)$$

The discrete forward dynamic current component is:

$$i_{df}^{(k)}[n] = c_{df}^{(k)}[n] [h_{df}^{(k)}[n] * s[n]], \quad (4.13)$$

using the synthesis expression for the dynamic filters of Eq. (4.10) it can be written as:

$$i_{df}^{(k)}[n] = c_{df}^{(k)}[n] (\mathbf{B} \mathbf{A}_{df})_{*k}[n] * s[n]. \quad (4.14)$$

Considering Eq. (4.3) and Eq. (4.10) the discrete modulation forward coefficients are given by:

$$c_{df}^{(k)}[n] = (\mathbf{B} \mathbf{A}_{fdf})_{*k}[n] * s[n] + (\mathbf{B} \mathbf{A}_{fdb})_{*k}[n] * \rho[n-1], \quad (4.15)$$

these coefficients can be organized in a vector:  $\mathbf{c}_{df}[n] = [c_{df}^{(1)}[n] \ \dots \ c_{df}^{(C)}[n]]^T$  such that it can be written as:

$$\mathbf{c}_{df}[n] = [(\mathbf{B} \mathbf{A}_{fdf})[n] * s[n] + (\mathbf{B} \mathbf{A}_{fdb})[n] * \rho[n-1]]^T, \quad (4.16)$$

where the convolution of a matrix with a signal is made by convolving every column of the matrix with the signal, resulting in a row vector.<sup>1</sup>

Using Eq. (4.16) the dynamic forward current in Eq. (4.14) can be written as:

$$i_{df}[n] = [\mathbf{B} \mathbf{A}_{df}[n] * s[n]] \mathbf{c}_{df}^T[n]. \quad (4.17)$$

---

<sup>1</sup>  $\mathbf{A}_{n \times m}[n] * s[n] = [(\mathbf{A}_{*1}[n] * s[n]) \ \dots \ (\mathbf{A}_{*m}[n] * s[n])]^T$ .

In a similar way, the discretization of the modulation backward coefficients of Eq. (4.5) are given by:

$$\mathbf{c}_{db}[n] = [(\mathbf{BA}_{bdf})[n] * s[n] + (\mathbf{BA}_{bdb})[n] * \rho[n-1]]^T, \quad (4.18)$$

so that the discrete dynamic backward current can be written as:

$$i_{db}[n] = [\mathbf{BA}_{db}[n] * \rho[n-1]] \mathbf{c}_{db}^T[n]. \quad (4.19)$$

The following sequence is obtained by sampling the integrator filter of Eq. (3.63) at  $n = 0, 1, \dots$  for  $t = nT_s$ , with  $T_s$  as the sampling period:

$$h_{LIF}[n] = e^{-nT_s/\tau_m} \mathbf{H}(nT_s), \quad (4.20)$$

which  $z$ -transform is given by:

$$H_{LIF}(z) = \frac{1}{1 - e^{-T_s/\tau_m} z^{-1}}, \quad |z^{-1}| < e^{T_s/\tau_m}. \quad (4.21)$$

For convenience, the integrator discrete linear filter used in the model has the  $z$ -transform

$$H_{LIF}(z) = \frac{1 - \beta}{1 - \beta z^{-1}}, \quad (4.22)$$

which is a **IIR** filter with unitary **DC** gain.  $\beta = e^{-T_s/\tau_m}$  defines a pole at  $s_p (= -1/\tau_m) = \frac{1}{T_s} \log(\beta)$ . In order for the filter to remain stable and causal one must have  $\beta \in ]0, 1[$ , so that  $s_p < 0$ . By taking the inverse  $z$ -transform of Eq. (4.22), the causal discrete impulse response of the filter is found to be:

$$h_{LIF}[n] = (1 - \beta)\beta^n \mathbf{H}[n]. \quad (4.23)$$

For a spike occurring at discrete time  $n_i T_s$  where the potential is reset to  $V_{reset}$  and the noise influences only the next sample,  $(n_i + 1)T_s$ , the discrete counterpart of Eq. (3.61) can be written as:

$$\begin{aligned} V_m[n] &= h_{LIF}[n] * [V_{reset}\delta[n - n_i] + I_s[n] \mathbf{H}[n - (n_i + 1)] + \sigma_d \xi[n] \mathbf{H}[n - (n_i + 1)]] \\ &= (1 - \beta)\beta^n \mathbf{H}[n] * [V_{reset}\delta[n - n_i] + I_s[n] \mathbf{H}[n - (n_i + 1)] + \sigma_d \xi[n] \mathbf{H}[n - (n_i + 1)]] \end{aligned} \quad (4.24)$$

where  $\sigma_d$  is the discrete noise power (variance), and the discrete current amplitude takes into account the membrane resistance. The discrete impulse response of the filter, given by Eq. (4.23), is used here only for analysis purposes.

The mean trajectory of the potential and its variance can be calculated. Considering that after a spike the reset potential is drawn randomly from a probability distribution with mean  $E\{V_{reset}\} = \mu_{V_{reset}}$  and variance  $E\{[V_{reset} - \mu_{V_{reset}}]^2\} = \sigma_{V_{reset}}^2$ , and that the discrete white noise process has the mean and discrete autocorrelation function:

$$E\{\xi[n]\} = 0, \quad E\{\xi[n]\xi[n']\} = \delta[n - n'], \quad (4.25)$$

#### 4. Modeling the Retina

the expected subthreshold potential trajectory is given by:

$$\begin{aligned}
\mu_{V_m}[n] &= E \{V_m[n]\} \\
&= h_{LIF}[n] * [V_{reset}\delta[n - n_i] + I_s[n] H[n - (n_i + 1)]] \\
&= (1 - \beta)\beta^{n-n_i}\mu_{V_{reset}} + \sum_{k=n_i+1}^n \beta^{n-k} I_s[k].
\end{aligned} \tag{4.26}$$

while the variance of the membrane potential is given by:

$$\begin{aligned}
\sigma_{V_m}^2[n] &= E \left\{ [V_m[n] - \mu_{V_m}[n]]^2 \right\} = (1 - \beta)^2 \beta^{2(n-n_i)} \sigma_{V_{reset}}^2 + (1 - \beta)^2 \sigma_d^2 \sum_{k=n_i+1}^n \beta^{2(n-k)} \\
&= (1 - \beta)^2 \beta^{2(n-n_i)} \sigma_{V_{reset}}^2 + (1 - \beta)^2 \frac{1 - \beta^{2(n-n_i)}}{1 - \beta^2} \sigma_d^2.
\end{aligned} \tag{4.27}$$

#### 4.2.3. Model Tuning

The complete set of parameters of the dynamic model in Fig. 4.1 is composed by:

$$\Theta' = \{\mathbf{a}_{sf}, \mathbf{a}_{sb}, \mathbf{A}_{df}, \mathbf{A}_{db}, \mathbf{A}_{fdf}, \mathbf{A}_{fdb}, \mathbf{A}_{bdf}, \mathbf{A}_{bdb}, \beta, \sigma_d, \sigma_{V_{reset}}, V_\theta, \mu_{V_{reset}}\}. \tag{4.28}$$

This set of parameters can be reduced since some parameters depend on others. Namely,  $V_\theta$  and  $\mu_{V_{reset}}$  are directly related with the integration time for a spike generation, therefore changing their values is equivalent to change the gains of the filters in the model. Also, changing the value of  $\beta$  is equivalent to change the shape of the other filters, which have more degrees of freedom hence  $\beta$  could be kept constant. Additionally, the value of  $\sigma_{V_{reset}}$  does not greatly influence the potential trajectory since it decays exponentially, as indicated by Eq. (4.27), and can be removed from the set of parameters to tune. Therefore, the set of model's parameters to be tuned is:

$$\Theta = \{\mathbf{a}_{sf}, \mathbf{a}_{sb}, \mathbf{A}_{df}, \mathbf{A}_{db}, \mathbf{A}_{fdf}, \mathbf{A}_{fdb}, \mathbf{A}_{bdf}, \mathbf{A}_{bdb}, \sigma_d\}, \tag{4.29}$$

which is composed by the filter synthesis coefficients vectors/matrices and the discrete input noise standard deviation. The non-trainable parameters were set to the values:  $\beta = 0.9$ ,  $\sigma_{V_{reset}} = 0$ ,  $V_\theta = 1$ , and  $\mu_{V_{reset}} = 0$ .

A typical procedure used in the optimization of the parameters of neuron models is to use maximum likelihood estimation, both in spiking neuron models [Paninski et al., 2004; Mulleney and Iyengar, 2008], and Poisson neuron models [Pillow et al., 2005]. To optimize the parameters set, the likelihood of the parameters set  $\Theta$ , given the output spike sequence  $\rho[n]$  and the input stimuli  $s[n]$ , is maximized in order to the parameters:

$$\Theta_{opt} = \arg \max_{\Theta} \mathcal{L}(\Theta | \rho[n], s[n]), \tag{4.30}$$

where  $\mathcal{L}(\Theta | \rho[n], s[n])$  is the likelihood of the parameters set given the input stimuli sequence  $s[n]$  and the output discrete neural function  $\rho[n]$ , in terms of the set of

parameters in Eq. (4.29). Since the logarithm is a monotonic increasing function, the maximization of Eq. (4.30) is equivalent to maximize its logarithm:

$$\Theta_{opt} = \arg \max_{\Theta} \ln \mathcal{L}(\Theta | \rho[n], s[n]). \quad (4.31)$$

The discretization of the neural function in time bins with duration equal to the sampling period  $T_s$  gives the discrete neural function (see Sec. B.1.2):

$$\rho[n] = \sum_{i=1}^{n_{spk}} \delta[n - n_i], \quad (4.32)$$

where  $n_i$  are the time bins containing spikes, corresponding to the time instants  $n_i T_s$ , and  $n_{spk}$  is the total number of spikes on the sequence.

To compute the probability of the generated spike sequence a Bayesian approach is applied [Tomás and Sousa, 2007; Tomás, 2009]. Considering that the firing mechanism is a renewal process, so that the firing of a spike depends only on the firing of the previous spike, as the SLIF model establishes, the probability to have a spike in the time bin  $n_i$  given that a spike was fired at  $n_{i-1}$  is:

$$\begin{aligned} P(\text{spike at } n_i | \text{spike at } n_{i-1}) &= \\ &= P(\rho[n_i] = 1, \rho[n_i - 1] = 0, \dots, \rho[n_{i-1} + 1] = 0 | \rho[n_{i-1}] = 1) \\ &= P(\rho[n_i] = 1 | \rho[n_i - 1] = 0, \dots, \rho[n_{i-1} + 1] = 0, \rho[n_{i-1}] = 1) \times \\ &\quad \prod_{j=n_{i-1}+1}^{n_i-1} P(\rho[j] = 0 | \rho[j - 1] = 0, \dots, \rho[n_{i-1} + 1] = 0, \rho[n_{i-1}] = 1) \\ &= P(\rho[n_i] = 1 | \rho[n_{i-1}] = 1) \prod_{j=n_{i-1}+1}^{n_i-1} P(\rho[j] = 0 | \rho[n_{i-1}] = 1). \end{aligned} \quad (4.33)$$

The likelihood function for a spike train can be obtained from Eq. (4.33) considering every spike interval. Considering that the neuron output is a renewal process, and taking into account every spike in the output sequence, with a total of  $n_{spk}$  spikes, its likelihood is given by:

$$\mathcal{L}(\Theta | \rho[n], s[n]) = \prod_{i=1}^{n_{spk}} P(\rho[n_i] = 1 | \rho[n_{i-1}] = 1) \prod_{j=n_{i-1}+1}^{n_i-1} P(\rho[j] = 0 | \rho[n_{i-1}] = 1), \quad (4.34)$$

where it is assumed that the neuron is at the reset state for  $n = 0$ , which is equivalent to assume that the neuron fired a spike at  $n_0 = -1$ . The log-likelihood is:

$$\begin{aligned} \ln \mathcal{L}(\Theta | \rho[n], s[n]) &= l(\Theta | \rho[n], s[n]) = \sum_{i=1}^{n_{spk}} \ln P(\rho[n_i] = 1 | \rho[n_{i-1}] = 1) \\ &\quad + \sum_{i=1}^{n_{spk}} \sum_{j=n_{i-1}+1}^{n_i-1} \ln P(\rho[j] = 0 | \rho[n_{i-1}] = 1) = . \end{aligned} \quad (4.35)$$

#### 4. Modeling the Retina

Since the neuron firing is obtained from the **SLIF** model description of the neuron, the conditional probabilities  $P(\rho[n] = 0 | \rho[n_i] = 1), n > n_i$  and  $P(\rho[n] = 1 | \rho[n_i] = 1), n > n_i$ , which are related by:

$$P(\rho[n] = 1 | \rho[n_i] = 1) = 1 - P(\rho[n] = 0 | \rho[n_i] = 1), \quad n_{i+1} < n < n_i, \quad (4.36)$$

can be marginalized in relative to the membrane firing potential. The neuron fires a spike whenever the potential  $V_m[n]$  reaches the threshold  $V_\theta$  so that:

$$\begin{aligned} P(\rho[n] = 1 | V_m[n]) &= \begin{cases} 0, & V_m[n] < V_\theta \\ 1, & V_m[n] \geq V_\theta \end{cases} \\ &= \text{H}(V_m[n] - V_\theta), \end{aligned} \quad (4.37)$$

consequently:  $P(\rho[n] = 0 | V_m[n]) = 1 - P(\rho[n] = 1 | V_m[n]) = \text{H}(V_\theta - V_m[n])$ . The spike nonoccurrence probability at the time instant  $nT_s$ , given that the last spike occurred at the time instant  $n_iT_s$ , with  $n > n_i$ , in terms of the subthreshold potential is given by:

$$\begin{aligned} P(\rho[n] = 0 | \rho[n_i] = 1) &= \int_{-\infty}^{+\infty} P(\rho[n] = 0 | V_m[n]) p(V_m[n] | \rho[n_i] = 1) dV_m[n] \\ &= \int_{-\infty}^{+\infty} (1 - \text{H}(V_m[n] - V_\theta)) p(V_m[n] | \rho[n_i] = 1) dV_m[n] \\ &= \int_{-\infty}^{V_\theta} p(V_m[n] | \rho[n_i] = 1) dV_m[n], \quad n > n_i. \end{aligned} \quad (4.38)$$

In a **SLIF** model, with additive Gaussian white noise, the membrane potential follows a Gaussian probability density with mean and variance given by Eq. (4.26) and Eq. (4.27), respectively. Therefore Eq. (4.38) becomes:

$$\begin{aligned} P(\rho[n] = 0 | \rho[n_i] = 1) &= \int_{-\infty}^{V_\theta} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) dV_m[n] \\ &= \Phi(V_\theta; \mu_{V_m}[n], \sigma_{V_m}^2[n]), \quad n > n_i. \end{aligned} \quad (4.39)$$

where the mean  $\mu_{V_m}[n]$  and the variance  $\sigma_{V_m}^2[n]$  are functions of the model parameters  $\Theta$ .

The log-likelihood function in Eq. (4.35) can be maximized using a gradient ascent method or, equivalently, to the minimization of the negative log-likelihood by a gradient descent method [Boyd and Vandenberghe, 2004]. Therefore, its derivatives in relative to each component  $\theta_k$  of the parameters set  $\Theta$  must be obtained.

Applying Eq. (4.36) the derivative of Eq. (4.35) is:

$$\begin{aligned} \frac{\partial l(\Theta | \rho[n], s[n])}{\partial \theta_k} &= \sum_{i=1}^{n_{spk}} \frac{-\frac{\partial}{\partial \theta_k} P(\rho[n_i] = 0 | \rho[n_{i-1}] = 1)}{1 - P(\rho[n_i] = 0 | \rho[n_{i-1}] = 1)} \\ &+ \sum_{i=1}^{n_{spk}} \sum_{j=n_{i-1}+1}^{n_i-1} \frac{\frac{\partial}{\partial \theta_k} P(\rho[j] = 0 | \rho[n_{i-1}] = 1)}{P(\rho[j] = 0 | \rho[n_{i-1}] = 1)}. \end{aligned} \quad (4.40)$$

The derivative of the probability in Eq. (4.40) in relative to the parameters set can be obtained using Eq. (4.39):

$$\frac{\partial}{\partial \theta_k} P(\rho[n_i] = 0 | \rho[n_{i-1}] = 1) = \int_{-\infty}^{V_\theta} \frac{\partial}{\partial \theta_k} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) dV_m[n], \quad (4.41)$$

where the mean and variance parameters depend on  $\Theta$  such that

$$\begin{aligned} \frac{\partial}{\partial \theta_k} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) &= \frac{\partial}{\partial \mu_{V_m}} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \frac{\partial \mu_{V_m}}{\partial \theta_k} \\ &+ \frac{\partial}{\partial \sigma_{V_m}} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \frac{\partial \sigma_{V_m}}{\partial \theta_k} \end{aligned} \quad (4.42)$$

where

$$\begin{aligned} \frac{\partial}{\partial \mu_{V_m}} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) &= N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \left( \frac{V_m[n] - \mu_{V_m}[n]}{\sigma_{V_m}^2[n]} \right) \\ &= -\frac{\partial}{\partial V_m} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \end{aligned} \quad (4.43)$$

and

$$\begin{aligned} \frac{\partial}{\partial \sigma_{V_m}} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) &= -\frac{1}{\sigma_{V_m}^2[n]} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \\ &+ \frac{(V_m[n] - \mu_{V_m}[n])^2}{\sigma_{V_m}^3[n]} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \\ &= -\frac{\partial}{\partial V_m} \left( \frac{V_m[n] - \mu_{V_m}[n]}{\sigma_{V_m}[n]} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \right). \end{aligned} \quad (4.44)$$

Replacing Eq. (4.43) and Eq. (4.44) in Eq. (4.41) we get:

$$\begin{aligned} \frac{\partial}{\partial \theta_k} P[\rho[n_i] = 0 | \rho[n_{i-1}] = 1] &= \int_{-\infty}^{V_\theta} -\frac{\partial}{\partial V_m[n]} \left[ N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \frac{\partial \mu_{V_m}}{\partial \theta_k} \right. \\ &+ \left. \frac{V_m[n] - \mu_{V_m}[n]}{\sigma_{V_m}[n]} N(V_m[n]; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \frac{\partial \sigma_{V_m}}{\partial \theta_k} \right] dV_m[n] \\ &= -N(V_\theta; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \left[ \frac{\partial \mu_{V_m}}{\partial \theta_k} + \frac{V_\theta - \mu_{V_m}[n]}{\sigma_{V_m}[n]} \frac{\partial \sigma_{V_m}}{\partial \theta_k} \right], \end{aligned} \quad (4.45)$$

#### 4. Modeling the Retina

and

$$\frac{\partial}{\partial \theta_k} P(\rho[n_i] = 1 | \rho[n_{i-1}] = 1) = N(V_\theta; \mu_{V_m}[n], \sigma_{V_m}^2[n]) \left[ \frac{\partial \mu_{V_m}}{\partial \theta_k} + \frac{V_\theta - \mu_{V_m}[n]}{\sigma_{V_m}[n]} \frac{\partial \sigma_{V_m}}{\partial \theta_k} \right]. \quad (4.46)$$

Together Eq. (4.40) and Eq. (4.45) in conjunction with the derivatives of the mean and the variance in relative to the parameters set  $\Theta$  permits using the gradient ascent algorithm for optimization. The discrete mean of the membrane potential is given by Eq. (4.26) while the deviation from the mean is given by Eq. (4.27). The mean depends directly on the current  $I_s$  given by Eq. (4.6) where the discrete static current components are given by the expressions of Eq. (4.12a) and Eq. (4.12b), and the forward dynamic component is given by Eq. (4.17) while the backward dynamic component is computed from Eq. (4.19). The derivatives of the mean relative to each parameter vector/matrix are given by

$$\frac{d\mu_{V_m}[n]}{d\theta_k} = \left( \mathbf{H}[n - (n_i + 1)] \frac{di_x[n]}{d\theta_k} \right) * h_{LIF} \quad (4.47)$$

which have the expressions:

$$\frac{d\mu_{V_m}[n]}{d\mathbf{a}_{sf}} = \mathbf{B} * \mathbf{s}[n] \mathbf{H}[n - (n_i + 1)] * h_{LIF} \quad (4.48a)$$

$$\frac{d\mu_{V_m}[n]}{d\mathbf{a}_{sb}} = \mathbf{B} * \rho[n - 1] \mathbf{H}[n - (n_i + 1)] * h_{LIF} \quad (4.48b)$$

$$\frac{d\mu_{V_m}[n]}{d\mathbf{A}_{df}} = [\mathbf{B} * s[n] \mathbf{H}[n - (n_i + 1)]] \mathbf{c}_{df}^T * h_{LIF} \quad (4.48c)$$

$$\frac{d\mu_{V_m}[n]}{d\mathbf{A}_{db}} = [\mathbf{B} * \rho[n - 1] \mathbf{H}[n - (n_i + 1)]] \mathbf{c}_{db}^T * h_{LIF} \quad (4.48d)$$

$$\frac{d\mu_{V_m}[n]}{d\mathbf{A}_{fdf}} = \left( [\mathbf{B} * s[n] \mathbf{H}[n - (n_i + 1)]] [\mathbf{B}\mathbf{A}_{df}[n] * s[n]]^T \right) * h_{LIF} \quad (4.48e)$$

$$\frac{d\mu_{V_m}[n]}{d\mathbf{A}_{fdb}} = \left( [\mathbf{B} * \rho[n - 1] \mathbf{H}[n - (n_i + 1)]] [\mathbf{B}\mathbf{A}_{df}[n] * s[n]]^T \right) * h_{LIF} \quad (4.48f)$$

$$\frac{d\mu_{V_m}[n]}{d\mathbf{A}_{bdf}} = \left( [\mathbf{B}[n] * s[n] \mathbf{H}[n - (n_i + 1)]] [\mathbf{B}\mathbf{A}_{db}[n] * s[n]]^T \right) * h_{LIF} \quad (4.48g)$$

$$\frac{d\mu_{V_m}[n]}{d\mathbf{A}_{bdb}} = \left( [\mathbf{B}[n] * \rho[n - 1] \mathbf{H}[n - (n_i + 1)]] [\mathbf{B}\mathbf{A}_{db}[n] * \rho[n - 1]]^T \right) * h_{LIF} \quad (4.48h)$$

$$\frac{1}{\sigma_{V_m}[n]} \frac{d\sigma_{V_m}[n]}{d\sigma_d} = \frac{1}{\sigma_d}. \quad (4.48i)$$

Equation (4.48i) is the derivative of the membrane's potential variance in order to the noise power. Where, once again, the convolution of a matrix with a signal is made by convolving every column of the matrix with the signal, which results in a row vector.

The **IIR** filter with impulse response Eq. (4.23) is implemented considering the difference equation obtained by inverting the  $z$ -transform of Eq. (4.23) which leads to the difference equation:

$$y[n] = \beta y[n - 1] + (1 - \beta)x[n], \quad (4.49)$$

considering the  $y[n]$  and  $x[n]$  as the output and input filter signals.

The static filters were initialized with the **STA** while the dynamic filters were initialized using the spike triggered covariance (**STC**) kernels. By performing a **STA** analysis of the stimuli that elicited a spike, or set of spikes, the input filter waveform is obtained (see Sec. B.1.3) [de Ruyter van Steveninck and Bialek, 1988]. The **STA** is used to get a first approximation for this input linear filter [Chichilnisky, 2001; Keat et al., 2001]. The impulse response of the static forward filter  $h_{sf}$  is initialized by synthesizing the **STA** waveform using the base functions through Eq. (4.9). The static backward filter  $h_{sb}$  was initialized by performing a spike-triggered average on the spike history, which gives the most probable spike patterns that elicit new spikes.

The shape of the forward and backward dynamic filters can be initialized using spike triggered covariance analysis [Schwartz et al., 2002; Simoncelli et al., 2004]. The  $k$  components of the dynamic forward filter  $h_{df}^{(k)}$  can be initialized by performing a principal component analysis (**PCA**) of the spike triggered covariance matrix (see Sec. B.1.5), while the dynamic backward filter  $h_{db}^{(k)}$  can be initialized by performing a **PCA** of the spike train autocorrelation matrix (see Sec. B.1.4), using excitatory and/or inhibitory directions.

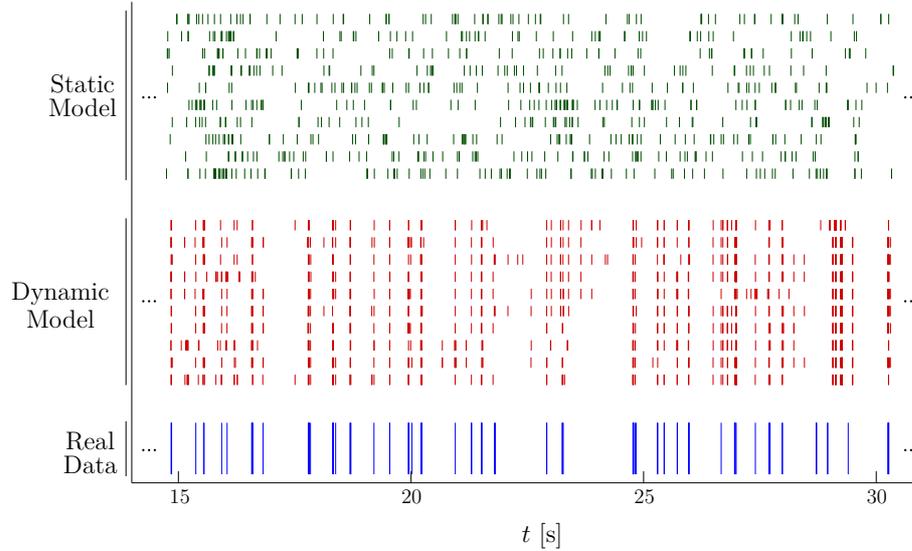
#### 4.2.4. Experimental Results

To optimize the log-likelihood of the parameters set given the neural function and the input stimuli, expressed by Eq. (4.40), where the derivatives of the probabilities are given by Eq. (4.45) and Eq. (4.46), it was used the Matlab. The maximum likelihood was obtained by optimization with the gradient descent as a function of  $\Theta$ , where the gradients in order to each model's parameter are given by Eq. (4.48a) to Eq. (4.48i). The model was discretized using a sampling period of  $T_s = 1$  ms.

The training algorithm was implemented and tested to estimate the responses of real rabbit retina ganglion cells. The data set consists of four trials of full field white noise stimulus, where each trial has a duration of  $\approx 50$ s with an average count of 6.58 spikes per second. While the stimuli values for these four trials is the same, small differences exist in the stimulation time: on average stimulation changed with a new random value at every 152 ms; the standard deviation of stimuli change was 238 ms. Section B.1.1 describes the apparatus used in the data acquisition. The visual stimuli was normalized by subtracting its mean value and by dividing by its standard deviation. Therefore, the resulting stimuli corresponding to the input  $s[n]$  applied to the model is a sequence of normally distributed random values with zero mean and unitary standard deviation.

To estimate the dynamic model, white noise analysis was initially performed where the neuron memory was set to 300 ms. Since the model includes feedforward and

#### 4. Modeling the Retina



**Figure 4.2.** Spike response of the dynamic and static model and response of rabbit RGC (vertical lines represent the timing of the elicited spikes).

feedback dynamic mechanisms, **STA** and **STC** analysis were performed using both the stimuli and the spike history. The static filters  $h_{sf}$  and  $h_{sb}$  in Fig. 4.1 were initialized with the shape of the feedforward and of the feedback **STA**, respectively. Similarly to what is done in the computation of the **STA** stimulus (see Fig. B.23), the feedback **STA** is obtained by averaging all spike responses before a spike in a window with 300 ms.

The initial shape of the dynamic filters  $h_{df}^{(k)}$  and  $h_{db}^{(k)}$  were obtained using the first 5 excitatory directions extracted using **PCA** from the **STC** covariance and from the cross-correlation between the training spike trains, respectively. Only excitatory directions were used because the **PCA** analysis on the data showed no strong suppressive directions. The modulating filters  $h_{fdf}$ ,  $h_{fdb}$ ,  $h_{bdf}$  and  $h_{bdb}$  were initially set to a small, non-zero value.

To estimate the dynamic model, many basis functions can be used to describe the linear filters. Typical examples are the Laguerre bases [Akçay and Ninness, 1999; Tomás and Sousa, 2007] or sinusoidal basis [Keat et al., 2001]. However, these basis functions are typically unable to describe delayed filters correctly and, for the used data, it considerably deteriorated the results. Therefore simple Gaussian kernels were used. Although not orthogonal, these basis functions allow a significant reduction of the number of trainable parameters, whilst achieving good results. A total of 31 kernels were employed each separated by 10 ms, and having a standard deviation of 5.6 ms.

To analyze the relevance of the dynamic blocks relative to using the static components only the performance of the complete dynamic model was compared to the static model (the number of dynamic blocks in Fig. 4.1 was set to zero). The optimization procedure used was the same for both the static and the dynamic model.

Metric		$d^{\text{time}}(q = 50 \text{ s}^{-1})$	$d^{\text{inter}}(q = 50 \text{ s}^{-1})$	NMSE
<i>Static Model</i>				
Training trial	mean	420.54	317.69	0.882
	std	10.39	7.05	
All four trials	mean	444.02	337.49	0.903
	std	12.03	6.95	
<i>Dynamic Model</i>				
Training trial	mean	251.63	242.62	0.625
	std	22.13	21.71	
All four trials	mean	325.85	281.32	0.808
	std	20.55	16.68	

**Table 4.1.** Error measures between the outputs of the trained static/dynamic model and real RGC responses.

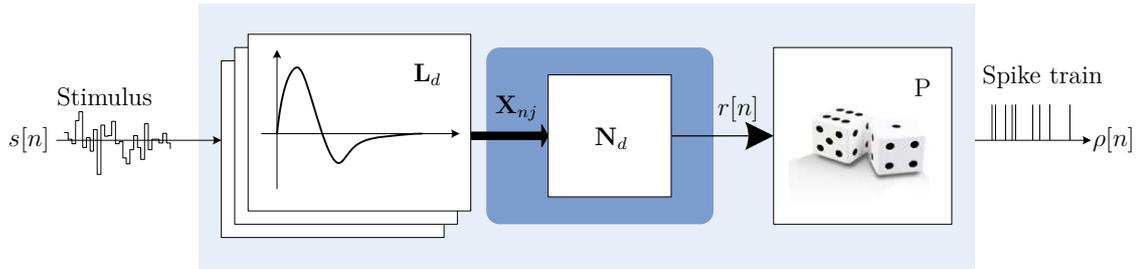
One of the ganglion cells response trial was used for training. The other three were used for comparison. Again, stimulation times are not exactly the same for all trials, which leads to slightly different neuron responses.

For evaluating the performance of the models, 30 spike response trials were produced by using both the tuned static and dynamic models. In Fig. 4.2, the first 10 response trials obtained using the training data set are presented. The figure also presents the real retina ganglion cell’s response. Examining the responses in Fig. 4.2 it can be observed that the static model is unable to model the structure of the ganglion cell’s response. On the other hand, the dynamic model is able to accurately reproduce the spike response pattern. However it does tend to fire 15% more spikes than the ganglion cell (average on the four trials).

To assess the dynamic and the static retina models two error metrics termed spike train metric and spike interval metric, that were proposed in [Victor and Purpura, 1997] are employed. These metrics are discussed more closely in Sec. 5.4. The former metric, denoted by  $d^{\text{time}}(q)$ , accounts for the cost associated with the absolute time of occurrence of the neuronal events in each spike train (see Sec. 5.4.1), while the latter metric is denoted by  $d^{\text{inter}}(q)$  and accounts for the cost of changing the time intervals between two spikes in a spike train (see Sec. 5.4.2). The parameter cost for the time metrics was set to  $q = 50 \text{ s}^{-1}$ . The NMSE firing rate metric was also applied (see Sec. 5.3.2). The firing rates were estimated for both the real and the estimated data, by convolving their peri-stimulus time histogram (PSTH) with a Gaussian window of zero mean and standard deviation of 20 ms.

The cross-evaluation between the models’ responses and the real ganglion cells responses using the error metrics are presented in Table 4.1. The table shows the results by comparing the model output with the training trial and with other 3 test trials. The error values shows that the dynamic blocks are essential to capture the

## 4. Modeling the Retina



**Figure 4.3.** Linear-Nonlinear-Poisson (LNP) model structure.

dynamics of real retina ganglion cells which cannot be described by a simple set of feedforward and feedback static filters. This can be seen by noticing that the mean values for all error metrics are much lower for the dynamic model than for the static model. The dynamic model also tends to achieve a higher degree of variability than the static model. However, this is due to the natural variability of the real **RGC** data.

### 4.3. A Multi Kernel LNP Model

Next a retina model based on the canonical **LNP** model is described and analyzed [Martins et al., 2007]. This retina model belongs to the firing rate class of models, and employs both excitatory and inhibitory kernels obtained from the **STA** and **STC** analysis of the **RGC** responses. The signals from the filtering of the stimulus with the excitatory and suppressive kernels are then mapped through a set of nonlinearities – polynomials whose order is chosen to optimize the output error – whose contributions are summed and mapped through a logistic function to produce the firing rate. A Poisson spiking mechanism is then added to generate the output spike train. The structure of this model follows the **LNP** model structure depicted in Fig. 3.13.

There are several strategies used for the estimation of the parameters of retina models, like spike-triggered analysis [Chichilnisky, 2001; Rust et al., 2004], information-theoretic approaches [Pillow and Simoncelli, 2006], and maximum-likelihood estimation [Paninski et al., 2004; Pillow et al., 2005; Mullowney and Iyengar, 2008]. Furthermore, along with this model an alternative strategy to estimate retina models is presented, relying on additive logistic regression with generalized additive models.

#### 4.3.1. Model Structure

The analysis of the responses of a spiking neuron obtained from its stimulation with a stimuli sequence drawn randomly from an ensemble gives insightful information about the stimulus subsets that elicited spikes [Rust et al., 2004]. From the analysis of this stimulus subsets, generally termed spike-triggered analysis, it is possible to characterize the neuron responses [Simoncelli et al., 2004], and was successfully applied in the development of retina models [Chichilnisky, 2001; Schwartz et al., 2002].

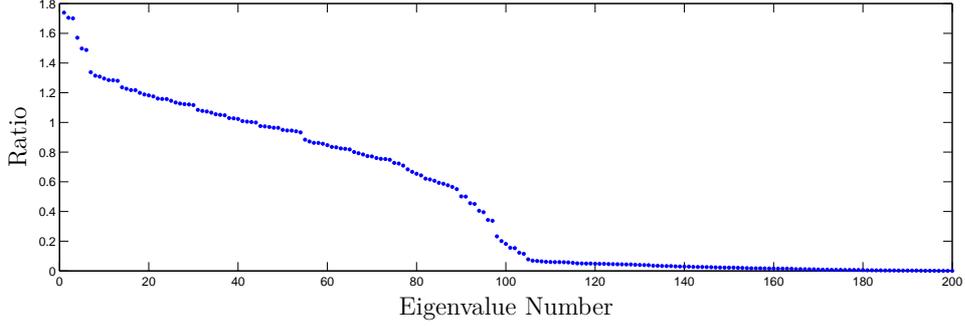
For example, by averaging the subset of stimuli that elicited a spike one obtains the **STA**, that is equal to the correlation of the firing rate with the input stimuli (see Sec. B.1.3), which can be interpreted as the most probable stimulus that elicit a spike. The **STA** provides an estimate for the first linear term in a polynomial series expansion of the neuron response function. If the system is linear the **STA** provides a complete characterization of the neuron [Rieke et al., 1997]. A simple but effective retina model that uses the **STA** as the unique input linear filter in a **LNP** model was described in Sec. 3.5.3 from [Chichilnisky, 2001]. A natural extension of the **STA** is to analyze the **STC** matrix to disclose how and what are the stimuli variations preferred by the neuron to generate a spike(s). The definition and computation of the **STC** matrix is presented in Sec. B.1.5. The analysis of the **STC** matrix provides a set of linear filters that describe better the subspace of stimuli that induces the neuron to fire, rather than using the **STA** alone. In this model the contributions of the linear filters resulting from the analysis of the **STC** matrix are used in conjunction with the **STA** to model the retina.

This model follows the general structure of a **LNP** model depicted in Fig. 3.13. However, the linear stage is extended to include several other characteristics of the neuron response so that it becomes multidimensional, with dimension  $d$ , as depicted in Fig. 4.3. Consequently, the nonlinear block is also multidimensional and combines the filters' outputs, which are individually processed by its respective nonlinear functional, and whose sum is fed to a logistic function to produce the firing rate that enters the Poisson spike generator. In the model's structure depicted in Fig. 4.3,  $\mathbf{X}_n$  represents the vector of values at time instant  $n$  resulting from the convolution of the stimulus vector  $\mathbf{s}_n$ , defined by Eq. (3.129), with the set of linear filters  $\mathbf{L}_d$ . The components of the vector  $\mathbf{X}_n$  are then nonlinearly combined in the nonlinear block to produce the firing rate  $r_n$ , which is supplied to a Poisson spike generator.

The principal component analysis (**PCA**) of the **STC** matrix provides the directions in the stimulus space along which the variance of the spike triggered ensemble differs from that of the raw stimulus ensemble. The increase of the variance in a given direction relatively to the raw stimuli indicates a stimulus dimension that is excitatory, while a decrease in variance relatively to the raw stimuli space indicates an inhibitory direction. These directions can be obtained by the eigendecomposition of the **STC** matrix in terms of eigenvectors/eigenvalues pairs by performing a **PCA** analysis. The excitatory directions correspond to the eigenvectors along which the eigenvalues are bigger than the respective values along the same eigenvectors for the raw stimulus, while the inhibitory or suppressive directions are given by the eigenvectors whose eigenvalues are smaller than the ones for the same eigenvectors for the raw stimulus [Rust et al., 2004]. In case the stimuli are driven from a Gaussian distribution with a covariance matrix equal to the identity matrix then the raw stimulus eigenvalues are all equal to unity, so that the excitatory **STC** matrix eigenvalues are bigger than one, and the suppressive directions eigenvalues are smaller than one while the remaining eigenvalues take values around one [Simoncelli et al., 2004].

The **STC** matrix,  $\mathbf{C}_{spk}$ , defined by Eq. (B.54) in vectorial form, is a  $M \times M$  symmetric positive definite matrix, where  $M$  is the neuron memory in terms of time

#### 4. Modeling the Retina



**Figure 4.4.** Eigenvalues ratio.

bins (the neuron time memory is equal to:  $M\Delta t$ ). Therefore, after performing an eigenanalysis the **STC** matrix can be decomposed as:

$$\mathbf{\Phi}_{spk}\mathbf{\Lambda}_{spk} = \mathbf{C}_{spk}\mathbf{\Phi}_{spk}, \quad (4.50)$$

where  $\mathbf{\Lambda}_{spk}$  is a diagonal matrix containing the eigenvalues  $\{\lambda_k\}$  of  $\mathbf{C}_{spk}$ , which are all real and positive, and  $\mathbf{\Phi}_{spk}$  is an orthogonal matrix whose columns are the eigenvectors  $\{\phi_k\}$  of  $\mathbf{C}_{spk}$ . This corresponds to the set of eigenvalues equations:

$$\mathbf{C}_{spk}\phi_k = \lambda_k\phi_k, \quad k = 1, \dots, M \quad (4.51)$$

where  $(\lambda_k, \phi_k), k = 1, \dots, M$  are the eigenvalue/eigenvector pairs, and  $M$  is the neuron memory. The set of distinct vectors  $\{\phi_k\}_{k=1, \dots, M}$  forms an orthogonal basis along which the **STC** can be expressed. To find the excitatory and suppressive stimulus directions the eigenvalues between the raw stimulus covariance matrix,  $\mathbf{C}_{stim}$ , and the **STC** matrix should be compared. Since the raw stimulus covariance matrix may not be equal to the identity matrix its projection along the **STC** eigenvectors is given by the diagonal matrix:

$$\mathbf{\Lambda}_{stim} = \mathbf{\Phi}_{spk}^T \mathbf{C}_{stim} \mathbf{\Phi}_{spk}. \quad (4.52)$$

The excitatory directions are given by the diagonal elements of the diagonal matrix  $\mathbf{\Lambda}_{rat}$  given by:

$$\mathbf{\Lambda}_{rat} = \mathbf{\Lambda}_{spk}\mathbf{\Lambda}_{stim}^{-1}, \quad (4.53)$$

which are bigger than one, while the suppressive or inhibitory directions are given by the directions whose respective values are smaller than one [Schwartz et al., 2002]. Figure 4.4 displays the eigenvalues ratio between the **STC** and the raw stimulus covariance matrix. The excitatory directions, for which the eigenvalues are greater than one, are evident while there are many directions along which the covariance matrix component's are zero. The data used was the response of a salamander type-ON **RGC** to Gaussian white noise stimulus displayed in Fig. B.17.

This model gives the firing rate, that is related to the firing probability as expressed by Eq. (B.38), by a generalized additive model, where the predictors are the stimulus projections along the **STA** and along the excitatory and suppressive directions

extracted from the **STC** matrix. Additive models are used in statistics for modeling the dependence of a random variable on various variables, or predictors [Hastie et al., 2009].

In order to apply the logistic regression the spike train can be interpreted as a realization of a binomial random variable that can have only two distinct values. The discrete spike sequence at the time bin  $n$  takes the value  $\rho[n] = 1$  for a spike occurrence, and  $\rho[n] = 0$  otherwise. Statistically speaking, we want to model the probability of a spike occurring,  $P(\rho[n] = 1|\mathbf{X}_n)$ , where  $\mathbf{X}_n = [x_{n1} \ x_{n2} \ \dots \ x_{nd}]^T$ , where each component  $x_{nj}$ ,  $j = 1, \dots, d$ , is the stimulus projections along each one of the  $d$  linear components, corresponding to excitatory and suppressive directions, such that in vectorial terms we have:

$$\mathbf{X}_n = \mathbf{L}_d^T \mathbf{s}_n, \quad (4.54)$$

where the columns of the matrix  $\mathbf{L}_d$  are the kernels' vectors.

The logistic regression model is particularly suited to model the posterior probabilities of classes of events. The logistic regression ensures that the probabilities sum to one and remain bounded in the interval  $[0, 1]$ . Additionally, it sets constraints on the distribution of the independent variables, that do not have to be normally distributed, linearly related, or have equal variance [Hastie et al., 2009]. In the logistic regression the relationship between the predictor variables and the response is not linear, instead it uses the logit transformation. While the linear logistic model assumes that the logs-odds are linear, the general additive logistic regression replaces each linear term by a more general term. The general additive logistic regression model for the retina has the form:

$$\text{logit} (P(\rho[n] = 1|\mathbf{X}_n)) = \alpha + \sum_{j=1}^d f_j(n_{nj}), \quad (4.55)$$

where  $d$  is the number of predictor variables used in the regression, specified in terms of the log-odds or logit transformation. Contrary to the linear regression model, in the generalized additive model the functions  $f_j$  are general, possible nonlinear, smooth functions, like basis functions, splines, or regular polynomials. The logit function is defined by:

$$\text{logit}(P(\rho[n] = 1|\mathbf{X}_n)) = \ln \left( \frac{P(\rho[n] = 1|\mathbf{X}_n)}{P(\rho[n] = 0|\mathbf{X}_n)} \right) = \ln \left( \frac{P(\rho[n] = 1|\mathbf{X}_n)}{1 - P(\rho[n] = 1|\mathbf{X}_n)} \right). \quad (4.56)$$

By inverting the logit transformation we get the logistic or sigmoid function, such that the probability of finding a spike at the time bin  $n$  is given by:

$$P(\rho[n] = 1|\mathbf{X}_n) = \frac{1}{1 + \exp \left( - \left[ \alpha + \sum_{j=1}^d f_j(x_{nj}) \right] \right)}, \quad (4.57)$$

where the number of excitatory and suppressive kernels is  $d$ . We should note that Eq. (4.57) has an expression for the firing rate similar to the one given by the escape

#### 4. Modeling the Retina

---

**Algorithm 4.1** Local scoring algorithm for the generalized additive logistic regression model, `GALRlocalScor`.

---

```

1:  $N \leftarrow$  number of time bins
2:  $d \leftarrow$  number of kernels/functions
3:  $x_{ij}, i = 1, \dots, N; j = 1, \dots, d \leftarrow$  predictor variables
4:  $\rho_i, i = 1, \dots, N \leftarrow$  spike train
5:  $\hat{\alpha} \leftarrow \ln\left(\frac{r}{1-r}\right)$   $\{r = \text{spike-count rate}\}$ 
    $r = \frac{1}{N} \sum_{i=1}^N \rho_i$ 
6:  $\hat{f}_j^1 \leftarrow 0, j = 1, \dots, d$   $\{\hat{f}^1 = \{f_1^1, f_2^1, \dots, f_d^1\}\}$ 
7: repeat
8:    $\hat{f}_j^0 \leftarrow \hat{f}_j^1$ 
9:   for  $i = 1$  to  $N$  do
10:     $\hat{\eta}_i = \hat{\alpha} + \sum_{j=1}^d \hat{f}_j^0(x_{ij})$ 
11:     $\hat{r}_i = 1 / [1 + \exp(-\hat{\eta}_i)]$ 
12:     $z_i = \hat{\eta}_i + \frac{(\rho_i - \hat{r}_i)}{\hat{r}_i(1 - \hat{r}_i)}$   $\{\text{working target variable}\}$ 
13:     $w_i = \hat{r}_i(1 - \hat{r}_i)$   $\{\text{fitting weights}\}$ 
14:   end for
15:    $\hat{\alpha}, \{\hat{f}_j^1\}_{j=1, \dots, d} \leftarrow \text{GAMbackfitting}(x_{ij}, z_i, w_i)$ 
16: until  $\Delta(\hat{f}^0, \hat{f}^1) = \frac{\sum_{j=1}^d \|\hat{f}_j^1 - \hat{f}_j^0\|}{\sum_{j=1}^d \|\hat{f}_j^0\|} > \varepsilon$   $\{\text{where } \|f_j\|^2 = \sum_{i=1}^N f_j^2(x_{ij})\}$ 
17: return  $\hat{\alpha}, \{\hat{f}_j^1\}_{j=1, \dots, d}$ 

```

---

neuron models discussed in Sec. 3.3.4. The models parameters:  $\alpha$  and the functions  $f_j(\cdot)$ , need to be optimized, and this can be done relying on the tools used on the additive logistic regression.

#### 4.3.2. Model Tuning

The spike train can be viewed as a sequence of events belonging to two distinct classes. The discrete neural function,  $\rho[n]$ , can take one of two distinct values at each time bin:  $\rho[n] = 1$  if a spike is fired or  $\rho[n] = 0$  otherwise. The probability of a spike occurrence at the time bin  $n$  is given by:

$$P(\rho[n] = 1 | \mathbf{X}_n) = r[n] \Delta t, \quad (4.58)$$

where  $r[n]$  is the discretized firing rate, and  $\Delta t$  the width of the time bins (that is normally equal to the sampling period  $T_s$ ). Instead of a direct dependence on the input stimulus, Eq. (4.57) expresses the probability of firing a spike as a function of the projection of the stimulus  $\mathbf{s}_n$  along the kernels  $\mathbf{L}_d$ , which are obtained from **STA**

---

**Algorithm 4.2** Backfitting algorithm for general logistic regression additive models, **GAMbackfitting**.

---

```

1:  $x_{ij}, i = 1, \dots, N; j = 1, \dots, d \leftarrow$  predictor variables
   { $d$  – number of predictor variables/regression functions}
2:  $z_i, i = 1, \dots, N; \leftarrow$  target variables
3:  $w_i, i = 1, \dots, N; \leftarrow$  fitting weights
4:  $\hat{\alpha} \leftarrow \frac{1}{N} \sum_{i=1}^N z_i$ 
5:  $\hat{f}_j \leftarrow 0, j = 1, \dots, d$ 
6: repeat
7:   for  $j = 1$  to  $d$  do {For all functions  $f_j$ }
8:      $\hat{f}_j \leftarrow S_j \left[ \{x_{ij}; z_i - \hat{\alpha} - \sum_{\substack{k=1 \\ k \neq j}}^d \hat{f}_k(x_{ik}); w_i\}_{i=1, \dots, N} \right]$  { $S_j$  – weighting fitting operator}
9:   end for
10: until  $\frac{1}{N} \sum_{i=1}^N \left( z_i - \hat{\alpha} - \sum_{j=1}^d \hat{f}_j(x_{ij}) \right)^2$  fails to decrease
11: return  $\hat{\alpha}, \{\hat{f}_j\}_{j=1, \dots, d}$ 

```

---

and from the **PCA** analysis of the **STC** matrix. Employing Eq. (4.57) the firing rate of Fig. 4.3, at time instant  $n$ , is given by:

$$r[n] = \frac{1}{1 + \exp \left( - \left[ \alpha + \sum_{j=1}^d f_j(X_{nj}) \right] \right)}, \quad (4.59)$$

using  $\Delta t$  as the time unit. Therefore, the generalized additive logistic regression model is:

$$\ln \left( \frac{r[n]}{1 - r[n]} \right) = \alpha + \sum_{j=1}^d f_j(X_{nj}), \quad (4.60)$$

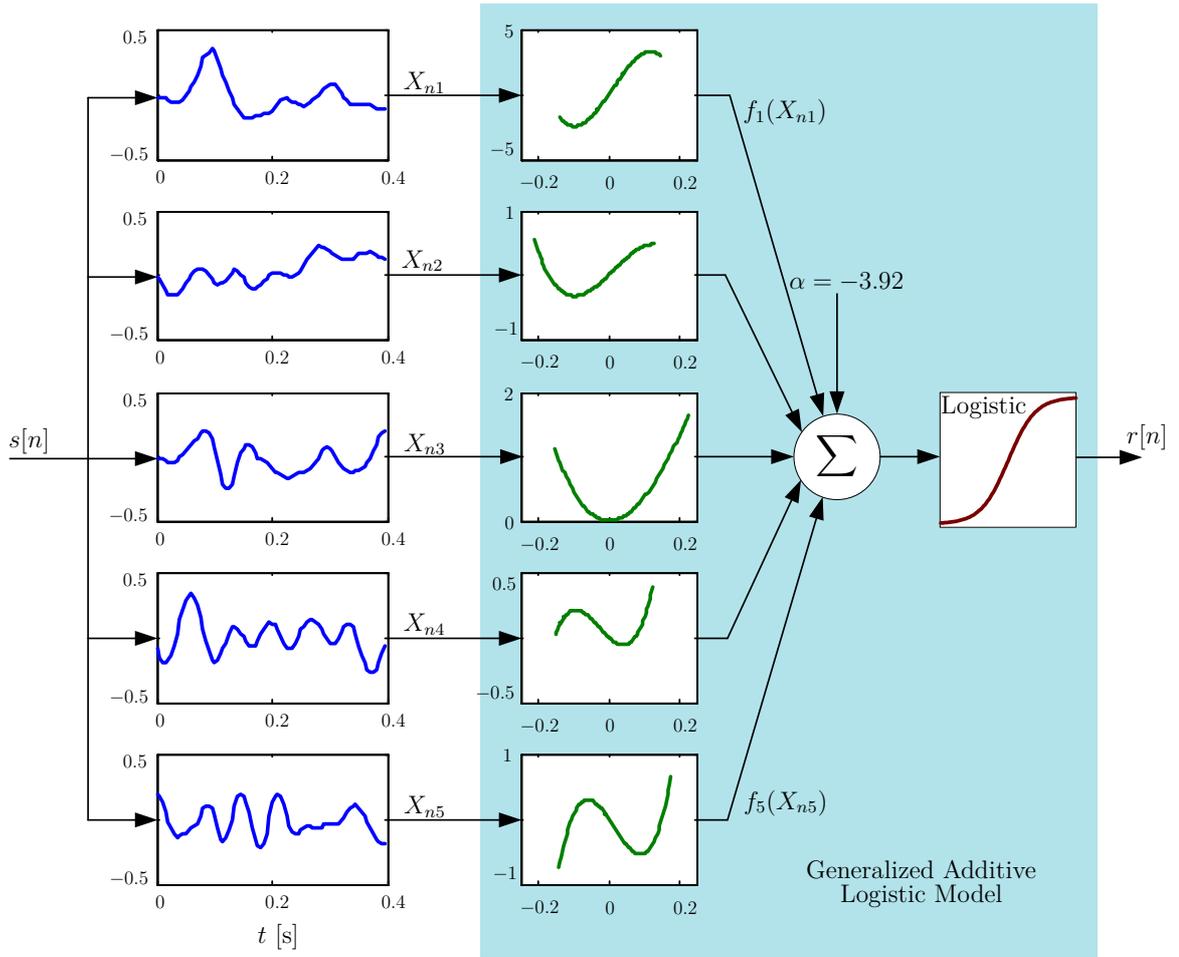
where the constant term  $\alpha$ , and the chosen functions  $f_j$  need to be adjusted given the stimulus projections  $\mathbf{X}_n$ , and the output firing rate  $r[n]$ .

To tune the model the combination of a local scoring algorithm for additive logistic regression models with the backfitting algorithm for additive models is applied [Hastie and Tibshirani, 1990]. The local scoring algorithm starts by computing an initial estimate for  $r[n]$  obtained by initializing the functions to zero,  $\hat{f}_j = 0, \forall j$  and approximating  $\hat{\alpha}$  with an initial estimate for the probability of a spike occurrence, such that  $r[n] = n_{spk}/N$ , where  $n_{spk}$  is the total number of spikes and  $N$  the number of time bins, such that  $N = T/\Delta t$ . So  $r[n] = r\Delta t$ , where  $r$  is the mean spike-count rate defined by Eq. (B.15). Therefore, the initial estimate for  $\hat{\alpha}$  is:

$$\hat{\alpha} = \ln \left( \frac{r}{1 - r} \right), \quad (4.61)$$

using  $\Delta t$  as the time unit. In the local scoring algorithm for logistic regression the

#### 4. Modeling the Retina



**Figure 4.5.** Additive logistic retina model. The linear filters components are depicted along with the estimated nonlinear functions.

model is fitted to the modified regression target variable given by:

$$z[n] = \hat{\alpha} + \sum_{i=1}^d \hat{f}_j(X_j[n]) + \frac{(\rho[n] - r[n])}{r[n](1 - r[n])}, \quad \text{for } n = 1, \dots, N, \quad (4.62)$$

with the regression weights  $w[n] = r[n](1 - r[n])$ . The steps of the local scoring algorithm are presented in Algorithm 4.1. The functions  $f_j$  are adjusted to  $z[n]$  with weights  $w[n]$  using the weighted backfitting algorithm for additive models. The weighted backfitting algorithm used in the local scoring algorithm is a general algorithm that can fit an additive model using any regression-type fitting mechanisms that supports weighting. Algorithm 4.2 describes the backfitting algorithm for fitting the generalized additive model.

The nonlinear functionals used in Eq. (4.55) are of the polynomial type with the

expression:

$$f_j(x) = \sum_{c=1}^Q \beta_c x^c, \quad (4.63)$$

where the coefficients  $\beta$  of the polynomial with degree  $Q$  are adjusted using a weighted least squares polynomial fit in step 8 of the backfitting algorithm in Algorithm 4.2. The zero order coefficients of each polynomial are imposed to be zero since the term  $\alpha$  in Eq. (4.60) takes the constant value into account. The retina model obtained with this technique is displayed in Fig. 4.5.

An inhomogeneous Poisson process is used for the spike generation block that assumes that the firing probability depends on the stimulus, and takes into account the spike history by introducing an absolute and a relative refractory period, whose values were previously calculated from the RGC responses following [Berry II and Meister, 1998].

### 4.3.3. Experimental Results

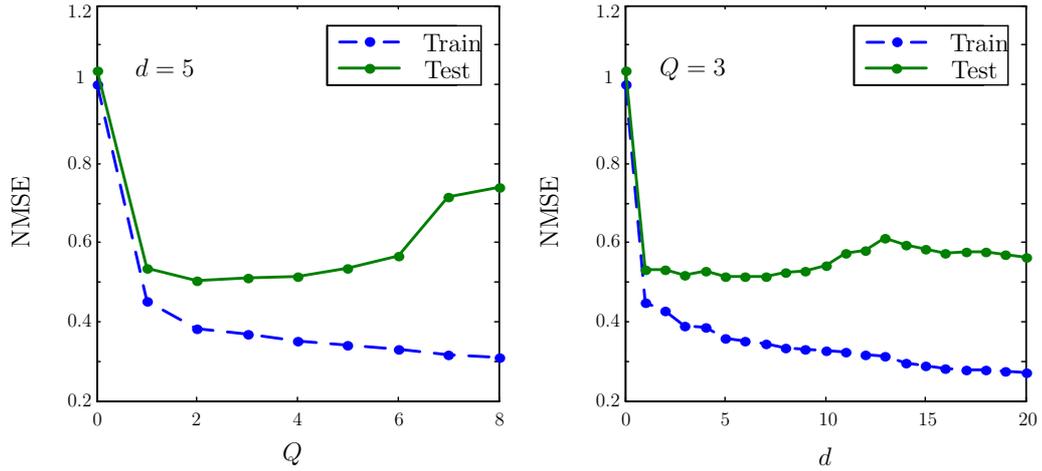
This model was implemented with a sampling period (and time bin width  $\Delta t$ ) equal to  $T_s = 1$  ms. The model was fitted and tested with the experimental data from [Keat et al., 2001] depicted in Fig. B.17, that consists of 12 trials of a salamander type-ON RGC responses to full field white noise stimulation. The trials have a duration of 10s and an average count of 8.34 spikes/s. The data was split in time into two groups, one for training (60%) and the other for testing (40%) purposes. From the STA analysis of the data it was observed that the neuron memory extends to a maximum of 400 ms, that was used as the time extension of the filters in the linear block.

To assess the results of the model the spike time metric, from [Victor and Purpura, 1997], and described in Sec. 5.4.1, that accounts for the cost associated with the absolute occurrence time of each spike is used. Additionally, to evaluate the firing rate estimation by the model the NMSE, described in Sec. 5.3.2, was used.

The model performance was evaluated while maintaining the number of linear components  $d$  in Eq. (4.55) fixed to  $d = 5$  and observing the values for the NMSE for different polynomial orders  $Q$  in Eq. (4.63). From the graphic in Fig. 4.6 it was found that the best value for the polynomials' order is  $Q = 3$  since for higher orders the model becomes overfitted. From the second graphic in Fig. 4.6 the minimum value for the NMSE is obtained for the test data with  $d = 5$ . For bigger values of  $d$  the model become overfitted to the train data. For  $d = 1$  the linear block contains only the STA filter kernel, and from the graphics Fig. 4.6 it becomes evident the importance of the use of additional excitatory/suppressive kernels.

From the results displayed in Fig. 4.6 the values for  $Q$  and  $d$ , were chosen to be  $Q = 3$  and  $d = 5$ . Figure 4.7 displays the firing rate obtained from the real data and the firing rate estimated by the model for the two time segments of the train and test data.

#### 4. Modeling the Retina



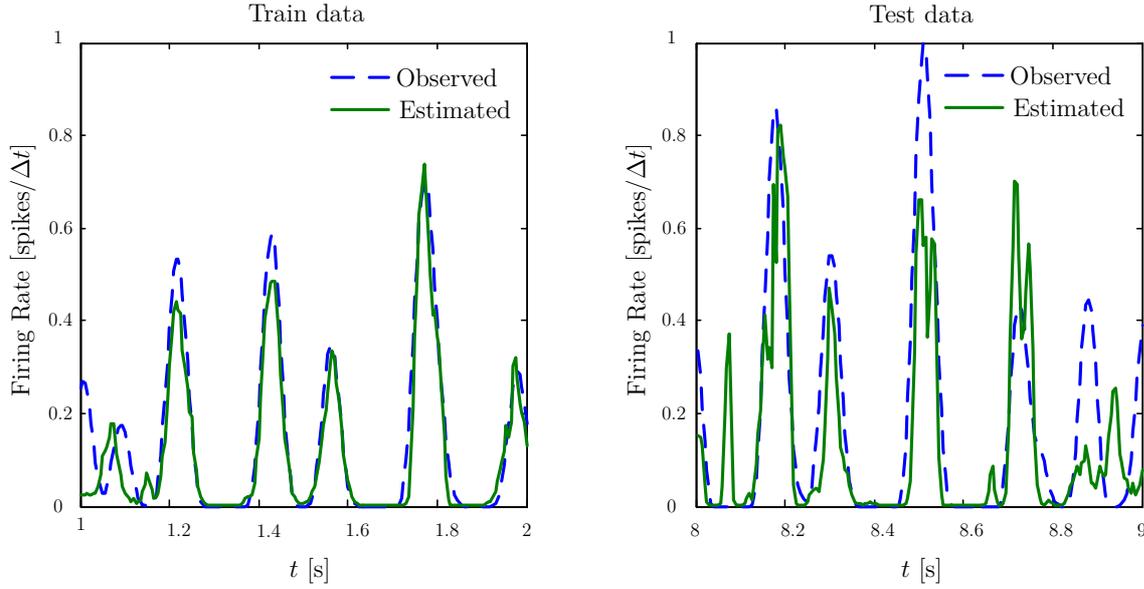
**Figure 4.6.** NMSE values for the train and test data sets with different values for the order polynomial  $Q$  and number of linear components  $d$ .

The Fig. 4.5 displays the additive logistic model with the linear kernels waveforms and the respective estimated nonlinear third order polynomials ( $Q = 3$  and  $d = 5$ ). The nonlinear functionals provide information about the contribution of each linear kernel giving some insight about the retina behavior when stimulated. From the nonlinearities waveforms it is possible to identify immediately if the contribution of each kernel to the retina response is excitatory or inhibitory.

Applying the firing rate estimate to a Poisson spike generator a set of spike trains are generated. Figure 4.8 displays a set of spike trains generated by the model and the spikes produced by the RGC. The spike time metric with  $q = 50\text{s}^{-1}$  was used to compare the the real data spike trains with the ones produced by the model. Table 4.2 displays the results obtained with the NMSE and the spike time error for the train and test data. The results for the NMSE shows that there is still some overfitting of the training data relative to the test data. However, the results are comparable with the ones obtained with other models, namely the dynamic model discussed in Sec. 4.2. The spike trains are very regular in time for this data so the spike time metric give small error both for the train and test data when compared to the dynamic model.

		Train Data	Test Data
NMSE		0.389	0.561
$d_{q=50\text{s}^{-1}}^{\text{time}}$	Observed Inter-trial	37.49	36.55
	Estimated Inter-trial	68.48	71.96
	Cross-trial	75.32	80.65

**Table 4.2.** Error measures for the generalized additive retina model.



**Figure 4.7.** Firing rate comparison between the observed and estimated with the multi kernel model for the train and test data.

## 4.4. Poisson Based Firing Rate Models versus Stochastic Integrate-and-Fire Models

The retina models based on the Poisson firing mechanism, like the one described in Sec. 3.5.3 and the SLIF model, described in Sec. 3.3.4, have a straight relationship and can even be considered identical under certain conditions [Tomás et al., 2008].

The membrane potential of the SLIF is given by Eq. (3.61). Due to the noise the value of the membrane potential cannot be exactly determined for  $t > t_i$ , the most that can be done is to calculate its probability density:  $p(V_m(t), t)$ . This probability density can be obtained as the solution of the partial differential equation:

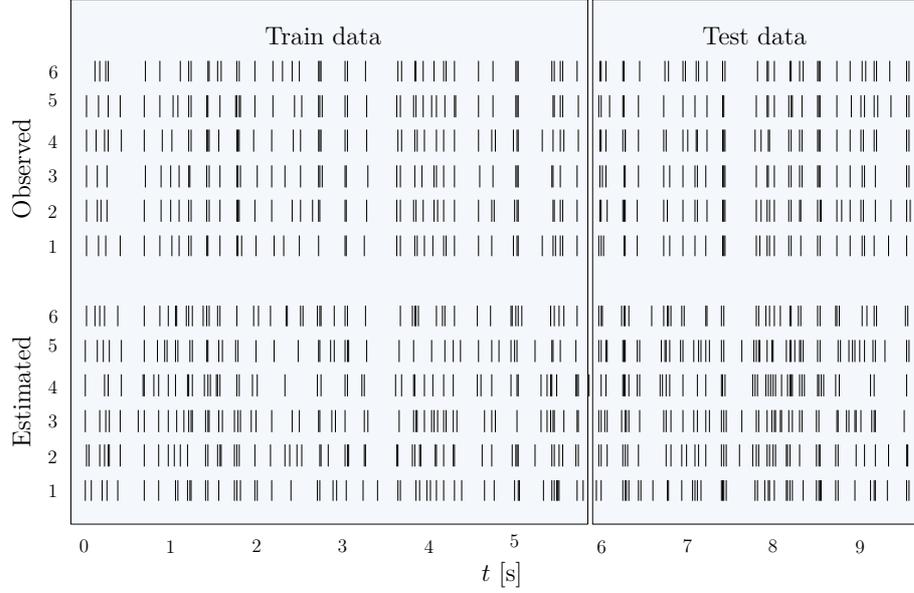
$$\tau_m \frac{\partial}{\partial t} p(V_m(t), t) = -\frac{\partial}{\partial t} [-V_m(t) + R_m I_s(t)] p(V_m(t), t) + \frac{1}{2} \sigma_\xi^2(t) \frac{\partial^2}{\partial t^2} p(V_m(t), t) \quad (4.64)$$

where  $\sigma_\xi^2(t)$  is the time dependent noise variance, with the firing threshold as a boundary condition  $p(V_\theta, t) = 0, \forall t$ . Equation (4.64) is a Fokker-Planck equation that describes the time evolution of the probability density [Gerstner and Kistler, 2002]. The solution of this Fokker-Planck equation with the initial condition  $p(V_m(t), t_i) = \delta(V_m(t_i) - V_\theta)$  gives a Gaussian probability density with mean  $\mu_{V_m}(t)$  and variance  $\sigma_{V_m}^2(t)$ :

$$p(V_m(t), t) = \frac{1}{\sqrt{2\pi\sigma_{V_m}^2(t)}} \exp\left(-\frac{[V_m(t) - \mu_{V_m}(t)]^2}{2\sigma_{V_m}^2(t)}\right), \quad (4.65)$$

where  $\mu_{V_m}(t)$  and  $\sigma_{V_m}^2(t)$  are given by Eq. (3.64) and Eq. (3.65) respectively. The probability density of Eq. (4.65) varies continuously with time and with the value of the membrane potential along its course.

#### 4. Modeling the Retina



**Figure 4.8.** Spike response of the additive logistic retina model.

The probability that the neuron fires a spike at the time instant  $t$  is given by:

$$\begin{aligned}
 P(\text{spike at } t | V_m(t), t) &= P(V_m(t) > V_\theta, t) \\
 &= 1 - \int_{-\infty}^{V_\theta} p(V_m(t), t) dV_m(t) \\
 &= 1 - \Phi(V_\theta | \mu_{V_m}(t), \sigma_{V_m}^2(t)) \\
 &= 1 - \Phi(V_\theta - \mu_{V_m}(t) | 0, \sigma_{V_m}^2(t)) \\
 &= 1 - \Phi(-\mu_{V_m}(t) | V_\theta, \sigma_{V_m}^2(t)) \\
 &= \Phi(\mu_{V_m}(t) | V_\theta, \sigma_{V_m}^2(t)).
 \end{aligned} \tag{4.66}$$

The function  $\Phi(x, \mu, \sigma^2)$  is the normal cumulative density function with mean  $\mu$  and variance  $\sigma^2$  defined by:

$$\begin{aligned}
 \Phi(x | \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-\frac{t^2}{2}} dt \\
 &= \Phi\left(\frac{x-\mu}{\sigma}\right),
 \end{aligned} \tag{4.67}$$

where the last equality shows the relation with the standard normal **CDF**. The last expression in Eq. (4.66) is identical to the generator potential mapping of Eq. (3.141) made in the white noise model to obtain the firing rate from the generator potential,

corresponding to a sigmoid. This expression is also similar to the logistic function of Eq. (4.59), which gives the firing rate as a function of the predictor variables in the generalized additive logistic regression model of the retina. For the **SLIF** model the slope of the sigmoid is controlled by the variance  $\sigma_{V_m}^2(t)$  and the mean value  $V_\theta$  controls the positioning of the sigmoid.

Furthermore, if we allow the reset potential to be set randomly with a given variance we can control the time evolution shape of the sigmoid. For a reset potential with variance  $\sigma_{V_{reset}}$  the expression for the variance of the membrane potential trajectory of the **SLIF** model, described in Sec. 3.3.4, given by Eq. (3.65), takes the form:

$$\sigma_{V_m}^2(t) = \frac{\sigma_\xi^2}{2} + \left( \sigma_{V_{reset}} - \frac{\sigma_\xi^2}{2} \right) e^{-2\frac{t-t_i}{\tau_m}}. \quad (4.68)$$

Although the value of the deviation converges to the limit case of Eq. (3.66), three cases can be distinguished for the initial shape of the sigmoid taking into account the relative values for the variances of the reset potential and input noise: for *i*)  $\sigma_{V_{reset}}^2 = \sigma_\xi^2/2$  the shape of the sigmoid is kept along the integration period; while if *ii*)  $\sigma_{V_{reset}}^2 < \sigma_\xi^2/2$  the initial deviation from the mean trajectory is  $\sigma_{V_{reset}}$  and grows exponentially with time to  $\sigma_\xi/\sqrt{2}$ ; and if *iii*)  $\sigma_{V_{reset}}^2 > \sigma_\xi^2/2$  the membrane potential deviation from the mean trajectory starts in  $\sigma_{V_{reset}}$  and decreases exponentially to  $\sigma_\xi/\sqrt{2}$ .

## 4.5. Conclusions

This chapter presented and discussed two novel retina models. The first model described belongs to the spiking neuron class of models, and the second one is a firing rate class' model.

The spiking neuron model has two versions which were analyzed. The first version has only a static structure, while a second version of the model includes the static plus a dynamic structure. The comparison between the performances of the static and dynamic versions for this model allows to conclude about the importance of the dynamic blocks in retina modeling. Using the stimulus and the spike history to modulate an additional dynamic contribution from filtering the stimulus and the spike train decreased the model's errors significantly in relation to the static version. These results emphasize the importance of dynamic mechanisms in modeling the response of the retina.

