

Giovanni Naldi
Thierry Nieus *Editors*

Mathematical and Theoretical Neuroscience

Cell, Network and Data Analysis



Springer

Springer INdAM Series

Volume 24

Editor-in-Chief

G. Patrizio

Series Editors

C. Canuto

G. Coletti

G. Gentili

A. Malchiodi

P. Marcellini

E. Mezzetti

G. Moscariello

T. Ruggeri

More information about this series at <http://www.springer.com/series/10283>

Giovanni Naldi • Thierry Nieus
Editors

Mathematical and Theoretical Neuroscience

Cell, Network and Data Analysis



Springer

Editors

Giovanni Naldi
Environment Science and Policy
University of Milan
Milan, Italy

Thierry Nieus
Department of Biomedical
and Clinical Sciences
University of Milan
Milan, Italy

ISSN 2281-518X

ISSN 2281-5198 (electronic)

Springer INdAM Series

ISBN 978-3-319-68296-9

ISBN 978-3-319-68297-6 (eBook)

<https://doi.org/10.1007/978-3-319-68297-6>

Library of Congress Control Number: 2018931066

© Springer International Publishing AG, part of Springer Nature 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Neuroscience is becoming increasingly quantitative and the need for theoreticians interested in collaborating with experimental neuroscientists will only increase in the coming years. In addition to a need for modeling, the kinds of problem which arise in neuroscience applications are mathematically interesting in their own right. Mathematical neuroscience here means an area of neuroscience where mathematics is the primary tool for elucidating the fundamental mechanisms responsible for experimentally observed behavior. As an example, of primary interest to neuroscientists are the roles of the highly nonlinear intrinsic properties of individual neurons and coupling properties between cells, which determine the dynamical activity of neuronal networks. Unique mathematical features of cell-based neuronal models that underlie complex spatiotemporal patterns are the multiple and vast time scales of ionic currents (from milliseconds to seconds) and the circuit properties, which include local as well as long-range coupling, possibly with random connectivity and with specific dynamics (i.e., time constants for decay of synaptic variables). The crucial modeling idea is to represent the electrical properties of biological membranes using an equivalent circuit consisting of capacitors and resistors in parallel. Some of these models can be simplified using phase-plane based methods and the dynamics on the plane can be further reduced to a scalar dynamical system on a circle. Simulations of the reduced model can then be used to explain the spiking statistics of single neurons driven by noisy stimuli. In this regard, the mean field approximation represents a fundamental approach that, by reducing the networks of excitatory and inhibitory neurons with defined connectivities, can account for experimental results obtained from recordings up to the entire brain.

This volume gathers contributions reflecting the topics presented during the INDAM Meeting held in Cortona (Italy) in September 2016. The Meeting brought together theoreticians, experimentalists and computational neuroscientists interested in combining different forms of expertise with their own research areas. In particular, the following aspects were covered during the meeting:

- Mathematical modelling in neuroscience, analytical and numerical aspects;
- Statistical analysis in neuroscience;

- Biological neural networks;
- Theoretical neuroscience.

Some contributions highlight new achievements and results, while others offer snapshots of the “state of the art” in certain topics. The book offers a unique resource, both for young researchers interested in quickly entering a given area of application and for more experienced ones seeking comprehensive overviews and extensive bibliographic references.

The first group of contributions in the volume deals with different types of biophysical model and the development of biological neural networks with their theoretical study. These works involve the study and simulation of local field potentials, cerebellar networks, neural field formalism, analysis from the dynamical system point of view, the development of a model of the neural system, and mathematical analysis of the networks related to visual perception. A second group of contributions focuses on topics in Neuro-informatics. These works involve the description of a project about autonomous rehabilitation and a big data-driven approach to the study of Alzheimer’s disease. A third group relates to the analysis of neural signals. These works present a set of metrics for spike train analysis and a space-by-time decomposition approach to network activities. The section also includes two overview chapters about nonlinear time series analysis and the inverse problem of MEG/EEG data.

The editors are grateful to Springer for the opportunity offered by the publication of this volume on these challenging research topics.

Milan, Italy
February 11, 2018

Giovanni Naldi
Thierry Nieus

Contents

From Single Neuron Activity to Network Information Processing: Simulating Cortical Local Field Potentials and Thalamus Dynamic Regimes with Integrate-and-Fire Neurons	1
Alberto Mazzoni	
Computational Modeling as a Means to Defining Neuronal Spike Pattern Behaviors	25
Siva Venkadesh and Giorgio A. Ascoli	
Chemotactic Guidance of Growth Cones: A Hybrid Computational Model	45
Iolanda Morana Roccasalvo, Silvestro Micera, and Pier Nicola Sergi	
Mathematical Modelling of Cerebellar Granular Layer Neurons and Network Activity: Information Estimation, Population Behaviour and Robotic Abstractions	61
Shyam Diwakar, Chaitanya Nutakki, Sandeep Bodda, Arathi Rajendran, Asha Vijayan, and Bipin Nair	
Bifurcation Analysis of a Sparse Neural Network with Cubic Topology ...	87
Diego Fasoli, Anna Cattani, and Stefano Panzeri	
Simultaneous Jumps in Interacting Particle Systems: From Neuronal Networks to a General Framework	99
Luisa Andreis, Paolo Dai Pra, and Markus Fischer	
Neural Fields: Localised States with Piece-Wise Constant Interactions ...	111
Aytül Gökçe, Stephen Coombes, and Daniele Avitabile	
Mathematical Models of Visual Perception Based on Cortical Architectures	123
Marta Favali, Giovanna Citti, and Alessandro Sarti	

Mathematical Models of Visual Perception for the Analysis of Geometrical Optical Illusions	135
Benedetta Franceschiello, Alessandro Sarti, and Giovanna Citti	
Exergaming for Autonomous Rehabilitation	151
Nunzio Alberto Borghese	
E-Infrastructures for Neuroscientists: The GAAIN and neuGRID Examples	161
Daniele Orlandi, Alberto Redolfi, Jérôme Revillard, David Manset, Stefan Teipel, and Giovanni B. Frisoni	
Theory and Application of Nonlinear Time Series Analysis	177
Angelo Di Garbo	
Measures of Spike Train Synchrony and Directionality	201
Eero Satuvuori, Irene Malvestio, and Thomas Kreuz	
Space-by-Time Tensor Decomposition for Single-Trial Analysis of Neural Signals	223
Ioannis Delis, Arno Onken, and Stefano Panzeri	
Inverse Modeling for MEG/EEG Data	239
Alberto Sorrentino and Michele Piana	

About the Authors

Giovanni Naldi studied mathematics at the University of Pavia and at the University of Milan, where he also received his Ph.D. in applied mathematics. He is currently a full professor of numerical analysis at the University of Milan and the director of the ADAMSS (ADvanced Applied Mathematical and Statistical Sciences) Center at the same University. His research work mainly focuses on the numerical analysis of partial differential equations, wavelet-based methods, multiscale models, nonlinear evolution phenomena, biomathematics, and computational neuroscience. He has supervised eight doctoral theses and is the author of more than 60 papers.

Thiery Nieus received his Ph.D. in applied mathematics at the Department of Mathematics “F. Enriques” in Milan (Italy). His research focuses on the computations performed by neuronal networks. His work involves the analysis and modeling of multiscale data, ranging from single synapses to population recordings. In September 2016, he joined Marcello Massimini’s laboratory (University of Milan, Italy), working on computational models of the thalamocortical circuit and on complexity measures of TMS/EEG data.

From Single Neuron Activity to Network Information Processing: Simulating Cortical Local Field Potentials and Thalamus Dynamic Regimes with Integrate-and-Fire Neurons



Alberto Mazzoni

Abstract Cortical and subcortical neural activity has been modeled for decades by means of recurrent networks of integrate and fire neurons (IFN). Such networks have proved to be able to capture a variety of neural functions ranging from sensory processing to short term memory and decision making. Some network scale phenomena, however, were considered too complex to be simulated with simple basic components as IFN. Namely, Local Field Potentials (LFP) were simulated with multi-compartmental models, as IFN lacked the ability to cope with the spatial features of the signal, and thalamic oscillations were modeled with Hodgkin and Huxley and similar neuron models, as IFN were thought not be able to mimic the rebound properties of the neurons of the area. Here we will show instead how it is possible to capture both phenomena with IFN networks. First, in a series of works spanning almost a decade, we were able to build flexible LFP proxies simulating the extracellular signal from IFN network dynamics. We identified several key properties of the neuronal structure, based on the dipolar approximation of LFP generation, and we implemented them in the IFN model. This led to a dimensionality reduction of the system parameters of two order of magnitudes, while capturing almost entirely the LFP dynamic. Actually, the resulting model of LFP is so efficient to be able to reconstruct spiking dynamics from LFP recorded in the primary visual cortex. Second, we were able to reproduce with IFN two dynamic regimes of the thalamus displaying different characteristic oscillations bands and different information processing properties. The key advancement was to highlight the connection between single neuron dynamics and network regimes. Spindle oscillations preventing information transmission to the cortex occur when the neurons fire exclusively because of rebound from inhibition. When the external excitatory inputs are sufficiently strong, instead, thalamocortical relay neurons start firing because of depolarization, weakening spindle oscillations and leading to information transmission. Starting from these results we will discuss the possibility

A. Mazzoni (✉)

The Biorobotics Institute, Scuola Superiore Sant'Anna, Pontedera, Pisa, Italy

e-mail: alberto.mazzoni@santannapisa.it

of capturing the whole process of encoding of sensory stimuli by means of a chain of spiking neuron model covering the whole path from peripheral sensors to primary sensory cortex. Building the whole structure with simple and elegant IFN will make the analysis and the comparison with experimental results sharper.

1 The Map and the Territory

The number of neuronal models available to the computational neuroscientist is striking. From binary cellular automata to morphological models, through mean field and spiking neuron models, the computational neuroscientist can select any desired level of realism and pick the single neuron features needed for the specific function to be modeled.

The rule of thumb should be to pick the simplest neuron model able to reproduce the phenomenon under investigation, but identifying such neuron model is usually part of the investigation. This might explain the success of flexible models as IFN or Izhikevich models that, beside representing an excellent tradeoff between efficiency and realism [1] can be tuned according to the investigation needs. The research described in this chapter focus on IFN models. A comprehensive introduction to the model goes beyond the scope of this text, but the reader not familiar with the model can find several complete reviews [2] and textbooks [3]. Here we will highlight three key properties of IFN:

Modularity

The starting point is a simple combination of one equation and one condition describing the basic Leaky Integrate and Fire model [4]

$$\tau_m \frac{dV}{dt} = -(V - V_L) + I_{syn} \quad (1)$$

Where V is the membrane potential, τ_m is the membrane time constant, V_L is the Leak equilibrium potential, and I_{syn} is the total contribution of synaptic currents to the potential. Equation (1) is complemented with the extra condition that if $V > V_{thr}$ then a spike is generated and V is held at a given reset value for a given refractory time.

Then for instance the scientist can decide whether to modify add an exponential or quadratic term $\Phi(V)$ to Eq. (1) in order to take into account action potential shape [5]

$$\tau_m \frac{dV}{dt} = -(V - V_L) + \Phi(V) + I_{syn}$$

Another option is to introduce adaptation to obtain the widely used aeIF [6]

$$\begin{aligned} \tau_m \frac{dV}{dt} &= -(V - V_L) + \Phi(V) + I_{syn} - w \\ \tau_w \frac{dw}{dt} &= -w + a(V - V_L) \end{aligned} \quad (2)$$

Where w is the adaptation current, having characteristic time τ_w , and a defines the strength of subthreshold adaptation. At each firing time, the variable w is increased by an amount b , which accounts for spike-triggered adaptation.

Finally, in defining I_{syn} , one can chose whether current or conductance synapses are to be preferred

$$CURR : I_{syn} = J_{syn} S(t) \sum_{t^*} \delta(t - t^*)$$

$$COND : I_{syn} = g_{syn} (V(t) - V_{syn}) S(t) \sum_{t^*} \delta(t - t^*)$$

Where the sum over δ simply identifies spike times. In the current model J_{syn} defines synaptic strength and $S(t)$ the time evolution of synaptic potentials (usually characterized as a difference of exponentials). In the conductance model g_{syn} defines a conductance value but the actual amplitude of the post-synaptic potential depends on the distance between current value of membrane potential and V_{syn} , the reverse potential of the synapse. See [7] for the relevant consequences of this last choice.

This is just a quick overview of several versions of IF neurons, but note that IF neurons can be modified to include virtually any other feature requested, as for instance short term depression due to synaptic depletion [8].

Availability of Analytical Solutions

Analytical mean field solutions are available for most networks of LIF current neurons [9]. These kind of solutions are particularly quick to simulate and have a limited set of parameters which allow (i) achieving a clear, explicit description of the relationship between such parameters and the dynamics of the network [10] (ii) reconstructing a quantitative fit of network dynamics form a limited number of experimental data, as detailed in Sect. 2.2.

Hardware Implementation

Neural models main purpose was for decades to reproduce the observed dynamics of single neurons and neural networks in order to explain them and predict phenomena yet to be tested. This simulations were mainly performed in software environments. Recently, a novel field of research, sometimes called neuro-robotics, emerged, in which spiking neuron models are implemented in hardware [11].

Three are the main reasons for these studies, which follow the “three missions” of biorobotics [12]. First, neuromorphic hardware device are developed to reproduce neural networks characteristic qualities as high energetic efficiency, robustness to noise and so on, which can be used also for processing non-biological information (as for instance in image recognition and data mining). Second, in the field of neuroprostheses, neuromorphic devices should generate stimulation patterns closely mimicking spike patterns generate by biologic sensors (e.g. mechanosensors in the fingertip). Whether this leads to an increase in decoding performance is still debated, but preliminary results suggest that patients learn to discriminate these stimulation

patterns much faster than continuous stimulations mimicking the evolution in time of the sensory stimuli. Last but not least using the stimulation patterns generated by neuromorphic devices can be used to test hypothesis about neural coding [13].

As a matter of fact, most of these neuromorphic devices are based on IFN, even those that are not designed to perform tasks mimicking biological functions [11]. Hence, even in hardware neuron simulations IFN seem to be the optimal trade-off between realism and simplicity.

For all the aforementioned reasons, the simple model invented more than one century ago by Lapicque [14] seems to be a crucial tool for present and future studies and applications in neural computation. Here, we try to push the boundaries of the applications of such models in two ways. First, showing how IFN neurons can be properly used to simulate Local Field Potentials, and then to reproduce and understand network dynamics observed through this signal and reconstruct from it microscopic activity [15–19]. Second, designing and testing an efficient model of thalamic oscillations with IFN [20], paving the way to build a chain of IFN networks modeling of the whole sensory pathway from peripheral mechanosensors to primary sensory cortex.

2 Simulating Local Field Potential with Integrate and Fire Neurons

2.1 Problems and Solutions

The Local Field Potential is operatively defined as the low passed (<500 Hz) signal recorded by extracellular electrodes located inside the cortex [21]. For decades this component of the extracellular signal was discarded or neglected in the analysis, as the main focus of the research was the high frequency (>500 Hz) component containing action potentials. At the end of the century, however, studies pointed out that also the LFP was information rich, and that this information was complementary to the one carried by spikes [22]. Moreover, as detailed in the introduction, the LFP is relatively more easy to record than spikes. The last years, then, witnessed a rise in high-profile neuroscience studies based on LFP analysis.

Crucially, network computational studies lagged behind, as no standard method to predict LFP from spiking neuron models was devised. Few IFN studies provided LFP estimates to be compared with experimental results and they mostly assumed that the LFP corresponded to the average membrane potential [23] or the global firing rate [10]. The lab of Prof. Einevoll in Aas University of life sciences did an amazing series of studies simulating the LFP generated by synaptic activity of accurate morphological models of different neurons [24] and released LFPy, a Python Toolbox for neuron to perform such simulations [25]. These studies showed how computational studies could lead to dramatic advancements in the understanding of the origin of extracellular signals.

2.2 Combining Integrate-and-Fire Neurons and Morphological Models

In 2008, a seminal paper by Belitski et al. [22] showed that delta and gamma bands of LFP in the primary visual cortex of anesthetized monkeys presented with movies carried independent information about movie scenes. The mechanism of such information encoding, and the relationship between the information and the different movie features, however, were unclear.

In order to shed light on these issues, we decided to model a single layer of primary visual cortex (V1) as an integrate and fire network which was known to display gamma band oscillations [10]. Full details of the network can be found in [18]. Briefly, the network was composed by 1000 IFN with inhibitory synapses (interneurons), 4000 IFN with excitatory synapses (pyramidal neurons) sparsely connected and injected with thalamic signals as well as cortical noise (Fig. 1a).

The key question was: once we know what each neuron is doing, what would the LFP be? We thought that simulating the LFP of a single-point neurons would have been insufficient then we introduced a basic version of two spatial components: neuron morphology and network architecture (Fig. 1b).

We hypothesized that, as GABA synapses are perisomatic and AMPA synapses are apical, their dipoles across dendrites will sum. Moreover, we assumed that, as pyramidal neurons in V1 have parallel dendrites their dipoles will sum, while, as interneuron dendrites are arranged almost isotropically, their global dipole over a sufficiently large volume goes to zero. This leads to a very simple LFP proxy for

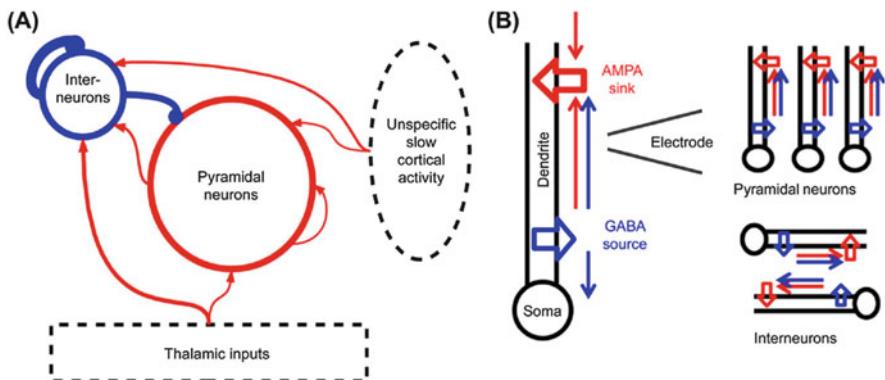


Fig. 1 (a) Schematic of network structure. Figure reproduced with permission from Mazzoni et al. [19]. (b) Schematic of the computation of the simulated LFP. Left side: AMPA synapses are usually apical and GABA synapses are usually peri-somatic and thus their dipoles sum with the same sign along the dendrite. Right side: currents from synapses of pyramidal neurons, due to their approximate open field arrangement, do contribute to LFP more than interneurons, which have a much less regular dendritic spatial organization. Figure reproduced with permission from Mazzoni et al. [16]

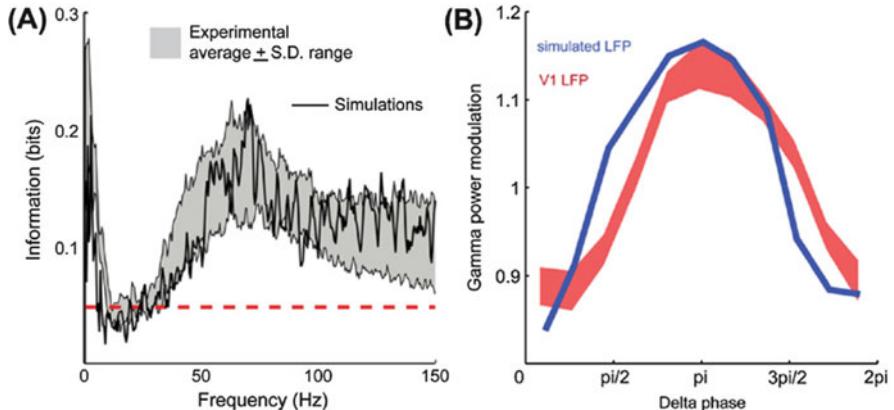


Fig. 2 (a) Information about naturalistic visual stimulus carried by the LFP power for both experimental recordings (gray area, representing the mean \pm STD area across the different channels) and simulations (black line). The red dashed line represents the $p = 0.05$ (bootstrap test) significance line of information values. Panel reproduced from Mazzoni [18] with permission. (b) Cross frequency phase-power coupling in experiment and simulations. The plot shows the modulation of gamma amplitude as a function of delta phase in the LFP recorded from primary visual cortex (red shaded area representing mean and SEM across recording sessions) compared to the modulation of the simulated LFP obtained in response to a naturalistic input (blue line, mean across all simulation points). Panel modified from Mazzoni et al. [19] with permission

IFN networks

$$LFP = \sum_{i \in pyr} |AMPA_i| + |GABA_i|$$

In particular, in Mazzoni et al. [18] we showed that the LFP obtained from this proxy combined with the IFN network displayed the same spectral information as the LFP actually recorded from monkeys' V1 cortex (Fig. 2a). Briefly, we also showed that this was due to the fact that the low frequency component of the LFP spectrum encoded the low frequency component of the thalamic input, while the gamma component of the LFP encoded the local mean value of the input. From these two simple rules we were able to predict Delta phase Gamma power cross-frequency coupling, that was later found in anesthetized and awake monkeys (Fig. 2b) [19].

As we mentioned in the introduction, a crucial open issue of LFP studies is to understand how the observed LFP relates to the underlying neural activity. In Mazzoni et al. [19], indeed, we used our IFN LFP model to improve existing algorithms predicting neurons firing activity based on LFP spectrum. In a later work [15] we used the LFP proxy introduced above and the aforementioned timescale separation to predict excitatory and inhibitory firing rates and even some features of the visual stimuli presented. Briefly, we derived a mean field equation describing LFP spectrum from 10 parameters: four synaptic strengths, first and second order of excitatory and inhibitory firing rate, and amplitude and characteristic time of input (Fig. 3). The values of these parameters were estimated for each experimental session, taking into account that synaptic strengths are network properties that are

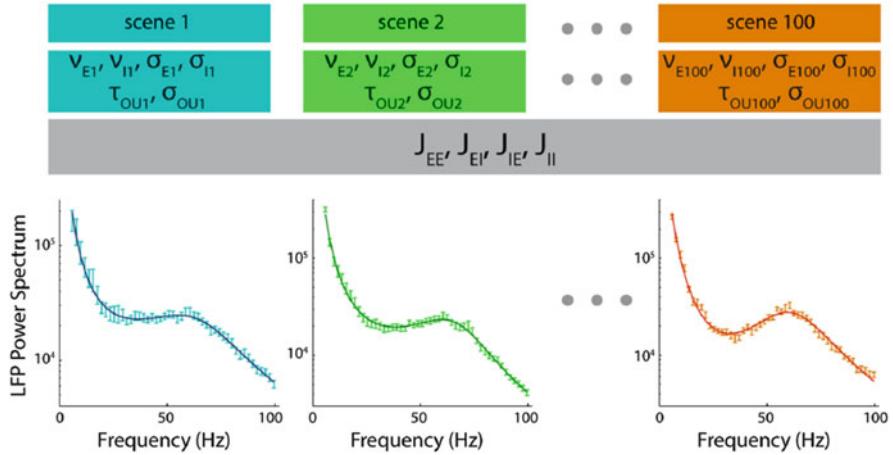


Fig. 3 Fitting procedure. For each of the 100 scenes of a recording, the six scene-dependent parameters (v_E , v_I , σ_E , σ_I , σ_{OU} , τ_{OU}) were determined using the algorithm described in the text. The values of the synaptic efficacies (J_{EE} , J_{EI} , J_{IE} , J_{II}) were the same for all scenes and equal to the median values found in the first step of the fitting procedure. Therefore, for each scene, we evaluated six parameters using 50 values of the LFP power spectrum at the different frequencies. Reproduced from [15] with permission

constant through the whole session, while input and consequently firing rates varied for each scene.

First we aimed at reconstructing firing activity from LFP. This is particularly important given the fact that LFP is much easier to record than firing activity. For instance, in chronic implants for clinical purposes LFP recordings are much more stable than recordings of spiking activity. Hence it is important to be able to reconstruct the later from the former. We found that predicted firing rates and observed Multi-Unit Activity (MUA) correlated remarkably well as shown in Fig. 4, in particular in the third line of Fig. 4c showing the strong correlation between the total MUA and the weighted sum of excitatory and inhibitory firing rates. This means that we are able to give a robust estimate of the firing activity underlying a given LFP spectrum.

We tested then the ability of our system to estimate the input to the local network from LFP spectra. Even though we were recording from the monkey primary visual cortex, and not from the periphery, we aimed at reconstructing the properties of the images presented to the monkeys. We found that the estimated size of input fluctuations correlated very well with image features and in particular with the temporal contrast of the images presented to the monkey (Fig. 4d).

To be make use of the LFP of a primary visual cortex network to estimate its underlying firing rate and even some properties of the presented images is in our opinion an interesting result per se. What we would like to stress is that this was achieved with an analytical solution of a simple IFN network combined with a simple IFN to LFP proxy. This suggest that our model captures the key elements of the relationship between LFP, neuron activity and stimuli.

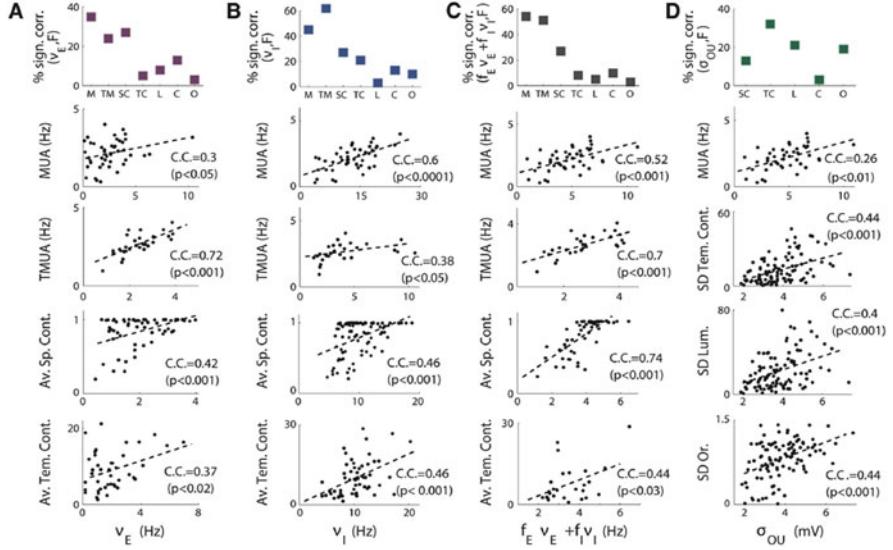


Fig. 4 Correlations between the parameters extracted from fit and the experimental observables. **(a)** Top, Percentages of recordings with significant correlation between the excitatory firing rate found through the fits and the experimental observables on the x-axis: MUA of a single channel (M); MUA of all the channels of a single experimental session (TM); average spatial contrast (SC), average temporal contrast (TC), luminance (L), orientation (O), and color (C). On the other four panels are displayed some examples in which these correlations were found. **(b)** Same as in **(a)**, but for correlations with inhibitory firing rates extracted from fits. **(c)** Same as in **(a)**, but for correlations with total firing rate. **(d)** Same as in **(a)**, but for the correlations between the amplitude of the fluctuations of the OU process, OU, that was found through the fits and the SDs of the different features of the movie. Reproduced from [15] with permission

2.3 Combining IFN Networks and Morphological Simulations

The LFP model described in the previous section, however, was inherently limited, as could not answer questions about, for instance, attenuation of LFP intensity with distance from the source, the effect of different neuron's morphologies on the LFP and such issued depending on the spatial architecture of single neurons and neuron networks. Moreover, it was based on two strong assumptions that are intuitive but not proved.

We thought then about a method to assess the ideal IFN LFP proxy capturing as much as possible the complexity of neuronal spatial organization by combining IFN dynamics with the ability of the LFPy simulator mentioned in the introduction [25] to accurately simulate the LFP generate by arbitrary neuron morphologies [17]. The main features of the IFN network, together with a summary of the main proxies are reproduced in Tables 1 and 2. The optimal proxy identification was performed in three steps: (i) simulation of network dynamics with the standard, well-tested IFN network (Fig. 5a), (ii) generation of a 3D network of realistic neuronal morphologies

Table 1 Summary of the structure of the IFN network used to test the LFP proxies (listed in G section of the table)

(A) Leaky integrate and fire model summary			
Populations		Excitatory, inhibitory	
Topology			-
Connectivity			Random and sparse
Neuron model			Leaky integrate and fire, fixed threshold, fixed refractory time
Synapse model			Difference of exponential functions defined by rise and decay time. Current-based synapses
Plasticity			-
Input			Sum of independent Poisson processes with same time-varying rate for all neurons
Measurements			For each population: firing rate, GABA and AMPA currents, membrane potential
(B) Populations			
Type	Elements	Pattern	
Interneurons (GABA synapses)	LIF neurons	dir. conn. Pdc weight: J _{GABA,cor/pr}	
Pyramidal neurons (AMPA synapses)	LIF neurons	dir. conn. Pdc weight: J _{AMPA,cor/pr}	
(C) Connectivity			
Name	Source	Target	Pattern
AMPA_cor-Pyr	Pyramidal	Pyramidal	dir. conn. Pdc weight: J _{AMPA,cor/pr}
AMPA_cor-Inter	Pyramidal	Interneuron	dir. conn. Pdc weight: J _{AMPA,cor/inter}
GABA-Pyr	Interneuron	Pyramidal	Dir. conn. Pdc weight: J _{GABA/pyr}
GABA-Inter	Interneuron	Interneuron	dir. conn. Pdc weight: J _{GABA/inter}
AMPA_th-Pyr	External	Pyramidal	Uniform, J _{AMPA,th/pyr}
AMPA_th-Inter	External	Interneuron	Uniform, J _{GABA,th/pyr}
(D) Neuron			
Type	Leaky integrate and fire		
Description	Subthreshold dynamics: $I_m V_m(t) = -V_m(t) = + \sum P S C_{syn}(t)$		
	If $V_m > V_{th}$ & $t > t^* + \tau_{refractory}$		
	\{t^* = t; \text{spike emitted with time stamp } t^*; V_m = V_{reset}\}		

(continued)

Table 1 (continued)

(E) Synapse	
Type	Current synapse
Description	$\tau_{dsyn} PSC(t) = -PSC(t) + x(t)$ $I_{psyn}x(t) = -x(t) + \tau_m \left(J_{syn} \sum_{sym} \delta(t - t_{sym} - \tau_t) \right)$
(F) Input	
Type	Description
Poisson	“Thalamic”: time-constant input with rate λ
Poisson	“Long range cortico-cortical”: Ornstein Uhlenbeck process (OU) with zero mean $\tau_n OU = -OU + \sigma/(2\tau_n)W$ where W is a white noise process with zero mean
(G) Measurements (for each population)	
Type	Description
Firing rate	Sum of spikes
AMPA	Sum of AMPA PSCs (cortical and thalamic)
GABA	Sum of GABA PSCs
V_m	Mean of membrane potential
Σ	Sum of AMPA and GABA PSCs. Note that AMPA and GABA have opposite signs
$\sum I $	Sum of absolute values of AMPA and GABA PSCs
Weighted sum (WS)	$\left[\sum_{pyr} AMPA(t - \tau_{AMPA}) - \alpha \sum_{pyr} GABA(t - \tau_{GABA}) \right]$
Reference weighted sum (RWS)	$\left[\sum_{pyr} AMPA(t - 6\ ms) - 1.65 \left(\sum_{pyr} GABA(t) \right) \right]$

Reproduced with permission from [17]

Table 2 Parameters of the IFN network used to test LFP proxies

Leaky integrate and fire model parameters	Pyramidal neurons	Interneurons
Population		
Size	4000	1000
Connectivity		
P _{dc}	0.2	0.2
Neuron		
V _{thr} (mV)	18	18
V _{reset} (mV)	11	11
τ _m (ms)	20	10
τ _{refractory} (ms)	2	1
Synapse		
τ _{rGABA} (ms)	0.25	0.25
τ _{dGABA} (ms)	5	5
τ _{rAMPA} (ms)	0.4	0.2
τ _{dAMPA} (ms)	2	2
J _{GABA} (mV)	-1.7	-2.7
J _{AMPA-cort} (mV)	0.42	0.7
J _{AMPA-th} (mV)	0.55	0.95
Inputs		
Thalamic input (spikes/ms)	[0.5:0.5:3, 6]	[0.5:0.5:3, 6]
OU τ _n (ms)	16	16
OU σ (mV)	0.25	0.25

Reproduced with permission from [17]

closely reproducing not only the composition but even the connectivity matrix of the original IFN network, (iii) injection of the activity of the IFN network (Fig. 5d) into the 3D morphological network, generating the ground truth LFP, varying across depths ad depicted in Fig. 5e, (iv) design of a broad set of LFP proxies from IFN network parameters, (v) comparison of the ability of the proxies in reproducing the ground truth LFP (Fig. 6). We found that, not surprisingly, the performance of the two most used LFP proxies, i.e. the firing rate and the membrane potential, was relatively poor. The proxy used in Mazzoni et al. [18] was good, but could be improved by a delayed linear combination of the two synaptic variables.

$$LFP(t) = \sum_{i \in pyr} |AMPA_i(t - 6)| + 1.65 * |GABA_i(t)|$$

Notably, both the negligible contribution of LFP generated by interneurons and the correctness of the dipole approximation for synaptic distributions were confirmed by morphological simulations. We tested the robustness of the ideal proxy in different dynamical regimes, different morphologies and different synaptic distributions (Table 3). Regarding dynamic regimes and morphologies the result were the same: whenever a significant LFP was generated, the proxy was working

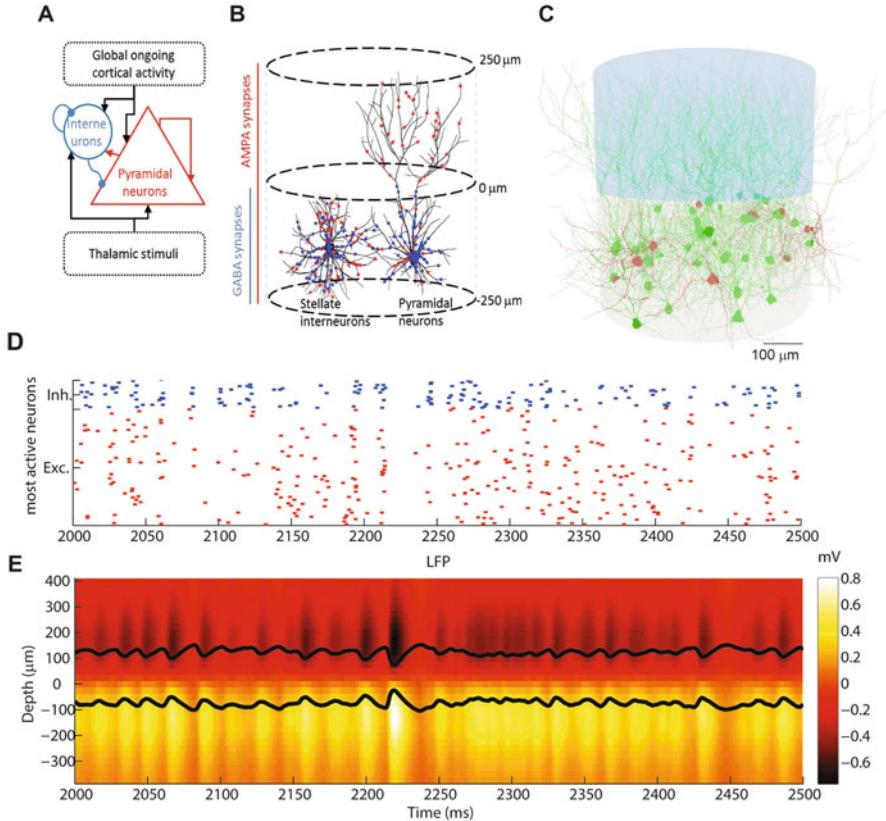


Fig. 5 LIF network and 3D morphological network. **(a)** Same network as Fig. 1a. **(b)** Sketch of the morphological 3D network made of two stacked cylinders with 250 μm radius and 250 μm height. A representative interneuron and pyramidal cell are depicted. Interneuron dendrites remain in the lower cylinder while the pyramidal neuron dendrites reach out to the upper cylinder. Dendrites in the lower cylinder receive both AMPA and GABA synapses while dendrites in the upper cylinder receive only AMPA synapses. **(c)** Graphical rendering of a subset of the 3D network composed of 10 interneurons and 40 pyramidal neurons. **(d)** Raster plot of the spiking activity of the 10 interneurons (blue, top) and the 40 pyramidal neurons (red, bottom) with the highest spiking activity in the LIF network, for a thalamic stimulation of 1.5 spikes/ms. **(e)** Depth-resolved LFP signal as simulated by injecting the spikes generated by the whole network during thalamic stimulation of 1.5 spikes/ms into the 3D network. Color shades represent evolution in time of LFP for each depth (see color bar). Black lines show LFP for 100 and $-100\mu\text{m}$ depth

very good. If the LFP goes to zero, be it because external stimulation is so low that recurrent activity is not triggered, or because the neuron dendritic morphologies does not lead to the formation of a dipole, the signal is dominated by local fluctuations that cannot be captured by any proxy.

The situation is completely different if the distribution of synapses over the surface of pyramidal neurons is altered: in this case the relative weight of AMPA

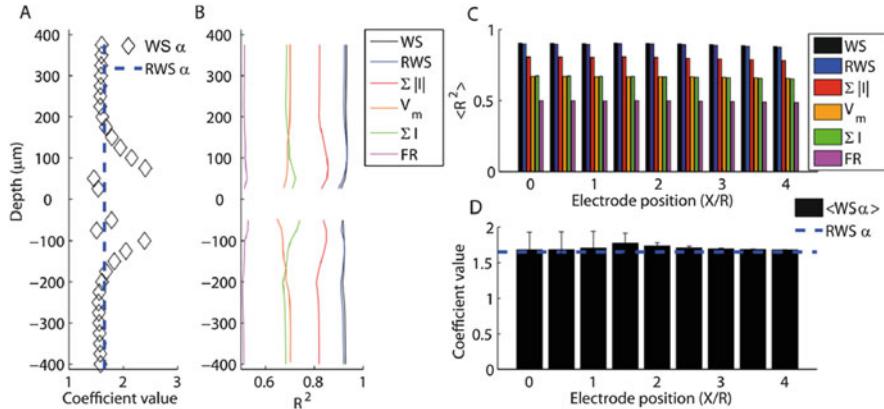


Fig. 6 (a) Value of relative contribution of AMPA and GABA currents optimizing correlation between WS proxy and ground truth LFP. Dashed line indicates average value over depths $\alpha = 1.65$ used for RWS proxy. (b) Fraction of LFP signal variance explained by different LFP proxies with optimal delay as a function of depth. (c) Fraction of LFP variance explained, averaged over all depths, by different proxies for different distances from the center of the 3D network. From best to worst: weighted sum (WS), reference weighted sum (RWS), sum of absolute values of the synaptic currents ($\sum |I|$), membrane potential (V_m), sum of synaptic currents ($\sum I$), firing rate (FR). Error bars are not displayed since they would not be visible in the figure. (d) Mean and standard deviation across depths of optimal coefficients of α in the WS proxy as a function of electrode lateral position. Dashed line indicates the fixed coefficient of the RWS proxy that was not optimized but kept constant for all depths and distances. Reproduced from [17] with permission

and GABA synapses must be adapted to the resulting dipole, and we can predict exactly in which way (see [17] for more details).

As summarized in Table 3, our work is able to provide an excellent LFP proxy for IFN network in a very broad range of situations. The main limitation of this work is not to take into account the effect of active channels in LFP generation. Not only the spike shape per se is known to “leak” into the high frequency LFP, but the after spike hyperpolarization is thought to play a significant role even at lower frequencies [21]. Future studies will focus on fixing this issue and determine how to take into account this in the determination of the optimal LFP proxy for IFN networks.

3 Integrate and Fire Neurons Model of the Thalamus

IFN networks have been used mainly to simulate the dynamics of cortical circuits, and only rarely to simulate peripheral nervous system and subcortical structures [26, 27]. Surprisingly, there is not even an established IFN model of the thalamus.

Briefly, the thalamus is at the same time a relay and a gate for sensory stimuli: neurons from peripheral sensors project directly (the retina) or through other steps (mechanoreceptors) to thalamic neurons which project to the primary sensory

Table 3 Summary of results for proxies and suggested adaption of results to other situations

Reference table for LFP proxy use	Action required
For the following reference conditions	
• Layer 2/3 pyramidal cell (two neighboring dendritic bushes (Fig. 1b, left)	Model temporal part of LFP as sum of AMPA and GABA currents or simply as GABA currents (Fig. 4f, g)
• Synaptic distribution with GABA synapses in lower (basal) bush and AMPA synapses in both (basal + apical) bushes (Fig. 9a, central panel)	Use the RWS (Fig. 5c) unless you are within 50 μm from the inversion point, in which the LFP largely cancels out (Fig. 2a, b)
• Current-based synapses	Use the RWS and determine amplitude $f_{RWS}(r, d)$ by means of Fig. 2b
• 1.5 spikes/ms input	Use the RWS (as long as the synaptic input is sufficiently strong to generate a sizable LFP, cf. Fig. 6c)
• Electrode recording from the center of the 3D network at depth = 100 μm relative to inversion point	Use the same RWS, since interneurons have a negligible effect on the LFP (Fig. 3)
The optimal proxy (RWS) is	
$LFP_{RWS}(r, d, t) = f_{RWS}(r, d) * \text{Norm} \left[\sum_{pyr} \text{AMPA}(t - 6 \text{ ms}) - 1.65 \left(\sum_{pyr} GABA(t) \right) \right]$	
The proxy is, however, robust or easily adaptable to a variety of conditions as described below	
Aim	
Use a simpler proxy	
Consider a different recording depth	
Simulate LFP recorded with electrode radially displaced from the center of the 3D network	
Vary rate of external synaptic input (input intensity)	
Include/remove LFP contribution from transmembrane currents of stellate interneurons	

Simulate neurons with morphologies different from pyramidal cell of layer 2/3

Use the RWS for all morphologies in which two dendritic bushes are vertically more distant than 150 pm (i.e., for all cells except stellate cells that do not contribute to LFP), see Fig. 8e

Simulate neurons with synaptic distributions different from our reference case

As shown in Fig. 9f, g

- Both synapses in both bushes: no appreciable LFP and no good proxy available

- AMPA synapses only in upper bush: RWS performs well ($R^2 = 0.81$), but to get better results increase relative weight of AMPA currents as follows:

$$LFP_{RWS}(r, d) = f_{RWS}(r, d)$$

$$* Norm \left[\sum_{PV} AMPA(t - 6 \text{ ms}) - \left(0.69 \sum_{PV} GABA(t) \right) \right]$$

- AMPA synapses only in lower bush: RWS performs well ($R^2 = 0.77$)

Simulate neurons with conductance synapses
Change AMPA and GABA coefficients as a function of input rate

Figure indications refer to original paper. Reproduced with permission from [17]

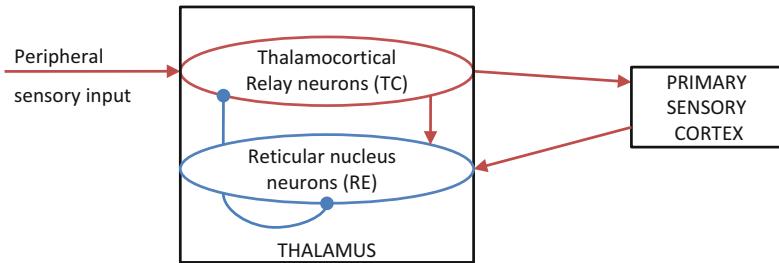


Fig. 7 Abstract representation of functional connectivity of the thalamus. Red lines indicate AMPA synapses, blue lines GABA synapses

cortex. The population of thalamic neurons projecting into the cortex is called thalamo-cortical relay neurons (TC neurons): they have excitatory AMPA synapses and they do not form TC-TC connections, but they do only form local connections with neurons in the Reticular nucleus of the thalamus (RE neurons). RE neurons have inhibitory GABA synapses. Beside TC-RE and RE-TC connections, they also form RE-RE connections and receive inputs from the layer 6 of the primary sensory cortex. Figure 7 illustrates these functional connections.

The interplay between excitatory transmission to the cortex and inhibitory feedback from the cortex to the thalamus is thought to play a role in the mechanism of attention and stimuli salience. Moreover, this interplay is at the core of spindle oscillations: a bursting activity in the 7–15 Hz frequency range characterizing the activity of the thalamus during sleep and preventing sensory information from reaching the cortex in that phase.

In vitro studies [28, 29] have revealed that thalamic neurons display two different firing modes. Near resting membrane potential, TC neurons fire with frequency proportional to the amplitude of the input current, due to voltage-dependent currents that generate action potentials [30]. This is usually called tonic mode. When TC neurons are hyperpolarized, on the contrary, they fire in a bursting mode, characterized by high-frequency bursts of action potentials (300 Hz). Spindles are then originated as follows: TC bursts depolarize RE cells. When RE start firing, their inhibitory inputs hyperpolarize TC cells, which consequently stop firing. When RE cells, lacking excitatory drive, stop firing too, TC cells fire due to the rebound from hyperpolarization and the cycle starts again. The overall process takes about 100 ms and generates rhythmic spindle oscillations.

3.1 Thalamic Neurons Modeling

The properties of rebound currents cannot be properly captured by the standard Leaky IFN, hence most studies to date have adopted neuron models at least as complex as the Hodgkin and Huxley model [31, 32]. Such thalamic models have

been investigated for more than 20 years [31, 33], and were able to capture a wealth of thalamic phenomena [34]. However, as motivated in the introduction, we do think that modeling brain circuits with IFN can be useful for a number of reasons. In particular, as both primary sensory cortex circuits (see above) and peripheral sensors [27, 35], having a IFN model of the thalamus would allow an homogeneous modeling of the whole path of sensory information from peripheral to central nervous system.

We are aware of only one attempt to model realistically thalamic interactions with integrate-and-fire (IF) neurons [36]. However, in our tests these model was extremely sensitive to the choice of the parameter, and was hence not suited for our purposes. We decided then to develop our own IFN model of the thalamus [20].

3.2 Integrate-and-Fire Model of the Thalamus Reproduces Sleep/Wake Information Processing Transition

We selected aEIF neuron models to capture the single neuron dynamics of thalamic neurons. We tuned adaptation parameters such in a way to obtain that RE and TC neurons adapt respectively quickly and slowly to incoming excitatory inputs (Fig. 8a–b, e–f) and both fire a short burst of spikes as a result of rebound from inhibition (Fig. 8c–d, g–h). We determined the range of synaptic strengths for which TC-RE and RE-RE cycle appear, we characterized them in terms of frequency and intensity of oscillations, and we identified the conditions of synchronization of heterogeneous cycles [20].

Under weak assumptions about neuronal characteristic timescales we found that these cycles operate in the range of spindle oscillations frequencies observed in animals. Crucially, these oscillations were stable at the network level only if a non-zero level of clustering of RE-RE connections was introduced [20].

Once both single neuron activity and network architecture were tuned such in a way to obtain self-sustained spindle oscillations in absence of sustained external input, we found that the network became also able to reproduce the key features of thalamic sensory processing. We studied the level of activity of both kinds of neurons as a function of the input received by TC neurons. The network displayed indeed two information processing regimes: below a critical level of the input both TC and RE neurons are weakly affected by the level of the input, while above that RE neurons become completely insensitive and TC neurons firing rate increases proportionally to the input (Fig. 9a). The shift in sensitivity of TC neurons, and hence the shift in information transmitted to the cortex, qualitatively matches the transition between sleep and wake: you don't wake up if your neighbors talk quietly, you do if they throw a party. Moreover, coherently with experimental observations, sensory information during wake is carried by TC neurons (Fig. 9d).

Our model, however, did not simply reproduce experimental results, but was able to link them with the underlying activity at the level of single neuron. We found that

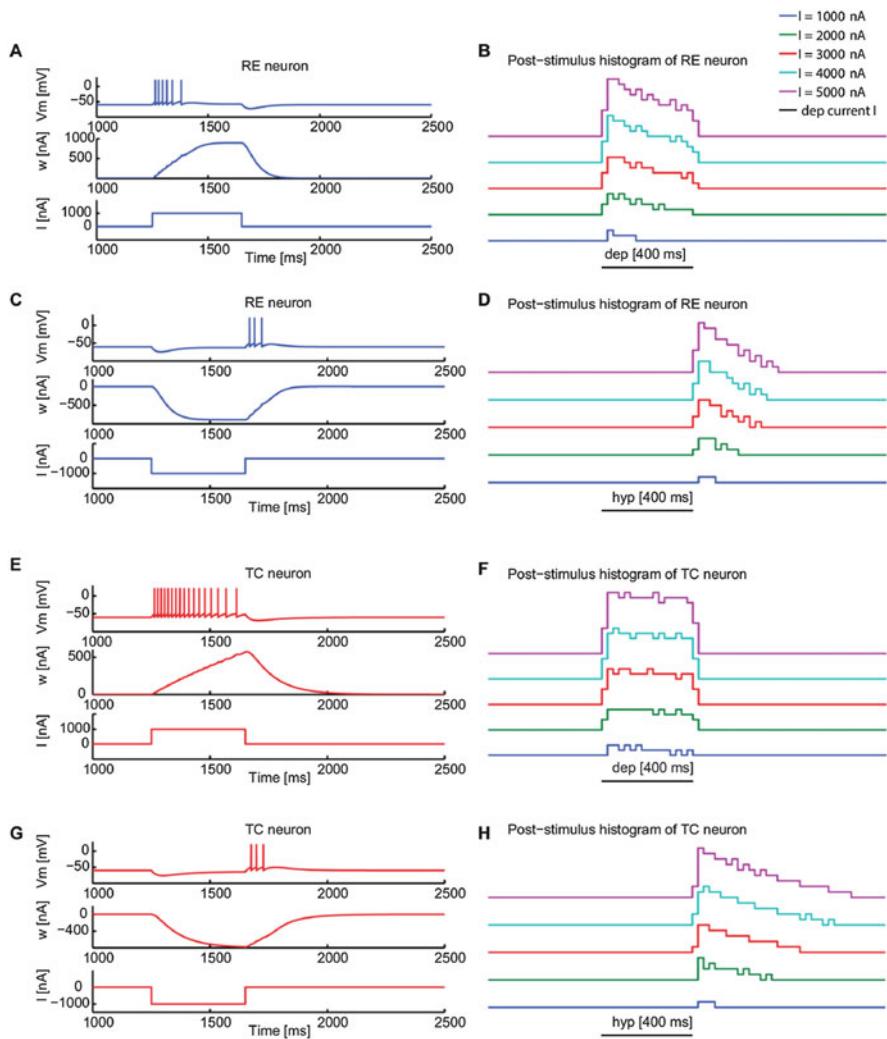


Fig. 8 Dynamical properties of single RE and TC neurons as a function of input current. **(a)** Depolarization induced activity of a RE neuron. Membrane voltage (top) and adaptation variable (middle) of a RE neuron in response to a depolarizing current (bottom). **(b)** Corresponding poststimulus time histograms for increasing depolarizing currents. **(c)** Hyperpolarization-rebound activity of a RE neuron and **(d)** corresponding post-stimulus time histograms for increasing hyperpolarizing currents. Parameters a and b , representing respectively the dynamics and the strength of adaptation [see Eq. (2)] of RE neurons are defined in this way: $a = 0.4 \mu\text{S}$ and $b = 0.02 \text{ nA}$. **(e)** Depolarization induced activity of a TC neuron and **(f)** corresponding post-stimulus time histograms for increasing depolarizing currents. **(g)** Hyperpolarization-rebound activity of a TC neuron and **(h)** corresponding post-stimulus time histograms for increasing hyperpolarizing currents. The values a and b are $0.2 \mu\text{S}$ and 0 nA . The current intensity in **(a, c, e, g)** is 1000 nA , while it varies between 1000 nA and 5000 nA in panels **(b, d, f, h)**. $VT = -50 \text{ mV}$ is the threshold potential for both types of neurons. Reproduced from [20] with permission

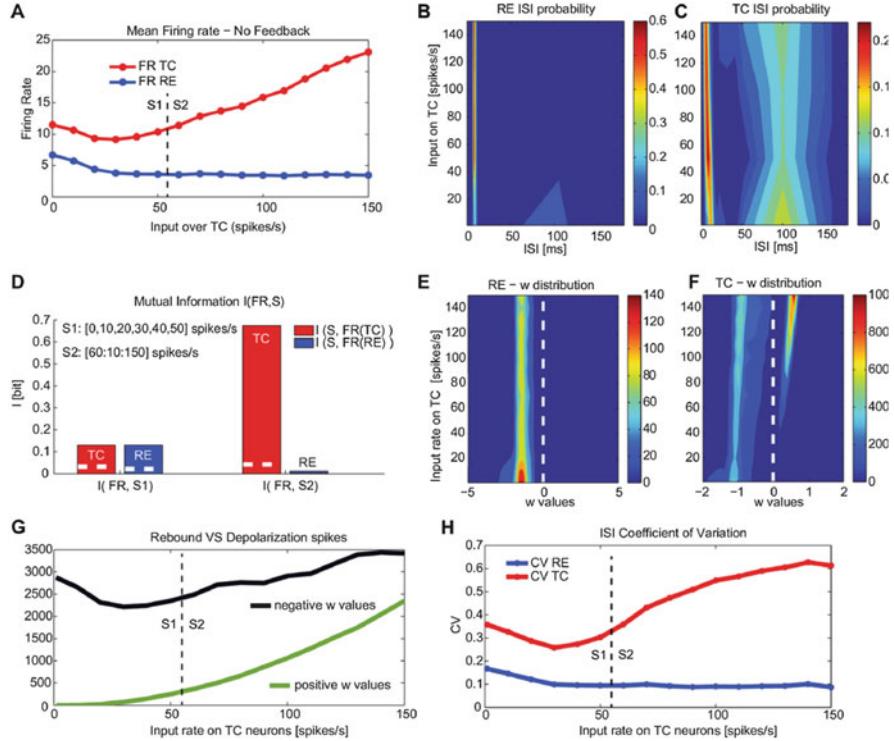


Fig. 9 Bursting and tonic modes displayed by a TC-RE network with RE-RE clustering as a function of external input on TC neurons. **(a)** Firing rate of TC (red) and RE (blue) neurons as a function of external driving input impinging on TC neurons. **(b, c)** ISI distribution as a function of external driving input on TC neurons of RE **(b)** and TC **(c)** neurons. **(d)** Mutual information between the set of increasing external stimulus (0–150 spikes/s) and the neural response given by the firing rate of TC and RE neurons. Different external sensory inputs are considered for the two regimes, following panel **(a)**: 0–50 spikes/s for the bursting mode and 60–150 spikes/s for the tonic mode. The white dashed line in the bar plots refers to significance threshold ($p < 0.05$, bootstrap test). The measures are averaged over 100 trials for each external stimulus. **(e, f)** Adaptation variable w of RE **(e)** and TC **(f)** neurons (color coded) as a function of the external input on TC neurons, averaged across 100 trials for each external stimulus. **(g)** Number of positive w values (depolarizing events, green) and negative w values (rebound events, black) of TC neurons. **(h)** Coefficient of variation of the ISI for both TC and RE cells as the input rate on TC neurons increases. The synaptic conductance is 300 μ S for RE \rightarrow TC and RE \rightarrow RE synapses, and 200 μ S for TC \rightarrow RE synapses. Reproduced from [20] with permission

while RE dynamics was not significantly different in the two regimes (Fig. 9e), the distribution of the recovery variable w in TC neurons was unimodal during “sleep” and became bimodal during “wake” (Fig. 9f). What does that mean? Spikes fired in the bursting/tonic mode are associate to a negative/positive value of w (Fig. 8). Hence, during the “sleep” state all neurons fire due to inhibition rebound and spindle oscillations dominate, while in wake excitation to TC neurons is sufficiently high to

induce also firing rate due depolarization and hence the two firing modes coexist (Fig. 9g).

Overall we found that our relatively simple IFN network was able to reproduce the salient features of thalamic activity: the two firing modes, spindle oscillations and the two information processing regimes, and also to explain how all these elements are correlated. This paves the way to further studies of the thalamic network, be it in the frame of sensory processing or deep brain stimulation.

3.3 Perspectives

We have reviewed here a small set of works focusing on extending the domain of application of IFN networks, specifically to include LFP recordings simulation and thalamic dynamics modeling. This chapter does not have the ambition to be a comprehensive review of applications on IFN networks, yet we would like to conclude addressing two interesting state-of-the-art lines of research that complement and go beyond the work presented here.

First, cortical IFN simulations moved from simulating single networks to simulating the whole column with layer-specific interactions [37, 38]. In Mazzoni et al. [17] we discussed how to simulate LFP for arbitrary cortical populations of neurons, but having only one underlying network. Simulating the interactions among layers, computing the LFP for each layer and finally sum them over the cortical volume would really predict the recordings performed by laminar extracellular electrodes. A first step toward this direction was accomplished in Hagen et al. [39]. Note also that layer-specific simulations are necessary to capture the dynamics of corticothalamic feedback, as the layers receiving TC cells inputs are different from those sending feedbacks to RE cells, and their interaction is not linear.

Second, a reason why IFN networks are a convenient choice to model network dynamics, in particular along the sensory pathway, is that they are currently the most frequent choice for neuromorphic hardware implementations (see [40] for a recent review and [41] for a recent application). In other words, being able to capture the dynamics of an area of the brain with an IFN network means to be able to build an efficient neuromorphic hardware reproducing that same dynamics, with limitless applications, not only in sensory processing but also in other neuroengineering fields as brain-machine interfaces [42].