The second model introduced is a multi kernel linear-nonlinear-Poisson. This model follows a generalized additive logistics regression modeling of the retina response, and it shows the importance of the use of more than the **STA** filter in the linear block, and that the use of excitatory and suppressive kernels improves the estimation of the firing rate. This model introduces the use additive logistic regression in conjunction with the generalized additive models in retina modeling that showed to be effective in the estimation of the nonlinear effects in the **LNP** model, and permits

#### 4. *Modeling the Retina*

to identify immediately the effect of each linear filter as being excitatory or suppressive. Additionally, this modeling technique easily allows the inclusion and removal of the excitatory and suppressive kernels in the modeling and analysis of the retina response.

This chapter ended with an analytical comparison between Poisson based firing rate models and spiking models. Based on the stochastic integrate-and-fire model and a Poisson based firing rate models, it was shown that these two types of models are closely related by comparing their spike generation probability functions.

*The purpose of computing is insight, not numbers.*

Hamming, R. W. (1987). *Numerical Methods for Scientists and Engineers*. Dover Publications, second edition

# 5

## Neural Activity Metrics

### 5.1. Introduction

THE selection of a retina model is an indispensable task in the development of a bioelectronic vision system. Therefore, to make a rational choice it is important to measure the performance of the different available models, and consequently there is the need of a metric, or set of metrics. Even in the analysis of natural neural responses it is useful in many situations to evaluate and compare the reliability of the spike sequences.

Several neural metrics have been proposed to analyze neural responses, and to tune and evaluate neural models. The neural metrics are sensitive to different characteristics of the neural code, and can be grouped in distinct classes according to the firing rate or time-code perspective. These metrics should be closely related to neural code supposedly used by the actual neuron. The neural metrics can be classified as belonging to one of the classes: *firing rate* metrics, *spike train* metrics or *firing event* metrics.

Many of the neural metrics proposed in the literature have been developed with the purpose of studying particular aspects of the neural code or to compare the responses of specific neural systems. Therefore, these metrics are mainly used in particular situations; their application is not generalized and, in many cases, no quantitative measurements are provided for their performance. In the search to solve this problem, a set of neural code metrics is herein presented with a unified perspective, i.e., both for the tuning and assessment of neural retina models and by performing a direct comparison of neural responses. The application of these metrics should be wisely made, since some of these metrics include parameters that regulate their sensitivity, so that the incorrect selection of its values can lead, for example, to consider that two completely different neural responses are similar, or that two responses of the same

## 5. Neural Activity Metrics

neuron to the same input stimulus are far apart. In some cases, the metrics do not directly compare the neural responses. Instead these responses are transformed first into a point in the metric's domain. By applying an incorrect transformation, i.e., by incorrectly estimating the point in the metric's domain, a bias can be introduced in the analysis. The application of neural code metrics in the tuning of neural models must also be done judiciously. The incorrect choice of the metric's parameters in the tuning of a model affects the quality of its output, as not all metrics are suitable for neuron model tuning.

This chapter describes the application of the metrics to compare real retina neural responses, and to train models, with the purpose of revealing their constraints. To emphasize the use of the metrics and to reveal their advantages and drawbacks, two different sets of retina neural responses are used. The application of the neural metrics to retina neural responses allows us to draw conclusions about the choice of its parameters and their use in the tuning and assessment of retina models. The most relevant results were published in [Martins et al., 2009].

After introducing the concept of a metric, several metrics used to compare spike trains, firing rates, and sets of spike trains, are described and their main characteristics discussed. Afterwards, the metrics are applied to tune and assess a set of canonical retina models with the goal of comparing their performances, and providing relevant guidelines for their usage.

### 5.2. The Definition of Metric

A metric is a function,  $d$ , that maps a pair of points,  $\mathbf{x}, \mathbf{y} \in \mathbb{S}$ , of a vectorial space  $\mathbb{S}$ , to a nonnegative real number representing the distance between these two points:  $d : \mathbb{S}^2 \rightarrow \mathbb{R}_0^+$ . To be regarded as a metric in the mathematical sense the distance function  $d(x, y)$  must have the properties  $\forall \mathbf{x}, \mathbf{y} \in \mathbb{S}$ :

$$d(\mathbf{x}, \mathbf{y}) \geq 0, \quad \text{where} \quad d(\mathbf{x}, \mathbf{y}) = 0 \Rightarrow \mathbf{x} = \mathbf{y}, \quad (5.1)$$

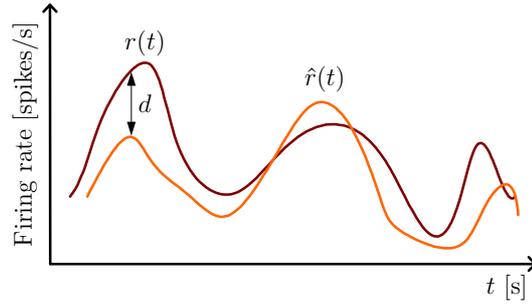
$$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}), \quad (5.2)$$

$$d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}) \quad (\text{triangle inequality}). \quad (5.3)$$

An example of a common metric is the Euclidian distance defined in the  $\mathbb{R}^n$  Euclidian space. The Euclidian distance corresponds to the distance between two  $n$ -dimensional points with coordinates  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  and  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ , defined by  $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_0^+$  as:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_n - x_n)^2}.$$

However, the neural response cannot be represented as a vector on the real space, therefore a suitable set of metrics have been devised. Different metrics and error measures have been proposed to evaluate the performance of neuronal models and to analyze neuronal activity. Some of these distance measures clearly depart from the straight mathematical definition of a metric since they do not possess some of



**Figure 5.1.** Comparison of two firing rates.

the properties enumerated in Eq. (5.1)–Eq. (5.3), though they are used as a distance measure between neural responses.

The different perspectives of the neural code led to the development of several metrics that can be classified into three distinct classes: *i*) firing rate metrics, which compare firing rates; *ii*) spike train metrics, that compare directly two spike trains; and *iii*) spike events metric, which measure the (dis)similarity between bursts of spikes within two sets of spike trains. These metrics look to the neural code from different perspectives (see Sec. 3.2), and each one weighs differently a set of characteristics of the neural response.

### 5.3. Firing Rate Metrics

Firing rate metrics measure the distance, or error, between two sets of spike trains by comparing their mean firing rates. Because these rates (number of elicited spikes per second) are not known *a priori*, they must be estimated from the neural responses. To estimate the mean number of elicited spikes per second, time is discretized into small time bins, and the rate is computed as the number of spikes found in each time bin normalized by the discretization period. To further improve the estimation, multiple trials obtained under similar conditions (i.e., with the same stimulus) are typically used:

$$\hat{r}[n] = \frac{1}{M} \sum_{j=1}^M \frac{1}{\Delta t} \underbrace{\int_{n\Delta t - \frac{\Delta t}{2}}^{n\Delta t + \frac{\Delta t}{2}} \rho_j(t) dt}_{\text{Number of spikes in bin } n} \quad (5.4)$$

where  $n$  is the time bin number,  $\Delta t$  is the discretization period and  $M$  represents the number of spike trains used to estimate the firing rate. In Eq. (5.4) it is assumed that the sampling occurs in the middle of the time bin. The firing rate calculated from Eq. (5.4) can be plotted as a histogram (see Fig. 5.5a), which is usually known as the peristimulus time histogram (PSTH). If an infinitesimally small time bin is used in Eq. (5.4), the estimated firing rate is given by the ensemble average of the

## 5. Neural Activity Metrics

neural response functions:

$$\hat{r}(t) = \frac{1}{M} \sum_{j=1}^M \rho_j(t) . \quad (5.5)$$

As the size of the time bin,  $\Delta t$ , in Eq. (5.4) decreases the firing rate becomes sharper. The major problem with the use of an extremely small discretization period is that the rate becomes impulsive, leading to large errors both when the error metrics are applied to compare different neural responses and when the objective is to tune the parameters of a model. This problem becomes more relevant when the number of trials available to compute the ensemble average is small.

Typically, to overcome this problem, some type of low pass filtering is applied to the ensemble average [Berry et al., 1997; Ventura et al., 2002], where the most common method is to convolve it with a Gaussian function  $g(t; \sigma)$  with zero mean and standard deviation  $\sigma$ :

$$\hat{r}(t) = \frac{1}{M} \sum_{j=1}^M \rho_j(t) * g(t; \sigma) . \quad (5.6)$$

The drawback of the application of a low pass filter is that it attenuates high frequency components that may exist in the true firing rate  $r(t)$ . Thus, in the specific case of Eq. (5.6), the Gaussian parameter  $\sigma$  must be chosen carefully so that the ensemble average is smoothed without removing important information about the firing rate. In fact, it is known that there exists an optimal value for the  $\sigma$  parameter that leads to the minimum error between the true and the estimated firing rates [Nawrot et al., 1999]. A method suggested to compute the Gaussian standard deviation is by making it equal to the time scale of the modulations in the firing rate [Berry et al., 1997]. This method generally selects an acceptable value for  $\sigma$ , although it ignores the number of trials used in the ensemble average. Another method to estimate  $\sigma$  is based on the optimization of the mean-squared-error between the estimated and the true firing rate [Shimazaki and Shinomoto, 2010].

The firing rate metrics are well suited for the analysis of rate code models. In the assessment of neuron models the comparison of two firing rates comprises a reference firing rate, usually obtained from the responses of a retinal neuron, and an estimated firing rate, corresponding to the model output.

### 5.3.1. Mean Squared Error

The *mean squared error* is a common metric in engineering that is also used in neuroscience in the comparison of firing rates. Given two firing rates  $r(t)$  and  $\hat{r}(t)$  the distance following the mean squared error (**MSE**) between the two firing rates is obtained from:

$$\text{MSE}(r, \hat{r}) = E\{[r(t) - \hat{r}(t)]^2\} = \frac{1}{T} \int_0^T [r(t) - \hat{r}(t)]^2 dt, \quad (5.7)$$

where  $\hat{r}(t)$  is the firing rate that is compared with the reference firing rate  $r(t)$  during the interval of time  $T$ . The first equality in Eq. (5.7) denotes the expectation operator  $E$ . In the case of an ergodic process a time average can be used instead of the ensemble average [Papoulis and Pillai, 2002; Therrien, 1992].

For actual data analysis the firing rate is discretized in time so that the MSE is obtained from

$$\text{MSE}(r, \hat{r}) = \frac{1}{N} \sum_{n=0}^{N-1} (r[n] - \hat{r}[n])^2, \quad (5.8)$$

where  $N$  is the number of time bins into which the spike trains were discretized.

### 5.3.2. Normalized Mean Squared Error

Another statistic used to measure how well a given model captures the fast modulations in the observed firing rate is obtained by dividing the MSE by the variance<sup>1</sup> of the observed firing rate. This statistic is called *normalized mean squared error* [Berry II and Meister, 1998]. The normalized mean squared error (NMSE) is defined by:

$$\text{NMSE}(r, \hat{r}) = \frac{\text{MSE}(r, \hat{r})}{\text{VAR}(r)} = \frac{\int_0^T [r(t) - \hat{r}(t)]^2 dt}{\int_0^T [r(t) - \langle r \rangle]^2 dt}. \quad (5.9)$$

When the firing rate is discretized into time bins of finite length, the expression for the NMSE can be written as:

$$\text{NMSE}(r, \hat{r}) = \frac{\sum_{n=0}^{N-1} (r[n] - \hat{r}[n])^2}{\sum_{n=0}^{N-1} (r[n] - \langle r \rangle)^2}, \quad (5.10)$$

where  $\langle r \rangle$  is the average reference firing rate. This error measure violates the metric properties in Eq. (5.2) and Eq. (5.3) due to the biased denominator term.

### 5.3.3. Percent Variance Accounted For

Another error measure with an expression just slightly different from the NMSE is the *percent-Variance-Accounted-For*. The percent-Variance-Accounted-For (%VAF) was originally proposed to evaluate how well a model describes a biological system in terms of the firing rate [Westwick and Kearney, 2003; Pillow et al., 2005]. In terms of the firing rate, the %VAF compares the observed firing rate  $r(t)$  with the estimated

---

<sup>1</sup> $\text{VAR}(x) = E\{(x - \mu_x)^2\} = E\{x^2\} - \mu_x^2$ , where  $\mu_x = E\{x\}$ .

## 5. Neural Activity Metrics

firing rate  $\hat{r}(t)$  using the expression:

$$\begin{aligned} \%VAF(r, \hat{r}) &= 100 \times \frac{\text{VAR}(r - \hat{r})}{\text{VAR}(r)} \\ &= 100 \times \frac{E\{[(r(t) - \hat{r}(t)) - (\langle r \rangle - \langle \hat{r} \rangle)]^2\}}{E\{[r(t) - \langle r \rangle]^2\}}, \end{aligned} \quad (5.11)$$

where the quantities between angle brackets are the mean firing rates of the observed and predicted trains that are calculated using Eq. (B.37). Expanding the expectation operator in Eq. (5.11), the %VAF becomes:

$$\begin{aligned} \%VAF(r, \hat{r}) &= \\ 100 \times &\frac{E\{r^2(t)\} + E\{\hat{r}^2(t)\} - \langle r \rangle^2 - \langle \hat{r} \rangle^2 + 2\langle r \rangle \langle \hat{r} \rangle - 2E\{r(t)\hat{r}(t)\}}{E\{r^2(t)\} - \langle r \rangle^2}, \end{aligned} \quad (5.12)$$

where the last term in the numerator of Eq. (5.12) is the correlation between the observed and the estimated firing rates. If the two processes are independent, then  $E\{r(t)\hat{r}(t)\} = E\{r(t)\}E\{\hat{r}(t)\} = \langle r \rangle \langle \hat{r} \rangle$ , and the last two terms in the numerator of Eq. (5.12) cancel each other. In this case, the two firing rates are uncorrelated, and the %VAF reduces to the sum of the variances of the observed and estimated firing rates normalized by the variance of the reference firing rate.

For a firing rate discretized into  $N$  time bins, the %VAF reduces to the expression:

$$\begin{aligned} \%VAF(r, \hat{r}) &= 100 \times \\ &\frac{\frac{1}{N} \sum_{n=0}^{N-1} r^2[n] + \frac{1}{N} \sum_{n=0}^{N-1} \hat{r}^2[n] - \langle r \rangle^2 - \langle \hat{r} \rangle^2 + 2\left(\langle r \rangle \langle \hat{r} \rangle - \frac{1}{N} \sum_{n=0}^{N-1} r[n]\hat{r}[n]\right)}{\frac{1}{N} \sum_{n=0}^{N-1} r^2[n] - \langle r \rangle^2}. \end{aligned} \quad (5.13)$$

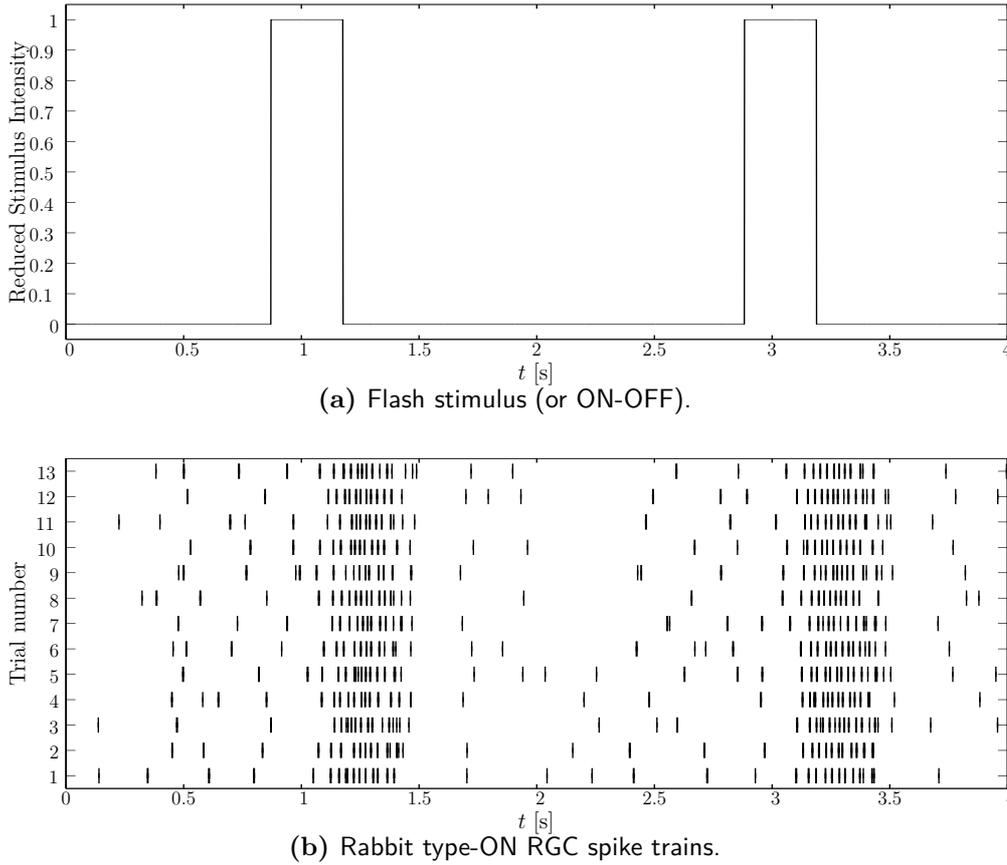
Using the discrete variance operator the expression for the %VAF metric can simply be written as:

$$\%VAF = 100 \times \frac{\text{VAR}(r[n] - \hat{r}[n])}{\text{VAR}(r[n])}. \quad (5.14)$$

The biased term in the denominator of Eq. (5.11) leads this error measure to violate Eq. (5.2) and Eq. (5.3).

### 5.3.4. Analysis of the Firing Rate Metrics

For the evaluation of the metrics two data sets resulting from the measurement of real retina responses were employed. The first set, provided by Prof. Eduardo Fernández [Unidad de Neuroprótesis y Rehabilitación Visual, 2008], was obtained by recording the responses of a type-ON rabbit retina RGC when stimulated with a full-field ON-OFF stimulus. This data corresponds to 13 trials from a single neuron, each lasting four seconds, with an average of 62 spikes per trial (see Fig. 5.2). The second set of data, from [Keat et al., 2001] and provided by Prof. Markus Meister, corresponds to



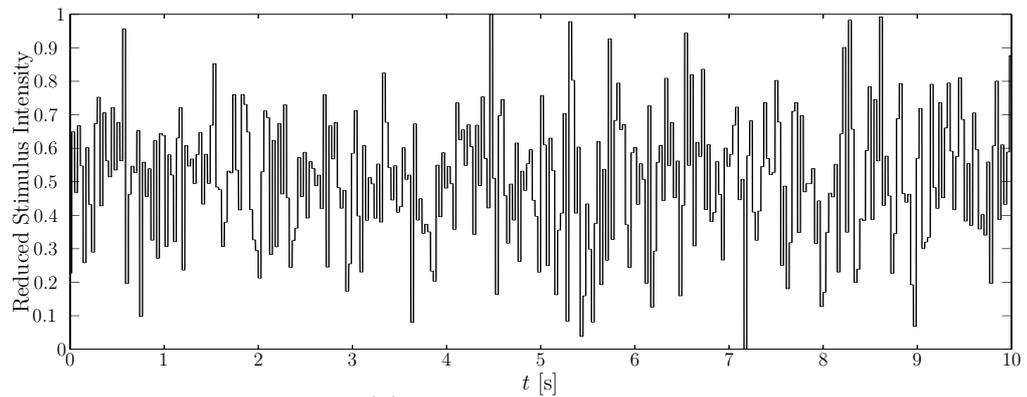
**Figure 5.2.** Rabbit type-ON RGC responses for a ON-OFF full-field stimulus ( $T_s = 1$  ms).

the recording of salamander type-ON RGC responses when stimulated by a sampled white noise spatially uniform stimulus. The 12 trials of responses are from a single neuron, where each trial lasts a total of 10s and has an average of 83 spikes per trial (see Fig. 5.3). The RGC neural functions and the stimuli were sampled with a period of  $T_s = 1$ ms.

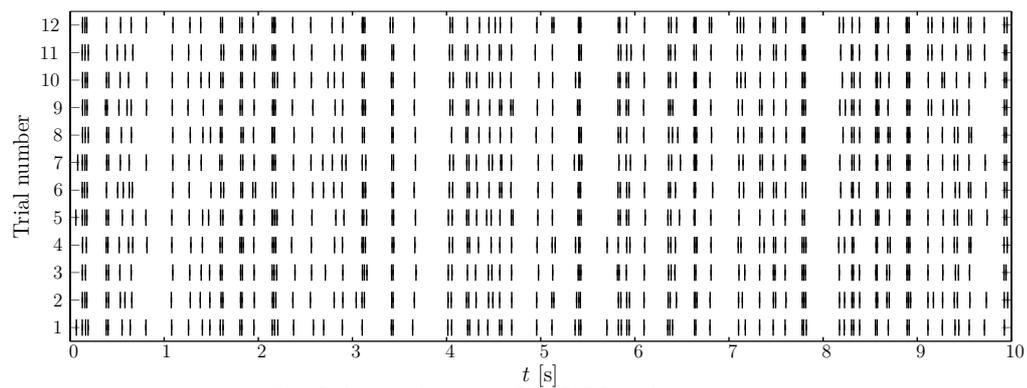
The main drawback in the usage of the firing rate metrics is that, given a set of neural responses, one does not know the mean firing rate of these responses. Instead, this rate must be estimated, which naturally leads to additional bias sources when the rate metrics are applied.

Figure 5.4 displays the firing rate error measures for the different metrics for the salamander type-ON RGC as a function of the Gaussian smoothing filter width. The values in the figure were obtained by *i*) dividing the 12 available trials into two sets of 6 trials; *ii*) estimating the firing rates by convolving the two ensemble averages with a Gaussian function with zero mean and standard deviation  $\sigma$ ; and *iii*) applying the MSE, NMSE and %VAF. To obtain the mean and standard deviation values, all possible combinations of trials were considered. The curves in Fig. 5.4 show that if no smoothing is applied,  $\sigma = 0$  s, the error is very large, even though the cell

## 5. Neural Activity Metrics

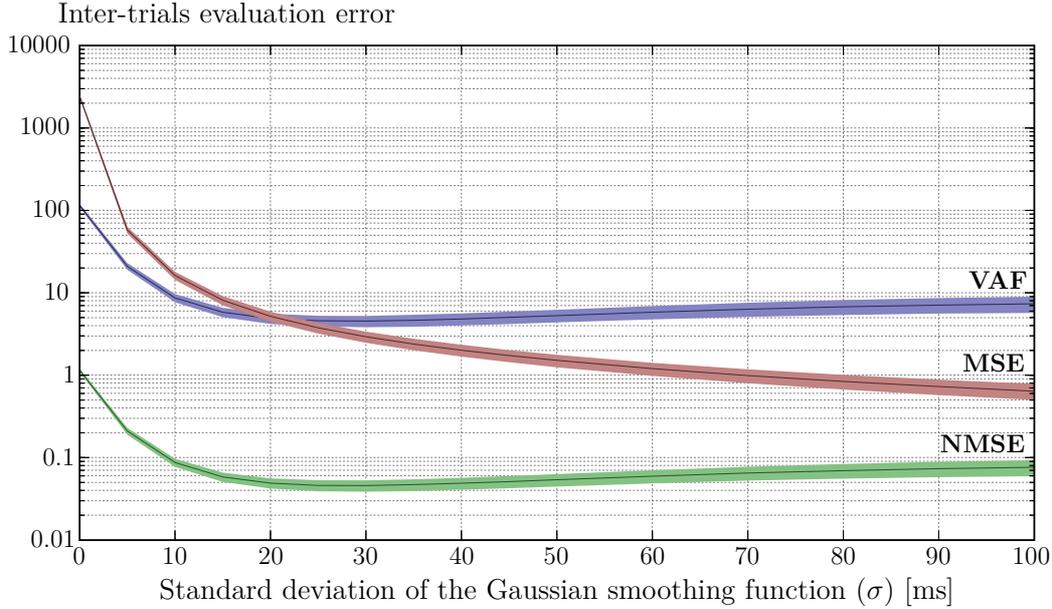


(a) White noise stimulus.



(b) Salamander type-ON RGC spike trains.

**Figure 5.3.** Salamander type-ON RGC responses for sampled white-noise full-field stimulus ( $T_s = 1$  ms).



**Figure 5.4.** Inter-trials evaluation of salamander RGC responses varying the standard deviation values of the Gaussian smoothing function (line – mean value; shaded area – mean  $\pm$  standard deviation).

is always encoding the same stimulus. This shows that some smoothing must be done in order to have a reasonable estimation of the firing rates. However, the larger the smoothing parameter  $\sigma$ , the more the high frequency components are attenuated, possibly hiding the true characteristics of the responses' firing rate. This is illustrated by the **NMSE** and **%VAF** curves. If the rates are over-smoothed, the temporal variance becomes too low and the error values start to increase. Between these two extremes, the values  $\sigma \in [15 \text{ ms}; 30 \text{ ms}]$  is defined with the minimum errors for the **NMSE**; this region also defines the values for the smoothing parameter, which should be used to estimate the firing rate.

The above procedure for estimating the value for the smoothing parameter  $\sigma$  can be generalized to other neural responses. To determine an appropriate value for this smoothing parameter, one can perform the inter-trial evaluation of a neuron's responses with the **NMSE** metric using different values of  $\sigma$ . This will define the range of values that can be used for the parameter  $\sigma$ . It should be noted that if the number of trials used in the ensemble average is large, the firing rate estimation is better, and the minimum moves towards a lower value of  $\sigma$ . Therefore, this technique has an advantage over the one proposed by Berry, *et al.* [Berry et al., 1997]: as the number of trials used in the ensemble average increases (which improves the estimation of the firing rate), the chosen value for  $\sigma$  decreases.

From the analysis of the **%VAF** metric it can be concluded that it should not be used to evaluate neural responses. By expanding the **%VAF** expression in Eq. (5.11)

## 5. Neural Activity Metrics

one gets::

$$\begin{aligned}
 \%VAF(r, \hat{r}) &= 100 \times \frac{\text{VAR}[r(t) - \hat{r}(t)]}{\text{VAR}[r(t)]} \\
 &= 100 \times \frac{E\{(r(t) - \hat{r}(t))^2\} - E^2\{r(t) - \hat{r}(t)\}}{\text{VAR}[r(t)]} \\
 &= 100 \times \left( \underbrace{\text{NMSE}(r, \hat{r})}_{T1} - \underbrace{\frac{(E\{r(t)\} - E\{\hat{r}(t)\})^2}{\text{VAR}[r(t)]}}_{T2} \right), \tag{5.15}
 \end{aligned}$$

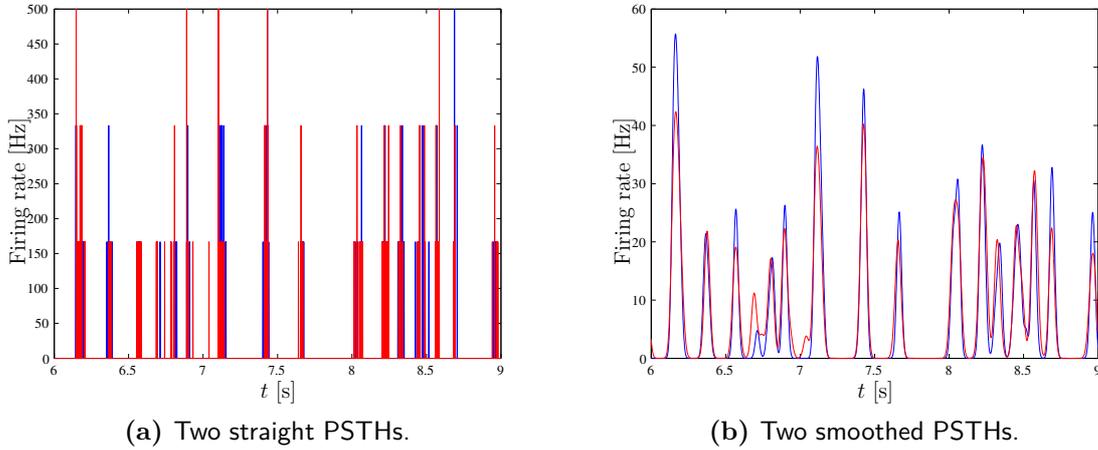
which states that the **%VAF** metric consists of a term T1 proportional to the **NMSE** metric minus a term T2 that increases with the difference of the mean firing rates. This expansion shows a higher difference in the mean temporal values of the rates indicates a lower **%VAF** error so that the second term is undesirable. Because the removal of the second term results in the **NMSE** metric, one concludes that the **%VAF** metric adds no information in the evaluation of neural responses relatively to the **NMSE** metric.

Table 5.1 contains the values obtained by applying the firing rate metrics to the salamander ON-type **RGC** responses. These values were obtained by composing and comparing different sets of spike trains from the set of 12 trials shown in Fig. 5.3b, with each set composed by 6 trials. A Monte-Carlo method was used to randomly select 6 trials among the 12 trials to compose the sets of spike trains to be compared. From these results it is possible to observe noticeable differences between the values of the **MSE** and its normalized version, the **NMSE**, and also between the use of straight **PSTH** and the smoothed **PSTH**. The **PSTH** was smoothed with a gaussian filter with a standard deviation  $\sigma$  that was computed using the Algorithm 5.3. To be considered to have a good performance in terms of the firing rate metrics the errors of a model should be within the range given by the values in Table 5.1 for this particular  $\sigma$ .

Another important characteristic of the firing rate metrics is their sensitivity to phase (time shift). That is, by comparing two sets of spike trains whose only difference is a shift in the time occurrence of the peaks in their firing rates the error will be significant. This error could be even lower if the metric is applied to compare a firing rate obtained from a set of spike trains and a firing rate equal to zero, which is not a desirable result. Figure 5.5 displays the firing rates for two different sets of trials from the twelve trains, and it can be observed from Fig. 5.5a that the non smoothed **PSTH** has big amplitude peaks so that a shift in the phase would result in a large jump in the error value. For the smoothed **PSTHs** of Fig. 5.5b the peaks' amplitudes are smaller and spread, so a phase shift only leads to a more moderate increase in the error, (one must recall that the integral of the straight and of the smoothed **PSTHs** are equal, so that the difference for the metrics values is due to phase errors only). Therefore, the values in Table 5.1 obtained with the smoothed **PSTHs** are much lower precisely because smoothing decreases the errors due to phase differences.

Firing Rate Metrics			
Straight PSTH			
Metric	mean $\pm$ std	min	max
MSE	$2454.6 \pm 138.2447$	2041.7	2875.0
NMSE	$1.1450 \pm 0.0542$	0.9833	1.3234
%VAF	$114.4976 \pm 5.4213$	98.3281	132.3359
Smoothed PSTH			
Metric	mean $\pm$ std	min	max
MSE	$8.0686 \pm 1.1295$	5.5717	12.1673
NMSE	$0.0593 \pm 0.0088$	0.0405	0.0951
%VAF	$5.9054 \pm 0.8836$	4.0269	9.5090

**Table 5.1.** Values of the firing rate metrics applied to the salamander ON-type RGC responses.

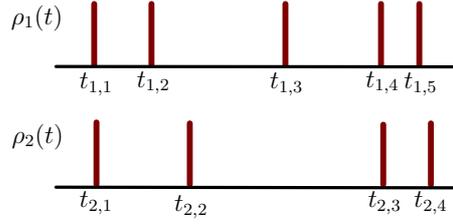


**Figure 5.5.** Comparing smoothed and non-smoothed PSTHs.

## 5.4. Spike Train Metrics

Spike train metrics measure directly the distance between two spike trains. These metrics are particularly suited for the assessment of time-code models by measuring the distance, or (dis)similarity, between the spike trains from a neuron and the ones predicted by a model. Three suitable metrics that can be found in the literature are: the *spike time metric*, the *interspike interval metric*, and the *spike distance metric*. For these metrics each point on the metric space  $\mathbb{S}$  is the neuronal response  $\rho(t)$ , and it gives the distance between the spike trains  $\rho_1(t)$  and  $\rho_2(t)$ , composed of the spikes at time instants  $t_{1,1}, t_{1,2}, \dots, t_{1,n_1}$  and  $t_{2,1}, t_{2,2}, \dots, t_{2,n_2}$ , respectively, as depicted in Fig. 5.6.

## 5. Neural Activity Metrics



**Figure 5.6.** Comparison of two spike trains.

### 5.4.1. Spike Time Metric

A pair of spike train metrics were proposed in [Victor and Purpura, 1996, 1997], and revisited in [Victor, 2005]. These metrics consider the spike trains as points in a vector space, and the metric is applied to calculate their distance. They include the *spike time metric*, denoted here by  $d^{\text{time}}(\cdot)$ , which considers the absolute time occurrence of the individual spikes in a train as the significant quantity that distinguishes two spike trains.

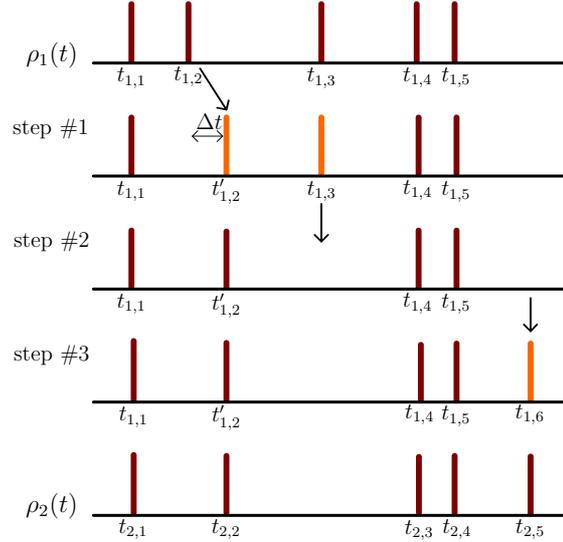
The distance between the spike trains  $\rho_1(t)$  and  $\rho_2(t)$  is measured as the minimum cost path followed to transform one spike train into the other by iteratively applying a set of elementary operations. The cost to transform a spike train  $\rho_i(t)$  into  $\rho_j(t)$ , like the ones depicted in Fig. 5.6 (where  $i = 1$  and  $j = 2$ ), denoted by the distance  $d(\rho_i(t), \rho_j(t))$ , is equal to the sum of the individual costs associated with a particular sequence of elementary steps that successively transform the spike train  $\rho_i(t)$  into  $\rho_j(t)$ :

$$d(\rho_i(t), \rho_j(t)) = \min \left\{ d(\rho_i(t), \rho_{i+1}(t)) + \dots + d(\rho_{j-1}(t), \rho_j(t)) \right\}, \quad (5.16)$$

where the difference between  $\rho_i(t)$  and  $\rho_{i+1}(t)$  corresponds to a single elementary operation. The minimum is calculated over all possible paths of spike trains,  $\rho_i(t), \rho_{i+1}(t), \dots, \rho_{j-1}(t), \rho_j(t)$ , beginning in  $\rho_i(t)$  and ending at  $\rho_j(t)$ .

The metric  $d^{\text{time}}(\cdot)$  is a function of a parameter  $q$  that measures its sensitivity to the timing of the spikes occurrence. The allowed elementary steps are: *i*) the insertion and *ii*) the deletion of an individual spike, both of these operations have a unitary cost per spike; and *iii*) the shift in the time occurrence of a spike, with a cost of  $q$  per second, which leads to a cost of  $q|\Delta t|$  for a shift of  $\Delta t$  seconds on the occurrence of a spike. The further away a spike is relative to the correspondent spike on the other train, the more costly it is to shift the spike to the right place. The shifting of a spike can be so costly that it can be less costly to delete it and raise another spike coincident with the comparing spike, which has a total cost of 2 ( $d^{\text{time}}(\text{spike delete} + \text{spike raise}) = 2$ ).

Extending the previous reasoning two antipodal limit cases are found. The first case occurs when  $q = 0$ , meaning that the shifting of a spike is costless, so that the contributions to the cost are due only to the deletion or raising of spikes. If the spike train  $\rho_1(t)$ , with  $n_1$  spikes, is being compared with the spike train  $\rho_2(t)$ , with  $n_2$  spikes, the minimum path to transform one spike train into the other can



**Figure 5.7.** Spike train transformation in the spike time metric.

be done with  $|n_1 - n_2|$  steps by creating (or deleting) a spike at each step with a cost of  $|n_1 - n_2|$ . This corresponds to a spike-count metric,  $d_{q=0}^{\text{time}}(\cdot) = d^{\text{count}}(\cdot)$ , where the single relevant characteristic of the spike trains is its total number of spikes. According to  $d^{\text{count}}$ , two spike trains with the same number of spikes have an error equal to zero, although the spikes can be positioned at different time instants.

The second extreme case occurs when  $q$  takes a high value. Considering two spike trains with a single spike each positioned at times  $t$  and  $t'$ , respectively, the cost for shifting the spike to the right place is  $d^{\text{time}}(\rho_1(t), \rho_2(t)) = q|t - t'|$ . The alternative path is to delete the spike at the time instant  $t$  and create another spike at time  $t'$ , which has a total cost of 2. The latter path can be less costly than the former if  $|t - t'| > 2/q$ . When  $q$  is very high, ( $q \rightarrow \infty$ ) the distance between two spike trains  $\rho_1(t)$  and  $\rho_2(t)$  is  $d_{q \rightarrow \infty}^{\text{time}} = n_1 + n_2 - 2c$ , where  $c$  is the number of coincident spikes in time on the two trains under comparison.

In between the two extreme situations described,  $d_q^{\text{time}}(\cdot)$  defines a family of metrics where the cost of displacing a spike by an amount of  $\Delta t = 2/q$  [s] is equal to its deletion and recreation. Therefore, the parameter  $q$  can be viewed as a measure of the precision of the temporal coding, determining how far a spike can be shifted without increasing the distance between the two spike trains with regard to the spike time metric with  $q \rightarrow \infty$ . For  $q \neq 0$  and finite  $d^{\text{time}}$  will always take values in the interval  $d^{\text{time}}(\rho_1(t), \rho_2(t)) \in \{|n_1 - n_2|, n_1 + n_2 - 2c\}$ , where  $n_1$  and  $n_2$  are the number of spikes on each spike train, and  $c$  the number of coincident spikes.

In the search for a minimal path between two spike trains there are restrictions in the operations that can be performed. Namely, the path cannot comprise the shift and deletion of the same spike, because the cost will be lower by just deleting the spike; a spike cannot be included and then shifted, since the direct insertion of the spike in the right position is less costly; and the insertion and deletion of the same

## 5. Neural Activity Metrics

---

### Algorithm 5.1 Spike Time Metric, $d^{\text{time}}$

---

```

1:  $\mathbf{t}_1 \leftarrow$  vector of spike time instants for train  $\rho_1$ 
2:  $\mathbf{t}_2 \leftarrow$  vector of spike time instants for train  $\rho_2$ 
3:  $n_1 \leftarrow$  number of elements in  $\mathbf{t}_1$ , i.e., number of spikes in train  $\rho_1$ 
4:  $n_2 \leftarrow$  number of elements in  $\mathbf{t}_2$ , i.e., number of spikes in train  $\rho_2$ 
5:  $q \leftarrow$  penalty to shift one spike by one second (in any direction)
6:  $\mathbf{D} \leftarrow$  matrix of size  $[(n_1 + 1) \times (n_2 + 1)]$  where
    • first row:  $\mathbf{D}_{0,:} = [0, 1, \dots, n_2]$ 
    • first column:  $\mathbf{D}_{:,0} = [0, 1, \dots, n_1]^T$ 
7: for  $i = 1$  to  $n_1$  do {For all spikes in train  $\rho_1$ ,  $\mathbf{t}_1$ }
8:   for  $j = 1$  to  $n_2$  do {For all spikes in train  $\rho_2$ ,  $\mathbf{t}_2$ }
9:      $\mathbf{D}_{i,j} = \min \left\{ \mathbf{D}_{i-1,j} + 1, \mathbf{D}_{i,j-1} + 1, \mathbf{D}_{i-1,j-1} + q|\mathbf{t}_{1,i} - \mathbf{t}_{2,j}| \right\}$ 
10:   end for
11: end for
12: return  $d^{\text{time}} \leftarrow \mathbf{D}_{n_1,n_2}$ 

```

---

spike cannot be part of the minimal path between the two spike sequences. Also, a minimal path cannot include both leftward and rightward spike shifts of a spike, nor the crossing of two spikes. This reduces the number of operations that can be exerted on a certain spike when comparing trains  $\rho_1(t)$  and  $\rho_2(t)$  to three: *i*) deletion of the last analyzed spike in  $\rho_1(t)$ , *ii*) insertion of a spike in  $\rho_1(t)$ , or *iii*) the two last spikes in  $\rho_1(t)$  and  $\rho_2(t)$  are related by a shift. Figure 5.7 illustrates a set of possible operations that can be exerted to transform one spike train into another, where in the first step a spike is shifted by  $\Delta t$  in train  $\rho_1(t)$ , in the second step a spike is deleted from  $\rho_1(t)$ , and in the third step a spike is inserted in the train, so that we arrive at  $\rho_2(t)$ .

The process to calculate the distance  $d^{\text{time}}(\rho_1(t), \rho_2(t))$  between two spike trains is formalized in Algorithm 5.1. Figure 5.7 illustrates the computation of the spike distance metric by considering the spike trains  $\rho_1(t) = \sum_{i=1}^{n_1} \delta(t - t_{1i})$ , with  $n_1$  spikes positioned at times  $\{t_{1,1}, t_{1,2}, \dots, t_{1,n_1}\}$ , and  $\rho_2(t) = \sum_{i=1}^{n_2} \delta(t - t_{2i})$  that has  $n_2$  spikes located at  $\{t_{2,1}, t_{2,2}, \dots, t_{2,n_2}\}$ . If  $d_{i,j}^{\text{time}}$  is the spike time distance between the two spike trains composed of the first  $i$  spikes of train  $\rho_1(t)$  and of the first  $j$  spikes of train  $\rho_2(t)$ , the allowed operations listed above imply that the minimum path must be:

$$d_{i,j}^{\text{time}} = \min \left\{ \begin{array}{l} \text{Add spike to } \rho_1 \text{ at } t_{2,j} \\ \text{(Erase spike in } \rho_2 \text{ at } t_{2,j}) \end{array} \overbrace{d_{i-1,j}^{\text{time}} + 1} \quad ; \quad \begin{array}{l} \text{Add spike to } \rho_2 \text{ at } t_{1,i} \\ \text{(Erase spike in } \rho_1 \text{ at } t_{1,i}) \end{array} \overbrace{d_{i,j-1}^{\text{time}} + 1} \quad ; \quad \underbrace{d_{i-1,j-1}^{\text{time}} + q|t_{1,i} - t_{2,j}|}_{\substack{\text{Shift spike in } \rho_1 \text{ from } t_{1,i} \text{ to } t_{2,j} \\ \text{(Shift spike in } \rho_2 \text{ from } t_{2,j} \text{ to } t_{1,i})}} \right\}. \quad (5.17)$$

By successively applying the expression in Eq. (5.17) to every new spike in each spike

train  $\rho_1(t)$  and  $\rho_2(t)$ , a two-dimensional array can be constructed where the value in the  $i$ th row and  $j$ th column contains the distance  $d_{i,j}^{\text{time}}$ . The first row of this two-dimensional array is filled with  $d_{0,j}^{\text{time}} = j$ , corresponding to have zero spikes in the first train and  $j$  spikes in the second train, so that all spikes are created in the first train (or deleted in the second train). The first column of this array is filled with  $d_{i,0}^{\text{time}} = i$ , corresponding to  $i$  spikes in  $\rho_1$  and zero spikes in  $\rho_2$ , so that all spikes are deleted on the first train (or created in the second train). By iteratively filling the  $(n_1 + 1) \times (n_2 + 1)$  array, the overall cost  $d^{\text{time}}(\rho_1(t), \rho_2(t))$  is located in the array cell  $(n_1, n_2)$ . The Algorithm 5.1 summarizes the steps needed to implement this metric in a computer, where the matrix  $\mathbf{D}_{(n_1+1, n_2+1)}$  is recursively filled with the distances given by Eq. (5.17). This procedure was first proposed and established to compare gene sequences [Sellers, 1974], this procedure is also used to measure distances between strings known as the edit or Levenshtein distance. This metric respects Eq. (5.1) – Eq. (5.3), so that it is a metric in the mathematical sense.

### 5.4.2. Interspike Interval Metric

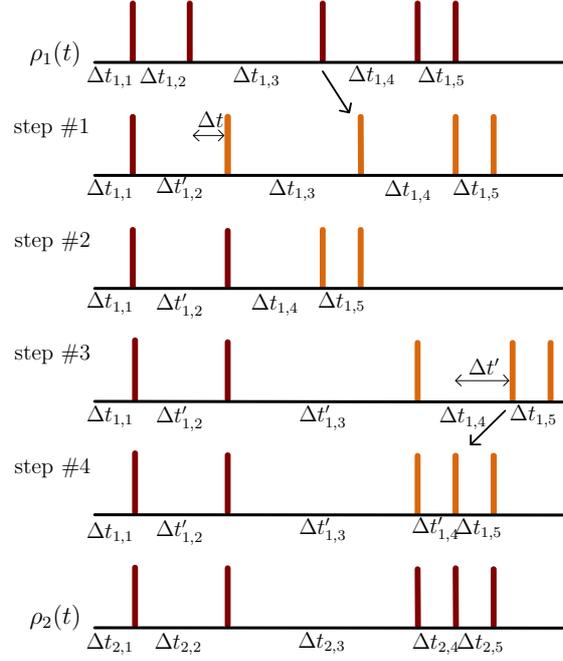
In the *interspike interval metric*,  $d^{\text{inter}}(\cdot)$ , the time intervals between two consecutive spikes convey the relevant information in the trains comparison. The neurobiological justification to use this metric is related to the fact that, due to the dynamics of the ions' channels in the neuron, the effect of an action potential depends on the length of the time interval since the last spike was fired [Victor and Purpura, 1996, 1997].

Similarly to the spike time metric, there is a limited number of allowed operations. The allowed elementary operations are the addition of an interspike interval with the associated spike, and the removal of an interspike interval with the associated spike; both operations have a cost equal to 1. The other allowed operation is the shift of a spike, that changes the duration of the associated **ISI**, with a cost  $q$  per second. As illustrated in Fig. 5.8 a set of elementary steps are applied to transform the spike train  $\rho_1(t)$  into  $\rho_2(t)$ . The deletion and insertion of an interspike interval are exemplified in the second and third steps of Fig. 5.8, respectively, while the shift of a spike, with the consequent change of its interval duration and the shift of all subsequent spikes, is exemplified in the first and fourth steps. The changing of an **ISI** by the amount  $\Delta t$  has a cost of  $q|\Delta t|$ , and has the consequence of changing the absolute time occurrence of all subsequent spikes by  $\Delta t$ ; however, the interspike intervals remain unchanged.

Like in the spike time metric, for  $q = 0$  this metric degenerates into the spike count metric:  $d_{q=0}^{\text{inter}}(\cdot) = d^{\text{count}}(\cdot)$ . Since the cost of changing the duration of an **ISI** is null ( $q = 0$ ), the distance is due only to the removal (or insertion) of **ISIs** with the associated spikes, so that  $d_{q=0}^{\text{inter}}(\rho_1, \rho_2) = d^{\text{count}}(\rho_1, \rho_2) = |n_1 - n_2|$ , for two trains  $\rho_1$  and  $\rho_2$  with a total of  $n_1$  and  $n_2$  spikes, respectively.

For values of  $q$  different from zero, the distance depends on the temporal pattern of spikes composing the sequence, namely on their **ISIs**. For  $q \rightarrow \infty$  an **ISI** duration will never be adjusted since it is less costly to remove the spike with its **ISI**, and insert another **ISI** with the right duration, which has a total cost of 2 ( $d^{\text{inter}}(\text{ISI removal} + \text{ISI insertion}) = 2$ ), and only the coincident **ISIs** between the

## 5. Neural Activity Metrics



**Figure 5.8.** Spike train transformation path in the interspike interval metric.

two trains will not be replaced. Therefore, similarly to the spike time metric, the distance between the spike trains  $\rho_1$  and  $\rho_2$ , with  $n_1$  spikes and  $n_2$  spikes (and consequently with  $n_1$  and  $n_2$  spike intervals, respectively), for  $q \rightarrow \infty$  will be  $d_{q \rightarrow \infty}^{inter}(\rho_1, \rho_2) = n_1 + n_2 - 2c$ , where  $c$  now is the number of interspike intervals with the same duration. Two **ISIs** are equal if the interspike interval before spike  $i$  in train  $\rho_1$  is equal to the interspike interval before spike  $j$  in  $\rho_2$ :  $\Delta t_{1,i} = \Delta t_{2,j}$ . For two consecutive interspike intervals that are equal by  $\Delta t_{1,i} = \Delta t_{2,j}$  and  $\Delta t_{1,k} = \Delta t_{2,l}$ , one must have  $k > i$  and  $l > j$ , because the order of the intervals cannot be reversed. It could happen that we have a different number of coincident **ISIs** depending on the matching sequence.

The Algorithm 5.2 is used to iteratively calculate the distance between two spike trains using  $d^{inter}(\cdot)$ , based on the time intervals between successive spikes [Victor and Purpura, 1996; Sellers, 1974]. Considering the spikes in  $\rho_1(t)$  located at the time instants  $\{t_{1,1}, t_{1,2}, \dots, t_{1,n_1}\}$ , and in  $\rho_2(t)$  located at  $\{t_{2,1}, t_{2,2}, \dots, t_{2,n_2}\}$ , and that the spike intervals are defined for  $\rho_1$  by  $\Delta t_{1,i} = t_{1,i} - t_{1,i-1}$ , and for  $\rho_2$  by  $\Delta t_{2,j} = t_{2,j} - t_{2,j-1}$ . Figure 5.8 illustrates this process. Denoting by  $d_{i,j}^{inter}$  the interspike interval distance between two spike trains composed of the first  $i$  spikes of train  $\rho_1(t)$  and the first  $j$  spikes of the train  $\rho_2(t)$ , the allowed operations imply that the minimum path can

be constructed iteratively by applying the expression:

$$d_{i,j}^{\text{inter}} = \min \left\{ \begin{array}{l} \text{Add ISI } \Delta t_{2,j} \text{ to } \rho_1 \\ \text{(Remove ISI } \Delta t_{2,j} \text{ in } \rho_2) \end{array} \overbrace{d_{i-1,j}^{\text{inter}} + 1} \quad ; \quad \begin{array}{l} \text{Add ISI } \Delta t_{1,i} \text{ to } \rho_2 \\ \text{(Remove ISI } \Delta t_{1,i} \text{ in } \rho_1) \end{array} \overbrace{d_{i,j-1}^{\text{inter}} + 1} \quad ; \quad \underbrace{d_{i,j}^{\text{inter}} + q|\Delta t_{1,i} - \Delta t_{2,j}|}_{\substack{\text{Change ISI } \Delta t_{1,i} \text{ in } \rho_1 \text{ to } \Delta t_{2,j} \\ \text{(Change ISI } \Delta t_{2,j} \text{ in } \rho_2 \text{ to } \Delta t_{1,i})}} \right\}. \quad (5.18)$$

By the successive application of this procedure, a two-dimensional array, with dimensions  $(n_1 + 1) \times (n_2 + 1)$ , can be constructed with the interspike distances,  $d_{i,j}^{\text{inter}}$ , between the first  $i$  ISIs of  $\rho_1(t)$  and the first  $j$  ISIs of  $\rho_2(t)$ . Algorithm 5.2 summarizes the steps needed to implement this metric in a computer, where the matrix  $\mathbf{D}_{(n_1+1, n_2+1)}$  is iteratively filled with the distances given by Eq. (5.18), and the overall distance  $d_{\rho_1, \rho_2}^{\text{inter}}$  ends in the element  $(n_1, n_2)$  of the matrix  $\mathbf{D}$ . The duration of the first ISIs is unknown, since one does not know when the previous spikes occurred. However, they must be at least equal to the time between the start of the data recording until the first spike occurs so that we can make  $\Delta t_{1,1} = t_{1,1}$  and  $\Delta t_{2,1} = t_{2,1}$  to initialize the algorithm; there are other possibilities [Victor and Purpura, 1997]. The first row of the two dimensional array can be filled with  $d_{0,j}^{\text{inter}} = j$ , meaning that all interspike intervals have to be removed from  $\rho_1$  (or inserted in  $\rho_2$ ), while the first column can be filled by taking into account that  $d_{i,0}^{\text{inter}} = i$ , meaning that all interspike intervals have to be removed from  $\rho_2$  (or inserted in  $\rho_1$ ). This metric obeys to Eq. (5.1) – Eq. (5.3), being a metric in the mathematical sense.

### 5.4.3. Spike Train Distance Metric

The *spike train distance metric*,  $d^2(\cdot)$  is a metric less general than the previous spike train metrics, since it is particularly suited to compare spike trains, but easier to analyze and compute [van Rossum, 2001]. This metric measures the dissimilarity between two spike trains by taking into account the temporal structure of the train, and depends on a temporal parameter  $\tau$  that measures the contribution of the displacement of spikes in the two trains under comparison. This metric has a plausible physiological reasoning behind, and is easy to implement than the previous spike train metrics. The spike train distance metric compares a modified version of the spike trains resulting from the convolution of the neural function with an exponentially decaying kernel. This kernel can be interpreted as the postsynaptic potentials in higher order-neurons for a small decay rate, while for a large decaying rate it can be interpreted as calcium-induced currents in the neuron membrane [Dayan and Abbot, 2001].

In this metric the neural response function  $\rho(t)$ , given by Eq. (B.7), is modified by replacing every spike, represented as a delta function, at times  $t_i$ ,  $i = 1, \dots, n$ , by the decaying exponential function:

$$h(t) = e^{-t/\tau} \mathbf{H}(t), \quad (5.19)$$

**Algorithm 5.2** Interspike interval metric,  $d^{\text{inter}}$ 

- 
- 1:  $\mathbf{t}_1 \leftarrow$  vector of spike time instants for train  $\rho_1$
  - 2:  $\mathbf{t}_2 \leftarrow$  vector of spike time instants for train  $\rho_2$
  - 3:  $n_1 \leftarrow$  number of elements in  $\mathbf{t}_1$ , i.e., number of spikes in train  $\rho_1$
  - 4:  $n_2 \leftarrow$  number of elements in  $\mathbf{t}_2$ , i.e. number of spikes in train  $\rho_2$
  - 5:  $q \leftarrow$  penalty to increase/decrease one second the interspike interval
  - 6:  $\mathbf{D} \leftarrow$  matrix of size  $[(n_1 + 1) \times (n_2 + 1)]$  where
    - first row:  $\mathbf{D}_{0,:} = [0, \dots, n_2]$
    - first column:  $\mathbf{D}_{:,0} = [0, \dots, n_1]^T$
  - 7:  $\{\mathbf{I}_1, \mathbf{I}_2\} \leftarrow$  vector of interspike intervals for trains  $\rho_1$  and  $\rho_2$ , respectively:
    - $\mathbf{I}_{1,1} \leftarrow \mathbf{t}_{1,1}; \quad \mathbf{I}_{2,1} \leftarrow \mathbf{t}_{2,1};$
    - $\mathbf{I}_{1,i} \leftarrow \mathbf{t}_{1,i} - \mathbf{t}_{1,i-1}, \quad \mathbf{I}_{2,j} \leftarrow \mathbf{t}_{2,j} - \mathbf{t}_{2,j-1}, \quad \forall_{i,j>1}$
  - 8: **for**  $i = 1$  to  $n_1$  **do** {For all spikes in train  $\rho_1$ ,  $\mathbf{t}_1$ }
  - 9:   **for**  $j = 1$  to  $n_2$  **do** {For all spikes in train  $\rho_2$ ,  $\mathbf{t}_2$ }
  - 10:      $\mathbf{D}_{i,j} = \min \left\{ \mathbf{D}_{i-1,j} + 1, \mathbf{D}_{i,j-1} + 1, \mathbf{D}_{i-1,j-1} + q|\mathbf{I}_{1,i} - \mathbf{I}_{2,j}| \right\}$
  - 11:   **end for**
  - 12: **end for**
  - 13: **return**  $d^{\text{inter}} \leftarrow \mathbf{D}_{n_1, n_2}$
- 

where the decay rate is given by  $\tau$ . The modified spike train becomes

$$\rho'(t) = \rho(t) * h(t) = \sum_{i=1}^n e^{-(t-t_i)/\tau} \mathbf{H}(t - t_i). \quad (5.20)$$

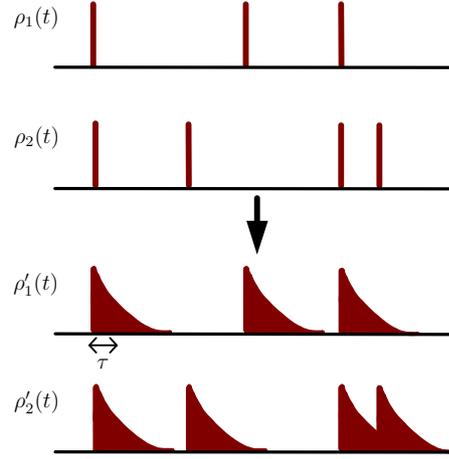
Figure 5.9 displays the schematic result of the convolution of spike trains with the exponentially decaying kernel. The distance between the two spike trains,  $\rho_1(t)$  and  $\rho_2(t)$ , is given by the expression:

$$d^2(\rho_1(t), \rho_2(t)) = \frac{1}{\tau} \int_0^{\infty} [\rho'_1(t) - \rho'_2(t)]^2 dt, \quad (5.21)$$

which corresponds to the Euclidean distance between the two filtered spike trains  $\rho'_1(t)$  and  $\rho'_2(t)$ . In this expression, the modified spike trains  $\rho'_1(t)$  and  $\rho'_2(t)$  are subtracted, and the square of those values is integrated in time.

This metric shows different behaviors depending on the values of the decay constant  $\tau$ . For a value of  $\tau$  much smaller than the **ISI**, the smoothed spike trains in Eq. (5.20) will contribute to the integral only if the spikes in both sequences are separated by more than  $\tau$ , and the metric works as a coincidence detector. In the case of continuous time spike trains, the number of coincident spikes can be neglected when  $\tau \rightarrow 0$ , so that the cross terms obtained by expanding the integrand in Eq. (5.21) can be neglected, and the expression for the distance becomes:

$$\lim_{\tau \rightarrow 0} d^2(\rho_1(t), \rho_2(t)) = \frac{1}{\tau} \int_0^{\infty} [\rho_1'^2(t) + \rho_2'^2(t)] dt = \frac{1}{2}(n_1 + n_2), \quad (5.22)$$



**Figure 5.9.** Spike waveform shape comparison with the spike distance metric.

where  $n_1$  and  $n_2$  are the number of spikes in the trains  $\rho_1(t)$  and  $\rho_2(t)$ , respectively. Thus, the metric just counts the number of noncoincident spikes. If the spike trains have  $c$  coincident spikes, the distance metric gives:

$$\lim_{\tau \rightarrow 0} d^2(\rho_1(t), \rho_2(t)) = \frac{1}{2}(n_1 + n_2 - 2c), \quad (5.23)$$

which is equal (excluding the multiplicative factor  $1/2$ ) to the result obtained with the spike time metric when a large cost for the spike shifting operation ( $d_{q \rightarrow \infty}^{\text{time}}$ ) is considered.

On the other hand, when a very large value for  $\tau$  ( $\tau \rightarrow \infty$ ) is used the main contribution to the integral in Eq. (5.21) comes from the times when the last spike has passed but the exponent has still not decayed. In this case, the metric can be approximated by:

$$\begin{aligned} \lim_{\tau \rightarrow \infty} d^2[\rho_1(t), \rho_2(t)] &= \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^{\infty} \left( \sum_{i=1}^{n_1} \text{H}(t - t_i) e^{-(t-t_i)/\tau} \right. \\ &\quad \left. - \sum_{j=1}^{n_2} \text{H}(t - t_j) e^{-(t-t_j)/\tau} \right)^2 dt \\ &= \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^{\infty} \left( n_1 e^{-t/\tau} - n_2 e^{-t/\tau} \right)^2 dt \\ &= \frac{1}{2}(n_1 - n_2)^2, \end{aligned} \quad (5.24)$$

where the second equality results from the fact that:

$$\lim_{\tau \rightarrow \infty} \sum_{i=1}^n \text{H}(t - t_i) e^{-(t-t_i)/\tau} = \sum_{i=1}^n \underbrace{\lim_{\tau \rightarrow \infty} \text{H}(t - t_i) e^{t_i/\tau}}_{=1} e^{-t/\tau}. \quad (5.25)$$

## 5. Neural Activity Metrics

The expressions for some special operations on the spike trains can also be derived, like the insertion (or deletion) of a spike, and the shift of a spike. If the spike train  $\rho_2(t)$  differs from  $\rho_1(t)$  just for a spike placed at time  $t_i$ , such that their convolved versions are related by:

$$\rho_2'(t) = \rho_1'(t) + \text{H}(t - t_i) e^{-(t-t_i)/\tau}, \quad (5.26)$$

the distance metric gives

$$d^2(\rho_1(t), \rho_2(t)) = \frac{1}{\tau} \int_{t_i}^{\infty} e^{-2(t-t_i)/\tau} dt = \frac{1}{2}, \quad (5.27)$$

and the deletion, or removal, of a spike produces the same value that is independent of the decay constant  $\tau$ .

For two spike trains whose only difference is the shift of a spike from  $t_i$  to  $t_i + \Delta t$  in train  $\rho_2(t)$  relative to  $\rho_1(t)$ , such that the convolved spike trains are related by:

$$\rho_2'(t) = \rho_1'(t) - \text{H}(t - t_i) e^{-(t-t_i)/\tau} + \text{H}(t - t_i - \Delta t) e^{-(t-t_i-\Delta t)/\tau}, \quad (5.28)$$

the distance between these spike trains is

$$\begin{aligned} d^2(\rho_1(t), \rho_2(t)) &= \frac{1}{\tau} \int_{t_i}^{t_i+\Delta t} e^{-2(t-t_i)/\tau} dt \\ &\quad + \frac{1}{\tau} \int_{t_i+\Delta t}^{\infty} \left[ e^{-(t-t_i)/\tau} - e^{-(t-t_i-\Delta t)/\tau} \right]^2 dt \\ &= 1 - e^{-|\Delta t|/\tau}, \end{aligned} \quad (5.29)$$

which approaches the value one for a large distance  $\Delta t$  between spikes compared to the decay  $\tau$ , and zero otherwise. Analytical expressions can be derived for other particular relations between the spike trains under analysis [van Rossum, 2001; Tomás and Sousa, 2008]. An interesting case considers the distance between two uncorrelated homogeneous Poisson spike trains generated with the same constant firing rate  $r$ . According to Eq. (5.22) the distance between two spike trains approaches  $(n_1 + n_2)/2$  for a small  $\tau$ , where  $n_1$  and  $n_2$  are the number of spikes in each trial. However, as stated by Eq. (B.77), the number of expected spikes in a Poisson trial with a constant firing rate  $r$  and a time duration  $T$  is equal to  $rT$ , so that

$$d_{\tau \rightarrow 0}^2(\rho_1(t), \rho_2(t)) = rT. \quad (5.30)$$

On the other hand, for large  $\tau$ , the distance values tend towards  $(n_1 - n_2)^2/2$ , as stated by Eq. (5.24). The expectation value  $\langle (n_1 - n_2)^2 \rangle / 2 = \langle n_1^2 \rangle / 2 + \langle n_2^2 \rangle / 2 - \langle n_1 n_2 \rangle$  for a Poisson process can be calculated with the help of Eq. (B.74) and Eq. (B.75). Thus,  $\langle n_1^2 \rangle = \langle n_2^2 \rangle = rT + (rT)^2$  and  $\langle n_1 n_2 \rangle = (rT)^2$ , which results in

$$d_{\tau \rightarrow \infty}^2(\rho_1(t), \rho_2(t)) = rT. \quad (5.31)$$

Spike Train Metrics				
	Spike Time Metric		Interspike Interval Metric	
$q$	min	max	min	max
$q = 0$	0	9	0	9
$q = \infty$	132	165	126	147

**Table 5.2.** Limit values for the spike train metrics using the neuronal responses of a ON-type salamander RGC.

From Eq. (5.30) and Eq. (5.31) it can be observed that the average value for the distance metric is equal for small and for large values of  $\tau$  in the case of a spike train described by a homogeneous Poisson process with mean firing rate  $r$ .

This metric can be computed in discrete time by convolving the involved discrete spike trains,  $\rho_1[n]$  and  $\rho_2[n]$ , with the discrete kernel

$$h[n] = e^{-nT_s/\tau} H[n], \quad (5.32)$$

where  $H[n]$  is the discrete Heaviside unit step function and  $T_s$  is the sampling period. The convolved spike trains are subtracted from each other, and the result is squared and summed (the discrete equivalent to integration). The final expression for the discrete spike train distance is:

$$d^2(\rho_1[n], \rho_2[n]) = \frac{1}{\tau} \sum_{i=0}^N (\rho'_1[i] - \rho'_2[i])^2 T_s, \quad (5.33)$$

where  $\rho'[n] = \rho[n] * h[n]$  and  $N$  is the length of the discretized spike trains.

#### 5.4.4. Spike Train Metrics Analysis

The spike train metrics compare two spike trains directly. The spike time metric and the spike interval metric were applied to classify a set of neural responses, namely with the objective of finding the metric that aggregates the neural responses in a more compact subspace of the response space [Victor and Purpura, 1996]. The metrics gave different results and it was noted by its proponents that the spike time metric and the interspike interval metric do not refine each other in the topological sense, meaning that the sequence of successive steps that minimize the distance between two spike trains according to one metric can have the opposite effect with respect to the other metric [Victor and Purpura, 1997].

Figure 5.10 presents two graphics that display the evolution of the spike time metric and of the interspike interval metric as a function of  $q$  by doing the inter comparison of the twelve trials displayed in Fig. B.6b. Table 5.2 shows the values for the maxima and minima for the limit values of  $q$ . For  $q = 0$  the minimum and maximum values are equal for both metrics, and both metrics reduce to the spike count metric –  $d_{q=0}^{\text{time}}(\rho_1, \rho_2) = d_{q=0}^{\text{inter}}(\rho_1, \rho_2) = d^{\text{count}}(\rho_1, \rho_2)$ . Both metrics have a minimum zero value for  $q = 0$  since there are several spike trains within the set of analyzed trials with

## 5. Neural Activity Metrics

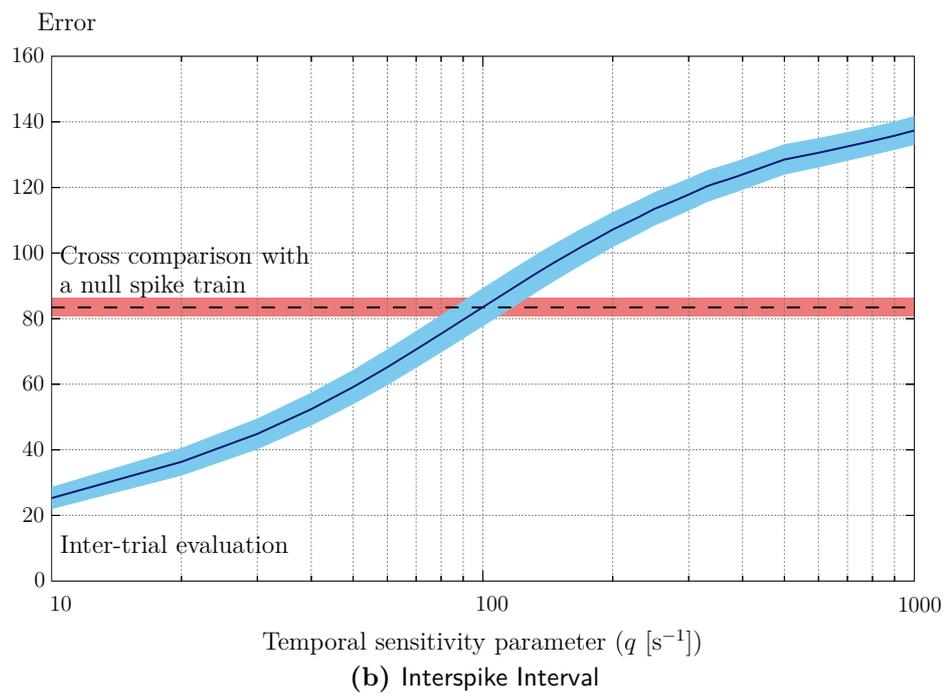
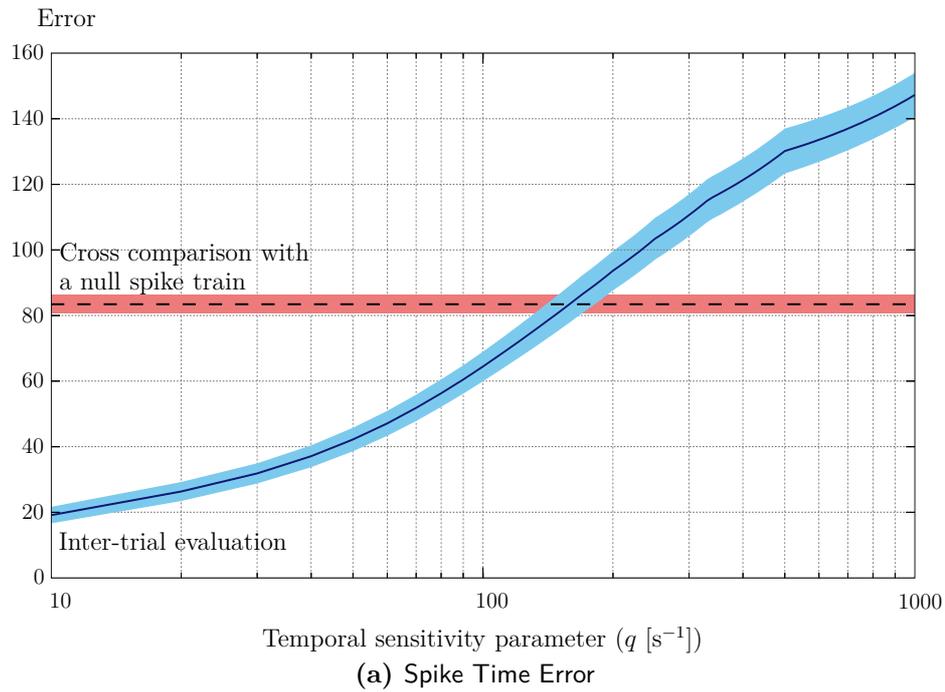
the same number of spikes. From the values for the maximum one can observe that two spike trains can differ by 9 spikes at most, in terms of the number of spikes. For  $q \rightarrow \infty$ , the values for the spike time metric indicate that there are, at the minimum, 132 and at the maximum 165 non-coincident spikes in the spike trains. For the interspike interval metric, the values for  $q \rightarrow \infty$  mean that the spike trains under comparison possess 126 non-coincident spike intervals at minimum, and 147 at maximum.

Figure 5.10 shows the evolution of the spike time metric,  $d^{\text{time}}$ , and of the interspike interval metric,  $d^{\text{int}}$ , versus the shifting cost  $q$ . The curves in Fig. 5.10 were obtained by performing an inter-trial evaluation of the salamander type-ON RGC responses, i.e., by comparing all 66 pairs of different spike trains and extracting the mean value and the standard deviation of the error. For  $q \geq 2/T_s$ , where  $T_s$  is the sampling period ( $T_s = 1$  ms in this case), the distance becomes constant because the cost of shifting a spike to the next time bin is bigger than deleting it and raising a new one in the correct time bin. For small values of  $q$ , the cost tends to the difference in the number of spikes in the two trains.

If the distance between one of the trials from the ON salamander RGC and another spike train with no spikes (a null spike train) is calculated, as illustrated in Fig. 5.10, an insightful result is obtained: for large values of  $q$ , the null spike train is closer to the RGC neural responses than the two response trials of the same cell for the same stimulus are. This result sheds some light onto the limit values that must be used for the parameter  $q$  by defining its maximum value, i.e., the maximum temporal sensitivity for the metrics.

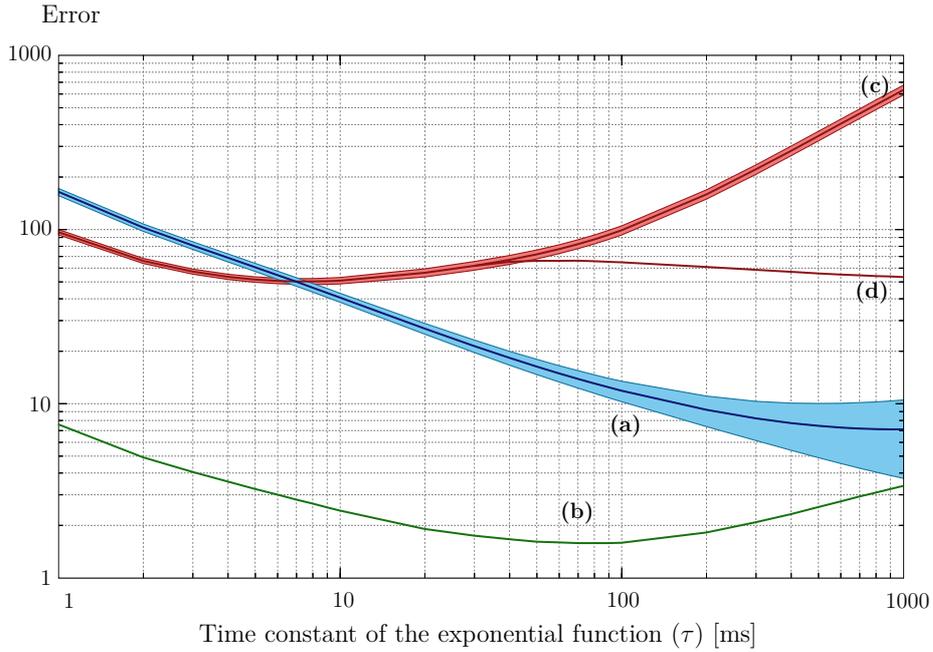
Generally, since not all cells present the same temporal sensitivity, the parameter  $q$  must be chosen according to the neural responses. A method for picking the value of  $q$  is to trace the curves illustrated in Fig. 5.10 for a given set of neural responses. For the case of the salamander type-ON RGC responses, these values are  $q_{\text{max}} \approx 110 \text{ s}^{-1}$  for the spike time, and  $q_{\text{max}} \approx 70 \text{ s}^{-1}$  for the interspike interval, which define a maximum spike/interspike interval displacement of  $1/q_{\text{max}} \approx 9$  ms and  $1/q_{\text{max}} \approx 14$  ms, respectively.

As for the case of the spike time and interspike interval metrics, by performing an inter-trial evaluation of a cell's responses, a suitable range of values to be used for the parameter  $\tau$  of the spike train distance metric,  $d^2(\cdot)$ , can be found. Fig. 5.11 shows several curves as a function of  $\tau$ , namely: *a)* the spike train distance error when applied to the inter-trial comparison of the salamander RGCs responses; *b)* the variance of the inter-trial comparison; *c)* the comparison of the RGCs responses with a null spike train; and *d)* the comparison of the RGCs responses with a spike train generated by a homogeneous Poisson process (i.e., with constant firing rate). In the last case, the firing rate was defined for each value of  $\tau$  as the one leading to the minimum error. Comparing the curves in Fig. 5.11, one can see that the maximum temporal precision of the cell's responses occurs at  $\tau = 8$  ms, so that  $\tau$  should be larger than 8 ms for this cell; otherwise the comparison with the null spike train will give better results. Moreover, the minimum standard deviation of the inter-trial comparison is achieved at approximately  $\tau = 80$  ms. As in the previous cases,



**Figure 5.10.** Inter-trial evaluation of salamander RGC responses for varying temporal sensitivities (line – mean value; shaded – mean  $\pm$  standard deviation).

## 5. Neural Activity Metrics

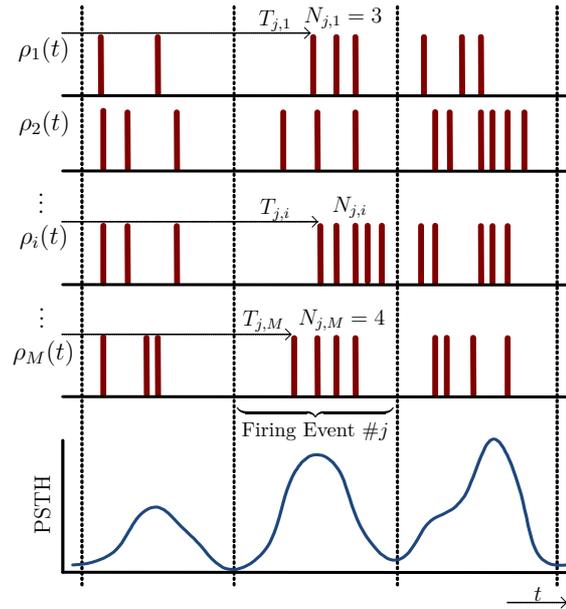


**Figure 5.11.** Inter-trial evaluation of salamander RGC responses with the spike distance metric as a function of the temporal sensitivity; (line – mean value; shaded area – mean  $\pm$  standard deviation): (a) inter-trial evaluation of cell responses; (b) standard deviation of the inter-trial evaluation of cell responses; (c) cross-comparison with a null spike train; (d) cross-comparison with the best constant rate model.

this analysis allows the definition of a range of values for the precision parameter:  $\tau \in [8; 100]$  ms. It should be noted that the superior limit of the range is also defined by the fact that the spike train distance operates by measuring the mean squared error of the low-pass filtered spike trains. Thus, for large values of  $\tau$ , high-frequency information is lost and only the average firing rate over time is measured; it will therefore lead to apparent small errors.

This section presented a method to determine the range of values for the parameter  $q$  for the spike time and the interspike interval metrics, and the parameter  $\tau$  for the spike train distance metric. However, the obtained values are not fixed for all neural responses, as some cells exhibit higher temporal precision than others do. Thus, in general, this method should be applied to the target neural responses before applying the neural metrics to assess models and responses.

The computational implementation of the spike train distance metric ( $d^2$ ) is straightforward. A difficulty of this metric is the correct choice of the right  $\tau$ , due to the fact that it scales the kernel in Eq. (5.32). For  $\tau \ll T_s$  the discrete kernel in Eq. (5.32) will always be equal to the discrete sequence  $[1 \ 0 \ 0 \ \dots]$ ; and the convolution will leave the discrete neuronal response function  $\rho[n]$  unchanged, while the value of the numeric integration is divided by  $\tau$  (see Eq. (5.33)), leading to erroneous values that can increase unboundedly for small  $\tau$ . Thus,  $\tau$  must be several times bigger than the sampling period  $T_s$ , so that the numerical integration expressed by Eq. (5.33)



**Figure 5.12.** Parsing a set of spike trains into firing events.

gives meaningful results. For values of  $\tau \rightarrow \infty$  the values of the metric should tend to Eq. (5.24), however this tendency is very slow, and for big values of  $\tau$  the kernel sequence will be quite long; the calculations will become lengthy, and the memory requirements can become prohibitive.

## 5.5. Spike Events Metrics

The *spike events metric*,  $d^{events}(\cdot)$ , has been proposed to measure the distance between two sets of spikes trains: a reference and an estimated set [Keat et al., 2001]. This metric starts by defining the firing events on each set of spike trains and employs the matching principle used in [Victor and Purpura, 1996] to match the events.

This metric is intended to compare sets of spike trains corresponding to the responses of a neuron to the same stimulus – the set of reference trials – with a second set of spike trains – the predicted set of trials – predicted by a model, for example. The comparison is made in terms of firing events, where a firing event is obtained by grouping identical bursts of spikes within the spike trains. Figure 5.12 represents a set of  $M$  spike trains with the events classification.

The events metric is based on the assumption that the neuron responds, within a certain range, with similar spike trains to the same stimulus. These spike trains are characterized by regions where no spikes are fired, followed by a burst of spikes – a spike event – in response to stimuli (see Fig. 5.3). Moreover, it was observed, particularly in the case of the retina, that these trains are reproducible from trial to trial, not only in terms of the time occurrence of the spikes but also regarding the number of spikes [Reinagel, 2001]. The spike train metrics described previously were

## 5. Neural Activity Metrics

found to be unsuited to measure the reliability of such trains. Therefore, the spike events metric takes into account: the time occurrence of spikes, the number of spikes, the variation in the time occurrence of each spike in the event, and the variation of the number of spikes in each event.

The computation of the spike events metric starts by parsing the spike trains into firing events (see Fig. 5.12), and then by measuring the cost of matching the events between the two sets. Firing events correspond to bursts of spikes delimited by regions where the firing rate is nearly zero. To compute the boundary values between firing events, the PSTH is obtained from the set of spike trains. This PSTH is usually smoothed by convolving it with a Gaussian filter whose width,  $\sigma$ , is defined as the time scale of modulations in the firing rate. Specifically, the value of  $\sigma$  is obtained by adjusting a Gaussian function to the histogram of time differences between all pairs of spikes trains within a trial. The interspike interval histogram can be obtained by correlating all pairs of trials from the set (see Sec. B.1.4). The value of  $\sigma$  is made equal to the width of the Gaussian curve fitted to the interspike time histogram divided by  $\sqrt{2}$ , because both spikes from the two spike trains contribute to the interspike time jitter [Berry et al., 1997]. Algorithm 5.3 details the steps to compute the width of the smoothing filter.

Figure 5.13a shows a segment with a 2 s duration of a set of 13 trials of a rabbit RGC when excited by ON-OFF type stimuli (see Fig. B.5a), and Fig. 5.13b shows the real and the smoothed PSTH. The minima of the smoothed PSTH are calculated and they correspond to the locations where the firing event boundaries should be positioned. The spikes between a pair of consecutive such minima are considered to be part of the same firing event. Since not all minima are equal to zero, a rule must be used to distinguish between real firing event boundaries and just a local decrease in the firing rate. Therefore, a given minimum  $m_i$ , between two consecutive maxima  $M_i$  and  $M_{i+1}$  in the PSTH, is considered to be a firing event boundary if it is smaller than its neighboring maxima according to the rule:

$$\frac{\sqrt{M_i M_{i+1}}}{m_i} \geq \phi. \quad (5.34)$$

If the condition in Eq. (5.34) is verified, then  $m_i$  is considered a true minimum that identifies a boundary between two adjacent firing events; otherwise, the two initial firing events are, in fact, a single event. The original values used for  $\phi$  are  $\phi \in \{1.5, 3\}$ , but experimental results have shown that the chosen value does not have too much influence in the result [Keat et al., 2001]. Figure 5.13b shows the firing events obtained by the application of the criteria decision expressed by Eq. (5.34) with  $\phi = 3$ . After being defined, each firing event  $j$  from every response  $i$  is characterized by two numbers: the time of the first spike,  $T_{j,i}$ ; and the number of spikes,  $N_{j,i}$  (see Fig. 5.12). Four quantities are calculated for each firing event, in order to obtain the correspondence between the firing events of the two sets of trials. The event  $j$  from a set comprising  $M$  trials (refer to Fig. 5.12) is characterized by:

- the average across trials of the time occurrence of the first spike within the

**Algorithm 5.3** Gaussian filter width  $\sigma$ , GaussFilterWidth

---

```

1:  $T \leftarrow$  sampling period
2:  $M \leftarrow$  number of spike trains
3:  $\rho_{M,:} \leftarrow$  matrix whose rows are the discretized reference spike trains
   {Accumulate the cross-correlation between all trials of spikes for a maximum lag
    $l_{max}$ .}
4: for  $i = 1$  to  $M - 1$  do {For all trials in  $\rho$ }
5:   for  $j = i$  to  $M$  do {For trials in  $\rho$  not equal to  $\rho_{i,:}$ }
6:      $\mathbf{c} = \mathbf{c} + \text{corr}(\rho_{i,:}, \rho_{j,:})$  {correlate the spike trains}
7:   end for
8: end for { $\mathbf{c}$  as dimensions  $2l_{max} + 1$ }

   {Take into account only the main lobe of the histogram  $\mathbf{c}$ }
9:  $m \leftarrow \mathbf{c}_1$ 
10: for  $i = 1$  to  $l_{max}$  do
11:   if  $m < \mathbf{c}_i$  then
12:      $l = i; m = \mathbf{c}_i$ 
13:   end if
14: end for
15:  $\mathbf{c} \leftarrow \mathbf{c}_l$  to  $\mathbf{c}_{2l+1}$ 

   {Calculate the mean of the histogram:  $\mu$ }
16:  $\zeta \leftarrow 0; n \leftarrow 0$ 
17: for  $i = 1$  to  $2l + 1$  do
18:    $\zeta = \zeta + \mathbf{c}_i(i - (l + 1))T$ 
19:    $n = n + \mathbf{c}_i$ 
20: end for
21:  $\mu = \frac{\zeta}{n}$ 

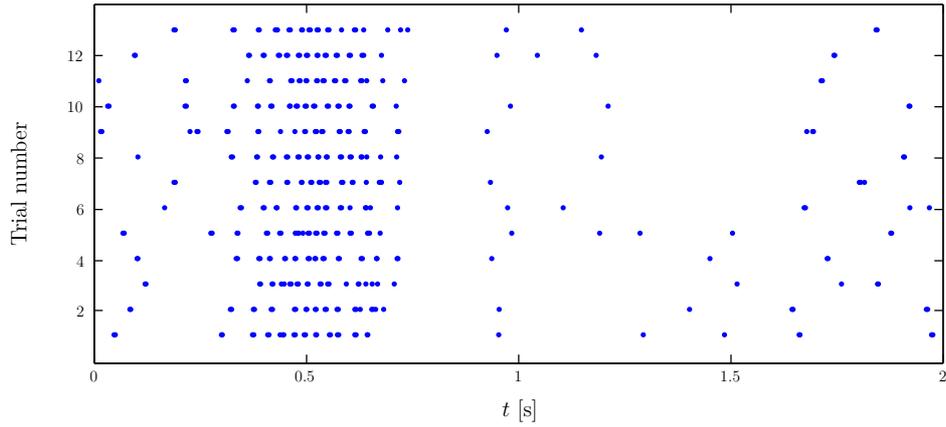
   {Variance calculation:  $\sigma^2$ }
22:  $\epsilon \leftarrow 0$ 
23: for  $i = 1$  to  $2l + 1$  do
24:    $\epsilon = \epsilon + (\mathbf{c}_i(i - (l + 1))T)^2$ 
25: end for
26:  $\sigma^2 = \frac{\epsilon}{n} - \mu^2$ 
27:  $\sigma = \sqrt{\sigma^2}$  //Standard deviation

   {Finally, taking into account that the contribution to the interspike interval is
   due to spikes from both trials}
28:  $\sigma = \frac{\sigma}{\sqrt{2}}$ 
29: return  $\sigma$ 

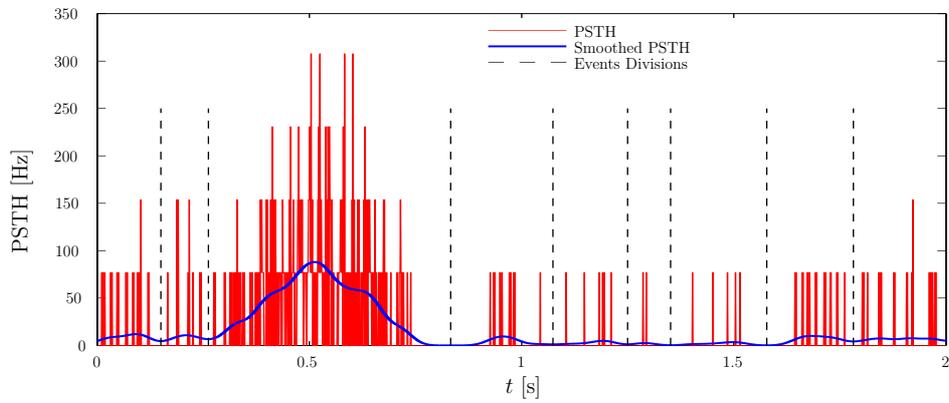
```

---

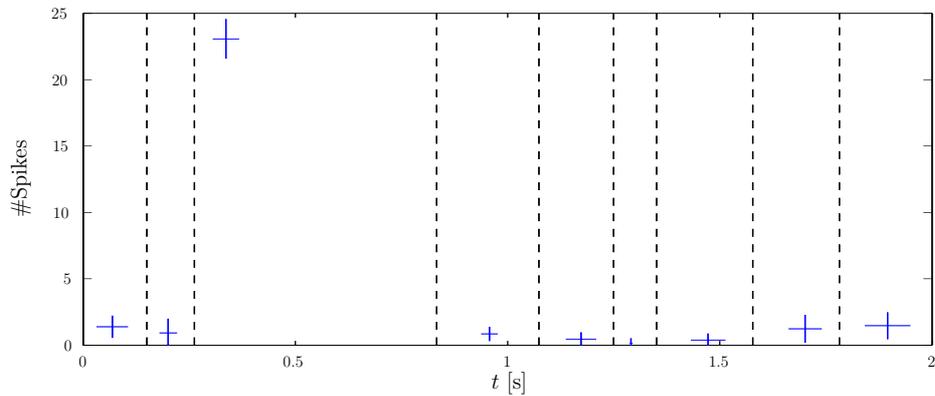
## 5. Neural Activity Metrics



(a) Trains of spikes from a rabbit ON-type cell. Each spike is represented by a dot.



(b) Real and smoothed PSTH with the firing events divisions.



(c) Firing events characterization: the values of the cross center  $(x, y)$  correspond to the mean time of first spike,  $T$ , and to mean number of spikes,  $N$ , respectively; the cross width represents the standard deviation of the time occurrence of the first spike,  $\delta T$ , while its height is the standard deviation of the number of spikes,  $\delta N$ .

**Figure 5.13.** Characterization of 13 observed trials into firing events.

firing event:

$$T_j = \frac{1}{M} \sum_{i=1}^M T_{ji} ; \quad (5.35)$$

- the average number of spikes across trials in the event:

$$N_j = \frac{1}{M} \sum_{i=1}^M N_{ji} ; \quad (5.36)$$

- the standard deviation of the time occurrence of the first spike, in an event, across trials, which measures the time jitter of the first spike in the event:

$$\delta T_j = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (T_{ji} - T_j)^2} ; \quad (5.37)$$

- the standard deviation of the number of spikes across the trials within a firing event, which measures the deviation from the mean spike number:

$$\delta N_j = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (N_{ji} - N_j)^2} . \quad (5.38)$$

This procedure maps the set of spike trains into a sequence of firing events – a train of events – where each event is characterized by the four quantities:  $T$ ,  $N$ ,  $\delta T$ , and  $\delta N$ . Figure 5.13c shows the characterization of the firing events from the rabbit ON-type RGC responses of Fig. 5.13a. The algorithm described in Algorithm 5.4 describes the steps followed to parse a set of trials into a sequence of firing events. The next step in computing the spike event metric is to match the firing events from the observed retina responses with the predicted neural responses.

If the firing events of the observed set are ordered according to index  $i$  and the ones from the estimated set according to index  $j$ , the overall expression for the error between the two event trains is given by

$$d^{event} = e_T E_T + e_N E_N + e_{\delta T} E_{\delta T} + e_{\delta N} E_{\delta N} - e_M E_M . \quad (5.39)$$

This distance measure includes five sources of error between the observed and predicted spike trains, namely:  $E_T$ ,  $E_N$ ,  $E_{\delta T}$ ,  $E_{\delta N}$ ,  $E_M$ . The error terms in Eq. (5.39)

## 5. Neural Activity Metrics

---

### Algorithm 5.4 Parsing spike trains to events, Spike2Events

---

- 1:  $\boldsymbol{\rho} \leftarrow$  matrix whose rows are the spike trains  
 {Compute:
    - The firing rate  $\mathbf{r}$  from  $\boldsymbol{\rho}_{M,N}$
    - Smooth  $\mathbf{r}$  with a Gaussian filter with a width equal to  $\sigma \leftarrow GaussFilterWidth(\boldsymbol{\rho})$
    - Compute the minima and maxima of the rate  $r$ ; every minimum  $m_i$  is delimited by the maxima  $M_i$  and  $M_{i+1}$ , such that the following conditions are true:
      - $m_i \leq r(k)$ , for all  $\forall_{N_i < k < N_{i+1}}$ , where  $N_i$  is the timestamp of the maxima  $M_i$
      - $M_i \geq r(k)$ , for all  $\forall_{n_{i-1} < k < n_i}$ , where  $n_i$  is the timestamp of the minima  $m_i$
      - $M_i \geq m_i \quad \wedge \quad M_{i+1} \geq m_i$
  - 2:  $(\mathbf{m}, \mathbf{n}) \leftarrow$  vectors of values and time stamps of the minima of rate  $r$ , i.e.  $\mathbf{m} = r(\mathbf{n})$
  - 3:  $(\mathbf{M}, \mathbf{N}) \leftarrow$  vectors of values and time stamps of the maxima of rate  $r$ , i.e.,  $\mathbf{M} = r(\mathbf{N})$  {Remove fake minima according to criteria of Eq. (5.34) or that leads to an event with no spikes}
  - 4: **for all**  $m_i \in \mathbf{m}$  **do**
  - 5:   **if**  $\sqrt{M_i M_{i+1}}/m_i < \phi$  **then**
  - 6:     Remove minimum  $m_i$  and its index  $n_i$  from the vectors  $\mathbf{m}$  and  $\mathbf{n}$ , respectively
  - 7:     Remove the smallest of the maxima  $\{M_i, M_{i+1}\}$  and its index from the vectors  $\mathbf{M}$  and  $\mathbf{N}$
  - 8:   **end if**
  - 9: **end for**  
 {Compute event statistics}
  - 10: **for all**  $m_i \in \mathbf{m}$  **do**
  - 11:    $\mathcal{T}_i \leftarrow$  average time stamp of the first spike in the interval  $[n_{i-1} \ n_i]$
  - 12:    $\mathcal{N}_i \leftarrow$  average number of spikes in the interval  $[n_{i-1} \ n_i]$
  - 13:    $\delta\mathcal{T}_i \leftarrow$  standard deviation in the time of first spike in  $[n_{i-1} \ n_i]$
  - 14:    $\delta\mathcal{N}_i \leftarrow$  standard deviation in the number of spikes in  $[n_{i-1} \ n_i]$
  - 15: **end for**
  - 16: **return**  $\mathcal{T}, \mathcal{N}, \delta\mathcal{T}, \delta\mathcal{N}$
- 

are defined as:

$$E_T = \sum_{\substack{\text{matched} \\ \text{event pairs } (i,j)}} |T_i - \hat{T}_j|; \quad (5.40)$$

$$E_N = \sum_{\substack{\text{matched} \\ \text{event pairs } (i,j)}} |N_i - \hat{N}_j| + \sum_{\substack{\text{unmatched} \\ \text{events } i}} N_i + \sum_{\substack{\text{unmatched} \\ \text{events } j}} \hat{N}_j; \quad (5.41)$$

150

$$E_{\delta T} = \sum_{\substack{\text{matched} \\ \text{event pairs } (i,j)}} |\delta T_i - \delta \hat{T}_j|; \quad (5.42)$$

$$E_{\delta N} = \sum_{\substack{\text{matched} \\ \text{event pairs } (i,j)}} |\delta N_i - \delta \hat{N}_j|; \quad (5.43)$$

The hat over the quantities in Eq. (5.40) – Eq. (5.44) refers to the values obtained from the estimated firing events, while the quantities without a hat are related to the observed firing events from a real neuron cell. The expression in Eq. (5.44) indicates that the matching of two events between sequences is rewarded by adding a negative contribution to the error in Eq. (5.39).

The weights for each error component in Eq. (5.39) are given by:

$$\begin{aligned}
 e_T &= \frac{1}{E\{\delta T\}} ; \\
 e_N &= \frac{1}{E\{\delta N\}} ; \\
 e_{\delta T} &= \frac{1}{2}e_T ; \\
 e_{\delta N} &= \frac{1}{2}e_N ; \\
 e_M &= 2;
 \end{aligned} \tag{5.45}$$

where the averages are calculated across all events in the reference event sequence. That is, if we have a total of  $Q$  events in the reference set of trials, these values are:

$$E\{\delta T\} = \frac{1}{Q} \sum_{j=1}^Q \delta T_j ; \quad E\{\delta N\} = \frac{1}{Q} \sum_{j=1}^Q \delta N_j. \tag{5.46}$$

The mean of the standard deviation of the first spike occurrence within the trials ( $E\{\delta T\}$ ) is used to scale the error related to the time jitter differences between the neurons and the predicted trials, while the mean of the standard deviation of the number of spikes ( $E\{\delta N\}$ ) plays the same role by scaling the differences in the number of spikes. The coefficients  $e_{\delta T}$  and  $e_{\delta N}$  are equal to half of  $e_T$  and  $e_N$ , respectively, meaning that the spike occurrence time and the spike number are twice as important to the error measurement as their variation. The value of  $e_M$  is constant and gives a negative contribution to the error in order to reward matching two events.

To match the events a recursive procedure similar to the one described to match different spike trains in the spike time metric (and in the interspike interval metric) is employed. To obtain a recursive procedure, the restriction for the alignment of two firing events sequences apply: two events in one train cannot be matched to two events in the other train that occur in reverse order. Representing the sequence of firing events from the reference trials by  $R$ , which possesses  $n_1$  events, and the sequence of firing events from the estimated trials by  $\hat{R}$ , which has  $n_2$  events, the number of possible matches is limited. Matching the event  $i$  from  $R$  with the event  $j$  from  $\hat{R}$ , one of three cases can occur: *i*) the last analyzed event in  $R$  is unmatched; *ii*) the last analyzed event in  $\hat{R}$  is unmatched; or *iii*) the last events in  $R$  and  $\hat{R}$  match each other. If  $d_{i,j}^{events}$  represents the error incurred in matching the first  $i$  events of  $R$  with the first  $j$  events of  $\hat{R}$ , taking into account the previous restrictions, the

## 5. Neural Activity Metrics

events can be matched iteratively. Each one of the previous three possibilities leads to three different values for the matching error, and the smallest value should be chosen to match event  $i$  with event  $j$  according to:

$$d_{i,j}^{events} = \min \left\{ d_{i-1,j}^{events} + e_N N_i ; d_{i,j-1}^{events} + e_N \hat{N}_j ; d_{i-1,j-1}^{events} + M_{i,j} \right\}, \quad (5.47)$$

where the cost  $M_{i,j}$  for matching events  $i$  and  $j$  is obtained from the expression:

$$M_{i,j} = e_T |T_i - \hat{T}_j| + e_N |N_i - \hat{N}_j| + e_{\delta T} |\delta T_i - \delta \hat{T}_j| + e_{\delta N} |\delta N_i - \delta \hat{N}_j| - e_M. \quad (5.48)$$

Based on Eq. (5.47), a recursive procedure can be initiated that will iteratively match the events from the two sets. By starting with  $d_{0,0}^{events} = 0$ , a matrix can be filled for the match of the first  $i$  events of one train to the first  $j$  events of the other train by recursively calculating  $d_{i,j}^{events}$ . The last element in the diagonal contains the total error given by Eq. (5.39). Algorithm 5.5 presents the sequence of operations to match the events from the reference sequence with the ones from the observed sequence.

Actually, not all terms of the matrix need to be calculated, since for two events far apart in time, the error  $M_{i,j}$  is so large that the two events will never match as pointed out in [Keat et al., 2001]. Specifically, events that occur farther apart than the following condition need not to be calculated:

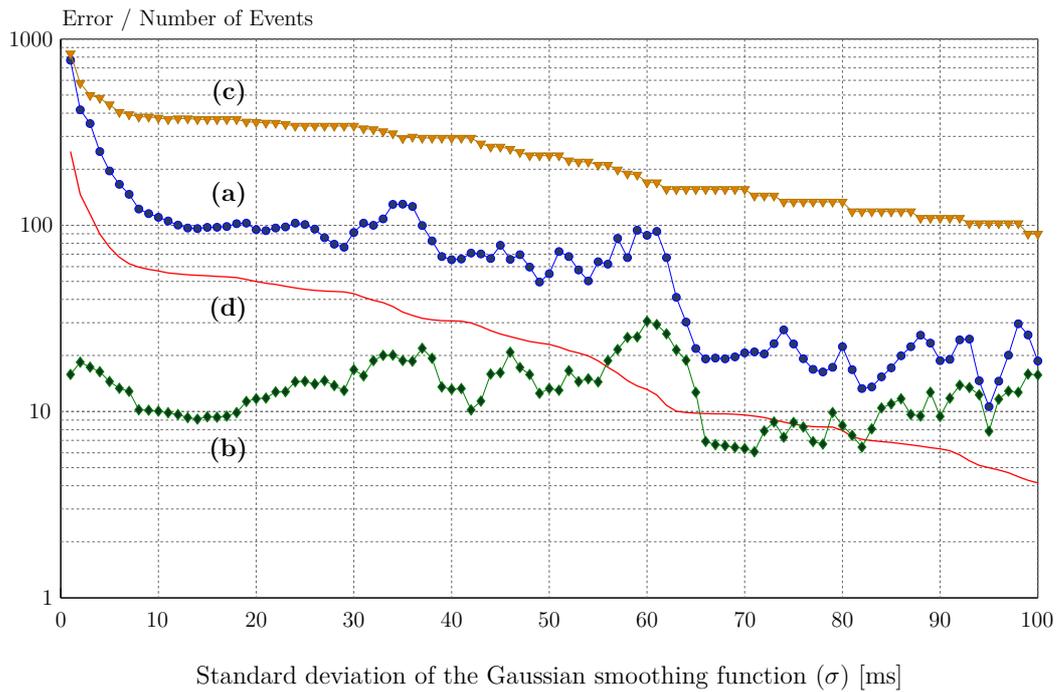
$$|T_i - \hat{T}_j| > \frac{1}{e_T} (2e_N N_{max} + e_M), \quad (5.49)$$

where  $N_{max}$  is the largest number of spikes in an event from the event sequences under comparison.

### 5.5.1. Spike Events Metric Analysis

The firing events metric compares trains of events. The matching of two events gives a negative contribution to the overall error (see Eq. (5.44) and Eq. (5.48)), so that the distance between a set of spike trains and itself is  $-Me_M \leq 0$ , where  $M$  is the number of identified events. While this result violates the typical non-negativity property of general metrics, an offset value equal to  $Me_M$  can be added (which violates the interchangeability property in this case).

The delimitation of events is a critical step that goes through the calculation of the minima of the smoothed PSTH. A real maximum can exist, or there can be a simple transition to/from a constant firing rate. Therefore, extra care must be taken in these cases. After calculating the PSTH minima, the events should be identified using Eq. (5.34); this avoids the use of barely pronounced minima to generate an event division. Even taking these cases into consideration, the metric is very sensitive to the identification of events; with small variations, such as joining two events into one, the metric can significantly vary its result. To demonstrate this the parameter  $\sigma$  can be varied, which corresponds to the standard deviation of the Gaussian function used to smooth the PSTH, and perform an inter-trial comparison of the responses of



**Figure 5.14.** Inter-trial evaluation of salamander *RGC* responses varying the standard deviation values of the Gaussian smoothing function: **(a)** average value of the inter-trial evaluation of cell responses; **(b)** standard deviation of the inter-trial evaluation of cell responses; **(c)** cross-comparison with a null spike train; **(d)** number of identified events.

**Algorithm 5.5** Spike events metric,  $d^{events}$ 

- 
- 1:  $\rho(t) \leftarrow$  set of observed spike trains
  - 2:  $\hat{\rho}(t) \leftarrow$  set of estimated spike trains  
 {Compute event statistics, as described in Algorithm 5.4; the function results in four vectors for each set of trials}
  - 3:  $\{\mathcal{T}, \mathcal{N}, \delta\mathcal{T}, \delta\mathcal{N}\} \leftarrow$  Spike2Events( $\rho$ )
  - 4:  $\{\hat{\mathcal{T}}, \hat{\mathcal{N}}, \delta\hat{\mathcal{T}}, \delta\hat{\mathcal{N}}\} \leftarrow$  Spike2Events( $\hat{\rho}$ )
  - 5:  $n_1 \leftarrow$  number of events in  $\rho(t)$ , i.e., number of elements in the vectors  $\mathcal{T}$  (and in the other statistics' vector)
  - 6:  $n_2 \leftarrow$  number of events in  $\hat{\rho}(t)$ , i.e., number of elements in the vectors  $\hat{\mathcal{T}}$  (like in the other statistics)
  - {Compute weights for each penalty as stated in Eq. (5.45)}
  - 7:  $e_T \leftarrow 1/E\{\delta\mathcal{T}\}$ ;  $e_N \leftarrow 1/E\{\delta\mathcal{N}\}$ ;
  - 8:  $e_{\delta T} \leftarrow 1/(2E\{\delta\mathcal{T}\})$ ;  $e_{\delta N} \leftarrow 1/(2E\{\delta\mathcal{N}\})$ ;  $e_M \leftarrow 2$   
 {Compute the distance between the first  $i$  events of  $\rho(t)$  and the first  $j$  events of  $\hat{\rho}(t)$ }
  - 9:  $\mathbf{D} \leftarrow$  matrix of size  $[(n_1 + 1) \times (n_2 + 1)]$  where
    - first row:  $\mathbf{D}_{0,:} = [0, e_N \hat{\mathcal{T}}^T]$
    - first column:  $\mathbf{D}_{:,0} = [0, e_N \mathcal{T}^T]$
  - 10: **for**  $i = 1$  to  $n_1$  **do** {For all events of  $\rho$ }
  - 11:   **for**  $j = 1$  to  $n_2$  **do** {For all events of  $\hat{\rho}$ }
  - 12:      $M = e_T |\mathcal{T}_i - \hat{\mathcal{T}}_j| + e_N |\mathcal{N}_i - \hat{\mathcal{N}}_j| + e_{\delta T} |\delta\mathcal{T}_i - \delta\hat{\mathcal{T}}_j| + e_{\delta N} |\delta\mathcal{N}_i - \delta\hat{\mathcal{N}}_j| - e_M$
  - 13:      $\mathbf{D}_{i,j} = \min \left\{ \mathbf{D}_{i-1,j} + e_N N_i; \mathbf{D}_{i,j-1} + e_N \hat{N}_j, \mathbf{D}_{i-1,j-1} + M \right\}$
  - 14:   **end for**
  - 15: **end for**
  - 16: **return**  $d^{events} \leftarrow \mathbf{D}_{n_1, n_2}$
- 

the salamander retinal ganglion cell. It should be noticed that the use of different  $\sigma$  values leads to variations in the number of identified events. For the inter-trial evaluation values presented in Fig. 5.14,  $Me_M$  has been added to the metric result; this shifts the baseline of comparison from  $-Me_M$  to 0, i.e., the point where two sets of spike trains are identified as identical.

To obtain these results, the 12 trials were divided into two groups of 6 trials; the mean values and standard deviation were computed by calculating the events metric error for all possible combinations of 6 trials. As shown in Fig. 5.14, as the standard deviation of the Gaussian smoothing function increases the number of identified events decreases. This is similar to varying the parameter  $\phi$  in Eq. (5.34), where a larger value for  $\phi$  implies that more events are merged together.

Analyzing the curves presented in Fig. 5.14, we can conclude that the mean error value varies considerably with the number of identified events. This problem comes mainly from the method of computing the weights in Eq. (5.45): by making them

proportional to the inverse on the average value of the variances, a small misalignment on some spikes or a small variability of the number of spikes leads to a large variability on the weights. Therefore, a slight change in the value of either  $\sigma$  or  $\phi$  results in: a variation in the number of identified events; and in a change of the weights; and in different error values. Another result from this variability is the large variance of the error, which in some cases is even larger than its mean value.

The spike events metric evaluates the spike trains from a perspective that is not directly considered by the other presented error metrics. However, its instability with a slight variation on the parameters and the event detection algorithm makes its application difficult. This problem will be further evidenced in the next section when it is applied to neural modeling.

The spike events metric has the particularity that, every time two events match with each other, there is a negative contribution – a compensation – to the overall error value. However, the application of this metric requires several trials of the response of a given cell for the same stimulus in order to define the events. Moreover, this metric violates the condition in Eq. (5.1) to define a metric function: it can give a negative distance between two sets of spike trains. Nevertheless, this problem could be solved by adding an offset to the origin by knowing the number of firing events in the reference set. The value to add to the firing events error in Eq. (5.39) would be  $Me_M$ , where  $M$  is the number of events present in the reference data. Although this leads to the desired situation that if the two sets are equal their distance will be zero, as Eq. (5.1) states. However, with this new term the property of Eq. (5.2) is not respected. Moreover, using the statistics of the reference spike train to compute the weights in Eq. (5.45) this metric does not fulfill the condition in Eq. (5.2).

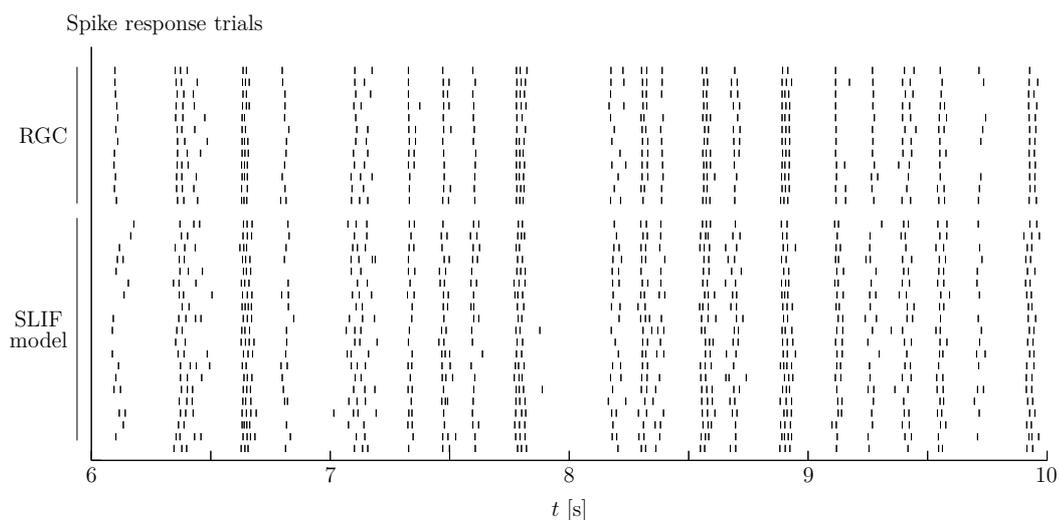
Table 5.3 shows the values obtained by the application of the spike events metrics to the salamander ON-type RGC responses and to the white noise stimuli. The firing events are classified using Eq. (5.34) and a threshold for the events discrimination equal to  $\phi = 3$ .

As it was previously remarked,  $d^{events}(\cdot)$  does not possess the metrics' property in Eq. (5.1); by comparing the set of spike trains with itself, the error is not zero but equal to  $-2 \times M$ , where  $M$  is the number of events. This is due to the negative term in Eq. (5.39) and all events are perfectly matched between the two sets of spike trains.

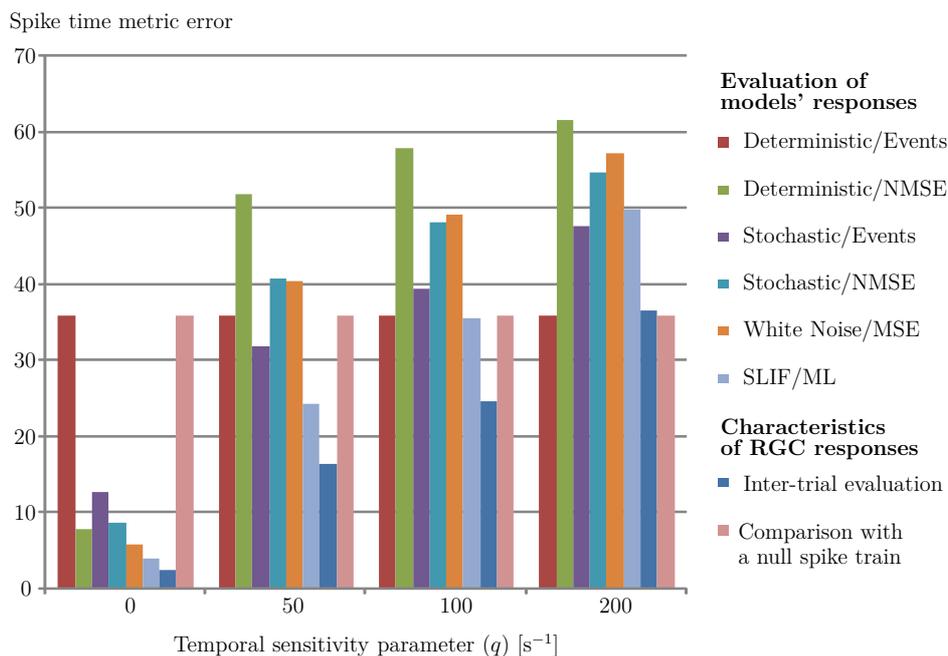
Spike Events Metric		
mean $\pm$ std	min	max
-0.8469 $\pm$ 9.9248	-26.1630	29.7635

**Table 5.3.** Mean and limit values for the spike events metric applied to the salamander ON-type RGC responses.

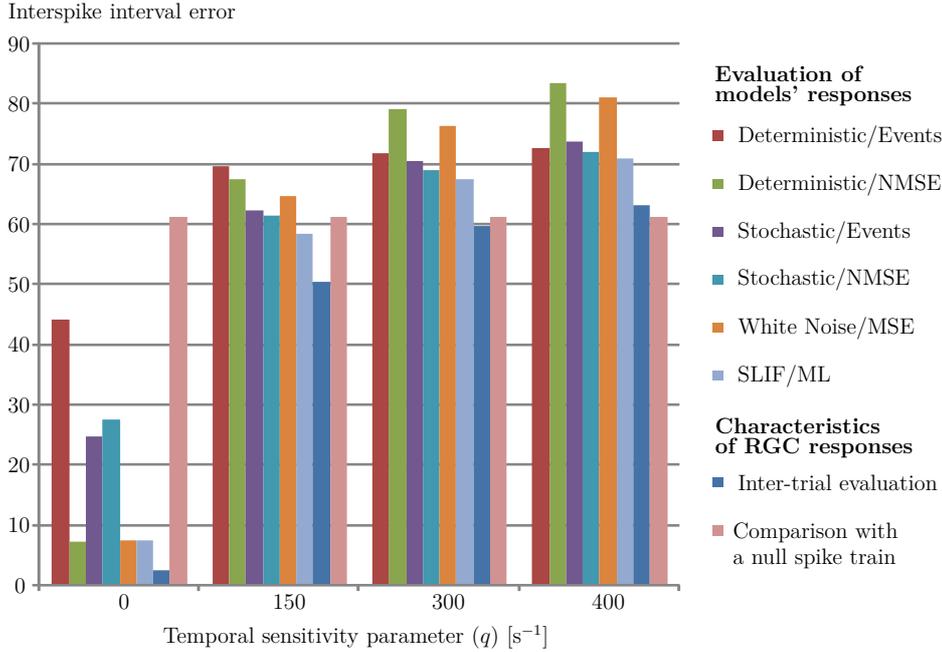
## 5. Neural Activity Metrics



**Figure 5.15.** Salamander **RGC** and **SLIF** model responses.



**Figure 5.16.** Evaluation of model responses with the spike time metric; using the salamander **RGC** responses for comparison (legend: *model name/training method*).



**Figure 5.17.** Evaluation of model responses with the interspike interval metric; using the rabbit RGC ON-OFF responses for comparison (legend: *model name/training method*).

## 5.6. Tuning and Assessment of Retina Models

For a further evaluation of the neural code metrics, they were applied for the tuning and posterior analysis of the output of neural models. This analysis employed four retina models that are briefly described in the next section. To illustrate the drawbacks of the metric’s application, four neural models of the retina are analyzed: *i*) a deterministic model Wilke et al. [2001]; Thiel et al. [2003] that outputs the firing rate,  $r(t)$ , this model is described and analyzed in Sec. 3.5.1; *ii*) a stochastic model Keat et al. [2001] that uses two internal noise sources to model the variability of the retina response and outputs the spike train itself, whose details are described in Sec. 3.5.2; *iii*) a LNP model Chichilnisky [2001], also designated as white-noise model due to the procedure used for fitting its parameters, that also produces a firing rate that is presented in Sec. 3.5.3; and *iv*) a SLIF model that also outputs the sequence of spikes Tomás et al. [2008], whose description and tuning is presented in Sec. 4.2. All models include a linear temporal filtering input block that outputs a temporal generator potential that drives a nonlinear block. In a more general approach, these models should also include a spatial filter to model the spatial processing properties of the retina (as is the case of the deterministic model Wilke et al. [2001]; Thiel et al. [2003]). However, for the present analysis and since the neuronal retina data does not have spatial information, the focus is on the temporal dimension and the spatial component is disregarded.

To assess the retina models, the two sets of data presented in the Fig. 5.2 and Fig. 5.3 were divided into two subsets: a training subset, used to optimize the models’

## 5. Neural Activity Metrics

parameters, consisting of 60% of the total number of available neural responses; and a validation subset consisting of the remaining 40%, used to evaluate the results. The search for the optimal values in the parameters' space was performed using unconstrained nonlinear optimization with the Neelder-Mead simplex algorithm [Flannery et al., 2002], complemented with a simulated annealing scheme based in [Efstratiadis and Koutsoyiannis, 2002]. The RGCs responses were discretized with a sampling period of  $T_s = 1$  ms.

The metrics used for tuning the models were the MSE for the white noise model, and the NMSE and  $d^{\text{events}}$  for the stochastic and deterministic models, respectively; the SLIF model was trained using a ML approach [Tomás et al., 2008]. The training with the rate-based metrics, namely the MSE and the NMSE was made using a Gaussian filter with a standard deviation of  $\sigma = 30$  ms and  $\sigma = 20$  ms for the ON-OFF and sampled white noise stimuli, respectively. When tuning the model parameters using the spike events metric, the PSTH was smoothed with Gaussian filters of  $\sigma = 47$  ms and  $\sigma = 17$  ms, for the ON-OFF and sampled white noise stimuli, respectively. These values correspond to the original algorithm implementation, where  $\sigma$  is made equal to the time scale of the modulations in the firing rate [Berry et al., 1997]. To illustrate the results of model tuning, Fig. 5.15 qualitatively compares 20 spike trains produced by the SLIF model against the salamander testing data set.

The use of the spike time and interspike interval metrics in the tuning of neuron models is unable to determine proper parameter sets. This problem was not noticed only in the neuronal responses and with the models analyzed, but seems to be a general problem of these metrics. For moderate values of the temporal precision parameter  $q$ , a reasonable solution for the model is to fire no spikes at all. Once the model reaches this point, it becomes difficult to evolve to a new set of parameters that results in firing spikes in "reasonable" time bins, i.e., such in a way that it reduces the error. This situation is even worse for high values of  $q$ , where its value is beyond the limit defined in section Sec. 5.4.4, when the null spike train has a lower error than the inter-trial error.

Nonetheless, the spike time metric and the spike interval metric are useful in the analysis of the models' output, for example to assess the variability of neural systems [Kreiman et al., 2000; Chichilnisky and Rieke, 2005; Reinagel and Reid, 2002]. This however must be made with care, as shown in Sec. 5.4, if an incorrect set of parameters is chosen, the null spike train appears to be closer to the RGC responses than two responses from the same cell. A similar problem is identified when evaluating the responses the models.

Figures 5.16 and 5.17 present the evaluation of model responses with the spike time and interspike interval metrics, for the salamander RGC responses to white noise stimulus for the rabbit RGC responses to ON-OFF stimulus, respectively. The figures present four values for the temporal precision of the metrics, where the first corresponds to the spike count metric ( $q = 0$  s<sup>-1</sup>). The three other values,  $q \in \{50, 100, 200\}$  s<sup>-1</sup> in Fig. 5.16 and  $q \in \{150, 300, 400\}$  s<sup>-1</sup> in Fig. 5.17, were chosen based on the method proposed in Sec. 5.4.4, i.e. using the null spike train for comparison. In both figures, from left to right, the second value in the horizontal

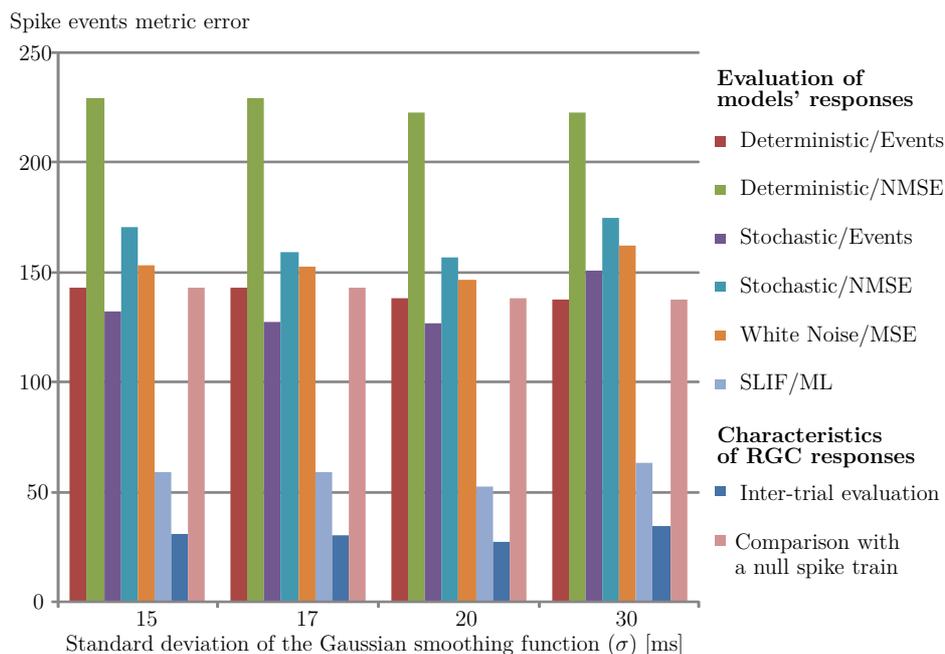
axis is well within the range of precision of the **RGCs**, the third is on the limit and the fourth is beyond the temporal precision limit exhibited by the cells. Thus, in the last case,  $q = 200 \text{ s}^{-1}$  and  $q = 400 \text{ s}^{-1}$ , the model is asked to have a higher temporal precision than the real neurons. As can be seen by analyzing the graphically depicted values in the fourth column ( $q = 200 \text{ s}^{-1}$  in Fig. 5.16 and  $q = 400 \text{ s}^{-1}$  in Fig. 5.17), the results are misleading: since all error bars are large, it seems that all models have an erroneous behavior. This is however an erroneous conclusion. By comparing the results using a smaller value of  $q$ , it can be concluded that some of the models have a good behavior. Even the third column in Fig. 5.17, for  $q = 300 \text{ s}^{-1}$ , may induce to error: since the value was chosen to be too close to the cell temporal precision limit, the null spike train seems to be identical to one of the responses of the rabbit **RGC**. These results further illustrate that the method presented in Sec. 5.4.4 should be used to select the value of the metrics' temporal precision; otherwise, the spike time and interspike interval metrics give meaningless results. This conclusion can also be drawn by using the spike time and the interspike interval metrics for the rabbit and salamander **RGCs**' responses, respectively.

To assess spike firing precision using large values of  $q$  for the spike time and interspike interval metrics, i.e., close to the **RGCs** temporal precision, the spike count metric must be used as an auxiliary measure. For example, consider measuring the temporal precision of the models with the interspike interval metric using the ON-OFF data (Fig. 5.17). By comparing the bars with  $q = 300 \text{ s}^{-1}$ , the deterministic/events model appears to be more precise than the deterministic/NMSE model. However, by assessing the results with the spike count metric ( $q = 0 \text{ s}^{-1}$ ), it can be concluded that this is not true. In fact, the deterministic/events model fires almost no spikes, so the error value for  $q = 300 \text{ s}^{-1}$  is due to the creation of almost all spikes. On the other hand, the deterministic/NMSE model has a spike count similar to the one exhibited by the rabbit **RGC**, thus the error comes from shifting non-coincident spikes (or, when these spikes are farther apart than 6.6 ms, to its removal and re-creation). While other measures could detect this problem, like the spike train distance with a large value for  $\tau$ , the spike count metric is easier to implement and provides insightful information.

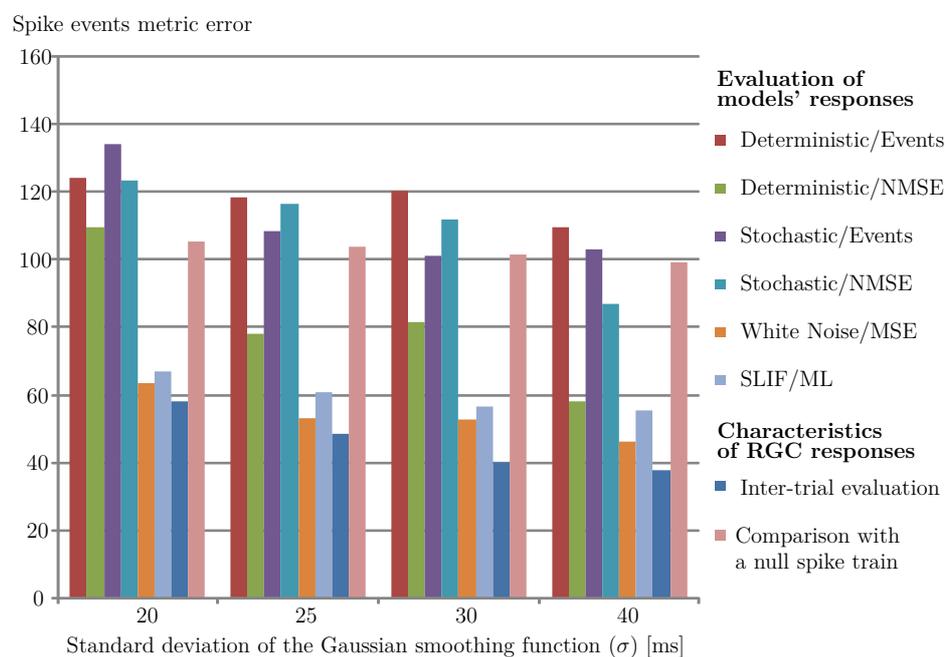
The use of the spike events metric to tune the model parameters is also problematic. As shown in subsection Sec. 5.5.1, a small variation on spike timing or on the identification of spike events can lead to large differences in the errors. In most situations, the adopted simulated annealing scheme was able to overcome the local minima. This was not however the case of the deterministic model trained with the spike events metric, where the best obtained solution was to fire no spikes at all.

While the metric does not seem to be the most adequate for model tuning, it can be used to assess model responses. Figure 5.18 and Fig. 5.19 present the obtained results by evaluating the models with the spike events metric. As shown by the graphic, the metric evaluates an interesting characteristic of the neural code: the capability for the model to reproduce the bursts of spikes exhibited in the real neural responses. This conclusion is drawn because the obtained error values are not correlated with the results of the application of other neural metrics, namely: the spike time and

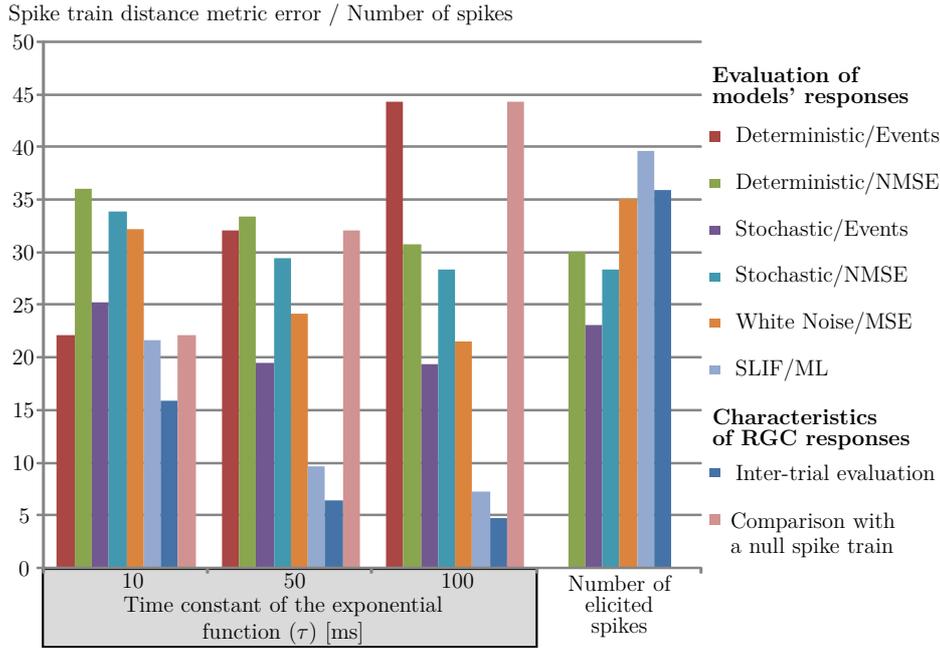
## 5. Neural Activity Metrics



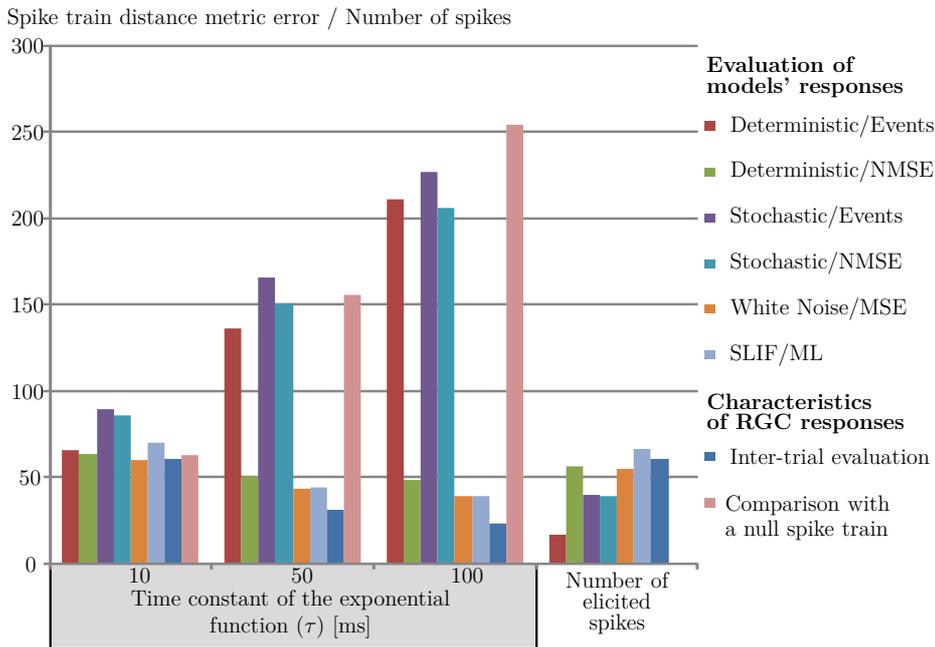
**Figure 5.18.** Evaluation of model responses with the spike events metric; using the salamander RGC responses for comparison (legend: *model name/training method*).



**Figure 5.19.** Evaluation of model responses with the spike events metric; using the rabbit RGC ON-OFF responses for comparison (legend: *model name/training method*).

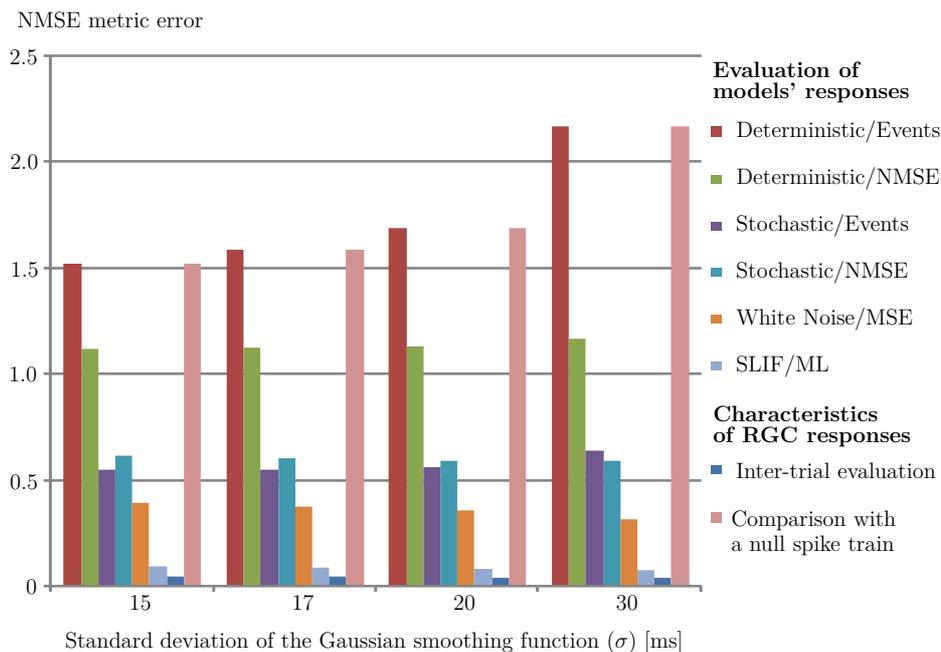


**Figure 5.20.** Evaluation of model responses with the spike train distance metric; using the salamander RGC responses for comparison (legend: *model name/training method*).

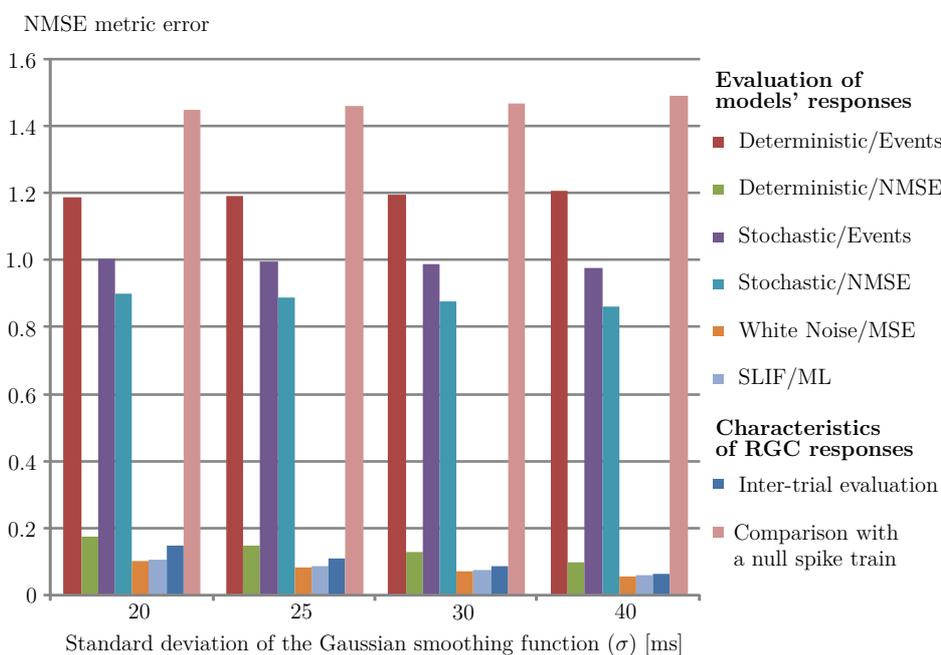


**Figure 5.21.** Evaluation of model responses with the spike train distance metric; using the rabbit RGC ON-OFF responses for comparison (legend: *model name/training method*).

## 5. Neural Activity Metrics



**Figure 5.22.** Evaluation of model responses with NMSE metric; using the salamander RGC responses for comparison (legend: *model name/training method*).



**Figure 5.23.** Evaluation of model responses with the NMSE metric; using the rabbit RGC ON-OFF responses for comparison (legend: *model name/training method*).

interspike interval metrics; the spike train distance metric, whose results are presented in Fig. 5.20 and Fig. 5.21; and the **NMSE** metric, presented in Fig. 5.22 and Fig. 5.23. The results for the **%VAF** metric are not presented herein because the metric does not only give no additional information regarding the **NMSE** metric, and also degrades the results; this was mathematically shown in Sec. 5.3.4.

Analyzing the obtained experimental results using the spike train distance metric, the relevance of the proposed algorithms for selecting adequate parameters is further demonstrated. For example, observing the error values for  $\tau = 10$  ms for the ON-OFF data, presented in Fig. 5.21, for this small value of  $\tau$ , i.e., for a high temporal precision, larger than that of the real **RGC**, all models seem to have similar behaviors. However, by increasing the value of  $\tau$ , therefore also increasing the difference between the inter-trial evaluation of the **RGC** responses and the cross-comparison with the null spike train, it can be seen that the models have different behaviors. It should be noted that, if the objective is to analyze the spiking temporal precision, the value of  $\tau$  should be small, since for a large value of  $\tau$  this metric measures not the temporal precision but the difference in the number of spikes [van Rossum, 2001; Tomás and Sousa, 2008].

Another characteristic common to several neural code metrics is that the absolute value of the error presents no information by itself. Instead, in order to have a correct understanding of the results, they must be compared with the inter-trial evaluation of cell responses and the null spike train. A good example is the spike train distance whose values are graphically presented in Fig. 5.20 and Fig. 5.21. While for the white noise data (salamander **RGC**), the inter-trial errors are around 20, for the ON-OFF data (rabbit **RGC**) these are around 60. This is due to the number of spikes in the trains and to the cells' variability. When these quantities increase so does the inter-trial error; as a consequence, the error measures for the models must also increase. It should be noted that the "perfect model" should reproduce the statistics of a real neuron. Thus, the error measure when comparing with the real data should have the same statistics as the inter-trial evaluation.

Analyzing all the graphically represented results, it can be concluded that, in general, the usage of neural metrics for model training may lead to results that are not good. To overcome this problem, other tuning methods, like the maximum likelihood (**ML**) approach used to tune the **SLIF** model should be employed (see Sec. 4.2). In this case, the optimization algorithm tries to maximize the likelihood for the model to reproduce every spike, given the history of spikes, i.e. the timing of all previously elicited spikes. While the maximum likelihood by itself is not a neural metric, it represents a general statistical measure that can also be used for neural models. In these cases, after parameter optimization, the model performance is evaluated using the presented neural metrics. The experimental results obtained show that the maximum likelihood approach can escape local minima generated by the neural metrics, thus resulting in well-trained models.

In general, it can be concluded from the presented results that a given metric assess the performance of a model from a particular standpoint. As a result, more than one metric should be used to address different aspects of the neural behavior, namely: *i)*

## 5. Neural Activity Metrics

a rate metric to compare the mean firing rate; *ii*) the spike count metric to compare the difference in the number of spikes; *iii*) a spike train metric to assess the model's firing accuracy; and *iv*) the spike events metric to evaluate the model's capacity to reproduce the spike bursts. However, for the firing metrics, the %VAF should not be used since its error decreases whenever the spike count metric error increases.

### 5.7. Conclusions

This chapter analyzes and compares different neural metrics to disclose their advantages and drawbacks by using sets of neuronal responses from the retina. Subsequently, its application to the tuning and assessment of a set of retina models provides important hints about its usefulness and applicability in the training and evaluation of the models. In most cases the metric depends on a free parameter that establishes its sensitivity to a particular characteristic of the neural code, so that an incorrect choice of these parameters can lead to meaningless results. Hence, a set of methods are proposed to define a valid range for the values of the different metrics' parameters.

These methods are based on the statistical analysis of the inter-trials errors of the real spike trains. The application of the neural metrics to the tuning and assessment of retina models from distinct classes reveals important results. Some of the analyzed metrics possess pronounced minima, specifically around the origin, which makes the optimization process more difficult; nonetheless, generally these metrics provide insightful results for the evaluation of models.

The understanding and choice of a neural metric is a crucial issue in the neural sciences and neural modeling. In the evaluation of neural responses, the choice of a metric is constrained by the neural features under analysis. In the optimization and assessment of neural models, the chosen metric greatly influences the parameter estimation and model performance. In most cases, the metrics are influenced by a parameter that affects its analysis of neural responses. In the case of spike train metrics, this parameter defines the temporal precision of the metric; on the other hand, the rate metrics are influenced by a parameter, or function, that defines the smoothing mechanism used to estimate the neural responses' firing rate. One of the problems in the application of the error metrics is selecting the values for these parameters. Therefore, methods to disclose limiting values for the parameters are proposed. In all cases, the method consist of observing the evolution of the metrics' values when applied in the inter-trial comparison of neural responses. However, the limiting values obtained for the parameters are dependent on the neural responses; since different cells present different temporal precisions, the values obtained for the parameters also depend on the responses. By applying the neural metrics for assessing models' responses, the importance of selecting adequate values for the metrics' parameters is revealed: when the parameters are not within the range of values defined by the presented methods, the metrics produce meaningless results.