References

1. Izhikevich, E.M.: Which model to use for cortical spiking neurons? *IEEE Trans. Neural Netw.* **15**, 1063–1070 (2004). <https://doi.org/10.1109/TNN.2004.832719>
2. Burkitt, A.N.: A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biol. Cybern.* **95**, 1–19 (2006). <https://doi.org/10.1007/s00422-006-0068-6>
3. Quiroga, R., Panzeri, S.: Principles of neural coding. CRC Press, Boca Raton (2013)
4. Knight, B.W.: Dynamics of encoding in a population of neurons. *J. Gen. Physiol.* **59**, 734–766 (1972). <https://doi.org/10.1085/jgp.59.6.734>
5. Fourcaud-Trocme, N., Brunel, N.: Dynamics of the instantaneous firing rate in response to changes in input statistics. *J. Comput. Neurosci.* **18**, 311–321 (2005). <https://doi.org/10.1007/s10827-005-0337-8>
6. Brette, R.: Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol.* **94**, 3637–3642 (2005). <https://doi.org/10.1152/jn.00686.2005>
7. Cavallari, S., Panzeri, S., Mazzoni, A.: Comparison of the dynamics of neural interactions between current-based and conductance-based integrate-and-fire recurrent networks. *Front. Neural Circuits.* **8**, 12 (2014). <https://doi.org/10.3389/fncir.2014.00012>
8. Barbieri, F.: Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? *Front. Neurosci.* **2**, 114–122 (2008). <https://doi.org/10.3389/neuro.01.003.2008>
9. Brunel, N.: Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000)
10. Brunel, N., Wang, X.-J.: What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *J. Neurophysiol.* **90**, 415–430 (2003). <https://doi.org/10.1152/jn.01095.2002>
11. Furber, S.: Large-scale neuromorphic computing systems. *J. Neural Eng.* **13**, 51001 (2016). <https://doi.org/10.1088/1741-2560/13/5/051001>
12. Dario, P., Hannaford, B., Takanishi, A.: Guest editorial special issue on biorobotics. *IEEE Trans. Robot.* **24**, 3–4 (2008). <https://doi.org/10.1109/TRO.2008.918929>
13. Oddo, C.M., Raspopovic, S., Artoni, F., Mazzoni, A., Spigler, G., Petrini, F., Giambattistelli, F., Vecchio, F., Miraglia, F., Zollo, L., Di Pino, G., Camboni, D., Carrozza, M.C., Guglielmelli, E., Rossini, P.M., Faraguna, U., Micera, S.: Intraneuronal stimulation elicits discrimination of textural features by artificial fingertip in intact and amputee humans. *eLife.* **5**, e09148 (2016). <https://doi.org/10.7554/eLife.09148>
14. Brunel, N., van Rossum, M.C.W.: Lapicque's 1907 paper: from frogs to integrate-and-fire. *Biol. Cybern.* **97**, 337–339 (2007). <https://doi.org/10.1007/s00422-007-0190-0>
15. Barbieri, F., Mazzoni, A., Logothetis, N.K., Panzeri, S., Brunel, N.: Stimulus dependence of local field potential spectra: experiment versus theory. *J. Neurosci.* **34**, 14589–14605 (2014). <https://doi.org/10.1523/JNEUROSCI.5365-13.2014>
16. Mazzoni, A., Brunel, N., Cavallari, S., Logothetis, N.K., Panzeri, S.: Cortical dynamics during naturalistic sensory stimulations: experiments and models. *J. Physiol. Paris.* **105**, 2–15 (2011). <https://doi.org/10.1016/j.jphysparis.2011.07.014>
17. Mazzoni, A., Lindén, H., Cuntz, H., Lansner, A., Panzeri, S., Einevoll, G.T.: Computing the local field potential (LFP) from integrate-and-fire network models. *PLoS Comput. Biol.* **11**, e1004584 (2015). <https://doi.org/10.1371/journal.pcbi.1004584>
18. Mazzoni, A., Panzeri, S., Logothetis, N.K., Brunel, N.: Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. *PLoS Comput. Biol.* **4**, e1000239 (2008). <https://doi.org/10.1371/journal.pcbi.1000239>
19. Mazzoni, A., Whittingstall, K., Brunel, N., Logothetis, N.K., Panzeri, S.: Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. *Neuroimage.* **52**, 956–972 (2010). <https://doi.org/10.1016/j.neuroimage.2009.12.040>

20. Barardi, A., Garcia-Ojalvo, J., Mazzoni, A.: Transition between functional regimes in an integrate-and-fire network model of the thalamus. *PLoS One.* **11**, e0161934 (2016). <https://doi.org/10.1371/journal.pone.0161934>
21. Buzsáki, G., Anastassiou, C.A., Koch, C.: The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012). <https://doi.org/10.1038/nrn3241>
22. Belitski, A., Gretton, A., Magri, C., Murayama, Y., Montemurro, M.A., Logothetis, N.K., Panzeri, S.: Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. *J. Neurosci.* **28**, 5696–5709 (2008). <https://doi.org/10.1523/JNEUROSCI.0009-08.2008>
23. Bazhenov, M., Stopfer, M., Rabinovich, M., Huerta, R., Abarbanel, H.D., Sejnowski, T.J., Laurent, G.: Model of transient oscillatory synchronization in the locust antennal lobe. *Neuron.* **30**, 553–567 (2001)
24. Lindén, H., Tetzlaff, T., Potjans, T.C., Pettersen, K.H., Grün, S., Diesmann, M., Einevoll, G.T.: Modeling the spatial reach of the LFP. *Neuron.* **72**, 859–872 (2011). <https://doi.org/10.1016/j.neuron.2011.11.006>
25. Lindén, H., Hagen, E., Łęski, S., Norheim, E.S., Pettersen, K.H., Einevoll, G.T.: LFPy: a tool for biophysical simulation of extracellular potentials generated by detailed model neurons. *Front. Neuroinform.* **7**, 41 (2013). <https://doi.org/10.3389/fninf.2013.00041>
26. Capogrosso, M., Wenger, N., Raspopovic, S., Musienko, P., Beauparlant, J., Bassi Luciani, L., Courtine, G., Micera, S.: A computational model for epidural electrical stimulation of spinal sensorimotor circuits. *J. Neurosci. Off. J. Soc. Neurosci.* **33**, 19326–19340 (2013). <https://doi.org/10.1523/JNEUROSCI.1688-13.2013>
27. Kim, S.S., Sripati, A.P., Bensmaia, S.J.: Predicting the timing of spikes evoked by tactile stimulation of the hand. *J. Neurophysiol.* **104**, 1484–1496 (2010). <https://doi.org/10.1152/jn.00187.2010>
28. Jahnsen, H., Llinás, R.: Electrophysiological properties of guinea-pig thalamic neurons: an in vitro study. *J. Physiol.* **349**, 205–226 (1984)
29. Llinás, R., Jahnsen, H.: Electrophysiology of mammalian thalamic neurones in vitro. *Nature.* **297**, 406–408 (1982)
30. Destexhe, A., Sejnowski, T.J.: Interactions between membrane conductances underlying thalamocortical slow-wave oscillations. *Physiol. Rev.* **83**, 1401–1453 (2003). <https://doi.org/10.1152/physrev.00012.2003>
31. Destexhe, A., Contreras, D., Sejnowski, T.J., Steriade, M.: A model of spindle rhythmicity in the isolated thalamic reticular nucleus. *J. Neurophysiol.* **72**, 803–818 (1994)
32. Willis, A.M., Slater, B.J., Gribkova, E.D., Llano, D.A.: Open-loop organization of thalamic reticular nucleus and dorsal thalamus: a computational model. *J. Neurophysiol.* **114**, 2353–2367 (2015). <https://doi.org/10.1152/jn.00926.2014>
33. Golomb, D., Wang, X.J., Rinzel, J.: Propagation of spindle waves in a thalamic slice model. *J. Neurophysiol.* **75**, 750–769 (1996)
34. Muller, L., Destexhe, A.: Propagating waves in thalamus, cortex and the thalamocortical system: experiments and models. *J. Physiol. Paris.* **106**, 222–238 (2012). <https://doi.org/10.1016/j.jphospharis.2012.06.005>
35. Delorme, A., Gautrais, J., van Rullen, R., Thorpe, S.: SpikeNET: a simulator for modeling large networks of integrate and fire neurons. *Neurocomputing.* **26–27**, 989–996 (1999). [https://doi.org/10.1016/S0925-2312\(99\)00095-8](https://doi.org/10.1016/S0925-2312(99)00095-8)
36. Destexhe, A.: Self-sustained asynchronous irregular states and Up-Down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. *J. Comput. Neurosci.* **27**, 493–506 (2009). <https://doi.org/10.1007/s10827-009-0164-4>
37. Bos, H., Diesmann, M., Helias, M.: Identifying anatomical origins of coexisting oscillations in the cortical microcircuit. *PLoS Comput. Biol.* **12**, e1005132 (2016). <https://doi.org/10.1371/journal.pcbi.1005132>

38. Potjans, T.C., Diesmann, M.: The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cereb. Cortex.* **24**, 785–806 (2014). <https://doi.org/10.1093/cercor/bhs358>
39. Hagen, E., Dahmen, D., Stavrinou, M.L., Lindén, H., Tetzlaff, T., van Albada, S.J., Grün, S., Diesmann, M., Einevoll, G.T.: Hybrid scheme for modeling local field potentials from point-neuron networks. *Cereb. Cortex.* **26**, 4461–4496 (2016). <https://doi.org/10.1093/cercor/bhw237>
40. Indiveri, G., Liu, S.-C.: Memory and information processing in neuromorphic systems. *Proc. IEEE.* **103**, 1379–1397 (2015). <https://doi.org/10.1109/JPROC.2015.2444094>
41. Osswald, M., Ieng, S.-H., Benosman, R., Indiveri, G.: A spiking neural network model of 3D perception for event-based neuromorphic stereo vision systems. *Sci. Rep.* **7**, 40703 (2017). <https://doi.org/10.1038/srep40703>
42. Boi, F., Moraitis, T., De Feo, V., Diotalevi, F., Bartolozzi, C., Indiveri, G., Vato, A.: A bidirectional brain-machine interface featuring a neuromorphic hardware decoder. *Front. Neurosci.* **10**, 563 (2016). <https://doi.org/10.3389/fnins.2016.00563>

Computational Modeling as a Means to Defining Neuronal Spike Pattern Behaviors



Siva Venkadesh and Giorgio A. Ascoli

Abstract The intrinsic dynamical diversity exhibited by neurons likely plays a major role in information processing in the nervous system. Experimentally observed neuron behaviors show a wide variety of qualitatively distinguishable spike patterns such as delayed spiking, fast spiking, spiking with frequency adaptation, and bursting. Defining a neuron type based on its spike pattern behaviors is valuable for a systematic investigation of collective dynamics emerging from a network of neurons. In this chapter, we present and discuss a strategy to define neuron types using compact mathematical representations. First, we compare several widely used neuronal modeling systems such as Integrate and Fire and Hodgkin-Huxley type models and discuss their benefits and limitations. Then, we describe a framework to systematically classify experimentally observed spike pattern types. Finally, we discuss various aspects of optimization techniques that are relevant to the problem of representing a neuron type in an abstract model space.

1 Introduction

Information processing in the nervous system is facilitated by the rich, intrinsic computational properties of the neurons. These properties are revealed in their excitability and various temporal patterns of activation. The precise timing of voltage spikes during this activation is crucial for the representation of information [1–3]. The patterns of these voltage spikes are distinguished by several qualitative features such as fast spiking, bursting, spike frequency adaptation, bursting followed by spiking, and latency to spike. In many cases, a neuron exhibits several of these features under different experimental conditions. Thus, a neuron can be qualitatively characterized by the set of all different experimentally observed spike pattern types, which defines its spike pattern behavior.

S. Venkadesh • G.A. Ascoli (✉)

Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA

e-mail: ascoli@gmu.edu

Computational simulations have been invaluable to the investigation of neuronal dynamics. Biophysically detailed Hodgkin-Huxley [4] type models have provided great insights into the computational properties of the neurons. Other simpler spiking models such as the Leaky Integrate and Fire models have been used to investigate the collective dynamics of neurons in large networks. Several such abstract models have been proposed for different applications. A particular model becomes the choice for a study by taking into account the factors such as its biophysical meaning, simulation cost, and computational possibilities. Section 2 provides a brief background on the computational properties of a neuron and a survey of the existing abstract models that capture these properties in varying levels of detail.

Traditionally, morphology has been the major characterizing factor that defines a neuron type. From a modeling perspective, characterization based on electrophysiological properties such as membrane capacitance, time constant and spike amplitude might seem relevant. However, these properties are highly sensitive to the experimental conditions and hence not very reliable. In addition, there could be several possible configurations of morphological and electrophysiological parameters that result in a single computational property. Therefore, in order to investigate the information processing in the nervous system, it is more useful and relevant to characterize the neuron types based on their spike pattern behaviors.

Of particular interest to this chapter is the representation of a neuron type as *possibilities* in an abstract model space, where these possibilities encapsulate its known behaviors. Such a representation is valuable for the following reasons: Firstly, it might provide key insights into the existence of computational subtypes within a certain morphologically characterized neuron type, or, it might suggest a new direction towards characterizing neurons entirely based on their detailed neuro-computational properties. Secondly, in order to create a biologically realistic large scale spiking neural network (SNN) model of a brain region, the individual neuronal models in that SNN should accurately represent the complex behaviors exhibited by the real neurons. Definition of these behaviors in the model space offers a sampling region for the neurons in such large scale SNN simulations and allows one to investigate the collective dynamics of several neuron types with distinct behaviors. However, this requires a framework that enables a systematic analysis and categorization of different experimentally observed behaviors. One such framework is discussed in Sect. 3. Finally, Sects. 4 and 5 discuss the overall strategy to model the spike pattern behaviors and the issues that need consideration.

2 Computational Model of a Neuron

Several abstract models have been proposed to capture the computational properties of a neuron in varying levels of detail. On one end, the Hodgkin-Huxley (HH) model [4], which is a four dimensional system, incorporates the persistent K^+ and transient Na^+ current dynamics in order to explain the generation of an action potential

in a neuron. The HH model has been traditionally used to study the dynamics of a single neuron in detail at the level of ion channel gating mechanisms. On the other end, the leaky-integrate-and-fire model (LIF), which is described by a single linear differential equation, simply defines a spike event based on a voltage threshold. LIF models have been widely used in SNN simulations to investigate the network level dynamics and to solve practical problems in machine learning [3, 5, 6]. In between these two, there is a wide spectrum of models that capture the computational properties of real neurons to different extents. In these models, the biophysical accuracy is usually traded off for a lower model simulation cost. The preference for a particular model might be made by weighing its biophysical details against its cost of simulation. Biophysical details are useful and relevant when one is interested in investigating the factors that affect the single neuron dynamics such as the morphology of the neuron [7], back propagation of action potentials [8] or ionic currents [9]. For other investigations, it is more meaningful to discuss the model preference in terms of the neuro-computational possibilities allowed in that system rather than its biophysical details.

2.1 Neuro-computational Properties

Neuron is a dynamical system [10] and the excitability of a neuron can be described by the bifurcation mechanisms revealed in its phase portrait [11]. The bifurcation from resting state of the membrane voltage (stable equilibrium) to repetitive spiking can be caused by a single parameter—the strength of the injected current. Two major bifurcation mechanisms broadly define the neuro-computational properties [12]: In a neuron that undergoes Andronov-Hopf bifurcation, perturbing the stable equilibrium results in a damped oscillation towards the resting state. In the case of saddle-node bifurcation, such perturbation results in an exponential convergence towards the resting state. The former is known as resonator neurons and the latter as integrator neurons. The resonators do not have a clearly defined voltage threshold to generate an action potential, whereas integrators do. An integrator's action potential is an *all-or-none* response, whereas a resonator can generate an action potential with large or intermediate amplitude. An integrator's firing response is directly proportional to the frequency of the input signal. The resonators require specific frequency components in the input signal to elicit a response. In addition, inhibitory input signals can promote spiking response in resonators, whereas inhibition always impedes an integrator's response. Thus, these two classes differ in some of the fundamental neuro-computational properties. Apart from being an integrator or a resonator, the biological neurons also exhibit different qualitative neuro-computational features in their activation patterns such as bursting, bursting followed by spiking, spike frequency adaptation, and latency to spike. Therefore, a good abstract model of a neuron should be able to reproduce all of these features.

2.2 Biophysically Meaningful Models

The Hodgkin-Huxley model [4] precisely captures the mechanisms of action potential generation in a neuron. It describes the dynamics of membrane voltage (V) using voltage-gated Na^+ and K^+ currents and an Ohmic leak current (I_{Na} , I_K and I_l respectively):

$$\begin{aligned} C \cdot \frac{dV}{dt} &= I - I_{Na} - I_K - I_l \\ I_{Na} &= g_{Na} \cdot m^3 \cdot h \cdot (V - E_{Na}) \\ I_K &= g_K \cdot n^4 \cdot (V - E_K) \\ I_l &= g_l \cdot (V - E_l) \end{aligned}$$

Here, C is the membrane capacitance and I is the externally injected current. The transient I_{Na} is further described by the state variables corresponding to three activation gates (m) and one inactivation gate (h). Similarly, the persistent I_K is described by the state variable corresponding to four activation gates (n). The variables m , h and n define the probabilities of opening/closing of their respective ion channels. g_{Na} , g_K and g_l are maximum conductances for Na^+ , K^+ and leak currents respectively and E_{Na} , E_K and E_l are the Nernst equilibrium potentials. Although this four dimensional system can capture the rich neuro-computational properties of a biological neuron, it is computationally very expensive to simulate. Consequently, the HH type models are not ideal to be used to build large scale neural networks.

Another biophysically meaningful model was suggested by Morris and Lecar (ML) [13], which reduced the HH model and included only two state variables. This model omitted the I_{Na} inactivation and approximated its activation by including an instantaneous Ca^{2+} current as follows:

$$\begin{aligned} C \cdot \frac{dV}{dt} &= I - I_{Ca} - I_K - I_l \\ I_{Ca} &= g_{Ca} \cdot M_\infty(V) \cdot (V - E_{Ca}) \\ I_K &= g_K \cdot W \cdot (V - E_K) \\ I_l &= g_l \cdot (V - E_l) \end{aligned}$$

Here, the state variable W describes the probability that a K^+ channel is in open state and $M_\infty(V)$ describes the instantaneous I_{Ca} activation. g_{Ca} , g_K and g_l are maximum conductances for Ca^{2+} , K^+ and leak currents respectively. Although the ML model reduces the dimensionality of the HH model, it requires a significantly small time step for the simulation in order to accurately reproduce the spike times

[14], which makes it still computationally expensive. In addition, this model does not capture spike frequency adaptation or tonic bursting.

2.3 Integrate and Fire (IF) Models

The simplest and one of the most widely used spiking models of a neuron is the leaky integrate and fire (LIF) model, which describes the membrane voltage (V) using only a leak current and a membrane time constant (τ). When V reaches a threshold level (V_t), it is instantaneously reset to a resting level (V_r), which marks a spike event.

$$\tau \cdot \frac{dV}{dt} = -(V - E_l)$$

$$\text{if } V \geq V_t \text{ then } V = V_r$$

Although widely used in large scale SNN simulations due to its simplicity, this model cannot reproduce several important neuro-computational properties including bursting, spike frequency adaptation, or resonance. A few variations of the LIF model have been introduced in order to capture some of these properties. For instance an LIF with Adaptation adds another differential equation to incorporate spike frequency adaptation. Here, the dynamics of an activation gate is introduced (g), which decays with time constant τ after a spike event.

$$\begin{aligned} \tau \cdot \frac{dV}{dt} &= -(V - E_l) - g \cdot (V - E_K) \\ \frac{dg}{dt} &= \frac{e \cdot \delta(t) - g}{\tau} \end{aligned}$$

Other variations of LIF such as IF-or-Burst (IFB) [15], Resonate-and-Fire (RF) [12], and Quadratic IF (QIF) [16] models add some neuro-computational properties to the basic LIF model. However, the QIF cannot reproduce bursting or resonance, the RF cannot reproduce spike frequency adaptation or be an integrator, and the IFB cannot reproduce subthreshold resonance or bursting followed by spiking. Brette and Gerstner incorporated the features of Izhikevich model (described below) into an IF model and created a two dimensional model known as Adaptive Exponential IF (AdEx) [17]. This model has not only been shown to reproduce several neuro-computational properties qualitatively [17], but also performed well in a quantitative single-neuron modeling competition [18].

2.4 Izhikevich Model

Izhikevich proposed a simple model (IM) of spiking neuron with two differential equations [19] and showed that this model was able to reproduce many neuro-computational properties qualitatively. In addition to the membrane voltage (V), a membrane recovery variable (U) was included in the equation as follows:

$$\frac{dV}{dt} = 0.04V^2 + 5 \cdot V + 140 - U + I$$

$$\frac{dU}{dt} = a(b \cdot V - U)$$

$$\text{if } V \geq 30\text{mV} \text{ then } V = c, U = U + d$$

The recovery variable U approximates the activation of I_K and the inactivation of I_{Na} in the HH model. Similar to the IF models, it defines a spike cutoff value to reset the state variables. The parameter ‘ a ’ is the time constant for the recovery variable U . The parameters ‘ b ’ and ‘ a ’ collectively determine whether the model is an integrator or resonator. Typically, the model is an integrator, when $b < a$, and resonator when $b > a$. The parameters ‘ c ’ and ‘ d ’ are after-spike reset values for V and U respectively. The resting voltage in this model is between -70 and -60 mV depending on the value of ‘ b ’. A slight variation of this model conveniently includes additional parameters such as resting voltage (V_r), threshold voltage (V_t) and cell capacitance (C) without affecting its computational efficiency [20]:

$$C \cdot \frac{dV}{dt} = k \cdot (V - V_r) \cdot (V - V_t) - U + I$$

$$\frac{dU}{dt} = a \cdot \{b \cdot (V - V_r) - U\}$$

$$\text{if } V = V_{peak} \text{ then } V = c, U = U + d$$

The parameter ‘ k ’ defines the shape of the spike upstroke and V_{peak} defines the spike cutoff value.

Unlike the biophysically detailed models, the IM has been shown to capture several neuro-computational properties with a very low simulation cost [14]. Consequently, this model has been used in a number of large scale SNN simulations [21–23]. Therefore, the IM is an appropriate choice if one wants to build a large scale SNN with neuro-computationally diverse single neuron models. However, to create such diversity, the nine parameters of this model that collectively determine its neuro-computational properties need to be known.

Qualitatively characterizing the neuro-computational properties as tonic spiking or bursting is useful to evaluate different abstract models as done in [14]. However, the real neurons exhibit more complex behaviors like bursting and spiking

with frequency adaptation under different current stimulation levels (e.g. CA1 Neurogliaform neuron). Therefore, before truly exploring the neuro-computational possibilities in the parameter space of the IM, one needs a systematic approach to classify the known spike pattern types and categorize different behaviors of real neurons. The following section discusses one previously developed such framework.

3 Spike Pattern Behaviors

In a previous work [24], more than 100 neuron types in the rodent hippocampal formation were identified based on their axonal and dendritic locations across the hippocampal layers. A knowledge base (Hippocampome.org) was created and it includes the morphological, electrophysiological and molecular marker evidences for all neuron types, where information is available in the literature. Experimentally recorded voltage traces of these neuron types were analyzed and a firing pattern classification scheme was developed based on the presence of transient/steady-state and silence/spiking/bursting components in a firing pattern [25] (Table 1). Consequently, more than 10 different firing pattern types were identified among all the hippocampal neuron types. Representative cases for different spiking and bursting/stuttering pattern types are listed in Figs. 1a, b respectively. It should be noted that a neuron type can exhibit more than one of these types under different experimental conditions (e.g. different levels of current injection). For instance, a CA1 Bistratified interneuron elicited spike trains of the types D.PSTUT and D.NASP at 400pA and 600pA current injection levels respectively [34]. The set of all different spike pattern types known to be elicited by a neuron type defines its behavior. However, since these neuron types have been primarily characterized based on morphology, one must consider the possibility of electrophysiological subtypes in the context of spike pattern behaviors.

As listed in Table 2, a neuron type belongs to the ‘multi-behavior’ category when a single neuron of that type is known to elicit two different firing pattern types (e.g. D.NASP and NASP for two different current injection levels). If different neurons of the same type elicited different firing pattern types for the same current injection level under identical experimental conditions, then electrophysiological ‘subtypes’ exist for that neuron type. If the current injection levels were different in this scenario, the existence of subtypes is not conclusive, since it could also be a multi-

Table 1 A scheme for firing pattern classification (reproduced from [25])

	Transient response	Steady-state response
Silence	Delay (D)	Silence (SLN)
Spiking	Adapting Spiking (ASP)	Non-Adapting Spiking (NASP)
Bursting	Transient Stuttering (TSTUT)	Persistent Stuttering (PSTUT)
	Transient Slow Wave Bursting (TSWB)	Persistent Slow Wave Bursting (PSWB)

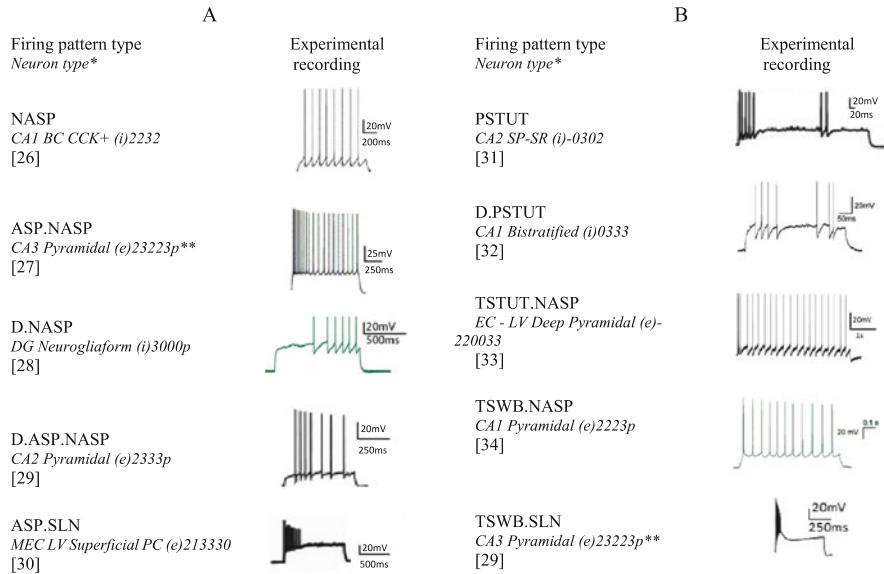


Fig. 1 Exemplar spiking (**a**) and bursting/stuttering (**b**) pattern types in the hippocampus (reproduced from [25]). *Naming scheme for neuron types is based on [24]. **Neuron subtypes

Table 2 Possible categories of behavior for a neuron type ‘*N*’ that exhibited more than one firing pattern type (adopted from [25])

	Scenario	Behavior category	Response of neuron N_1		Response of neuron N_2	
			For I_1	For I_2	For I_1	For I_2
1	Recording from a single neuron	Multi-behavior	FPT_1	FPT_2	—	—
2	Recordings from multiple neurons	Subtypes	FPT_1	—	FPT_2	—
3	Recordings from multiple neurons	Subtypes and Multi-behavior	FPT_1	FPT_3	FPT_2	—
4	Recordings from multiple neurons	Subtypes or Multi-behavior ^a	FPT_1	—	—	FPT_2

FPT Firing pattern type, I Input current injection level, N_1 and N_2 are two different neurons of the type N

^aNon conclusive category. It also includes the cases where the recordings were made under different experimental conditions regardless of the current injection levels (e.g. different room temperatures, different animal species)

behavior neuron type (‘subtypes *or* multi-behavior’). Finally, a neuron type can also satisfy the criteria for both ‘subtype *and* multi-behavior’. Definition of spike pattern behaviors in the IM parameter space might give useful insights into the ‘subtypes *or* multi-behavior’ category. For instance, if different spike pattern types of a behavior from the ‘subtypes *or* multi-behavior’ category cannot be reproduced in a single IM, it might suggest the existence of computational ‘subtypes’ within that neuron type.

However, one must also take into account the possible limitations of the IM in these cases. From a modeling perspective, this requires optimization and exploration of the IM parameters. The next two sections discuss various methodologies to achieve this and the associated challenges in the context of parameter optimization.

4 Evolutionary Algorithm as a Tool for Modeling Neuronal Dynamics

An Evolutionary algorithm (EA) is a powerful tool to tackle optimization problems. The EA is guided by a fitness function, which is used to evaluate the quality of the points in the parameter search space. The objective of using an EA for an optimization problem is usually concerned with choosing the appropriate type of the algorithm and its configuration, and the definition of the fitness function. The possible choices for an EA include, but are not limited to, genetic algorithm (GA), evolutionary strategies (ES), genetic programming (GP), and particle swarm optimization (PSO). Each of these nature-inspired algorithms has their own set of features, but they all have a common core template: A population of candidate solutions goes through small changes for several iterations in order to explore and exploit the search space with the guidance of a fitness function.

4.1 Model Optimization Using the EA

A number of studies have addressed the optimization problem in the context of neuronal models and different approaches for tuning the model parameters have been discussed in the literature. With the goal of fitting the model responses to voltage traces recorded from the biological neurons, these studies mainly differ in the following details: (a) the choice of the optimization algorithm, (b) the type of the neuronal model used and (c) the fitness function definition.

On the model spectrum, simple models such as the adaptive IF, adaptive threshold IF, adaptive exponential IF and Izhikevich models have been optimized using the GA and PSO [35–37]. The parameters of the biophysically meaningful models such as ML and HH with multiple compartments have also been optimized using the GA and ES [38–41]. However, the major aspect that characterizes these studies is the fitness function which defined the objectives. As mentioned before, the goal of this optimization problem is to create models whose responses are close to the experimentally recorded voltage traces. There are different ways to capture this goal in the fitness function. One approach is to directly compare the points of the experimental voltage trace with those of the model response. The problem with this approach is that the fitness function becomes very sensitive to small time shifts in the spike train [42, 43]. Another similar approach called the phase plane

trajectory density (PPTD) method avoids this problem [43]. The PPTD method excludes the time parameter and plots two time-dependent variables such as the voltage $V(t)$ and its first derivative dV/dt for the comparison. Using this method as a fitness measure, a Purkinje cell model with 1600 compartments and 24 free parameters has been tuned [39]. The effects of using different fitness functions on the performance of the GA for a HH model with 19 free parameters has been studied in [40]: These functions were based on direct comparison of voltage traces, the PPTD method, and the comparison of inter-spike intervals. The fitness function that combined all these three modes of analysis was shown to be the most effective. Another approach is to use the features of a spike pattern such as spike frequency, AP width, and AP overshoot in the fitness function [41]. The major advantage of this approach is that the features can be selectively included in the fitness function. For instance, if one decides that the shape of the spike upstroke is not relevant to their study, and is instead only interested in reproducing a certain spike frequency, the fitness function can be configured accordingly. In [38], the input current—firing frequency curve was used to tune the ML model with six free parameters. Finally, the possibility of using spike coincidences as the fitness measure has been explored with IF and Izhikevich models [35–37]. This approach was based on the number of coincidences between the experimentally recorded spikes and the model spikes, where coincidences were defined using a small temporal window. On a benchmark dataset, this approach worked well with the IF models, but performed poorly with the Izhikevich model.