Another critical point in the application of neural metrics concerns the analysis

of the results. As seen in the examples presented, when comparing neural responses of the same neuron to the same stimulus, the error values are not zero, especially in the case of spike train metrics. To assess the quality of a neural model using the neural metrics, the errors obtained in the comparison of the model's responses with the real neuron responses must have the same statistics as the inter-trial errors of the neuron's responses. Moreover, for the modeling and tuning of a neural system, a question can be raised: is the model unable to mimic the neural system under study according to the intended metric or is the training algorithm simply unable to find the correct set of parameters? In general, each metric is tied to a particular view of the neural code. Consequently, the development and tuning of a neural model must always have a particular underlying neural perspective. However, some metrics, such as the spike time and the spike interval metric, which are suited for analyzing the precision of neural responses, are difficult to use in the tuning of neural models. This difficulty comes from the fact that they typically lead to local minima that, from the experience with several neural responses and several models, are hard to overcome. An alternate solution consists of applying other optimization techniques, like the maximum likelihood optimization procedure, and apply the neuronal metrics to assess the results.

Finally, it should be stressed that, in order to fully assess neurons' and models' responses, one must consider different metrics depending on the relevant quantity under measure: rate metrics are used to compare mean firing rates, spike train metrics assess temporal precision, and spike events metrics measure how well spike bursts are reproduced. However, the %VAF metric should not be used, and the spike events metric should be used cautiously due to its sensitivity to the event identification procedure. The metrics should be applied with care, using the proposed methods to select their parameter ranges.

The results presented in this chapter from the analysis of several retina models also show that there is not a universal model capable of describing meaningful responses for all types of visual stimuli existent in everyday life. In assessing the models, two different types of visual stimuli were used. The deterministic flash or ON-OFF stimulus represents geometrical images better, such as artificial landscapes like buildings or text, while the random stimuli possess characteristics similar to natural landscapes, like a forest [Dong and Atick, 1995]. Maybe the final answer to what model is better can only be undoubtedly answered by performing experiments with real prosthesis in humans. However, for sure the plasticity of the nervous system and its ability to adapt and learn would play an important part [Fernández et al., 2005; Carmena, 2012].



... a review of history establishes the fact that the process of scientific progress requires a multitude of viewpoints and reveals that even unsuccessful efforts contribute to the advancement of knowledge by identifying dead ends.

Marmarelis, V. Z. (2004). *Nonlinear Dynamic Modeling of Physiological Systems*. IEEE Series on Biomedical Engineering. Wiley-Interscience

# 6

## System Identification and Parameters Estimation

### 6.1. Introduction

SYSTEM modeling is a central task in science, and particularly in engineering. There is no scientific knowledge area where modeling does not play an important role. Moreover, an independent scientific field is established when its body of knowledge contains firstly a series of models, and subsequently a set of analysis techniques and modeling tools so that it can develop independently *per se*, e.g., physics, electronics and informatics.

This chapter is dedicated to the presentation and analysis of a system identification technique that is based on multiple model adaptive estimators. This novel technique can be applied in the identification of nonstationary linear and nonlinear systems. This technique was briefly mentioned for a single parameter in [Martins, 2006]. Afterwards it was extended for the multidimensional case and applied to nonstationary linear systems [Martins et al., 2011a,b], and it was also applied to the estimation of nonlinear systems [Martins et al., 2013].

### 6.2. System Modeling and Identification

The goal of system modeling is to obtain an explicit mathematical expression for the relation between the input and output signals, using experimental or natural input-output data and possibly other available knowledge about the system. The modeling process is usually achieved in two steps. First, a suitable mathematical model, containing unknown parameters and/or functions, is selected to model the

## 6. System Identification and Parameters Estimation

relationship between the input and output signals, this phase is termed model specification. In a second step, termed model estimation, the model's parameters and functions are estimated by using the input-output data. The whole process is referred as system identification. In the model specification process all prior knowledge and information regarding the system is used to select an appropriate model. The model estimation task employs estimation methods to obtain the parameters values in order to maximize the accuracy of the resulting model predictions.

Various modeling tools and identification techniques have been developed and used to model systems, linear or nonlinear [Ljung, 1999; Nelles, 2001]. Several techniques have been specially extended and applied to model physiological systems that are mainly nonlinear [Westwick and Kearney, 2003; Marmarelis, 2004]. Modeling of physiological system is of paramount importance since it enables the understanding of the mechanisms involved in the biologic process and to infer unobserved behaviors, fundamental in the development of prosthesis.

Although that the model and the estimation technique are not completely independent, sometimes the type of model used in system modeling is confused with the technique employed in the estimation process. A linear or nonlinear model can be classified in two major categories: static and dynamic models. Static models perform a direct mapping between the input signal vector  $\mathbf{u} \in \mathbb{R}^n$  and the output vector  $\mathbf{y} \in \mathbb{R}^r$ . This relation is usually obtained by mapping the input into the output possibly using basis functions like:

$$\mathbf{y} = \sum_{i=1}^M \alpha_i \Phi_i(\mathbf{u}, \boldsymbol{\beta}_i), \quad (6.1)$$

where the output vector  $\mathbf{y}$  is modeled as a weighted sum of the  $M$  basis functions  $\Phi_i(\cdot)$ , with weights  $\alpha_i$ . The basis functions depend on the input and are parameterized by a set of (non)linear parameters gathered in the vector  $\boldsymbol{\beta}_i$ . If the mapping is nonlinear the basis functions are nonlinear and the parameters are nonlinear. Typical basis functions are harmonic functions (e.g. sines and cosines functions [Davis, 1989]), polynomials (e.g. splines [De Boor, 2001], Laguerre polynomials [Rugh, 1981]), sigmoidal functions (used in neural networks, e.g. logistic function, hyperbolic functions [Haykin, 1999]). The basis functions formulation can be further extended by replacing each weight by a function [Nelles, 2001].

An alternative to static models are dynamic models. Typically a dynamic model describes the system using differential (or integral) equations in the continuous case, or difference equations, in the discrete-time case. The word dynamic relates to the fact that the output at a given time instant depends on the input signal at the current and on previous time instants, such as the retina model presented in Sec. 4.2.

The typical approaches to system modeling are the parametric and nonparametric modeling [Nelles, 2001; Ljung, 1999]. In the parametric approach the input/output relation is described with a finite number of parameters. A typical example is to model this relationship by differential, or difference, equations for dynamic systems and by algebraic models for static systems. The models' parameters can be constant

	Parametric	Nonparametric	Modular	Connectionist
Model specification	-	+	±	±
Interpretability	+	±	+	-
Robustness to noise	-	+	+	±
Compactness	+	-	+	-
Adaptable to time variance	±	+	±	+

**Table 6.1.** System modeling methodologies: strengths (+) and weaknesses (-) (from [Marmarelis, 2004]).

or time-varying depending if the model is stationary or nonstationary. Frequently, the parameters are related with some physical process that describes the system. In the nonparametric approach the system can be modeled analytically with an integral equation where the unknown quantity are the kernel functions, like in Volterra-Wiener expansions [Rugh, 1981], or through a computing mapping like look-up tables or an infinite impulse response.

In the modeling of physiological systems two additional approaches are distinguished, namely the modular and the connectionist approach [Marmarelis, 2004]. The modular approach is an hybrid between the parametric and nonparametric approach. It uses block structured models composed of parametric and nonparametric components reflecting the knowledge about the functional organization of the system. The connectionist approach makes use of generic model configurations and architectures, like artificial neural networks, to represent nonlinear mappings. The connectionist models are fully parameterized but lack the interpretability of the models' parameters opposed to the parametric modeling.

Table 6.1 resumes the strengths and weaknesses of different modeling methodologies. Desirably, a model should be: accurate –reproduce the observed data from the input signal; global – accurate under natural working conditions; compact – low mathematical and computational complexity; interpretable – it should give insight into the working mechanisms of the system; robust to the presence of external and internal noise processes. In the model specification stage the prior information about the system is used to select an adequate model, and a suitable estimation method for each case is applied to optimize the accuracy of the model for different data types and noise scenarios. The model accuracy can be accessed using different measures and norms, depending on the system type and purpose of the model.

The method to be proposed for the identification of is based on a parametric dynamic model for the system. This approach has the advantage that if the parameters are related with some physical process describing the system one can have an initial guess for its values.

### 6.2.1. State Space Representation

The state-space representation of dynamic systems provides a consistent framework for describing and analyzing systems of any degree of complexity that has shown to be very convenient. Models of physical, chemical, economic, and biological systems

## 6. System Identification and Parameters Estimation

can be described as assemblages of interconnected first-order differential equations or by higher order differential equations that can be transformed in the former. A scalar  $n$ th-order differential equation, or difference equation, can be decomposed into  $n$  first order ordinary differential, or difference, equations or, equivalently, by a single first order differential, or difference, vector equation of dimension  $n$ . This constitutes the core representation of a dynamic system in the state-space framework.

In the state-space representation the dynamic model of a general nonlinear discrete time-varying system is described by the vector difference equation:

$$\mathbf{x}[t + 1] = \mathbf{f}(\mathbf{x}[t], \mathbf{u}[t], \boldsymbol{\xi}[t], \boldsymbol{\gamma}[t], t), \quad t = 0, 1, 2, \dots, \quad (6.2)$$

where  $\mathbf{x}[t] \in \mathbb{R}^n$  is the vector of the system's state variables at the time instant  $t$ , and  $\mathbf{u}[t] \in \mathbb{R}^m$  is the input signal that drives/controls the system, and  $\boldsymbol{\xi}[t] \in \mathbb{R}^p$  accounts for the system's disturbance noise. The vector  $\boldsymbol{\gamma}[t] \in \mathbb{R}^r$  represents the time-variable system's parameters vector.

The system's output vector is written as:

$$\mathbf{y}[t] = \mathbf{g}(\mathbf{x}[t], \mathbf{u}[t], \boldsymbol{\xi}[t], \boldsymbol{\gamma}[t], t), \quad (6.3)$$

where  $\mathbf{y}[t] \in \mathbb{R}^r$ . The nonlinear observation of the system's state variables degraded with noise is modeled as:

$$\mathbf{z}[t] = \mathbf{h}(\mathbf{y}[t], \boldsymbol{\theta}[t], t), \quad (6.4)$$

where  $\mathbf{z}[t] \in \mathbb{R}^r$  is the observations vector, and  $\boldsymbol{\theta}[t] \in \mathbb{R}^r$  is the noise associated with the measurements. The uncertainties in the process and the system and observation noises are characterized by:

$$E\{\mathbf{x}[0]\} = \bar{\mathbf{x}}_0 \quad \text{cov}[\mathbf{x}[0]; \mathbf{x}[0]] = \boldsymbol{\Sigma}_0 \quad (6.5)$$

$$E\{\boldsymbol{\xi}[t]\} = \mathbf{0} \quad \text{cov}[\boldsymbol{\xi}[t]; \boldsymbol{\xi}[\tau]] = \boldsymbol{\Xi}[t]\delta_{t\tau} \quad (6.6)$$

$$E\{\boldsymbol{\theta}[t]\} = \mathbf{0} \quad \text{cov}[\boldsymbol{\theta}[t]; \boldsymbol{\theta}[\tau]] = \boldsymbol{\Theta}[t]\delta_{t\tau} \quad (6.7)$$

where  $\mathbf{x}[0]$ ,  $\boldsymbol{\xi}[t]$ ,  $\boldsymbol{\theta}[t]$  are assumed to be independent for all  $t$ . The statistical independence between samples of the process and observation noises indicate that they are white noise. In the case of non-white noise the system and/or observation dynamics can be extended to include a pre-whitening process [Orfanidis, 1990; Simon, 2006].

For the case of a linear system with linear observations Eq. (6.2) and Eq. (6.3), can be written as:

$$\mathbf{x}[t + 1] = \mathbf{A}\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] + \mathbf{L}\boldsymbol{\xi}[t], \quad t = 0, 1, 2, \dots, \quad (6.8)$$

and

$$\mathbf{z}[t] = \mathbf{C}\mathbf{x}[t] + \boldsymbol{\theta}[t], \quad (6.9)$$

respectively. Thus, as for the output, the simplest case occurs when the output is equal to the state variables vector:  $\mathbf{y}[t] = \mathbf{I}_n\mathbf{x}[t]$ , where  $\mathbf{I}_n$  is the identity matrix of order  $n$ .

## 6.3. Multiple Model Adaptive Estimators

The identification method proposed is based on the multiple-model adaptive estimation (MMAE) algorithm. The MMAE algorithm is a versatile and powerful algorithm frequently used in system identification and state estimation problems. It was developed some decades ago [Magill, 1965; Lainiotis, 1971] but conquered a renewed and consistent interest over time [Athans and Chang, 1976; Hanlon and Maybeck, 2000] and as seen some substantial developments in the recent years [Aguilar, 2007; Hassani et al., 2009c,a]. The MMAE has been used in a profusion of practical applications like navigation control [Eide and Maybeck, 1996], adaptive robust control [Fekri et al., 2006], surveillance systems [Reid, 1979], biomedical applications [Yu et al., 1992], to name a few.

By following a Bayesian analysis of the system state using the input and output signals, and possibly some available observations of the state variables, the MMAE provides a way to recursively calculate the probability of each model, from a set of models, to represent the system under identification. In the most usual application the set of models is pre-established and fixed, and the MMAE combines the posterior probabilities and the state estimates obtained from each model to estimate the overall system state and its parameters [Anderson and Moore, 2005].

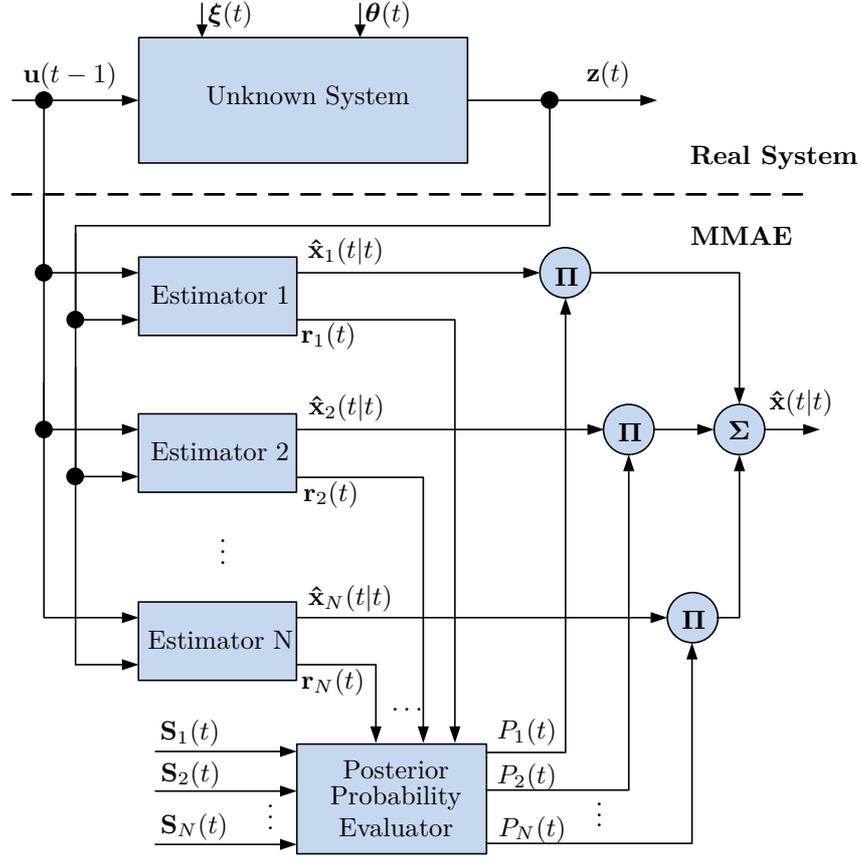
Departing from the traditional utilization the constellation based multiple-model adaptive estimation (CBMMAE) iteratively constructs a set of models that, based on the values of their posterior probabilities, is adaptively conformed to lower the posterior error covariance matrix. This new method is capable of identifying with precision the system under analysis and provides an estimate for the unknown parameters. Since the set of models is adaptively adjusted to search the parameter space for the model that best mimics the system, when the system's parameters change the algorithm is able to adjust the models to the new system configuration that is eventually identified, and its new parameters estimated.

### 6.3.1. MMAE Algorithm Setup and Properties

The MMAE is composed by a bank of state estimators based on a set of models for the system, also called local observers. These estimators run in parallel and provide a local estimate for the system state. Additionally, the MMAE contains a (posterior) probability evaluator block that based on the individual estimates and on the error covariance matrices, computes the likelihood of each model of representing the system, and computes the overall system state by weighting the individual estimates. Each estimator is designed using one element of a set of models that represents the possible different system's behavior patterns. A crucial issue in the utilization of the MMAE is the adequate selection of models to be used in the bank of estimators.

The MMAE algorithm follows a Bayesian approach to obtain the posterior probabilities for the set of models based on the state and covariance estimate from each model. The combination of the posterior probabilities gives the estimate of the true system state and its error covariance matrix [Athans and Chang, 1976; Anderson and Moore,

## 6. System Identification and Parameters Estimation



**Figure 6.1.** The MMAE algorithm block diagram.

2005].

Employing  $N$  estimators in the MMAE setup the posterior probability for the model  $k$  to represent the system is computed recursively through the expression:

$$P_k[t+1] = \frac{\beta_k[t+1] e^{-\frac{1}{2}w_k[t+1]}}{\sum_{j=1}^N \beta_j[t+1] e^{-\frac{1}{2}w_j[t+1]} P_j[t]}, \quad k = 1, \dots, N, \quad (6.10)$$

where the scalar quantity  $w_k[t]$  is given by:

$$w_k[t] = \|\mathbf{r}_k[t]\|_{\mathbf{S}_k^{-1}[t]} = \mathbf{r}_k^T[t] \mathbf{S}_k^{-1}[t] \mathbf{r}_k[t], \quad (6.11)$$

which defines a matrix induced metric, and  $\beta_k[t]$  is defined by:

$$\beta_k[t] = \frac{1}{(2\pi)^{m/2} \sqrt{|\mathbf{S}_k[t]|}}. \quad (6.12)$$

The vector  $\mathbf{r}_k[t]$  in Eq. (6.11) is the residual between the observed  $\mathbf{z}$  and the predicted vector  $\hat{\mathbf{z}}$  of observations from model  $k$ :

$$\mathbf{r}_k[t] = \mathbf{z}[t] - \hat{\mathbf{z}}_k[t], \quad (6.13)$$

and  $\mathbf{S}_k[t]$  is the residual covariance matrix with determinant  $|\mathbf{S}_k|$ .

Based on  $N$  different models, each estimator generates a state estimate  $\hat{\mathbf{x}}_k[t]$ , that are combined to compute the overall system state estimate given by:

$$\hat{\mathbf{x}}[t] = \sum_{k=1}^N P_k[t] \hat{\mathbf{x}}_k[t], \quad (6.14)$$

where  $P_k[t]$  is the posterior probability for model  $k$  given by Eq. (6.10). The state covariance error matrix is updated according to the recursive expression:

$$\Sigma[t] = \sum_{k=1}^N P_k[t] \left( \Sigma_k[t] + (\hat{\mathbf{x}}_k[t] - \hat{\mathbf{x}}[t]) (\hat{\mathbf{x}}_k[t] - \hat{\mathbf{x}}[t])^T \right). \quad (6.15)$$

An important property of the MMAE algorithm is that if the  $i$ th model matches the system then the posterior probabilities of the models evolve according to:

$$\lim_{t \rightarrow \infty} P_i[t] = 1; \quad \text{while} \quad \lim_{t \rightarrow \infty} P_k[t] = 0; \quad \forall k \neq i. \quad (6.16)$$

Equation (6.16) states that if there is a model that matches the system then it is identified with probability one, while the posterior probabilities of the other models go to zero [Baram and Sandell, 1978; Hassani et al., 2009b]. It should be noted that from Eq. (6.10) if one starts with  $\sum_{k=1}^N P_k[0] = 1$  then

$$\sum_{k=1}^N P_k[t] = 1, \quad \forall t. \quad (6.17)$$

A remarkable key property of the MMAE algorithm is that if none of the models match the system then the posterior probability of the closest model according to the matrix induced metric, expressed by Eq. (6.11), tends to one [Moore and Hawkes, 1975; Athans and Chang, 1976].

The system parameters vector estimate and its covariance matrix are computed from:

$$\hat{\boldsymbol{\gamma}}[t] = \sum_{k=1}^N P_k[t] \boldsymbol{\gamma}_k, \quad (6.18)$$

$$\Sigma_{\hat{\boldsymbol{\gamma}}}[t] = \sum_{k=1}^N P_k[t] (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}[t]) (\boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}[t])^T. \quad (6.19)$$

### 6.3.2. The Multiple Model Adaptive Estimator for Nonlinear Systems

Similarly to its application in linear systems the MMAE algorithm for nonlinear systems comprises three stages: estimation of the system state for each vector of parameters corresponding to each different configuration, the evolution of the individual probabilities for each system configuration, and combination of the individual estimates to obtain the overall estimation of the system state and of the parameter

## 6. System Identification and Parameters Estimation

vector. The update of the different hypothesis probabilities, given in Eq. (6.10), is obtained by developing the dynamic evolution of the probabilities given by:

$$P_k[t+1] = \frac{p(\mathbf{z}[t+1]|\boldsymbol{\gamma}_k, \mathbf{Z}[t])}{\sum_{i=1}^N P_i[t]p(\mathbf{z}[t+1]|\boldsymbol{\gamma}_i, \mathbf{Z}[t])} P_k[t], \quad (6.20)$$

where  $\mathbf{Z}[t] = \{\mathbf{u}[0], \mathbf{u}[1], \dots, \mathbf{u}[t-1], \mathbf{z}[1], \dots, \mathbf{z}[t-1]\}$  is the set of previous inputs and past observations, and  $\boldsymbol{\gamma}_i$  is the current vector of parameters. This expression can be derived by making no assumption on which type of system is considered so that it is valid both for linear and nonlinear systems [Athans and Chang, 1976]. However, the posterior probability density  $p(\mathbf{z}[t+1]|\boldsymbol{\gamma}_k, \mathbf{Z}[t])$  cannot be obtained exactly for nonlinear systems, and it can only be approximated with a sub-optimal filter, like the extended Kalman filter (EKF) or the unscented Kalman filter (UKF), for computing the state estimates  $\hat{\mathbf{x}}_k[t|t]$  and  $\boldsymbol{\Sigma}[t|t]$ . The optimum state estimate and its covariance matrix are also given by Eq. (6.14) and Eq. (6.15). The selection of the type of the non-optimal filter is usually based on the physical problem, on the performance required, and on the available computer resources.

An algorithm similar to MMAE is the Gaussian Sum (GS) filter [Anderson and Moore, 2005]. The GS filter was developed for estimation of the non-gaussian probability density of the system state propagated by nonlinear systems by a finite sum of Gaussian functions, each having a different mean and covariance [Alspach and Sorenson, 1972]. Although the GS filter is based on different assumptions and follows a different philosophy concerning the MMAE algorithm, its structure is identical to the MMAE. The original GS filter explicitly employs the EKF for each estimator in Fig. 6.1 to obtain the individual state estimates,  $\hat{\mathbf{x}}_k[t|t]$ , and the state covariance matrices,  $\boldsymbol{\Sigma}[t|t]$ , of each system model that are then mixed according to:

$$p(\mathbf{x}[t]|\mathbf{Z}[t]) = \sum_{k=1}^N \alpha_k N(\mathbf{x}[t]; \hat{\mathbf{x}}_k[t|t], \boldsymbol{\Sigma}[t|t]), \quad (6.21)$$

to compute the current state probability density function. Recent developments of the GS filter also uses the UKF as the state estimate [Gabriel et al., 2012]. The values for the weights,  $\alpha_k$ , of each Gaussian density in the mixture are obtained by the same expression of Eq. (6.10) (where now  $\alpha_k(t) = P_k(t)$ ), although  $\alpha_k$  is no longer interpreted as a probability. Similarly to Eq. (6.17) the properties:

$$\sum_{k=1}^N \alpha_k[t] = 1, \text{ and } \alpha_k[t] \geq 0, \quad \forall t, \quad (6.22)$$

still hold. In the GS the overall state estimate and the error covariance matrix are also obtained from Eq. (6.14) and Eq. (6.15).

These results allow us to anticipate that by employing sub-optimal state estimators the MMAE can be applied to the identification and parameters estimation of nonlinear systems.

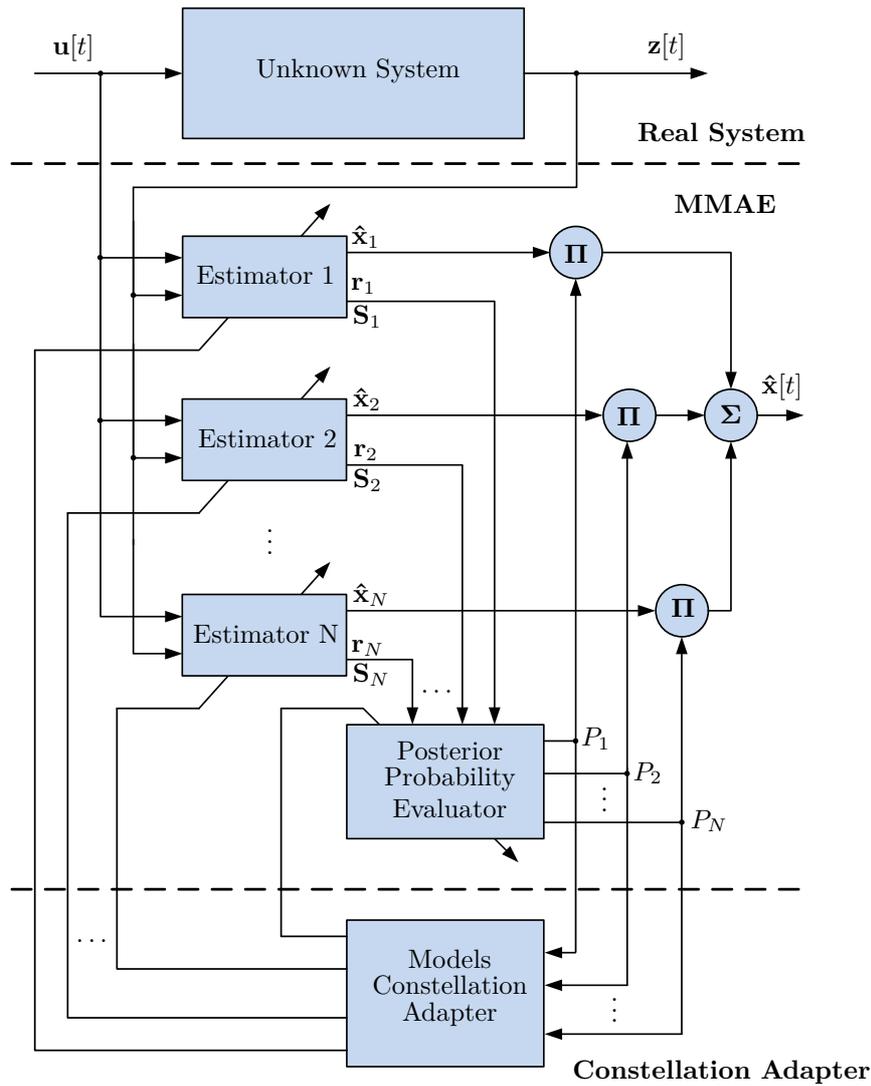


Figure 6.2. Block diagram of the multiple model adaptive estimator and constellation adapter.

## 6.4. Constellation Based Multiple Model Adaptive Estimators

The standard approach to system identification using the MMAE algorithm is to provide an exhaustive set of models that represent every possible system configuration and let the MMAE pick up the true one by observing the evolution of the individual posterior probabilities [Anderson and Moore, 2005]. A disadvantage of this approach is that the convergence of the MMAE can be slow when the system has a large number of possible different configurations, since each model equally compete for the posterior probabilities even knowing that some of the modes are unlikely, with the additional drawback that, depending on the number of models used, the computational burden

can become very high. An approach proposed to circumvent this drawback is the use of a variable structure for the estimators bank in the **MMAE** by including along the way only the most probable models from a preestablished set based on a Markov law [Li and Bar-Shalom, 1996]. This approach is somewhat complex since the probabilities for each intermediate step need to be computed from the Markov law, and the number of models and possible different structures can be large.

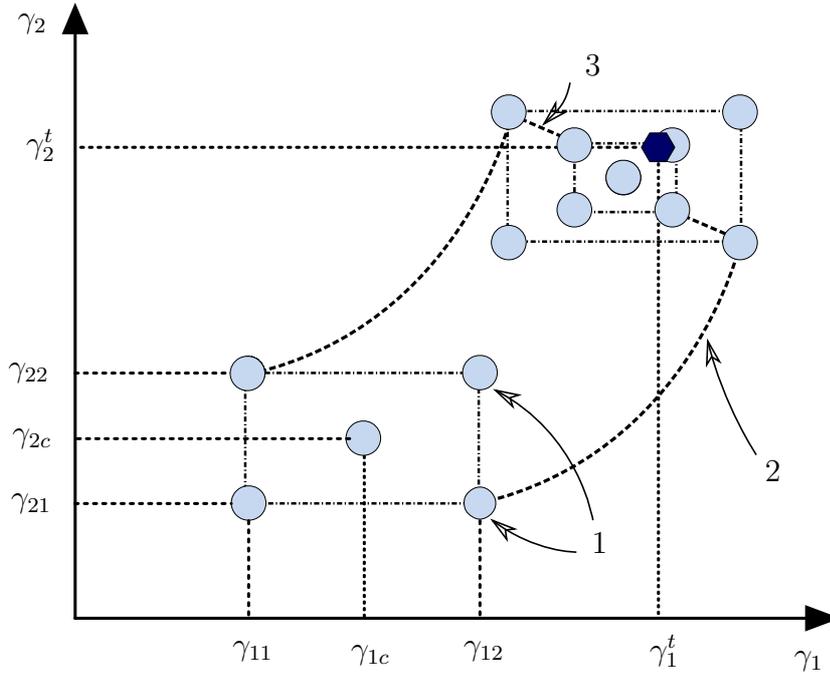
The approach followed in the constellation based multiple-model adaptive estimation (**CBMMAE**) is to establish a minimum set of models in the unknown parameter space, to be used in the estimators bank of Fig. 6.1, that are properly designed, whose parameters are adapted according to the convergence path followed by the posterior probabilities for each model. Therefore, relying on the properties of the **MMAE** algorithm expressed by Eq. (6.16), a method is proposed to establish a minimum set of models to be employed in the estimators bank, and a strategy to update its parameters so that the system is eventually identified and its parameters tuned [Martins et al., 2011a].

The performance of the algorithm is analyzed for different scenarios, that includes a linear stationary and nonstationary system and a nonlinear system.

The algorithm starts with a base model for the system and a suitable range of values for each of the unknown parameters. Therefore, with this base model, a set of models with values for their parameters positioned in specific positions in the space spanned by the unknown parameters, termed a constellation, is designed. The cardinality of the models constellation depends only on the number of unknown parameters. This constellation of models is used on the bank of estimators of the **MMAE**. Thus, by analyzing the posterior probabilities given by the **MMAE**, the constellation of models is succeedingly shifted on the parameter space until the region containing the system's parameters is found. When the region where the system's parameters dwell is identified the constellation's volume is shrunk by reducing the intervals' range for its parameters, fine-tuning this way the parameters estimation.

Due to its adaptive nature the algorithm is able to detect changes in the system's parameters so that it can adjust to new system configurations for the case of nonstationary systems. The use of the **CBMMAE** algorithm is not restricted to linear systems and can be used in the identification of nonlinear systems by using the appropriate state estimator, optimal or non-optimal, respectively, and by taking into account the common constraints related to the nonlinear systems.

Establishing a convenient set of models, termed a constellation, for the estimators bank of the **MMAE** algorithm one knows that the posterior probability of the model closest to the unknown system according to the induced matrix metric expressed by Eq. (6.11), will tend to one according to the property expressed by Eq. (6.16). Starting with a constellation of models in the unknown parameter space, whenever the posterior probability of a given model is identified with probability close to one, the constellation is moved on the parameter space into the direction of that point. Repeating this process the unknown system parameters' point ends inside the constellation's volume. At this phase the system parameters are identified with an error equal to the radius of the constellation. After bracketing the system parameters'



**Figure 6.3.** Constellation topology and evolution in a two-dimensional parameter space: 1- initial constellation setup; 2- tracking and bracketing; 3- shrinking.

point, the constellation can be shrunk to refine the parameters estimation. In the meantime, in the case of a nonstationary system, if the system parameters' point changes its position on the parameter space then it is no longer bracketed and the algorithm can be initialized to track the new system parameters' point until it is bracketed again, and afterwards the constellation volume can be reduced to refine the estimate of the unknown parameters.

The **CBMMAE** algorithm can be divided into three main stages : *i*) constellation design – establish a proper set of models in the space of unknown parameters; *ii*) tracking and bracketing – search the parameter space, by moving the models constellation, to localize the region containing the system parameters; *iii*) shrinking process – reduce the volume of the constellation. These steps are illustrated graphically in Fig. 6.3 for the case of a search in a two-dimensional parameter space.

### 6.4.1. Design of the Models Constellation

The steps of **CBMMAE** algorithm are illustrated in Fig. 6.3. The first step is the design of a constellation of models in the space spanned by the unknown parameters. Afterwards, the constellation of models is shifted on the parameter space by analyzing the posterior probabilities for each of its models until the system's parameters point is enclosed inside the constellation's volume. To track and enclose the system parameters point the constellation's topology must have the properties: *i*) possess a finite volume, different from zero; *ii*) and have an interior point. The first

## 6. System Identification and Parameters Estimation

property assures that the position of the system parameters' point can be localized inside a delimited region in the parameter space, which implies that for a space of  $n$  parameters one needs to have a constellation with at least  $n + 1$  distinct points. The interior point enables the algorithm to detect when the system's parameters point is localized inside the constellation volume by checking when the posterior probability of the corresponding model approaches one.

A simple topology that has the characteristics listed before is an hypercube with a center point. For a  $n$ -dimensional space an hypercube has  $2^n$  points, corresponding to its vertices, plus a center point, giving a constellation with a total of  $N = 2^n + 1$  points. Figure 6.3 displays an hypercube topology in step 1, with a central point for the initial constellation, for a two-dimensional parameters space. Hence, in the two-dimensional case one has a constellation with  $N = 5$  models with parameters pairs' coordinates in the space spanned by the unknown parameters denoted by:

$$\left\{ \left[ \gamma_{11} \ \gamma_{21} \right]; \left[ \gamma_{11} \ \gamma_{22} \right]; \left[ \gamma_{12} \ \gamma_{21} \right]; \left[ \gamma_{12} \ \gamma_{22} \right]; \left[ \gamma_{1c} \ \gamma_{2c} \right] \right\}, \quad (6.23)$$

where  $\left[ \gamma_{1c} \ \gamma_{2c} \right]$  is the center point, and  $\left[ \gamma_1^t \ \gamma_2^t \right]$  is the vector of system parameters,  $\boldsymbol{\gamma}$ , to be estimated.

If the parameters' ranges are known *a priori* the two-dimensional topology points can be initialized with the extreme values for their intervals. For example, if it is known that the parameter  $\gamma_1$  is in the interval  $\gamma_1 \in [a, b]$ , and  $\gamma_2$  is in the interval  $\gamma_2 \in [c, d]$ , the hypercube constellation can be initialized with the values:

$$\left[ \gamma_{11} \ \gamma_{12} \right] = \left[ a \ b \right], \quad \text{and} \quad \left[ \gamma_{21} \ \gamma_{22} \right] = \left[ c \ d \right], \quad (6.24)$$

and the center point of the topology can be chosen to have the values:

$$\left[ \gamma_{1c} \ \gamma_{2c} \right] = \left[ \frac{\gamma_{11} + \gamma_{12}}{2} \ \frac{\gamma_{21} + \gamma_{22}}{2} \right]. \quad (6.25)$$

The choice of values for the parameters of the initial constellation is flexible since the tracking and shrinking stages automatically adjust the parameters' range to more suitable values. Hence, if the initial parameters' range are unknown they can be initialized with a suitable interval.

### 6.4.2. System Identification and Parameters Estimation

To explore the parameter space each point in the topology is associated with one of the  $N$  estimators in the MMAE algorithm. Figure 6.2 displays the CBMMAE architecture, where a constellation adapter block is added to the standard MMAE structure. Like in the utilization of the MMAE in classical multiple-model adaptive control [Fekri et al., 2006], this block is added in a purely *ad-hoc* manner. Due to the property of the MMAE of Eq. (6.16), the posterior probability of the model whose parameters are closer to the true system with respect to the induced matrix metric tends to one. When the posterior probability of a model reaches a pre-establish

threshold value close to one,  $P_i \geq P_{th}$ , meaning that this particular model is identified as the closest to the true system. Taking as an example the initial constellation depicted in Fig. 6.3, and considering the more intuitive euclidian metric for the case, it is the model corresponding to the point  $[\gamma_{12} \ \gamma_{22}]$  that is identified as being the closest to the true system point,  $[\gamma_1^t \ \gamma_2^t]$ , by the MMAE algorithm. Therefore the constellation is moved towards the system point by translating the whole constellation by positioning its center point,  $[\gamma_{1c} \ \gamma_{2c}]$ , at the position of the last point identified as the closest to the system point. By repeating this procedure the models constellation will move on the parameter space until the center point,  $[\gamma_{1c} \ \gamma_{2c}]$ , is identified as being the closest to the system parameters' point,  $[\gamma_1^t \ \gamma_2^t]$ , that ends located inside the constellation's volume – the system parameters' point is bracketed. During the process of translation the points of the constellation keep their relative positions. The tracking and bracketing stage is illustrated by the step 2 in Fig. 6.3.

If the goal is to identify the system among a set of several possible configurations the tracking process is able to identify the system by identifying the region in the parameter space where the system point is located. This is useful when the system can have a high number of possible different configurations, which in many cases precludes the application of the standard MMAE algorithm directly by using every possible system configuration. In the CBMAE just  $2^n + 1$  models are needed, where  $n$  is the number of variable parameters, and the system is identified with an error equal to the intervals used for the model's parameters in the constellation, which is preestablished so that the system can be identified with the required precision.

However, if it is intended to estimate the system's parameters the models parameters' ranges in the constellation are systematically reduced. Whenever the true system is bracketed the volume of the constellation can be reduced by a shrinking process to refine the parameters estimation. A straight method for shrinking the constellation is to multiply each parameter interval by a scale factor  $\lambda < 1$ , and by keeping the previous constellation's interior point in the same position. Figure 6.3 illustrates the constellation shrinking process in step 3. Using the shrunken constellation in the MMAE the system point is tracked and bracketed again, as before, by analyzing the evolution of the posterior probabilities. By successively repeating the tracking and shrinking stages it is possible to identify the system and refine the estimation of the unknown system parameters. For a single parameter the shrinking process can be optimized by successively dividing the parameter's interval using the golden ratio [Martins, 2006].

The tracking process can be made faster by using a constellation with a high volume rather than by using a small constellation, however the subsequent shrinking process would take more steps. Another important remark of the CBMAE approach is that the number of models used in the constellation does not depend on the number of possible different configurations of the system, as in the standard application of the MMAE, but on the number of system's parameters that can actually vary. The hypercube constellation is optimal in terms of the number of models since it has the least number of models (plus one) relative to the number of unknown parameters.

## 6. System Identification and Parameters Estimation

In the usual application of the **MMAE** a system with  $n$  independent parameters that can take  $m$  different values, uses:

$$N = m^n \quad (6.26)$$

models for the estimators bank, that excludes its usability in many situations. However, in the **CBMMAE** using an hypercube constellation one always needs

$$N = 2^n + 1 \quad (6.27)$$

models, irrespective of the different values that the system's parameters can have, which corresponds to the minimum value that  $m$  can take in Eq. (6.26):  $m = 2$ . Therefore, the **CBMMAE** is also suited to the identification and parameters' estimation of systems whose parameters can vary continuously. Another remark about the **CBMMAE** is that its extension to a higher dimensional parameter space is straightforward by simply using, for example, an hypercube constellation with a central point.

### 6.4.3. Application to Nonstationary Systems

Due to the adaptive nature of the constellation the **CBMMAE** can be applied to the estimation of time-varying system parameters provided that the system's parameters change is less frequent than the time convergence of the algorithm.

In the **CBMMAE** whenever a vertex of the hypercube is identified as being the nearest to the true model the hypercube is re-centered on that point and its volume is kept unchanged. This process is repeated until the tracking process is completed. If the system parameters keep changing the constellation keeps moving around the parameter space until the system's parameters remain fixed sufficient time to be bracketed and estimated.

Once the true system parameter point is bracketed the constellation is shrunk to refine the parameters estimation. In the meanwhile, if the system parameters change to another point in space, the **MMAE** algorithm identifies successively one of the models on the hypercube's vertices as the nearest to the true system, which is an indication that the system point has moved.

After being tracked, and for a shrinking scale factor of  $\lambda = 0.5$ , which is equivalent to reduce the parameters intervals to half, if the constellation is shifted twice after being bracketed it means that the system's parameters point has moved. So the changing of the system parameters can be detected by counting the number of times the constellation is translated after the bracketing of system parameters' point. By using a general shrinking scale factor  $\lambda < 1$ , the change of the system parameters' point can be detected after being bracketed if a hypercube vertex is consecutively selected  $L \geq \lceil 1/\lambda \rceil$  times as the nearest to the system point. In a nonstationary system, after detecting that the system parameters point has moved, the constellation can be re-initialized to its initial volume, and optionally to its original position in the parameter space, to track the new system parameters' vector.

## 6.5. System State Estimators

The **MMAE** algorithm, and consequently the constellation based identification algorithm, relies on the estimation of the state of the system under analysis using a series of local estimators. In the case of linear systems the Kalman filter (**KF**) is the most commonly applied optimal state estimator [Gelb, 1974].

Optimal state estimation is considerably more difficult when the system contains nonlinear elements since the shape of the probability density functions of the signals and noise are altered as they are processed by the nonlinear elements. Therefore, the mean and standard deviation are incomplete descriptors of the probability density function, and the state estimate based on the conditional mean may be different from the one obtained based on the mode or median [Stengel, 1994].

For the state estimate of nonlinear systems there are several possible options, some more adequate than others depending on the characteristics of the problem [Simon, 2006]. To apply the **CBMMAE** algorithm to the identification and parameter estimation to nonlinear systems the extended Kalman filter (**EKF**) and the unscented Kalman filter (**UKF**) are employed as the state estimators. The adoption of suboptimal state estimators can produce satisfactory results because the random signals are summed in the estimators, and it is assured by the central limit theorem that the probability density functions of the sums tend to become Gaussian irrespective of the individual distributions of the signals. Moreover, the estimation process contains integration and/or summations, which tend to average out the non-Gaussianities, smoothing the effects of the nonlinearities in producing the state estimators.

Next, the different used types of optimal (**KF**) and suboptimal (**EKF** and **UKF**) state estimators are briefly reviewed.

### 6.5.1. The Kalman Filter

For a discrete time-invariant linear system, with linear observations, Eq. (6.2) and Eq. (6.4) can be written as:

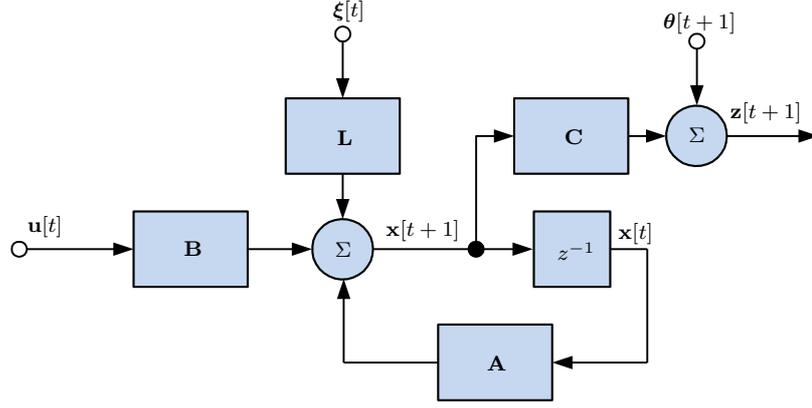
$$\mathbf{x}[t+1] = \mathbf{A}\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] + \mathbf{L}\boldsymbol{\xi}[t], \quad t = 0, 1, 2, 3, \dots, \quad (6.28)$$

$$\mathbf{z}[t] = \mathbf{C}\mathbf{x}[t] + \boldsymbol{\theta}[t], \quad t = 1, 2, 3, \dots \quad (6.29)$$

The system's dynamics and the associated observations are represented in the block diagram of Fig. 6.4, where  $z^{-1}$  represents a unit discrete-time delay.

The Kalman filter (**KF**) provides a recursive way to obtain the estimate of the state and the associated error covariance matrix of a linear system using the previous estimate of the system state and the current observations (of some) of the state variables [Gelb, 1974]. The estimation process is separated into a *predict cycle* and an *update cycle*. In the predict cycle the predicted state estimate, denoted by  $\hat{\mathbf{x}}[t+1|t]$ , and the predicted covariance error matrix, denoted by  $\boldsymbol{\Sigma}[t+1|t]$ , at the time instant

## 6. System Identification and Parameters Estimation



**Figure 6.4.** Block structure of a linear system dynamics in state-space.

$t + 1$ , are predicted using the previous state estimate  $\hat{\mathbf{x}}[t|t]$  and the last input to the system,  $\mathbf{u}[t]$ , by:

$$\hat{\mathbf{x}}[t + 1|t] = \mathbf{A}\hat{\mathbf{x}}[t|t] + \mathbf{B}\mathbf{u}[t], \quad (6.30)$$

$$\Sigma[t + 1|t] = \mathbf{A}\Sigma[t|t]\mathbf{A}^T + \mathbf{L}\Xi\mathbf{L}^T, \quad (6.31)$$

respectively, where  $\hat{\mathbf{x}}[t|t]$  is the updated state estimate, and  $\Sigma[t|t]$  is the updated error covariance matrix at the time instant  $t$ .

In the update cycle, the state estimate is updated using the new observation vector,  $\mathbf{z}[t + 1]$ , according to:

$$\hat{\mathbf{x}}[t + 1|t + 1] = \hat{\mathbf{x}}[t + 1|t] + \mathbf{H}[t + 1]\mathbf{r}[t + 1], \quad (6.32)$$

where the Kalman filter gain,  $\mathbf{H}[t + 1]$ , and the residual,  $\mathbf{r}[t + 1]$ , defined in Eq. (6.13), are given respectively by:

$$\mathbf{H}[t + 1] = \Sigma[t + 1|t + 1]\mathbf{C}^T\Theta^{-1}, \quad (6.33)$$

and

$$\mathbf{r}[t + 1] = \mathbf{z}[t + 1] - \mathbf{C}\hat{\mathbf{x}}[t + 1|t]. \quad (6.34)$$

The residual vector  $\mathbf{r}[t]$  has zero mean:

$$E\{\mathbf{r}[t]\} = \mathbf{0}, \quad (6.35)$$

and covariance:

$$\mathbf{S}[t + 1] = \mathbf{C}\Sigma[t + 1|t]\mathbf{C}^T + \Theta. \quad (6.36)$$

The state covariance matrix is updated according to:

$$\Sigma[t + 1|t + 1] = \Sigma[t + 1|t] - \Sigma[t + 1|t]\mathbf{C}^T\mathbf{S}[t + 1]^{-1}\mathbf{C}\Sigma[t + 1|t]. \quad (6.37)$$

The state estimate and the error covariance matrix equations can be propagated using the initial conditions  $\hat{\mathbf{x}}[0|0] = E\{\mathbf{x}[0]\}$ , and  $\Sigma[0|0] = \text{cov}[\mathbf{x}[0]; \mathbf{x}[0]]$ .

### 6.5.2. The Extended Kalman Filter

When the system dynamics and/or the observation equations are not linear a suboptimal state estimator must be used. The extended Kalman filter (**EKF**) is a widely used nonlinear estimation technique. It follows the Kalman filter reasoning by linearizing the system dynamics and the observation equations by picking up the linear terms of its Taylor series expansion about the current state estimate [Anderson and Moore, 2005; Simon, 2006].

Considering the nonlinear system dynamics of Eq. (6.2) the predicted state estimate and the error covariance matrix are given by:

$$\hat{\mathbf{x}}[t+1|t] = \mathbf{f}(\hat{\mathbf{x}}[t|t], \mathbf{u}[t], \mathbf{0}, t), \quad (6.38)$$

$$\Sigma[t+1|t] = \hat{\mathbf{A}}[t]\Sigma[t|t]\hat{\mathbf{A}}^T[t] + \hat{\mathbf{L}}[t]\Xi[t]\hat{\mathbf{L}}^T[t], \quad (6.39)$$

where the Jacobian matrices of the dynamics are defined as:

$$\begin{aligned} \hat{\mathbf{A}}_{n \times n}[t] &= \left. \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\xi}, t)}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}[t|t], \boldsymbol{\xi}=\mathbf{0}}, \\ \hat{\mathbf{L}}_{n \times p}[t] &= \left. \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\xi}, t)}{\partial \boldsymbol{\xi}} \right|_{\mathbf{x}=\hat{\mathbf{x}}[t|t], \boldsymbol{\xi}=\mathbf{0}}. \end{aligned} \quad (6.40)$$

After a nonlinear observation, expressed by Eq. (6.3), the state estimate is also updated according to Eq. (6.32), however, the Kalman filter gain is computed from:

$$\mathbf{H}[t+1] = \Sigma[t+1|t+1]\hat{\mathbf{C}}^T[t+1] \left[ \hat{\mathbf{D}}[t+1]\boldsymbol{\Theta}(t+1)\hat{\mathbf{D}}^T[t+1] \right]^{-1} \quad (6.41)$$

where the Jacobian matrices for the nonlinear observation equation are given by:

$$\begin{aligned} \hat{\mathbf{C}}_{m \times n}[t+1] &= \left. \frac{\partial \mathbf{h}(\mathbf{x}, \boldsymbol{\theta}, t)}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}[t+1|t], \boldsymbol{\theta}=\mathbf{0}}, \\ \hat{\mathbf{D}}_{m \times m}[t+1] &= \left. \frac{\partial \mathbf{h}(\mathbf{x}, \boldsymbol{\theta}, t)}{\partial \boldsymbol{\theta}} \right|_{\mathbf{x}=\hat{\mathbf{x}}[t+1|t], \boldsymbol{\theta}=\mathbf{0}}. \end{aligned} \quad (6.42)$$

The residual is:

$$\mathbf{r}[t+1] = \mathbf{z}[t+1] - \mathbf{h}(\hat{\mathbf{x}}[t+1|t], \mathbf{0}, t+1), \quad (6.43)$$

whose covariance is obtained from:

$$\mathbf{S}[t+1] = \hat{\mathbf{C}}[t+1]\Sigma[t+1|t]\hat{\mathbf{C}}^T[t+1] + \hat{\mathbf{D}}[t+1]\boldsymbol{\Theta}[t+1]\hat{\mathbf{D}}^T[t+1]. \quad (6.44)$$

The system state error covariance matrix is updated according to:

$$\Sigma[t+1|t+1] = \Sigma[t+1|t] - \Sigma[t+1|t]\hat{\mathbf{C}}^T[t+1]\mathbf{S}[t+1]^{-1}\hat{\mathbf{C}}[t+1]\Sigma[t+1|t]. \quad (6.45)$$

Like in the **KF** the initial conditions to propagate the state estimate equation are  $\hat{\mathbf{x}}[0] = E\{\mathbf{x}[0]\}$ , and  $\Sigma[0|0] = \text{cov}\{\mathbf{x}[0]; \mathbf{x}[0]\}$  for the error covariance matrix.

### 6.5.3. The Unscented Kalman Filter

Instead of propagating the state statistics through the nonlinear system dynamics and/or through the nonlinear observation equation, the unscented Kalman filter (**UKF**) obtains the state statistics from a minimal set of sample points, called sigma points, that are propagated through the nonlinearities [Julier and Uhlmann, 1997, 2004].

The **UKF** filter is based on the unscented transform that permits to calculate the mean and covariance of a nonlinear transformation of a random variable [Julier, 2002]. Starting with a  $n$ -dimensional random variable  $\mathbf{x}$ , with mean vector  $\hat{\mathbf{x}}$  and covariance matrix  $\Sigma_{\mathbf{xx}}$ , a set of  $2n + 1$  points, termed sigma points, are obtained from the expressions:

$$\begin{aligned} \mathbf{x}_0 &= \hat{\mathbf{x}}, & W_0 &= \kappa/(n + \kappa); \\ \mathbf{x}_i &= \hat{\mathbf{x}} + \left( \sqrt{(n + \kappa)\Sigma_{\mathbf{xx}}} \right)_i^T, & W_i &= 1/2(n + \kappa); \\ \mathbf{x}_{i+n} &= \hat{\mathbf{x}} - \left( \sqrt{(n + \kappa)\Sigma_{\mathbf{xx}}} \right)_i^T, & W_{i+n} &= 1/2(n + \kappa); \end{aligned} \quad (6.46)$$

where  $\mathbf{x}_i$  are the sigma points and  $W_i$  are their respective weights to be used in the unscented transform. The parameter  $\kappa \in \mathbb{R}$  is used to fine tune the higher order moments statistics, and  $\left( \sqrt{(n + \kappa)\Sigma_{\mathbf{xx}}} \right)_i$  is the  $i$ th row of the square root matrix of  $(n + \kappa)\Sigma_{\mathbf{xx}}$ .

Transforming each sigma point by the nonlinearity:

$$\mathbf{y}_i = \mathbf{f}[\mathbf{x}_i], \quad (6.47)$$

the mean and covariance of the transformed random variable  $\mathbf{y}$  are approximated by:

$$\hat{\mathbf{y}} = \sum_{i=0}^{2n} W_i \mathbf{y}_i, \quad (6.48)$$

and

$$\Sigma_{\mathbf{yy}} = \sum_{i=0}^{2n} W_i (\mathbf{y}_i - \hat{\mathbf{y}})(\mathbf{y}_i - \hat{\mathbf{y}})^T. \quad (6.49)$$

For a nonlinear system with nonlinear observation equation, as described by Eq. (6.2) and Eq. (6.4), respectively, the **UKF** filter starts by calculating the sigma points and weights according to Eq. (6.46), where the current state has the mean vector  $\hat{\mathbf{x}}[t|t]$ , and the covariance matrix  $\Sigma[t|t]$ . If the system noise  $\boldsymbol{\xi}[t]$ , and/or the observation noise  $\boldsymbol{\theta}[t]$ , are not additive the state vector can be augmented to  $\mathbf{x}_a^T[t] = [\mathbf{x}^T[t] \quad \boldsymbol{\xi}^T[t] \quad \boldsymbol{\theta}^T[t]]$ , to include the noise dynamics, so that the state mean vector becomes:

$$\hat{\mathbf{x}}_a[t|t] = \begin{bmatrix} \hat{\mathbf{x}}[t|t] \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad (6.50)$$

and the respective covariance matrix is:

$$\Sigma_a[t|t] = \begin{bmatrix} \Sigma[t|t] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Theta[t] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Xi[t] \end{bmatrix}. \quad (6.51)$$

The predicted state estimate is computed from:

$$\hat{\mathbf{x}}[t+1|t] = \sum_{i=0}^{2n} W_i \hat{\mathbf{x}}_i[t+1|t], \quad (6.52)$$

where  $\hat{\mathbf{x}}_i[t+1|t]$  are the transformed sigma points by the nonlinearity:

$$\hat{\mathbf{x}}_i[t+1|t] = \mathbf{f}(\hat{\mathbf{x}}_i[t|t], \mathbf{u}[t], \mathbf{0}, t). \quad (6.53)$$

The predicted state covariance matrix is given by:

$$\Sigma[t+1|t] = \sum_{i=0}^{2n} W_i (\mathbf{x}_i[t+1|t] - \hat{\mathbf{x}}[t+1|t]) (\mathbf{x}_i[t+1|t] - \hat{\mathbf{x}}[t+1|t])^T + \Xi[t]. \quad (6.54)$$

To update the estimate given a new observation a new set of sigma points is generated from Eq. (6.46) using the predicted state estimate  $\hat{\mathbf{x}}[t+1|t]$  and the predicted covariance matrix  $\Sigma[t+1|t]$ . Then the predicted observation is obtained from:

$$\hat{\mathbf{z}}[t+1|t] = \sum_{i=0}^{2n} W_i \hat{\mathbf{z}}_i[t+1|t], \quad (6.55)$$

where the observations vectors  $\hat{\mathbf{z}}_i$  are obtained with the sigma points transformed by the observation equation:

$$\hat{\mathbf{z}}_i[t+1|t] = \mathbf{h}(\hat{\mathbf{x}}_i[t+1|t], \mathbf{0}, t+1). \quad (6.56)$$

The residual covariance matrix is obtained from:

$$\mathbf{S}[t+1] = \sum_{i=0}^{2n} W_i (\mathbf{z}_i[t+1|t] - \hat{\mathbf{z}}[t+1|t]) (\mathbf{z}_i[t+1|t] - \hat{\mathbf{z}}[t+1|t])^T + \Theta[t] \quad (6.57)$$

In statistical terms the Kalman gain is given by:

$$\mathbf{H}[t+1] = \Sigma_{\mathbf{xz}}[t+1|t+1] \mathbf{S}^{-1}[t+1]. \quad (6.58)$$

The updated error covariance matrix is computed from:

$$\begin{aligned} \Sigma[t+1|t+1] &= \Sigma[t+1|t] - \mathbf{H}[t+1] \mathbf{S}[t+1] \mathbf{H}^T[t+1] \\ &= \Sigma[t+1|t] - \mathbf{H}[t+1] \Sigma_{\mathbf{xz}}^T[t+1], \end{aligned} \quad (6.59)$$

where the cross covariance of the state and the observation is estimated with:

$$\Sigma_{\mathbf{xz}}[t+1] = \sum_{i=0}^{2n} W_i (\mathbf{x}_i[t+1|t] - \hat{\mathbf{x}}[t+1|t]) (\mathbf{z}_i[t+1|t] - \hat{\mathbf{z}}[t+1|t])^T. \quad (6.60)$$

As before, the updated state estimate is given by Eq. (6.32) with the Kalman gain given by Eq. (6.58) and using the predicted measurement Eq. (6.55) in the residual. For the case of an augmented state vector the system and the observation noise should be dropped from Eq. (6.54) and Eq. (6.57), since it was already included in the estimates of Eq. (6.50) and Eq. (6.51).

## 6.6. Experimental Results

The **CBMMAE** is applied to the identification and parameters estimation of a **RGC** neuron model. The stochastic leaky integrate-and-fire (**SLIF**) model is employed to evaluate the algorithm performance when applied to linear systems, and, afterwards the **SLIF** model is complemented with a nonlinear block to obtain the firing rate to test the algorithm performance in the estimation of a nonlinear system's parameters. In the context of physiological modeling this approach follows a modular viewpoint, where the blocks of the model reflects the knowledge about the system under identification.

### 6.6.1. Linear System Identification

The **SLIF** model neuron model was presented in Sec. 3.3.4, and is used to illustrate the application of the algorithm to a linear system. The neuron can be divided into three main functional components, as described in Sec. 2.2. First, the neuronal dendrites collect the inputs signals - in the form of spikes - from the presynaptic neurons and transmit them to the soma. In the soma the collected signals are integrated, and when a potential threshold is surpassed an action potential - a spike, is generated and conducted through the axon to other neurons.

In the junction between two neurons the output axonal branch and the input dendrite are brought very close forming a synapse. When the action potential reaches the synapses of the axonal terminal of the presynaptic neuron it provokes the un-leash of neurotransmitter molecules into the synaptic cleft that are collected by the postsynaptic neuron, triggering the opening of ions channels in its membrane. As detailed in Sec. 3.4.1 the ionic current induced in a neuron's synapse in response to the presynaptic action potential can be modeled as an  $\alpha$ -function:

$$I_s(t) = \alpha_0 t e^{-t/\tau_\alpha} H(t). \quad (6.61)$$

The  $\alpha$ -function can be obtained as the solution of the differential equation<sup>1</sup>:

$$\frac{d^2 I_s(t)}{dt^2} + \frac{2}{\tau_\alpha} \frac{dI_s(t)}{dt} + \frac{1}{\tau_\alpha^2} I_s(t) = \alpha_0 \rho(t), \quad (6.62)$$

where  $\alpha_0$  establishes the peak amplitude of the postsynaptic current pulse, and  $\tau_\alpha$  is the pulse decay time constant. The input spike train signal  $\rho(t)$ , defined by the series of Dirac delta functions:

$$\rho(t) = \sum_i \delta(t - t_i), \quad (6.63)$$

models the barrage of incoming action potentials from different presynaptic neurons that impinge the post-synaptic neuron at distinct time instants  $t_i$ .

---

<sup>1</sup>Equation (6.62) can be obtained by calculating the Laplace transform of Eq. (6.61).

The **SLIF** model, presented in Sec. 3.3.4, models the subthreshold potential of the neuron membrane through the differential equation:

$$\tau_m \frac{dV_m(t)}{dt} = -V_m(t) + R_m I_s(t) + \sigma_m \xi_m(t), \quad (6.64)$$

where  $\tau_m = R_m C_m$  is the membrane time constant, and  $\sigma_m$  gives the power of the noise  $\xi_m(t)$ . The noise  $\xi_m(t)$  in Eq. (6.64) is considered to be Gaussian white noise with mean and covariance:

$$\begin{aligned} E\{\xi_m(t)\} &= 0, \\ E\{\xi_m(t)\xi_m(t')\} &= \delta(t - t'). \end{aligned} \quad (6.65)$$

The noise term,  $\sigma_m \xi_m(t)$ , in Eq. (6.64), models the influence of the synaptic current on the membrane potential resulting from the stochastic spike arrival from background activity in the neural network. The trajectory of the membrane's voltage, given by the solution of Eq. (6.64), for a single input current pulse with the shape of an  $\alpha$ -function, as given by Eq. (6.61), and not considering the influence of the noise ( $\sigma_m = 0$ ), is:

$$V_m(t) = \frac{\alpha_0 \Delta^2}{C_m} \left[ e^{-t/\tau_m} - e^{-t/\tau_\alpha} \left[ 1 + \frac{t}{\Delta} \right] \right] H(t), \quad (6.66)$$

where  $1/\Delta = 1/\tau_\alpha - 1/\tau_m$ .

Using the state-space formalism, the subthreshold dynamics of the neuron according to the **SLIF** model, expressed by the first order differential equation of Eq. (6.64), and considering the input synaptic current generated by the input spikes modeled by the second order differential equation Eq. (6.62), can be written in the vectorial form as the third order differential equation system:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{L}\boldsymbol{\xi}(t), \quad (6.67)$$

by considering the following convenient vector of state-space variables:

$$\mathbf{x}(t) = \begin{bmatrix} dI_s(t)/dt + 1/\tau_\alpha I_s(t) \\ I_s(t) \\ V_m(t) \end{bmatrix}, \quad (6.68)$$

where the dynamic's matrix of the system is:

$$\mathbf{A} = \begin{bmatrix} -1/\tau_\alpha & 0 & 0 \\ 1 & -1/\tau_\alpha & 0 \\ 0 & 1/C_m & -1/\tau_m \end{bmatrix}. \quad (6.69)$$

The input vector is equal to the neural function:

$$\mathbf{u}(t) = [\rho(t)], \quad (6.70)$$

## 6. System Identification and Parameters Estimation

Parameter	Value
Membrane capacitance $C_m$	250 pF
Membrane time constant $\tau_m$	10 ms
Synapse current peak $\alpha_0$	0.38089 $\mu\text{A/s}$
Synapse current decay time $\tau_\alpha$	0.32564 ms

**Table 6.2.** Neuron model parameters

where the input matrix is just equal to:

$$\mathbf{B} = \begin{bmatrix} \alpha_0 \\ 0 \\ 0 \end{bmatrix}. \quad (6.71)$$

The system noise vector,  $\boldsymbol{\xi}(t)$  in Eq. (6.67), contains two different components to model the spontaneous activity observed in neurons in the absence of a controlled input stimulus. A component that mimics the random arrival of spikes at the neuron synapses,  $\xi_s(t)$ , that establishes an input background current noise. This component is equivalent to the sum of a noise component in the second member of Eq. (6.62). This noise term is also characterized by Gaussian white noise with mean and covariance:

$$\begin{aligned} E\{\xi_s(t)\} &= 0; \\ E\{\xi_s(t)\xi_s(t')\} &= \delta(t - t'). \end{aligned} \quad (6.72)$$

The noise vector also comprises the noise voltage component that models the membrane's potential oscillations due to the variations in the ionic currents through the membrane, expressed by the noise term,  $\xi_m(t)$ , in the second member of Eq. (6.64). Therefore, the system noise vector is composed by:

$$\boldsymbol{\xi}(t) = \begin{bmatrix} \xi_s(t) \\ \xi_m(t) \end{bmatrix}. \quad (6.73)$$

The noise vector is multiplied by the noise gain matrix:

$$\mathbf{L} = \begin{bmatrix} \sigma_s & 0 \\ 0 & 0 \\ 0 & \sigma_m \end{bmatrix}, \quad (6.74)$$

where  $\sigma_s$  and  $\sigma_m$  are the synaptic input current and membrane voltage noise gains.

The values for the parameters of the neuron model are taken to conform with the literature [Gewaltig et al., 2001]. The membrane's capacitance and time constant are given in Table 6.2. The parameters of the differential equation for the  $\alpha$ -function in Eq. (6.62) were also chosen to obtain physiologically realistic values. The current peak amplitude  $\alpha_0$ , and its time decay constant  $\tau_\alpha$ , were chosen so that the amplitude of the voltage in Eq. (6.66) is 0.14 mV with a rise time of 1.7 ms which give the values displayed on Table 6.2.

Considering that the synaptic current is sampled at regular intervals of time like the neuron membrane voltage, is as is commonly found in many experimental setups [Picchini et al., 2008; Martins and Sousa, 2009], the discrete time measurement equation:

$$\mathbf{z}(t) = \mathbf{C}\mathbf{x}(t) + \boldsymbol{\theta}(t), \quad t = nT_s, \quad n \in \mathbb{Z}, \quad (6.75)$$

can be joined to the state-space equations, where  $T_s$  is the sampling period. The observation matrix is:

$$\mathbf{C} = \begin{bmatrix} 0 & A_I & 0 \\ 0 & 0 & A_V \end{bmatrix} \quad (6.76)$$

where  $A_I$  and  $A_V$  are the current,  $I_s(t)$ , and voltage,  $V_m(t)$ , measurements' amplification. The noise vector  $\boldsymbol{\theta}(t)$  in Eq. (6.75) models the errors in the observations.

For the numerical simulation of the continuous-time neuron model Eq. (6.67) must be converted into its discrete-time equivalent. A common procedure is to discretize the independent variable and apply the Euler approximation for the derivative [Rotter and Diesmann, 1999; Martins and Sousa, 2009]<sup>2</sup>. For a sampling period equal to  $T_s$  Eq. (6.67) becomes:

$$\mathbf{x}[nT_s + T_s] = [\mathbf{I} + T_s\mathbf{A}]\mathbf{x}[nT_s] + \mathbf{B}\mathbf{u}[nT_s] + \boldsymbol{\xi}[nT_s], \quad n \in \mathbb{Z}, \quad (6.77)$$

with the discretization process the independent variable,  $t$ , is a multiple of the sampling period,  $T_s$ , that takes only values on the sampling grid:  $0, T_s, 2T_s, \dots, nT_s$ . The approximation in Eq. (6.77) is valid for the input signal,  $\mathbf{u}[t]$ , composed by a series of Dirac delta functions positioned at the sampling grid, as expressed by Eq. (6.70), so that the input matrix  $\mathbf{B}$  is not affected by the discretization process. Dropping the dependence of the independent variable on  $T_s$ , Eq. (6.77) can be written as:

$$\mathbf{x}[t + 1] = \mathbf{A}_d\mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] + \boldsymbol{\xi}[t], \quad t \in \mathbb{Z}, \quad (6.78)$$

where  $\mathbf{A}_d = \mathbf{I} + T_s\mathbf{A}$ .

The discrete system intensity noise matrix must be calculated so that the noise power of the discrete system is independent of the sampling period [Gelb, 1974]. Thus, the covariance of the discrete white noise is given by:

$$\boldsymbol{\Xi}_d = \text{cov} [\boldsymbol{\xi}[kT_s]; \boldsymbol{\xi}[nT_s]] = T_s\mathbf{L}\boldsymbol{\Xi}\mathbf{L}^T\delta_{kn}, \quad (6.79)$$

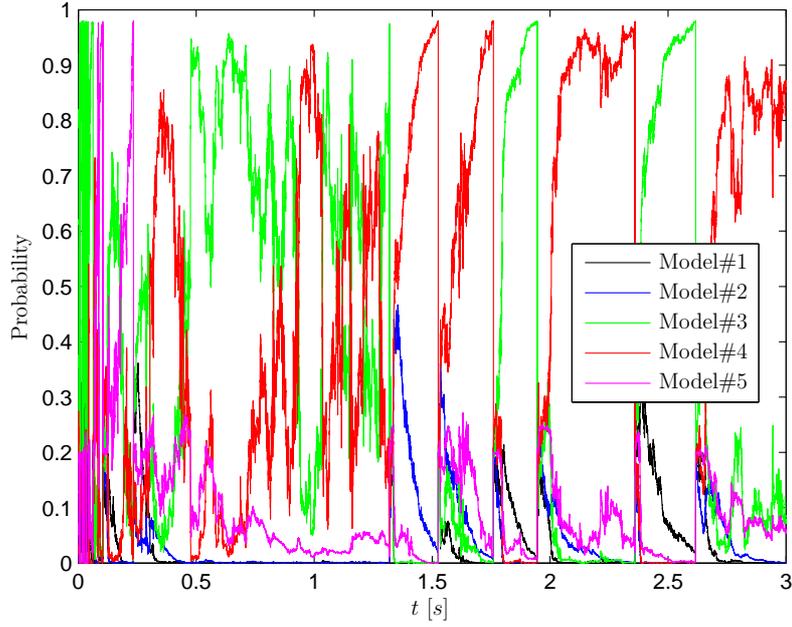
where  $\delta_{kn}$  is the Kronecker delta function.

The neuron's currents and voltages are very small so that in order to avoid possible numerical instabilities a similarity transformation should be used. A transformation matrix is used so that the neuron current is measured in nanoampere and the membrane voltage is measured in millivolt. The matrix for this similarity transform is:

$$\mathbf{T} = \begin{bmatrix} 10^9 & 0 & 0 \\ 0 & 10^9 & 0 \\ 0 & 0 & 10^3 \end{bmatrix}. \quad (6.80)$$

---

<sup>2</sup>The Euler approximation for the first derivative of the function  $f(t)$  with respect to time, between the time instants  $t$  and  $t + T_s$ , is:  $df(t)/dt = [f(t + T_s) - f(t)]/T_s$



**Figure 6.5.** Probabilities evolution of a set of models ( $n = 5$ ) with the CBMMAE.

Applying the similarity transform  $\mathbf{T}$ , the matrix equation of Eq. (6.78) becomes:

$$\mathbf{x}[t + 1] = \mathbf{T}\mathbf{A}_d\mathbf{T}^{-1}\mathbf{x}[t] + \mathbf{T}\mathbf{B}\mathbf{u}[t] + \mathbf{T}\boldsymbol{\xi}[t], \quad t \in \mathbb{Z}, \quad (6.81)$$

where now the current and the voltage variables of the state-vector of Eq. (6.68) come in nanoampere and millivolt.

The CBMMAE algorithm was implemented and tested using the MATLAB environment. The third order vector differential equation was discretized and different configurations for the unknown parameters were used to test the algorithm. The true model's parameters are given in Table 6.2. It was used a discretization step equal to  $T_s = 0.1$  ms. The input spike train,  $\rho[t]$ , was generated with a Poisson distribution with an expected number of 45 spikes/s. The system was simulated using the noises variances:  $\sigma_s^2 = 10^{-18}$  A<sup>2</sup> and  $\sigma_m^2 = 10^{-6}$  V<sup>2</sup> in Eq. (6.79). The intensity of observation noise is  $\boldsymbol{\Theta} = 10^{-4}\mathbf{I}_{2 \times 2}$ , and the current gain is  $A_I = 1$  and the voltage amplification gain is  $A_V = 1$  in the measurement matrix of Eq. (6.76).

For  $n$ -unknown parameters the initial posterior probabilities used for each model are:

$$P_k = 1/N, \quad k = 1, \dots, N; \quad \text{with } N = 2^n + 1. \quad (6.82)$$

Whenever the algorithm signals that the posterior probability of one of the models has surpassed the threshold probability:  $P_i \geq P_{th}$ , the models constellation is shifted as described in Sec. 6.4.1, and the posterior probabilities are re-initialized according to Eq. (6.82). It was used  $P_{th} = 0.98$  for the threshold probability to update the models constellation. When the models constellation is moved/shrunked

the predicted and the updated stationary covariance matrices are computed by solving the respective Ricatti equation. The state estimate for each model is initialized to  $\hat{\mathbf{x}}_k[0|0] = \mathbf{0}_{3 \times 1}$ ,  $k = 1, \dots, N$ , where  $N = 2^n + 1$ , and with the initial covariance matrix  $\Sigma_k[0|0] = 10^{-3} \mathbf{I}_{3 \times 3}$ . Whenever the constellation is shifted, the state estimate for each new model is initialized to the previous system state estimate:  $\hat{\mathbf{x}}_k[t|t] = \hat{\mathbf{x}}[t-1|t-1]$ ,  $k = 1, \dots, N$ , obtained by Eq. (6.14). Figure 6.5 displays the time evolution of the posterior probabilities for 5 models used in the estimation of  $n = 2$  parameters:  $\boldsymbol{\gamma} = [1/\tau_\alpha \quad 1/\tau_m]^\top$ .

The results displayed in Fig. 6.6 to Fig. 6.7 were obtained by averaging the results for  $M = 25$  Monte-Carlo runs of the experiment with a number of unknown parameters  $n = 3$ , that implies a constellation of  $N = 9$  models. The unknown parameters' vector is:

$$\boldsymbol{\gamma} = \begin{bmatrix} 1/\tau_\alpha \\ 1/\tau_m \\ C_m \end{bmatrix} \quad (6.83)$$

The parameters of the models used in the estimators bank were initialized, as described in Sec. 6.4.1, by using the intervals:

$$\begin{aligned} \gamma_1 &\in [600, 1200] \\ \gamma_2 &\in [400, 600] \\ \gamma_3 &\in [10^{-11}, 10^{-10}]. \end{aligned} \quad (6.84)$$

The shrinking scale factor is equal to  $\lambda = 0.5$  in the simulations. The errors of the parameters' estimates are displayed in Fig. 6.7, and were obtained with the expression:

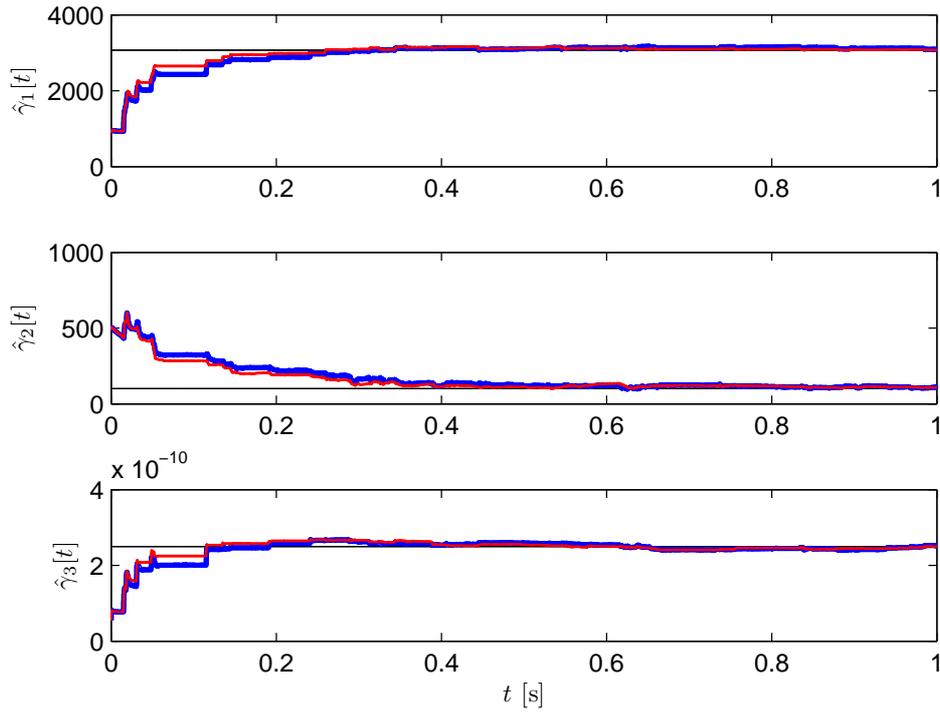
$$\tilde{\gamma}[t] = \frac{1}{M} \sum_{i=1}^M (\gamma[t] - \hat{\gamma}_i[t]). \quad (6.85)$$

The plots of Fig. 6.6 and Fig. 6.7 show that the CBMMAE is effective in the identification of the system and in the tuning of the 3 unknown parameters. From Fig. 6.7 it can be seen that even starting with a set of parameters far from the true system point, the estimation error goes to zero and is kept low along time for the parameter scale.

Figure 6.8 displays a zoom of one parameter estimate with the associated estimated error obtained with Eq. (6.19), corresponding to the standard deviation of the parameter estimate, that is within the limits of the true parameter value.

Figure 6.9 shows the algorithm behavior in the identification and parameter tuning of a time-variant system, and the associated error. From this figure it can be seen that the CBMMAE follows the change of the parameter, with the estimation error decaying rapidly. In this simulation the constellation topology is reset to its initial form whenever the algorithm detects that the system parameters' point is no longer bracketed.

## 6. System Identification and Parameters Estimation



**Figure 6.6.** Time evolution of the system parameters' estimates for the linear system (— KF; — UKF ).

### 6.6.2. Nonlinear System Identification

To evaluate the performance of the algorithm in the identification and parameter estimation of nonlinear systems, the SLIF neuron model is extended. The neuron model is composed by the linear block described in Sec. 6.6.1, that generates the neuron subthreshold membrane potential from the set of evoked potentials (or spikes) based on the SLIF neuron model, complemented by a nonlinear block to generate the firing rate from the membrane potential. The parameters of the linear components of the model are given in Table 6.2.

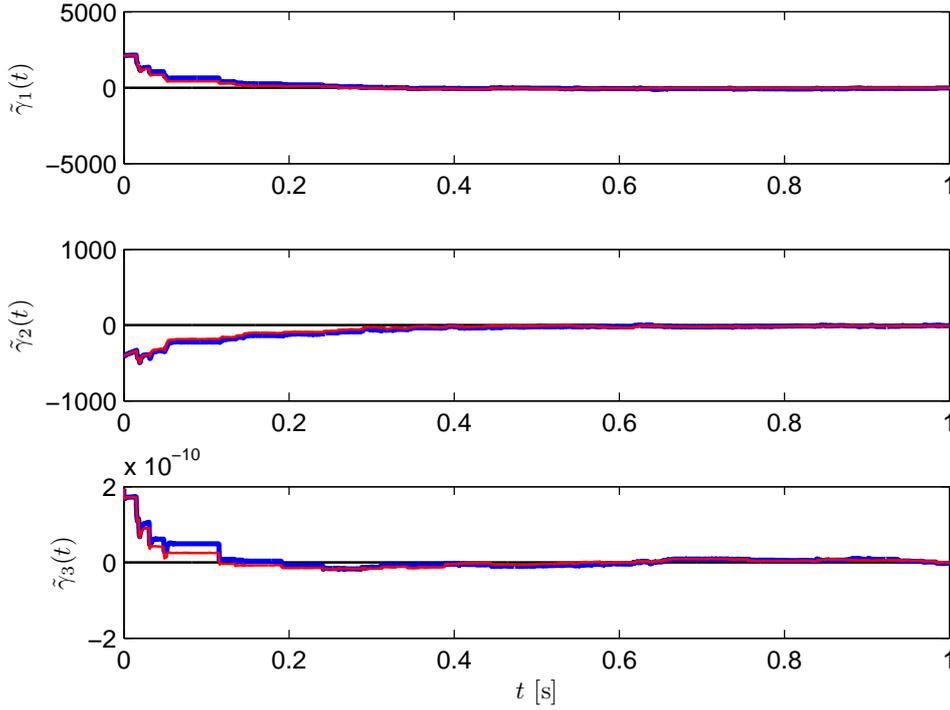
The discrete firing rate  $r[t]$  is obtained from the membrane potential  $V_m[t]$  given by the linear model of Eq. (6.67), with the system matrices of Eq. (6.69) and Eq. (6.71), by applying a sigmoid such that:

$$r[t] = \mathcal{S}(V_m[t]). \quad (6.86)$$

The sigmoid has the form:

$$\mathcal{S}(V_m[t]) = \frac{r_0}{1 + e^{-\frac{V_m[t] - V_0}{\tau_r}}}, \quad (6.87)$$

where  $r_0$  establishes the maximum value for the firing rate,  $V_0$  shifts the sigmoid along the horizontal axis, and  $\tau_r$  establishes the slope of the sigmoid between its zero



**Figure 6.7.** Estimation error for the 3 unknown linear system parameters (— KF; — UKF ).

minimum value and its maximum value  $r_0$ . The discrete state-space nonlinear system model can be written as:

$$\mathbf{x}[t+1] = \begin{bmatrix} \mathbf{A}_d & \mathbf{0}_{3 \times 1} \\ \mathcal{S}(\mathbf{x}_3[t]) \end{bmatrix} \mathbf{x}[t] + \mathbf{B}\mathbf{u}[t] + \boldsymbol{\xi}[t], \quad (6.88)$$

where the state vector of Eq. (6.68) is augmented to include the firing rate:

$$\mathbf{x}[t] = \begin{bmatrix} dI_s(t)/dt + 1/\tau_\alpha I_s(t) \\ I_s(t) \\ V_m(t) \\ r[t] \end{bmatrix}. \quad (6.89)$$

The linear and nonlinear blocks of the model are apparent from Eq. (6.88). The parameters of the sigmoid are adjusted to the firing rate of the **RGC** data set referenced in [Gerstner and Naud, 2009] by following the methodology described in the adjustment of the nonlinearity in the model presented in Sec. 3.5.3. The noise matrix of Eq. (6.73) is augmented to include the firing rate noise:

$$\boldsymbol{\xi}(t) = \begin{bmatrix} \xi_s(t) \\ \xi_m(t) \\ \xi_r(t) \end{bmatrix}, \quad (6.90)$$

## 6. System Identification and Parameters Estimation

so that, besides the current and voltage, now the firing rate is also corrupted by noise. The noise power matrix is now:

$$\mathbf{L} = \begin{bmatrix} \sigma_s & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \sigma_m & 0 \\ 0 & 0 & \sigma_r \end{bmatrix}. \quad (6.91)$$

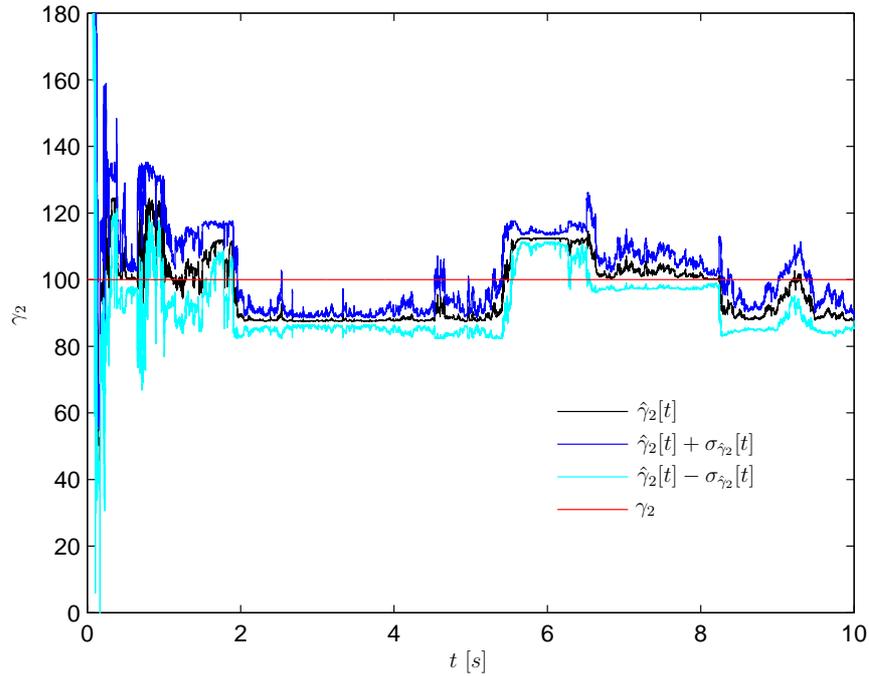
The values for the noise variances are:  $\sigma_s^2 = 10^{-18} \text{ A}^2$ ,  $\sigma_m^2 = 10^{-6} \text{ V}^2$  and  $\sigma_r^2 = 10^{-2} \text{ Hz}^2$ .

The input signal  $\mathbf{u}[t] = \sum \delta[t - t_i]$  is generated from a Poisson distribution with a mean rate of 45 spikes/s. (The **RGC** data presented in [Gerstner and Naud, 2009] has an average of 42 spikes/s). The observation matrix used is:

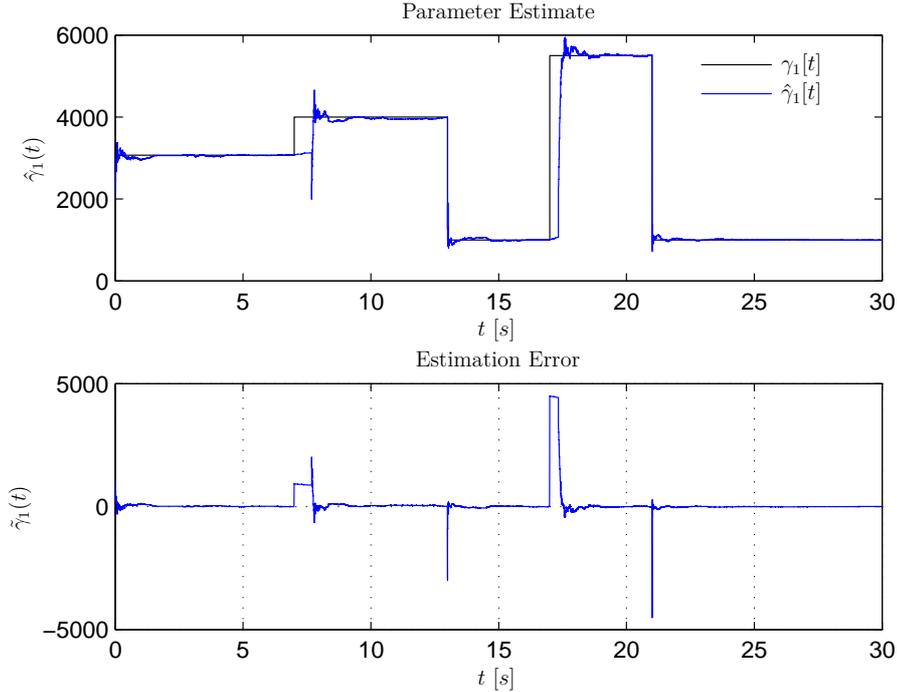
$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}, \quad (6.92)$$

meaning that only the nonlinear firing rate is observed so that the other state variables are inferred through the nonlinearity. The observation noise has the intensity:  $\Theta = 10^{-4}$ .

The constellation models' probabilities were initialized with  $P_k = 1/N$ ,  $k = 1, \dots, N$  with  $N = 2^n + 1$ , with initial the mean estimate  $\hat{\mathbf{x}}_k[0|0] = \mathbf{0}_{4 \times 1}$  and the



**Figure 6.8.** Zooming of the time evolution of a parameter estimation and the associated error.



**Figure 6.9.** Parameter estimation for a linear time-variant system and respective estimation error for parameter  $\gamma_1 = 1/\tau_\alpha$ .

initial covariance  $\Sigma_k[0|0] = 10^{-3}\mathbf{I}_{4 \times 4}$ . The threshold probability was set to  $P_{th} = 0.95$ . The shrinking scale factor is  $\lambda = 0.5$ . The results were obtained by averaging  $M = 25$  Monte-Carlo runs of the experiment. Figure 6.10 displays the evolution of the estimates of two parameters of the linear block computed with Eq. (6.18) using a constellation of  $N = 2^2 + 1 = 5$  models. The estimation error displayed in Fig. 6.11 was computed with from Eq. (6.85). Figure 6.10 shows the results obtained in the estimation of the two parameters vector:

$$\boldsymbol{\gamma} = \begin{bmatrix} 1/\tau_\alpha \\ 1/\tau_m \end{bmatrix}, \quad (6.93)$$

for the nonlinear system that results in a constellation of  $N = 5$  models, by applying the CBMAE algorithm.

The results displayed in Fig. 6.10 and Fig. 6.11, where the EKF and the UKF are used as the state-estimators in the MMAE, show that the algorithm is robust in the estimation of the parameters even when they are initialized with values quite different from their real values, and are observed through the nonlinearity.

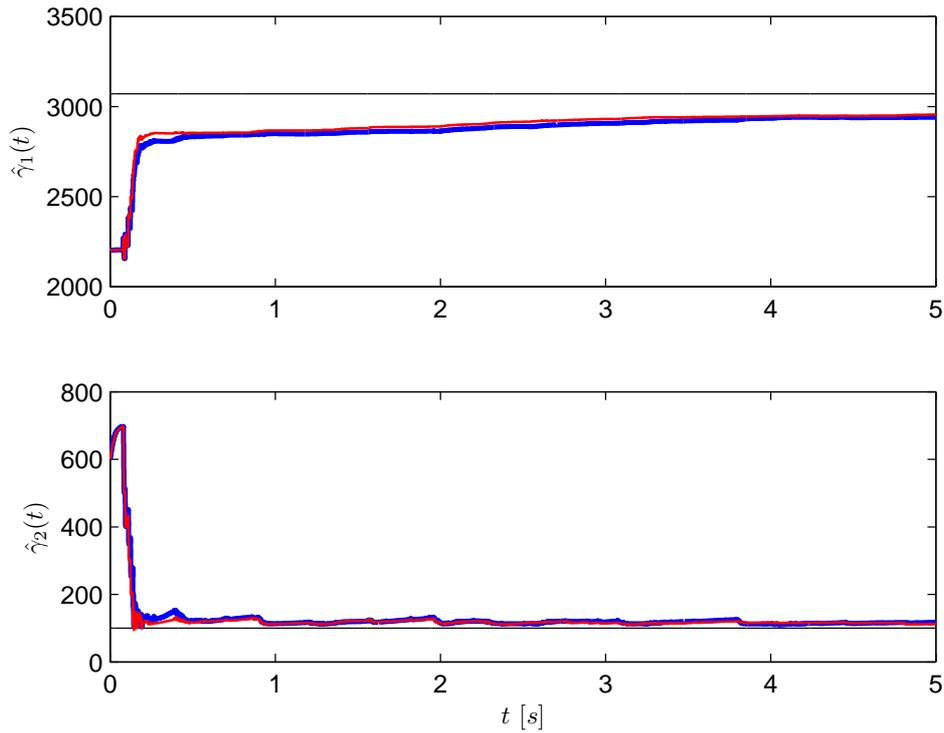


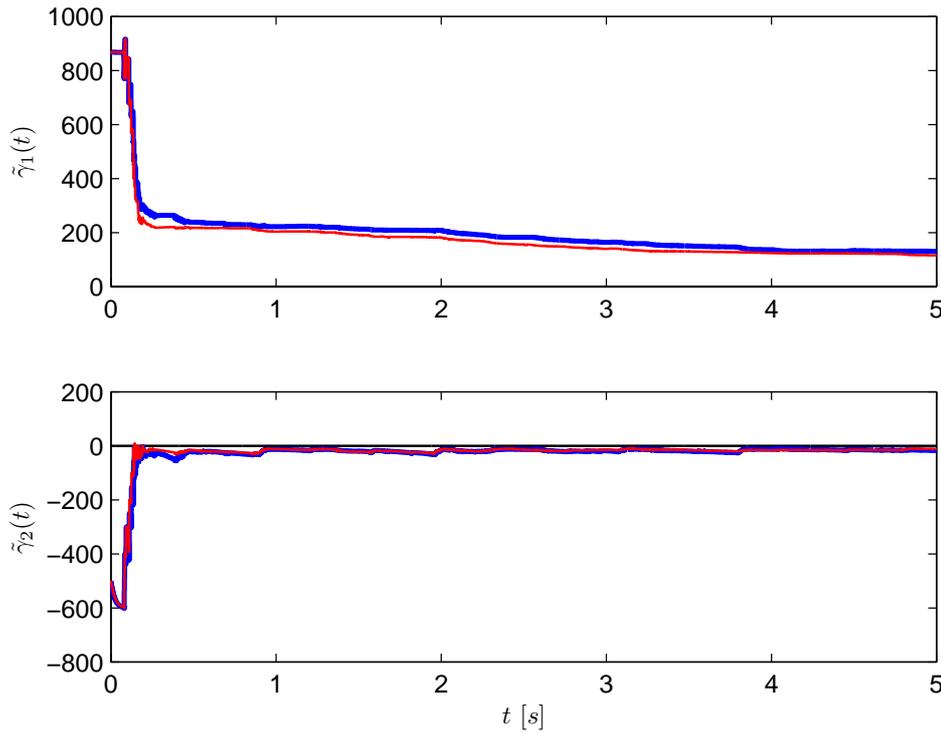
Figure 6.10. Estimates of the nonlinear system parameters (— EKF; — UKF ).

## 6.7. Conclusions

This chapter proposes a new algorithm for system identification and simultaneously for parameter estimation based on the **MMAE**, by applying a properly designed constellation of models that are adapted according to the posterior probabilities of each model, that is designated by **CBMMAE**.

The **CBMMAE** algorithm was first applied in the identification and parameters' estimation of a linear neuron model. Afterwards, the linear neuron model was complemented with a sigmoidal nonlinearity to produce the firing rate. In both situations the **CBMMAE** was able to estimate the system parameters, even in the nonlinear case where the system state variables used in the state estimators are inferred through the nonlinearity.

The experimental results show the effectiveness of the constellation method in the identification of the system and in the estimation of its parameters. Due to its adaptive nature the **CBMMAE** is directly applicable to time-varying systems. In the case of time-variant systems, the change of the parameters are detected and the algorithm is able to estimate the new system's parameters by reinitializing the constellation to its starting configuration. The only restriction, that is also shared by the standard **MMAE** method, is that the parameters should not change more frequently than the convergence time of the algorithm. However, even if the system's



**Figure 6.11.** Estimates errors for the nonlinear system parameters (— EKF; — UKF ).

parameters keep changing frequently the **CBMMAE** algorithm continues to search the parameter space towards the true point, that is eventually determined when the system stabilizes in a functioning point.

An additional advantage of the **CBMMAE** over the traditional application of the **MMAE** algorithm is that the number of models depend on the number of parameters than can change, and not on the number of possible configurations of the system, and the number of models in the estimators bank is optimal in terms of the number of parameters. Considering that only 1 parameter can change the algorithm needs  $2^1 + 1 = 3$  models, independently of the number of different values that this parameter can take.

An additional advantage is that there is no need to know the exact values of the real system parameters. If even if their range is not known, despite one usually have some clues from the physical problem under analysis, the algorithm can be initialized with an arbitrary interval for the system parameters. The only cost is that the tracking process and possibly the shrinking process are done in more steps and take more time.

The experiments reported used the same model for every element of the constellation, however the bank of estimators can have different models with different structures, or several bank of filters can be used in parallel with different structures.

In the constellation shrinking process the same scale factor is applied to every di-

## *6. System Identification and Parameters Estimation*

mension of the constellation, a further development of this algorithm is to analyze the dependence of the system dynamic's matrix on each parameter in order to optimize the adaptation of the parameters constellation according to this dependence. An additional issue related with this algorithm concerns with the optimal constellation volume. Starting the algorithm with a wider initial constellation topology, the search in the parameter space is faster, while the sequel parameters tuning is slower. Therefore, the obtention of the optimal values between the constellation dispersion and the shrinking scales is another challenge.

*Every limit is a beginning as well as an ending.*

George Eliot (1819-1880)

# 7

## Conclusions

**T**HE advances of signal processing techniques and of electronics, in conjunction with the development of biocompatible materials, opened the possibility to design prosthesis capable to directly interface to the nervous systems, even at the brain level. However, the challenges are enormous, there are several technological issues to solve, namely those related with the biological interfaces, and with the ethical problems connected to the clinical human testing of such devices. In the meanwhile, these systems are continuously perfected by including the most recent technological advances. This chapter presents the main conclusions of this thesis and points some future work directions.

The thesis addresses several problems within the broad field of bioelectronic vision. Although, the emphasis is on the engineering issues related to this subject, there was a deliberate concern to fill the gaps between the biological and the electronics field. There was a historical distance between these two fields, that has been narrowing in the recent years due to the increasing interest from electrical engineers to apply the advances achieved in electronics and in computers systems for biomedical applications. Therefore, several incursions are made along the thesis into different biological topics considered relevant to understand the challenges at hand. However, there was an effort to keep an engineering perspective. The biological background is introduced in chapter two, and chapter three presents the most relevant neuron models, in general, and retina models, in particular, that are used and analyzed in the subsequent chapters.

A requirement in the development of bioelectronic vision is the modeling of the signal processing occurring in the retina. The retina converts the incoming stimulus in the form of a pattern of light into a series of spikes. The contribution to this topic was the development of two new retina models. These models contribute to the family of spiking neuron models and to the firing rate models. The first model,

## 7. Conclusions

belonging to the spiking neuron models' family, is termed dynamic retina model. The model uses both the spike and the stimulus history to modulate the spike generation current. It can have two different configurations: it can use only the static filtering of the stimulus and of the spike history to model the spike generating current, or use an additional dynamic filtering of the stimulus and spike history to modulate the spike generation current. The comparison of the model performance with these two alternative configurations discloses the importance of the use of dynamic blocks in retina modeling, since the output error decreases significantly with the inclusion of the dynamic blocks on the model.

The second model is based on the classic linear-nonlinear-Poisson (**LNP**) model. However, this new model departs from the traditional structure by selecting and including several excitatory and inhibitory kernels in the linear block. The kernels, which are obtained from the analysis of the spike triggered covariance matrix, are included by analyzing its influence on the **NMSE** and resulted in a significant decrease of the output error compared to the traditional single use of the **STA**. This model introduced the use of additive logistic regression in conjunction with the generalized additive models in retina's modeling. This approach has an additional advantage because it provides a direct way to clearly identify the type of influence of each kernel in the model: if it is excitatory or suppressive, and its relative contribution to the retina model response, by analyzing the shape of the nonlinear functionals obtained.

How the neuronal metrics can be used in the tuning and assessment of neuronal models in general, and of retina models in particular is also part of the thesis. After the organization of several neuronal metrics that are classified into three different classes, specifically: firing rate metrics, spike train metrics and firing events metrics, these are analyzed and its application to the tuning and assessment of a set of representative retina models is discussed. The analysis of the results of the application of the different metrics to sets of real spike trains permitted to propose a method to select the values for its parameters, so that they provide meaningful results both in the context of model tuning and in terms of the evaluation of the model's performance. Specifically, the spike time metrics depend on a free parameter that establishes its sensitivity to a particular characteristic of the encoding neural mechanism in terms of the spike trains. If the value of this parameter is not chosen correctly the produced results are misleading and the metrics are useless to evaluate neural responses.

Another relevant result is that, in general, the metrics do not refine each other, so that to fully assess a neuron model different metrics should be used depending on the relevant coding characteristics under evaluation. We show that there is even a metric that is useless in the evaluation of neural response's and in the models assessment.

The evaluation of the neural metrics to the assessment of the retina models also unveiled the fact that there is not an optimal universal retina model irrespective of the type of stimulus and the metric used to evaluate the results. The answer to the question of what are the relevant neuronal encoding features used by the nervous system does not reunite the consensus among researchers. While the knowledge about the neuronal encoding mechanisms increases and the discovery of what are the *de*

*facto* important features of the neural code relevant for carrying neuronal information, the development of models that implement and mimic the coding mechanisms and metrics that reflect and evaluate the codification features must be pursued. Certainly, the final conclusion can only be drawn by performing experiments with real prosthesis in humans, where the plasticity of the nervous system will definitely play an important role. Hence, the effort to develop robust retina models that reflect the new discoveries about the neural coding mechanism must be continued.

Another area addressed in this thesis is the identification of systems. The contribution to this area is the proposal of a novel algorithm for system identification and parameter tuning, termed constellation based multiple-model adaptive estimation (CBMMAE) algorithm. The CBMMAE algorithm relies on the classical multiple-model adaptive estimation (MMAE) algorithm but uses a different structure for the bank of filters and adds a models constellation adapter block. The CBMMAE algorithm starts by the design of an adaptive constellation of models whose parameters are refined by the constellation adapter until a model close to the real system is identified, and/or its parameters estimated. The results show the effectiveness of the constellation method to identify the parameters both for linear and nonlinear systems. The adaptability of the models' constellation makes this algorithm appropriate for the identification of time-varying systems. It is also shown that the number of models in the constellation is optimal relatively to the number of unknown parameters when compared to the common use of the multiple-model adaptive estimators algorithm. Also, due to the adaptive nature of the constellation, the algorithm deals easily with high uncertainties in the parameters values. The regular structure of the constellation makes the algorithm suitable to be further exploited to be used in parallel computational systems to further reduce the computational time.

Besides the further development of retina models and metrics that should reflect the new findings about the neural encoding of information there is also room for the improvement of the new identification algorithm proposed. The CBMMAE algorithm proved to be effective in the identification and estimation of linear and nonlinear (neuron) models. However, it opened a vast field of applications and several improvements can be made and tested. For the identification of nonlinear systems, there are several types of nonlinearities present in nature, and the performance of the constellation can behave differently depending on the nonlinearity.

The research and test of alternative constellation topologies is another topic that can be further developed. Presently the constellation has  $2^n + 1$  points, for an  $n$ -dimensional space (each one relative to a different model) which corresponds to a hypercube with a center point. In a high dimensional unknown parameters' space it can be advantageous to reduce the number of models used in the estimators bank. This can be achieved by using a simplex constellation, which has  $n + 1$  points. However, in its present form the CBMMAE algorithm does not give a quantitative measure of the distance between the models and the unknown system, only the nearest one. To use a simplex constellation the history of the movement of the constellation by the tracking and bracketing process along the parameter's space must be retained so that the shift of the simplex can be adjusted according to its movement history and

## 7. Conclusions

will need a careful research on details of the matter.

Another topic that should be further developed is the independent adjustment of the scale shrinking factor for each system's parameter to be identified. In the **CBMMAE** algorithm proposed the scale shrinking factor applied to the constellation is equal for every direction in the parameter space (type of parameter). However, by analyzing the dependence of the system dynamic's matrix on each unknown parameter it should be possible to derive a rule to obtain the scale shrinking factor for each individual parameter to optimize the tracking/shrinking process. Moreover, the formulation of the **CBMMAE** algorithm is quite general so that it is expected to see its natural spread to other areas of application and to different types of systems.

# Appendices





# Blindness In The World

**I**N the International Statistical Classification of Diseases, Injuries and Causes of Death, produced by World Health Organization (**WHO**), *low vision* is defined as visual acuity of less than 6/18 but equal to or better than 3/60, or a corresponding visual field loss to less than 20°, in the better eye with the best possible correction. *Blindness* is defined as visual acuity of less than 3/60, or a corresponding visual field loss to less than 10°, in the better eye with the best possible correction. *Visual impairment* includes both low vision and blindness [World Health Organization, 2010].

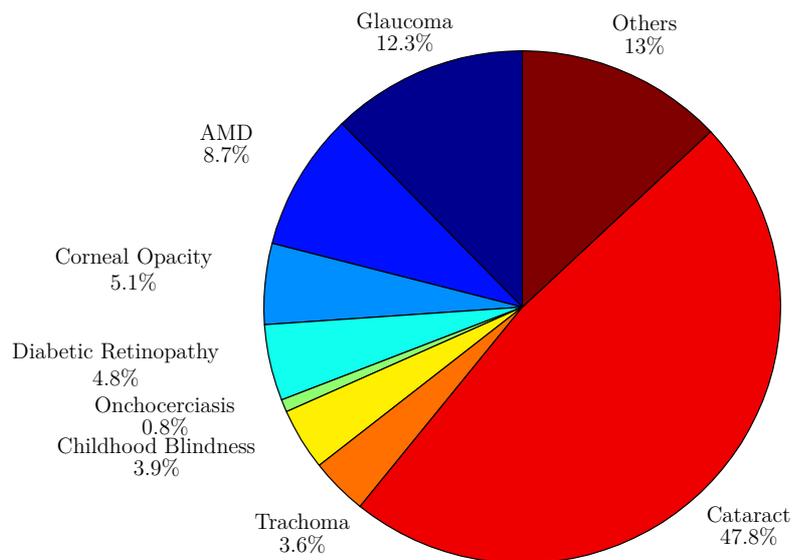
This appendix review the main causes of blindness and the associated diseases. At the end the principal consortia labs involved in the development of visual prosthesis are listed.

## A.1. Extent and Main Causes of Visual Impairment

In 2002, it was estimated by the **WHO** that over than 161 million people were visually impaired, of whom, 124 million people had low vision and 37 million were blind [World Health Organization, 2004]. However, refractive error as a cause of vision impairment is not included in these figures, which implies that the actual global magnitude of vision impairment is bigger. The latest WHO estimates including refractive errors add to the previous number and effectively double the estimated total number of visually-impaired people worldwide, bringing it to some 314 million people globally. The estimates also confirm that uncorrected refractive errors are a leading cause of visual impairment worldwide [World Health Organization, 2006].

The impairments of profound blindness may have origin in degenerative retinal diseases, or in brain injuries that affect the superior vision centers due to accidents or to direct surgical intervention (e.g. for a tumor removal).

## A. Blindness In The World



**Figure A.1.** Global causes of blindness due to eye diseases, excluding refractive errors (data from [World Health Organization, 2004]).

**Cataract** One of the major causes of vision impairments in the world is cataract. Cataract is normally related with the ageing process and is characterized by the opacity of the eye's lens which impedes the regular flow of light. The actual treatment for this disease consists of a surgical intervention to replace the eye opaque lens with an artificial intraocular lens.

**Glaucoma** The second major cause of blindness is glaucoma. Glaucoma occurs when the aqueous humor does not drain out correctly and the pressure within the eye becomes too high, compromising the blood vessels of the optic nerve head and eventually the axons of the ganglion cells, causing the death of these vital cells. The reduction of intraocular pressure is imperative to avoid total blindness. This disease affects the retinal nervous system and can cause permanent damages.

**AMD** The third cause of blindness worldwide is the **AMD**. In some persons the macula, responsible for fine detail in the center of the visual area, degrades with age due to unknown reasons. The pigment epithelium behind the retina degenerates forming drusen and leaks fluid behind the foveal macular area. The cones in the fovea die causing central vision loss and making impossible to read or see fine details. The **AMD** is a major cause of blindness in the developed countries due to the number of elderly people above 70 years of age.

**Corneal opacities** The graphic in Fig. A.1 shows that the next cause of blindness is corneal opacity, which occurs when the cornea becomes scarred, preventing light from passing through the cornea to the retina, and causing, in some cases, the cornea to appear white or clouded over. Corneal opacity can be caused by infection, like conjunctivitis, or caused by the herpes virus, measles, injury, or inflammation of the eye caused by a stroke or by a chemical agent. In many cases it can be reversed by adequate treatment, that may include surgery.

**Trachoma** Trachoma, other cause of blindness, is an infection caused by an organism called *Chlamydia trachomatis* that can be treated with antibiotics. It is a common cause of blindness worldwide but rare in developed countries.

**Diabetic retinopathy** A significant percentage of blindness is caused by diabetes, which is a serious problem in the industrialized countries. Approximately 90% of all diabetic patients have retinopathy after twenty years. Diabetic retinopathy is characterized by anomalies in the blood vessels that get blocked, leak, or multiply in an uncontrolled manner, leading to irreversible blindness.

**Children blindness** A major cause of blindness among children is the deficiency of vitamin A, particularly in children under 5 years. Within this percentage is blindness caused by premature born, infant retinopathy and cataract. Blindness among children is a major problem due to their lifetime expectation. It is estimated that 1.4 million children below age 15 are blind.

**Onchocerciasis** Onchocerciasis is responsible for blindness particularly in the African and Latin America countries. Onchocerciasis is a disease transmitted by a parasite spread by flies in river side areas.

**RP** RP is a terrible disease presently with no cure. RP is an inherited disease that causes degeneration of the retina and pigment excess. First it provokes night blindness, then tunnel vision and, as more and more peripheral retina becomes damaged and the rods die, progresses gradually to total blindness.

Figure A.1 displays the distribution of the principal causes of blindness in around the world, and their prevalence based on the data collected from the World Health Organization, authority for health within the United Nations Organization, in 2002 [World Health Organization, 2004]. The RP is included within others causes of blindness grouped in the general class "others".

The blindness distribution is not geographically uniform. About 90% of visually impaired people live in developing countries. Statistics says that females have a higher risk of being visually impaired. In terms of age it is estimated that about 82% of visually impaired people have more than 50 years old. The "Vision 2020:

## A. *Blindness In The World*

The Right to Sight<sup>1</sup> is a global initiative for the elimination of avoidable blindness [World Health Organization, 2007], launched jointly by the United Nations World Health Organization (WHO), the International Agency for the Prevention of Blindness (IAPB) and international eye care institutions and corporations. One of the largest and most productive eye care facilities in the world is the Aravind Eye Care System that was established in 1976 in Madurai, India. It has treated over 2.3 million out patients and performed over 270 0000 surgeries, mainly people living in rural India, and it was the recipient of the first edition of the António Champalimaud<sup>1</sup> Vision Award in 2007.

---

<sup>1</sup>The Champalimaud Foundation (<http://www.fchampalimaud.org>), based on Lisbon and created in 2004 in the bequest of the late portuguese industrialist and entrepreneur António Champalimaud, supports individual researchers and research teams working at medical science, namely in the field of neuroscience

# B

## Characterization of the Neural Response

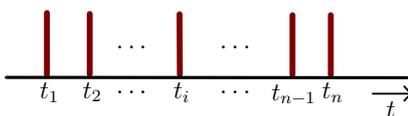
**T**HIS appendix presents several quantities and mathematical tools that allow a description of the neural response. For an engineering audience, the literature on computational neuroscience in general, and for retina modeling in particular, is full of somewhat unfamiliar terms, definitions, notions and concepts. Despite the fact that, in the essence, many of the concepts are coincident (so that their main differences reside primarily in the terminology), sometimes there are some slight changes in the mathematical definitions. In other cases, however, completely new concepts exist.

In the early days of experimental neuroscience, the neural response was measured by counting the number of spikes occurring in a given time window applied at the onset of the stimulus presentation. In the experimental setups in use today, the retina is repeatedly stimulated with the same stimulus, and the acquired results are averaged over the experimental trials.

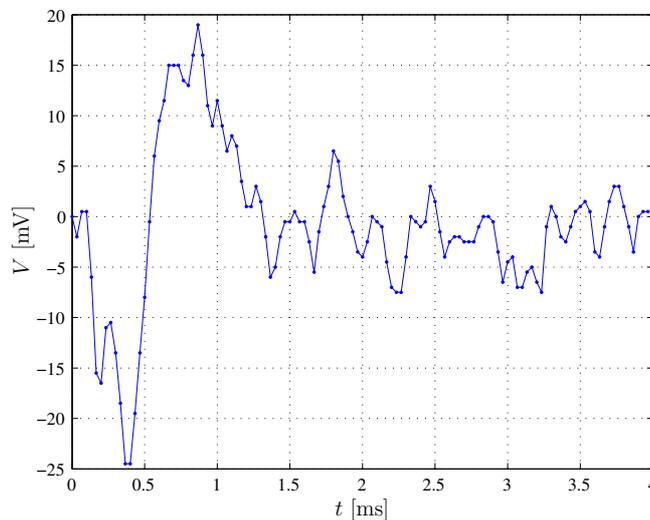
We start by describing the main issues in the stimulation and recording of the responses of the retina. A first look at the resulting spike sequences reveals that the neurons' responses are not equal for the same stimulus. Instead, the responses show some variability from trial to trial, meaning that the neural behavior has a certain level of randomness. This randomness implies that the neural code does not possess a one-to-one correspondence, but instead the same stimulus triggers different neural responses despite the fact that neurons have a certain degree of similarity. To uncover the neural code, it is essential to quantify the degree of randomness in the neural responses. Thus, we will present the main tools from probability and stochastic process theory used to describe neural responses.

The remainder of the chapter is devoted to presenting the methods and quantities

## B. Characterization of the Neural Response



**Figure B.1.** Representation of a spike train.



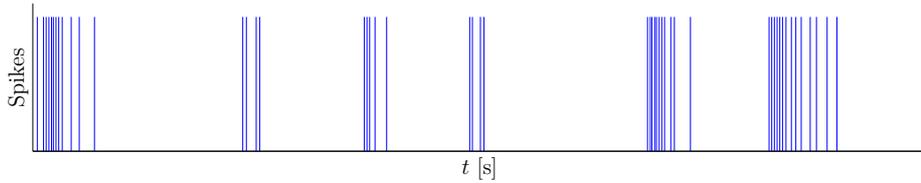
**Figure B.2.** Spike waveform of a rabbit's retinal ganglion cell.

used to analyze and characterize the neural response of a nervous retinal cell, so that we can quantify and thus reproduce its behavior. Although the methods and tools presented in the subsequent sections could be applied to the analysis and characterization of the responses of any neural cell, we will specifically target the neural response of the retinal cells.

### B.1. Spikes: The Essence of the Neural Code

Regardless of the input stimulus type – continuous or discrete – the neural output is always a discrete sequence of voltage pulses, also called evoked potentials and briefly termed *spikes*, that are positioned at time instants  $\{t_i\}$ , with  $i = 1, \dots, n$ , in the train, like the representation in Fig. B.1. The waveform of the evoked potentials has a stereotyped shape for a given class of neural cells. Moreover, it is common practice to classify the ganglion cells by the shape of their evoked potentials [Wandell, 1995]. Figure B.2 shows the spike waveform of a rabbit OFF-type RGC, sampled with a frequency  $f_s = 30$  kHz, with the sampled points dotted.

Because of the stereotypical form of the spikes, despite the fact that their time lengths, amplitudes, and shapes have slight variations, the information carried to the brain must be encoded in the spikes' temporal occurrence instants. With regard to this fact, a spike train can be represented by a time series of equal amplitude bars,



**Figure B.3.** Neuronal response function of a retinal ganglion cell.

with a bar located at every instant where a spike occurs. Figure B.3 shows a graphical representation of a spike train segment from the response of a rabbit transient brisk OFF-type ganglion cell when excited with a Gaussian random stimulus [Keat et al., 2001].

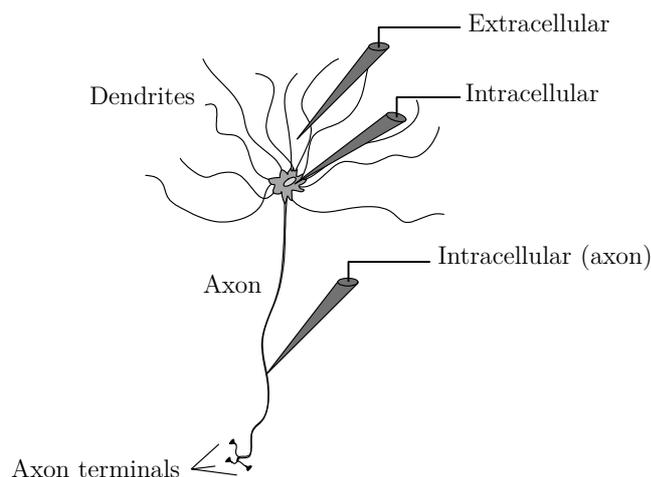
A common, and not completely answered, question is: how does a spike train represent the sensory input, the internal states of the brain and the subsequent motor control? The answer to this question can be viewed as the search for a translation dictionary for this peculiar language [Meister and Berry II, 1999]. An interesting, and funny, illustration used to introduce this problem is the concept of the *homunculus* [Rieke et al., 1997]. The homunculus is a metaphor in which a little man, placed in the brain, receives the spike trains produced by the sensory organs in response to stimuli, and tries to figure out what stimulus the organism is sensing. After decoding the stimuli from the spike trains, the homunculus has to generate a spike sequence to communicate with the organism's members as a reaction to the environmental changes. This metaphor encompasses the two processes involved in the neural code: coding and decoding. The homunculus receives the sensory data encoded as a spike train, decodes it in order to perceive the stimulus, and has to encode the response again as a spike train to communicate its reaction.

In the development of a retina model, the main goal is to understand how stimuli are encoded. This process is intimately related with the decoding process, because when a stimulus is classified, by decoding the spike train it generates, the underlying coding mechanism is revealed.

A current topic in the research community is the identification of the relevant characteristics of the spike train that convey information [Eggermont, 1998]. Owing to the variability of the neurons' responses to the same stimulus, several researchers claim that the only significant feature in the spike train is the firing rate, and not the individual time instants of each spike occurrence. This viewpoint is usually called the *rate coding* approach [Berry II and Meister, 1998].

On the other hand, several physiologists claim, based on some recent studies, that by stimulating the retinal ganglion cells repeatedly, with the same visual stimulus, the cells' responses are quite regular with only a limited variability [Uzzell and Chichilnisky, 2004]. Thus, some recent retina models group the set of spikes from the neuron's responses into firing events, and assume that the time occurrence of the first spike, the time interval between the first and second spike, and the total number of spikes within each event are relevant to encode information [Berry et al., 1997; de Ruyter van Steveninck and Bialek, 1988]. This perspective of the neural code is

## B. Characterization of the Neural Response



**Figure B.4.** Cellular recording of neuronal signals.

termed a *time coding* approach [Berry II and Meister, 1998]. Thus, in the time coding approach, the precise time relations between the spikes from the same neuron are considered to be meaningful.

Another current debate relates to the information conveyed by a retinal ganglion cell; whether it is independent of nearby cells, or if the information coded by the cell population response is relevant. Some studies have shown that 90% of the information is coded by the ganglion cell alone, and only 10% of the information is encoded in the population response [Nirenberg et al., 2001].

### B.1.1. Retina Stimulation and Responses Recording

The retina data have been obtained by sampling the responses of stimulated animals RGCs. The most commonly used retinas for data recording have provenance from different animals, mainly vertebrates such as rabbits, salamanders, turtles, monkeys, and even humans. Some invertebrates, like the blowfly, are also used [Keat et al., 2001; Berry II et al., 1999; Chichilnisky, 2001; de Ruyter van Steveninck and Bialek, 1988].

The spike's voltage course can be recorded extracellularly or intracellularly. Figure B.4 depicts the possible different locations of the recording electrode to measure neural activity. The signals recorded intracellularly are stronger and low-noise, while the extracellular recordings are weaker and subject to noise from neighboring cells [Dayan and Abbot, 2001]. Usually, an extracellular electrode collects electrical signals originating on different neurons, which causes some difficulties in its analysis, particularly for classifying the cells according to the waveforms associated with spikes. Thus, if the recorded signals come from different neural cells of the same type, it is impossible to perform their classification, and so it is assumed that the spikes were produced by the same cell.

Currently, in the most common apparatus, the photoreceptor layer of the retina is

excited with visual light patterns, and the neural responses coming from the RGCs are measured with an array of electrodes (see for example [Meister et al., 1994]). Another, less common, alternative is to directly stimulate the retina with electrical currents induced through the electrodes, like in [Grumeta et al., 2000].

The visual stimuli used to stimulate the retina RGC can have many variants and the choice of stimulus is usually tied to the particular retinal response that we want to characterize. In its most general form, a visual stimulus can have a spatial, temporal and light wavelength dependency:  $s(x, y, t, \lambda)$ , or  $s(\mathbf{r}, t, \lambda)$  where  $\mathbf{r}$  is the column vector  $\mathbf{r} = [x \ y]^T$ . Usually, the analysis of the retina's neuronal response is restricted to a particular subset of this general ensemble. The visual stimuli can be divided into two main categories in terms of their spatial and temporal variability. The retina's response is also commonly investigated for a particular wavelength, so that the dependence of the stimulus on the light wavelength can be dropped and we can focus only on the image luminance. The analysis of the chromaticity response of the retina is made by varying the stimulus wavelength.

In terms of their time variation, the stimuli are usually classified as deterministic or stochastic. In terms of the spatial behavior of stimuli, they can be organized as full-field or uniform type, meaning that they are spatially constant and do not convey any spatial information, or they can carry spatial information by changing spatially. The deterministic stimuli can have a closed mathematical description, such as the ON-OFF-type, or can be composed of simple patterns changing in time, such as a moving bar drifting along the visual field with a given speed. The stochastic stimuli can change both spatially and temporally in a random fashion.

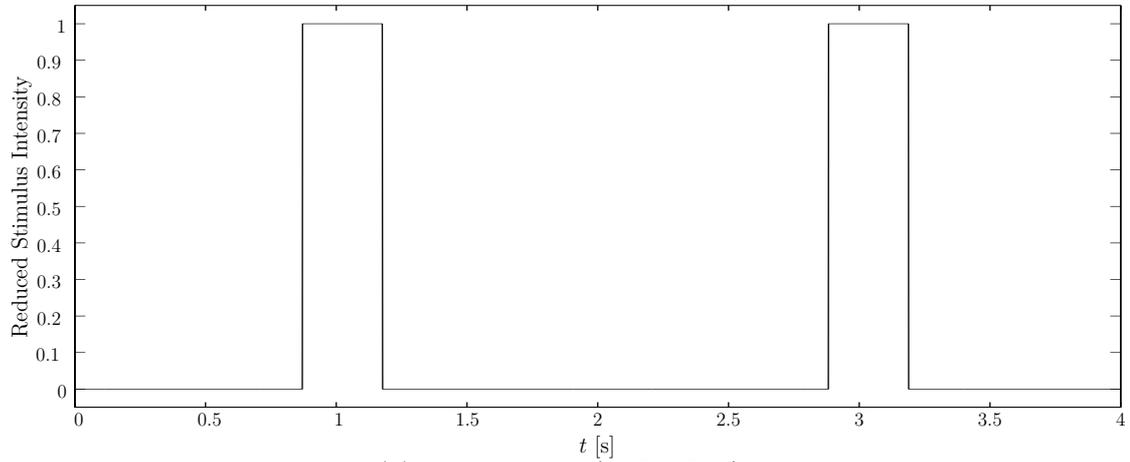
The simplest subset, called the binary subset, comprises spatially uniform stimuli with only two different intensity levels, like that illustrated in Fig. B.7a. If the time duration of the presentation of each intensity level is constant, we have a periodic waveform and the stimulus is called ON-OFF. Otherwise, if the time period for each intensity level varies in a random way, it is called a flash stimulus.

Still, in the spatially uniform ensemble, a different stimulus intensity can be chosen for each plate, which is called a random stimulus and is shown in Fig. B.7b. The light intensity levels can be tabulated previously or can be obtained by randomly sampling a probability distribution (like a Gaussian distribution, for example [Keat et al., 2001]). This kind of stimulus is generally called white noise. A software library with a wide variety of types of stimuli and functionalities available to drive and control the images produced in a computer can be found at [Vision Egg, 2007].

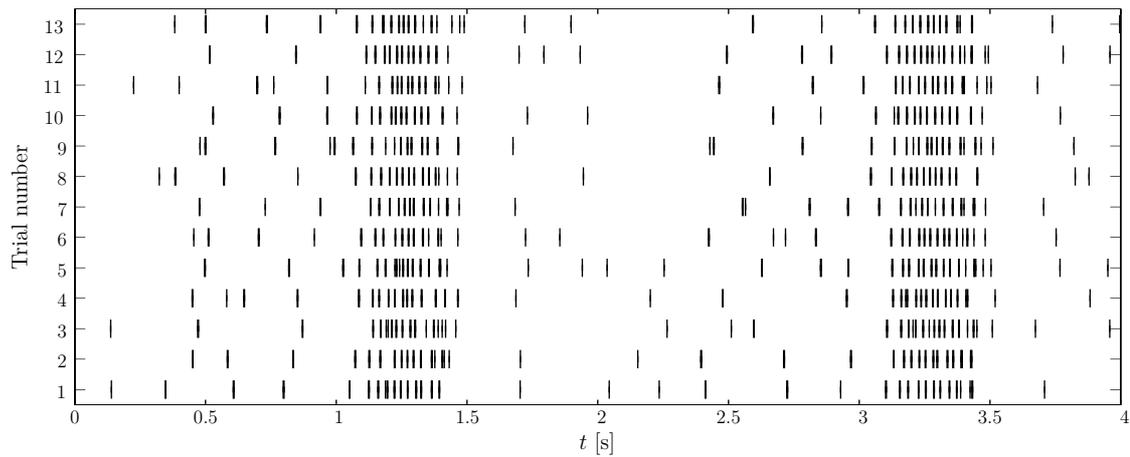
In the ensemble of spatially non-uniform stimuli, there are many variants, ranging from simple bars to films of natural scenes, depending on the retina response characteristics one wants to unveil (see [Victor and Shapley, 1979; Demb et al., 1999]). Figure B.12a represents simple horizontal and vertical bars that sweep the visual field at a predetermined velocity. The bars can be viewed as a particular case of the more general stimulus called sinusoidal grating displayed in Fig. B.8a. The sinusoidal grating has the mathematical expression:

$$s_{\text{simu}}(x, y, t) = A \cos(kx \cos \theta + ky \sin \theta) \cos(\omega t) , \quad (\text{B.1})$$

B. Characterization of the Neural Response

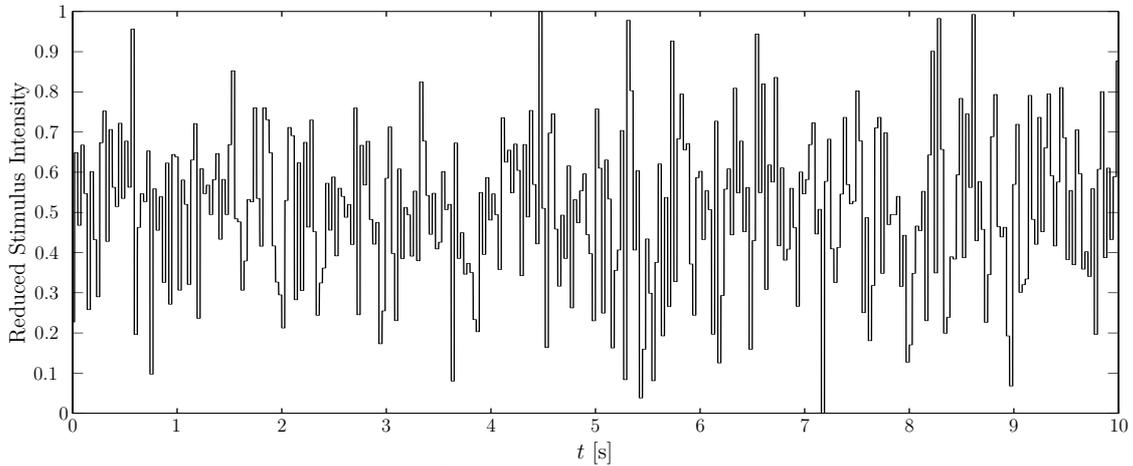


(a) Flash stimulus (or ON-OFF).

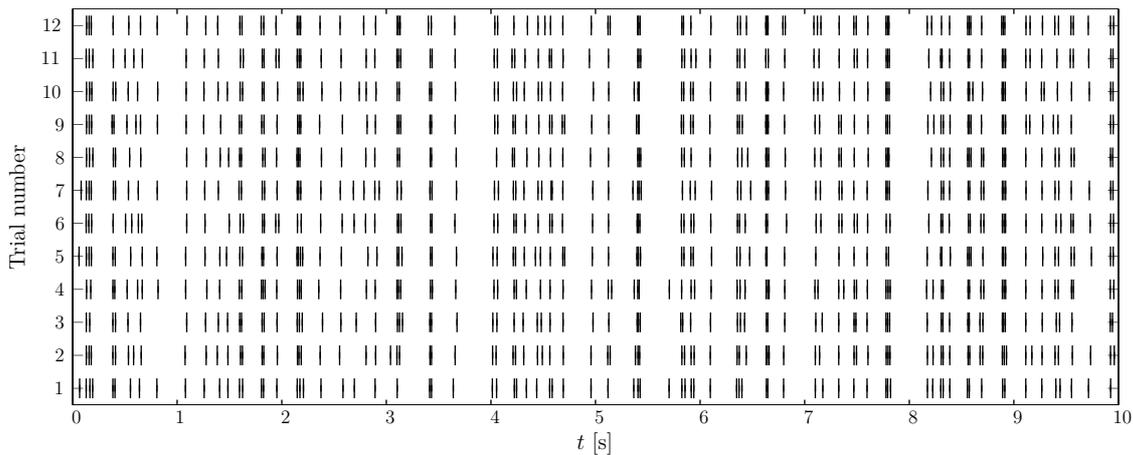


(b) Rabbit type-ON RGC spike trains.

**Figure B.5.** Rabbit ON-type RGC responses for a ON-OFF full-field stimulus ( $T_s = 1$  ms).



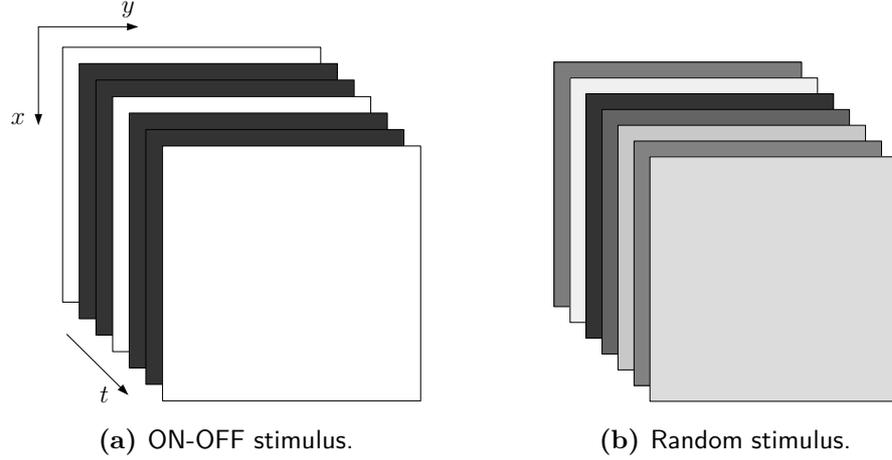
(a) White noise stimulus.



(b) Salamander type-ON RGC spike trains.

**Figure B.6.** Salamander type-ON RGC responses for sampled white-noise full-field stimulus ( $T_s = 1$  ms).

## B. Characterization of the Neural Response



**Figure B.7.** Spatially uniform visual stimuli.

where the parameter  $k$  represents the spatial frequency, in radians per meter ( $\text{rad}\cdot\text{m}^{-1}$ ), so that  $\lambda = 2\pi/k$  is the wavelength of the grating. The parameter  $\theta$  controls the direction of stimulus movement. We can see Fig. B.8a as if we had taken a photograph of the stimulus at a given time instant, but the stimulus amplitude in a given position  $(x, y)$  changes sinusoidally from the maximum  $A \cos(kx \cos \theta + ky \sin \theta)$ , to the minimum  $-A \cos(kx \cos \theta + ky \sin \theta)$  according to the temporal frequency  $\omega$ , in radians per second ( $\text{rad}\cdot\text{s}^{-1}$ ), so that  $T = 2\pi/\omega$  is the period of the sinusoidal modulation wave. Figure B.8b shows the time evolution of the stimulus amplitude at a given point  $(x_0, y_0)$  where  $s_0 = s(x_0, y_0, 0)$ . To model moving bars from the sinusoidal grating, we simply have to consider that whenever  $s_{\text{sinu}}(x, y, t)$  is positive it is equal to  $A$ , and when it is negative it is equal to  $-A$ . By introducing the signum function,  $\text{sgn}(x)$ , defined by:

$$\text{sgn}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases} . \quad (\text{B.2})$$

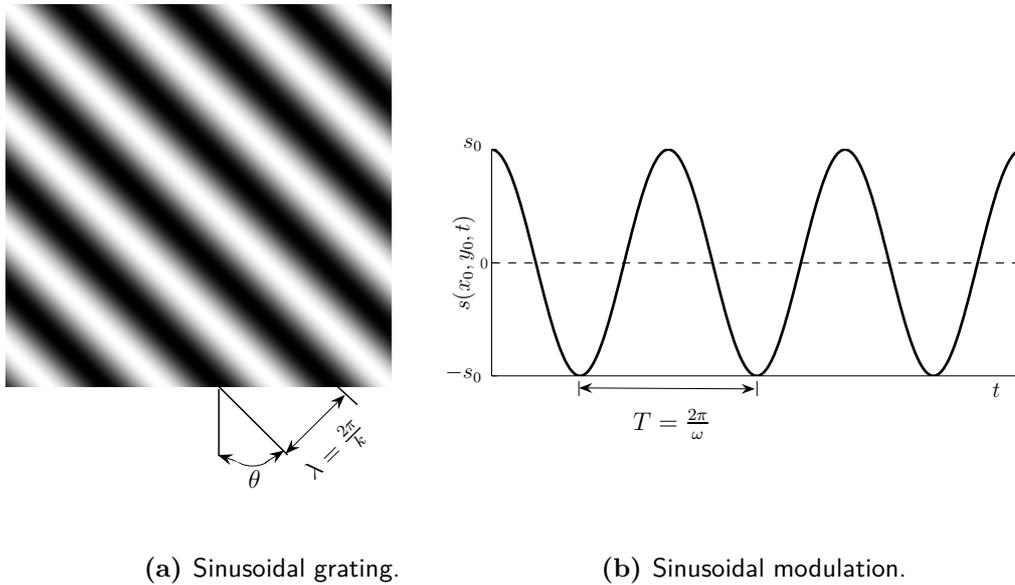
The moving bars can be expressed based on Eq. (B.1) like:

$$s_{\text{bars}}(x, y, t) = A \text{sgn}(\cos(kx \cos \theta + ky \sin \theta)) \cos(\omega t) . \quad (\text{B.3})$$

Another set of commonly used visual stimuli are based on the Gabor functions. These kinds of stimuli are predominantly used in the characterization of the **RF** of ganglion cells, and in some simple **RGCs** where the structure of their **RF** has a Gabor function shape [Dayan and Abbot, 2001]. The expression for the Gabor function is:

$$s_{\text{Gabor}}(x, y, t) = \frac{A}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)} \cos(kx \cos \theta + ky \sin \theta) \cos(\omega t) , \quad (\text{B.4})$$

and it can be seen as a two-dimensional, independent Gaussian function, with the span controlled by the parameter  $\sigma_x$  in the  $x$ -direction, and by  $\sigma_y$  in the  $y$ -direction,



**Figure B.8.** Spatially non-uniform visual stimuli.

multiplied by a sinusoidal grating. Thus, the two-dimensional Gaussian function is modulated in space by a sinusoidal function with vector number  $k$  (the spatial wavelength), and is also modulated in time with a sinusoidal function with temporal frequency  $\omega$ . Figure B.9 shows an image of a horizontal and of a rotated Gabor function. The dark and lighter bands spatially move from image to image in time.

The stochastic stimulus class comprises stimuli whose spatial and temporal information are random. Figure B.10 shows two such kinds of stimuli. Figure B.10a shows a frame of white dots chosen randomly against a dark background, while in Fig. B.10b each dot has a variable amplitude obtained by randomly sampling a Gaussian distribution. This type of stimulus is commonly referred to as a white noise stimulus, and the analysis of the respective neural responses is recognized to have several interesting features such as: exploration of a larger portion of the input space; insensitivity to the strong adaptation of the retina to a particular deterministic stimuli; and a receptive field estimation that is more robust to the fluctuations in the responsiveness of the neuron [Pillow and Simoncelli, 2003]. A whole set of analysis tools, particularly white noise analysis, are based in the study of the retinal responses to such stimuli [Westwick and Kearney, 2003; Rieke et al., 1997; Rugh, 1981]

In the current experimental apparatus, a computer monitor, with the cathode ray tube (CRT) or thin-film transistor (TFT) technology, is used to present the stimulus that is guided through an optical system to stimulate the retina. Figure B.11 displays a more complex stimulus, where each frame is divided into small squares that, in the limit case, can be the image picture elements, whose RGB intensities are driven

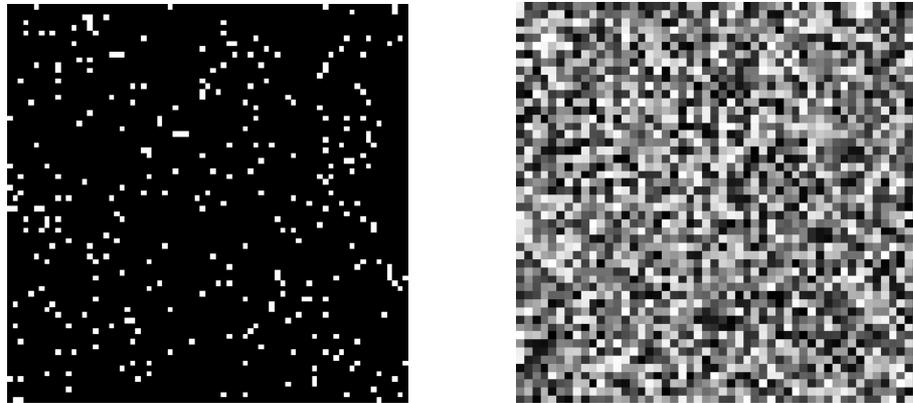
## B. Characterization of the Neural Response



(a) Horizontal Gabor function.

(b) Rotated Gabor function.

**Figure B.9.** Spatially nonuniform Gabor functions.



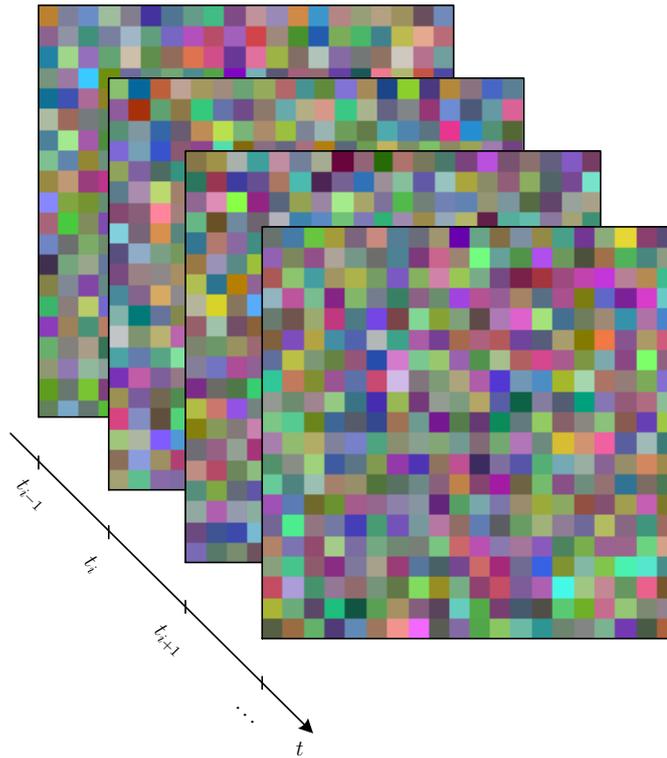
(a) Random dots.

(b) White noise.

**Figure B.10.** Stimuli with spatial and temporal modulation.

by randomly sampling a Gaussian distribution. In order to display the stimulus on a computer monitor, the stimulus must be discretized in time. The stimulus in Fig. B.11 is discretized, where  $t_{i+1} = t_i + \Delta t$  so that  $\Delta t$  is the time bin width, that establishes the sampling period of the stimuli,  $T_s$ . The images displayed by modern computer monitors are digital, so each frame is composed of a set of small picture elements, called picture element, which correspond to an inherent fine-grained spatial sampling.