4.2 Feature-Based Fitness Function

As an optimization problem, the objective here can be broadly defined as follows: implement a fitness function that can reduce the model error in terms of spike times. However, as mentioned earlier, the spike pattern behaviors are characterized by the higher level temporal features such as first spike latency (fsl) and spike frequency adaptation (sfa). Therefore, it is useful to include these features in the fitness function rather than the spike times. This allows one to give importance to one feature over the other by assigning different weights to their respective errors. In addition, the right combination of these features can implicitly represent the original goal of reproducing accurate spike times. As an example, consider the experimentally recorded voltage trace from a CA3 pyramidal neuron given in Fig. 2a [44]. The spike times of this pattern can be implicitly represented as objectives by simply including the following features as illustrated in Fig. 2a: fsl , pss , sc and sfa during the first five spikes. Thus, the fitness function for such an ASP.NASP pattern could be defined as $-\sum_{f \in S} w_f E_f$, where $S : \{fsl, sfa, pss, sc\}$, and w and E are the weight and the error associated with a feature. However, this fitness function will not work for a spike pattern of the type D.PSTUT, because this definition does not capture post-burst interval (pbi), which is a characterizing feature of this spike pattern type as illustrated in Fig. 2b. For such a stuttering or bursting spike pattern,

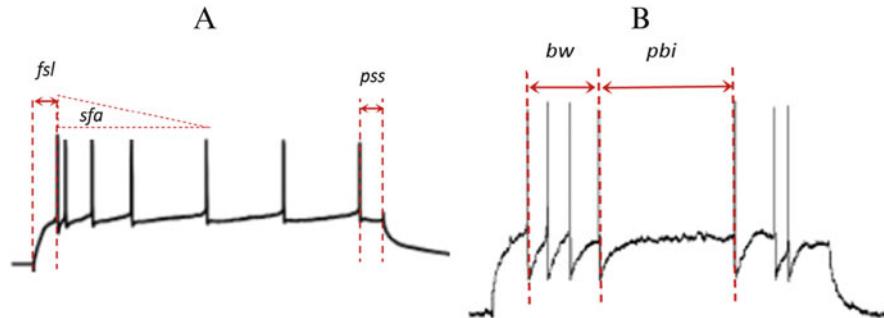


Fig. 2 Possible quantitative features of spike patterns to be included in the fitness function. First spike latency (*fsl*): Duration between the onset of input current stimulation step and the first spike, Post spike silence (*pss*): Duration between the last spike and the end of input current stimulation step, Spike count (*sc*): Number of spikes recorded, Spike frequency adaptation (*sfa*): Reduction in the frequency of spikes following the first spike, Burst count (*bc*): Number of bursts recorded. A burst is a cluster of high frequency spikes followed by a period of quiescence. Burst width (*bw*): Duration of a burst, Post burst interval (*pbi*): Duration of quiescence following the burst

additional features such as *bc*, *bw*, and *pbi* as illustrated in Fig. 2b should be included in the fitness function. Therefore, for each of the different spike pattern types, a minimal set of higher level temporal features needs to be identified to be used as fitness measures.

4.3 Fitness Landscape with a Feature Based Function

To our knowledge, optimization algorithms have not been successfully used to fit the IM responses to experimental voltage recordings. As mentioned before, on a benchmark optimization test, the IM performed poorly using an EA compared to the Adaptive, Adaptive threshold, and Adaptive exponential IF models [35–37]. The major factor that affects the performance of an EA is the nature of the problem fitness landscape. The fitness landscape is the hypersurface in the parameter search space and it defines the relationship between the parameters and the fitness value. As the search for the optimal parameters is guided by this landscape, its topographical features directly affect an EA’s performance. These features can be used to assess the computational complexity of an optimization problem [45] and knowledge about the fitness landscape will be helpful in choosing the appropriate algorithm and its configuration.

The major problem that any EA faces is the premature convergence to local optima [46]. This happens when the population of candidate solutions reaches fitness peaks that are better than the neighboring regions, but are not the globally best. If all the candidate solutions reach the same local optimum, there is very little chance of further exploration of the search space by the algorithm. This problem is

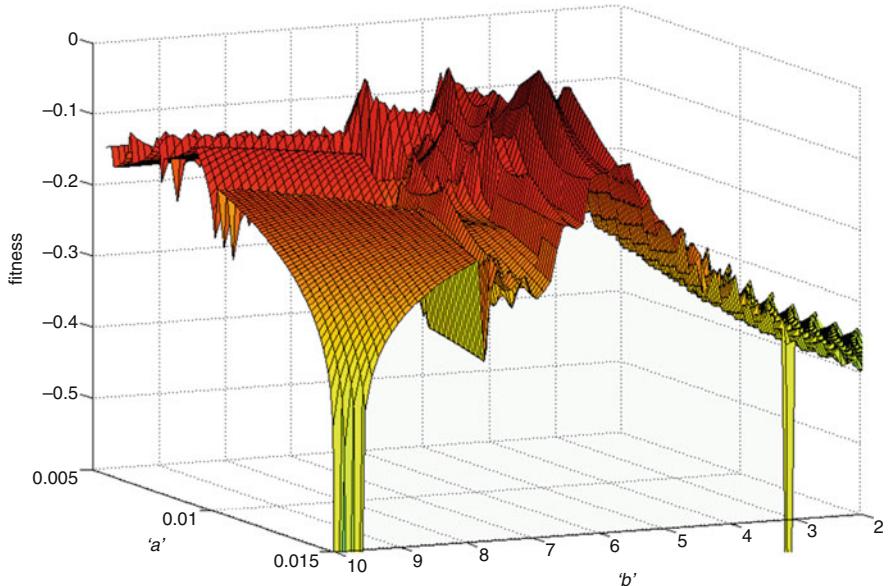


Fig. 3 A rugged region in the fitness landscape created by the IM parameters ‘ a ’ and ‘ b ’. The fitness function included errors on features fsl , sfa , pss , sc as described in Sect. 4.1 for the trace given in Fig. 2a. Fitness values range from 0 to -1 , where 0 denotes the best fitness

more severe in the rugged landscapes [47], which contain several of these locally optimal peaks. The parameters of the IM create such rugged regions in the fitness landscape (Fig. 3). In these landscapes, the gradient information about the fitness is not very reliable and this has a negative effect on the performance of the EA.

Another problem is the epistatic interaction among the parameters in creating the fitness landscape. In the context of optimization, epistasis is defined as the dependency of a parameter’s contribution to the fitness value on the state of other parameters. The higher the epistatic interactions among the parameters, the harder it is for the EA to optimize them [48]. There are strong interactions among the Izhikevich model parameters to create a fitness landscape, as exemplified in Fig. 4. The region shaded in blue corresponds to the model responses that are not biologically valid, hence it is assigned the lowest fitness value. As obvious from this illustration, this region is not defined independently by the parameters ‘ k ’ or ‘ Vt ’, but collectively by both ‘ k ’ and ‘ Vt ’. Therefore, independently varying one of these genes might break the building blocks representing such interaction. Finally, the noise introduced in the fitness landscape creates a different problem. Fitness is evaluated by comparing the model responses with the experimentally recorded spikes. Since these electrophysiological recordings are not completely devoid of measurement errors, the resulting fitness landscape might not be an accurate representation of the desired theoretical objective.

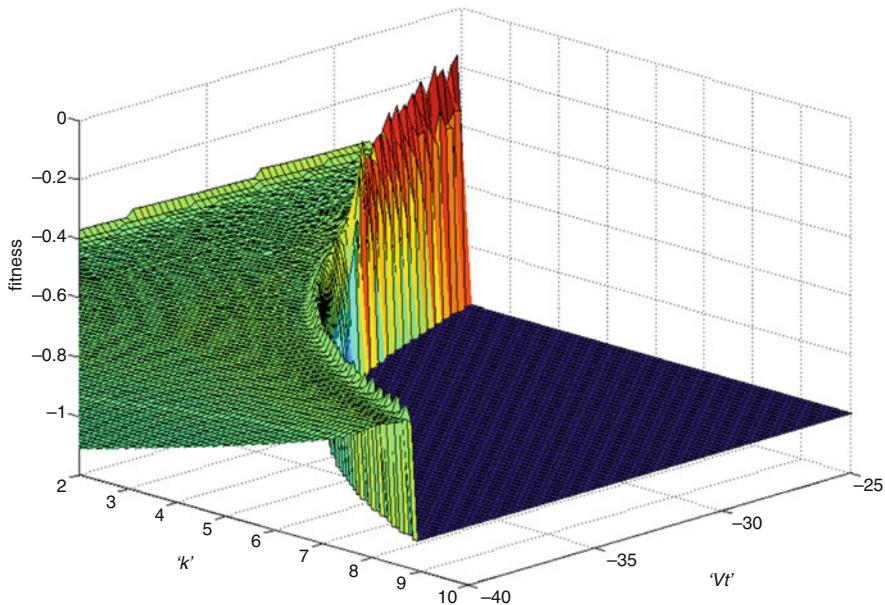


Fig. 4 Epistatic interaction between the parameters ' k' ' and ' Vt ' to define invalid region (shaded in blue) in the parameter space

It is necessary to consider the facts discussed above when configuring the EA to optimize the IM parameters. There is no practicable method that directly solves the problem of premature convergence caused by the rugged landscapes. Maintaining diversity in the population has been shown to be a general strategy to avoid premature convergence to the local optima. However, it is crucial to achieve the right balance between exploration and exploitation of the search space [49]. For instance, too much selection pressure with very little mutation might exploit only the local regions without really exploring other regions of the search space. On the other hand, too much mutation with very little selection pressure would always explore different regions without really climbing to their peaks. Therefore, a proper balance between the mutation/crossover operations and the selection pressure might mitigate the effects of rugged fitness landscapes. Knowledge of epistatic interactions among the parameters will be helpful to devise schemes that could prevent building blocks from being destroyed by the variation operators such as crossover. Although there is no guarantee, appropriately configuring an EA with the knowledge of the problem landscape maximizes its chances of reaching the global optimum.

5 Modeling Spike Pattern Behaviors

The goal of the problem under discussion is not just to fit IM responses to the experimental voltage recordings. It is to characterize the neuron types in terms of IM parameters and their relationships that define their behaviors. Although developing a strategy that can fit IM responses to individual voltage traces (as discussed in Sect. 4.1) is a necessary objective, modeling a behavior requires further considerations, which are discussed below.

5.1 Optimization Objectives with a Behavior

A neuron type can exhibit more than one firing pattern type as listed in Table 2. The set of all different firing pattern types exhibited by a neuron type under different experimental conditions (such as different input current injection) defines its behavior. For example, the CA1 Bistratified neuron is known to produce the firing pattern type D.PSTUT (Fig. 2b [34]). In addition, a firing pattern of the type D.NASP has also been recorded from this neuron type [34]. The objectives of modeling such behaviors can be defined in the fitness function in one of many ways. One method is to extend the approach described in Sect. 4.1, and simply sum up the weighted feature errors across multiple firing pattern types. However, this approach will not be very useful and informative with cases where there are conflicting objectives in the fitness function. Before pointing out and illustrating an alternate approach, it is useful to consider the scenarios that could possibly create such conflicting objectives. As explained in Sect. 2.2, modeling the spike pattern behaviors might give useful insights into the ‘subtype *or* multi-behavior’ class of neuron types. If a neuron type indeed has multiple subtypes, then the voltage traces recorded from these different subtypes might be computationally impossible to be represented in a single IM, and thus creating conflicting objectives. With this consideration, an alternate approach to define such objectives is to use Multi Objective Optimization (MOO) [50, 51].

The MOO is a useful technique to deal with conflicting objectives. Rather than trying to find a single optimal solution for a problem, the MOO attempts to find a Pareto front of optimal solutions, where optimality is defined by the notion of ‘dominance’: A candidate solution ‘A’ dominates another candidate solution ‘B’, if and only if ‘A’ is not worse than ‘B’ in all objectives, and ‘A’ is strictly better than ‘B’ in at least one objective. Thus, in a MOO with conflicting objectives, the candidate solutions that are *not* dominated by others form the Pareto front of solutions. In principle, features of a single experimental voltage trace (e.g. *fsl*,

sc, and *pss* from a recording of type D.NASP) could be viewed as the multiple objectives for the MOO. However, considering the rich dynamics of Izhikevich model, there might not necessarily be any trade-off between the features of a single spike pattern. In other words, the features *fsl* and *sc* are not conflicting objectives in a D.NASP voltage trace when Izhikevich model is used. Such tradeoffs are more likely to occur among the objectives across different firing pattern types. For instance, one may try to fit a single model to both NASP and TSWB.NASP pattern types by only varying the input current, which might not be possible in the IM. However, trying to do so with the MOO might give useful insights into the nature of trade-off among the features that characterize these spike pattern types.

5.2 Parameter Space Exploration

Besides conflicting objectives, one must also take into account the possible inadequacy of objectives while constraining a model. Several neuron types have not been investigated in detail in the literature and, as a result, the available experimental evidences might not be sufficient to create accurate model representations. These cases should be viewed as a parameter exploration problem rather than an optimization. The EA's have been largely used as parameter optimizers with the goal of reaching a single global optimum. However, their optimization capability highly relies on their ability to effectively explore the parameter space. In principle, due to their stochastic nature, an appropriately configured EA should be able to reach multiple global optima across different trials.

As an illustration, the EA was run to optimize the IM parameters (including the input current) for arbitrarily defined feature constraints for four different spike pattern types. Across different stochastic trials, the EA identified multiple points in the search space that reproduced these constraints (Fig. 5a). Representative model response for each spike pattern type is given in Fig. 5b. Some observations could be made from this figure from the point of view of the spike pattern behaviors. Although only a weak adaptation was added to the NASP constraints in order to produce an ASP.NASP pattern, these two patterns were clearly separated in the IM parameter space. Very few neuron types have been known to exhibit ‘multi-behavior’ with these two spike pattern types. However, more definite constraints need to be defined for each spike pattern type to explore the possibilities of such behaviors in the IM parameter space.

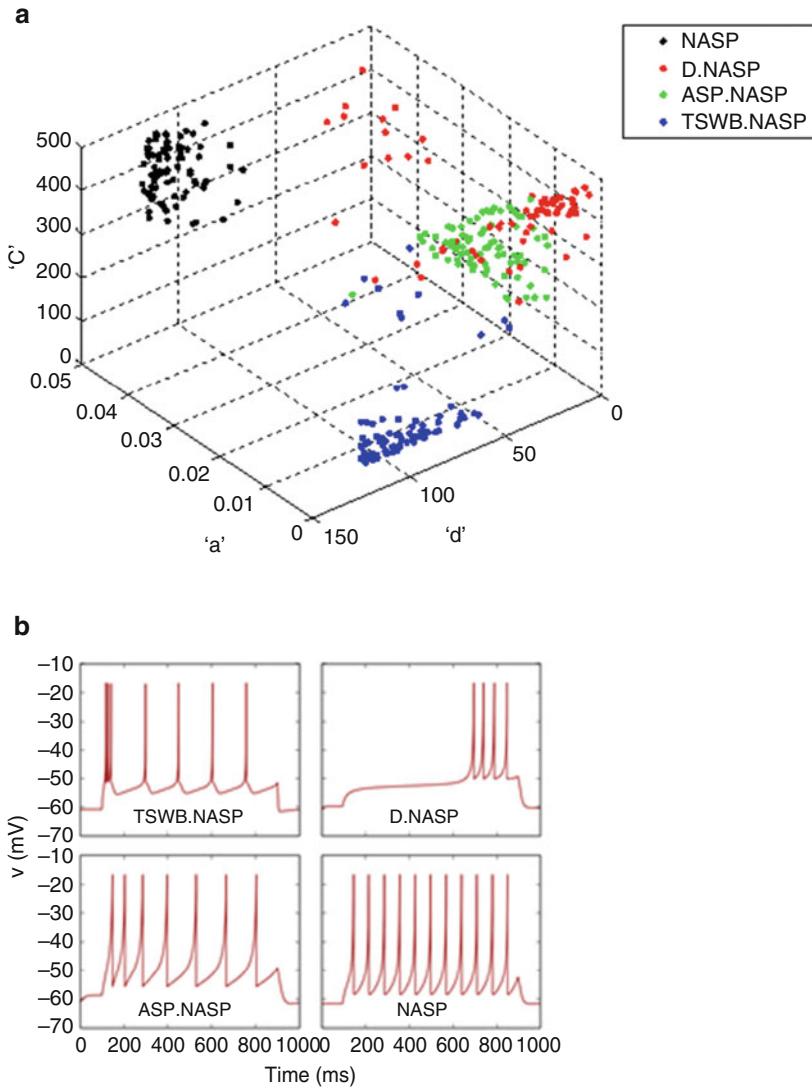


Fig. 5 Parameter space exploration with stochastic EA trials (a) Regions in the IM parameter space representing four different spike pattern types. Only the dimensions ‘ C ’, ‘ a ’, and ‘ d ’ are shown. (b) Representative IM responses for four spike pattern types TSWB.NASP, D.NASP, ASP.NASP and NASP. The models were constrained using arbitrarily defined features for each spike pattern type

6 Summary

In sum, defining spike pattern behaviors in the IM parameter space might reveal their characterizing features and will help to fill the gap in the knowledge about certain behaviors. In addition, this will be valuable to large scale SNN simulations that aim to investigate the collective dynamics of several neuron types with distinct behaviors. In order to create a biologically realistic yet computationally efficient SNN, its individual neuron models should capture the complex behaviors of the real neurons. Achieving this goal will provide a platform to develop theories that could causally link the neuro-computational properties at neuronal level to network behavior and to the emerging computational properties at the network level. This requires techniques that are not only capable of optimizing the model parameters, but also capable of effectively exploring the parameter space.

Acknowledgements We thank Alexander Komendantov and Eric Siggy Scott for their contribution and suggestions.

Support NINDS (NIH) R01NS39600, NSF IIS-1302256

References

1. Buzsáki, G., Chrobak, J.J.: Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks. *Curr. Opin. Neurobiol.* **5**(4), 504–510 (1995)
2. O’Keefe, J., Recce, M.L.: Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*. **3**(3), 317–330 (1993)
3. Gütig, R., Sompolinsky, H.: The tempotron: a neuron that learns spike timing-based decisions. *Nat. Neurosci.* **9**(3), 420–428 (2006)
4. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**(4), 500–544 (1952)
5. Yu, Q., Tang, H., Tan, K.C., Yu, H.: A brain-inspired spiking neural network model with temporal encoding and learning. *Neurocomputing*. **138**, 3–13 (2014)
6. Ghosh-Dastidar, S., Adeli, H.: Spiking neural networks. *Int. J. Neural Syst.* **19**(04), 295–308 (2009)
7. Saraga, F., Wu, C.P., Zhang, L., Skinner, F.K.: Active dendrites and spike propagation in multicompartment models of oriens-lacunosum/moleculare hippocampal interneurons. *J. Physiol.* **552**(3), 673–689 (2003)
8. Yu, Y., Shu, Y., McCormick, D.A.: Cortical action potential backpropagation explains spike threshold variability and rapid-onset kinetics. *J. Neurosci.* **28**(29), 7260–7272 (2008)
9. Goldwyn, J.H., Shea-Brown, E.: The what and where of adding channel noise to the Hodgkin-Huxley equations. *PLoS Comput. Biol.* **7**(11), e1002247 (2011)
10. Izhikevich, E.M.: *Dynamical systems in neuroscience*. MIT Press, Cambridge (2007)
11. Izhikevich, E.M.: Neural excitability, spiking and bursting. *Int. J. Bifurcation Chaos*. **10**(06), 1171–1266 (2000)
12. Izhikevich, E.M.: Resonate-and-fire neurons. *Neural Netw.* **14**(6), 883–894 (2001)
13. Morris, C., Lecar, H.: Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.* **35**(1), 193 (1981)

14. Izhikevich, E.M.: Which model to use for cortical spiking neurons? *IEEE Trans. Neural Netw.* **15**(5), 1063–1070 (2004)
15. Smith, G.D., Cox, C.L., Sherman, S.M., Rinzel, J.: Fourier analysis of sinusoidally driven thalamocortical relay neurons and a minimal integrate-and-fire-or-burst model. *J Neurophysiol.* **83**(1), 588–610 (2000)
16. Ermentrout, B.: Type I membranes, phase resetting curves, and synchrony. *Neural Comput.* **8**(5), 979–1001 (1996)
17. Brette, R., Gerstner, W.: Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol.* **94**(5), 3637–3642 (2005)
18. Jolivet, R., Schürmann, F., Berger, T.K., Naud, R., Gerstner, W., Roth, A.: The quantitative single-neuron modeling competition. *Biol. Cybern.* **99**(4–5), 417–426 (2008)
19. Izhikevich, E.M.: Simple model of spiking neurons. *IEEE Trans. Neural Netw.* **14**(6), 1569–1572 (2003)
20. Izhikevich, E.M.: Hybrid spiking models. *Philos. Trans. R. Soc. Lond. A.* **368**(1930), 5061–5070 (2010)
21. Izhikevich, E.M., Edelman, G.M.: Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci.* **105**(9), 3593–3598 (2008)
22. Beyeler, M., Richert, M., Dutt, N.D., Krichmar, J.L.: Efficient spiking neural network model of pattern motion selectivity in visual cortex. *Neuroinformatics.* **12**(3), 435–454 (2014)
23. Nageswaran, J.M., Dutt, N., Krichmar, J.L., Nicolau, A., Veidenbaum, A.V.: A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Netw.* **22**(5), 791–800 (2009)
24. Wheeler, D.W., White, C.M., Rees, C.L., Komendantov, A.O., Hamilton, D.J., Ascoli, G.A.: [Hippocampome.org](#): a knowledge base of neuron types in the rodent hippocampus. *Elife.* e09960 (2015)
25. Komendantov, A.O., Wheeler, D.W., Reese, C.L., Hamilton, D.J., White, C., Venkadesh, S., Ascoli, G.A.: Firing pattern classification of hippocampal neurons. *Soc. Neurosci. Conf. Poster.* **560**, 14 (2014)
26. Cope, D.W., Maccaferri, G., Marton, L.F., Roberts, J.D.B., Cobden, P.M., Somogyi, P.: Cholecystokinin-immunopositive basket and Schaffer collateral-associated interneurones target different domains of pyramidal cells in the CA1 area of the rat hippocampus. *Neuroscience.* **109**(1), 63–80 (2002)
27. Podlogar, M., Dietrich, D.: Firing pattern of rat hippocampal neurons: a perforated patch clamp study. *Brain Res.* **1085**(1), 95–101 (2006)
28. Armstrong, C., Szabadics, J., Tamas, G., Soltesz, I.: Neurogliaform cells in the molecular layer of the dentate gyrus as feed-forward γ -aminobutyric acidergic modulators of entorhinal–hippocampal interplay. *J. Comp. Neurol.* **519**(8), 1476–1491 (2011)
29. Chevaleyre, V., Siegelbaum, S.A.: Strong CA2 pyramidal neuron synapses define a powerful disynaptic cortico-hippocampal loop. *Neuron.* **66**(4), 560–572 (2010)
30. Canto, C.B., Witter, M.P.: Cellular properties of principal neurons in the rat entorhinal cortex. II. The medial entorhinal cortex. *Hippocampus.* **22**(6), 1277–1299 (2012)
31. Mercer, A., Botcher, N.A., Eastlake, K., Thomson, A.M.: SP–SR interneurones: a novel class of neurones of the CA2 region of the hippocampus. *Hippocampus.* **22**(8), 1758–1769 (2012)
32. Hamam, B.N., Amaral, D.G., Alonso, A.A.: Morphological and electrophysiological characteristics of layer V neurons of the rat lateral entorhinal cortex. *J. Comp. Neurol.* **451**(1), 45–61 (2002)
33. Zemankovics, R., Káli, S., Paulsen, O., Freund, T.F., Hájos, N.: Differences in subthreshold resonance of hippocampal pyramidal cells and interneurons: the role of h-current and passive membrane characteristics. *J. Physiol.* **588**(12), 2109–2132 (2010)
34. Pawelzik, H., Hughes, D.I., Thomson, A.M.: Physiological and morphological diversity of immunocytochemically defined parvalbumin-and cholecystokinin-positive interneurones in CA1 of the adult rat hippocampus. *J. Comp. Neurol.* **443**(4), 346–367 (2002)
35. Rossant, C., Goodman, D.F., Platkiewicz, J., Brette, R.: Automatic fitting of spiking neuron models to electrophysiological recordings. *Front. Neuroinform.* **4**, 2 (2010)

36. Lynch, E.P., Houghton, C.J.: Parameter estimation of neuron models using in-vitro and in-vivo electrophysiological data. *Front. Neuroinform.* **9**, 10 (2015)
37. Rossant, C., Goodman, D.F., Fontaine, B., Platkiewicz, J., Magnusson, A.K., Brette, R.: Fitting neuron models to spike trains. *Front. Neurosci.* **5**, 9 (2011)
38. Gerken, W.C., Purvis, L.K., Butera, R.J.: Genetic algorithm for optimization and specification of a neuron model. *Neurocomputing*. **69**(10), 1039–1042 (2006)
39. Van Geit, W., Achard, P., De Schutter, E.: Neurofitter: a parameter tuning package for a wide range of electrophysiological neuron models. *BMC Neurosci.* **8**(Suppl 2), P5 (2007)
40. Keren, N., Peled, N., Korngreen, A.: Constraining compartmental models using multiple voltage recordings and genetic algorithms. *J. Neurophysiol.* **94**(6), 3730–3742 (2005)
41. Druckmann, S., Banitt, Y., Gidon, A., Schürmann, F., Markram, H., Segev, I.: A novel multiple objective optimization framework for constraining conductance-based neuron models by experimental data. *Front. Neurosci.* **1**(1), 7 (2007)
42. Van Geit, W., De Schutter, E., Achard, P.: Automated neuron model optimization techniques: a review. *Biol. Cybern.* **99**(4-5), 241–251 (2008)
43. LeMasson, G., Maex, R.: Introduction to equation solving and parameter fitting. *Computational neuroscience: realistic modeling for experimentalists*, pp. 1–23. CRC Press, London (2001)
44. Hemond, P., Epstein, D., Boley, A., Migliore, M., Ascoli, G.A., Jaffe, D.B.: Distinct classes of pyramidal cells exhibit mutually exclusive firing patterns in hippocampal area CA3b. *Hippocampus*. **18**(4), 411 (2008)
45. Chiong, R. (ed.): *Nature-inspired algorithms for optimisation*, vol. 193. Springer, Berlin (2009)
46. Eshelman, J.D.S.L.J.: Spurious correlations and premature convergence in genetic algorithms. *Found. Genet. Algorithm*. **1**, 102 (2014)
47. Kolarov, K.: Landscape ruggedness in evolutionary algorithms. In: *Proceedings of IEEE International Conference on Evolutionary Computation*, pp. 19–24. IEEE (1997)
48. Davidor, Y.: Epistasis variance: a viewpoint on GA-hardness. *Found. Genet. Algorithm*. **1**, 23–35 (1991)
49. De Jong, K.A.: *Evolutionary computation: a unified approach*. MIT Press, Cambridge (2006)
50. Deb, K.: *Multi-objective optimization using evolutionary algorithms*, vol. 16. Wiley, New York (2001)
51. Coello, C.A.C., Van Veldhuizen, D.A., Lamont, G.B.: *Evolutionary algorithms for solving multi-objective problems*, vol. 242. Kluwer Academic, New York (2002)

Chemotactic Guidance of Growth Cones: A Hybrid Computational Model



Iolanda Morana Roccasalvo, Silvestro Micera, and Pier Nicola Sergi

Abstract The strategy used by axons to find the correct paths during the nervous system development is not yet completely understood. In addition, some emergent counterintuitive axonal responses were described in chemical active environments. Here, a computational model (Roccasalvo et al. Sci Rep 5:11340, 2015, <https://doi.org/10.1038/srep11340>) is presented to reproduce the key role of the diffusion-driven instability to form intracellular patterns during axonal pathfinding. The dependence of the pattern geometry on the diffusion constants of chemicals was studied for a simple case. In addition, a suitable non standard iterative Gierer-Meinhardt system was implemented (Roccasalvo et al. Sci Rep 5:11340, 2015, <https://doi.org/10.1038/srep11340>) to account for the interaction with extracellular gradients. The formation of neuritic paths was quantitatively reproduced in regular conditions, while emergent and counterintuitive phenomena, resulting from interactions between intracellular and extracellular environments, were qualitatively predicted (Roccasalvo et al. Sci Rep 5:11340, 2015, <https://doi.org/10.1038/srep11340>).