The retina has a response to the light stimulus intensity spanning several orders of magnitude, from a single photon to an influx of several millions of photons per second. The retina, similarly to other sensory organs, performs stimulus intensity compression, adapting to the mean level of the stimulus, and senses only deviations from the stimulus mean following the Weber-Fechner law (see Sec. 3.2). In order to simplify the study of the responses of retinal ganglion cells, the stimuli can be described only in terms of their fluctuations around their mean level. For a continuous-time stimulus



**Figure B.11.** Gaussian white noise stimulus sequence with spatial, temporal and chromatic variation.

this means that:

$$\frac{1}{T} \int_0^T s(t) dt = 0, \quad (\text{B.5})$$

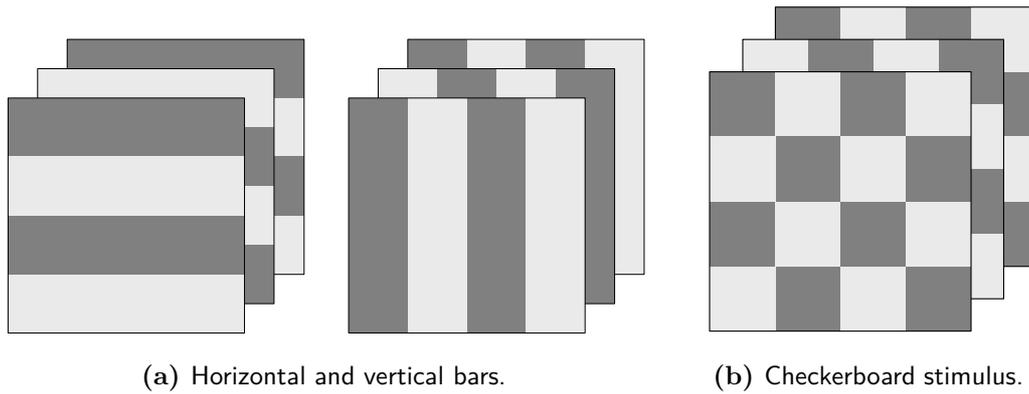
where  $T$  is the stimulus time duration. For a discrete time stimulus used in signal processing systems, the mean value of the stimulus is:

$$\frac{1}{N} \sum_{n=1}^N s[n] = 0. \quad (\text{B.6})$$

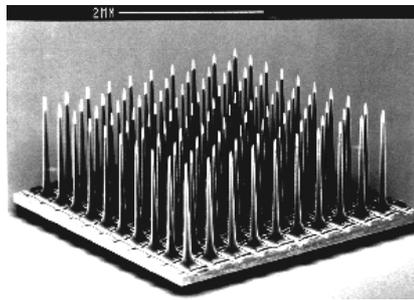
We can obtain the stimulus property expressed in Eq. (B.6) by subtracting its average value. Departing from a stimulus where Eq. (B.6) applies, a value can be added to the stimulus value in order to establish the desired mean level for stimulating the retina.

Next, we will provide an overview of the apparatus used at the University Miguel Hernandez to gather retina data from several types of vertebrates, including humans. This experimental apparatus is composed of three main parts: a stimulation system, a retina holder and positioning device, and a data acquisition and recording system. Figure B.14 displays the experimental apparatus for retina data acquisition used at the University Miguel Hernandez.

## B. Characterization of the Neural Response



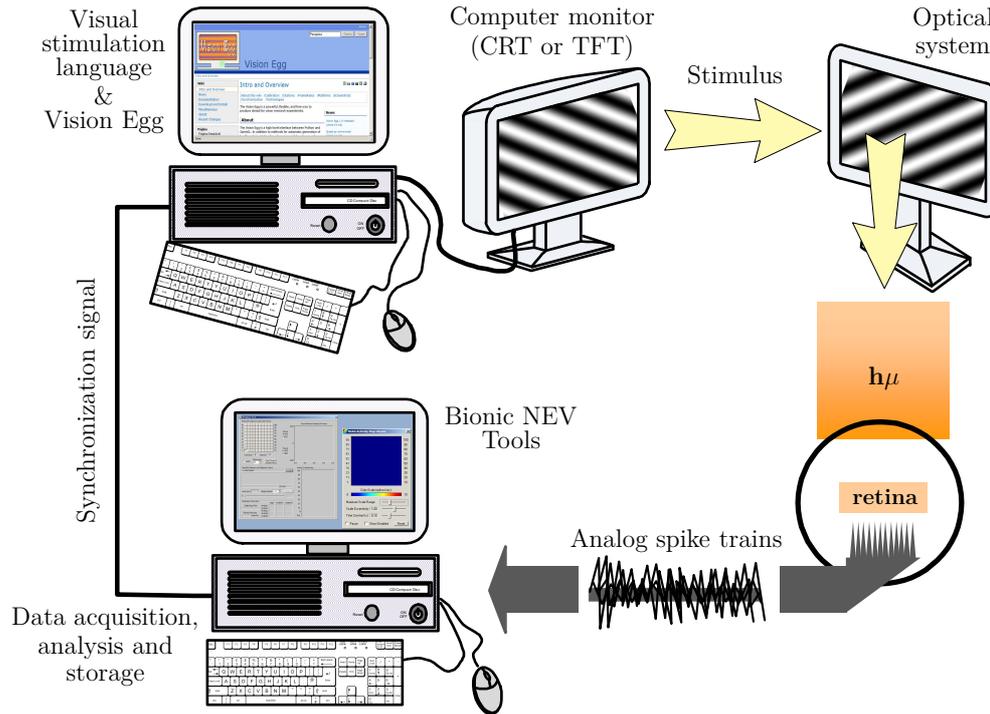
**Figure B.12.** Spatially non-uniform visual stimuli



**Figure B.13.** The Utah Microelectrode Array.

The most important component of the stimulation equipment is a computer with a graphics card capable of driving two monitors: one used for the experimental stimulus control and the other to present the stimuli. The stimulus displayed by the monitor is reflected by a mirror and deflected by an optic prism in order to stimulate the retina positioned in an appropriate holder.

The visual stimulation is generated using a 17" **CRT** high-resolution RGB monitor attached to a computer running a specifically designed program, written in the Python programming language using the VisionEgg libraries [Vision Egg, 2007]. The images are focused with the help of a lens onto the photoreceptor layer. To set up the whole system for data acquisition the retina is first flashed periodically with full field white light, whereas a microelectrode array (**MEA**), like the Utah microelectrode array (depicted in Fig. B.13), is lowered into the retina until a significant number of electrodes detect light-evoked single and multi-unit responses. This allows the recording from 60-70 microelectrodes on average out of a total of 100 microelectrodes during each experiment. The electrode array is connected to a 100-channel amplifier (low and high corner frequencies of 250 and 7500 Hz) and to a digital signal processor-based data acquisition system. The neural spike events are detected by comparing the instantaneous electrode signal to the thresholds set for each data channel. When a supra-threshold event occurs, the signal window surrounding the



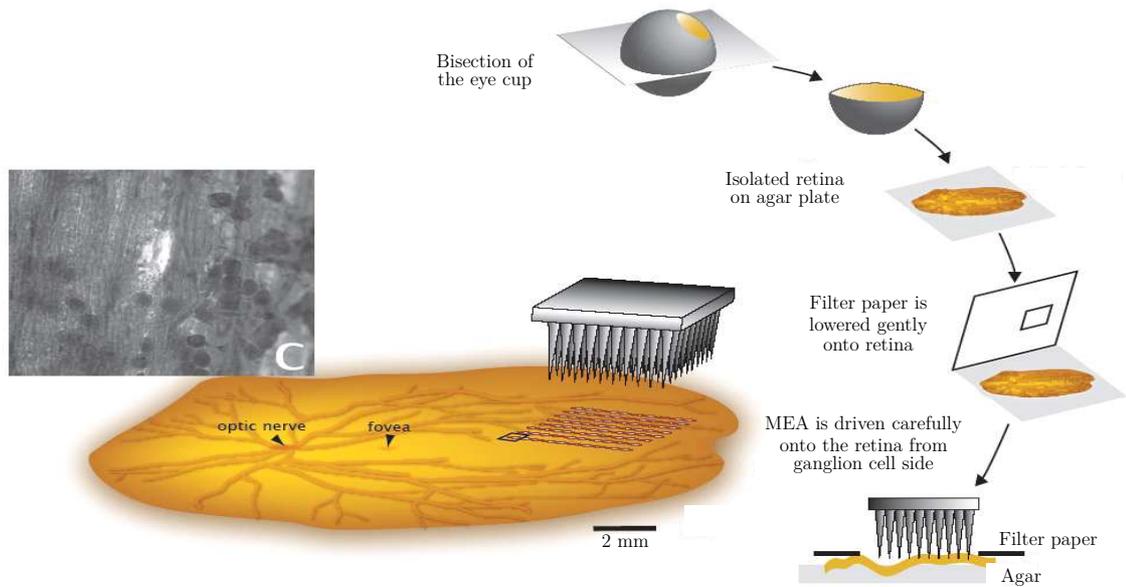
**Figure B.14.** Experimental apparatus for retina data acquisition and analysis.

event is time-stamped (see Fig. B.2), and is stored for later offline analysis. All the selected channels of data, as well as the state of the visual stimulus, are digitized with a commercial multiplexed A/D board data acquisition system, from Bionic Technologies, Inc, (now part of Cyberkinetics, Inc [Cyberkinetics, 2008]) and stored digitally.

The preparation of the retina for the data acquisition process needs to follow a delicate and sensitive process in order to obtain an effective and meaningful response from its RGCs. Figure B.15 illustrates the whole process for the preparation of a rabbit retina. First, the animal is sacrificed with an injection of an overdose of anesthetic solution, followed by the enucleation of its eye. The eyeball is hemisected with a razor blade and the cornea and lens are separated from its posterior half. The retina is then carefully removed from the remaining eyecup with the pigment epithelium, and is mounted on a glass slide, with the ganglion cell layer side up, which is then covered with a millipore filter. This preparation is then mounted on a recording chamber and superfused with bicarbonate-buffered Ames solution at a temperature of 35° C, to postpone the retina death.

The recorded neuron cell responses are classified according to the spike waveform shapes, by means of the classification software SAC [Shoham, 2001], which uses a decomposition algorithm based on the Expectation-Maximization (EM) algorithm, in which the distribution of waveforms from each unit is modeled as a multivariate  $t$ -student distribution. As seen in Fig. B.17, the neuron responses to the same stimulus

## B. Characterization of the Neural Response



**Figure B.15.** Retina preparation for data acquisition.

are not unique; instead they have a certain degree of randomness.

### B.1.2. Spike Trains and Firing Rates

The spike train of Fig. B.3 is described mathematically by a series of Dirac delta functions,  $\delta(t)$ , or unit impulse functions, using the signal processing terminology [Oppenheim et al., 1999b], each one positioned at a spike time occurrence  $t_i$ , where  $i = 1, \dots, n$ , where  $n$  is the total number of spikes in the train:

$$\rho(t) = \sum_{i=1}^n \delta(t - t_i) , \quad (\text{B.7})$$

where  $\rho(t)$  is the termed *neural response function*. The neural response function neglects the height and shape of the action potentials, so that all information is contained in the time arrival of the spikes; thus, the spike train is considered to be a point process [Brenner et al., 2002].

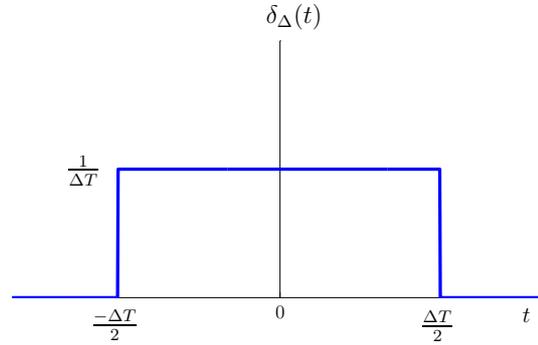
The Dirac delta function  $\delta(t)$  has a set of important properties [Arfken and Weber, 2005]:

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1 , \quad (\text{B.8})$$

and

$$\delta(t) = 0 \quad \text{for } t \neq 0 . \quad (\text{B.9})$$

Another useful property of the delta function, which is a direct consequence of the



**Figure B.16.** The  $\delta_{\Delta}$  function.

previous ones and is known as the sifting property, is:

$$\int_{-\infty}^{+\infty} \delta(t - t_0) f(t) dt = \int_{-\infty}^{+\infty} \delta(t - t_0) f(t_0) dt = f(t_0). \quad (\text{B.10})$$

The Dirac delta function is not strictly a function since it does not have a closed mathematical definition. To define the continuous Dirac delta function, we can use the  $\delta_{\Delta}(t)$  function depicted in Fig. B.16 and defined as:

$$\delta_{\Delta}(t) = \begin{cases} \frac{1}{\Delta T} & -\frac{\Delta T}{2} < t < \frac{\Delta T}{2} \\ 0 & \text{otherwise} \end{cases}. \quad (\text{B.11})$$

By taking the limit when  $\Delta T \rightarrow 0$  we have:

$$\delta(t) = \lim_{\Delta T \rightarrow 0} \delta_{\Delta}(t). \quad (\text{B.12})$$

Given the definition in Eq. (B.12), we can see that the delta function has an area equal to one, and, by making  $\Delta T \rightarrow 0$ , the nonzero function values become concentrated around the origin, having an infinitesimal duration and an infinite amplitude. Although it has a peculiar definition, the enumerated properties make the delta function very useful.

The discrete-time counterpart of the delta function is the unit impulse. Unlike its continuous form, the discrete unit impulse has a closed mathematical formula (see [Oppenheim et al., 1999a]):

$$\delta[n] = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0 \end{cases}, \quad (\text{B.13})$$

## B. Characterization of the Neural Response

where  $n$  in Eq. (B.13) represents the discrete independent variable, usually time<sup>1</sup>.

From Eq. (B.7), we can obtain the number of spikes,  $n$ , present in the recording of a neuron response with a total time duration  $T$  and where each spike occurs at instants  $t_i$ , so that  $0 \leq t_i \leq T$  for all  $i$  by integrating the neuron response:

$$\begin{aligned} n &= \int_0^T \rho(\tau) d\tau \\ &= \sum_{i=1}^n \int_0^T \delta(\tau - t_i) d\tau, \end{aligned} \quad (\text{B.14})$$

where the integral in the last term evaluates to one due to the property in Eq. (B.8).

The neural response can be characterized by several quantities computed from the neural response function. One such quantity is the *spike-count rate*. The spike-count rate,  $r$ , is the number of spikes,  $n$ , appearing during a trial, divided by the total time duration,  $T$ , of the trial, and has the expression:

$$r = \frac{n}{T} = \frac{1}{T} \int_0^T \rho(\tau) d\tau. \quad (\text{B.15})$$

The spike-count rate is the time average of the neural response over a particular trial. The spike-count rate does not give any temporal information about the neural response.

We can average the neural response function over many experimental trials obtained with the same stimulus, such as the trials displayed in Fig. B.17. This average is denoted by  $\langle \rho(t) \rangle$ , and is computed by summing all the individual neural response functions for each trial and dividing the result by the number of trials  $M$ . If the neural response function for the trial  $j$ , where  $j = 1, \dots, M$ , is represented by  $\rho_j(t)$ , the expression for the *average neural response* is:

$$\langle \rho(t) \rangle = \frac{1}{M} \sum_{j=1}^M \rho_j(t). \quad (\text{B.16})$$

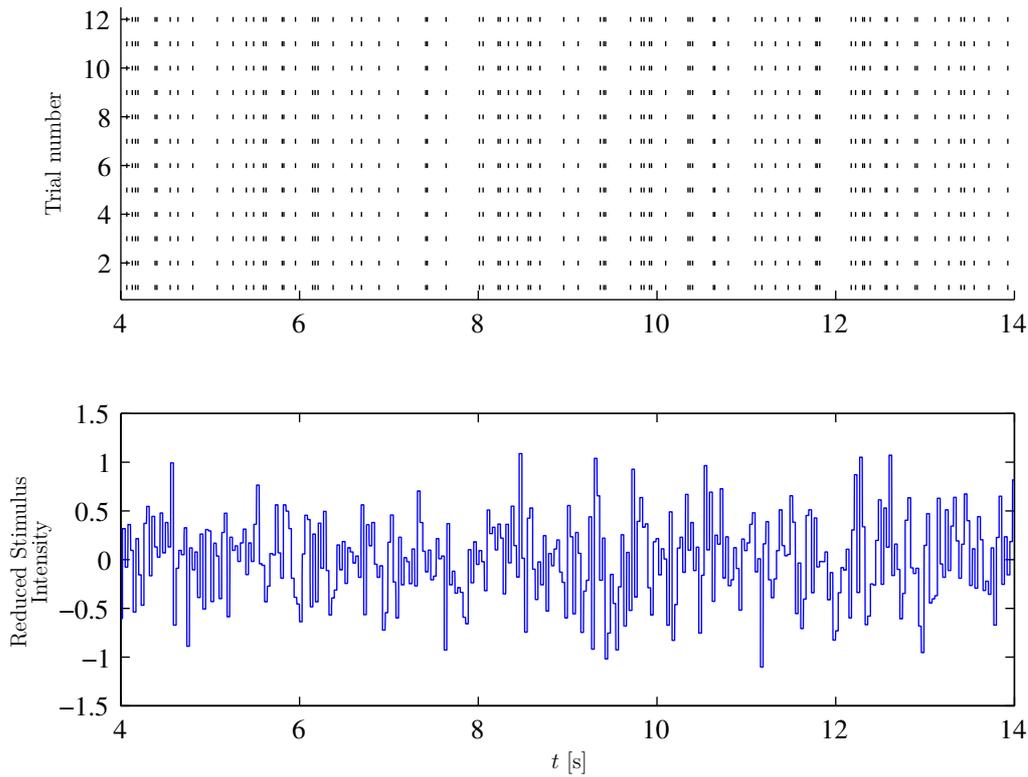
If we represent the occurrence of a spike at time  $t_i$ , with  $0 \leq t_i \leq T$ , in the trial  $j$ , for  $j = 1, \dots, M$ , as  $t_{ij}$  then, with this nomenclature, the neural response average can be written as:

$$\langle \rho(t) \rangle = \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{n_j} \delta(t - t_{ij}), \quad (\text{B.17})$$

where  $n_j$  represents the total number of spikes in the trial  $j$ , which is usually different from trial to trial.

---

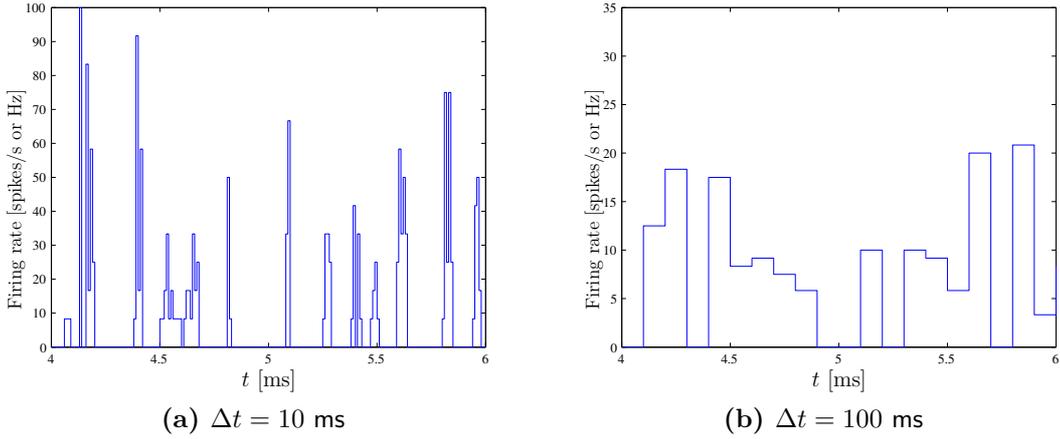
<sup>1</sup>Due to the fact that  $n$  is frequently used for the independent variable in discrete-time, this convention is applied. However,  $n$  is also employed to denote the total number of spikes in a train. The two meanings should be distinguished according to the context.



**Figure B.17.** Neural spike trains from a Salamander ON-type retinal ganglion cell (*Top*) when driven by the uniform white noise stimulus obtained from sampling a Gaussian distribution (*Bottom*), (data from [Keat et al., 2001]).

A *time-dependent firing rate* can be obtained by counting spikes over short intervals of time, which is done more precisely with a large number of trials, which results from repeatedly presenting the same stimulus to the neuron. The firing rate at time  $t$  can be computed by counting the number of spikes that occur between the time instants  $t$  and  $t + \Delta t$  and dividing the result by  $\Delta t$ ; thus, the firing rate is measured in spikes/s or Hz. The precision of the firing rate computation increases by using a narrower time bin width  $\Delta t$ ; as a result, a higher temporal resolution can be achieved. If we use only a spike train from a single trial and a narrow time bin, the firing rate will be either zero or one, so an average over multiple trials should be made. The time-dependent firing rate is defined as the average number of spikes over several trials appearing in an interval between time instants  $t$  and  $t + \Delta t$ , divided by the interval length  $\Delta t$ . The time-dependent firing rate is defined as:

## B. Characterization of the Neural Response



**Figure B.18.** Retinal ganglion cell ON-type firing rate using different bin widths for the neuron response division.

$$\begin{aligned}
 r(t) &= \frac{1}{M} \sum_{j=1}^M \frac{1}{\Delta t} \int_t^{t+\Delta t} \rho_j(\tau) d\tau \\
 &= \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle \rho(\tau) \rangle d\tau,
 \end{aligned} \tag{B.18}$$

where the integral in the first equality computes the number of spikes between the times  $t$  and  $t + \Delta t$  in the  $j$ th trial, while the integral in the second equality captures the number of spikes in the time range  $t$  to  $t + \Delta t$  of the neural response average.

Formally, we can make the time interval length of Eq. (B.18) go to zero. Thus, by taking limit we have

$$\begin{aligned}
 r(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \langle \rho(\tau) \rangle d\tau \\
 &= \left\langle \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_t^{t+\Delta t} \rho(\tau) d\tau \right\rangle,
 \end{aligned} \tag{B.19}$$

where we can recognize the derivative of the integral of  $\rho(t)$  inside the angle brackets in the last expression. Therefore, the formal definition of the firing rate can be written as:

$$r(t) = \langle \rho(t) \rangle, \tag{B.20}$$

that states that the firing rate is the average of the neural function over many trials.

To calculate the firing rate from the observed data, a time interval with a finite length must be used in order to obtain a reliable estimate of the average. Figure B.18 displays a parcel of the firing rate for the data in Fig. B.17. For a small  $\Delta t$ , the

quantity  $r(t)\Delta t$  is the average number of spikes present at the interval  $[t, t + \Delta t]$ :

$$\begin{aligned}\langle n \rangle_{[t, t+\Delta t]} &= \int_t^{t+\Delta t} r(t) dt \\ &\cong r(t)\Delta t.\end{aligned}\tag{B.21}$$

By integrating  $r(t)$  over a given time interval, we obtain the average number of spikes over that interval.

If  $\Delta t$  is small in Eq. (B.18), there will be no more than one spike in the time interval  $t \in [t, t + \Delta t]$  for each neural response function, so that  $r(t)\Delta t$  gives the fraction of trials in which a spike occurred between those times. In this way,  $r(t)\Delta t$  can be interpreted as the probability of a spike occurring in the interval  $\Delta t$  around  $t$ . That is,

$$P(n = 1 \text{ in } \Delta t) = r(t)\Delta t.\tag{B.22}$$

The spike-count firing rate can be averaged over several trials, yielding a another quantity called the *average firing rate* and denoted by  $\langle r \rangle$  with the expression

$$\langle r \rangle = \frac{\langle n \rangle}{T} = \frac{1}{T} \int_0^T \langle \rho(\tau) \rangle d\tau = \frac{1}{T} \int_0^T r(\tau) d\tau.\tag{B.23}$$

The first equality in Eq. (B.23) indicates that  $\langle r \rangle$  is the average number of spikes per trial divided by the trial duration, where the second and third equalities result from Eq. (B.14) and Eq. (B.20), respectively.

The three different quantities:  $r$ ,  $r(t)$ , and  $\langle r \rangle$ , are frequently used in the literature without distinction as *firing rate*, which causes some misunderstanding; therefore, special care should be taken to understand from the context which definition is being employed.

The formal firing rate  $r(t)$  defined by Eq. (B.19) should, theoretically, be calculated with an infinite set of trials. This is impossible, but the estimate becomes more accurate as more and more trials are included in the set. Indeed, we have only a finite set of experimental trials that must be discretized into time bins, and the firing rate can be obtained by counting spikes within the time bins that compose a trial and by averaging over trials.

In practice, the firing rate is estimated by dividing the spike trains into time bins with a finite width  $\Delta t$ , then the number of spikes within each time bin is counted, and the result divided by  $\Delta t$ . Furthermore, if we have several trials as the response to the same stimulus, we can perform this operation for all trials and average the results over these trials. This procedure generates a staircase-like spike-count firing rate, which is a piecewise constant time function for each time bin, resembling a histogram (see Fig. B.18). Each bar amplitude is equal to  $1/\Delta t$  times the number of spikes in each time bin. By decreasing the width of the time bins,  $\Delta t$ , the temporal resolution increases, and the firing rate can be estimated at a finer time scale, but at the cost of reducing the number of possible different rates. In the limiting case,

## B. Characterization of the Neural Response

for a time bin narrower than the neuron's refractory period, there can be only 0 or 1 spike in each bin, so that the height of each histogram bar can be only 0 or  $1/\Delta t$ .

One way to avoid the quantization of the firing rate, which is always proportional to the inverse of the time bin width, is to divide the trial into bins with a variable width by fixing the number of spikes within each time bin, so that the firing rate is approximated by the fixed number of spikes in the bin divided by the variable bin width.

Both procedures described make the firing rate dependent on the size of the bin and/or the spikes' locations. To prevent these effects a time window,  $w(t)$ , can be used to smooth the firing rate. This window slides along the spike trains and counts the number of spikes within the window at each new location. This is the same as convolving the spike trains with the window. Generalizing, if we have  $M$  trials with spikes positioned at times  $t_{ij}$ , with  $i = 1, \dots, n_j$ , where  $n_j$  is the number of spikes in the trial  $j$ , and  $j = 1, \dots, M$ , the firing rate can be approximated by

$$\begin{aligned} r_{\text{approx}}(t) &= \int_{-\infty}^{+\infty} w(\tau) \langle \rho(t - \tau) \rangle d\tau \\ &= \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{n_j} w(t - t_{ij}) , \end{aligned} \quad (\text{B.24})$$

for a window with unit area (otherwise the result must be divided by the window area). This procedure corresponds to the convolution of the average neural response with the filter kernel  $w(t)$ . The last equality in Eq. (B.24) results from the Dirac delta property in Eq. (B.10).

Convolution is one of the most important operations in signal processing and system analysis since many systems can be completely specified by their convolving properties. The convolution operator is denoted by an asterisk, and it represents the integral in continuous time:

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau) h(t - \tau) d\tau . \quad (\text{B.25})$$

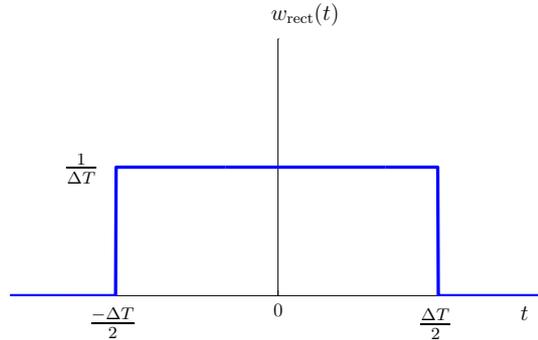
The correspondent operation in discrete-time is defined as [Oppenheim et al., 1999b]:

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{+\infty} x[k] h[n - k] . \quad (\text{B.26})$$

Using this notation, the approximation for the firing rate in Eq. (B.24) can be written as:

$$r_{\text{approx}}(t) = w(t) * \langle \rho(t) \rangle . \quad (\text{B.27})$$

One of the simpler filter kernels is the rectangular window, or boxcar filter. A rectangular window with duration  $\Delta T$  and unit area has the expression in continuous



**Figure B.19.** The rectangular (boxcar) filter window.

time:

$$w_{\text{rect}}(t) = \begin{cases} \frac{1}{\Delta T} & -\frac{\Delta T}{2} < t < \frac{\Delta T}{2} \\ 0 & \text{otherwise} \end{cases}. \quad (\text{B.28})$$

Figure B.19 graphically represents a rectangular window. As a consequence of using a window to convolve the spike trains, the values obtained for the firing rates separated in time by less than one window width are correlated since they include common spikes in their calculation.

The main problem with the rectangular window is that it is not a continuous function. As a result, filtering the neural function average, which is also a series of discontinuous delta functions, results in an irregular waveform for the firing rate. To avoid these effects, a continuous window function that goes to zero as the distance to the origin increases can be used. A widespread filter window is the ubiquitous bell-shaped Gaussian function. The Gaussian function is defined as

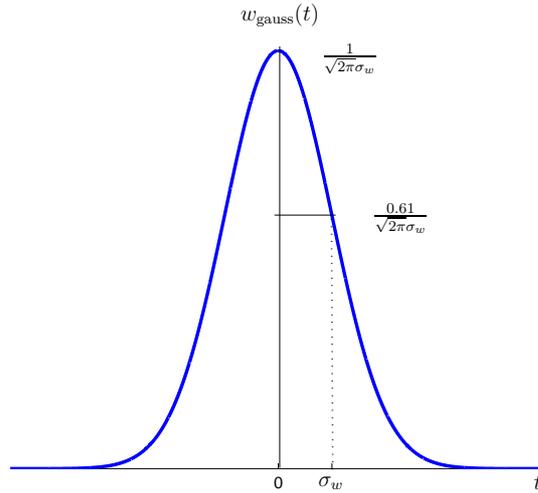
$$w_{\text{gauss}}(t) = \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{t^2}{2\sigma_w^2}}, \quad (\text{B.29})$$

where the parameter  $\sigma_w$  controls the width of the function and, consequently, the temporal resolution of the rate. Figure B.20 displays the graph of the Gaussian window.

Both the rectangular and Gaussian windows are not causal, meaning that they approximate the firing rate at a given time instant by taking into account spikes that were fired before and after that instant. To make the firing rate at time  $t$  dependent only on spikes fired before  $t$ , we must use a causal filter for the window. A commonly used causal filter is the  $\alpha$  window (see [Dayan and Abbot, 2001]), which is expressed as

$$w_{\alpha}(t) = \alpha^2 t \exp(-\alpha t) H(t), \quad (\text{B.30})$$

## B. Characterization of the Neural Response



**Figure B.20.** The Gaussian filter window.

where  $1/\alpha$  controls the temporal resolution of the firing rate estimate. The expression of Eq. (B.30) introduces the continuous Heaviside function,  $H(t)$ , or the *unit step function*, in signal processing terminology [Ziemer et al., 1998], defined as:

$$H(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}. \quad (\text{B.31})$$

The Dirac delta function, defined in Eq. (B.12) as the limit of a rectangular window, can also be defined as the derivative of the Heaviside unit step function:

$$\delta(t) = \frac{dH(t)}{dt}. \quad (\text{B.32})$$

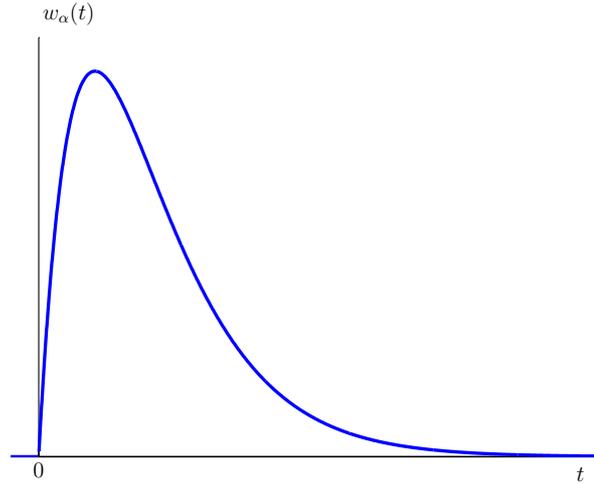
The delta function properties, introduced in Eq. (B.8) and Eq. (B.9), still hold. The delta function is called a generalized function that can be defined as the limit of several alternative functions as long as its properties are satisfied [Ziemer et al., 1998; Arfken and Weber, 2005]. The discrete-time counterpart of the continuous Heaviside unit step function, Eq. (B.31), is the *unit step sequence* with the expression:

$$H[n] = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases}, \quad (\text{B.33})$$

where  $n$  is the discrete independent variable. The unit step sequence is related with the unit impulse, defined in Eq. (B.13), by

$$\delta[n] = H[n] - H[n - 1] \quad \text{or} \quad H[n] = \sum_{k=-\infty}^n \delta[k]. \quad (\text{B.34})$$

In order to process the neuronal data in a computer, the data must be discretized. In discrete-time, the spikes in the continuous neural function of Eq. (B.7) cannot be



**Figure B.21.** The  $\alpha$  function filter.

positioned at any arbitrary instant of time. In the process of discretization, if we use a sampling period  $\Delta t$ , the spikes are placed within a specific time bin,  $n_i$ , with width  $\Delta t$ . In discrete-time, the neuronal function of Eq. (B.7) is represented by a series of discrete delta functions, or unit impulses, defined in Eq. (B.13), as:

$$\rho[n] = \sum_{i=1}^n \delta[n - n_i], \quad (\text{B.35})$$

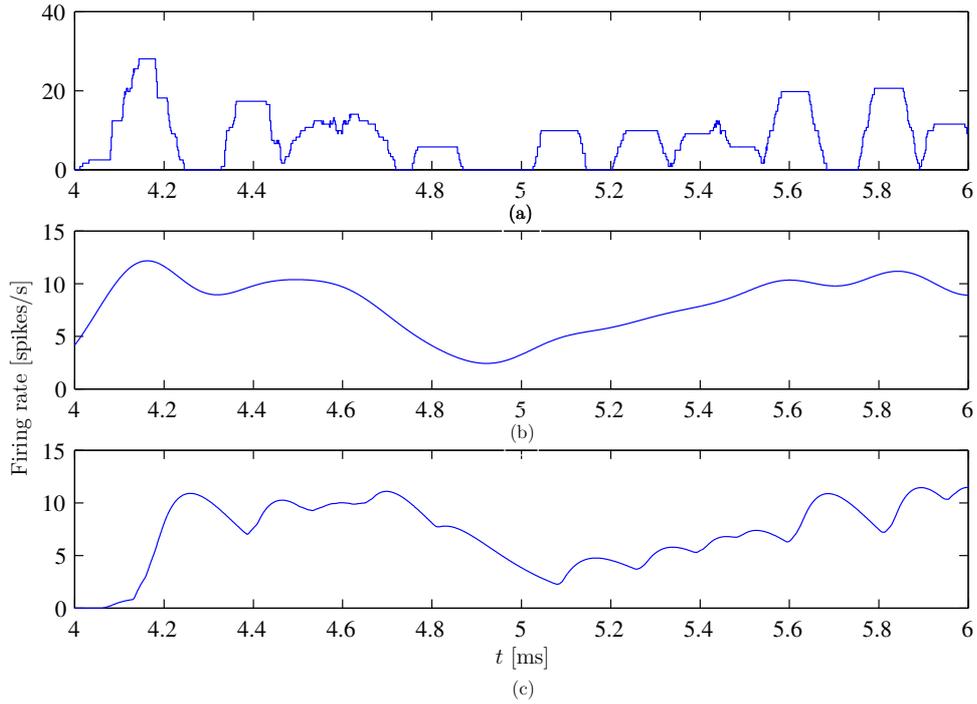
where  $n_i$  are the discrete time instants corresponding to the time bins containing a spike. The discrete neural function can be obtained by dividing the total time duration of the response,  $T$ , into intervals of width  $\Delta t$ , so that the sequence of the neural function has a length equal to  $N = T/\Delta t$ . It is convenient to make  $\Delta t$  small, which means that the sampling frequency has to be sufficiently high in order to have at most one spike in each time bin, which can be accomplished by choosing a time sampling period smaller than the refractory period of the neuron. It was shown that a sampling frequency within the range 10 kHz to 20 kHz is enough for the case of primate retinal ganglion cells [Uzzell and Chichilnisky, 2004].

The firing rate, defined in Eq. (B.20) for the continuous case, becomes in discrete-time

$$r[n] = \frac{1}{\Delta t} \frac{1}{M} \sum_{j=1}^M \sum_{i=1}^{n_j} \delta[n - n_{ij}], \quad (\text{B.36})$$

where the limit  $n_j$  in the sum represents the number of spikes in the trial  $j$  and the limit  $M$  is the total number of trials in the experiment. The resulting sequence  $r[n]$  gives the number of fired spikes per second (Hz) in the time bin  $n$ , with  $0 \leq n \leq N-1$ . The discrete firing rate  $r[n]$  is frequently plotted against time, giving the PSTH, as depicted in Fig. B.22.

## B. Characterization of the Neural Response



**Figure B.22.** Firing rate obtained by filtering the neural response with different types of filter windows: (a) rectangular window with  $\Delta t = 100$  ms; (b) Gaussian window with  $\Delta t = 100$  ms; (c)  $\alpha$ -window with  $1/\alpha = 100$  ms. The sampling rate is  $F_s = 1$  kHz.

The average firing rate is obtained from the discrete firing rate by the expression:

$$\langle r \rangle = \frac{1}{N} \sum_{n=1}^N r[n], \quad (\text{B.37})$$

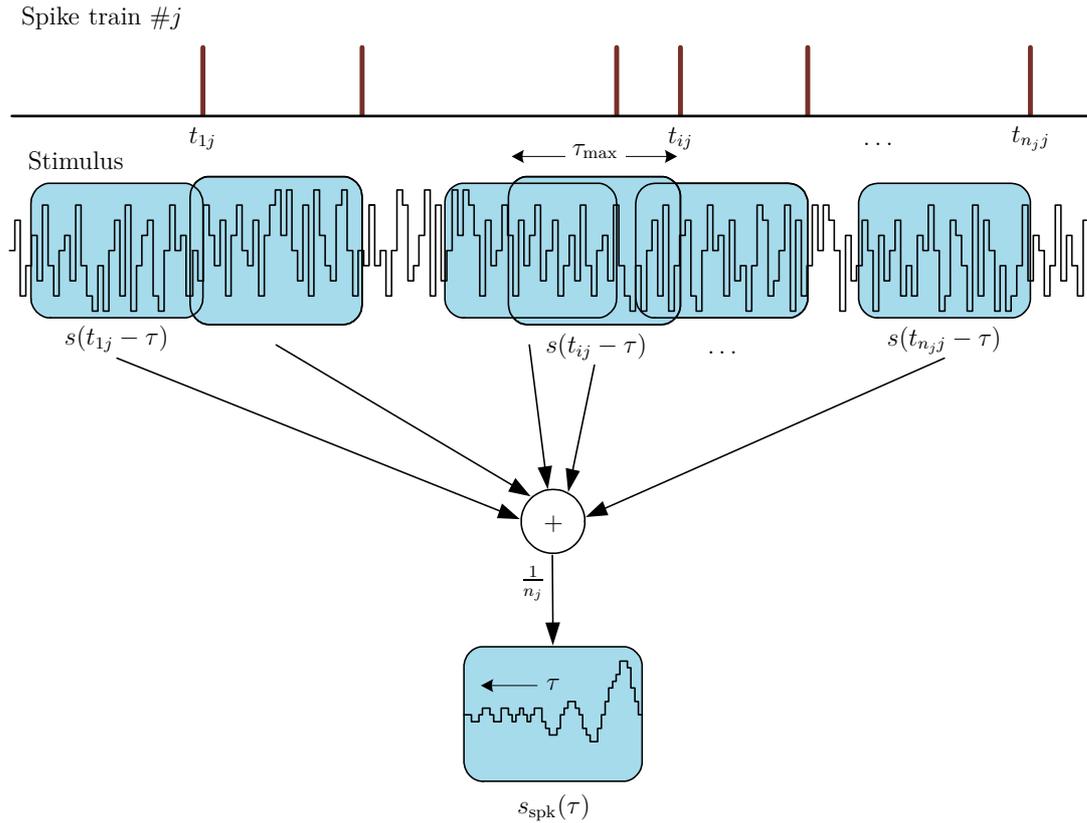
where  $N$  is the number of time bins into which the spike train was sampled. The average number of spikes  $\langle n \rangle$  within the time bin  $n$ , corresponding to the time interval  $[n\Delta t, (n+1)\Delta t]$ , is given by

$$\langle n \rangle_{[n\Delta t, (n+1)\Delta t]} = r[n]\Delta t. \quad (\text{B.38})$$

For a small  $\Delta t$ , so that there is at most one single spike per time bin, the firing rate corresponds to the probability that the neuron will fire a spike within that time bin.

### B.1.3. Spike Triggered Average

A useful and enlightening characteristic of a neuron is the waveform of the stimuli that produces a given response. We can choose different patterns of spikes and compute the stimuli that originate it. There are several possibilities, ranging from the stimulus that elicits a single spike to the stimulus that produces a more complicated sequence



**Figure B.23.** Procedure for the STA computation.

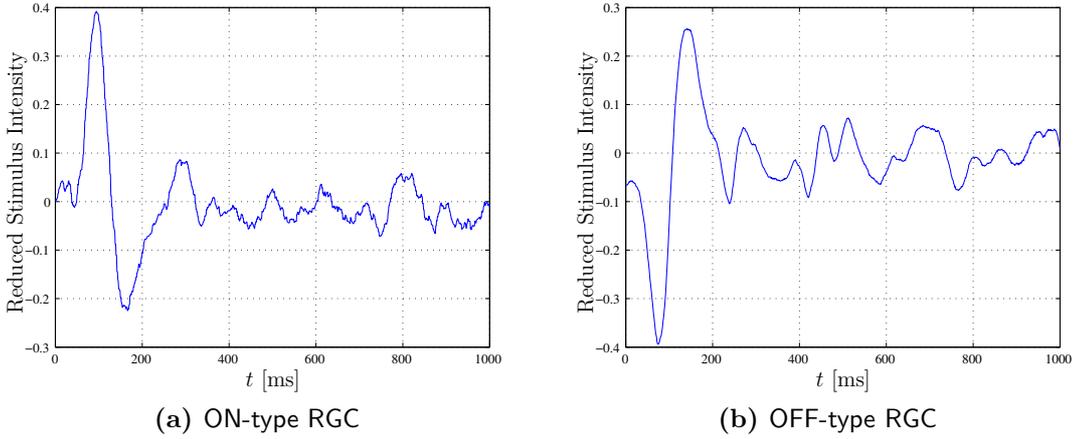
of spikes [de Ruyter van Steveninck and Bialek, 1988]. However, the most simple and common situation is to analyze the stimuli that elicited a single spike.

With the single spike response, we try to figure out what the stimulus looks like, on average, before this single action potential was fired. The resulting quantity is termed the spike triggered average (**STA**). The **STA** provides a very useful technique to characterize the neural selectivity, and it constitutes the basic framework for several retina models [Chichilnisky, 2001; de Ruyter van Steveninck and Bialek, 1988]. The **STA** is also referred to in the literature as the reverse correlation function [Simoncelli et al. [2004], the mean effective stimulus, the triggered correlation function, or even the first Wiener kernel [Rieke et al., 1997].

The procedure to compute the **STA** is: *i*) pick the stimulus segments before every fired spike in a trial, *ii*) add all these segments, *iii*) normalize the resulting waveform by the total number of spikes, and then *iv*) repeat the previous procedure for all trials available and normalize the result by the number of trials. Figure B.23 illustrates the computing process of the **STA** for a single spike train # $j$  with  $n_j$  spikes.

If we represent the stimulus segment with length  $\tau_{\max}$ , which occurs just before a spike located at the time instant  $t_i$  in trial  $j$  by  $s(t_{ij} - \tau)$ , with  $0 < \tau \leq \tau_{\max}$  for  $i = 1, \dots, n_j$ , where  $n_j$  is the number of spikes in trial  $j$ , and  $j = 1, \dots, M$ , where

## B. Characterization of the Neural Response



**Figure B.24.** Time reversed spike triggered average (STA) of a ON-type retinal ganglion cell (salamander), and of an OFF-type cell (rabbit), (data from [Keat et al., 2001]).

$M$  is the number of trials evolved in the computation, the **STA** takes the value

$$s_{\text{spk}}(\tau) = \frac{1}{M} \sum_{j=1}^M \frac{1}{n_j} \sum_{i=1}^{n_j} s(t_{ij} - \tau) \quad 0 \leq \tau \leq \tau_{\text{max}} . \quad (\text{B.39})$$

This expression is equivalent to a weighted sum over several trials.

Although it was not imposed a limit for the value of  $\tau_{\text{max}}$  in Eq. (B.39), the neural response depends on the stimulus only within a time window with a few hundred milliseconds wide before the spike occurrence, which corresponds to the neuron memory. This happens because  $s_{\text{spk}}(\tau)$  goes to zero for positive values of  $\tau$  larger than the correlation time between the stimulus and the response. If, however, the stimulus is not temporally autocorrelated, we can state that  $s_{\text{spk}}(\tau)$  will be 0 for  $\tau < 0$  because the neuron response should not be dependent on future stimuli – it is a causal system.

To compute the **STA**, the stimuli segment is acquired over a finite time period before every spike occurrence; these different stimuli are summed, and the result is normalized by the number of spikes considered. Figure B.24 shows the plot of the **STA**, time reversed, for two distinct types of **RGC** cells. These cells are categorized according to the waveform of their **STA**. As we can see from Fig. B.24a, the **RGC** of the salamander fires preferably in response to the positive onset of stimulus, so it is classified as an ON-type cell. The rabbit **RGC** in Fig. B.24b fires predominantly when the stimulus has a negative offset, so it is classified as an OFF-type cell.

Taking into account the property expressed in Eq. (B.10) we can write

$$s(t_i - \tau) = \int_0^T \delta(t - t_i) s(t - \tau) dt , \quad (\text{B.40})$$

so that Eq. (B.39) becomes:

$$\begin{aligned}
 s_{\text{spk}}(\tau) &= \frac{1}{M} \sum_{j=1}^M \frac{1}{n_j} \sum_{i=1}^{n_j} \int_0^T \delta(t - t_{ij}) s(t - \tau) dt \\
 &= \frac{1}{M} \sum_{j=1}^M \int_0^T \frac{1}{n_j} \sum_{i=1}^{n_j} \delta(t - t_{ij}) s(t - \tau) dt
 \end{aligned} \tag{B.41}$$

By assuming that each spike train has a similar number of spikes, we can approximate the number of spikes in each trial by  $n_j \simeq \langle n \rangle$ ; for example, this is a good approximation when the number of spikes in each trial is high. With this assumption, Eq. (B.41) can be written as:

$$\begin{aligned}
 s_{\text{spk}}(\tau) &= \frac{1}{\langle n \rangle} \left\langle \int_0^T \sum_{i=1}^n \delta(t - t_i) s(t - \tau) dt \right\rangle \\
 &= \frac{1}{\langle n \rangle} \left\langle \int_0^T \rho(t) s(t - \tau) dt \right\rangle
 \end{aligned} \tag{B.42}$$

The last equation above relates the **STA** with the neural response function defined by Eq. (B.7). Furthermore, if the same stimulus is used for all trials, Eq. (B.42) can be further simplified to:

$$s_{\text{spk}}(\tau) = \frac{1}{\langle n \rangle} \int_0^T \langle \rho(t) \rangle s(t - \tau) dt = \frac{1}{\langle n \rangle} \int_0^T r(t) s(t - \tau) dt . \tag{B.43}$$

By definition, the correlation function of the firing rate with the stimulus is:

$$C_{rs}(\tau) = \frac{1}{T} \int_0^T r(t) s(t + \tau) dt . \tag{B.44}$$

Equation (B.43) relates the **STA** with the correlation function of the stimulus with the firing rate. Comparing the previous equation with Eq. (B.43), we have the relation

$$s_{\text{spk}}(\tau) = \frac{1}{\langle r \rangle} C_{rs}(-\tau) , \tag{B.45}$$

where  $\langle r \rangle = \langle n \rangle / T$  is the average firing rate over the set of trials. From Eq. (B.45), we can see that  $s_{\text{spk}}(\tau)$  is proportional to the correlation function of the firing rate with the stimulus at preceding times, so that the **STA** is also called the *reverse correlation function*.

Whenever the input stimulus has a shape similar to the **STA**, the neuron has a high probability of firing a spike. This is the genesis of several neuron models, and

## B. Characterization of the Neural Response

of retina models in particular. Considering that the firing rate of a visual neuron is a functional of a filtered version of the input stimulus  $s(t)$ ,

$$r(t) = r_0 F \left[ \int_{-\infty}^{\infty} h(\tau) s(t - \tau) d\tau \right], \quad (\text{B.46})$$

where  $h(t)$  is the filter's impulse response and  $F[\cdot]$  is a nonlinear memoryless functional. It has been shown that the **STA** is proportional to the filter  $h(t)$  [Rieke et al., 1997; Chichilnisky, 2001]. This result is very useful since it allows to separate the calculus of the linear filtering properties of the neuron response from the nonlinearities in the spike generation [Keat et al., 2001]. The **STA** can also be interpreted as the neuron receptive field since it represents the preferred stimulus of the ganglion cell in terms of time profile [Rust et al., 2004].

We must be aware that, in the expression for the **STA** of Eq. (B.39), the stimulus segments gathered from different trials are weighted differently according to the number of spikes contained in the trial – the stimulus segments belonging to a train with fewer spikes have a greater weight in the overall sum. However, the important relations between Eq. (B.39) and other quantities, such as the neural function in Eq. (B.42), the firing rate in Eq. (B.43), and the correlation function in Eq. (B.45), are valid only if the approximation  $n_j \simeq \langle n \rangle$  is fulfilled.

We can simplify the calculation of the **STA** for discrete-time. If the sequence  $s[n]$  represents the discrete input stimulus of a **RGC**, whose response is the binary sequence  $\rho[n]$ , which is composed by a sequence of unit impulse functions, then we can calculate the **STA** using the expression

$$s_{\text{spk}}[l] = \frac{\sum_{j=1}^M \sum_{n=1}^N \rho_j[n] s[n-l]}{\sum_{j=1}^M \sum_{n=1}^N \rho_j[n]}, \quad 0 \leq l \leq l_{\text{max}}, \quad (\text{B.47})$$

where  $N$  is the length of the stimulus sequence and  $\rho_j[n]$  is the neural response of the trial  $j$  of a total of  $M$  responses obtained in the experiment by applying the same stimulus. In Eq. (B.47), all stimulus segments before every spike in every available trial are summed, and the result is divided by the total number of spikes occurring in all trials. In matrix notation, the **STA** can be represented by the vector:

$$\mathbf{s}_{\text{spk}} = \left[ s_{\text{spk}}[0] \quad s_{\text{spk}}[1] \quad \cdots \quad s_{\text{spk}}[l_{\text{max}}] \right]^T, \quad (\text{B.48})$$

with dimension  $l_{\text{max}}$  corresponding to the neuron memory.

The **STA** is particularly meaningful if the stimulus is composed of a white noise sequence, meaning that it is uncorrelated. For a stimulus sequence that is not autocorrelated, the resulting **STA** sequence goes to zero for a time lag larger than the neuron memory. The time lag,  $l_{\text{max}}$ , is the discrete dual of  $\tau_{\text{max}}$  in Eq. (B.39) for the continuous, and they are related by  $l_{\text{max}} = \tau_{\text{max}}/\Delta t$ , where  $\Delta t$  is the sampling period used in the discretization of the stimulus and of the neural response. Depending

on the species, a time lag between 40 ms and 950 ms is used [Keat et al., 2001] to calculate the STA of the retinal ganglion cells.

Although we have considered only time in showing how to calculate the STA, we can also include space. To calculate the spatial spike-triggered average, we must average the sequence of images (and not only a single luminance value for each time instant) that generated a fire for every spike occurrence. This is the procedure to obtain, for example, the spatial and temporal form of the receptive field of a neuron [Chichilnisky, 2001].

#### B.1.4. Spike Train Autocorrelation Function

The *spike train autocorrelation function* gives the time relation between any two spikes in a given spike train. It gives the distribution of times between any two spikes in a train.

The spike train autocorrelation function is the autocorrelation of the quantity obtained by subtracting from the neural function, defined in Eq. (B.7), the mean firing rate  $\langle r \rangle$  averaged over time and over trials:

$$C_{\rho\rho}(\tau) = \frac{1}{T} \int_0^T \langle (\rho(t) - \langle r \rangle)(\rho(t + \tau) - \langle r \rangle) \rangle dt . \quad (\text{B.49})$$

In fact, the expression in Eq. (B.49) follows the autocovariance definition since the mean firing rate is subtracted from the neural function before the autocorrelation computation [Therrien, 1992]. Nonetheless, it is called autocorrelation in the neuroscience literature [Dayan and Abbot, 2001].

We can define, analogously to the spike train autocorrelation, a cross-correlation between spike trains generated by different neurons. The *cross-correlation function* between a pair of spike trains is useful when we are looking for synchronicity in the firing from different neurons. For example, if the cross-correlation between two spike trains from different neurons have a peak at zero lag means that the neurons are firing synchronously, while a shift of the peak from zero indicates that the neurons are firing synchronously with a phase shift. The autocorrelation function is an even function of the lag  $\tau$  so that  $C_{\rho\rho}(\tau) = C_{\rho\rho}(-\tau)$ , while the cross-correlation function is not an even function of the lag, which means that  $C_{\rho_1\rho_2}(\tau) \neq C_{\rho_1\rho_2}(-\tau)$ .

In practice, the spike train autocorrelation is computed by: *i*) sampling the continuous neural function into time bins of width  $\Delta t$ ; *ii*) the number of spikes in each time bin is recorded into a sequence; *iii*) the spike count sequence, corresponding to the mean, is subtracted from this sequence; finally, *iv*) the autocorrelation of the sequence is computed. The resulting autocorrelation can be plotted in the form of a histogram as a function of the lag between the time bins.

In discrete-time, the spike train autocorrelation function is calculated by the autocorrelation of the neural response with the average firing rate removed, as shown

## B. Characterization of the Neural Response

by the expression:

$$C_{\rho\rho}[l] = \frac{1}{N} \frac{1}{M} \sum_{j=1}^M \sum_{n=1}^N (\rho_j[n] - \langle r \rangle)(\rho_j[n+l] - \langle r \rangle), \quad (\text{B.50})$$

where  $l$  represents the lag between two spikes measured in number of time bins. As before,  $N$  is the total number of time bins, corresponding to the neural response length, and  $M$  is the number of trials included in the calculation.

### B.1.5. The Spike Triggered Covariance

The *spike-triggered covariance* is another important quantity that characterizes the neural response. Its application is gaining increasing relevance in recent studies that examine the modeling of neuronal responses, particularly of the retinal response [Schwartz et al., 2002; Simoncelli et al., 2004]. The **STC** is used to obtain parameters from the firing rate second order statistics, and it can also be used to recover a series of linear filters that resemble the neural response in the presence of both symmetric and asymmetric nonlinearities [Rust et al., 2004].

Qualitatively, the **STC** function tells us how the stimulus varies with itself before a spike is fired. It is obtained by computing the covariance between the stimulus segments before every spike. If the occurrence of spike  $i$  in the trial  $j$  is represented by  $t_{ij}$ , where  $1 \leq i \leq n_j$  and  $1 \leq j \leq M$ , the **STC** is defined in continuous time by

$$\begin{aligned} C_{ss}(\tau_1, \tau_2) &= \frac{1}{\sum_{j=1}^M \int_0^T \rho_j(t)} \\ &\times \sum_{j=1}^M \sum_{i=1}^{n_j} [(s(t_{ij} - \tau_1) - s_{\text{spk}}(\tau_1)) \cdot (s(t_{ij} - \tau_2) - s_{\text{spk}}(\tau_2))] \\ &= \frac{1}{\sum_{j=1}^M n_j} \sum_{j=1}^M \sum_{i=1}^{n_j} s(t_{ij} - \tau_1) \cdot s(t_{ij} - \tau_2) - s_{\text{spk}}(\tau_1) s_{\text{spk}}(\tau_2), \end{aligned} \quad (\text{B.51})$$

where  $s_{\text{spk}}(\tau)$  is the spike triggered average and the denominator corresponds to the total number of spikes from all trials. The **STC** is a bi-dimensional function that returns the stimulus variation at time  $\tau_1$ , before the spike is fired, as a function of its value at time  $\tau_2$ , before the spike is fired. In discrete-time, the expression for the **STC** is

$$C_{ss}[l_1, l_2] = \frac{1}{\sum_{j=1}^M n_j} \sum_{j=1}^M \sum_{i=1}^{n_j} [s[n_{ij} - l_1] - s_{\text{spk}}[l_1]] \cdot [s[n_{ij} - l_2] - s_{\text{spk}}[l_2]], \quad (\text{B.52})$$

where  $n_{ij}$  is the bin of the trial  $j$  where the spike  $i$  occurs and  $s_{\text{spk}}[l]$  is the discrete spike-triggered average.

If the stimulus vector before spike  $i$  in trial  $j$  is represented by  $\mathbf{s}_{ij}$ ,

$$\mathbf{s}_{ij} = \begin{bmatrix} s[n_{ij}] \\ s[n_{ij} - 1] \\ \vdots \\ s[n_{ij} - l_{\max}] \end{bmatrix}, \quad (\text{B.53})$$

the **STC** matrix can be written as:

$$\mathbf{C}_{\text{spk}} = \frac{1}{\sum_{j=1}^M n_j} \sum_{j=1}^M \sum_{i=1}^{n_j} (\mathbf{s}_{ij} - \mathbf{s}_{\text{spk}})(\mathbf{s}_{ij} - \mathbf{s}_{\text{spk}})^T, \quad (\text{B.54})$$

where  $\mathbf{s}_{\text{spk}}$  is the **STA** vector.

## B.2. Stimulus and Response Statistics, and Firing Probabilities

As stated before, to estimate a probability with some reliability, we need a large number of data vectors. Or, if we are interested only in some common statistical parameters, such as the mean and variance, we need a very long experiment in time for the case of an ergodic process. The former case is more common in experimental neuroscience: the same stimulus is repeatedly presented to the neuron, and the respective responses are recorded. For the case of the retina neural circuit, the responses to the same stimulus show some variability that justifies, and even enforces, the application of probability tools to characterize and describe the neural code.

In a common experimental setup, the retina is excited with a visual stimulus,  $s(t)$ , chosen by the experimenter. As a consequence, a sequence of spikes, occurring at times  $t_1, t_2, \dots, t_n$ , which correspond to a particular neural response,  $\rho_j(t)$ , is recorded. Then, the probability that the neuron will fire conditioned to that particular stimulus,  $P(\rho_j(t)|s(t))$ , can be estimated from the collected responses.

The stimulus itself can be drawn from a given probability distribution  $P[s(t)]$ , such as a Gaussian distribution, for example, defining an ensemble of stimuli signals. The stimulus distribution may even resemble the statistical properties of natural scenes [Yu and de Sa, 2004].

Since the stimulus is random and the neuron response has also a random nature, the neuron activity can be described by the joint probability distribution of the stimuli signals and the spike trains:  $P(\rho_j(t), s(t))$ . This joint distribution quantifies the likelihood that, in the course of the experiment, the stimulus  $s(t)$  and the spike train  $\rho_j(t)$  will both be observed. By employing the relationship between the joint, the marginal, and the conditional probability distributions, the joint distribution can be written using the response distribution conditioned to a given stimulus as

$$P(\rho_j(t), s(t)) = P(\rho_j(t)|s(t)) P(s(t)), \quad (\text{B.55})$$

## B. Characterization of the Neural Response

where  $P(s(t))$  is the stimulus marginal probability distribution. We can also try to find which spike train a given stimulus will trigger. As mentioned before, there is not a one-to-one relationship between a given stimulus and the generated spike train, but we can write the following joint probability distribution:

$$P(s(t), \rho_j(t)) = P(s(t)|\rho_j(t)) P(\rho_j(t)) . \quad (\text{B.56})$$

By equating the probability distributions of Eq. (B.55) and Eq. (B.56), we have

$$P(\rho_j(t)|s(t)) P(s(t)) = P(s(t)|\rho_j(t)) P(\rho_j(t)) . \quad (\text{B.57})$$

From Eq. (B.57), we can follow two different perspectives about the neural code. From a modeling, or encoding, perspective we would like to know the distribution  $P(\rho_j(t)|s(t))$  so that, given a certain stimulus, the most likely neural response can be obtained. We can then generate the spike train to stimulate the superior parts of the nervous system. The desired distribution has the expression

$$P(\rho_j(t)|s(t)) = \frac{P(s(t)|\rho_j(t)) P(\rho_j(t))}{P(s(t))} . \quad (\text{B.58})$$

This last equality expresses Bayes' rule that relates marginal and conditional probabilities.

By taking the brain decoding perspective, which perceives the stimulus from the received spike train, the distribution of interest is:

$$P(s(t)|\rho_j(t)) = \frac{P(\rho_j(t)|s(t)) P(s(t))}{P(\rho_j(t))} . \quad (\text{B.59})$$

Next, we will present the most common probability distributions employed in the analysis of the neural response.

### B.2.1. Spike Train Statistics

The structure of a spike train can be perfectly regular, as in the trains generated by neurons controlling the heart beat where spikes are fired regularly at almost constant time intervals. On the opposite extreme, they can show a completely random behavior, where spikes are fired independently of past history, as is observed in the brain.

The retinal ganglion cells fire spontaneously, even in the dark, such that those spike trains obviously do not convey any visual information and should be considered as pure noise. However, when properly stimulated, retinal ganglion cells can produce very reproducible spike trains from trial to trial [Berry et al., 1997], which means that the spike train variability is controlled.

The stochastic process that generates a sequence of stereotyped discrete events, such as spikes, is called a *point process*. The probability  $P(t_1, t_2, \dots, t_n)$  that a sequence of  $n$  spikes occurring at the time instants  $\{t_1, t_2, \dots, t_n\}$  is proportional

to the probability density of the spike occurrences at those times:  $p(t_1, t_2, \dots, t_n)$ . Specifically, the probability of occurrence of a sequence of  $n$  spikes, with spike  $i$  occurring between time instants  $t_i$  and  $t_i + \Delta T$ , with  $i = 1, 2, \dots, n$ , is given by:

$$P(t_1, t_2, \dots, t_n) = p(t_1, t_2, \dots, t_n) \cdot (\Delta T)^n . \quad (\text{B.60})$$

However, the possible number of different spike sequences is so huge that we would need an infinite amount of data to fully characterize the probability density  $p(t_1, t_2, \dots, t_n)$ . The impossibility of reliably characterizing the probability density in Eq. (B.60) to a desirable degree led to the development of statistical models that simplify the description of a spike train.

In general, the generation of a spike can depend on the full history of the process. That is, the generation of a spike at time  $t_n$  can depend on all the spikes generated previously, so that the probability of a neuron firing a spike at the time instant  $t_n$  is dependent on all the previously generated spikes. This is described mathematically as the conditional probability

$$P(t_n | t_1, t_2, \dots, t_{n-1}) . \quad (\text{B.61})$$

By observing the firing dynamics of a neuron, it was noticed that, for most cases, the firing of a spike depends mainly on the last generated spike. If the probability of spike generation depends only on the last generated spike, the conditional probability of Eq. (B.61) can be simplified to

$$P(t_n | t_1, t_2, \dots, t_{n-1}) = P(t_n | t_{n-1}) , \quad (\text{B.62})$$

and the point process is deemed a renewal process.

The opposite case of Eq. (B.61) is when the generation of a spike is considered to be independent of the history of the whole process, so that the firing of a spike is statistically independent of the previous spikes occurrences, such that

$$P(t_n | t_1, t_2, \dots, t_{n-1}) = P(t_n) . \quad (\text{B.63})$$

This stochastic process is called a Poisson process. Poisson processes play an important role in the description of spike trains statistics, providing a useful approximation of the stochastic neuronal firing.

As we have seen in Sec. B.1.2, the firing rate  $r(t)$ , defined formally by Eq. (B.20), is proportional to the probability of the neuron to fire a single spike around the time  $t$ , and, if it is considered that the occurrence of one given spike is independent of the occurrence of other spikes, it can be used to compute the probabilities for all possible action potentials. Specifically, the probability of a neuron firing a spike between the time instants  $t$  and  $t + \Delta t$  is:

$$P(\text{spike in } [t, t + \Delta t]) = r(t)\Delta t. \quad (\text{B.64})$$

Poisson processes can be classified into two main categories: homogeneous Poisson processes, where the firing rate is constant along the spikes' generation,  $r(t) = r$ , and inhomogeneous Poisson process, where the firing rate is time dependent,  $r(t)$ .

## B. Characterization of the Neural Response

The inhomogeneous Poisson processes can be further adapted in several different ways to better model the spike train statistics, namely, the inclusion of a model of the absolute refractory period, where the probability of firing is null, and the relative refractory period, where the probability of firing is initially low but increases continually [Berry II and Meister, 1998]. The existence of the refractory period means that the spike events are not independent and that the firing rate by itself is unable to fully describe a spike train correctly.

The majority of neuron models, and retina models in particular, are rate models. This means that their goal is to deliver at the output an estimate of the cell firing rate, but what they convey to the brain is a train of spikes. So, the characterization of the spike train in statistical terms is very important in order to have a model to generate spike trains while departing from an estimation of the firing rate.

### B.2.2. Homogeneous Poisson Model of Spike Trains

As stated before, a spike train is described by the neural response function,  $\rho(t)$ , with spikes placed at time instants  $t_1, t_2, \dots, t_n$ .

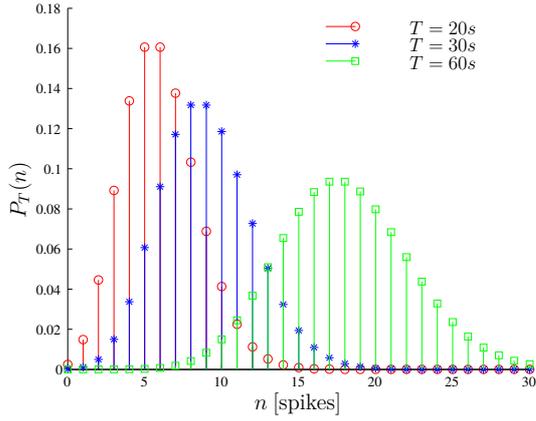
First, we divide the spike train with a total time duration  $T$  into time bins sufficiently small so that there is at most one spike in each bin. If the width of each of these time bins is  $\Delta t$ , then there are  $N = T/\Delta t$  time bins. Note that the probability of having a spike in a time bin is equal to  $r\Delta t$ . If we observe a particular spike train with  $n$  spikes in the  $N$  bins ( $n \leq N$ ), the probability of the occurrence of this very particular sequence is equal to the probability of having  $n$  bins with spikes, which is  $(r\Delta t)^n$  for a constant firing rate, times the probability of having the other  $N - n$  time bins without any spike, which is  $(1 - r\Delta t)^{N-n}$ . By the previous reasoning, a particular spike sequence has the probability

$$P(\underbrace{101110 \dots 001}_{n \text{ spikes}}) = (r\Delta t)^n (1 - r\Delta t)^{N-n} . \quad (\text{B.65})$$

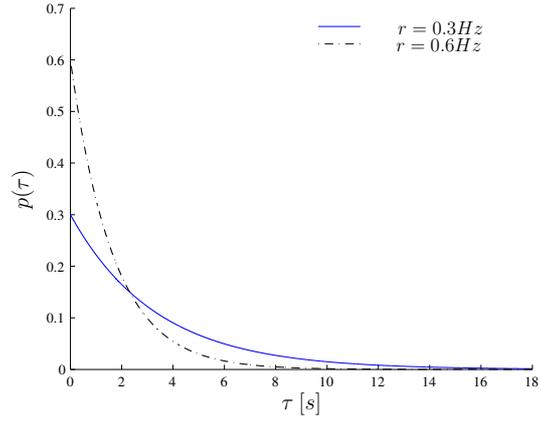
The argument of  $P(\cdot)$  in Eq. (B.65) represents a particular spike train in which we are interested, where a spike is represented by a 1, while an empty bin is represented by 0.

#### Spike Count Distribution

Considering first a description of the spike train that disregards the precise time occurrences of each spike, where we are concerned only with the probability of having exactly  $n$  spikes within the trial of duration  $T$ . We denote this probability by  $P_T(n)$ , called the *spike count distribution*. If we are not interested in a specific spike train, but instead with the probability of having  $n$  spikes within a spike train, independent of the particular time bins in which they appear, then we can have  $N! / (n!(N - n)!)$  different spike trains with exactly  $n$  spikes distributed among the  $N$  bins. Therefore, the probability of getting  $n$  spikes in a time period equal to  $T$ , for a constant firing



**Figure B.25.** Spike number probability density for a train described by a homogeneous Poisson process with  $r = 0.3$  Hz.



**Figure B.26.** Interspike time interval exponential probability density for a spike train described by a homogeneous Poisson process.

rate  $r$ , is given by Eq. (B.65) multiplied by the combinatorial factor

$$P_T(n) = \frac{N!}{n!(N-n)!} (r\Delta t)^n (1-r\Delta t)^{N-n}. \quad (\text{B.66})$$

Equation (B.66) is a discrete probability function, or probability mass function (pmf), corresponding to a binomial distribution. The binomial distribution reduces to the Bernoulli distribution for the particular case of  $n = 1$  [Mood et al., 1974; Papoulis and Pillai, 2002], which corresponds to the probability of the occurrence of a single spike during the time period  $T$ . Figure B.25 shows a plot of the probabilities for the number of spikes using the same firing rate  $r = 0.3$  spikes/s for different values of the trial duration  $T$ . We can note that, as the value of  $T$  increases, the probability of having a larger number of spikes also increases.

In Eq. (B.66), each spike is placed in a given time bin, within the  $N$  bins available, which corresponds to a time precision of  $\Delta t$ . However, in continuous terms, the spikes can be placed at any time instant so that we can calculate the limit when the width of the time bins goes to zero. Since  $N = T/\Delta t$  as  $\Delta t \rightarrow 0$ , the number of time bins,  $N$ , will grow, but  $N\Delta t = T$  remains constant, so that Eq. (B.66) becomes:

$$\begin{aligned} P_T(n) &= \lim_{\Delta t \rightarrow 0} \frac{N!}{n!(N-n)!} (r\Delta t)^n (1-r\Delta t)^{N-n} \\ &= \lim_{N \rightarrow \infty} \frac{N(N-1)\cdots(N-n+1)}{n!} \left(r\frac{T}{N}\right)^n \left(1-r\frac{T}{N}\right)^{N-n}. \end{aligned} \quad (\text{B.67})$$

## B. Characterization of the Neural Response

By unfolding the terms in the previous equation, it follows:

$$P_T(n) = \lim_{N \rightarrow \infty} \frac{N(N-1) \cdots (N-n+1)}{N^n} \frac{1}{n!} (rT)^n \times \left( \left( 1 + \frac{-rT}{N} \right)^{\frac{N}{-rT}} \right)^{-rT} \left( 1 - r \frac{T}{N} \right)^{-n}. \quad (\text{B.68})$$

Taking the limit and knowing that  $\lim_{\epsilon \rightarrow 0} (1 + \epsilon)^{1/\epsilon} = e$ , we obtain the following equation for the discrete probability density:

$$P_T(n) = \frac{(rT)^n}{n!} e^{-rT}, \quad (\text{B.69})$$

corresponding to a Poisson distribution, where the constant firing rate  $r$  appears explicitly.

Two important statistical parameters of a random variable are the mean, or expected value, and the variance. The mean gives the most probable value of the random variable after many observations. For a discrete random variable  $x$ , such as the number of spikes in a spike train given by the discrete probability density function (pdf) of Eq. (B.69), with discrete density  $P_X(x)$ , the mean value, or average, is calculated from

$$\mu_x = E\{x\} = \langle x \rangle = \sum_{k=-\infty}^{+\infty} x_k P_X(x_k), \quad (\text{B.70})$$

where  $x_k$  denote the values that the discrete random variable  $x$  can take and  $P_X(x_k)$  their probability of occurrence. The variance gives a measure of the dispersion of the values of the random variable  $x$  around its mean  $\mu_x$ . It has the definition

$$\begin{aligned} \sigma_x^2 &= \text{VAR}(x) = E\{(x - \mu_x)^2\} \\ &= E\{x^2 - 2\mu_x x + \mu_x^2\} = E\{x^2\} - \mu_x^2, \end{aligned} \quad (\text{B.71})$$

where  $\sigma_x^2 \geq 0$  is a positive quantity. The last equality in Eq. (B.71) results from the fact that the expectation operator  $E\{\cdot\}$  defined in Eq. (B.70) is linear. This relation holds both for discrete and continuous random variables and is very useful to calculate the variance. For a discrete random variable, the variance can be computed directly through the expression

$$\sigma_x^2 = \sum_{k=-\infty}^{+\infty} (x_k - \mu_x)^2 P_X(x_k). \quad (\text{B.72})$$

The expected number of spikes, and its variance, within a trial can be calculated using Eq. (B.69) based on Eq. (B.70) and Eq. (B.72), respectively. However, these quantities are more easily calculated with the help of the *moment generating function*. The moment generating function,  $M(t)$ , for a discrete probability density  $P(n)$  is defined as [Papoulis and Pillai, 2002]

$$M(t) = \sum_{n=0}^{+\infty} e^{tn} P(n). \quad (\text{B.73})$$

## B.2. Stimulus and Response Statistics, and Firing Probabilities

The moment generating function is useful because the different statistical moments,  $m_k$ , can be easily calculated by using the relation:

$$m_k = M^{(k)}(0), \quad k = 1, 2, \dots, +\infty, \quad (\text{B.74})$$

where  $M^{(k)}(0)$  is the  $k$ -order derivative calculated at the origin:

$$M^{(k)}(0) = \frac{d^k}{dt^k} M(t)|_{t=0}. \quad (\text{B.75})$$

The pdf in Eq. (B.69) has the moment generating function:

$$\begin{aligned} M(t) &= \sum_{n=0}^{+\infty} e^{tn} \frac{(rT)^n}{n!} e^{-rT} = e^{-rT} \sum_{n=0}^{\infty} \frac{(rT e^t)^n}{n!} \\ &= e^{-rT} e^{rT e^t}. \end{aligned} \quad (\text{B.76})$$

The expected number of spikes,  $\langle n \rangle$ , in a trial with duration  $T$  described by the Poisson pdf in Eq. (B.69), corresponds to the first moment,  $m_1$ , of Eq. (B.76):

$$\begin{aligned} E\{n\} = \langle n \rangle &= \sum_{n=0}^{\infty} n P_T(n) = m_1 \\ &= rT, \end{aligned} \quad (\text{B.77})$$

which states that the expected number of spikes generated at a constant firing rate is equal to the firing rate  $r$  times the considered time interval  $T$ . This can be seen in Fig. B.25. The variance of the spike count in  $T$  is

$$\begin{aligned} \sigma_n^2 &= E\{(n - \langle n \rangle)^2\} = \langle n^2 \rangle - \langle n \rangle^2 = m_2 - m_1^2 \\ &= rT, \end{aligned} \quad (\text{B.78})$$

which is equal to the expected number of spikes. A parameter that characterizes a probability distribution is the Fano factor, defined as

$$F = \frac{\sigma^2}{\mu^2}, \quad (\text{B.79})$$

where  $\sigma^2$  is the distribution variance and  $\mu$  the mean. The Fano factor characterizes the spike count variability, and, for the case of the Poisson distribution, it takes the value

$$F = \frac{\sigma_n^2}{\langle n \rangle} = 1. \quad (\text{B.80})$$

The fact that the spike count mean and variance are equal is a distinguishing characteristic of a Poisson process. In practice, to analyze if a spike train is adequately described by a Poisson process, one can check if its Fano factor is approximately one [Berry II and Meister, 1998].

### Interspike Interval Distribution

The probability distribution of the time intervals between adjacent spikes is called the *interspike interval distribution*, and it constitutes another useful characterization of spike patterns.

Given a spike occurring at time  $t_i$ , for some value of  $i$ , the **ISI** distribution gives the waiting time for the next spike to occur. This probability is equal to the probability that no spikes are generated during a certain interval of time times the probability that a spike is generated in the next time interval.

The probability of not generating a spike in the time interval  $[t_i, t_i + \tau]$ , given that a spike occurred at  $t_i$ , of a homogeneous Poisson process, can be obtained by plugging  $n = 0$  into Eq. (B.69), which gives

$$P_\tau(n = 0) = e^{-r\tau} . \quad (\text{B.81})$$

The probability of generating a spike in the time interval  $\Delta t$ , for a small  $\Delta t$ , is given by:

$$P_{\Delta t \rightarrow 0}(n = 1) = r\Delta t . \quad (\text{B.82})$$

Therefore, the probability that a spike was fired at instant  $t_i$  and that the next spike is generated at  $t_{i+1}$ , such that  $t_i + \tau \leq t_{i+1} < t_i + \tau + \Delta t$  is:

$$P(\tau \leq t_{i+1} - t_i < \tau + \Delta t) = r\Delta t e^{-r\tau} . \quad (\text{B.83})$$

By definition, for a small  $\Delta t$ , the probability density of the **ISI** is equal to the probability in Eq. (B.83) divided by  $\Delta t$ , which gives

$$p_{ISI}(\tau) = r e^{-r\tau} , \quad \tau = t_{i+1} - t_i, \quad (\text{B.84})$$

which shows that the interspike-time probability density for a homogeneous Poisson spike train is exponential. As Fig. B.26 shows, the short interspike intervals are more likely to occur while the long ones have an exponentially decaying probability that is a function of their duration.

An interesting characteristic of the exponential **pdf** relevant to the description of **ISI** distribution is that it is memoryless, meaning that the time that we have to wait for a new spike is independent of the time that we have already been waiting so far. Let us calculate the probability of a spike occurring in the next period of time  $\tau_0 + \tau_1$ , given that we already have been waiting for it during  $\tau_0$ , which means that we want to compute

$$P_{ISI}(t > \tau_0 + \tau_1 | t > \tau_0) . \quad (\text{B.85})$$

By using Bayes' law, this probability is equal to:

$$P_{ISI}(t > \tau_0 + \tau_1 | t > \tau_0) = \frac{P_{ISI}(t > \tau_0 + \tau_1)}{P_{ISI}(t > \tau_0)} . \quad (\text{B.86})$$

To calculate the probabilities in the numerator and denominator of Eq. (B.86), we have to integrate the continuous pdf of Eq. (B.84)

$$P_{ISI}(t > \tau_0 + \tau_1) = \int_{\tau_0 + \tau_1}^{+\infty} r e^{-r\tau} d\tau = e^{-r(\tau_0 + \tau_1)}, \quad (\text{B.87})$$

and

$$P_{ISI}(t > \tau_0) = \int_{\tau_0}^{+\infty} r e^{-r\tau} d\tau = e^{-r\tau_0}, \quad (\text{B.88})$$

giving

$$P_{ISI}(t > \tau_0 + \tau_1 | t > \tau_0) = e^{-r\tau_1} = P_{ISI}(t > \tau_1), \quad (\text{B.89})$$

which means that the fact that no spike was fired during the period  $\tau_0$  does not influence the probability of getting one in the next period  $\tau_1$ . From the interspike interval pdf in Eq. (B.84), we can calculate the mean interspike interval.

For a continuous pdf,  $f(x)$ , like the ISI density in Eq. (B.84) the mean, or expected value, of the random variable  $x$  is defined as

$$E\{x\} = \mu_x = \int_{-\infty}^{+\infty} x f(x) dx, \quad (\text{B.90})$$

and the variance of the random variable  $x$  has the definition:

$$\sigma_x^2 = E\{(x - \mu_x)^2\} = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x) dx. \quad (\text{B.91})$$

Using the definition of Eq. (B.90), the expected value for the ISI is:

$$E\{\tau\} = \mu_{ISI} = \langle \tau \rangle = \int_0^{\infty} \tau r e^{-r\tau} d\tau = \frac{1}{r}, \quad (\text{B.92})$$

and, by Eq. (B.91), the variance of the interspike intervals gives

$$\sigma_{ISI}^2 = \langle \tau^2 \rangle - \langle \tau \rangle^2 = \int_0^{\infty} \tau^2 r e^{-r\tau} d\tau - \left(\frac{1}{r}\right)^2 = \frac{1}{r^2}. \quad (\text{B.93})$$

The coefficient of variation, defined as the ratio between the standard deviation,  $\sigma$ , and the mean,  $\mu$ , is given by

$$C_V = \frac{\sigma}{\mu}. \quad (\text{B.94})$$

For the ISI distribution of a Poisson spike train, it takes the value:

$$C_V = \frac{\sigma_{ISI}}{\mu_{ISI}} = 1. \quad (\text{B.95})$$

Equation (B.95) is a required, but not sufficient, condition to identify a Poisson spike train. For a renewal process, the Fano factor, evaluated over long time intervals, approaches  $C_V^2$  [Dayan and Abbot, 2001].

### B.2.3. Inhomogeneous Poisson Model of Spike Trains

As stated in Sec. B.1.2, the time-dependent firing rate  $r(t)$  gives the probability, per unit of time, of a spike occurrence. Equation (B.22) states that the probability of a spike occurring within a time bin of width  $\Delta t$ , at time  $t$ , is  $r(t)\Delta t$ .

Due to the statistical independence of the spike occurrences for a Poisson process, stated by Eq. (B.63), the occurrence probability of a sequence, with spikes placed at time instants  $t_1, t_2, \dots, t_n$ , is equal to the probability of finding the spikes in those specific  $n$  time bins times the probability of not finding any spike in the remaining  $N - n$  time bins ( $N = T/\Delta t$ ).

The probability of finding a spike in a bin with width  $\Delta t$ , centered at the time instant  $t_i$ , is  $r(t_i)\Delta t$ , and the probability of not having a spike at the time bin  $t_j$  is  $1 - r(t_j)\Delta t$ . The probability of the particular spike train occurring with spikes located exactly at times  $t_1, t_2, \dots, t_n$ , where  $0 \leq t_i \leq T$ , is given by

$$\begin{aligned} P(t_1, t_2, \dots, t_n)(\Delta t)^n &= \frac{1}{n!} \prod_{j=1; j \neq i}^N (1 - r(t_j)\Delta t) \prod_{i=1}^n r(t_i)\Delta t \\ &= \frac{1}{n!} \prod_{j=1}^N (1 - r(t_j)\Delta t) \prod_{i=1}^n \frac{r(t_i)\Delta t}{1 - r(t_i)\Delta t}, \end{aligned} \quad (\text{B.96})$$

where the index  $j$  represents all  $N$  possible time bins with width  $\Delta t$ . The factor  $1/n!$  is introduced because the spikes are indistinguishable and there are  $n!$  different ways of assigning the labels. The first product in the second equality of Eq. (B.96) can be simplified by considering that

$$\prod_{j=1}^N (1 - r(t_j)\Delta t) = \exp\left(\sum_{j=1}^N \ln(1 - r(t_j)\Delta t)\right), \quad (\text{B.97})$$

and by simplifying the logarithm computation. Since  $\Delta t$  can be made very small, and since, for small  $x$ , the Taylor's expansion of the logarithm in the neighborhood of one is  $\ln(1 + x) = x - 1/2x^2 + 1/3x^3 - \dots$ , then Eq. (B.97) can be approximated to:

$$\begin{aligned} &\prod_{j=1}^N (1 - r(t_j)\Delta t) = \\ &= \exp\left(\sum_{j=1}^N \left(-r(t_j)\Delta t - \frac{1}{2}(-r(t_j)\Delta t)^2 + \frac{1}{3}(-r(t_j)\Delta t)^3 + \dots\right)\right) \\ &= \exp\left(-\sum_{j=1}^N r(t_j)\Delta t + \frac{1}{2}\Delta t \sum_{j=1}^N r^2(t_j)\Delta t - \frac{1}{3}(\Delta t)^2 \sum_{j=1}^N r^3(t_j)\Delta t + \dots\right), \end{aligned} \quad (\text{B.98})$$

By applying the property

$$\lim_{\Delta t \rightarrow 0} \sum_{j=1}^{T/\Delta t} f(t_j)\Delta t = \int f(t)dt, \quad (\text{B.99})$$

replacing the terms of the exponential of Eq. (B.98), and computing the limit when  $\Delta t \rightarrow 0$ , we arrive at

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \prod_{j=1}^N (1 - r(t_j)\Delta t) \\ &= \lim_{\Delta t \rightarrow 0} \exp \left( - \int_0^T r(t)dt + \frac{1}{2}\Delta t \int_0^T r^2(t)dt - \frac{1}{3}(\Delta t)^2 \int_0^T r^3(t)dt + \dots \right) \quad (\text{B.100}) \\ &= \exp \left( - \int_0^T r(t)dt \right) . \end{aligned}$$

The second product in Eq. (B.96) can also be simplified to:

$$\prod_{i=1}^n \left( \frac{r(t_i)\Delta t}{1 - r(t_i)\Delta t} \right) = (\Delta t)^n \prod_{i=1}^n \left( \frac{r(t_i)}{1 - r(t_i)\Delta t} \right) \quad (\text{B.101})$$

and the limit when  $\Delta t \rightarrow 0$  is:

$$\lim_{\Delta t \rightarrow 0} (\Delta t)^n \prod_{i=1}^n \left( \frac{r(t_i)}{1 - r(t_i)\Delta t} \right) = (\Delta t)^n \prod_{i=1}^n r(t_i) . \quad (\text{B.102})$$

Replacing the simplifications obtained in Eq. (B.100) and Eq. (B.101), the discrete probability density of Eq. (B.96) becomes

$$P(t_1, t_2, \dots, t_n) = \frac{1}{n!} \exp \left( - \int_0^T r(t)dt \right) \prod_{i=1}^n r(t_i) , \quad (\text{B.103})$$

which corresponds to the probability density of a spike train with spikes occurring exactly at the time instants  $t_1, t_2, \dots, t_n$ .

The homogeneous Poisson process is a particular case of the inhomogeneous process where the firing rate is held constant during the entire trial duration,  $r(t) = r$ . The joint probability density for a given number of spikes occurring at specific instants  $t_i$ , for a constant firing rate, is

$$P(t_1, t_2, \dots, t_n) = \frac{1}{n!} e^{-rT} r^n , \quad (\text{B.104})$$

which is independent of the time occurrence of the spikes. This density can be written as a function of the spike count density,  $P_T(n)$ , for the homogeneous case given in Eq. (B.69) as:

$$P(t_1, t_2, \dots, t_n) = P_T(n) \left( \frac{1}{T} \right)^n . \quad (\text{B.105})$$

### Spike-count statistics

The spike-count distribution characterizes the distribution of the number of spikes from a Poisson model. The joint probability density obtained in expression Eq. (B.103) gives the probability of the occurrence of  $n$  spikes at the time instants  $t_1, t_2, \dots, t_n$ . To characterize the distribution of the number of spikes along a spike train, Eq. (B.103) must be integrated for all possible time occurrences of the  $n$  spikes within  $T$ . That is,

$$\begin{aligned} P(n) &= \int_0^T \int_0^T \cdots \int_0^T P(t_1, t_2, \dots, t_n) dt_1 dt_2 \cdots dt_n \\ &= \int_0^T \int_0^T \cdots \int_0^T \frac{1}{n!} \exp\left(-\int_0^T r(t) dt\right) \prod_{i=1}^n r(t_i) dt_1 dt_2 \cdots dt_n \\ &= \frac{1}{n!} \exp\left(-\int_0^T r(t) dt\right) \int_0^T \int_0^T \cdots \int_0^T \prod_{i=1}^n r(t_i) dt_1 dt_2 \cdots dt_n, \end{aligned} \quad (\text{B.106})$$

resulting in

$$P(n) = \frac{1}{n!} \exp\left(-\int_0^T r(t) dt\right) \left(\int_0^T r(t) dt\right)^n. \quad (\text{B.107})$$

It should be noted that Eq. (B.69) can be generated from the previous expression considering a constant firing rate equal to  $r$ . The average number of spikes can be obtained with a similar procedure used for the homogeneous case, leading to

$$\langle n \rangle = \int_0^T r(t) dt, \quad (\text{B.108})$$

which, for the particular case of a constant firing rate  $r$ , gives the result presented in Eq. (B.77).

### Interspike Interval Distribution

The ISI distribution for the inhomogeneous Poisson process can be obtained by calculating the probability of not finding a spike during a time interval with length  $\tau$ , multiplied by the probability of firing a spike at  $t = \tau + \Delta t$ , with a small  $\Delta t$ . The probability that a spike is not fired during the time  $\tau$  can be obtained from Eq. (B.107) by making  $n = 0$  and assuming that the trial has a duration  $\tau$ , which results in:

$$P_\tau(n = 0) = \exp\left(-\int_0^\tau r(t) dt\right). \quad (\text{B.109})$$

As before, the probability of generating a single spike within the small time interval  $t \in [\tau, \tau + \Delta t]$  is:

$$P_{\Delta t}(\tau \leq t_i \leq \tau + \Delta t) = r(\tau) \Delta t. \quad (\text{B.110})$$

So the **ISI** probability density has the following density function:

$$p_{ISI}(\tau) = \exp\left(-\int_0^\tau r(t)dt\right) r(\tau) . \quad (\text{B.111})$$

### B.2.4. Firing Rates and Neuron Refraction

Normally the firing of a spike affects the firing of the next spike since during a time period known as the refractory time, when the neuron is restoring its ion levels, it is unable to fire again [Berry II and Meister, 1998]. Therefore, so that the description of the neuron firing as a Poisson process can be improved to take into account this effect.

So usually the firing rate depends on the time  $t_i$  the last spike was fired:  $r(t|t_i)$ . The simplest way to introduce a refractory period is to consider the firing rate to be given by [Gerstner and Kistler, 2002]:

$$r(t|t_i) = r(t)T(t - t_i) \quad (\text{B.112})$$

where  $T(t - t_i)$  is refraction function.

The probability of a neuron not firing a spike can be generalized for refractory firing rates. Supposing that a spike is fired at the time instant  $t_i$ , the probability that a spike is not fired during the time interval  $[t_i, t], t \geq t_i$ , which is equivalent to not fire a spike during  $\tau = t - t_i$ , can be obtained from Eq. (B.107) by making  $n = 0$  and assuming that the trial has a duration  $\tau$ , and by considering now that the firing rate is given by Eq. (B.112). By integrating the probability density from  $t_i$  to  $t, t \geq t_i$ , results:

$$P(\text{no spike in } [t_i, t]|t_i) = P(n = 0) = \exp\left(-\int_{t_i}^{t_i+t} r(\tau|t_i)d\tau\right) . \quad (\text{B.113})$$

This is an important quantity in the renewal process theory known by the survivor function:  $S(t|t_i)$ . The survivor function,  $S(t|t_i)$ , gives the probability that the neuron survives from  $t_i$  to  $t$  without firing any spike. For a nonhomogeneous Poisson process the survivor function is given by Eq. (B.109). Generalizing for a process with firing rate intensity, or hazard,  $r(t|t_i)$  the survivor function is:

$$S(t|t_i) = \exp\left(-\int_{t_i}^{t_i+t} r(\tau|t_i)d\tau\right) . \quad (\text{B.114})$$

The **ISI** distribution for a general firing rate  $r(t|t_i)$  can be obtained by multiplying the probability of not finding a spike during the time interval  $[t_i, t], t \geq t_i$ , given by Eq. (B.113), by the probability of firing a spike at  $t_{i+1} = t + \Delta t$ , with a very small  $\Delta t$ . As seen in Eq. (B.64), the probability of generating a single spike at time  $t_{i+1}$

## B. Characterization of the Neural Response

within the small time interval  $t_{i+1} \in [t, t + \Delta t]$ , now with the refractory firing rate, is:

$$P(\text{spike in } [t, t + \Delta t]) = r(t|t_i)\Delta t, \quad (\text{B.115})$$

so the **ISI** probability distribution is:

$$P_{ISI}(t|t_i) = \exp\left(-\int_{t_i}^{t_i+t} r(\tau|t_i)d\tau\right) r(t|t_i)\Delta t. \quad (\text{B.116})$$

By definition the **ISI** probability density is given by dividing the **ISI** probability of Eq. (B.116) by the infinitesimal time interval  $\Delta t$ , leading to the following density function:

$$p_{ISI}(t|t_i) = \exp\left(-\int_{t_i}^{t_i+t} r(\tau|t_i)d\tau\right) r(t|t_i). \quad (\text{B.117})$$

Note that using Eq. (B.114) the **ISI** probability density can be written as:

$$p_{ISI}(t|t_i) = S(t|t_i)r(t|t_i). \quad (\text{B.118})$$

The integration of **ISI** probability density gives the probability that a neuron which as fired a spike at  $t_i$  will fire the next spike at  $t$ . So the survivor function is directly related with the **ISI** probability density by:

$$S(t|t_i) = 1 - \int_{t_i}^t p_{ISI}(\tau|t_i)d\tau. \quad (\text{B.119})$$

The derivative Eq. (B.119) gives that the **ISI** is also given by:

$$p_{ISI}(t|t_i) = -\frac{d}{dt}S(t|t_i), \quad (\text{B.120})$$

in conjunction with Eq. (B.118) gives that the firing rate is:

$$r(t|t_i) = -\frac{d}{dt}S(t|t_i)/S(t|t_i). \quad (\text{B.121})$$

From (B.121) and Eq. (B.120) the firing rate with the **ISI** probability density can be related by:

$$r(t|t_i) = \frac{p_{ISI}(t|t_i)}{S(t|t_i)} = \frac{p_{ISI}(t|t_i)}{1 - \int_{t_i}^t p_{ISI}(\tau|t_i)d\tau}, \quad (\text{B.122})$$

where the last equality is due to Eq. (B.119).

The previous relations between the firing rate, survivor probability, and interspike interval probability density are very useful since if any of these is obtained experimentally the other quantities can be obtained easily [Dayan and Abbot, 2001].

### Refractory Firing Rates

The simplest refractory firing rate is to introduce an absolute refractory period after the emission of a spike. The refraction function would be:

$$T(t - t_i) = H(t - (t_i - T_{ref})) \quad (\text{B.123})$$

where the firing rate is zero during a period  $T_{ref}$  after the firing of a spike. The firing rate intensity given by (B.112) is termed hazard function in the terminology of renewal process.

Considering a spike train generated by a homogeneous Poisson with constant firing rate:  $r(t) = r$  (see Sec. B.2.2). Taking into account the absolute refractory period of the neuron, a time period after the firing of a spike is introduced according to Eq. (B.112) where the firing rate is zero, so that the neuron cannot fire within this time interval. The firing intensity, becomes:

$$\begin{aligned} r(t|t_i) &= r H(t - (t_i + T_{ref})) \\ &= \begin{cases} 0, & t < t_i + T_{ref} \\ r, & t \geq t_i + T_{ref} \end{cases} \end{aligned} \quad (\text{B.124})$$

From this firing rate the ISI probability density can be obtained with the help of (B.117), which gives

$$\begin{aligned} p_{ISI}(t) &= \begin{cases} 0, & t \leq t_i + T_{ref} \\ r e^{-r[t - (t_i + T_{ref})]}, & t > t_i + T_{ref} \end{cases} \\ &= r e^{-r[t - (t_i + T_{ref})]} H(t - (t_i + T_{ref})). \end{aligned} \quad (\text{B.125})$$

For the homogeneous Poisson with absolute refractory process with the firing rate of Eq. (B.124) the survivor function is:

$$\begin{aligned} S(t|t_i) &= \begin{cases} 1, & t < t_i + T_{ref} \\ e^{-r[t - (t_i + T_{ref})]}, & t \geq t_i + T_{ref} \end{cases} \\ &= H(t) + [e^{-r[t - (t_i + T_{ref})]} - 1] H(t - (t_i + T_{ref})). \end{aligned} \quad (\text{B.126})$$

The refractory firing rate of Eq. (B.124) is nonrealistic since the firing rate changes instantly from zero to  $r$  at  $t = t_i + T_{ref}$ . A more natural firing rate is given by:

$$\begin{aligned} r(t|t_i) &= r [1 - e^{-\lambda[t - (t_i + T_{ref})]}] H(t - (t_i + T_{ref})) \\ &= \begin{cases} 0, & t < t_i + T_{ref} \\ r [1 - e^{-\lambda[t - (t_i + T_{ref})]}], & t \geq t_i + T_{ref} \end{cases} \end{aligned} \quad (\text{B.127})$$

that grows smoothly from 0, at  $t = t_i + T_{ref}$ , to  $r$  with a time constant  $1/\lambda$ . Following (B.114) the survivor function for this firing rate is:

$$\begin{aligned} S(t|t_i) &= \begin{cases} 1, & t \leq t_i + T_{ref} \\ e^{-r[t - (t_i + T_{ref})]} e^{r[1 - e^{-\lambda[t - (t_i + T_{ref})]}]/\lambda}, & t > t_i + T_{ref} \end{cases} \\ &= e^{-r[t - (t_i + T_{ref})]} e^{r(t|t_i)/\lambda} H(t - (t_i + T_{ref})). \end{aligned} \quad (\text{B.128})$$

## B. Characterization of the Neural Response

From Eq. (B.118) gives for (B.127) and (B.128) the ISI density:

$$\begin{aligned}
 p_{ISI}(t|t_i) &= \begin{cases} 0, & t \leq t_i + T_{ref} \\ e^{-r[t-(t_i+T_{ref})]} e^{r[1-e^{-\lambda[t-(t_i+T_{ref})]}]/\lambda} r [1 - e^{-\lambda[t-(t_i+T_{ref})]}], & t > t_i + T_{ref} \end{cases} \\
 &= e^{-r[t-(t_i+T_{ref})]} e^{r(t|t_i)/\lambda} r(t|t_i) H(t - (t_i + T_{ref})).
 \end{aligned}
 \tag{B.129}$$

Other typical refraction functions include several sigmoid likewise functions.

## B.3. Spiking Mechanisms

Since many existing retina models produce as their outputs an estimate of the firing rate produced by the RGC, it is important to have a method to generate the spike trains from the firing rate. To generate the spike trains from the firing rate, we can follow the two distinct views already mentioned: we can generate a spike train from the firing rate considering that  $r(t)\Delta t$  is the probability of the neuron to fire a spike in the time interval  $\Delta t$ , or we can look to  $r(t)\Delta t$  as the number of spikes that we must generate in the time interval  $\Delta t$ . We have seen that the two interpretations given previously are equivalent for a small  $\Delta t$ , but they can lead to two different spike generation mechanisms.

### B.3.1. Generation of Poisson Spike Trains

The spike trains are very often considered to be well described by a Poisson process. We will see how to numerically generate a spike train from a variable firing rate  $r(t)$  by considering the spike train as a Poisson process. There are two common procedures for generating Poisson spike trains.

A first approach uses the probability of a spike occurring during a short time interval, presented in Eq. (B.22). For a Poisson process, whether it is inhomogeneous or homogeneous, the probability of generating a spike within the small time interval is

$$P(n = 1 \text{ in } [t, t + \Delta t]) = r(t) \cdot \Delta t. \tag{B.130}$$

By using this equation, a spike train can be generated by first dividing the time duration of the trial into a sequence of small intervals of width  $\Delta t$ , and also by sampling the firing rate  $r(t)$  in intervals of width  $\Delta t$ , composing the sequence  $r[n]$ . Then, a sequence of random numbers,  $x[n]$ , is generated with a uniform distribution in the interval  $[0, 1]$ , and the two sequences are compared. Whenever  $r[n]\Delta t \geq x[n]$ , a spike is placed at time bin  $n$ ; otherwise no spike is generated. This procedure is useful when  $r[n]\Delta t \ll 1$ , and the generated spike trains have a discrete time bin assigned to the spikes. The use of a time bin width of  $\Delta t = 1$  ms is usually enough [Dayan and Abbot, 2001].

The second approach used to generate Poisson spike trains with a constant firing rate is to choose the interspike intervals duration randomly from the exponential

density for the interspike interval distribution [Berry II and Meister, 1998], given by Eq. (B.84). Each successive spike is placed at a time equal to the previous one plus a random value drawn randomly from the interspike density:

$$t_{i+1} = t_i - \ln(x)/r , \quad (\text{B.131})$$

where  $x$  is a random number generated from a uniform distribution in the interval  $[0, 1]$ . However, in practice, when the obtained spike train is discretized to serve as the input to the next block of a processing system, this procedure becomes equivalent to the previous one.

To extend this last approach to generate a spike train with a time-varying firing rate, a spike train is generated by considering a constant maximum firing rate  $r_{\max}$

$$t_{i+1} = t_i - \ln(x)/r_{\max} , \quad (\text{B.132})$$

where  $r_{\max} > r(t)$ , for all  $t$ , and  $x$  is a random number with uniform distribution in the interval  $[0, 1]$ . Then, a thinning process is applied to the spike train generated at  $r_{\max}$  by keeping or deleting the spike posted at each  $t_i$ . The thinning is carried out by generating another random number  $x$  with uniform distribution for each  $i$ , and if  $r(t_i)/r_{\max} < x$  the spike at time  $t_i$  is removed; otherwise it is maintained [Dayan and Abbot, 2001].

The spiking mechanism of a real neuron breaks the assumption of independent firing, as stated by Eq. (B.63), namely during the absolute and relative refractory periods, where the probabilities of firing are null or very low, respectively; these are the principal features of neuronal firing not modeled by a Poisson model. To take into account the refractory effects in the Poisson model, the firing rate can be modulated by a function with zero value just after a spike is fired and during the absolute refractory period, and with an increasing exponential value tending to one during the relative refractory period. The absolute refractory period is variable, and it can be drawn randomly from a Gaussian distribution with a mean equal to the mean absolute refractory period and variance equal to the variance of the absolute refractory period [Dayan and Abbot, 2001].

The Poisson model can be further extended to include the neuron's refractory periods [Berry II and Meister, 1998], and to model bursting cells that appear predominantly in the visual cortex, where a neuron cell fires a burst of spikes in response to one event [Bair et al., 1994].

### B.3.2. Integrate-and-Fire Spike Generation

A straightforward interpretation of the firing rate as the number of spikes fired per second allows for the conversion of  $r(t)$  to  $\rho(t)$  by generating a spike train with a simple integrate-and-fire method. To code the firing rate  $r(t)$  into a set of spikes pulses, as displayed in Fig. B.27, let us suppose that we start integrating the firing

## B. Characterization of the Neural Response

rate,  $r(t)$ , at the time instant  $t_i$  such that

$$v(t) = \int_{t_i}^t r(t) dt . \quad (\text{B.133})$$

Whenever the activation signal  $v(t)$  crosses a predefined threshold  $\phi$  from below, that is when  $v(t) \geq \phi$ , a spike is fired and the integrator is reset to its rest value

$$\rho(t_i) = \begin{cases} \delta(t - t_i) & v(t_i) \geq 0 \\ 0 & \text{otherwise} \end{cases} . \quad (\text{B.134})$$

Whenever a spike is fired, the feedback loop in Fig. B.27 is activated and the integrator value is reset to zero ( $v(t + \delta t) = 0$ ), or, alternatively, a function of the threshold,  $f(\phi)$ , with a negative value can be added to it. For the case of a constant firing rate  $r$ , we have that  $v(t) = r \times (t - t_i)$ , and a spike is fired when  $v(t)$  reaches the threshold. Considering that the threshold is reached at the time instant  $t_{i+1}$ , such that  $v(t_{i+1}) = r(t_{i+1} - t_i) = \phi$ , we have that the time between two spikes is proportional to the inverse of the firing rate,  $t_{i+1} - t_i = \phi/r$ . Choosing the right value for the threshold (1 in the previous case), we can generate a sequence of spike trains whose spikes are generated according to the desired firing rate.

The computational implementation of an integrator can be accomplished by taking the relationship between  $r(t)$  and  $v(t)$  into account, which can be written as

$$r(t) = \frac{dv(t)}{dt} . \quad (\text{B.135})$$

By applying the Euler approximation to the first derivative,

$$\frac{dy(nT_s)}{dt} = \frac{y[(n+1)T_s] - y[nT_s]}{T_s} , \quad (\text{B.136})$$

where  $T_s$  is the sampling period used in the discretization process, to Eq. (B.135), we have that

$$r[nT_s] = \frac{v[(n+1)T_s] - v[nT_s]}{T_s} , \quad (\text{B.137})$$

which results in a recursive expression for the computation of  $v(t)$  at the time instants  $(n-1)T_s, nT_s, (n+1)T_s, \dots$ :

$$v[n+1] = v[n] + r[n] T_s \quad (\text{B.138})$$

where the argument dependence on  $T_s$  was dropped.

The comparator block in Fig. B.27 compares  $v[n]$  with the threshold, and whenever it crosses  $\phi$ , the output spike sequence is one ( $\hat{\rho}[n] = \delta[n] = 1$ ) and the integrator is reset to zero, or discharged by an amount equal to  $f(\phi)$ , so that  $v[n+1] = 0$ , or  $v[n+1] = v[n] - f(\phi)$ , respectively.

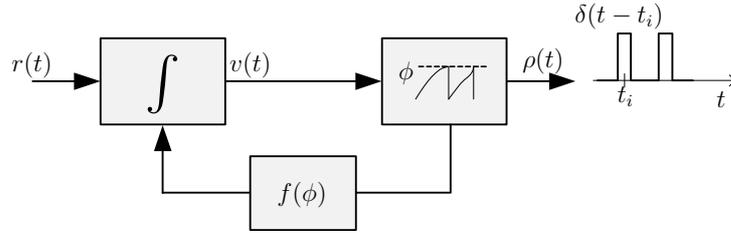


Figure B.27. Integrate-and-fire spike generation from firing rate.

## B.4. Integration in the Leaky Integrate-and-Fire Model

Despite quite straight, the result of the integration of the leaky integrate and fire neuron model appears wrongly in several publications and in various sites in the internet. We present next the detailed steps for the integration of the differential equation for the leaky integrate-and-fire (LIF) model.

The leaky integrate-and-fire model is depicted in Fig. B.28. The differential equation for the LIF has the form:

$$C_m \frac{dV_m(t)}{dt} + \frac{1}{R_m} (V_m(t) - E_r) = I_s(t) \quad (\text{B.139})$$

for a membrane resting potential equal to  $E_r$ . This equation states that if the input stimulus current is equal to zero ( $I_s(t) = 0$ ) for a long period of time the membrane potential stabilizes ( $dV_m(t)/dt = 0$ ), then the membrane potential goes to the rest potential:  $V_m(t) = E_r$  in Eq. (B.140).

The differential equation Eq. (B.139) can be written in the form:

$$\frac{dV_m(t)}{dt} + \frac{1}{\tau_m} V_m(t) = \frac{1}{\tau_m} E_r + \frac{1}{C_m} I_s(t), \quad (\text{B.140})$$

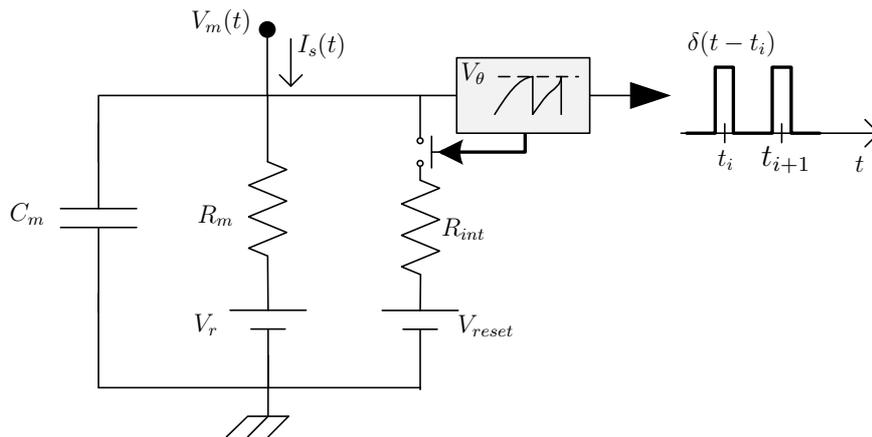


Figure B.28. The integrate-and-fire (I&F) model.

## B. Characterization of the Neural Response

by introducing the membrane time constant  $\tau_m = R_m C_m$ . By recalling the appropriate integration factor from differential equations theory, both sides of Eq. (B.140) are multiplied by  $e^{t/\tau_m}$  and the obtained equation is:

$$e^{t/\tau_m} \frac{dV_m(t)}{dt} + e^{t/\tau_m} \frac{V_m(t)}{\tau_m} = \frac{1}{\tau_m} E_r e^{t/\tau_m} + \frac{1}{C_m} e^{t/\tau_m} I_s(t). \quad (\text{B.141})$$

Integrating the left hand side of Eq. (B.141) between the initial time instant  $t_i$  and  $t$ , we get:

$$\begin{aligned} \int_{t_i}^t \left( e^{\frac{x}{\tau_m}} \frac{dV_m(x)}{dx} + \frac{e^{\frac{x}{\tau_m}}}{\tau_m} V_m(x) \right) dx &= e^{\frac{x}{\tau_m}} V_m(x) \Big|_{t_i}^t \\ &= e^{\frac{t}{\tau_m}} V_m(t) - e^{\frac{t_i}{\tau_m}} V_m(t_i). \end{aligned} \quad (\text{B.142})$$

where it was used the fact that the expression under the integral is the derivative of the product of the two functions:  $e^{\frac{t}{\tau_m}} V_m(t)$ . If the neuron has fired a spike at the time instant  $t_i$  then its membrane potential is at the reset value  $V_m(t_i) = V_{reset}$  (in general, this value is different from the membrane resting potential  $E_r$  that is often considered to be equal to zero:  $E_r = 0$ ). The integration of the right hand side of Eq. (B.141) gives:

$$\int_{t_i}^t \left( \frac{1}{\tau_m} E_r e^{x/\tau_m} + \frac{1}{C_m} e^{x/\tau_m} I_s(x) \right) dx = E_r (e^{t/\tau_m} - e^{t_i/\tau_m}) + \frac{1}{C_m} \int_{t_i}^t e^{\frac{x}{\tau_m}} I_s(x) dx, \quad t \geq t_i. \quad (\text{B.143})$$

Equating the results of Eq. (B.142) and Eq. (B.143), and by simplifying the terms, the membrane potential between two spikes assuming that a spike was fired at  $t = t_i$ , so that  $V_m(t_i) = V_{reset}$ , is given by

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{(t-t_i)}{\tau_m}} + \frac{1}{C_m} \int_{t_i}^t e^{-\frac{t-x}{\tau_m}} I_s(x) dx, \quad t \geq t_i, \quad (\text{B.144})$$

by making a change of variables Eq. (B.144) takes the common form

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{(t-t_i)}{\tau_m}} + \frac{1}{C_m} \int_0^{t-t_i} e^{-x/\tau_m} I_s(t-x) dx, \quad t \geq t_i. \quad (\text{B.145})$$

Using the Heaviside unit step function, defined by

$$H(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}, \quad (\text{B.146})$$

Eq. (B.144) can be rewritten as

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{(t-t_i)}{\tau_m}} + \frac{1}{C_m} \int_{-\infty}^{+\infty} e^{-x/\tau_m} H(x) I_s(t-x) H((t-t_i)-x) dx, \quad t \geq t_i. \quad (\text{B.147})$$

Recurring to the convolution operation Eq. (B.147) can be written as:

$$V_m(t) = E_r + \left[ e^{-t/\tau_m} H(t) \right] * \left[ (V_{reset} - E_r) \delta(t - t_i) + \frac{1}{C_m} I_s(t) H(t - t_i) \right]. \quad (\text{B.148})$$

Alternatively, by introducing a change of variables Eq. (B.144) can also be written as:

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) + \frac{1}{C_m} \int_{-\infty}^{+\infty} e^{-\frac{(t-t_i)-x}{\tau_m}} H((t - t_i) - x) I_s(x + t_i) H(x) dx, \quad t \geq t_i. \quad (\text{B.149})$$

where the integral in Eq. (B.149) corresponds to a convolution operation, so that the final expression for the solution of Eq. (B.140) can be written as

$$V_m(t) = E_r + (V_{reset} - E_r) e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) + \frac{1}{C_m} \left[ e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) \right] * [I_s(t + t_i) H(t)], \quad (\text{B.150})$$

for a spike occurring at time instant  $t_i$  and for  $t \geq t_i$ . The left shift of the input stimulus current in Eq. (B.150) reflects the fact that only the current after the spike occurrence  $t_i$  contributes to the membrane potential, as the integral in Eq. (B.144) states.

If the resting membrane potential is equal to the reset potential after a spike,  $E_r = V_{reset}$ , Eq. (B.150) becomes:

$$V_m(t) = E_r + \frac{1}{C_m} \left[ e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) \right] * [I_s(t + t_i) H(t)], \quad t \geq t_i, \quad (\text{B.151})$$

meaning that the neuron membrane stays at the rest potential in the absence of an input current. If the resting potential is zero Eq. (B.150) becomes:

$$\begin{aligned} V_m(t) &= V_{reset} e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) + \frac{1}{C_m} \left[ e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) \right] * [I_s(t + t_i) H(t)] \\ &= \left[ e^{-\frac{t-t_i}{\tau_m}} H(t - t_i) \right] * \left[ V_{reset} \delta(t) + \frac{1}{C_m} I_s(t + t_i) H(t) \right], \quad t \geq t_i, \end{aligned} \quad (\text{B.152})$$

However, extra care must be taken for the value given to  $V_{reset}$  since in the absence of an external stimulus, the membrane potential will always decay to zero exponentially.

The expression of Eq. (B.148) states that the neuron membrane potential is at the rest state,  $V_m(t) = E_r$ , until an input stimulus excites the neuron so that the membrane potential is incremented by the convolution of the input stimulus current,  $I_s(t)$ , with the low-pass filter,  $e^{-t/\tau_m} H(t)$ , and when a spike is fired the membrane potential is reset to  $V_m(t_i) = V_{reset}$  that decays exponentially to the resting potential.



# C

## Bioelectronic Vision Systems

THE ultimate goal for researching computational models for the retina is to develop bioelectronic vision systems to aid patients suffering from blindness. In the last years, research projects and consortia have been setup by joining people from different fields, such as biomedical engineering, computer science and electrical engineering. This appendix provides a brief overview of the most important projects and prostheses, which are categorized according to the two main types of bioelectronic vision: retinal prostheses, based on epiretinal and subretinal implants, and cortical visual prostheses [Javaheri et al., 2006]. It also describes concisely a prototype of a bioelectronic vision system designed in the scope of the Cortical Visual Neuroprosthesis for the Blind (CORTIVIS) European project [Project CORTIVIS, 2006].

The book [Martins and Sousa, 2009] develops this subject further and describes with more detail the process of design and implementation of a bioelectronic vision system.

### C.1. Bioelectronic Vision Labs

Several research groups, frequently involving multidisciplinary teams due to the subject, have been promoting projects to develop and demonstrate the feasibility of artificial vision systems. Around the world there are under a dozen laboratories devoting significant resources and attention to the design and development of visual neuroprosthesis. The huge challenge of artificially restoring vision to the blind poses engineering and biological problems hard to overcome and also requires the process of clinical human testing. This is an enormous effort for those who almost always in an academic environment struggle to push this research area. Figure C.1 dis-

C. Bioelectronic Vision Systems

plays a world map with the more relevant laboratories involved in the development of bioelectronic vision systems.

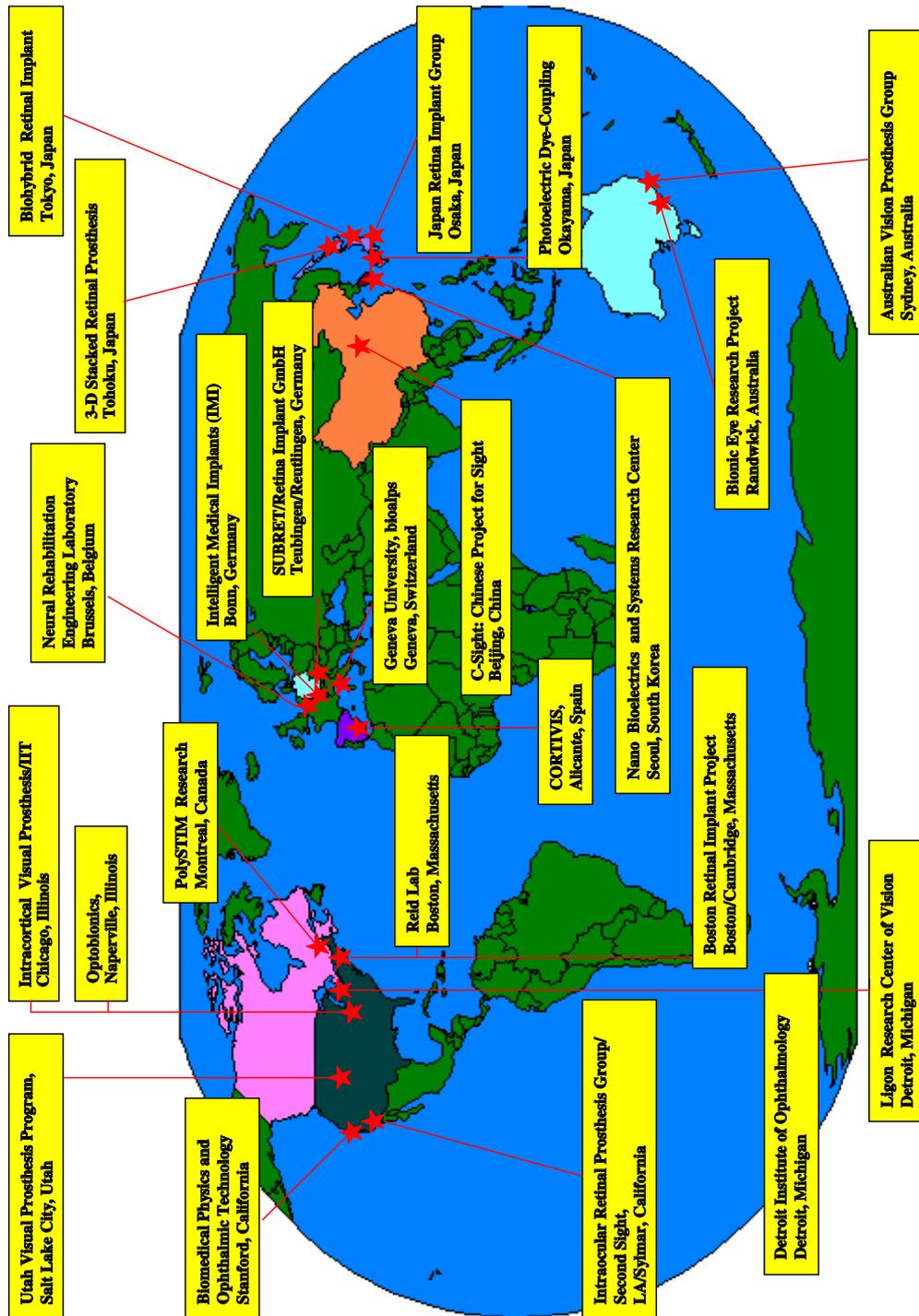


Figure C.1. World map of vision prosthesis research labs (adapted from [Hessburg and Rizzo III, 2007]).

## C.2. Components of a Bioelectronic Vision System

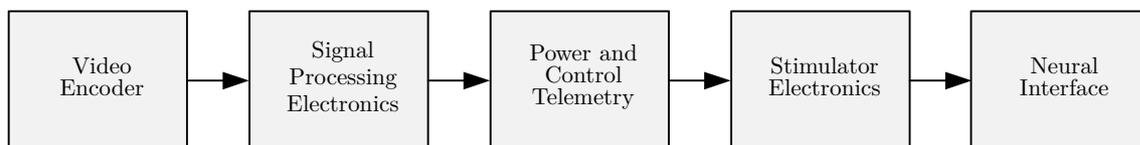
The foremost component of the vision system is the eye. The eye is responsible for gathering and transducing light energy to neural electrical signals, in the form of electrical impulses, that are posteriorly conducted to the brain for further information extraction. Roughly speaking, the eye is composed by an optical system that focuses light on an neuronal tissue – the retina. In the retina the light patterns are encoded into electrical signals where the neuronal processing starts, so that the retina can be seen as an extension of the brain. The visual neuroprostheses must model the optical eye system, which does not pose any major technological difficulties nowadays, and the neuronal processing occurring in the retina, where the true challenge lies, and, in case of a cortical prosthesis, mapping the electrical stimulus signals appropriately in the brain.

Bioelectronic vision systems can be classified in two main classes of visual neuroprostheses: *i*) retinal neuroprostheses, and *ii*) cortical neuroprosthesis. Retina neuroprostheses are suitable when the front end of the retina is functioning properly while a cortical neuroprosthesis is the last hope when the retina is not functioning, including the optic nerve, and only the superior brain vision centers remain intact (see Fig. 1.1). The cortical neuroprosthesis interfaces directly with the visual processing center in the brain, known as the visual cortex area V1, of profoundly blind people, whose optical neurotransmitters are irreversibly damaged, substituting, in this way, their entire vision system. The concept and components of a bioelectronic vision system supported on a complete visual neuroprosthesis that interfaces directly with the brain is depicted in Fig. C.4.

Figure C.2 displays the main blocks of a bioelectronic vision system. It includes a set of components which, depending on the type of visual neuroprosthesis, can be biological structures or their counterpart electronic circuits.

For example, an external video acquisition device is usually required for capturing the visual field image and to convert the light patterns into electrical signals. However, when an array of stimulation electrodes is placed in the subretinal space, the light photons of the image falling on the retina are converted into electrical currents by an array of microphotodiodes that directly replace the function of the damaged photoreceptors cells.

The digital signal processing system transforms the visual space image into a set of discrete signals, according to the retina model and taking into account the visuotopic organization of the target structure (the retina or the cerebral cortex, depending on the prosthesis type). A module to transmit power and control signals to the implanted



**Figure C.2.** Components of a visual neuroprosthesis.

electronics is usually required [Piedade et al., 2005; Santos et al., 2006]. This module provides energy and control signals to the stimulator that interfaces to the nervous system to induce the perception of phosphenes, which is an entoptic phenomenon characterized by the sensation of seeing light. The electronics, based on integrated circuits and (micro)electrodes, has the ultimate goal of replacing the function of the counterpart biological elements.

Image acquisition is a common task in engineering. For cortical or optic nerve neuroprostheses, a generic small and full functional digital camera is suited for a vision prosthesis in terms of dynamic range, sensitivity and depth of field and, as also important, in terms of aesthetic. In a retinal neuroprosthesis the image encoder can be integrated into the neural interface, and lay at the plane of the retina, with the advantage that the eye optics can be used to project the image in the encoder.

In the signal processing block the biggest challenge is the mapping of the visual space into the visuotopic organization of the target structure, particularly the visual cortex. This is a somewhat complicated task due to the uniqueness of this map among individuals, and because it is conformal only in low resolution; for high spatial resolutions this mapping seems to be locally random. Therefore, parameterizable models have to be developed for implementing this module and properly stimulating individuals. This is a somewhat more complicated task due to the plasticity of the visual pathways and the different possible combinations between electrodes and phosphenes elicited. Based on the developed models, the electronics of this module transforms the image into a discrete set of signals that drive the stimulators. To adapt the intensity of the incoming light signals into the range level of the neurons being stimulated an AGC can perform similarly to what is done by the photoreceptors. The first two components of the visual neuroprosthesis are susceptible of being included into a single device attached to a set of eyeglasses, and the elements of the visual neuroprosthesis described in the sequel are likely to be located inside the patient.

The information from the visual scene must be conducted to the implant, and the ways to do that are different for a retinal neuroprosthesis and a cortical neuroprosthesis. There are two main ways to transmit signals through the skin: percutaneous connectors [Dobelle, 2000] or using radio frequency (RF) telemetry [Piedade et al., 2005]. On one hand, the percutaneous connectors have the advantage of being more robustness and obviate the use of multiplexors, but on the other hand they have the drawback of being a source of infections. On the other hand, a radio frequency link have the challenge of communicating both power and control signals and have to be bidirectional. In a retinal neuroprosthesis, a percutaneous connection would have to pass to the outside of the eye through the sclera. For the case of a retinal neuroprosthesis it can also be used a laser to transmit power and information to the implanted circuits used to stimulate the target cells [Weiland et al., 2005]. All types of connections have a series of constraints like bandwidth, which increases with the number of electrodes in the neural interface, and also the transmitted power must be limited in order not to cause damage to the tissues by heating.

The next component in the chain of Fig. C.2 is the neural stimulator that must also be capable of exciting multiple electrodes at the same time in order to evoke consistent

phosphenes. It receives power and data, namely through a telemetry sub-system, and must be capable of controlling the amount of power delivered, to avoid the damage to neighboring tissues, and should also be able to circumvent malfunctioning electrodes, for example. The implementation of this module on chips presents a trade-off between the processing capabilities and power consumption: increasing the process capabilities of the chip diminishes the required link bandwidth but augments power consumption and potential failures [Maynard, 2001]. The most adequate technology, due to size and power consumption, is digital VLSI circuits [Warren and Normann, 2003].

The last element in a vision prosthesis is the interface with the nervous system. The neural interface establishes the bridge between the nervous system and the external electronics. It makes the transduction between the electrical currents generated by the electronic device into ionic currents that flow inside the human body. For the retinal neuroprosthesis the neural interface options range from silicon chips to specific developed ceramic materials [Wu, 2006]. For cortical interfaces the oxidized iridium is a candidate material because it has shown a good biocompatibility, and a good electronic to ionic current transducer. There are some more compatibility issues related with the neural interface that must be taken into account [Warren and Normann, 2003].

In conclusion, before reaching the ganglion cells layer in the retina or the visual cortex, the visual signal has already been subjected to a series of processing stages. When the interface to the visual stimulus is made at the level of the ganglion cell layer, as for the case of an epiretinal neuroprosthesis, the information at the output must be identical to the one produced by a healthy retina; the transformation of the visual space to the retinotopic space is done by modeling the neural processing of the retina. For the case of a cortical neuroprosthesis the signal processing occurring along the visual pathway should be in order to generate the adequate stimulus for the neural interface.

## C.3. Types of Visual Prosthesis

The efforts for conveying some kind of vision to profoundly blind people already have some history [Rizzo III and Wyatt, 1997]. Throughout the world, several research groups and consortia dedicate their efforts to design a vision prosthesis capable of conveying to the blind people some kind of vision.

Some non-conventional approaches and electronic devices have been proposed to convey vision to visually impaired people. In some of these devices the visual information is converted to auditory [Arno et al., 1999] or tactile signals that are afterwards communicated to the brain. A somewhat curious device is one that consists of a flexible cable with a matrix of electrodes at the end that is placed against the patient tongue and a pattern of electrical impulses stimulates its sensitive nerves [Weiss, 2001].

Bioelectronic vision systems are supported on visual neuroprostheses that inter-

face with the following neural structures: *i*) the photoreceptors layer of the retina; *ii*) the ganglion cells layer of the retina; *iii*) the optic nerve, and *iv*) the visual cortex [Warren and Normann, 2003]. Thus, the prosthesis that still uses some part of the human vision system are of three types: retinal neuroprosthesis and optic nerve neuroprosthesis at the eye level; and cortical neuroprosthesis, at the brain level. The retinal neuroprostheses use the remaining functioning parts of the retina to transmit the visual signals to the brain, the optic nerve neuroprostheses stimulate what is left of the optic nerve, whereas the aim with the cortical neuroprostheses is to inject visual signals directly into the visual cortex.

The type of approach used in a visual prosthesis is related with the type of blindness an individual suffers from. In one extreme, the blindness can derive from damages at superior retina layers, as a consequence of early stages of diseases like the retinitis pigmentosa, and from some kinds of macular degeneration, where the principal injuries occur at the photoreceptor layer but the ganglion cells layer remains mostly intact, which allows their (re)usage. In this type of blindness a retinal neuroprosthesis can be used. In the other extreme there is what is called profound blindness, where the ganglion cells layers and the optic nerve are irreversibly injured and incapable of transmitting any kind of nervous signals. In this case the reestablishment of some sort of vision can only be done by the circumvention of the optic nerve and the remaining option is to stimulate the visual cortex directly with an electronic device. This is where cortical neuroprosthesis come into play.

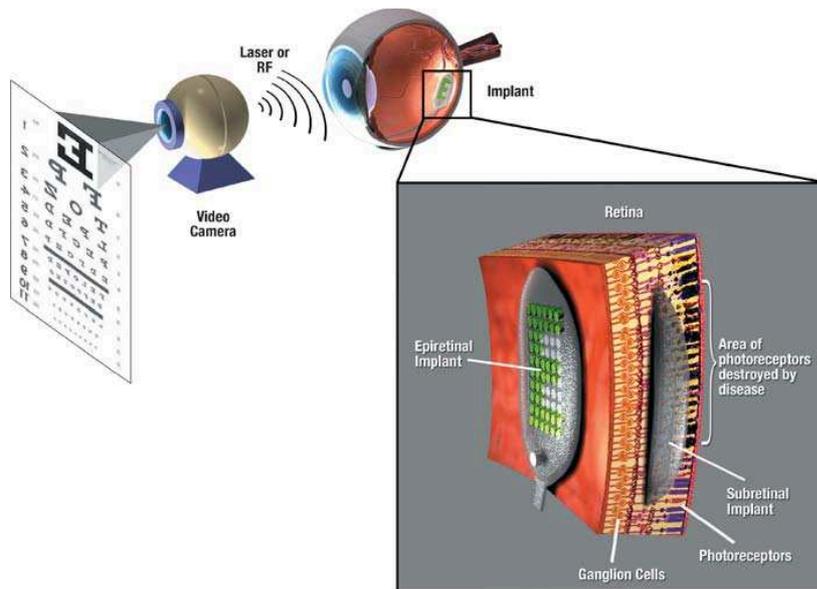
An intermediate situation is the utilization of the optic nerve to conduct the visual signals to the brain. Some of the retinal diseases leave a significant number of intact ganglion cells dispersed along the retina, whose axons concentrate at the optic nerve. The strategy is to induce the visual signals in this great bunch of ganglion cells at the optic nerve expecting to excite conductive axons, which is a difficult task since the axons are forming the optic nerve are arbitrarily organized and change from individual to individual.

In the sequel the main characteristics of the prostheses that interface with the retina and with the visual cortex are presented.

#### C.3.1. Retinal Neuroprosthesis

There are the two kinds of retinal implants as depicted in Fig. C.3. In a *subretinal* implant the prosthesis is placed between the pigment epithelial layer and the outer layer of the retina which contains the photoreceptors cells. Whereas the *epiretinal* device is placed directly against the ganglion cells and their axons layer, bypassing the rods and cones, stimulating directly the inner retina, being a more invasive technique to the eye.

Unlike the subretinal implant, the *epiretinal* implant does not use any remaining network of the retina for information processing. Thus, the epiretinal sensor has to encode visual information as trains of electrical impulses that are then conveyed by an electrode array directly into the ganglion cells axons, which unite to form the optic nerve. This spatiotemporal stimulation pattern of electrical impulses has to represent



**Figure C.3.** Epiretinal and a subretinal prosthesis [Weiland et al., 2005].

the visual information in such a way that it is understood by the brain visual cortex. On the other hand, the information-transfer characteristics of the epiretinal implant are more amenable to external control while the subretinal implant needs intact original optics [Zrenner, 2002]. A relevant example of the development and test of a subretinal implant is reported in [Chow et al., 2004], and an example of a epiretinal implant can be found in [Rizzo III and Wyatt, 1997; Humayun et al., 1999].

Nevertheless, retinal neuroprostheses require the presence of viable cells in the inner retina. Therefore, diseases limited primarily to the outer retina are potentially treatable with a retinal neuroprosthesis. The references [Margalit et al., 2002] and [Weiland et al., 2005] present an extended overview of retinal neuroprostheses.

One of these prostheses, termed "Bionic Eye", employs a new ceramic material to substitute the retina's photoreceptors, that act as a optic detector transducing light into electrical impulses, by means of the photo-ferroelectric effect [Wu, 2006]. This material is directly implanted in the patient eye and, under optic illumination, generates a photo-current that excites directly the retina ganglion cells. It seems to be bio-compatible and it can be used in retinal dystrophies, where the optic nerve and retinal ganglia are intact, such as in the Retinitis Pigmentosa, to directly stimulate retinal ganglia.

Another example of a retinal neuroprosthesis is the artificial silicon retina (ASR) microchip [Chow et al., 2004], which uses silicon technology [Optobionics Corporation, 2012]. The ASR microchip is a silicon-based 2mm in diameter device that contains approximately 5000 microelectrode-tipped microphotodiodes and is powered by the incident light. Eyes of patients with retinitis pigmentosa were implanted with the ASR microchip. Patients did not show signs of implant rejection or infection and for all them a vision function improvement have occurred.

There are other examples of subretinal implants, namely the one that consists of a chip ( $3 \times 3 \times 0.1\text{mm}$ , 1500 microphotodiodes, amplifiers and electrodes of  $50 \times 50$  micrometers, spaced 70 micrometers) and a  $4 \times 4$  array of identical electrodes, spaced 280 micrometers, for direct stimulation. They were chronically implanted next to the foveal rim of two blind retinitis pigmentosa patients [Zrenner et al., 2006]. The implant was removed in one patient after 4 weeks but the other decided to keep the implant. Patients reported small, yellowish or greyish phosphenes for individual electrode stimulation and they were able to differentiate spatial patterns such as lines, angles or bright squares.

There is a second type of approach at the low level of the visual system whose functioning principle is to stimulate directly the optic nerve, obviating the utilization of the retina layers, possibly damaged. These devices are implanted around the optic nerve stimulating electrically its fibers [Brélen et al., 2005]. This type of prosthesis also needs a way to model the function of the retina.

### C.3.2. Cortical Visual Neuroprosthesis

Cortical visual neuroprostheses are bioelectronic systems that use the visual cortex in the brain as the interface between the electronics components and the biological vision pathway. A last resource for blind individuals that can not benefit from a retinal neuroprosthesis is the direct stimulation of the visual cortex. This is the last hope when the retina is not functioning at all, including the optic nerve, because retinal neuroprostheses rely on it to transmit electrical signals from the eye to the visual cortex. The visual cortex is a brain vision center accessible and it is well positioned to be stimulated. This kind of neuroprostheses include all the electronic components presented in Fig. C.2 to substitute the biological counterparts, shown in Fig. 1.1.

The first permanent device developed and applied for chronic stimulation of neural tissue was accomplished in 1968 [Brindley and Lewin, 1968]. This device had 80 electrodes, each with its own controlling unit (receiver). Using this system, it was demonstrated the feasibility of a permanent cortical vision neuroprosthesis by observing that the electrical stimulation of the occipital lobe of the human cortex causes a subject to perceive phosphenes.

Despite this initial success [Dobelle and Mladejovsky, 1974] the surface electrodes have a number of drawbacks: currents necessary to induce a phosphene are relatively high (of the order of milliamperes), consequently the distance between the electrodes have to be considerable in order to minimize their interactions due to current spread, but that degrade the spatial resolution; moreover, current injection can produce short and long term complications depending upon the level of current injected [Agnew and McCreery, 1990].

Two main groups have been working during the 1990's towards cortical vision prosthesis. One was based at the National Institutes of Health (NIH) in Washington, D.C., and the other at the John Moran Laboratories in Applied Vision and Neural Sciences at the University of Utah. Both groups have tried to overcome the

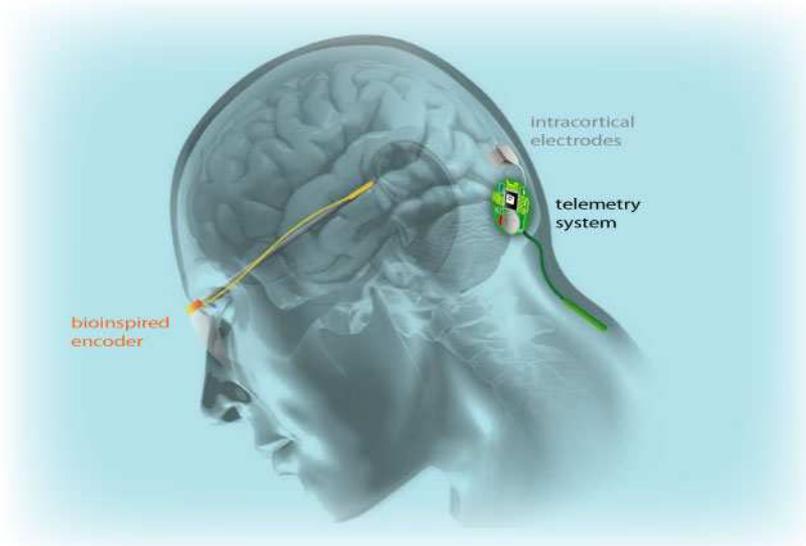
problems referred above by employing penetrating microelectrodes instead of using surface electrodes on the visual cortex. One example of these microelectrode arrays made available by the technologies in the semiconductor manufacturing [Maynard, 2001], was developed in the Utah University and is known as the Utah Electrode Array:  $10 \times 10$  microelectrodes, each with 1.0 – 1.5mm long, disposed in a square grid contained in a package with dimensions  $4.2 \times 4.2$ mm [Maynard et al., 1997; Normann et al., 1999]. The silicon micromachining and micromanufacturing technologies allow the fabrication of small arrays with a large number of microelectrodes capable of stimulating only the neurons nearest to the electrode and with a small amount of current (in the order of dozens of  $\mu A$ ). The major concerns with the microstimulation are related with biocompatibility and long term functioning of the inserted microelectrode array.

Research is going on to design and develop cortical visual neuroprostheses through intracortical stimulation, but no one of these prostheses has been permanently applied for chronic stimulation. Besides the technical and pathological issues, the consequences of direct stimulation of the brain and implications in terms of personal identity and personhood are important concerns [Mathews, 2011], notwithstanding that the stimulation of the visual cortex should mainly affect vision.

The European research project CORTIVIS aimed to design and develop a complete visual neuroprosthesis to restore useful vision sense to profoundly blind people [Project CORTIVIS, 2006]. It performs intracortical microstimulation through one or more Utah Electrode Array implanted (see Fig. C.9a) into the primary visual cortex. The system is composed by a primary unit located outside the body and a secondary unit, implanted inside the skull that communicate with each other using wireless communication technology. Figure C.4 displays a conceptual sketch of the prototype bioelectronic vision system. The book [Martins and Sousa, 2009] contains a more detailed description about the design and implementation of the prosthesis prototype.

## C.4. Retinal Prostheses

People with retinal degeneration diseases, such as retinitis pigmentosa or macular degeneration, lose their sight as the cells in the retina that normally sense light deteriorate. Retinal implants can take over for these lost cells, converting light into neural signals that are then interpreted by the brain. Simple versions of these devices have already been tested in humans, giving patients the ability to detect light and to distinguish simple objects. The epiretinal approach has the advantage of leaving the retina intact by placing the implant in the vitreous cavity, a natural fluid-filled space, and stimulating the ganglion cell layer. On the contrary, the subretinal implantation of a retinal prosthesis only replaces potentially damaged photoreceptors with a microstimulator by taking advantage of the remaining healthy cells of the retina. This type of implant requires detaching the retina in a more complex surgery, and the retina is a fragile tissue only 0.25 mm thick.



**Figure C.4.** The Cortivis intracortical visual prosthesis [Project CORTIVIS, 2006].

### C.4.1. Epiretinal implants

The epiretinal approach to the retinal prosthesis involves the capture and digitalization of images with an external device, such as a video camera. These images are transformed into patterns of electrical signals used to excite remaining viable inner retinal neurons. Power delivery and data telemetry sub-systems are required to drive this process. Several research groups have designed epiretinal implants based on intraocular and external elements with different characteristics: *i*) the Intraocular Retinal Prosthesis (IRP), developed by Mark Humayun and colleagues at the University of Southern California [Humayun et al., 1996]; *ii*) Joseph Rizzo and John Wyatt developed an epiretinal prosthesis at the Harvard Medical School and the Massachusetts Institute of Technology [Wyatt and Rizzo, 1996]; and *iii*) the Retina Implant (RI) was developed by Rolf Eckmiller with a consortium of 14 expert groups in Germany [Eckmiller, 1997].

The IRP uses an external camera to acquire the image, which is passed through a visual processing unit to generate the information that is coded in the form of electrical pulses patterns. These patterns are transmitted into the eye by an inductive link telemetry system, composed of magnetic coils implanted in the temporal skull. The electrical stimulation pattern is delivered, via a transscleral (across eye wall) cable, to an array of microelectrodes attached to the inner retinal surface to stimulate viable inner retina neurons (intraocular part of the prosthesis). The first array stimulates inner retinal neurons through 16 platinum microelectrodes, ranging in size from 250  $\mu\text{m}$  to 500  $\mu\text{m}$ . Recent technological advances allowed the number of microelectrodes to increase to 60 and to incorporate microelectromechanical systems (MEMS) in order to achieve a better fit of a planar electrode array onto a curved inner retinal surface.

Clinical trials testing chronic long-term implantation of the IRP have been performed in 6 patients implanted with 16-electrode arrays manufactured by the Second Sight Medical Products [Javaheri et al., 2006]. Reports of these trials show that patients are able to distinguish the direction of motion; they also have the ability to discriminate between percepts created by different electrodes, and the retinotopic organization is not lost when a patient loses sight [Lakhanpal et al., 2003].