I.M. Roccasalvo

Translational Neural Engineering Area, The Biorobotics Institute, Scuola Superiore Sant'Anna, Pisa 56025, Italy

Present address: AMET srl 10149, Torino, Italy

S. Micera

Translational Neural Engineering Area, The Biorobotics Institute, Scuola Superiore Sant'Anna, Pisa 56025, Italy

Bertarelli Foundation Chair in Translational Neuroengineering, Center for Neuroprosthetics and Institute of Bioengineering, School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland

P.N. Sergi (✉)

Translational Neural Engineering Area, The Biorobotics Institute, Scuola Superiore Sant'Anna, Pisa 56025, Italy

e-mail: pn.sergi@gmail.com; p.sergi@sssup.it

1 Introduction

Pathfinding axons establish connections to wire and develop the nervous system [20], navigating through the body towards specific targets by sensing environmental characteristics [16] and by following diffusible gradients of chemical cues [4, 6, 22, 32, 35]. In particular, the growth cone (GC) [25] is able to sense diffusible gradients and to move toward secretive targets [13, 14, 23, 29, 30] through a chemotactic guidance process, involving the amplification of external chemical signals through an internal transduction process [30]. Calcium (Ca^{2+}) [11] is a key regulator of several phenomena involved in axonal guidance, as the promotion and inhibition of the outgrowth of axons and the turning of the GC [12]. In addition, more complex, and unexpected axonal behaviours were described when the concentration of extracellular calcium changed, so the attractive/repulsive nature of guidance cues was related to both intracellular and extracellular calcium (i.e., $[\text{Ca}^{2+}]_i$ and $[\text{Ca}^{2+}]_e$) [37]. Although there is a wide interest to obtain a set of “main rules” governing neuritic chemotaxis to replicate biological behaviours through simplified in silico models [26–28], the nature of these rules is not totally clear and interdisciplinary studies have been performed in order to link traditional and in silico experiments [2, 21]. To further explore this field, a Java open-source software, CX3D [38] (Institute of Neuroinformatics of ETH, Zurich), was used to implement the developing neural networks in a three-dimensional physical space. The growth cones were biomimetically implemented as monoconnected domains, and the whole evolution of the internal chemical field was implemented through a non standard Gierer-Meinhardt (GM) system [1, 19]. This phenomenological approach has been used to model the emergence of complex patterns in living systems, starting from the reciprocal interaction of chemicals [19, 34]. Finally, the axonal outgrowth was reproduced in silico, accounting for the activity of growth cone and filopodia: the suitability of the generated trajectories was tested, for regular cases, through a quantitative comparison with literature experiments, while unexpected behaviours were qualitatively reproduced.

2 Methods

2.1 Evolution of Intracellular Chemical Fields Within the GC Domain

A reaction-diffusion PDE system, with suitable boundary conditions, was used to describe the evolution of the intracellular chemical fields within the GC. This system was written in a general form as:

$$\begin{cases} \dot{a} = f(a, h) + D_a \Delta a, & \mathbf{x} \in \Omega, t \geq 0 \\ \dot{h} = g(a, h) + D_h \Delta h, & \mathbf{x} \in \Omega, t \geq 0 \end{cases} \quad (1)$$

where $a(\mathbf{x}, t)$ and $h(\mathbf{x}, t)$ were the concentrations of activator and inhibitor in a simply connected bounded domain $\Omega \subset R^n$. The effects of diffusion were considered through the functions $\Delta a(\mathbf{x}, t)$, $\Delta h(\mathbf{x}, t)$, together with the diffusion constants D_a and D_h , while the reaction kinetics was described through the functions $f(a, h)$ and $g(a, h)$, which were non linear in a and h . First, to explore the behaviour of this kind of system, the following simplified problem was considered and numerically solved:

$$\begin{cases} \dot{a} = f(a, h) + D_a \Delta a, & \mathbf{x} \in \Omega \\ \dot{h} = g(a, h) + D_h \Delta h, & \mathbf{x} \in \Omega \\ \nabla a \cdot \mathbf{n} = \nabla h \cdot \mathbf{n} = 0, & \mathbf{x} \in \partial\Omega \end{cases} \quad (2)$$

where $\Omega := [0, L] \times [0, L] \subseteq R^2$, $L \in R$ was the size of the bidimensional domain, and \mathbf{n} was the unit outer normal on $\partial\Omega$. According to [10], the reaction kinetics was modelled through a couple of non linear relationships involving the activator and the inhibitor compounds: $f(a, h) = Aa^r/h^s - Ba + C$, $g(a, h) = Da^t/h^u - Eh + F$, where $A, B, C, D, E, F \in R$ and $r = t = 2$, $s = 1$, $u = 0$. Furthermore, for sake of simplicity, $D = A$, $E = B$ and $C = F = 0$. To explore the spatially homogenous solutions, the ODE part of System (2) was considered, while to account for the emergence of complex patterns due to diffusion-driven instability, the whole PDE system was investigated. First, System (2) was linearized about a uniform stationary state (a_0, h_0) then the condition to find asymptotically unstable solutions was expressed as:

$$D_a g_h + D_h f_a > 2 \sqrt{D_a D_h (f_a g_h - f_h g_a)} \quad (3)$$

where f_a , f_h , g_a , g_h , were the elements of the Jacobian matrix of the linearized system. Finally, the space of instability of the system was explored as a function of the parameter $d = D_h/D_a$ (i.e., the ratio between diffusibility of inhibitor and activator), and the occurrence of different kinds of patterns was tested.

Moreover, a multistep variant form of System (2) was implemented to explore the influence of extracellular chemical gradients. According to [1, 10], both $f(a, h)$ and $g(a, h)$ were qualitatively related to the activator and the inhibitor compounds, thus $A = \rho$, $B = \mu_a$, $D = A$, $E = \mu_h$. As a consequence, the previous system was rewritten as:

$$\begin{cases} \dot{a} = \rho \frac{a^2}{h} - \mu_a a + C + D_a \Delta a, & \mathbf{x} \in \Psi, 0 \leq t \leq t^* \\ \dot{h} = \rho a^2 - \mu_h h + F + D_h \Delta h, & \mathbf{x} \in \Psi, 0 \leq t \leq t^* \end{cases} \quad (4)$$

where, ρ was the production coefficient of both activator and inhibitor, while μ_a and μ_h , $C > 0$ and $F > 0$, D_a and D_h , were, respectively, the decay coefficients, the source terms and the diffusion coefficients for activator and inhibitor compounds. In addition, the GC domain was defined as $\Psi := \{(x, y) \in R^2 : x^2/L_a^2 + y^2/L_b^2 - 1 \leq 0\}$, $L_b > L_a > 0$ and $L_a, L_b \in R$. In other words, the activator concentration $a(\mathbf{x}, t)$ within the elliptical GC domain (Ψ) depended on the contribution of four processes: diffusion, degradation, production, and interaction with an antagonist morphogen,

whose concentration was $h(\mathbf{x}, t)$, for $0 \leq t \leq t^*$ (where t^* was the ending time of each step). The value of $a(\mathbf{x}, 0)$ was defined as follows:

$$a(\mathbf{x}, 0) = a_{\text{basal}}(\mathbf{x}, 0) + f_{\text{rand}}(\mathbf{x}, 0) + f_{\text{grad}}(\mathbf{x}, 0) \quad (5)$$

where $a_{\text{basal}}(\mathbf{x}, 0)$ was the basal concentration of activator at the beginning of each step (i.e., $t = 0$), $\forall \mathbf{x} \in \Psi$, while $f_{\text{rand}}(\mathbf{x}, 0)$ was a random term due to the growth cone intracellular activity, at the beginning of each step [12]. Finally, $f_{\text{grad}}(\mathbf{x}, 0)$ was related to the GC sensitivity to very low gradients of extracellular chemical compounds, thus it was written as follows:

$$f_{\text{grad}}(\mathbf{x}, 0) = q(\mathbf{x}, 0)c_1 \exp\left(\frac{-\mathbf{r}^2}{2\sigma^2}\right) \quad (6)$$

where $q(\mathbf{x}, 0)$ was the external concentration of chemical for $\mathbf{x} \in \partial\Psi$, $c_1 \in R$, $\sigma \in R$, while \mathbf{r} represented the distance of each point of the GC domain (i.e., $x \in \Psi$) from the maximum sensitivity point. The location of this point resulted from the interaction between the extracellular calcium field and the nature of the external chemical source. In particular, in presence of an attractive source and normal concentration of Ca_e^{2+} , as well as in presence of a repulsive source and high concentration of Ca_e^{2+} , this point was located at the base of the filopodium closest to the external source and extending in direction of the maximum gradient. On the contrary, for an attractive source with low extracellular concentration of Ca_e^{2+} , as well as for a repulsive source with normal concentration of Ca_e^{2+} , this point was located at the base of a contralateral filopodium, symmetrically with reference to the previous two cases. System (3) was iteratively implemented within the CX3D framework, where diffusive processes were handled fostering a computational approach, similar to the standard finite volumes method [1, 38].

2.2 Computational Model of Axonal Outgrowth Guided by Chemotaxis

In silico growth cones were implemented as elliptic sets of physical nodes (Fig. 1a). Initial random values of activator a and inhibitor h were assigned from uniform distributions, to account for biological fluctuation within the growth cone domain, for each node and for each growth step. Several filopodia, extending from the growth cone border were also added and modelled as monodimensional segments (see Fig. 1a-f). Moreover, for each growth cone and for each step of growth, the number and the angular disposition of filopodia were randomly chosen to reproduce biological variability and stochastic fluctuations of shape. The extracellular gradient was implemented as a point source (e.g., a single ejection of a chemical solution [18]), which was placed 100 μm away from the centre of the growth cone and at a 45° angle with respect to y-axis [33]. The black arrow in Fig. 1a shows the gradient

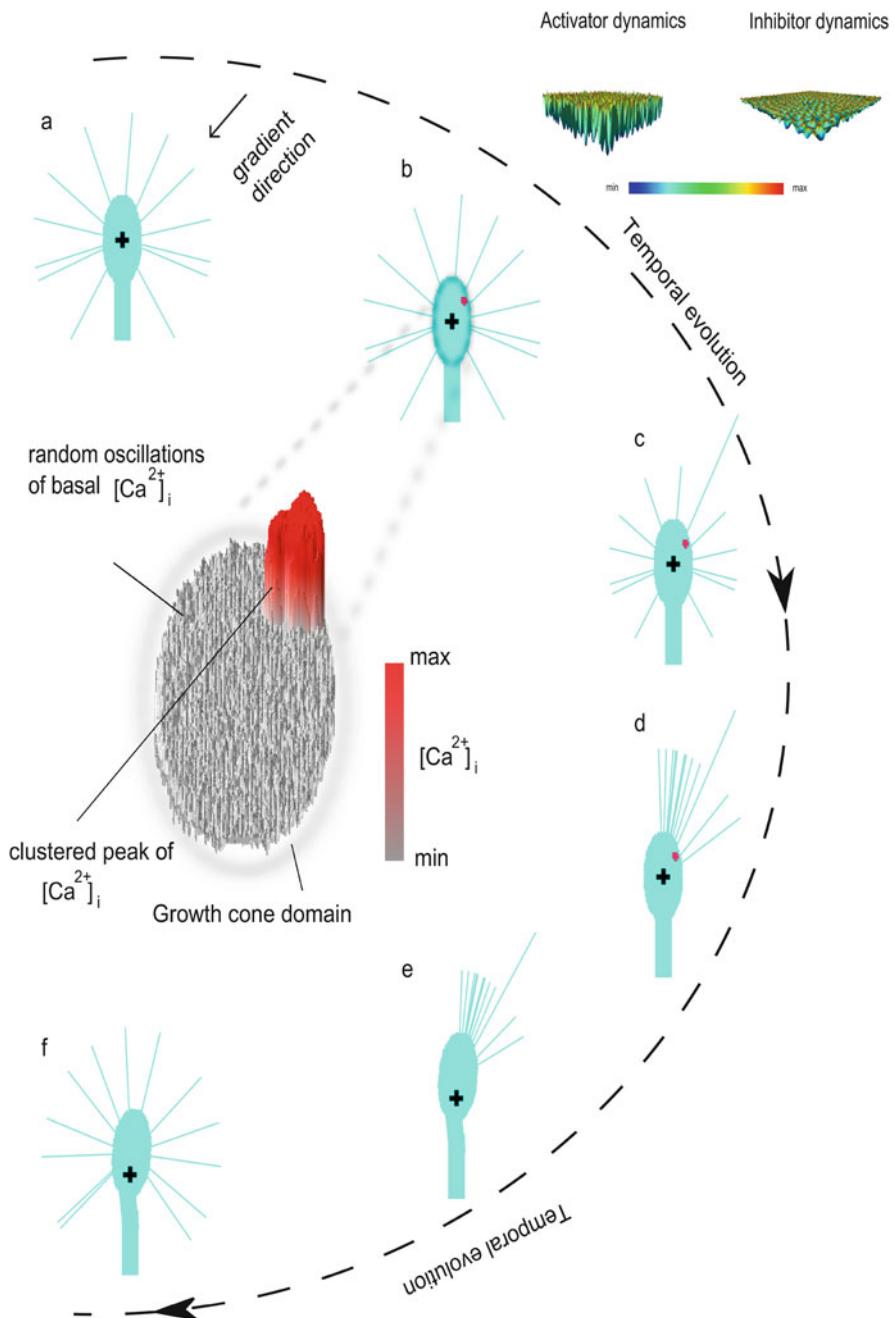


Fig. 1 Temporal evolution of the turning response of in silico growth cones. The temporal evolution (external dotted line) of an in silico growth cone exposed to an attractive gradient was

direction. The outgrowth of axons was modelled as a multistep process, accounting for the whole length of each trajectory (taken from a Gaussian distribution), as well as the number of “decision” points according to experimental trials [33]. More specifically, the turning response was cyclic for each step. First, intracellular gradients of morphogens (activator (i.e. $[Ca^{2+}]_i$), and inhibitor) evolved according to System (3), where $\rho = 35$ [1/s], $D_a = 8$ [$\mu m^2/s$], $D_h = 8000$ [$\mu m^2/s$], $\mu_a = 0.08$ [1/s], $\mu_h = 0.16$ [1/s], $C = F = 50$ [$\mu mol/(\mu l s)$]. All these values were taken from literature [1], since they were inside the space of instability of System (3), and therefore suitable to result in diffusion-driven patterns. The initial conditions of this system accounted for the basal amount of activator together with its random fluctuations as well as the influence the external chemical concentration through the sensitivity of the GC. The $a(x, t)$ converged to asymmetric patterns, where one or more clusters were polarized towards the chemical source (Fig. 1b). The elongation dynamics of filopodia was also implemented (Fig. 1c): the leading filopodium was chosen to be angularly close to the peak of activator, according to experimental observations, which link local intracellular calcium peaks to the enhanced outgrowth of filopodia [3]. On the contrary, the other filopodia retracted, while new protrusions sprouted close to the leading filopodium (Fig. 1d). Thus, the turning of *in silico* growth cones was preceded by the preferential protrusion of filopodia on the turning side, according to the biological behaviour [5, 15, 17, 36, 37]. Then, the whole growth cone simultaneously advanced (Fig. 1e) and turned so that its new orientation was a linear combination between its previous orientation and the actual angular direction of the leading filopodium. Finally, filopodia were symmetrically redistributed on the growth cone surface (Fig. 1f).

2.3 Quantitative Evaluation of Growth Cone Model Performance

The ability of the model in reproducing the biological behaviour of outgrowing axons was assessed through quantitative indexes. The displacement of the barycentre was used to trace the axonal paths and to reduce the trajectories to monodimensional lines. In *silico* turning angle and tortuosity of traces were

Fig. 1 (continued) shown (the arrow showed the gradient direction, see (a)). In particular, a cluster of activator (intracellular calcium) arose within the growth cone side facing the external gradient (b). Random oscillations of the basal intracellular calcium together with the calcium peak were shown within the growth cone domain (central magnification of b). ((c)–(f)) The position of the intracellular calcium peak influenced the choice of leading filopodium that guided the growth cone advance. In all CX3D images, the black cross indicates the growth cone barycentre at the beginning of assay. On the right corner of the figure modular surfaces of coupled activator and inhibitor dynamics of a Turing-like model simultaneously present within the simulation domain. Modified from [24] under a Creative Common Attribution 4.0 International License

analysed and compared with the same experimental parameters. More specifically, the turning angle was defined as the angle between the original direction of axon outgrowth and a straight line connecting the positions of growth cone barycentre at the onset and the end of turning assay. Finally, the tortuosity of traces was measured as ratio of curvilinear length of a trajectory to the straight length linking its beginning with its end [2].

3 Results

3.1 Diffusion-Driven Instability

A phase portrait of the ODE part of the PDE system (2) was used to explore the behavior of spatial homogeneous solutions. In particular, since $A = D = 0.01$, two equilibrium points were found at $(1, 1)$ and $(0, 0)$. The origin behaved as a “mixed” node, while $(1, 1)$ was a non linear center (Fig. 2a). In this case, the condition for the onset of activator patterns was:

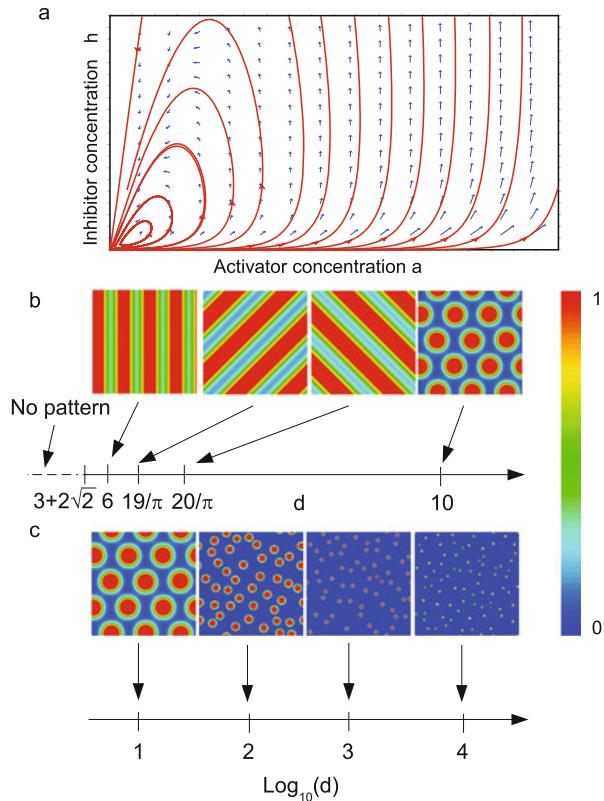
$$d - 1 > 2\sqrt{d} \quad (7)$$

which resulted in $d > 3 + 2\sqrt{2}$. In particular, an oscillatory behaviour, with no patterns, was found for $d = 3 + 2\sqrt{2}$, while for $d = 6$ several vertical stripes were created. Similarly, for $d = 19/\pi$ several ascending diagonal stripes arose, and, quite curiously, for the similar value of $d = 20/\pi$, descending diagonal stripes were present. Finally, for greater values of d stripes turned into a spotted pattern (see Fig. 2b), which was controlled by d . More specifically, the size of spots decreased with the increase of magnitude of d , as shown in Fig. 2c.

3.2 In Silico Paths of Outgrowing Axons

The trajectories of in silico growth cones ($n = 16$) were explicitly reproduced during a 1 hour exposure to a netrin gradient (see Fig. 3). The exposure to an attractive gradient first resulted in small deflections of trajectories in direction of the source, while, at a later stage, axons widely turned since they were more intensely guided by the shape of the extracellular gradient. Similarly, in silico growth cone traces ($n = 16$) were explicitly reproduced in control conditions (Fig. 3b). In this case, the simulated axons did not show any preferential advance direction or biased deflections, since no chemical gradients were present.

Fig. 2 Diffusion-driven instability. (a) The phase portrait of the ODE part of the PDE system (2). The plot of the direction field shows a center at the point (1,1) and a stable “mixed” node at the point (0,0). (b) The variability of the spatial pattern resulting from the PDE is shown as a function of the ratio d between the activator and the inhibitor diffusion constant. (c) The magnitude of d also influence the size of the spots



3.3 Quantitative Assessment of the Axonal Chemoattractive Response

The distributions of experimental [33] ($n = 16$) and simulated ($n = 16$) turning angles were reproduced after a 1-h exposure to a netrin gradient: the median values were similar (21.7° and 22.6° , respectively) as well as the whole distribution profiles. A Wilcoxon rank sum test was used to test the statistical significance of the difference between experimental and computational values (p -value = 0.8091, 95% confidence level) (Fig. 3c). In addition, the tortuosity of axonal paths was analysed (Fig. 3e) and resulted in 1.022 (median value) for the experimental samples [33] ($n = 10$), while for simulations the same value was 1.012 ($n = 16$). Again, a Wilcoxon rank sum test showed that no significant differences were found between experiments and simulations (p -value = 0.06045, 95% confidence level). Finally, a correlation ($R^2 = 0.95$) between experimental and in silico values of turning angles was found (as shown in Fig. 3g).

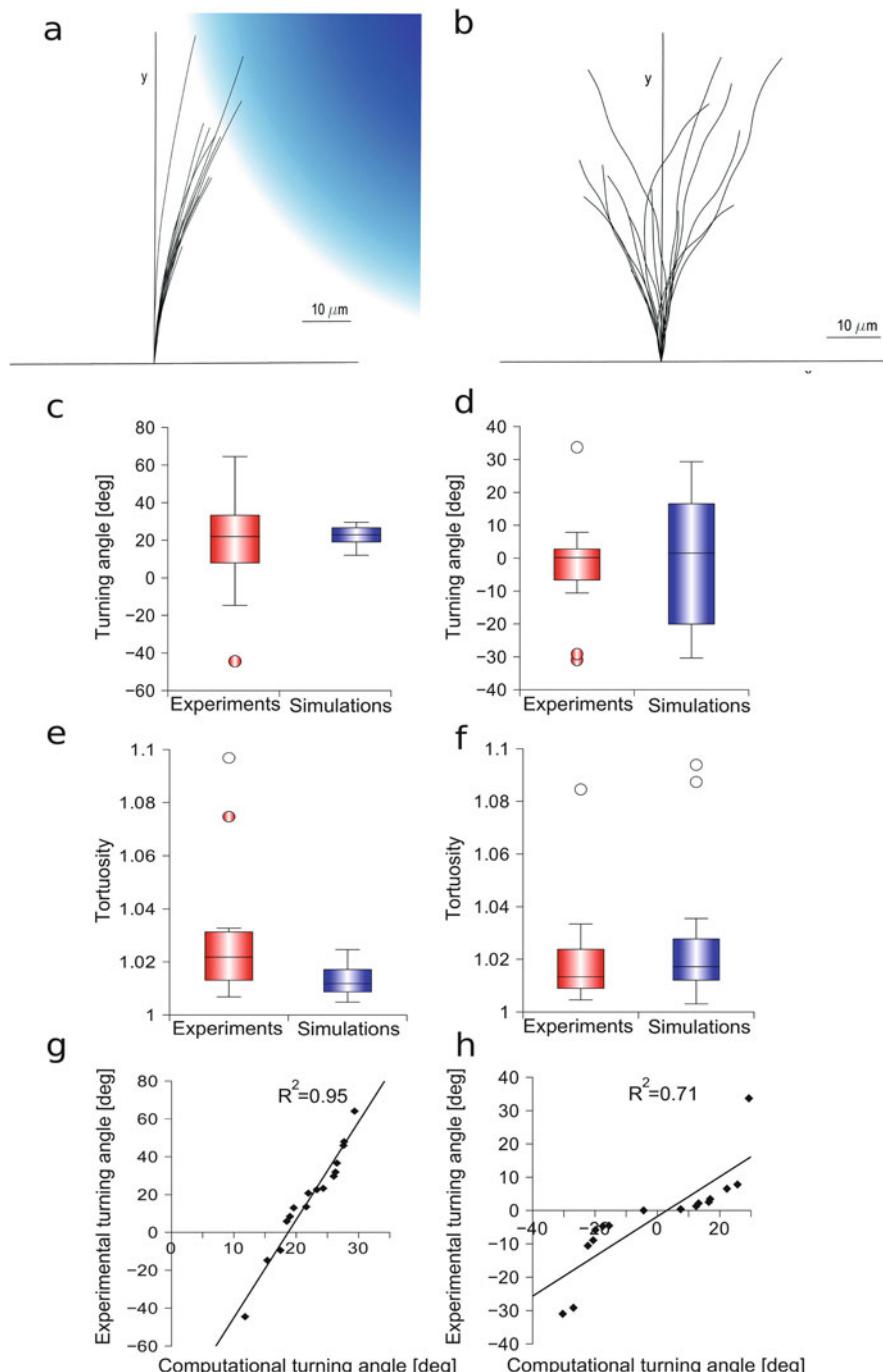


Fig. 3 In silico axonal outgrowth for different extracellular conditions and assessment of in silico model performance for chemoattractive turning and control conditions. The outgrowth trajectories

3.4 Quantitative Assessment of Axonal Outgrowth in Control Conditions

In Fig. 3d, the distributions of experimental turning angle [33] ($n=16$) and computational ($n=16$) samples were compared in control conditions (i.e., no extracellular gradients). The median values resulted, respectively, in 0.2° and 1.5° , while the range of variability was similar, and no statistically significant differences were found (Wilcoxon rank sum test, 95% confidence level, p -value = 0.6963). In Fig. 3f, experimental [33] ($n=7$) and computational ($n=16$) distributions of tortuosity are compared: the median values were respectively 1.013 and 1.017, with a similar range of variability. Finally, no statistically significant differences were found between simulations and experiments (Wilcoxon rank sum, 95% confidence level, p -value = 0.4921), while a positive correlation ($R^2 = 0.71$) was found between in silico and experimental distributions of turning angles (Fig. 3h).

3.5 Qualitative Predictions of Axonal Counterintuitive Behaviours

In silico axons accounted for the global effect of attractive/repulsive chemical sources coupled with the different levels of $[Ca^{2+}]_e$ [31]. The ability of the model in reproducing biological counterintuitive responses was then tested. In Fig. 4, the observed biological behaviours are schematically represented on the left side of the panel, while in silico traces ($n=3$, for each case) are shown on the right side. For each simulation, an extracellular chemical source was assumed to be placed 100 μm away from the growth cone and at a 45° angle with respect to the direction of y-axis (coloured gradients shows the nature of repulsive and attractive sources). When $[Ca^{2+}]_e$ was about 0.9 mM (i.e., normal culture conditions [9]),

Fig. 3 (continued) of in silico axons (a) for an attractive gradient of netrin (gradation of colour) and (b) in control conditions (without gradient). In both cases, 16 traces of outgrowing axons are shown after a 1 – hour assay. The explicit paths of in silico growth cones were measured in presence of an attractive netrin gradient and compared with the biological ones: (c) The difference between turning angles for experimental and simulated axons ($n=16$ axons, for each set) was not statistically significant. (d) Turning angles for experiments [33] and simulations in control conditions ($n=16$, for each set) were not significantly different. (e) The difference between tortuosity of experimental [33] ($n=10$) and computational ($n=16$) traces was not statistically significant. (f) Tortuosity values for experimental ($n=7$) and simulated ($n=16$) trajectories in control conditions were not significantly different. (g) Computational distribution of turning angles plotted versus experimental results with chemical gradient [33] ($n=16$ samples for each case): a linear relationship was found ($R^2 = 0.95$). (h) Computational distribution of turning angles plotted versus experimental results [33] ($n=16$ samples for each case): a linear relationship was found ($R^2 = 0.71$). Modified from [24] under a Creative Common Attribution 4.0 International License

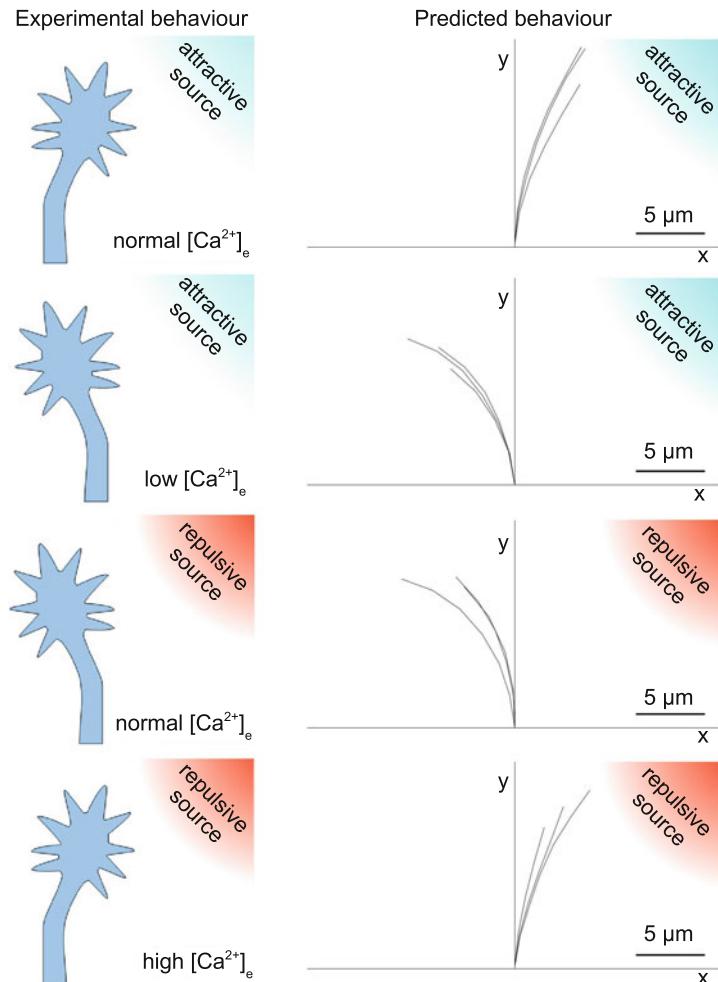


Fig. 4 Predictions of counterintuitive behaviours. Left side: cartoons of different biological behaviours of GC in presence of attractive and repulsive cues for different extracellular concentration of calcium ($[Ca^{2+}]_e$). Upper panel: regular attraction towards an attractive cue. Upper-middle panel: counterintuitive behaviour, that is GC repulsion away from an attractive cue. Lower-middle panel: regular behaviour, that is GC repulsion away from a repulsive cue. Lower panel: counterintuitive behaviour, that is GC attraction towards a repulsive cue. Right side: qualitative predictions of the model of the biological behaviours. Upper panel: regular attraction towards an attractive cue. Upper-middle panel: counterintuitive behaviour, that is GC repulsion away from an attractive cue. Lower-middle panel: regular behaviour, that is GC repulsion away from a repulsive cue. Lower panel: counterintuitive behaviour, that is GC attraction towards a repulsive cue. The traces of the growth cone trajectories ($n = 3$, for each case) are shown. Reprinted from [24] under a Creative Common Attribution 4.0 International License

the extracellular attractive gradient was able to attract the growth cone towards the source (upper panel). On the contrary, when the extracellular calcium level decreased (upper-middle panel), in silico growth cones turned away from the source, showing the influence of the decreasing level of $[Ca^{2+}]_e$ on the switch from attraction to repulsion, according to the biological behaviour. Moreover, when an extracellular repulsive gradient was implemented together with a normal level of $[Ca^{2+}]_e$ (0.9 mM), axons migrated away from the chemical source (lower-middle panel). However, an increment of the extracellular calcium level was able to elicit an attractive turning on the side of growth cone facing the chemical source (lower panel).