The retina implant team (EPI-RET) has developed the RI and founded the Intelligent Implants company in 1998. The implant is composed of a Retina Encoder (RE), a wireless Signal-and-Energy transmission system (SE), and a Retina Stimulator (RS). The RE, which is located outside the eye, consists of a photosensor array with around 100,000 pixels at the input. The RE approximates the typical receptive field properties of retinal ganglion cells by means of hundreds to thousands of tunable spatiotemporal filters. This cell output is encoded and transmitted via a wireless, signal and energy transmission system (electromagnetic and/or optoelectronic) to the implanted RS. The RS is a ring-shaped, soft microcontact foil centered about the fovea that is fixed to the epiretinal surface to be in contact with a sufficient number of retinal ganglion cells to elicit electrical spikes. The RE also provides a perception-based interaction between the RE and the human subject in order to tune the various receptive fields' filter properties with information "expected" by the central visual system. Eckmiller and his group have been testing the RE/RS mainly in animals [Walter et al., 1999]. They have chosen to focus their efforts on understanding the information processing requirements of both the retinal prosthesis and the brain in terms of a dialogue-based RE tuning [Eckmiller et al., 1999]. Clinical trials have been primarily focused on testing of the RI implant and dialogue-based RE tuning. Rizzo and Wyatt epiretinal implants also consist of independent intraocular and extraocular units, without batteries implanted within the body and no wires penetrating the eye. The extraocular unit is composed of a tiny charged couple device (CCD) camera, a signal processing unit and a fixed-direction laser; all mounted on a pair of glasses. The output of the CCD camera/signal processing unit modulates the amplitude of the laser beam (820 nm wavelength). The extraocular unit runs with replaceable batteries to be kept in the patient's pocket. The intraocular unit consists of a photodiode array and a stimulator chip sandwiched around a flexible thick polyamide strip that supports the electrodes. The photodiodes are used to capture the processed signal from a laser pulse emitted from the glasses. The stimulator chip then delivers this information to the microelectrode array on the epiretinal surface of the eye. Rizzo and Wyatt have applied implants in 5 blind patients with Retinitis Pigmentosa, and 1 normal-sighted patient who was scheduled for enucleation due to orbital cancer. Three different types of electrode arrays were tested, with different numbers of electrodes, size, and spacing of the peripheral electrodes. The reported results from short-term studies were not conclusive. By stimulating a single electrode above a threshold level, multiple phosphenes were often perceived by the blind subjects. However, by simultaneously stimulating multiple electrodes it was not possible to perceive even the simplest visual pattern, neither by blind nor by normal-sighted patients. Due to the problems that Rizzo and Wyatt found

in epiretinal stimulation, they have abandoned the epiretinal approach and are now developing a subretinal approach, which will be discussed in the next section.

### C.4.2. Subretinal implants

In the subretinal approach, a microphotodiode array has to be implanted between the bipolar cell layer and the retinal pigment epithelium. The main advantage of this type of implant is that the microphotodiodes of the subretinal prosthesis directly replace the functions of the damaged photoreceptor cells, while it is assumed that the retina's remaining intact neural network is still capable of processing the generated electrical signals.

Subretinal implants were proposed by: *i*) Alan and Vincent Chow, who have developed the Artificial Silicon Retina (ASR) microchip [Chow et al., 2004]; *ii*) Eberhart Zrenner and a consortium of research universities using a MicroPhotoDiode Array (MPDA) [Zrenner et al., 1998]; and, more recently, *iii*) Rizzo and Wyatt [Wyatt and Rizzo, 2006], who have developed a third type of subretinal prosthesis.

Alan and Vincent Chow of Optobionics Corp believed that a subretinal implant could function as a simple solar cell; therefore, their ASR Microchip was powered entirely by light entering the eye [Peyman et al., 1998]. With a diameter of two millimeters, the ASR contains approximately 5000 microelectrode-tipped microphotodiodes used to convert incident light into electrical signals. These electrical impulses, in turn, stimulate any viable retinal neurons, which then process and send these signals to the visual processing centers in the brain via the optic nerve. Ophthalmologist Alan Chow and his team at Optobionics tried this approach in people in the year 2000. They implanted a silicon disk with 5000 microscopic solar cells, or photodiodes in one eye in 30 people. Most of these implant recipients have reported moderate to significant improvements in at least one aspect of visual function, including light sensitivity, size of visual field, visual acuity, or movement or color perception. Many doubts arise about these results, because the amount of current needed to actively stimulate retina ganglion cells is not in the range of the current obtained from a photodiode [Wickelgren, 2006]. In fact, Chow abandoned the concept that the ASR Microchip is effective as a prosthetic device, and now he suggests that the insufficient levels of current delivered from the implant may have a therapeutic as well as neuroprotective effect on otherwise dying retinal photoreceptors. Therefore, this device is best classified as a therapeutic device, rather than as a true retinal prosthesis.

The SubRet consortium has designed and fabricated various types of ultrathin and flexible MPDA devices, as well as CMOS-based chips with different pixel sizes and electrode configurations. The first generation of developed MPDAs, similar to the original work of Chow, consists of  $20 \times 20 \mu\text{m}^2$  pixels on a 3 mm diameter crystalline silicon chip. After this first generation, this team developed a special deposition technique to produce very thin and flexible MPDAs that fit to the curvature of the eyeball. They used amorphous hydrogenated silicon which has a light absorption 20-30 times higher than crystalline silicon. Biomedical experiments conducted by Zrenner and his team made clear that a purely photovoltaic operation is not effec-

tive, so additional energy has to be input by near infrared or radio frequency power transmission. A charge transfer of 100-1000  $\mu\text{C}/\text{cm}^2$  within 1 ms is required for provoking a retina response, while light exposure at a retina location does not exceed 0.1  $\mu\text{W}/\text{cm}^2$ . Prototypes of their subretinal devices, with external power source to supply energy for the subretinal implant, have been proposed, namely using near infrared radiation.

By implanting their prosthesis in rabbits, cats, and pigs, they attempted to detect electrically stimulated activity in the visual cortex as a result of retinal stimulation, as well as investigate the long-term biocompatibility and stability of these implants in the subretinal space. In nearly half of the tested animals, cortical evoked potentials were recorded with chronically implanted epidural electrodes during stimulation with light flashes, as well as during electrical stimulation of the subretinal space.

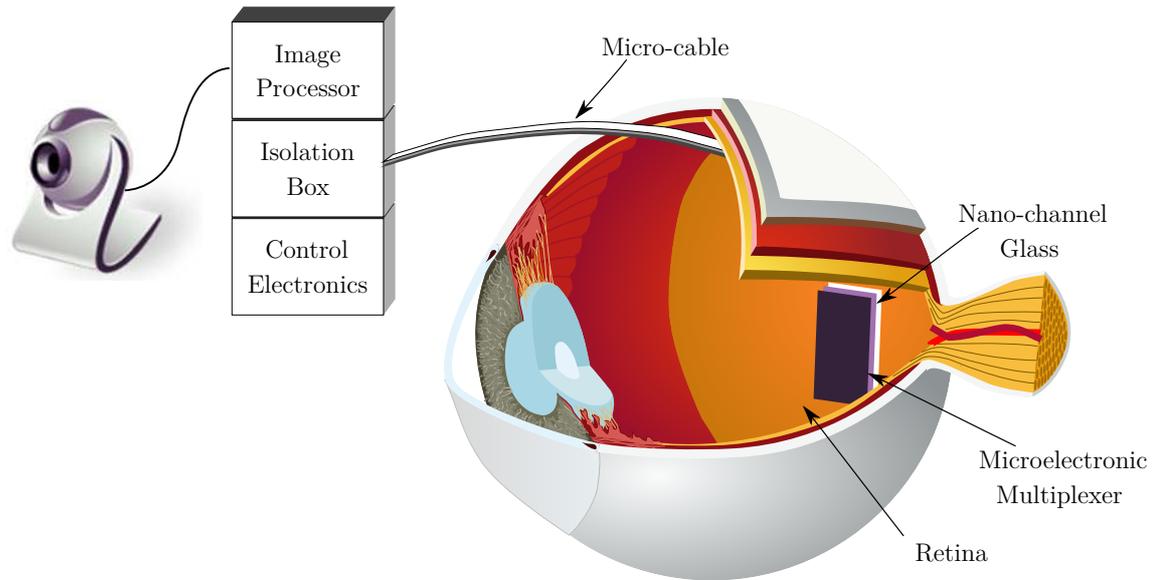
As stated in the last section, Rizzo and Wyatt decided to start working on subretinal implants. Although the Boston Retinal Implant Project is in the early stages of development, they have reported biocompatibility studies as well as the evaluation of surgical methods to implant their device in rabbits, pigs, and dogs. Minimally invasive surgical techniques have been tested, by using a posterior, ab externo approach to implant the prosthesis and to insert the stimulating electrode array in the subretinal space.

The main advantage of subretinal implants in comparison to the epiretinal approach is that the microphotodiodes replace directly the functions of the damaged photoreceptor cells, while the remaining intact neural network of the retina retina is still capable of processing electrical signals. However, the closer proximity of the subretinal prosthesis to inner retinal neurons predisposes the contacted retinal neurons to an increased likelihood of thermal injury resulting from heat dissipation. Together with the lack of external sources of energy for the microphotodiodes, this is one of the main problems of invasive neuroprostheses based on subretinal implants.

Open issues are the long-term biocompatibility of microelectronics in the saline environment of the eye, both in terms of hermetic packaging of the microfabricated electrode arrays, and the heat generated and dissipated with its use. Also included in these biocompatibility issues is the unknown effect of chronic electrical stimulation on the retina.

## **C.5. Retinal Bioelectronic Vision System Design**

The design and implementation of bioelectronic vision systems for epiretinal and subretinal prostheses have been a topic of research during the last few years. A number of issues have to be addressed in order to design and implement bioelectronic vision systems based on retinal neuroprostheses. One important aspect is the interface between the electrode array and the retina, namely regarding biocompatibility and the requirement to conform with the spherical, concave surface of the retina. It is also necessary to supply power to permanent implants through a wireless system, since no wires are expected to go through the eye wall. Moreover, the electrical stimulation's



**Figure C.5.** Intraocular test device.

pulse rate and the instant of occurrence need to be determined in a general way, but may need to be individually tuned for each patient.

Some numbers about visual acuity that can be useful to consider for designing bioelectronic vision systems are: normal visual acuity (20/20) corresponds to angular separation of lines about 1 min of arc or spatial separation on the retina of about  $10\ \mu\text{m}$ ; applying the Nyquist sampling frequency, for such visual acuity the maximum pixel size is  $5\ \mu\text{m}$ . Sufficient acuity (20/100) for reading with some visual aid requires pixels smaller than  $25\ \mu\text{m}$ . To achieve a useful reading performance, it has been estimated that about 600 pixels is the minimum for resolving images in the central field [Margalit et al., 2002].

The architectures of an intraocular epiretinal prosthesis test device [Scribner et al., 2001], and of a proposed system that can be used for both epiretinal and subretinal stimulation [Loudin et al., 2007] are overviewed. This latter system is based on a photodiode array implant, and video frames are processed and conveyed onto the retinal implant by a head-mounted near-to-eye projection system operating at near-infrared wavelengths.

The intraocular test device will enable short-term (less than an hour) human experiments to study issues related with interfacing electrode arrays with retinal tissue. The design combines two technologies: *i*) electrode arrays fabricated from NanoChannel Glass (NCG), and *ii*) Infrared Focal Plane Array (IRFPA) multiplexers. Figure C.5 shows an IRP test device to be used in acute human experiments. Ophthalmologists use standard retinal surgical techniques in an operating room environment to perform the experimental procedure. Local anesthesia is administered so that the patient is conscious during the procedure.

NGC uses fiber optic fabrication techniques to produce thin wafers of glass with

millions of very small channels with a diameter on the order of  $1\ \mu\text{m}$ , perpendicular to the plane of the wafer. These channels are filled with a good electrical conductor, and one surface of the glass is shaped with a spherical form consistent with the radius of the retina curvature. The image is serially input into the multiplexer via a very narrow flexible microcable. The real function of the micro-electronic multiplexer in Fig. C.5 is essentially the reverse of the IRFPAs microelectronic multiplexers, acting as a demultiplexer to read an image onto the stimulator array. The electrical connection to the silicon multiplexer is made so there is nothing protruding above the spherical curved envelope defined by the polished NCG surface, and therefore protects the retina from damage.

Because the test duration with the IRP experimental device are very short, there was no need to address the more difficult chronic issues that arise with permanent implants. Since patients are connected to external instrumentation during experiments, electrical shocks are prevented by isolating them using low voltage batteries and optocouplers to input signals.

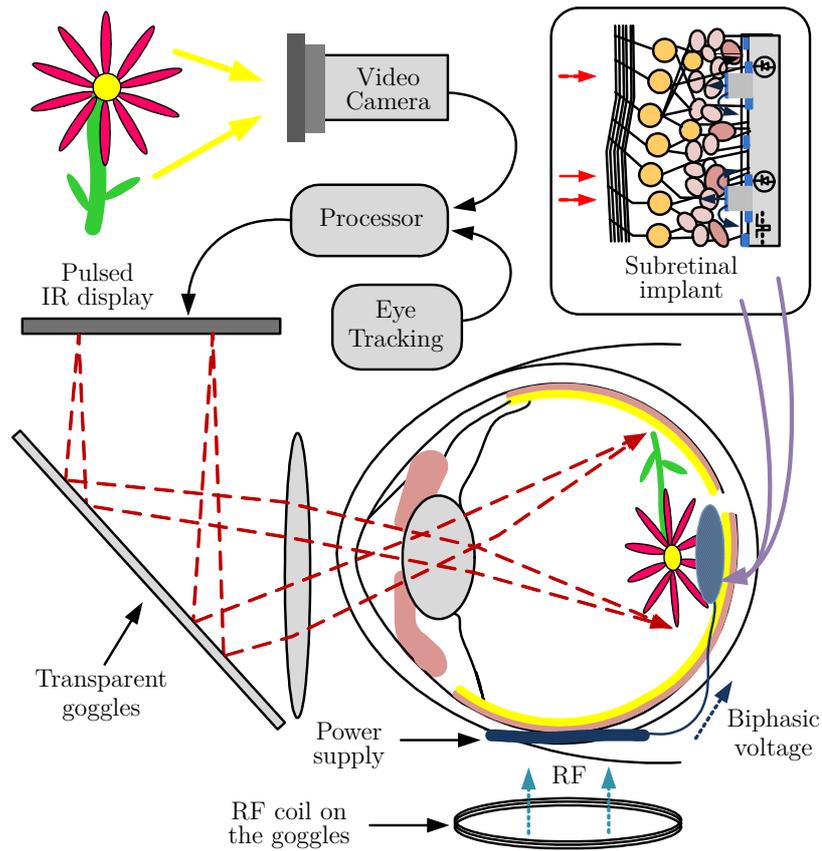
A simplified layout of the general system architecture for implementing both epiretinal and subretinal prostheses is presented in Fig. C.6. It includes a goggles-mounted video camera, an image processor and a near-infrared (NIR) display. An extraocular power supply is connected to the subretinal implant. The inset shows a magnified view of a small area of the retinal implant.

The proposed video camera acquires and transmits  $640\times 480$  pixel images at 25-50 Hz to a pocket PC. This computer processes data and displays the resulting video on an liquid crystal display (LCD) matrix mounted on goggles worn by the patient. The LCD screen is illuminated with pulsed NIR (NIR wavelength: 800-900nm) light, projecting each video image through the eye optics onto the retina. The NIR light is received by the photodiode array on an implanted chip, where each photodiode converts the NIR signal into a proportional electric current which is injected into the retina through an electrode placed in its center.

The projected NIR image is superimposed onto a normal image of the scene observed through the transparent goggles. Therefore, electrical stimulation introduces visual information into the retinal tissue above the implant, while any remaining peripheral vision responds normally to visible light. Such overlay is possible because NIR light does not activate normal photoreceptors, and the implant's response to natural visible light in the eye is negligible when compared to the bright and pulsed infrared image.

The prosthesis provides stimulation with a frame rate of up to 50 Hz in a central  $10^\circ$  visual field, with a full  $30^\circ$  field accessible via eye movements. Pixel sizes are scalable from  $100\ \mu\text{m}$  to  $25\ \mu\text{m}$ , which allows an acuity up to 20/100 to be achieved, which corresponds to 640–10,000 pixels on an implant with 3 mm in diameter.

Charge injection is maximized by biasing the photodiodes using a common pulsed biphasic power supply. Since the stimulation pulse must be synchronized with the IR light pulse, the system requires both power delivery and a trigger signal. Delivering  $20\ \mu\text{A}$ , 0.5 ms pulses to 640 electrodes at 50 Hz requires a peak current of about two tenths of a milliampere. The power transmission system is composed of a pair

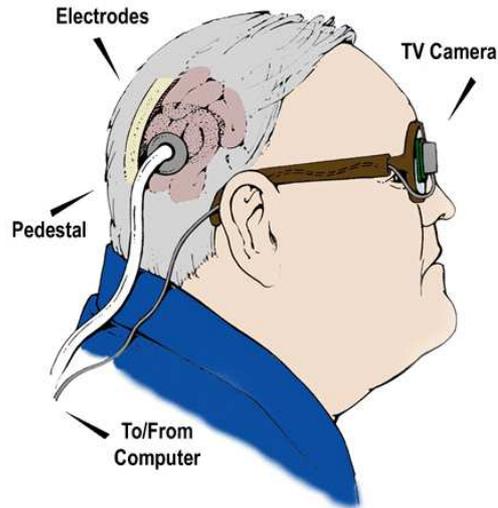


**Figure C.6.** A simplified layout of a retinal implant (adapted from [Loudin et al., 2007]).

of inductively coupled coils: the transmitter coil is mounted beside the eye on the goggles, while the receiving coil and associated electronic circuit are implanted on the eye. The operating frequency of this transmission system is limited to 1 MHz. The AC current from the receiving coil is rectified using a half-wave rectifier, which collects charge into a tantalum electrolytic capacitor to provide DC current to the rest of the circuit.

One of the most important characteristics of this bioelectronic vision system is the fact that it can be used both for epiretinal and subretinal stimulation. The optical projection of the images into the eye also preserves a natural link between eye movements and visual information. Given that video goggles project images onto a retinal area much larger than the chip itself, a larger field of view can be observed with natural eye movements. Moreover, the parallel optical transmission of information during stimulation avoids the use of multiple wires connecting the acquisition system to the electrode array. The main disadvantage of this system is the fact that the photodiodes are placed in series with the electrodes, which prevents the generation of some types of typical biphasic stimulation waveforms, such as the symmetric biphasic current pulses.



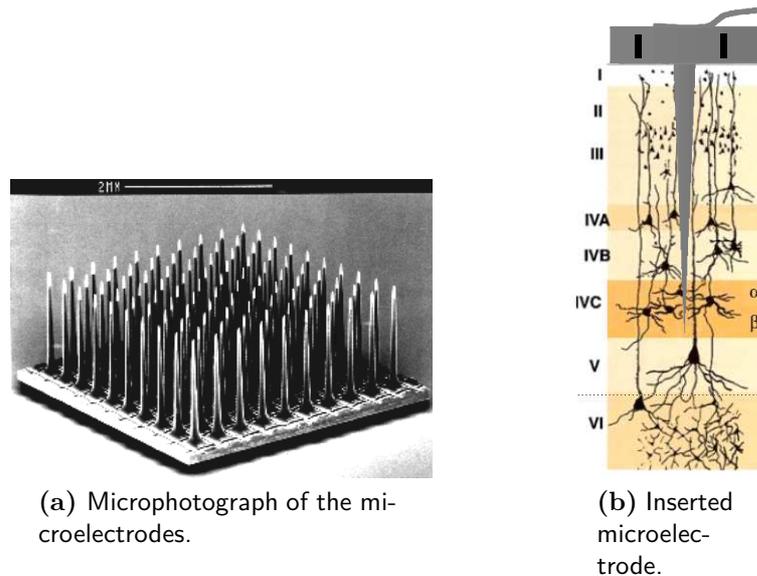


**Figure C.8.** A diagram of the Dobbie apparatus layout of a visual prosthesis.

per-second (*fps*). From experience, typical values are 4 *fps*, which includes trains containing only a single pulse. Biphasic symmetric 1 ms pulses are applied with threshold amplitudes between 10-20 V, which may vary up to  $\pm 20\%$  from day to day; the system was calibrated on a daily basis.

It has been shown that these cortical implants allow blind people to recognize patterns when several phosphenes are induced in parallel. Over time, they were even able to perform simple tasks, such as recognize and find objects with different forms in simple scenarios. However, to stimulate this type of surface electrode, which is placed at the surface of the visual cortex, the current has to be sufficiently high (on the order of mA) in order to induce an electrical field able to stimulate internal neurons of the visual cortex. Moreover, the distance between electrodes restricts the resolutions, preventing the perception of more complex patterns. Another main disadvantage of the surface cortical visual prosthesis described by Dobbie is that the apparatus includes a connector supported on a pedestal implanted in the skull, which receives the cables from the processing device and convey the signals to the implant in the cortex. This connector can be a source of infections, which can lead to serious health problems.

Intracortical microstimulation is a more recent technique for developing cortical visual prostheses. This technique is based on the insertion of microelectrodes in the visual cortex, with 1 mm to 2 mm long. The insertion allows the deep layer of neurons to be directly stimulated, decreasing the required current several orders of magnitude, from mA to  $\mu\text{A}$ . These arrays of microelectrodes not only considerably reduce the required current, but also increase safety and reduce the distance between electrodes; this allows the increase of spatial resolution of the phosphene patterns. The laboratories that have developed the most relevant research work in intracortical implants are from the Illinois Institute of Technology [Troyk et al., 2006], the National Institutes



(a) Microphotograph of the microelectrodes.

(b) Inserted microelectrode.

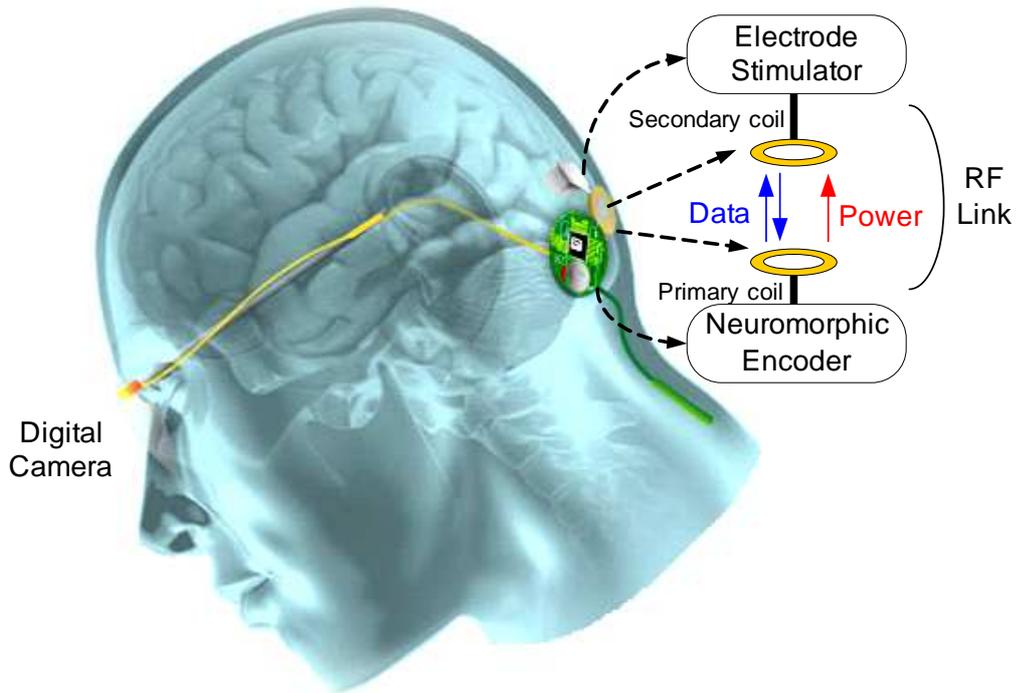
**Figure C.9.** The Utah Microelectrode Array.

of Health [Schmidt et al., 1996], and the University of Utah [Normann et al., 1999].

At the University of Utah, the team of Richard Normann has designed a microelectrode array for recording and stimulating single cells in the cortex. It consists of 100 microelectrodes with  $400\ \mu\text{m}$  spacing, each  $1.5\ \text{mm}$  long and fabricated on a silicon wafer that measures  $4\ \text{mm}$  on a side. An electron micrograph of this silicon-based microelectrode array is presented in Fig. C.9a. A technique was developed to insert the electrodes into the cerebral cortex with a single movement, and when inserted, it allows a high density of stimulation points to be achieved and it stimulates up to layer IV of the visual cortex (see Fig. C.9b).

This array has also been used in the Cortical Visual Neuroprosthesis for the Blind (CORTIVIS) project [Ahnelt et al., 2002] described with more fine details in [Martins and Sousa, 2009]. The main objective of this project was to show the feasibility of an artificial vision system capable of conferring to profoundly blind people some kind of vision, namely the discrimination of shape and location of objects, resulting in a substantial improvement in the standard of living of blind and visually impaired persons.

A scheme of a bioelectronic vision system, with the main modules of the CORTIVIS cortical neuroprosthesis, is represented in Fig. C.10. This system includes an image capture device, often a video camera, that captures the visual stimulus and converts it into electrical signals, usually electrical currents or voltages. These electrical signals are sampled, quantized and processed in the *Neuromorphic Encoder* module. This module is responsible for all the required digital processing, including the generation of the spike events used to stimulate the visual cortex neurons. The information about these events is serialized and can be transmitted to inside the skull through electrical connectors, as in [Dobelle, 2000], or through a wireless communication system. In



**Figure C.10.** Bioelectronic vision system based on the CORTIVIS neuroprosthesis (from [Piedade et al., 2005]).

both cases, the *Electrode Stimulator* module responsible for exciting the visual cortex cells through a microelectrode array has to be implanted in the visual cortex by a neurosurgeon. However, while the latter approach reduces the risk of infections, and therefore improves the patient health level; it poses the need for wireless delivery of both power and data [Piedade et al., 2005].

In the next sections, we describe and analyze in more detail the bioelectronic vision system developed to implement an intracortical visual neuroprosthesis in the scope of the CORTIVIS project.

Bioelectronic vision systems result from the contributions of several different areas of science and engineering, namely biology, neuroscience, signal processing and microelectronics. These bioelectronic vision systems can be supported in retinal implants or in cortical implants.

Retinal implants have advantages over cortical implants with regard to surgical implantation and access to target nerve cells, and offers less resistance from patients. Additionally, the mapping of the retina, which lies in the back of the eye, to a physical location in space is well known. The types of retinal implants are delineated primarily by the anatomical location of the electrode–neuron interface, which can be on the epiretinal surface or in the subretinal space. For the latter type of implant, the microphotodiodes directly replace the functions of the damaged photoreceptor cells, while the remaining intact retina’s neural network processes the electrical signals. Epiretinal implants rely on external imaging devices, that include a camera to acquire

the environment images, and an external image processor.

Regarding epiretinal prostheses, subretinal prostheses do not require external cameras or image processing units, and the patient's eye movements can be used to locate objects. However, the lack of an external source of energy for the microphotodiodes in the subretinal prosthesis is a significant drawback, since the level of ambient light is not sufficient to produce the current generated by the microphotodiodes necessary to stimulate the adjacent retinal neurons.

Clinical trials testing chronic longterm implants have been performed since the early 2000s in humans, both for subretinal and epiretinal prostheses [Javaheri et al., 2006]. In the trial periods, which vary from a few months to several years, it was demonstrated that patients perceived phosphenes in response to the electrical stimulation of the retina. Biocompatibility studies were performed in order to examine the effects of an extraneous body in the subretinal space, and surgical methods to implant the device have been evaluated as well. It has been demonstrated that the idea behind the simple subretinal approach is not effective because it lacks a viable source of power [Zrenner, 2002]. It is believed that the low levels of current delivered from the implant, although insufficient to electrically activate any remaining retinal neurons in a retina with damaged photoreceptors, may have therapeutic and neuroprotective effects for otherwise dying retinal photoreceptors.

Experiments have been reported in animals with prototypes of subretinal devices that contain external power sources. This power source supplies energy to the subretinal implant by means of very thin wires that run outside of the eye [Sachs et al., 2005]. Systems have been also designed to be used for both epiretinal and subretinal stimulation. In one of these systems [Loudin et al., 2007], near infrared light is sent to a photodiode array, which converts it into an electrical current that is injected into the retina. Charge injection is increased by biasing the diodes using a common power supply. In this system, the camera is mounted in the goggles, and the power transmission system consists of a pair of inductively coupled coils; the transmitter coil is mounted beside the eye on the goggles, while the receiving coil and associated electronic circuit are implanted in the eye.

For retinal blindness with degeneration of the ganglion cell neuron, which in turn gives rise to the optic nerve axons, a retinal prosthesis would not be helpful. Therefore, visual cortical prostheses have been pursued by a number of individual researchers and groups since the 1960s; it was experimentally shown that phosphenes could be evoked by stimulating electrodes implanted in the visual cortex.

Cortical prostheses are much more complex, not only because they have to account for the significant signal processing that must be implemented, but also due to the challenge of positioning electrodes precisely in the primary visual cortex. Dobbelle was the first to develop a cortical neuroprosthesis and to implant blind volunteers with permanent electrode arrays in the visual cortex. This cortical neuroprosthesis includes a television camera, which is mounted into a pair of sunglasses, linked to a sub-notebook computer in a belt pack. The belt pack also contains another microcontroller and associated electronics to stimulate the brain. This stimulus generator is connected through a percutaneous pedestal to the planar electrodes on the visual

cortex, in this case an array of about 64 surface electrodes. Although it has been shown that the cortical implants allow blind people to recognize patterns when several phosphenes are induced in parallel, significant drawbacks are associated with this approach to developing visual prostheses. Some of the difficulties include interactions between phosphenes and multiple phosphenes induced by a single electrode, as well as the usage of high currents and large electrodes. Occasionally, pain was caused by meningeal stimulation, and possible focal epileptic activity was induced by electrical stimulation [Pollen, 1977]. Another main disadvantage of the surface cortical visual prosthesis described by Dobbins is that the apparatus includes a connector implanted in the skull, which can lead to health problems due to infections.

A second generation of intracortical prostheses has been developed in order to overcome the main drawbacks identified in the first generation of intracortical visual prostheses. This type of visual neuroprosthesis performs intracortical microstimulation through an array of microelectrodes implanted into the primary visual cortex. The space between microelectrodes is quite reduced and the microelectrodes are inserted in depth to directly stimulate the neurons in the inner layers of the visual cortex. Microelectrode arrays are fabricated on a silicon wafer, and typically have a spacing of hundreds of micrometers and depths of about 1 mm [Maynard et al., 1997]. This new generation of intracortical prostheses makes use of wireless implants to multichannel microstimulation [Sawan et al., 2005]. Wireless inductive links carry power and stimulus information to inside the cranium, usually by means of a low-coupling transformer. This wireless link establishes a path between the primary unit, located outside the body, and the secondary unit, implanted inside the body. In this chapter, we have presented in detail the intracortical prosthesis developed in the scope of the *Cortical Visual Neuroprosthesis for the Blind* (CORTIVIS) project. In this prosthesis, the primary unit comprises a neuromorphic encoder, a forward transmitter and a backward receiver. The developed neuromorphic encoder generates the spikes to stimulate the cortex by mimicking the characteristics of the spatiotemporal receptive field response of ganglion cells. The secondary unit comprises a forward receiver, a microelectrode stimulation circuitry and a backward transmitter that is used to monitor the implant. A prototype of the proposed system was developed and tested in animals. However, clinical trials have to be further performed in order to test the implantation of the microelectrode array and the operation of an entire intracortical prosthesis.

All of this research has paved the way toward restoring useful vision to profoundly blind people by interfacing a cortical neuroprosthesis with the visual cortex. However, several issues have to be addressed by research groups such as: the biocompatibility of microelectronics and microelectrodes, the heat generated and dissipated by the intracranial module space, and the plasticity of the visual system in response to electrical stimulation. Attention also has to be given to understanding how the brain interprets a stimulation pattern resulting from an increasing number of electrodes, which is also a crucial issue in the evolution of vision prosthetic design. At last it is important to have the involvement of companies to produce, manufacture and test these devices medically. Hence the research community will be able to perform visual

psychophysical experiments, in order to develop stimulation algorithms that result in the "best" perceptions.

While still a mirage, blind individuals will certainly require high resolution visual prosthesis. This requirement is more likely to be fulfilled first by retinal prosthesis. Surveys with these patients indicate that mobility without a cane, face recognition and reading are the main necessary enabling capabilities [Weiland and Humayun, 2008]. However the results and implications of fully functional visual prostheses are exciting and the challenges ahead are demanding and promise to be exciting. Table C.1 resumes the main pros and contra of the different approaches and the challenges in the development of visual prosthesis.

Table C.1 summarizes the main pros and contra of the different approaches, referred in this chapter, to develop visual neuroprostheses.

**Table C.1.** Main pros and cons of visual prostheses approaches.

Advantages	Disadvantages
Visual Cortex	
<p>Only approach possible for a non-functional retina and/or optic nerve</p> <p>Implant site robust and protected by skull</p> <p>Easy surgical access</p> <p>High density electrode implantation</p> <p>Phosphene thresholds are low (1-10 <math>\mu A</math>)</p>	<p>Stimulation far from photoreceptors</p> <p>Possibly poor visuotopic organization</p> <p>Multiple feature representations in V1 (color, lines, motion, ocular dominance)</p> <p>Societal phobias about "brain implant"</p> <p>Consequences of surgical complications</p>
Epiretinal	
<p>Stimulating close to photoreceptors: uses native processing in thalamus and cortex</p> <p>Less surgical complications than in cortical</p> <p>Saccadic eye motions cause sheer loads on implanted arrays</p> <p>Difficult to adhere the electrode array to the retina</p>	<p>Requires functional optic nerve pathway</p> <p>May stimulate optic nerve fibers: greatly complicate visuotopic organization</p>
Subretinal	
<p>Stimulation closest to photoreceptors: uses retinal, thalamic and cortical processing</p> <p>If bipolar cells can be directly stimulated retinotopic organization should be preserved</p> <p>Less surgical complications than in cortical implants</p>	<p>Requires functional retina and optic nerve pathway</p> <p>Blockage of nutrients from choroid by the implant</p> <p>Very complex surgical access</p> <p>Can't stimulate cells passively with microimplants (requires external power)</p>
Optic Nerve	
<p>Less surgical complications than in cortical</p>	<p>Requires functional optic nerve pathway</p> <p>Visuotopic organization requires placing electrodes at many closely spaced regions of the optic nerve</p> <p>Complex electrode array to provide patterned vision</p> <p>Very difficult surgical access</p>

# Bibliography

- Abbott, L. F. and Kepler, T. B. (1990). Model neurons: From Hodgkin-Huxley to Hopfield. In Garrido, L., editor, *Statistical Mechanics of Neural Networks*, volume 368 of *Lecture Notes in Physics*, chapter 2, pages 5–18. Springer Berlin / Heidelberg, Berlin, Heidelberg.
- Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions*. Dover Publications, New York, USA.
- Agnew, W. and McCreery, D., editors (1990). *Neural Prostheses: Fundamental Studies*. Biophysics and Bioengineering Series. Prentice Hall, Englewood Cliffs, New Jersey, USA.
- Aguiar, A. P. (2007). Multiple-model adaptive estimators: Open problems and future directions. *European Control Conference*, 24:5544–5545.
- Ahnelt, P., Ammermüller, J., Pelayo, F., Bongard, M., Palomar, D., Piedade, M., Ferrandez, J., Borg-Graham, L., and Fernandez, E. (2002). Neuroscientific basis for the design and development of a bioinspired visual processing front-end. In *IFMBE Proceedings*, volume 3, pages 1692–1693, Vienna.
- Akçay, H. and Ninness, B. (1999). Orthonormal basis functions for modelling continuous-time systems. *Signal Processing*, 77(1):261–274.
- Alspach, D. L. and Sorenson, H. W. (1972). Nonlinear bayesian estimation using gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17(4):439–448.
- Anderson, B. D. O. and Moore, J. B. (2005). *Optimal Filtering*. Electrical Engineering Series. Dover Publications.
- Arfken, G. B. and Weber, H. J. (2005). *Mathematical Methods for Physicists*. Academic Press, San Diego, USA, 6th edition.
- Arno, P., Capelle, C., Wanet-Defalque, M.-C., Catalan-Ahumada, M., and Veraart, C. (1999). Auditory coding of visual patterns for the blind. *Perception*, 28(8):1013–1029.

## Bibliography

- Athans, M. and Chang, C. B. (1976). Adaptive estimation and parameter identification using multiple model estimation algorithm. Technical Note 1976-28, Lincoln Lab, MIT, Lexington, Massachusetts.
- Baccus, S. and Meister, M. (2002). Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36:909–919.
- Bair, W., Koch, C., Newsome, W., and Britten, K. (1994). Power spectrum analysis of bursting cells in area MT in the behaving monkey. *The Journal of Neuroscience*, 14(5):2870–2892.
- Bair, W. and O’Keefe, L. P. (1998). The influence of fixational eye movements on the response of neurons in area mt of the macaque. *Visual Neuroscience*, 15(4):779–786.
- Baram, Y. and Sandell, N. R. (1978). An information theoretic approach to dynamical systems modeling and identification. *IEEE Transactions on Automatic Control*, 23(1):61 – 66.
- Bernard, C., Ge, Y. C., Stockley, E., Willis, J. B., and Wheal, H. V. (1994). Synaptic integration of NMDA and non-NMDA receptors in large neuronal network models solved by means of differential equations. *Biological Cybernetics*, 70(3):267–273.
- Berry, M. J., Warland, D. K., and Meister, M. (1997). The structure and precision of retinal spike trains. *Neurobiology*, 94:5411–5416.
- Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., and Warland, S. (1991). Reading a neural code. *Science*, 252:1854–1857.
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Brélen, M. E., Duret, F., Gérard, B., Delbeke, J., and Veraart, C. (2005). Creating a meaningful visual perception in blind volunteers by optic nerve stimulation. *IOP Publishing, Journal of Neural Engineering*, 2:S22–S28.
- Brenner, N., Agam, O., Bialek, W., and de Ruyter van Steveninck, R. (2002). Statistical properties of spike trains: Universal and stimulus-dependent aspects. *Physical Review*, 66.
- Brindley, G. S. and Lewin, W. S. (1968). The sensations produced by electrical stimulation of the visual cortex. *The Journal of Physiology*, 196:479–493.
- Brunel, N. and van Mark C. W. Rossum (2008). Quantitative investigations of electrical nerve excitation treated as polarization: Louis lapicque 1907. *Biol. Cybern.*, 97(5):341–349.

- Brunel, N. and van Rossum, M. C. W. (2007). Lapicque's 1907 paper: from frogs to integrate-and-fire. *Biol Cybern*, 97(5-6):337–339.
- Burkitt, A. N. (2006). A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input. *Biological Cybernetics*, 95(1):1–19.
- Capela, S., Tomás, P., and Sousa, L. (2007). Stochastic integrate-and-fire model for the retina. In *15th European Signal Processing Conference (EUSIPCO 2007)*, pages 2514–2518. EURASIP.
- Carmena, J. M. (2012). How to control a prosthesis with your mind. *IEEE Spectrum*.
- Casti, A., Hayot, F., Xiao, Y., and Kaplan, E. (2008). A simple model of retina-lgn transmission. *Journal of Computational Neuroscience*, 24:235–252. 10.1007/s10827-007-0053-7.
- Chichilnisky, E. J. (2001). A simple white-noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12(2):199–213.
- Chichilnisky, E. J. and Rieke, F. (2005). Detection sensitivity and temporal resolution of visual signals near absolute threshold in the salamander retina. *The Journal of Neuroscience*, 2(25):318–330.
- Chow, A. Y., Chow, V. Y., Packo, K. H., Pollack, J. S., Peyman, G. A., and Schuchard, R. (2004). The artificial silicon retina microchip for the treatment of vision loss from retinitis pigmentosa. *Archives of Ophthalmology*, 122(4):460–469.
- Curcio, C. A., Sloan, K. R., Kalina, R. E., and Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, 4(292):497–523.
- Cyberkinetics (2008). Neurotechnology systems, inc. <http://www.cyberkineticsinc.com>.
- Davis, H. F. (1989). *Fourier Series and Orthogonal Functions*. Dover Books on Mathematics. Dover Publications.
- Dayan, P. and Abbot, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, Cambridge, Massachusetts, USA.
- De Boor, C. (2001). *A Practical Guide to Splines*. Number vol. 27 in Applied Mathematical Sciences. Springer.
- de Ruyter van Steveninck, R. and Bialek, W. (1988). Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spikes sequences. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, 234(1277):379–414.

## Bibliography

- Demb, J. B., Haarsma, L., Freed, M. A., and Sterling, P. (1999). Functional circuitry of the retinal ganglion cell's nonlinear receptive field. *The Journal of Neuroscience*, 19(22):9756–9767.
- Dobelle, W. and Mladejovsky, W. (1974). Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind. *Journal of Physiology*, 243:553–576.
- Dobelle, W. H. (2000). Artificial vision for the blind by connecting a television camera to the visual cortex. *American Society for Artificial Internal Organs (ASAIO) Journal*, 46:3–9.
- Dobelle, W. H., Mladejovsky, M. G., and Girvin, J. P. (1974). Artificial vision for the blind: Electrical stimulation of visual cortex offers hope for a functional prosthesis. *Science*, 183(4123):440–444.
- Dong, D. W. and Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–358.
- Donoghue, J. P. (2002). Connecting cortex to machines: recent advances in brain interfaces. *Nature Neuroscience*, 5:1085–1088.
- Dowling, J. E. (1987). *The retina: an approachable part of the brain*. Belknap Press.
- Eckmiller, R. (1997). Learning retina implants with epiretinal contacts. *Ophthalmic Research*, 29(5):281–289.
- Eckmiller, R., Hunermann, R., and Becker, M. (1999). Exploration of a dialog-based tunable retina encoder for retina implants. *Neurocomputing*, 26–27:1005–1011.
- Efstratiadis, A. and Koutsoyiannis, D. (2002). An evolutionary annealing-simplex algorithm for global optimisation of water resource systems. In *Proceedings of the Fifth International Conference on Hydroinformatics (Hydroinformatics 2002)*, pages 1423–1428, Cardiff, United Kingdom. IWA Publishing.
- Eggermont, J. J. (1998). Is there a neural code? *Neuroscience and Biobehavioral Reviews*, 22(2):355–370.
- Eide, P. and Maybeck, P. (1996). An MMAE failure detection system for the F-16. *IEEE Transactions on Aerospace and Electronic Systems*, 32(16):1125–1136.
- Fekri, S., Athans, M., and Pascoal, A. (2006). Issues, progress and new results in robust adaptive control. *International Journal of Adaptive Control and Signal Processing*, 20:519–579.
- Fernández, E., Ferrandez, J.-M., Ammermüller, J., and Normann, R. A. (2000). Population coding in spike trains of simultaneously recorded retinal ganglion cells. *Brain Research*, 887(1):222–229.

- Fisher, M. E. (1983). Scaling, universality and renormalization group theory. In Hahne, F. J., editor, *Lecture Notes in Physics*, volume 186, pages 1–139. Springer.
- FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445–466.
- Flannery, B. P., Teukolsky, S. A., Press, W. H., and Vetterling, W. T. (2002). *Numerical Recipes in C*. Cambridge University Press, Cambridge, Massachusetts, USA, 2nd edition. The Art of Scientific Computing.
- Gabriel, Singla, P., Singh, T., and Scott, P. D. (2012). Adaptive gaussian sum filter for nonlinear bayesian estimation. *IEEE Transactions on Automatic Control*.
- Gelb, A. (1974). *Applied Optimal Estimation*. The MIT Press.
- Gerstein, G. L. and Mandelbrot, B. (1964). Random Walk Models for the Spike Activity of Single Neuron. *Biophysical Journal*, 4:41–68.
- Gerstner, W. and Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.
- Gerstner, W. and Naud, R. (2009). How good are neuron models? *Science*, 326(5951):379–380.
- Gestri, G., Mastebroek, H. A. K., and Zaagman, W. H. (1980). Stochastic constancy, variability and adaptation of spike generation: Performance of a giant neuron in the visual system of the fly. *Biological Cybernetics (Springer-Verlag)*, 38(1):31–40.
- Gewaltig, M.-O., Diesmann, M., and Aertsen, A. (2001). Propagation of cortical synfire activity: survival probability in single trials and stability in the mean. *Neural Networks*, 14(6-7):657–673.
- Grill-Spector, K. and Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, 27:649–677.
- Grumeta, A. E., Jr., J. L. W., and Rizzo III, J. F. (2000). Multi-electrode stimulation and recording in the isolated retina. *Journal of Neuroscience Methods, Elsevier*, 101:31–42.
- Hamming, R. W. (1987). *Numerical Methods for Scientists and Engineers*. Dover Publications, second edition.
- Hanlon, P. and Maybeck, P. (2000). Multiple-model adaptive estimation using a residual correlation Kalman filter bank. *Aerospace and Electronic Systems, IEEE Transactions on*, 36(2):393–406.
- Hassani, V., Aguiar, A., Athans, M., and Pascoal, A. (2009a). Multiple model adaptive estimation and model identification using a minimum energy criterion. *Proceedings of American Control Conference (ACC)*, pages 518–523.

## Bibliography

- Hassani, V., Aguiar, A. P., Athans, M., and Pascoal, A. M. (2009b). Multiple model adaptive estimation and model identification using a minimum energy criterion. In *Proceedings of the 2009 Conference on American Control Conference*, pages 518–523, Piscataway, NJ, USA. IEEE Press.
- Hassani, V., Aguiar, A. P., Pascoal, A. M., and Athans, M. (2009c). Further results on plant parameter identification using continuous-time multiple-model adaptive estimators. *48th IEEE Conference on Decision and Control*, pages 7261–7266.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability. Taylor & Francis.
- Haykin, S. (1999). *Neural networks: a comprehensive foundation*. Prentice Hall.
- Hennig, M. H. and Funke, K. (2001). A biophysically realistic simulation of the vertebrate retina. *Neurocomputing*, 38-40:659 – 665.
- Hérault, J. and Durette, B. (2007). Modeling visual perception for image processing. In *Proceedings of the 9th international work conference on Artificial neural networks, IWANN'07*, pages 662–675, Berlin, Heidelberg. Springer-Verlag.
- Herz, A. V. M., Gollisch, T., Machens, C. K., and Jaeger, D. (2006). Modeling single-neuron dynamics and computations: A balance of detail and abstraction. *Science*, 314(5796):80–85.
- Hessburg, P. and Rizzo III, J. (2007). The eye and the chip. World congress on artificial vision 2006. *Journal of Neural Engineering*, 4(1):1–2.
- Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544.
- Humayun, M., Jr, E. J., Dagnelie, G., Greenberg, R., Propst, R., and Phillips, D. (1996). Visual perception elicited by electrical stimulation of retina in blind humans. *Archives of Ophthalmology*, 114:40–46.
- Humayun, M. S., de Juan Jr., E., Weiland, J. D., Dagnelie, G., Katona, S., Greenberg, R., and Suzuki, S. (1999). Pattern electrical stimulation of the human retina. *Vision Research*, 39(15):2569–2576.
- IEEE (2012). IEEE Xplore Digital Library. <http://ieeexplore.ieee.org>.
- Ifeachor, E. C. and Jervis, B. W. (2002). *Digital Signal Processing: A Practical Approach*. Prentice Hall, Harlow, England, 2nd edition.

- Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons. *IEEE Transactions on Neural Networks*, 15:1063–1070.
- Javaheri, M., Hahn, D., Lakhanpal, R., Weiland, J., and Humayun, M. (2006). Retinal prostheses for the blind. *Annals Academy of Medicine Singapore*, 35(3):137–144.
- Jolivet, R. and Gerstner, W. (2004). Predicting spike times of a detailed conductance-based neuron model driven by stochastic spike arrival. *Journal of Physiology*, 98(4-6):442–451.
- Jolivet, R., Lewis, T. J., and Gerstner, W. (2004). Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *Journal of Neurophysiology*, 92(2):959–976.
- Julier, S. J. (2002). The scaled unscented transformation. In *Proceedings of the American Control Conference*, volume 6, pages 4555 – 4559.
- Julier, S. J. and Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems. In *In Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, volume 3068, pages 182–193.
- Julier, S. J. and Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422.
- Keat, J., Reinagel, P., Reid, R. C., and Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron*, 30:803–817.
- Kim, K. J. and Rieke, F. (2001). Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *The Journal of Neuroscience*, 21(1):287–299.
- Kistler, W. M., Gerstner, W., and van Hemmen, J. L. (1997). Reduction of the Hodgkin-Huxley equations to a single-variable threshold model. *Neural Comput.*, 9(5):1015–1045.
- Kolb, H. (1997). Amacrine cells of the mammalian retina: neurocircuitry and functional roles. *Eye (London, England)*, 11(6):904–923.
- Kolb, H. (2003). How the retina works. *Scientific American*, 91:28–35.
- Kolb, H., Fernández, E., and Nelson, R. (2012). WebVision: The organization of the retina and visual system. <http://webvision.med.utah.edu>.
- Kreiman, G., Krahe, R., Metzner, W., Koch, C., and Gabbiani, F. (2000). Robustness and Variability of Neuronal Coding by Amplitude-Sensitive Afferents in the Weakly Electric Fish *Eigenmannia*. *Journal of Neurophysiology*, 84(1):189–204.

## Bibliography

- Kuriscak, E., Marsalek, P., Stroffek, J., and Wunsch, Z. (2012). The effect of neural noise on spike time precision in a detailed CA3 neuron model. *Computational and Mathematical Methods in Medicine*, 2012.
- Lainiotis, D. G. (1971). Optimal adaptive estimation: Structure and parameter adaption. *Automatic Control, IEEE Transactions on*, 16(2):160 – 170.
- Lakhanpal, R., Yanai, D., Weiland, J., Fujii, G., Caffey, S., Greenberg, R. R., Jr, J. E., and Humayun, M. (2003). Advances in the development of visual prostheses. *Current Opinion in Ophthalmology*, 14:122–127.
- Lewen, G. D., Strong, S. P., and Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, 275:1805–1808.
- Li, X.-R. and Bar-Shalom, Y. (1996). Multiple-model estimation with variable structure. *IEEE Transactions on Automatic Control*, 41(4):478–493.
- Lim, J. S. (1990). *Two-Dimensional Signal and Image Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, USA.
- Ljung, L. (1999). *System Identification: Theory for the User*. Pearson Education, New Jersey, USA.
- Loudin, J., Simanovskii, D., Vijayraghavan, K., Sramek, C., Butterwick, A., Huie, P., McLean, G., and Palanker, D. (2007). Optoelectronic retinal prosthesis: System design and performance. *Journal of Neural Engineering*, 4:S72–S84.
- Magill, D. (1965). Optimal adaptive estimation of sampled stochastic processes. *Automatic Control, IEEE Transactions on*, 10(4):434 – 439.
- Margalit, E., Maia, M., Weiland, J. D., Greenberg, R. J., Fujii, G. Y., Torres, G., Piyathaisere, D. V., O’Hearn, T. M., Liu, W., Lazzi, G., Dagnelie, G., Scribner, D. A., de Juan Jr, E., and Humayun, M. S. (2002). Retinal prosthesis for the blind. *Survey of Ophthalmology*, 47(4):335–356.
- Marmarelis, V. Z. (2004). *Nonlinear Dynamic Modeling of Physiological Systems*. IEEE Series on Biomedical Engineering. Wiley-Interscience.
- Marr, D. and Hildreth, E. (1980). Theory of Edge Detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 207(1167):187–217.
- Martins, J., Tomás, P., and Sousa, L. (2009). Neural code metrics: Analysis and application to the assessment of neural models. *Neurocomputing*, 72(10 – 12):2337 – 2350.
- Martins, J. C. (2006). Parameters identification with the multiple model adaptive estimation (MMAE) algorithm. In *Proceedings of the 25th IASTED international conference on Modeling, identification, and control, MIC’06*, pages 501–506, Anaheim, CA, USA. ACTA Press.

- Martins, J. C., Caeiro, J. J., and Sousa, L. A. (2011a). A new approach to system identification and parameter tuning with multiple model adaptive estimators. In *7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011)*, pages 72–77.
- Martins, J. C., Caeiro, J. J., and Sousa, L. A. (2011b). On the use of adaptive model constellations in multiple model adaptive estimators. In *Proc. of RecPad 2011 - 17th Portuguese Conference on Pattern Recognition*, Porto, Portugal.
- Martins, J. C., Caeiro, J. J., and Sousa, L. A. (2013). Constellation based multiple model adaptive estimators for system identification and parameters tuning. *Digital Signal Processing, Elsevier*. submitted.
- Martins, J. C. and Sousa, L. A. (2005). Comparison of computational retina models. In *5th IASTED Conference on Visualization, Imaging, and Image Processing*, pages 156–161, Benidorm, Spain.
- Martins, J. C. and Sousa, L. A. (2009). *Bioelectronic Vision: Retina Models, Evaluation Metrics, and System Design*, volume 3 of *Series on Bioengineering & Biomedical Engineering*. World Scientific, Singapore.
- Martins, S. F., Sousa, L. A., and Martins, J. C. (2007). Additive logistic regression applied to retina modelling. *IEEE International Conference on Image Processing, 2007. ICIP2007.*, 3:309–312.
- Mathews, D. J. H. (2011). Deep brain stimulation, personal identity and policy. *International Review of Psychiatry*, 23(5):486–492.
- Maynard, E. M. (2001). Visual prosthesis. *Annual Review in Biomedical Engineering*, 3:145–168.
- Maynard, E. M., Nordhausen, C. T., and Normann, R. A. (1997). The Utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalography and Clinical Neurophysiology*, 102(3):228–239.
- Berry II, M. J., Brivanlou, I. H., Jordan, T. A., and Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398:334–338.
- Berry II, M. J. and Meister, M. (1998). Refractoriness and neural precision. *The Journal of Neuroscience*, 18(6):2200–2221.
- Fernández, E., Pelayo, F., Romero, S., Bongard, M., Marin, C., Alfaro, A., and Merabet, L. (2005). Development of a cortical visual neuroprosthesis for the blind: the relevance of neuroplasticity. *Journal of Neural Engineering*, 2:R1–R12.
- Optobionics Corporation (2012). Optobionics: Technology for vision. <http://www.optobionics.com>.

## Bibliography

- Project CORTIVIS (2006). Cortical visual neuroprosthesis for the blind. <http://cortivis.umh.es/>.
- Rizzo III, J. F. and Wyatt, J. (1997). Prospects for a visual prosthesis. *The Neuroscientist*, 3(4):251–262.
- Vision Egg (2007). Visionegg. <http://www.visionegg.org>.
- Meister, M., Lagnado, L., and Baylor, D. A. (1995). Concerted signaling by retinal ganglion cells. *Science*, 270:1207–1210.
- Meister, M. and Berry II, M. J. (1999). The neural code of the retina. *Neuron*, 22:253–450.
- Meister, M., Pine, J., and Baylor, D. A. (1994). Multi-neuronal signals from the retina: acquisition and analysis. *Journal of Neuroscience Methods, Elsevier*, 51:95–106.
- Mihalaş, S. and Niebur, E. (2009). A generalized linear integrate-and-fire neural model produces diverse spiking behaviors. *Neural Comput.*, 21(3):704–718.
- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, Singapore, 3rd edition.
- Moon, T. K. and Stirling, W. C. (2000). *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, Upper Saddle River, New Jersey, USA.
- Moore, J. B. and Hawkes, R. M. (1975). Decision methods in dynamic system identification. In *Proc. of IEEE Conference on Decision and Control*, volume 14, pages 645 – 650.
- Mullowney, P. and Iyengar, S. (2008). Parameter estimation for a leaky integrate-and-fire neuronal model from ISI data. *Journal of Computational Neuroscience*, 24:179–194.
- Nagumo, J., Arimoto, S., and Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50:2061–2070.
- Nawrot, M., Aertsen, A., and Rotter, S. (1999). Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity. *Journal of Neuroscience Methods*, 94(1):81–92.
- Nelles, O. (2001). *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Engineering Online Library. Springer.
- Nelson, M. and Rinzel, J. (1995). *The book of GENESIS: exploring realistic neural models with the GEneral NEural SIMulation System*, chapter The Hodgkin-Huxley Model, pages 27–51. Springer-Verlag, New York.

- Nirenberg, S., Carcieri, S. M., Jacobs, A. L., and Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature*, 411:698–701.
- Nirenberg, S. and Latham, P. E. (1998). Population coding in the retina. *Current Opinion Neurobiology*, 8(4):488–493.
- Nirenberg, S. and Latham, P. E. (2003). Decoding neuronal spike trains: How important are correlations? *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 100(12):7348–7353.
- Normann, R. A., Maynard, E. M., Rousche, P. J., and Warren, D. J. (1999). A neural interface for a cortical vision prosthesis. *Vision Research*, 39(15):2577–2587.
- Oppenheim, A. V., Shafer, R. W., and Buck, J. R. (1999a). *Discrete-Time Signal Processing*. Prentice-Hall, Upper Saddle River, New York, USA, 2nd edition.
- Oppenheim, A. V., Willsky, A. S., and Nawab, S. H. (1999b). *Signals & Systems*. Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA, 2nd edition.
- Orfanidis, S. J. (1990). *Optimum Signal Processing: An Introduction*. McGraw-Hill Book Co., Singapore, 2nd edition.
- Paninski, L., Pillow, J. W., and Simoncelli, E. P. (2004). Maximum Likelihood Estimation of a Stochastic Integrate-and-Fire Neural Encoding Model. *Neural Computation*, 16(12):2533–2561.
- Paninsky, L. (2006). The most likely voltage path and large deviations approximations for integrate-and-fire neurons. *Journal of Computational Neuroscience*, 21(1):71–87.
- Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, USA, 4th edition edition.
- Passaglia, C. L., Enroth-Cugell, C., and Troy, J. B. (2001). Effects of remote stimulation on the mean firing rate of cat retinal ganglion cells. *The Journal of Neuroscience*, 21(15):5794–5803.
- Peyman, G., Chow, A., Liang, C., Chow, V., Perlman, J., and Peachey, N. (1998). Subretinal semiconductor microphotodiode array. *Ophthalmic Surgery and Lasers*, 29(3):234–241.
- Pezaris, J. S. and Reid, R. C. (2007). Demonstration of artificial visual percepts generated through thalamic microstimulation. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 104(18):7670–7675.
- Picchini, U., Ditlevsen, S., De Gaetano, A., and Lansky, P. (2008). Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. *Neural Computation*, 20(11):2696–2714.

## Bibliography

- Piedade, M., Gerald, J., Sousa, L., Tavares, G., and Tomás, P. (2005). Visual neuro-prosthesis: A non invasive system for stimulating the cortex. *IEEE Transactions on Circuits and Systems*, pages 2648–2662.
- Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., and Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *The Journal of Neuroscience*, 25(47):11003–11013.
- Pillow, J. W. and Simoncelli, E. P. (2003). Biases in white noise analysis due to non-Poisson spike generation. *Neurocomputing*, 52–54:109–115.
- Pillow, J. W. and Simoncelli, E. P. (2006). Dimensionality reduction in neural models: An information-theoretic generalization of spike-triggered average and covariance analysis. *Journal of Vision*, 6(4):414–428.
- Plesser, H. E. and Gerstner, W. (2000). Noise in Integrate-and-Fire Neurons: From Stochastic Input to Escape Rates. *Neural Computation*, 12(2):367–384.
- Pollen, D. (1975). *The Nervous System*, chapter Some perceptual effects of electrical stimulation of the visual cortex in man, pages 519–528. Raven Press, New York, USA.
- Pollen, D. (1977). Responses of single neurons to electrical stimulation of the surface of the visual cortex. *Brain, Behavior and Evolution*, 14:67–86.
- Project RNC (2008). Retina neural code: Accurate modelling towards an artificial visual system. <http://sips.inesc-id.pt/~rnc/>.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., Lamantia, A.-S., McNamara, J. O., and Williams, S. M. (2007). *Neuroscience*. Sinauer Associates, Inc., Sunderland, MA, USA, 4th edition.
- Randall, D. J., Burggren, W. W., French, K., and Eckert, R. (2002). *Eckert Animal Physiology: Mechanisms and Adaptations*. W.H. Freeman and Company.
- Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., and Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *The Journal of Neurophysiology*, 77(5):2836–2841.
- Reid, D. B. (1979). An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, 24:843–854.
- Reinagel, P. (2001). How do visual neurons respond in the real world? *Current Opinion in Neurobiology*, 11(4):437–442.
- Reinagel, P. and Reid, R. C. (2002). Precise firing events are conserved across neurons. *The Journal of Neuroscience*, 22(16):6837–6841.

- Rieke, F. (2001). Temporal contrast adaptation in salamander bipolar cells. *The Journal of Neuroscience*, 21(5):9445–9454.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. The MIT Press, Cambridge, Massachusetts, USA.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research, Pergamon Press*, 5(12):583–601.
- Rodieck, R. W. (1998). *The First Steps in Seeing*. Sinauer Associates, Sunderland, Massachusetts, USA.
- Rotter, S. and Diesmann, M. (1999). Exact digital simulation of time-invariant linear systems with applications to neuronal modeling. *Biological Cybernetics*, 81(5–6):381–402.
- Rugh, W. J. (1981). *Nonlinear System Theory: The Volterra Wiener Approach*. Johns Hopkins Univ. Press., Baltimore, USA.
- Rust, N. C., Schwartz, O., Movshon, J. A., and Simoncelli, E. P. (2004). Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1. *Neurocomputing*, pages 793–799.
- Sachs, H., Schanze, T., Brunner, U., Sailer, H., and Wiesenack, C. (2005). Transscleral implantation and neurophysiological testing of subretinal polyimide film electrodes in the domestic pig in visual prosthesis development. *Journal of Neural Engineering*, 2(S57–S64).
- Santos, M., Fernandes, J., and Piedade, M. (2006). A microelectrode stimulation system for a cortical neuroprosthesis. In *Proc. of Conference Design of Circuits and Integrated Systems (DCIS'06)*.
- Sawan, M., Hu, Y., and Coulombe, J. (2005). Wireless smart implants dedicated to multichannel monitoring and microstimulation. *IEEE Circuits and Systems Magazine*, 5(1):21–39.
- Schmidt, E., Bak, M., Hambrecht, F., Kufta, C., O'Rourke, D., and Vallabhanath, P. (1996). Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex. *Brain*, 119:507–522.
- Schwartz, O., Chichilnisky, E. J., and Simoncelli, E. P. (2002). Characterizing neural gain control using spike-triggered covariance. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advanced Neural Information Processing Systems (NIPS\*01)*, volume 14, pages 269–276, Cambridge, Massachusetts, USA. MIT Press.

## Bibliography

- Scribner, D., Humayun, M., Justus, B., Merritt, C., Klein, R., Howard, J., Peckerar, M., Perkins, F., Johnson, L., Bassett, W., Skeath, P., Margalit, E., Eong, K.-G. A., Weiland, J., de Juan Jr., E., Finch, J., Graham, R., Trautfield, C., and Taylor, S. (2001). Intraocular retinal prosthesis test device. *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.
- Sellers, P. H. (1974). On the theory and computation of evolutionary distances. *SIAM Journal Applied Mathematics*, 26(4):787–793.
- Shimazaki, H. and Shinomoto, S. (2010). Kernel bandwidth optimization in spike rate estimation. *Journal of computational neuroscience*, 29(1–2):171–182.
- Shoham, S. (2001). *Advances towards an implantable motor cortical interface*. PhD thesis, University of Utah.
- Simon, D. (2006). *Optimal state estimation: Kalman,  $H_\infty$  and nonlinear approaches*. Wiley-Interscience.
- Simoncelli, E. P., Paninski, L., Pillow, J., and Schwartz, O. (2004). *Characterization of Neural Responses with Stochastic Stimuli*, chapter 23, pages 327–338. MIT Press, 3rd edition.
- Stengel, R. F. (1994). *Optimal Control and Estimation*. Dover Publications, Inc., New York.
- Therrien, C. W. (1992). *Discrete Random Signals and Statistical Signal Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, USA.
- Thiel, A., Wilke, S. D., Greschner, M., Bongard, M., Ammermüller, J., Eurich, C. W., and Schwegler, H. (2003). *Temporally Faithful Representations of Salient Stimulus Movement Patterns in the Early Visual System*, chapter Visual Attention Mechanisms, pages 93–100. Kluwer Academic Publishers.
- Tomás, P. and Sousa, L. (2008). Statistical analysis of a spike train distance in Poisson models. *IEEE Signal Processing Letters*, 15(1):357–360.
- Tomás, P. (2009). *Neural Code: Tuning and Assessment of Retina Models*. PhD thesis, IST-TULisbon, Lisbon, Portugal.
- Tomás, P., Martins, J. C., and Sousa, L. A. (2008). Towards a unified model for the retina: Static vs dynamic integrate and fire models. In *International Conference on Bio-Inspired Systems and Signal Processing (BIOSIGNALS 2008)*, volume 2, pages 528–533, Funchal, Madeira, Portugal.
- Tomás, P. and Sousa, L. (2007). Feature selection for the stochastic integrate and fire model. In *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on*, pages 1–6.

- Troyk, P., Detlefsen, D., and DeMichele, G. (2006). A multifunctional neural electrode stimulation ASIC using neurotalk interface. In *Proceedings of the 28th IEEE EMBS Annual International Conference*, pages 2994–2997, Canada.
- Unidad de Neuroprótesis y Rehabilitación Visual (2008). Instituto de Bioingeniería, Universidad Miguel Hernández. <http://bioingenieria.umh.es/unidades.asp?unidad=8>.
- Uzzell, V. J. and Chichilnisky, E. J. (2004). Precision of spike trains in primate retinal ganglion cells. *Journal of Neurophysiology, The American Physiological Society*, 92:780–789.
- van Rossum, M. C. W. (2001). A Novel Spike Distance. *Neural Comp.*, 13(4):751–763.
- Ventura, V., Carta, R., Kass, R. E., Gettner, S. N., and Olson, C. R. (2002). Statistical analysis of temporal evolution in single spike-neuron firing rates. *Biostatistics*, 3(1):1–20.
- Victor, J. D. (1999). Temporal aspects of neural coding in the retina and lateral geniculate: a review. *IOP Publishing, Network: Comput. Neural Syst.* 10 (1999), 10:R1–R66.
- Victor, J. D. (2005). Spike train metrics. *Current Opinion in Neurobiology*, 15(5):585–592.
- Victor, J. D. and Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: A metric-space analysis. *Journal of Physiology*, 76(2):1310–1326.
- Victor, J. D. and Purpura, K. P. (1997). Metric-space analysis of spike trains: theory, algorithms and application. *Network: Computational Neural Systems*, 8(2):127–164.
- Victor, J. D. and Shapley, R. M. (1979). Receptive field mechanisms of cat X and Y retinal ganglion cells. *The Journal of General Physiology*, 74(2):275–298.
- Walter, P., Szurman, P., Vobig, M., Berk, H., Ludtke-Handjery, H., Richter, H., Mittermayer, C., Heimann, K., and Sellhaus, B. (1999). Successful long-term implantation of electrically inactive epiretinal microelectrode arrays in rabbits. *Retina*, 19(6):546–552.
- Wandell, B. A. (1995). *Foundations of Vision*. Sinauer Associates, Sunderland, Massachusetts, USA.
- Warland, D. K., Reinagel, P., and Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, 78(5):2336–2350.

## Bibliography

- Warren, D. J. and Normann, R. A. (2003). *Visual Neuroprosthesis*. CRC Press, Boca Raton, Florida, USA.
- Weiland, J. D. and Humayun, M. S. (2008). Visual prosthesis. *Proceedings of the IEEE*, 96(7):1076–1084.
- Weiland, J. D., Liu, W., and Humayun, M. S. (2005). Retinal prosthesis. *Annual Review Biomedical Engineering*, 7:361–401.
- Weiss, P. (2001). The seeing tongue. *Science News*, 160(9):140.
- Werblin, F. and Roska, B. (2007). The movies in our eyes. *Scientific American*, pages 73–79.
- Westwick, D. T. and Kearney, R. E. (2003). *Identification of Nonlinear Physiological Systems*. IEEE Press Series on Biomedical Engineering. IEEE Press, Piscataway, New Jersey, USA.
- Wickelgren, I. (2006). A vision for the blind. *Science*, 312:1124–1126.
- Wilke, S. D., Thiel, A., Eurich, C. W., Greschner, M., Bongard, M., Ammermüller, J., and Schwegler, H. (2001). Population coding of motion patterns in the early visual system. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, 187(7):549–558.
- World Health Organization (2004). Magnitude and causes of visual impairment. Technical Report Fact Sheet n° 282, United Nations Organization.
- World Health Organization (2006). Sight test and glasses could dramatically improve the lives of 150 million people with poor vision. Technical Report Fact Sheet n° 282, United Nations Organization.
- World Health Organization (2007). Vision 2020 - the right to sight.global initiative for the elimination of avoidable blindness: action plan 2006-2011. Technical report, United Nations Organization, Geneva.
- World Health Organization (Version for 2010). International statistical classification of diseases and related health problems. Technical Report 10th revision, United Nations Organization, Geneva.
- Wu, N. J. (2006). Bionic eye. <http://www.svec.uh.edu/BIONIC.html>.
- Wulf, M. (2001). *On Modeling the Spatiotemporal Processing Characteristics of the Retina: What is the Retina for?* PhD thesis, Universität Hamburg.
- Wyatt, J. and Rizzo, J. (1996). Ocular implants for the blind. *IEEE Spectrum*, pages 47–53.

- Wyatt, J. and Rizzo, J. (2006). Development of a wireless first generation boston retinal implant subretinal prosthesis. In *Proceedings of The Eye and the Chip - 2006 World Congress on Artificial Vision*.
- Yu, C., Roy, R., Kaufman, H., and Bequette, B. (1992). Multiple-model adaptive predictive control of mean arterial pressure and cardiac output. *Biomedical Engineering, IEEE Transactions on*, 39(8):765–778.
- Yu, H.-H. and de Sa, V. R. (2004). Nonlinear reverse correlation with synthesized naturalistic noise. *Neurocomputing, Computational Neuroscience: Trends in Research 2004*, 58–60:909–913.
- Ziemer, R. E., Tranter, W. H., and Fannin, D. R. (1998). *Signals and Systems: Continuous and Discrete*. Prentice-Hall, New York, USA, 4th edition.
- Zrenner, E. (2002). Will retinal implants restore vision? *Science*, 295(5557):1022–1025.
- Zrenner, E., Besch, D., Bartz-Schmidt, K. U., Gabel, F. G. V.-P., Kutteneuler, C., Sachs, H., Sailer, H., and Wilke, B. W. R. (2006). The active subretinal implant: 10 years of development to clinical application. In *104th DOG Annual Meeting*.
- Zrenner, E., Miliczek, K., Gabel, V., Graf, H., E.Guenther, Haemmerle, H., Hoeflinger, B., Kohler, K., Nisch, W., Schubert, M., Stett, A., and Weiss, S. (1998). The development of subretinal microphotodiodes for replacement of degenerated photoreceptors. *Ophthalmic Research*, 30(3):269–280.