4 Discussion

In this work, a phenomenological approach was used to reproduce the onset of intracellular patterns due to diffusible chemicals inside the GC. This behaviour was implemented through a GM algorithm, which involved the presence of two compounds, referred as activator and inhibitor. The activator was autocatalytic, and also able to activate the inhibitor, while the inhibitor was able to inhibit the activator. This line of research, which makes abstraction of the natural biochemical complexity, refers to foundational works [10, 19, 34], which, from a side, model chemical reactions through quite simple reaction kinetics (as in Systems (2) and (3)), from the other side, phenomenologically describe the presence of complex patterns due to diffusion. Indeed, from suitable sets of parameters, belonging to the so called “Instability space” or “Turing space”, the system dynamics was able to amplify fluctuations of some finite wavelengths more than greater or smaller ones. As a consequence, a diffusion-driven instability arose also in apparently simple kinetics. In other words, whether within Systems (2) and (3) eigenfunctions were in form of $\cos(mx)$ or $\cos(nx)$ ($m, n \in Z$) [34] the resulting spatial pattern was striped (as in Fig. 2b for, $d = 6$, $d = 19/\pi$ and $d = 20/\pi$), while whether they were in form $\cos(mx)\cos(nx)$ [34] the resulting spatial pattern was spotted (as in Fig. 2b, for $d = 10^n$, $n \in Z$). In this work, the essential dependence of the spatial pattern on the ratio between diffusion constants was verified for both Systems (2) and (3), even if they were implemented in different computational frameworks and with different sets of parameters. In both cases, the spotted pattern was predictable from the value of d , which was 10^n , $n \in Z$ in System (2) and 10^3 in System (3). Although a standard GM implementation was effective, it resulted in deterministic patterns, dependent on internal parameters and equal for all steps of outgrowth. In other words, it was not suitable to reproduce the outgrowth of axons in chemical active environments. As a consequence, the iterative non standard GM system (3) was implemented and updated for each outgrowth step. This approach resulted in intracellular fields simultaneously dependent on intracellular and extracellular

conditions. More specifically, in this case, CaMKII and CaN likely played the role of activator and inhibitor compounds [31]. Nevertheless, the proposed approach was not essentially dependent on their nature, thus it is general and suitable to be applied to different chemical pathways, resulting in different final chemical compounds, able to play the role of activators and inhibitors. In addition, to keep simulations biomimetic, the non standard GM system was simultaneously applied to the whole domain of the GC and resulted in an internal gradient of calcium ions. Although in this work the GC shape was constant through the time, time-variant modifications of the GC shape can be considered as resulting from exocytosis (due to attraction) or endocytosis due to repulsion. Indeed, *in silico* GCs were sensible to the coupled action of extracellular chemical gradient and intracellular sensitivity. So, they were able to reproduce the natural phenomenon of “adaptation” for a wide range of extracellular calcium concentration. In addition, the GM-like system was iteratively implemented over the whole GC domain for every new displacement of the GC, in order to implement sensing/desensing cycles. Moreover, since natural GCs undergo several changes (e.g., shape, number of protruding filopodia, intracellular concentration of calcium ions, etc.), *in silico* GCs accounted for random effects through stochastic fluctuations on initial concentration of intracellular calcium and by changing homogeneous distribution of filopodia, mimicking their natural disposition. In conclusion, this model was able to explicitly reproduce the trajectories of outgrowing axons, starting from the internal biochemical fields. The *in silico* traces were compared to biological paths through quantitative indexes (turning angle, tortuosity) and for both control and attractive conditions to avoid a simply heuristic and visual comparison. Finally, this model was capable to reproduce some counterintuitive biological behaviours, as the attraction/repulsion switch depending on the extracellular concentration of calcium as well as on both the steepness of the internal calcium gradient and on the nature of the chemical source. Although a quantitative comparison with experimental results was provided for regular cases, the comparison with unexpected (or “counterintuitive”) behaviours was qualitative, and more experimental data are needed to further calibrate the model. Nevertheless, the main achievement of this approach was to explicitly reproduce the axonal paths, consistently to the interplay between intra and extracellular environments. In addition, since *in silico* paths of outgrowing axons were iteratively implemented and, for each step, the growth cone accounted for the interactions between intracellular and extracellular environments, time-variant phenomena belonging to a multiphysics domain could be further added. In particular, different chemical sources, different topographies of substrates [8], electromagnetic fields as light or laser [7], as well as time-variant and coupled phenomena, could be used to extend the present model.

References

1. Bauer, R., Zubler, F., Hauri, A., Muir, D.R., Douglas, R.J.: Developmental origin of patchy axonal connectivity in the neocortex: a computational model. *Cereb. Cortex* **24**(2), 487–500 (2012). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3888370/>

2. Borisuk, R., al Azad, A.K., Conte, D., Roberts, A., Soffe, S.R.: A developmental approach to predicting neuronal connectivity from small biological datasets: a gradient-based neuron growth model. *PLoS One* **9**(2), 1–15 (2014). <https://doi.org/10.1371/journal.pone.0089461>
3. Cheng, S., Gessell, M.S., Rehder, V.: Local calcium changes regulate the length of growth cone filopodia. *J. Neurobiol.* **50**(4), 263–275 (2002). <http://dx.doi.org/10.1002/neu.10027>
4. Chilton, J.K.: Molecular mechanisms of axon guidance. *Dev. Biol.* **292**(1), 13–24 (2006). <http://doi.org/10.1016/j.ydbio.2005.12.048>. <http://www.sciencedirect.com/science/article/pii/S0012160605009437>
5. Davenport, R.W., Kater, S.B.: Local increases in intracellular calcium elicit local filopodial responses in helisoma neuronal growth cones. *Neuron* **9**(3), 405–416 (1992). [http://dx.doi.org/10.1016/0896-6273\(92\)90179-H](http://dx.doi.org/10.1016/0896-6273(92)90179-H)
6. Dickson, B.J.: Molecular mechanisms of axon guidance. *Science* **298**(5600), 1959–1964 (2002). <https://doi.org/10.1126/science.1072165>. <http://science.sciencemag.org/content/298/5600/1959>
7. Ehrlicher, A., Betz, T., Stuhrmann, B., Koch, D., Milner, V., Raizen, M.G., Käs, J.: Guiding neuronal growth with light. *Proc. Natl. Acad. Sci.* **99**(25), 16024–16028 (2002). <https://doi.org/10.1073/pnas.252631899>. <http://www.pnas.org/content/99/25/16024.abstract>
8. Ferrari, A., Cecchini, M., Dhawan, A., Micera, S., Tonazzini, I., Stabile, R., Pisignano, D., Beltram, F.: Nanotopographic control of neuronal polarity. *Nano Lett.* **11**(2), 505–511 (2011). <http://dx.doi.org/10.1021/nl103349s>. PMID: 21241061
9. Forbes, E., Thompson, A., Yuan, J., Goodhill, G.: Calcium and camp levels interact to determine attraction versus repulsion in axon guidance. *Neuron* **74**(3), 490–503. <http://dx.doi.org/10.1016/j.neuron.2012.02.035>
10. Gierer, A., Meinhardt, H.: A theory of biological pattern formation. *Kybernetik* **12**(1), 30–39 (1972). <http://dx.doi.org/10.1007/BF00289234>
11. Gomez, T.M., Spitzer, N.C.: Regulation of growth cone behavior by calcium: new dynamics to earlier perspectives. *J. Neurobiol.* **44**(2), 174–183 (2000). [http://dx.doi.org/10.1002/1097-4695\(200008\)44:2<174::AID-NEU7>3.0.CO;2-R](http://dx.doi.org/10.1002/1097-4695(200008)44:2<174::AID-NEU7>3.0.CO;2-R)
12. Gomez, T.M., Zheng, J.Q.: The molecular basis for calcium-dependent axon pathfinding. *Nat. Rev. Neurosci.* **7**(2), 115–125 (2006). <http://dx.doi.org/10.1038/nrn1844>
13. Goodhill, G.J.: Diffusion in axon guidance. *Eur. J. Neurosci.* **9**(7), 1414–1421 (1997). <http://dx.doi.org/10.1111/j.1460-9568.1997.tb01496.x>
14. Goodhill, G.J., Baier, H.: Axon guidance: stretching gradients to the limit. *Neural Comput.* **10**(3), 521–527 (1998). <http://dx.doi.org/10.1162/089976698300017638>
15. Gundersen, R.W., Barrett, J.N.: Characterization of the turning response of dorsal root neurites toward nerve growth factor. *J. Cell Biol.* **87**(3), 546–554 (1980). <https://doi.org/10.1083/jcb.87.3.546>. <http://jcb.rupress.org/content/87/3/546>
16. Katz, L.C., Crowley, J.C.: Development of cortical circuits: lessons from ocular dominance columns. *Nat. Rev. Neurosci.* **3**(1), 34–42 (2002). <http://dx.doi.org/10.1038/nrn703>
17. Lau, P.m., Zucker, R.S., Bentley, D.: Induction of filopodia by direct local elevation of intracellular calcium ion concentration. *J. Cell Biol.* **145**(6), 1265–1276 (1999). <https://doi.org/10.1083/jcb.145.6.1265>. <http://jcb.rupress.org/content/145/6/1265>
18. Lohof, A., Quillan, M., Dan, Y., Poo, M.: Asymmetric modulation of cytosolic camp activity induces growth cone turning. *J. Neurosci.* **12**(4), 1253–1261 (1992). <http://www.jneurosci.org/content/12/4/1253>
19. Meinhardt, H.: Orientation of chemotactic cells and growth cones: models and mechanisms. *J. Cell Sci.* **112**(17), 2867–2874 (1999). <http://jcs.biologists.org/content/112/17/2867>
20. Ming, G.I., Wong, S.T., Henley, J., Yuan, X.b., Song, H.j., Spitzer, N.C., Poo, M.m.: Adaptation in the chemotactic guidance of nerve growth cones. *Nature* **417**(6887), 411–418 (2002). <http://dx.doi.org/10.1038/nature745>
21. Mortimer, D., Pujic, Z., Vaughan, T., Thompson, A.W., Feldner, J., Vetter, I., Goodhill, G.J.: Axon guidance by growth-rate modulation. *Proc. Natl. Acad. Sci.* **107**(11), 5202–5207 (2010). <https://doi.org/10.1073/pnas.0909254107>. <http://www.pnas.org/content/107/11/5202.abstract>

22. Mueller, B.: Growth cone guidance: first steps towards a deeper understanding. *Annu. Rev. Neurosci.* **22** (1999)
23. O'Donnell, M., Chance, R.K., Bashaw, G.J.: Axon growth and guidance: receptor regulation and signal transduction. *Annu. Rev. Neurosci.* **32**, 383–412 (2009). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4765433/>
24. Roccasalvo, I.M., Micera, S., Sergi, P.N.: A hybrid computational model to predict chemotactic guidance of growth cones. *Sci. Rep.* **5**, 11340 (2015). <https://doi.org/10.1038/srep11340>
25. Rosoff, W.J., Urbach, J.S., Esrick, M.A., McAllister, R.G., Richards, L.J., Goodhill, G.J.: A new chemotaxis assay shows the extreme sensitivity of axons to molecular gradients. *Nat. Neurosci.* **7**(6), 678–682 (2004). <http://dx.doi.org/10.1038/nn1259>
26. Sergi, P.N., Cavalcanti-Adam, E.A.: Biomaterials and computation: a strategic alliance to investigate emergent responses of neural cells. *Biomater. Sci.* **5**, 648–657 (2017). <http://dx.doi.org/10.1039/C6BM00871B>
27. Sergi, P.N., Marino, A., Ciofani, G.: Deterministic control of mean alignment and elongation of neuron-like cells by grating geometry: a computational approach. *Integr. Biol.* **7**, 1242–1252 (2015). <http://dx.doi.org/10.1039/C5IB00045A>
28. Sergi, P.N., Morana Roccasalvo, I., Tonazzini, I., Cecchini, M., Micera, S.: Cell guidance on nanogratings: a computational model of the interplay between PC12 growth cones and nanostructures. *PLoS One* **8**(8), e70304 (2013). <https://doi.org/10.1371/journal.pone.0070304>
29. Shimozono, S., Iimura, T., Kitaguchi, T., Higashijima, S.I., Miyawaki, A.: Visualization of an endogenous retinoic acid gradient across embryonic development. *Nature* **496**(7445), 363–366 (2013). <http://dx.doi.org/10.1038/nature12037>
30. Song, H.j., Poo, M.m.: The cell biology of neuronal navigation. *Nat. Cell Biol.* **3**(3), E81–E88 (2001). <http://dx.doi.org/10.1038/35060164>
31. Sutherland, D.J., Pujic, Z., Goodhill, G.J.: Calcium signaling in axon guidance. *Trends Neurosci.* **37**(8), 424–432 (2014). <http://dx.doi.org/10.1016/j.tins.2014.05.008>
32. Tessier-Lavigne, M., Goodman, C.S.: The molecular biology of axon guidance. *Science* **274**(5290), 1123–1133 (1996). <https://doi.org/10.1126/science.274.5290.1123>. <http://science.sciencemag.org/content/274/5290/1123>
33. de la Torre, J.R., Höpker, V.H., li Ming, G., ming Poo, M., Tessier-Lavigne, M., Hemmati-Brivanlou, A., Holt, C.E.: Turning of retinal growth cones in a netrin-1 gradient mediated by the netrin receptor {DCC}. *Neuron* **19**(6), 1211–1224 (1997). [http://doi.org/10.1016/S0896-6273\(00\)80413-4](http://doi.org/10.1016/S0896-6273(00)80413-4). <http://www.sciencedirect.com/science/article/pii/S0896627300804134>
34. Turing, A.M.: The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **237**(641), 37–72 (1952). <https://doi.org/10.1098/rstb.1952.0012>. <http://rstb.royalsocietypublishing.org/content/237/641/37>
35. Yamamoto, N., Tamada, A., Murakami, F.: Wiring of the brain by a range of guidance cues. *Prog. Neurobiol.* **68**(6), 393–407 (2002). [http://doi.org/10.1016/S0301-0082\(02\)00129-6](http://doi.org/10.1016/S0301-0082(02)00129-6). <http://www.sciencedirect.com/science/article/pii/S0301008202001296>
36. Zheng, J., Wan, J., Poo, M.: Essential role of filopodia in chemotropic turning of nerve growth cone induced by a glutamate gradient. *J. Neurosci.* **16**(3), 1140–1149 (1996). <http://www.jneurosci.org/content/16/3/1140>
37. Zheng, J.Q.: Turning of nerve growth cones induced by localized increases in intracellular calcium ions. *Nature* **403**(6765), 89–93 (2000). <http://dx.doi.org/10.1038/47501>
38. Zubler, F., Douglas, R.: A framework for modeling the growth and development of neurons and networks. *Front. Comput. Neurosci.* **3**, 25 (2009). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2784082/>

Mathematical Modelling of Cerebellar Granular Layer Neurons and Network Activity: Information Estimation, Population Behaviour and Robotic Abstractions



Shyam Diwakar, Chaitanya Nutakki, Sandeep Bodda, Arathi Rajendran, Asha Vijayan, and Bipin Nair

Abstract Recent studies show cerebellum having a crucial role in motor coordination and cognition, and it has been observed that in patients with movement disorders and other neurological conditions cerebellar circuits are known to be affected. Simulations allow insight on how cerebellar granular layer processes spike information and to understand afferent information divergence in the cerebellar cortex. With excitation-inhibition ratios adapted from *in vitro* experimental data in the cerebellum granular layer, the model allows reconstructing spatial recoding of sensory and tactile patterns in cerebellum. Granular layer population activity reconstruction was performed with biophysical modeling of fMRI BOLD signals and evoked local field potentials from single neuron and network models implemented in NEURON environment. In this chapter, evoked local field potentials have been reconstructed using biophysical and neuronal mass models interpreting averaged activity and constraining population behavior as observed in experiments. Using neuronal activity and correlating blood flow using the balloon and modified Windkessel model, generated cerebellar granular layer BOLD response. With the focus of relating neural activity to clinical correlations such models help constraining network models and predicting activity-dependent emergent behavior and manifestations. To reverse engineering brain function, cerebellar circuit functions were abstracted into a spiking network based trajectory control model for robotic articulation.

1 Introduction

Neurons are regarded as fundamental processing units in the central nervous system (CNS), interacting at the synaptic level and forming neuronal microcircuits in the brain. Cognition and behavior has been attributed to neural and population

S. Diwakar (✉) • C. Nutakki • S. Bodda • A. Rajendran • A. Vijayan • B. Nair
Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Amritapuri Campus, Clappana PO, Kollam, Kerala 690525, India
e-mail: shyam@amrita.edu

responses [1]. Neurons communicate through spikes and the information conveyed by this spike trains can be quantified by considering the exact timing of action potentials. The responses of single cells to external stimuli are often averaged over many trials to minimize the effect of neuronal variability. Cerebellum is a crucial structure of brain known to be affected in several cognitive and motor coordination related disorders [2]. Large scale networks of the cerebellum have been conceived as internal models for motor control and process information attributed to stimuli processing and synaptic learning [3]. The cerebellar cortex, deep cerebellar nuclei and the inferior olive are the three centers of activity within the cerebellar circuit and contain different types of neurons. Understanding information from single cerebellar granule neuron perspective provides a mesoscopic view of the information encoded in the system.

Biophysically, neurons are considered as phase-change devices that map input spike trains to output response where transformation are representative of processing at the postsynaptic targets. The information from presynaptic neuron reaches the post synaptic neuron initiating a cascade of downstream process. Information processing at the single neuron level involves the transformation of input spike trains into an appropriate output spike train. Quantification of information [Mutual Information (MI)] in cerebellar granule neuron helps to understand information coding and processing capacity of cerebellum in both normal and diseased conditions. Information encoding mechanism at the cerebellar granular layer rely on estimates of synaptic inputs, to understand the excitability of neurons affected by these stimuli, and in identifying patterns that modify synaptic inputs for spike outputs that result in motor coordination and learning [4]. Granule cells are defined by a limited number of synaptic inputs (max. four excitatory and four inhibitory), this helps to understand the input space of granular layer broadly. In cerebellum, granule cell (GrC)-mossy fiber (MF) connections carry more temporal information than climbing fiber (CF) system [5–8]. Information transmission in cerebellum via MF-GrC relay [9] has also been considered as an important site for synaptic plasticity. MFs receive information from corticopontine, spinocerebellar and vestibulocerebellar tracts. The multimodal nature of sensory and tactile information flow to the cerebellum has raised crucial questions on the understanding of neural circuit functions in mammals [10], especially in areas of cerebellum including crus I, crus II and paraflocculus in rats [11].

Cerebellum granular layer activity have also been studied using different recording and imaging techniques. In this chapter, we used mathematical modeling to understand population activity from: (1) averaged single neuron responses as seen during fMRI and (2) network activity as perceived during local field potentials (LFP). Modeling functional magnetic resonance imaging (fMRI) blood-oxygen level dependent (BOLD) signals allow reconstructing local neural activity and the following changes in cerebral blood flow due to neurovascular coupling and hemodynamic relationships. Regional blood flow is controlled by variations in concentrations of different ionic and molecular by-products like K^+ , nitric oxide (NO), carbon dioxide (CO_2). These components may directly alter the blood flow by depolarization or hyperpolarization of neural membranes triggering vascular

dilation and constriction [12]. Although the neural activity and the blood flow are very closely related, it is important to understand how the neural activity drives the vascular response. Functional Magnetic Resonance Imaging (fMRI) is used to understand the neural activity of the brain by measuring the Blood oxygenation Level Dependent (BOLD) signal from active regions of the brain. fMRI measures the hemodynamic changes in brain to estimate neural activity with respect to a specific task. Hemodynamic response function (HRF) depends on the blood flow, cerebral oxygenation, blood oxygenation and blood volume. Functional brain maps are perceived as static representations of averaged dynamic activity over a longer period of time. Recent studies [13], have shown different parts of the brain connect, coordinate and interact while performing cognitive functions. Since fMRI BOLD studies relate more to behavior and cognition, modeling fMRI BOLD activity allows connecting synaptic and neural activity to behavioral responses as observed during imaging.

Estimation of neuronal ensemble activity can also be done by recording extracellular potentials called Local Field Potential (LFP), as seen experimentally from electrophysiological recordings. Mathematical estimations can be done using multi neuron or mass models. LFP is a neuronal population signal obtained by low pass-filtering (100–300 Hz) of the extracellular electrical potential recorded with intracranial electrodes. Amplitude and frequency of LFP depend upon the proportional contribution of multiple sources and various properties of the brain tissue. LFP is considered to be a measure of neural activity widely used in sensory processing, cognitive process and motor planning [14–21]. Studies have shown that signal like LFP and EEG of cortical neurons can be generated using models like neural mass model [22–24]. Neural mass model also referred to as lumped approach have been used to generate EEG signals [25]. Neural mass model have shown to produces various rhythms [23] in signals like EEG and MEG that mainly depends on kinetics of the populations [26]. We used mass models to reconstruct LFP signals of granular layer. LFP of granular layer includes ionic currents in the granule neurons and the in vivo response observed in anaesthetized rats include trigeminal (T) and cortical waves (C) [27].

Considering the different neuronal estimations, one could use this information in modelling the different functions of the brain and also to decipher its role in articulation control. In brain, the cerebellum plays a major role in timing of motor activities. With its modular architecture and movement-related functional roles, single neuron and population neural activity in the cerebellum have inspired control algorithms for several kinematic tasks. ‘Cerebellum’ has been referred to as a “non-stiff joint which works on low power actuators” [28] allowing cerebellum-based strategies for robotic articulators. Cerebellar microcircuits allow modeling forward and inverse kinematics for fast and precise movement and have enabled trajectory planning [29]. Perceptron model [30] resembling the cerebellar Purkinje neuron [31] has been a basis of pattern separation roles. These concepts along with the timing in neuronal spikes have proposed to encode fine tuning and learning of movement tasks based patterns [28, 32]. Reconstructing the cerebellar microcircuitry with pattern

classifying capability on robotic platform can help validate the neuronal estimation through reverse engineering of the cerebellum.

Cerebellum includes afferent and efferent connections related to motor and sensory areas of the brain including auditory, tactile and visual receptors and transmitting back to cerebral cortex. In this chapter, we focus on the quantification of mutual information (MI) in cerebellar granule neurons using two coding schemes; temporal and rate coding with respect to auditory and visual stimuli and the effect of plasticity was analyzed by changing the intrinsic excitability and release probability to understand how plasticity mechanisms affect the average amount of information transmitted through the neuron. The study helped to explain how the information transmitted by a granule cell is regulated by long term synaptic plasticity occurring at the MF-GrC relay during auditory and visual stimuli. We also modeled the population level activity of the cerebellar granular layer with a mathematical description for measuring the BOLD response and by computational reconstructions of evoked postsynaptic local field potentials using a neural mass model. Abstractions of the cerebellar granular layer activity have been demonstrated by reconstructing a spiking neural network mimicking the cerebellar articulation controller [33] for robotic applications.

2 Methods

2.1 Single Neuron Modeling

The methods involving biophysical modeling have been described in previous studies [3, 34, 35]. Single neuron dynamics was modeled using Adaptive Exponential Integrate and fire model (AdEx), a spiking neuron model with two variables [36]. Equations (1) and (2) were able to generate different patterns of firing and were used to simulate the neuronal dynamics of single neurons.

$$\frac{CdV}{dt} = -gl * (V - El) + gl * delT * \exp\left(v - \frac{vT}{\Delta T}\right) + I - W \quad (1)$$

$$\tau w \frac{dw}{dt} = a(V - EL) \quad (2)$$

when $V > 0$ mv, $V = V_r$ and $w = w + b$

AdEx model predicts the timing of 96% of the spikes (+2 ms) and in the model, V represents the membrane voltage, C is the membrane capacitance, E_L is the resting potential, g_l is the leakage conductance, $delT$ is the slope factor, V_T is the threshold potential, V_r is the resting potential, I is the synaptic current, τw is the time constant, a is the level of sub threshold adaptation, and b represents spike triggered adaptation.

Detailed computational biophysical granule neuron model with 52 compartments [37] (Fig. 1a) was used to quantify information in granule neurons for in vivo burst

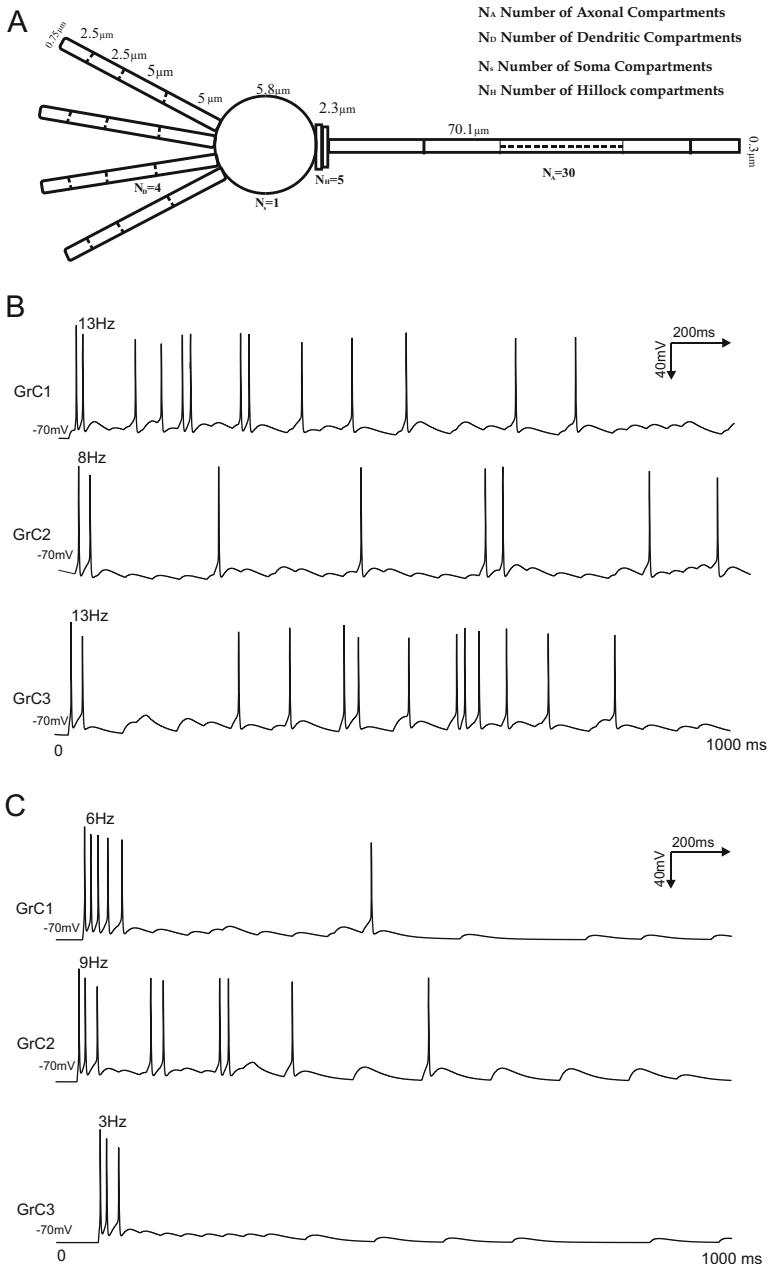


Fig. 1 Mathematical reconstruction of cerebellar granule neuron response: (a) Cartoon representation of granule neuron model with 1 soma compartment, 4 dendritic compartments, 30 axonal compartments and 5 hillock compartments as in [37]. GrC membrane potential were maintained at -70 mV , which was activated by MF stimulation at different times. Fifteen granule neurons with different auditory and visual stimuli were simulated and 3 GrC responses with respect to auditory (b) and visual stimuli (c) with firing frequency have been shown

stimuli from mossy fibers (MF). Applying different inputs including auditory and visual stimuli, we have mathematically reconstructed firing patterns of both auditory and visual responses of 30 granule neurons and decoded corresponding mossy fiber activities from these modalities [38].

2.2 Cerebellar Granular Layer Information Processing

Information reliability of single neurons was estimated by Shannon's mutual information [39] to highlight the significance of adaptation to new inputs and role of long term memory. Since information theory determined statistical significance within neural variability, information theoretic methods allowed exploring the distribution of MF-GrC relay. According to Claude E. Shannon, Mutual information and entropy are two important attributes in information theory. Entropy is the amount of uncertainty one observer has about the state of the system which is used to study the channel capacity required to reliably transmit the information. It can be calculated by

$$H(X) = -\sum_{i=1}^N P_i * \log_2(P_i) \quad (3)$$

As in Eq. (4), Mutual Information can be calculated from the neuronal response r for an input s .

$$MI(R, S) = \sum \sum p(s) p(r|s) \log_2 p(r|s) / p(r) \quad (4)$$

R and S are input and output spaces representing responses and stimuli, $p(r)$ and $p(s)$ are priori probability of both stimulus and response and $p(r|s)$ is conditional probability, which relating to the response r with the stimuli s [39]. For estimating MI, auditory and visual stimuli were applied through four mossy fiber synapses and simulated for 1000 ms and repeated the same with 100 different stimuli by changing the release probability and intrinsic excitability. MI was estimated for continuous temporal dataset by applying binning method [40], which completely discrete data into different clusters based on different bin points. Here, entropy, conditional probability and mutual information were estimated for five different bin sizes of width varying from 1 to 5 ms and average MI were taken. MI was estimated from the samples using two different attributes of the spike trains: the firing rate and spike timing. In rate coding, it was assumed that neuron reacts to the number of input spike it received while in temporal coding, it was assumed that the time of occurrences of the spikes convey information.

2.3 Model Based Methods for Hemodynamic Response

Several models have been proposed to model BOLD signal from neural activity based on the changes in blood flow, blood volume, and changes in oxygenation (Fig. 2). Balloon model developed by Buxton and colleagues [41] and delayed compliance model of Mandeville [42] are commonly used biomechanical models to predict the flow volume and oxygenation during brain activation (Fig. 2). These models predict the initial dip, overshoot and post-stimulus of the BOLD signal in terms of Cerebral Blood Volume (CBV), where recovery to the baseline is much slower than the changes with Cerebral Blood Flow (CBF) in modified windkessel model.

Modeled characterization of hemodynamic response:

- During brain activation Cerebral Blood flow (CBF) increased much more than CMRO₂, generating a reduction of E(t) (oxygen extraction) and the total deoxyhemoglobin content present in the voxel caused a BOLD signal change [43, 44].
- The flow and BOLD response to a given brief stimulus was delayed 1–2 s with a temporal width of 4–6 s. For a continuous stimulus of 30s or longer, the response reached a plateau value, although there could be substantial variation [45].
- A post stimulus undershoot of BOLD signal can be frequently observed and may last for 30 s or more, longer stimuli tending to have longer post-undershoots [46, 47], where CBF response is a shorter and weaker post-stimulus undershoot [41, 48].

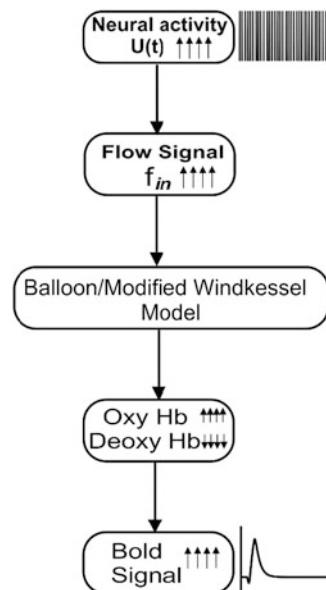


Fig. 2 Modeling of HRF and BOLD response from neural activity

- The initial dip of the BOLD signal lasted 1–2 s before the standard BOLD signal increase [49–51], and a corresponding transient increase of deoxyhemoglobin has been reported in optical imaging studies [52, 53].
- BOLD response shows a temporal nonlinearity and is mostly defined for a brief stimulus less than about 4 s and extended stimulus longer than 6 s. Comparing both short and long duration stimuli more than 4 s, temporal nonlinearity was reduced.
- Nonlinearity has been reported as a refractory period, when two identical stimuli presented close together in time produced a net response with less than twice the integrated response of a single stimulus [54–56]
- Studies shown that the baseline CBF effect to the particular stimulus can have a strong effect on the magnitude of the BOLD response to the same stimulus [44, 57–60].

2.3.1 Balloon Model Based Prediction

BOLD signal $y(t)$ was modeled as a nonlinear function of normalized volume (v), deoxyhemoglobin (q) and resting net oxygen extraction fraction (E_o).

$$y(t) = \lambda(v, q, E_o) = V_0 \left[k_1 (1 - q) + k_2 \left(1 - \frac{q}{v} \right) + k_3 (1 - v) \right] \quad (5)$$

$$k_1 = 7E_o$$

$$k_2 = 2$$

$$k_3 = 2E_o - 0.2$$

Where V_0 was resting blood volume fraction. The value of V_0 was taken from earlier studies [61]. BOLD signal is the sum of both intra and extra vascular signal. k_1 , k_2 and k_3 are constants [41]. In Eq. (5), the parameter k_1 contributed to intra vascular signal, k_2 contributed to extra vascular signal and k_3 contributed to balance the sum of signals.

The rate of change of volume is related as the difference of inflow and outflow, can be expressed as

$$T_v V = fin - fout(v) \quad (6)$$

$$fout(v) = v^{1/\alpha} \quad (7)$$

Table 1 BOLD physiological parameters and values

Parameter	Value	References
V_0	0.02	Friston et al. (2000)
k1, k2 and k3	$7E_0.2, 2E_0.0.2$	Buxton et al. (1998)
f_{in}, f_{out}	1(Rest)	
q	1(Rest)	
CBV	0.80	Grub et al. (1997)
CBF	0.38	Grub et al. (1997)
$1/\alpha = \alpha + \beta$	3.03	Friston et al. (2000)
T_v	1	Friston et al. (2000)
E_0	0.08	Friston et al. (2000)
T_c	7.5	Kong et al. (2004)
β	0.7	Kong et al. (2004)

Where f_{in} ($f_{in} = 1$ at rest) and f_{out} ($f_{out} = 1$ at rest) were blood inflow and outflow [62] established the relationship between CBV and CBF. It was observed that $1/\alpha = \alpha + \beta = 3.03$ [61]. The parameter T_v represents meant transit time and q the changes in deoxyhemoglobin content estimated (Eq. 8). The dynamics of the BOLD response changes as per the outflow function.

$$T_v q = f_{in} \frac{E(f_{in}, E_0)}{E_0} - f_{out}(v) \frac{q}{v} \quad (8)$$

Where E_o represented net oxygen extraction fraction and E_0 was taken as 0.08 [61]. The parameter $E(f_{in}, E_0)$ can be calculated as (Eq. 9)

$$(f_{in}, E_O) = 1 - (1 - E_o) \frac{1}{f_{in}} \quad (9)$$

2.3.2 Modified Windkessel Model Based Prediction

Modified Windkessel model was proposed by [63] was extended to model delayed compliance by incorporating a normalized state variable ‘c’ into the existed windkesel model (Table 1).

$$T_c c = V^\beta - c \quad (10)$$

Where T_c represented compliance time constant. As per the model, the rate of change of volume can be expressed using (Eq. 11)

$$V = f_{in} - \frac{v^{\alpha+\beta}}{A(t)} \quad (11)$$

Using ‘c’ in Eq. (11) led to

$$T_v V = f_{in} - \frac{v^\gamma}{c} \quad (12)$$

Where, $\gamma = \alpha + \beta$ [61]. Here, α is laminar flow and β is a constant representing the demised volume reserve at high blood pressures. The value of γ was set to 5.5 [63].

2.4 Evoked Local Field Potentials and Neural Mass Model

To estimate the role of spiking behavior in evoked granule neuron local fields, LFP was estimated with several methods including neural mass models as well as using biophysical models or spiking neural models. Neural mass models describes the mean activity of the population of neurons signified by average firing rates and membrane potentials. The neural mass model of the cerebellum (Fig. 3) featured the population activity average contributed by granule neurons receiving excitatory input from the mossy fibers. Pulse stimuli reproducing brain slice in vitro and anaesthetized in vivo conditions were simulated [27, 64].

2.4.1 Cerebellum Granular Layer Neural Mass Model with Mossy Fibers Input Patterns

Stimuli were provided to the cerebellum granular layer mass model using pulse generators. Pulses were convolved with Gaussian white noise representing spike variability and post-synaptic latency. The addition of noise influenced information processing, spike time reliability, stochastic resonance, firing irregularity and action potential initiation and conductance in models [65]. The convolved input was then transformed using transformation function which converts the pulses to action potentials which enables the neuron to fire. In this model, we only considered excitatory stimuli. Postsynaptic Potential (PSP) transformation was modeled as a linear transformation using a second order differential equation characteristic of

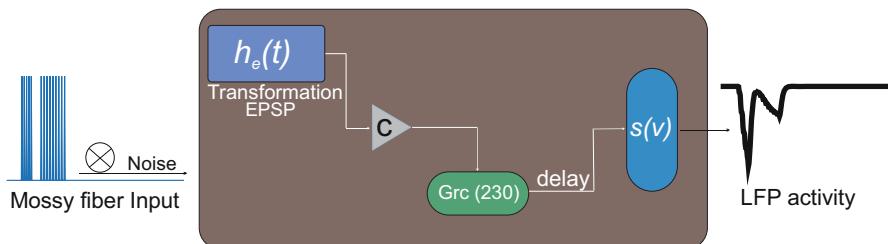


Fig. 3 Neural mass model for granular layer LFP reconstruction

impulse responses. Excitatory postsynaptic potential transformation function was calculated (Eq. 13) which was proposed by [66]

$$h_e(t) = \begin{cases} 0, & t < 0 \\ Aae^{-at}, & t \geq 0 \end{cases} \quad (13)$$

Where ‘A’ determined the maximum amplitude of excitatory post synaptic potential (EPSP) and ‘a’ was the lumped representation representing the sum of reciprocal with respect to the time constant of passive membrane. Each PSP block introduces two differential equations of the form (Eq. 14) [67]

$$\ddot{y}(t) = Aax(t) - 2ay(t) - a^2y(t) \quad (14)$$

Where $y(t)$ was the input to the transform function and $x(t)$ the output of the transformation. The connectivity constant representing the strength of the synaptic contacts from one population to another (mossy fiber to granular layer) which modeled the average number of active synapses within the population contributing to LFP. The potential-to-rate function was a non-linear function (Sigmoid) and converted the average membrane potential V [68] of a population to an average pulse density, the output of the neural mass model. This transformation is described by a non-linear function which is given and simplified to Eq. (15) [26, 67, 69]

$$s(v) = \frac{2e_0}{1 + e^{s(v_0-v)}} \quad (15)$$

Where e_0 is the maximum firing rate of a neuronal population v_0 is the PSP for which 50% firing rate is achieved, s is the sigmoid steepness parameter. Implementation values for each parameter while using transformation and sigmoid function are listed in Table 2.

2.4.2 Reconstruction of Local Field Potential from Spiking Models

VERTEX is a MATLAB based tool designed to aid the simulation of extracellular potentials generated by activity in SNN’s with input as position of neurons from virtual electrode [70]. Extracellular potentials were estimated by summing up the

Table 2 Mass model values for parameters used to estimate synaptic potential

Parameter	Description	Standard value
A	Average excitatory synaptic gain	3.25 mv
a	Reciprocal of excitatory time constant	100 Hz
v_0	Potential at half of maximum firing rate (Sigmoid function)	6 mv
e_0	Half of maximum firing rate (Sigmoid function)	2.5 Hz
S	Steepness parameter (Sigmoid function)	0.56 mV^{-1}

membrane currents of each compartment, weighted by distance from electrode tips as represented in data-based models. Line Source Approximation (LSA) method [71–74] was used to calculate the contribution of dendrites and soma compartments at a given spatial point and time. This method has been previously used in studies to reconstruct or to calculate Local Field potential [74–78].

3 Spiking Neural Network Based on Cerebellum for Kinematics

Spiking models were used to reconstruct a cerebellum-like spiking architecture that could implicitly model forward and inverse kinematics using the concept of internal models [79]. For reproducing the spiking dynamics of the neurons, Adaptive Exponential Integrate and Fire neuron model was used in this study [80]. Spiking neural network developed had three components: an encoder, network and a decoder. Gaussian kernel based encoder was developed to encode the different kinematic parameter to spiking information where the kinematic parameter was taken as the mean with a movement standard deviation of 3 cm as reported in [33, 80]. A scalable spiking neural network model (Fig. 4) consisting of 48 neurons [21 mossy fibre (Mf), 21 granule cells (GC) and 6 Purkinje neurons (PC)] were used to represent inverse kinematics and was applied on a sensor-free 6 DOF low-cost robotic arm, developed at the lab [33]. Adaptation of the network was done using

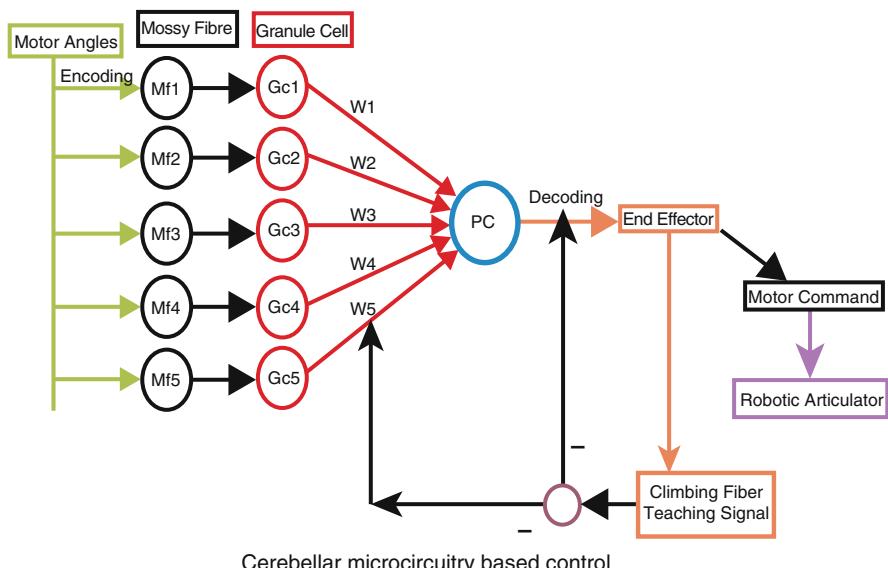


Fig. 4 Cerebellum-like spiking neural network for robotic kinematic control

a Hebbian like learning rule and synaptic weights were adapted at Mossy Fibre-Granule cell [81], and Parallel Fibre-Purkinje Cell [82].

The structured network was also able to separate input patterns and classify movement based on kinematic parameters [34]. Change in weights was calculated using a Hebbian like learning rule and the weights were updated when the error value decreased (Eq. 16).

$$\Delta w_t = \begin{cases} \eta e_t x_t, & \text{if } e_t < e_{t-1} \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

$$w_t = w_t + \Delta w_t$$

4 Results

4.1 Estimation of MI at MF-GrC Relay

Granule neuron responses from crus I and II and paraflocculus of the cerebellar cortex in rats were reconstructed for both auditory and visual stimuli (see Fig. 5). Each of the granule neuron patterns matched with previous experimental studies [10]. This was achieved by changing inter spike interval (ISI) and inter burst interval (IBI) of mossy fiber input. According to experimental data, the auditory stimuli was delivered with a calibrated speaker and the visual stimuli was presented using two white LEDs under sound-attenuating, light-proof box. High frequency firing was observed during auditory stimuli (20 Hz) when compared to visual (9 Hz) (Fig. 5). Simulations were carried out for duration of 1000 ms. It has been observed that the inter spike interval was lower in visual responses than auditory.

Communication of information from the presynaptic to postsynaptic neurons happens though the generation of action potential (spikes). As expected, models

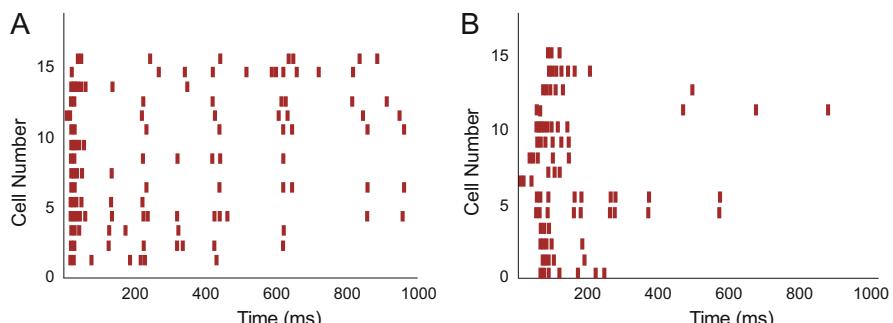


Fig. 5 Response spike trains for auditory and visual stimuli. **(a)** Granule neuron spike train for auditory stimuli and for visual stimuli **(b)**. Firing patterns of 15 granule neurons are shown

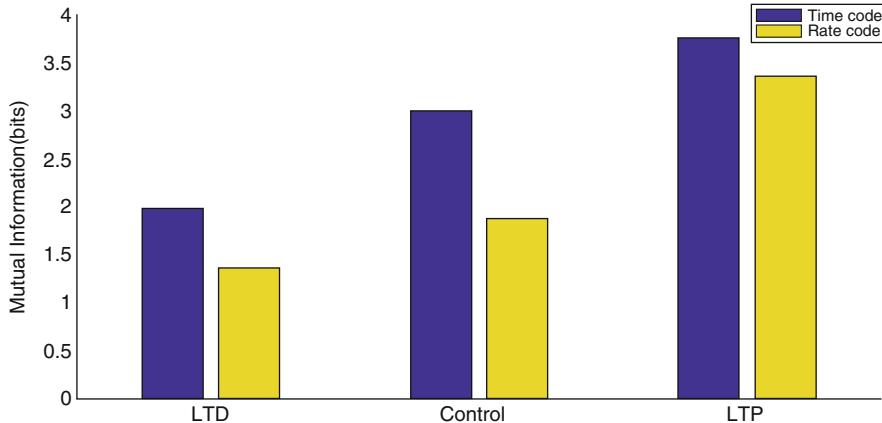


Fig. 6 Mutual information for auditory stimuli was estimated during LTD, Control and LTP. MI was estimated using two different coding schemes; using time and rate code. As expected, mutual information was higher during LTP than control and lower during LTD than control. MI estimates were taken for five different bin sizes from 1 to 5 ms for time code and the average MI is indicated

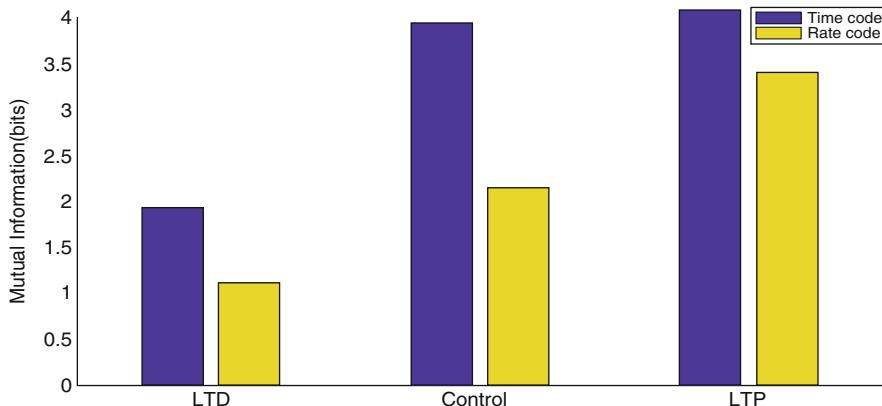


Fig. 7 Mutual information for visual stimuli was estimated during LTD, Control and LTP. MI was estimated using two different coding schemes; using time and using rate code. Like with auditory stimuli, mutual information was higher during LTP than control and lower during LTD than control. MI estimates were taken for five different bin sizes from 1 to 5 ms for time code and the average MI is indicated

showed a significant reduction in spike number during Long Term-Depression (LTD) and increment during Long Term-Potentiation (LTP). So the information conveyed by the granule neurons is high during LTP and low during LTD conditions (Figs. 6 and 7). MI was estimated by both temporal and rate coding methods. The plasticity experiments reveal how the release probability and intrinsic excitability changes, modulates firing behavior of granule neurons.

Models indicate synchronous firing activity in granule neuron response for both auditory and visual stimuli during both control and LTP conditions. Synchronization of neuronal activity plays an important role in memory formation and cellular mechanism of memory storage [82].

4.2 Variations in BOLD Response Measured Using Balloon Model and Modified Windkessel Model (MFWM)

BOLD response was modeled using Balloon model (Fig. 8b) and Modified Windkessel model (Fig. 8d) using spiking patterns from multi compartment granule cell neural activity for 2 s of simulation time. With increase in neural activity, the blood inflow and blood oxygenation level increased and eventually the volume of blood changes, leading to reduction in total deoxyhemoglobin. Transient features of the BOLD signal results from the difference in time courses of flow and volume

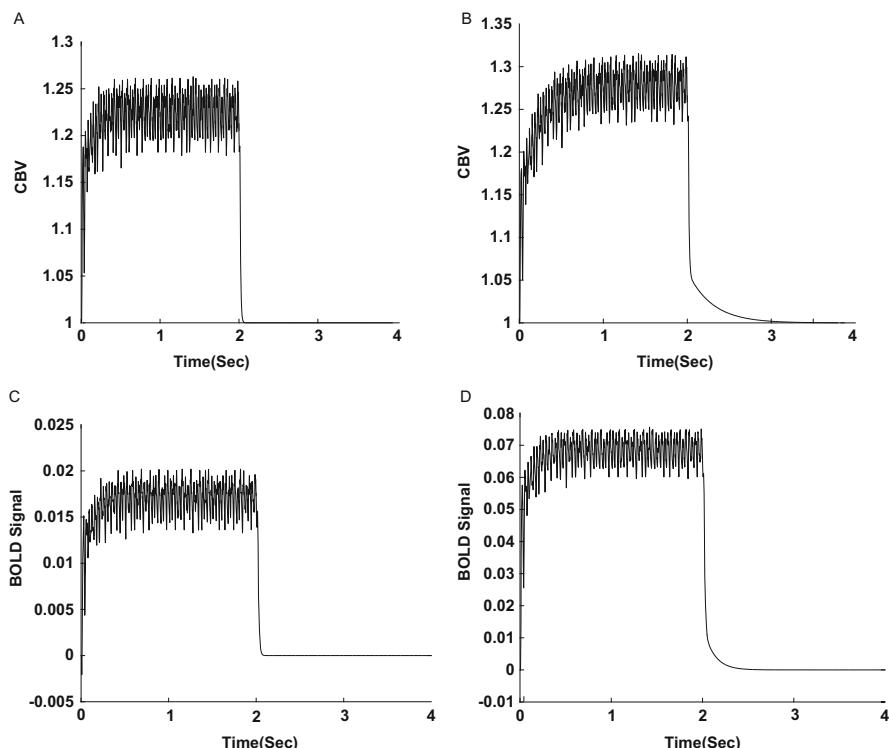


Fig. 8 Cerebral blood volume (CBV) and BOLD changes predicted using balloon and modified Windkessel model. (a) Predicted CBV using balloon model. (b) Predicted CBV using modified Windkessel model. (c) Predicted BOLD response using balloon model. (d) Predicted BOLD response using modified Windkessel model

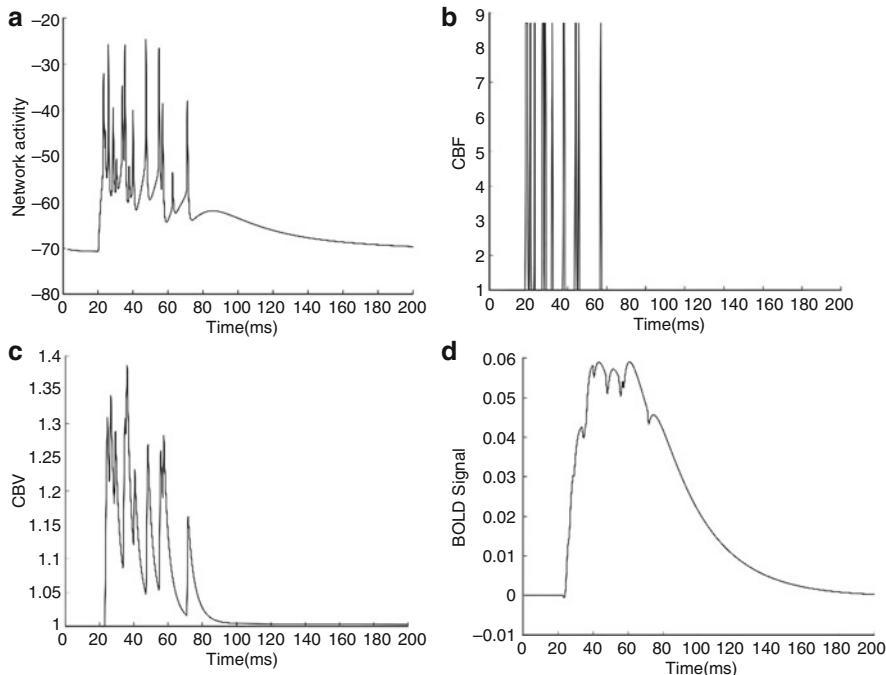


Fig. 9 Linking neural activity to the resulting physiological responses and the measured BOLD response. (a) Neural activity of 720 granule cells. (b) CBF with respect to the neural activity. (c) Predicted Cerebral Blood Volume (CBV). (d) Predicted BOLD response

changes. Predictions were compared under the same input condition and also looked how delayed compliance contribute to predict BOLD (Fig. 8b, d) and CBV (Fig. 8a, c) when blood flow was provided as input for the models.

MFWM is a combination of Balloon and Windkessel model. Incorporating compliance into the model makes more relevant for predicting flow volume relationship (Fig. 9). BOLD and CBV prediction analysis was done by varying the model parameters at different conditions using 720 granule cell mean population activity with a simulation time of 200 ms. Comparatively Modified Windkessel model shows the good predictions compared to Balloon model in returning to the baseline for all simulations (Fig. 9c, d). The model with added delayed compliance has captured the dynamic relationship between CBF-CBV returning to the baseline.

4.3 Simulating Extracellular Potentials Recordings in Neural Mass Model (NMM) and Spiking Neural Network (SNN)

Neural mass models allowed to reconstruct Local Field Potentials in-vitro generating N_{2a}, N_{2b} waveforms and in vivo behavior was observed as Trigeminal and Cortical waveforms for sensory inputs. The model allowed to correlate circuit level

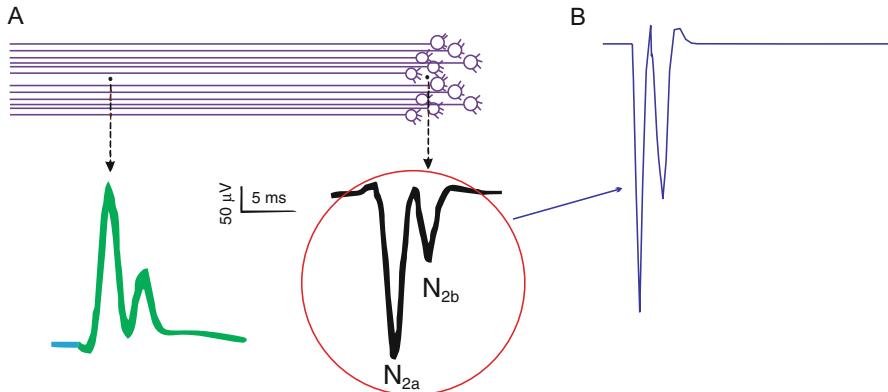


Fig. 10 Modeling LFP in vitro. (a) LFP simulated for single spike inputs as in brain slices using biophysical model in response to stimuli. (b) The Generation N_{2a} and N_{2b} waveform from Neural Mass model composed of 230 Granule Cell Network

activity to the underlying neural circuitry and helped to reconstruct N_{2a}, N_{2b} (Fig. 10b) and T, C waves as seen in experimental data (Fig. 10a) [83, 84].

Evoked LFP activity of granular layer was computed using detailed biophysical model, spiking neuron network model used in tools or algorithms including LFPsim [78] (Fig. 11a), VERTEX, ReConv. We used in vitro like inputs [74] to a cerebellum granular layer model to reconstruct LFPs. LFPsim simulated MEA activity. Virtual Electrode Recording Tool for Extracellular potentials (VERTEX) was used to model the field potential as perceived by multi-electrode arrays (MEA) with SNN (Fig. 11b, e), estimating population field activity across the circuit. Figure 11c shows the T and C wave reconstruction using the ReConv [74] algorithm.

4.4 Optimized Kinematic Control Using SNN

Reconstructed cerebellum-like SNN functions as a control algorithm for robotic articulation tasks. SNN could reproduce the firing behavior of granule neuron (see Fig. 12 and [33]) and Purkinje cell which could clearly bring out abstractions from sensorimotor control.

In the spiking neural network, 21 mossy fiber behaves as an input pulse generator while the granular layer consisting of 279 granule neuron models were modeled as a pattern recoder which could tune patterns from different stimuli. Purkinje neuron on the other hand generated the output joint angles of the trajectory. Optimized joint angles were used during training for the trajectory planning and was trajectory estimated by the SNN was the predicted output. It was observed that accuracy increased over the iterations suggesting adaptation and trained motor task (see Fig. 13). Generally, spiking neural network was made to stop training epochs when the prediction error decreased below 5%.

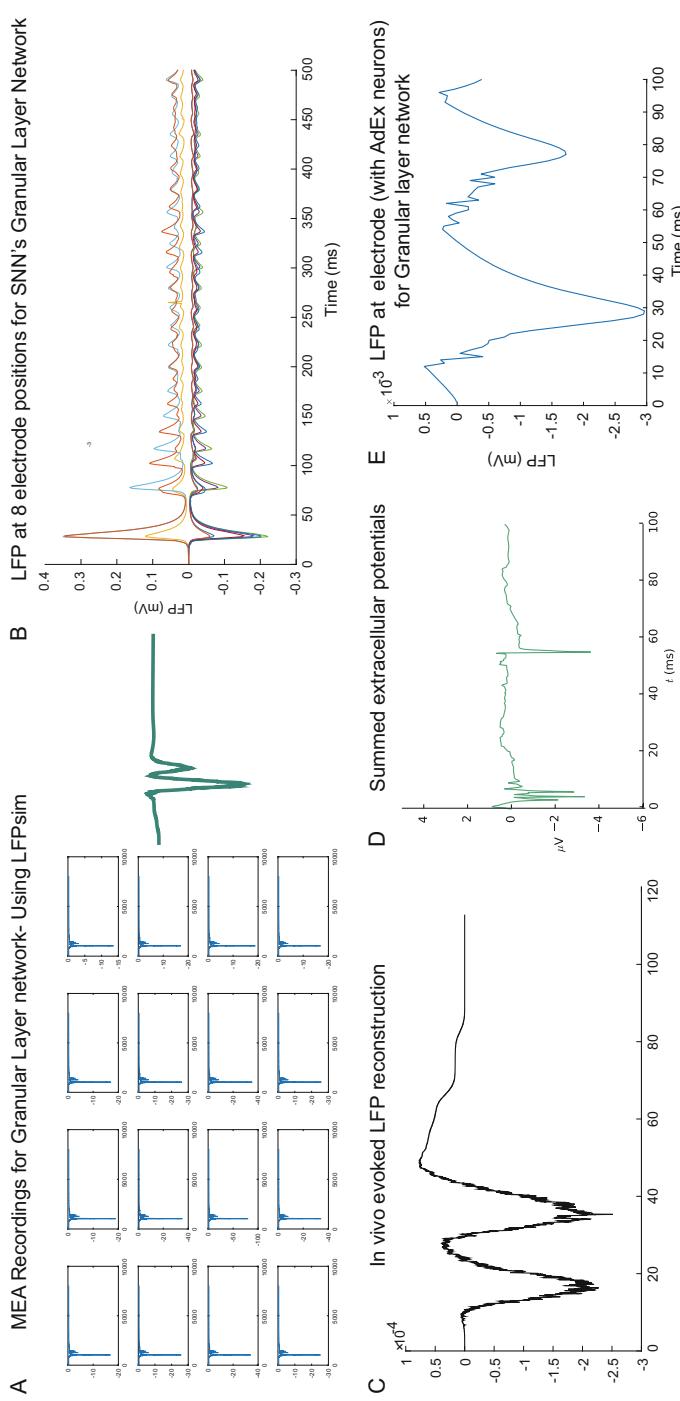


Fig. 11 LFP recordings from network activity **(a)** LPsim reconstruction of MEA activity of a 720-granule cell network **(b)** granule cell cluster activity using ReConv algorithm **(c)** network of 230 neuron with spiking dynamics of granular layer using VERTEX **(d)** summed up extracellular activity of granule cell **(e)** SNN modeling of granular layer network LFP showing T and C waveforms

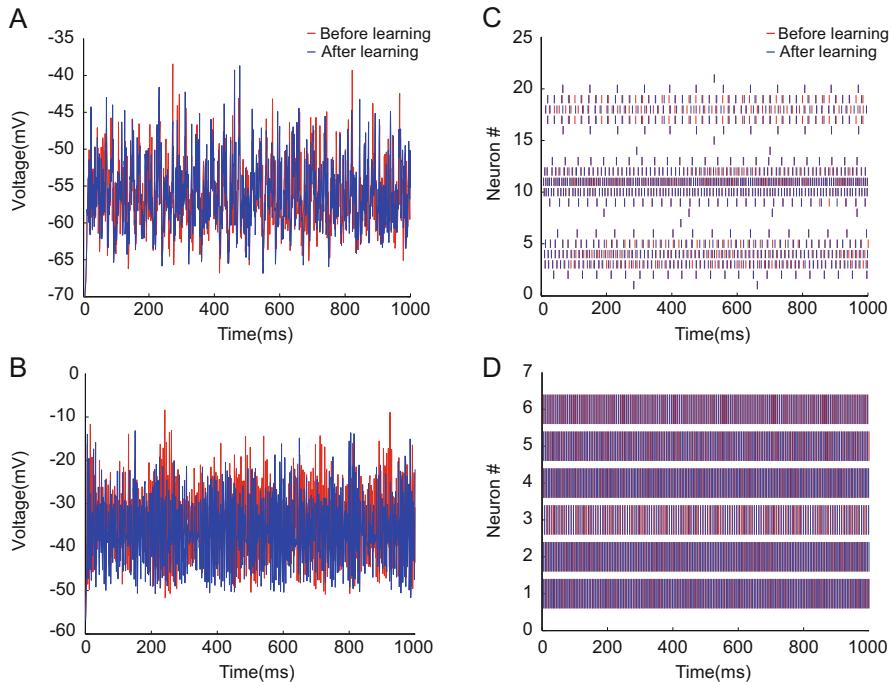


Fig. 12 Cerebellum-like spiking network for supervised learning. Mean firing rates of granule neurons (**a**) and Purkinje neurons (**b**) in the network. Raster plot of the granule cells (**c**) and Purkinje cells (**d**). Red lines represent the firing behavior of the neurons in the first iteration and blue lines represent the firing behavior of the neurons after network adapts to corrected weight in the final epoch

5 Discussion

In this chapter, cerebellar granular layer coding behavior [7, 8] has been mathematically modeled to demonstrate diverse yet significant roles in output information, shaping population responses and to abstract the granular layer properties into dynamics models for robotic platform which could aid in kinematic control. In this study, transmission reliability of single granule neurons for auditory and visual stimuli was estimated using mathematical modeling and mutual information theory. Spike time-coupling to LFP and BOLD allow estimating how envelopes of firing rate activity, BOLD behavior and evoked LFP changes connect response behavior and stimuli properties may be correlated at various levels. Through this multi-perspective study, we point to perspectives to correlate spike firing, fMRI BOLD and LFP which has been relevant to understanding brain function and fMRI research. By connecting spiking mutual information change to firing dynamics, spiking and burst patterns contribution of single neurons in activity can be interlinked to emergent behavior observed from BOLD or LFP reconstructions.

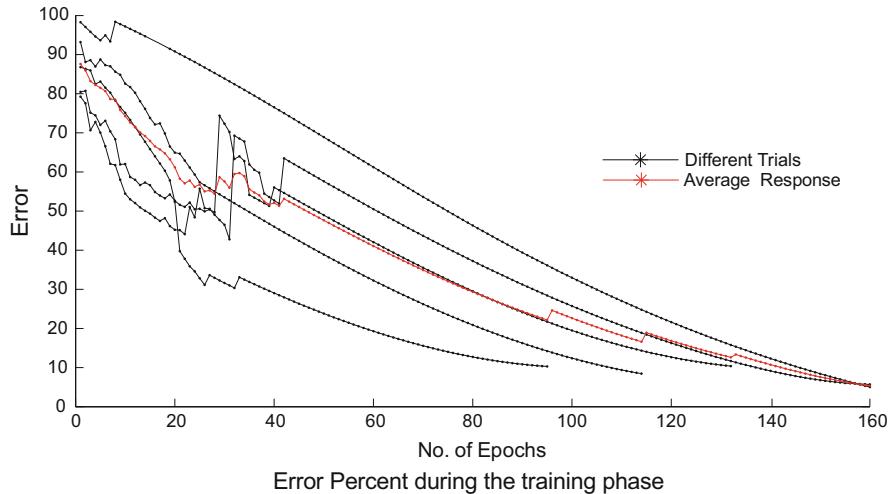


Fig. 13 Using SNN for robotic pattern abstractions. Mapping errors to training epochs, reliability in error percentage was estimated

Mutual information was estimated during both auditory and visual stimuli enabling a quantitative analysis of modeled coding strategies for different input patterns and synaptic parameters. Likewise, different models such as BOLD signal models and local field potential modeling helps to reconstruct the population activity of a set of neurons. Activation in granule neuron clusters [74] was modeled show clustered activation related to convergence-divergence from glomerular inputs and could be attributed to firing activity as implied in mass models. MI estimates suggest granule neurons may favor selective processing of spike bursts although not all heterogeneities were included [85, 86]. Induced LTP may allow significantly more information flow and related changes to downstream processing in the cerebellum.

Modeling BOLD signal suggests delayed compliance in MFWM plays a major role in predicting the dynamics of CBV returning to baseline. Evoked postsynaptic LFP was also reconstructed using spiking models suggesting spiking information had key roles in assessing emergent behavior. While assumptions of excitatory-inhibitory balance bridge the behavioral differences from the single neuron level to the neuronal ensemble activity, spiking models reveal the larger role of spiking information in local field estimations. Although the detailed biophysical model [84, 87] did not evaluate the time windowing roles on pattern encoding, excitatory stimuli and cluster activation may define sparse recoding of sensory and tactile inputs in the cerebellum input layer.

Large scale networks models for robotic controllers showcase the role of cerebellum as a pattern separator aiding extension to the Marr-Albus-Ito theories. Reconstructed SNN could act as trajectory predictor and can extract robotic kinematic parameters (manuscript in preparation). Abstractions using cerebellum-like spiking network models can be used to control robotic articulators for unsupervised

and supervised trajectory learning for articulators and can be adapted for timing roles and specific tasks [87, 88].

6 Conclusion

Several neurological conditions such as ataxia, autism, movement coordination conditions show cerebellum and granular layer associations. A brain structure with this social relevance would be fascinating to study. With the focus of functional role of such circuits in neuroinspired computing methods, estimates of population behavior was computed using modeling methods. Through analysis of visual and auditory stimuli it was observed that the cerebellum may perform specific roles including that of a pattern encoding device. This study also looked into cerebellar microcircuit as an applications to robotics which represents a control structure for kinematics.

Mutual information estimates from the cerebellar granular layer need to be further validated to estimate storage capacity of the underlying neurons. Although the cerebellar HRF was not estimated to exactly specify cerebellar function, future studies may need data-based validations of the averaging roles of granule cell clusters as encoded during the BOLD response. Properties connecting single neuron and population activity in cerebellum can help us abstract signals for novel robotic applications and can act as controllers for neuroprosthesis.

Acknowledgment This work derives direction and ideas from the Chancellor of Amrita University, Sri Mata Amritanandamayi Devi. Authors would like to acknowledge Egidio D’Angelo of University of Pavia, Giovanni Naldi and Thierry Nieus of University of Milan, for their support towards work in this manuscript. This work was supported by Grants SR/CSI/49/2010, SR/CSI/60/2011, SR/CSRI/60/2013, SR/CSRI/61/2014 and Indo-Italy POC 2012-2014 from DST, BT/PR5142/MED/30/764/2012 from DBT, and Sir Visveswaraya Faculty fellowship to SD from MeitY, Government of India and partially by Embracing the World.

References

1. Falk, E.B., Berkman, E.T., Lieberman, M.D.: From neural responses to population behavior neural focus group predicts population-level media effects. *Psychol. Sci.* **23**, 439–445 (2012)
2. Rajendran, A., Nutakki, C., Sasidharakurup, H., et al.: Cerebellum in neurological disorders: A review on the role of inter-connected neural circuits. *J. Neurol. Stroke.* **6**, 1–4 (2017). [10.15406/jnsk.2017.6.00196](https://doi.org/10.15406/jnsk.2017.6.00196)
3. D’Angelo, E., Solinas, S., Garrido, J., et al.: Realistic modeling of neurons and networks: towards brain simulation. *Funct. Neurol.* **28**, 153–166 (2013). [10.11138/FNeur/2013.28.3.153](https://doi.org/10.11138/FNeur/2013.28.3.153)
4. Medini, C., Nair, B., D’Angelo, E., et al.: Modeling spike-train processing in the cerebellum granular layer and changes in plasticity reveal single neuron effects in neural ensembles. *Comput. Intell. Neurosci.* **2012**, 359529 (2012). <https://doi.org/10.1155/2012/359529>
5. Eccles, J.C.: The cerebellum as a computer: patterns in space and time. *J. Physiol.* **229**, 1–32 (1973)

6. Ito, M.: Neural design of the cerebellar motor control system. *Brain Res.* **40**, 81–84 (1972)
7. Marr, D.: A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470 (1969)
8. Albus, J.S.: A theory of cerebellar function. *Math. Biosci.* **10**, 25–61 (1971)
9. Arleo, A., Nieuw, T., Bezzi, M., et al.: How synaptic release probability shapes neuronal transmission: information-theoretic analysis in a cerebellar granule cell. *Neural Comput.* **22**, 2031–2058 (2010). https://doi.org/10.1162/NECO_a_00006-Arleo
10. Ishikawa, T., Shimuta, M., Häusser, M.: Multimodal sensory integration in single cerebellar granule cells in vivo. *eLife.* **4**, e12916 (2015). <https://doi.org/10.7554/eLife.12916>
11. Huang, C.C., Sugino, K., Shima, Y., et al.: Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells. *eLife.* (2013). <https://doi.org/10.7554/eLife.00400>
12. Mapelli, L., Gagliano, G., Soda, T., et al.: Granular layer neurons control cerebellar neurovascular coupling through an NMDA receptor/NO-dependent system. *J. Neurosci.* **37**, 1340–1351 (2017). <https://doi.org/10.1523/JNEUROSCI.2025-16.2016>
13. Fox, M.D., Snyder, A.Z., Vincent, J.L., et al.: From the cover: the human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci.* **102**, 9673–9678 (2005). <https://doi.org/10.1073/pnas.0504136102>
14. Henrie, J.A., Shapley, R.: LFP power spectra in V1 cortex: the graded effect of stimulus contrast. *J. Neurophysiol.* **94**, 479–490 (2005). <https://doi.org/10.1152/jn.00919.2004>
15. Kandel, A., Buzsáki, G.: Cellular-synaptic generation of sleep spindles, spike-and-wave discharges, and evoked thalamocortical responses in the neocortex of the rat. *J. Neurosci.* **17**, 6783–6797 (1997)
16. Kreiman, G., Hung, C.P., Kraskov, A., et al.: Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron.* **49**, 433–445 (2006). <https://doi.org/10.1016/j.neuron.2005.12.019>
17. Liu, J., Newsome, W.T.: Local field potential in cortical area MT: stimulus tuning and behavioral correlations. *J. Neurosci.* **26**, 7779–7790 (2006). <https://doi.org/10.1523/JNEUROSCI.5052-05.2006>
18. Mitzdorf, U.: Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* **65**, 37–100 (1985)
19. Montgomery, S.M., Buzsáki, G.: Gamma oscillations dynamically couple hippocampal CA3 and CA1 regions during memory task performance. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14495–14500 (2007). <https://doi.org/10.1073/pnas.0701826104>
20. Roux, S., Mackay, W.A., Riehle, A.: The pre-movement component of motor cortical local field potentials reflects the level of expectancy. *Behav. Brain Res.* **169**, 335–351 (2006). <https://doi.org/10.1016/j.bbr.2006.02.004>
21. Scherberger, H., Jarvis, M.R., Andersen, R.A.: Cortical local field potential encodes movement intentions in the posterior parietal cortex. *Neuron.* **46**, 347–354 (2005). <https://doi.org/10.1016/j.neuron.2005.03.004>
22. Wendling, F., Bellanger, J.J., Bartolomei, F., Chauvel, P.: Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals. *Biol. Cybern.* **83**, 367–378 (2000). <https://doi.org/10.1007/s004220000160>
23. Lopes da Silva, F.H.: Biophysical issues at the frontiers of the interpretation of EEG/MEG signals. *Electroencephalogr. Clin. Neurophysiol. Suppl.* **45**, 1–7 (1996)
24. Jansen, B.H., Rit, V.G.: Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* **73**, 357–366 (1995). <https://doi.org/10.1007/BF00199471>
25. Zetterberg, L.H., Kristiansson, L., Mossberg, K.: Performance of a model for a local neuron population. *Biol. Cybern.* **31**, 15–26 (1978)
26. David, O., Friston, K.J.: A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage.* **20**, 1743–1755 (2003). <https://doi.org/10.1016/j.neuroimage.2003.07.015>
27. Roggeri, L., Rivieccio, B., Rossi, P., D'Angelo, E.: Tactile stimulation evokes long-term synaptic plasticity in the granular layer of cerebellum. *J. Neurosci.* **28**, 6354–6359 (2008). <https://doi.org/10.1523/JNEUROSCI.5709-07.2008>

28. Luque, N.R., Garrido, J.A., Carrillo, R.R., et al.: Cerebellarlike corrective model inference engine for manipulation tasks. *IEEE Trans. Syst. Man Cybern. B Cybern.* **41**, 1299–1312 (2011). <https://doi.org/10.1109/TSMCB.2011.2138693>
29. Wolpert, D.M., Kawato, M.: Multiple paired forward and inverse models for motor control. *Neural Netw.* **11**, 1317–1329 (1998)
30. Rosenblatt, F.: Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington (1962)
31. Albus, J.S.: A new approach to manipulator control: the cerebellar model articulation controller (CMAC). *J. Dyn. Syst. Meas. Control.* **97**(3), 220–227 (1975)
32. Wolpert, D.M., Diedrichsen, J., Flanagan, J.R.: Principles of sensorimotor learning. *Nat. Rev. Neurosci.* **12**, 739–751 (2011). <https://doi.org/10.1038/nrn3112>
33. Vijayan, A., Nutakki, C., Kumar, D., et al.: Enabling a freely accessible open source remotely controlled robotic articulator with a neuro-inspired control algorithm. *Int. J. Interact. Mob. Technol.* **13**, 61–75 (2017)
34. Nieujs, T., Sola, E., Mapelli, J., et al.: LTP regulates burst initiation and frequency at mossy fiber-granule cell synapses of rat cerebellum: experimental observations and theoretical predictions. *J. Neurophysiol.* **95**, 686–699 (2006). <https://doi.org/10.1152/jn.00696.2005>
35. D'Angelo, E., Nieujs, T., Maffei, A., et al.: Theta-frequency bursting and resonance in cerebellar granule cells: experimental evidence and modeling of a slow K^+ -dependent mechanism. *J. Neurosci.* **21**, 759–770 (2001). [pii]
36. Naud, R., Marcille, N., Clopath, C., Gerstner, W.: Firing patterns in the adaptive exponential integrate-and-fire model. *Biol. Cybern.* **99**, 335–347 (2008). <https://doi.org/10.1007/s00422-008-0264-7>
37. Diwakar, S., Magistretti, J., Goldfarb, M., et al.: Axonal Na^+ channels ensure fast spike activation and back-propagation in cerebellar granule cells. *J. Neurophysiol.* **101**, 519–532 (2009). <https://doi.org/10.1152/jn.90382.2008>
38. Medini, C., Rajendran, A.G., Jijibai, A., et al.: Computational characterization of cerebellum granule neuron responses to auditory and visual inputs. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 22–27 (2016)
39. Shannon, C.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948)
40. Ross, B.C., Leonenko, N., Carpena, P., et al.: Mutual information between discrete and continuous data sets. *PLoS One.* **9**, e87357 (2014). <https://doi.org/10.1371/journal.pone.0087357>
41. Buxton, R.B., Wong, E.C., Frank, L.R.: Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med.* **39**, 855–864 (1998)
42. Mandeville, J.B., Marota, J.J., Ayata, C., et al.: Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *J. Cereb. Blood Flow Metab.* **19**, 679–689 (1999). <https://doi.org/10.1097/00004647-199906000-00012>
43. Fox, P.T., Raichle, M.E.: Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1140–1144 (1986)
44. Hoge, R.D., Atkinson, J., Gill, B., et al.: Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9403–9408 (1999). <https://doi.org/10.1073/PNAS.96.16.9403>
45. Bandettini, P.A., Wong, E.C., Hinks, R.S., et al.: Time course EPI of human brain function during task activation. *Magn. Reson. Med.* **25**, 390–397 (1992)
46. Frahm, J., Krüger, G., Merboldt, K.D., Kleinschmidt, A.: Dynamic uncoupling and recoupling of perfusion and oxidative metabolism during focal brain activation in man. *Magn. Reson. Med.* **35**, 143–148 (1996)
47. Krüger, G., Kleinschmidt, A., Frahm, J.: Dynamic MRI sensitized to cerebral blood oxygenation and flow during sustained activation of human visual cortex. *Magn. Reson. Med.* **35**, 797–800 (1996). <https://doi.org/10.1002/mrm.1910350602>
48. Obata, T., Liu, T.T., Miller, K.L., et al.: Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: application of the balloon model to the interpretation of BOLD transients. *Neuroimage.* **21**, 144–153 (2004)

49. Ernst, T., Hennig, J.: Observation of a fast response in functional MR. *Magn. Reson. Med.* **32**, 146–149 (1994)
50. Hu, X., Le, T.H., Ugurbil, K.: Evaluation of the early response in fMRI in individual subjects using short stimulus duration. *Magn. Reson. Med.* **37**, 877–884 (1997). <https://doi.org/10.1002/mrm.1910370612>
51. Yacoub, E., Hu, X.: Detection of the early decrease in fMRI signal in the motor area. *Magn. Reson. Med.* **45**, 184–190 (2001)
52. Malonek, D., Grinvald, A.: Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science* **272**, 551–554 (1996)
53. Buxton, R.B.: The elusive initial dip. *Neuroimage*. **13**, 953–958 (2001). <https://doi.org/10.1006/nimg.2001.0814>
54. Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J.: Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16**, 4207–4221 (1996)
55. Buckner, R.L.: Event-related fMRI and the hemodynamic response. *Hum. Brain Mapp.* **6**, 373–377 (1998)
56. Huettel, S.A., McCarthy, G.: Regional differences in the refractory period of the hemodynamic response: an event-related fMRI study. *Neuroimage*. **14**, 967–976 (2001). <https://doi.org/10.1006/nimg.2001.0900>
57. Corfield, D.R., Murphy, K., Josephs, O., et al.: Does hypercapnia-induced cerebral vasodilation modulate the hemodynamic response to neural activation? *Neuroimage*. **13**, 1207–1211 (2001). <https://doi.org/10.1006/nimg.2001.0760>
58. Davis, T.L., Kwong, K.K., Weisskoff, R.M., Rosen, B.R.: Calibrated functional MRI: mapping the dynamics of oxidative metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1834–1839 (1998)
59. Kastrup, A., Li, T.Q., Krüger, G., et al.: Relationship between cerebral blood flow changes during visual stimulation and baseline flow levels investigated with functional MRI. *Neuroreport*. **10**, 1751–1756 (1999)
60. Kim, S.G., Rostrup, E., Larsson, H.B., et al.: Determination of relative CMRO₂ from CBF and BOLD changes: significant increase of oxygen consumption rate during visual stimulation. *Magn. Reson. Med.* **41**, 1152–1161 (1999)
61. Friston, K.J., Mechelli, A., Turner, R., Price, C.J.: Nonlinear responses in fMRI: the balloon model, Volterra kernels, and other hemodynamics. *Neuroimage*. **12**, 466–477 (2000). <https://doi.org/10.1006/nimg.2000.0630>
62. Grubb, R.L., Raichle, M.E., Eichling, J.O., Ter-Pogossian, M.M.: The effects of changes in PaCO₂ on cerebral blood volume, blood flow, and vascular mean transit time. *Stroke*. **5**, 630–639 (1974)
63. Kong, Y., Zheng, Y., Johnston, D., et al.: A model of the dynamic relationship between blood flow and volume changes during brain activation. *J. Cereb. Blood Flow Metab.* **24**, 1382–1392 (2004). <https://doi.org/10.1097/01.WCB.0000141500.74439.53>
64. Faisal, A.A., Selen, L.P.J., Wolpert, D.M.: Noise in the nervous system. *Nat. Rev. Neurosci.* **9**, 292–303 (2008). <https://doi.org/10.1038/nrn2258>
65. van Rotterdam, A., Lopes da Silva, F.H., van den Ende, J., et al.: A model of the spatial-temporal characteristics of the alpha rhythm. *Bull. Math. Biol.* **44**, 283–305 (1982)
66. Hebbink, G.J.: Activity types in a neural mass model. University of Twente (2014)
67. Jansen, B.H., Zouridakis, G., Brandt, M.E.: A neurophysiologically-based mathematical model of flash visual evoked potentials. *Biol. Cybern.* **68**, 275–283 (1993). <https://doi.org/10.1007/BF00224863>
68. Bhattacharya, B.S., Coyle, D., Maguire, L.P.: A thalamo-cortico-thalamic neural mass model to study alpha rhythms in Alzheimer's disease. *Neural Netw.* **24**, 631–645 (2011). <https://doi.org/10.1016/j.neunet.2011.02.009>
69. Lopes da Silva, F., van Rotterdam, A., Barts, P., et al.: Models of neuronal populations: the basic mechanisms of rhythmicity. *Prog. Brain Res.* **45**, 281–308 (1976)

70. Pettersen, K.H., Linden, H., Dale, A., Einevoll, G.T.: Extracellular spikes and current-source density. In: Brette, R., Destexhe, A. (eds.) *Handbook of Neural Activity Measurement*, pp. 92–130. Cambridge University Press, Cambridge (2012)
71. Gold, C., Henze, D.A., Koch, C., et al.: On the origin of the extracellular action potential waveform: a modeling study. *J. Neurophysiol.* **95**, 3113–3128 (2006). <https://doi.org/10.1152/jn.00979.2005>
72. Gold, C., Henze, D.A., Koch, C.: Using extracellular action potential recordings to constrain compartmental models. *J. Comput. Neurosci.* **23**, 39–58 (2007). <https://doi.org/10.1007/s10827-006-0018-2>
73. Holt, G.R., Koch, C.: Electrical interactions via the extracellular potential near cell bodies. *J. Comput. Neurosci.* **6**, 169–184 (1999)
74. Diwakar, S., Lombardo, P., Solinas, S., et al.: Local field potential modeling predicts dense activation in cerebellar granule cells clusters under LTP and LTD control. *PLoS One.* **6**, e21928 (2011). <https://doi.org/10.1371/journal.pone.0021928>
75. Lindén, H., Pettersen, K.H., Einevoll, G.T.: Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. *J. Comput. Neurosci.* **29**, 423–444 (2010). <https://doi.org/10.1007/s10827-010-0245-4>
76. Lindén, H., Hagen, E., Łęski, S., et al.: LFPy: a tool for biophysical simulation of extracellular potentials generated by detailed model neurons. *Front. Neuroinform.* **7**, 41 (2014). <https://doi.org/10.3389/fninf.2013.00041>
77. Parusuram, H., Nair, B., Naldi, G., et al.: A modeling based study on the origin and nature of evoked post-synaptic local field potentials in granular layer. *J. Physiol. Paris.* **105**, 71–82 (2011). <https://doi.org/10.1016/j.jphysparis.2011.07.011>
78. Parusuram, H., Nair, B., Hines, M., et al.: Computational modeling of single neuron extracellular electric potentials and network local field potentials using LFPsim. *Front. Comput. Neurosci.* **10**, 1–13 (2016). <https://doi.org/10.3389/FNCOM.2016.00065>
79. Kawato, M.: Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* **9**, 718–727 (1999). [https://doi.org/10.1016/S0959-4388\(99\)00028-8](https://doi.org/10.1016/S0959-4388(99)00028-8)
80. Vijayan, A., Nutakki, C., Medini, C., et al.: Classifying movement articulation for robotic arms via machine learning. *J. Intell. Comput.* **4**, 123–134 (2013)
81. Eccles, J.C., Ito, M., Szentágothai, J. The cerebellum as a neuronal machine. *Prog. Brain Res. Adv. Neural Popul. Coding* (1967). [https://doi.org/10.1016/0013-4694\(69\)90099-6](https://doi.org/10.1016/0013-4694(69)90099-6)
82. Jutras, M.J., Buffalo, E.A.: Synchronous neural activity and memory formation. *Curr. Opin. Neurobiol.* **20**, 150–155 (2010). <https://doi.org/10.1016/j.conb.2010.02.006>
83. Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biol. Cybern.* **13**, 55–80 (1973). <https://doi.org/10.1007/BF00288786>
84. Solinas, S., Nieus, T., D’Angelo, E.: A realistic large-scale model of the cerebellum granular layer predicts circuit spatio-temporal filtering properties. *Front. Cell. Neurosci.* **4**, 12 (2010)
85. Avila-Akerberg, O., Krahe, R., Chacron, M.J.: Neural heterogeneities and stimulus properties affect burst coding *in vivo*. *Neuroscience.* **168**, 300–313 (2010). <https://doi.org/10.1016/j.neuroscience.2010.03.012>
86. Li, Z., Ouyang, G., Yao, L., Li, X.: Estimating the correlation between bursty spike trains and local field potentials. *Neural Netw.* **57**, 63–72 (2014). <https://doi.org/10.1016/j.neunet.2014.05.011>
87. D’Angelo, E., De Zeeuw, C.I.: Timing and plasticity in the cerebellum: focus on the granular layer. *Trends Neurosci.* **32**, 30–40 (2008)
88. Antonietti, A., Casellato, C., Geminiani, A., et al.: Healthy and pathological cerebellar spiking neural networks in vestibulo-ocular reflex. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2514–2517. IEEE (2015)

Bifurcation Analysis of a Sparse Neural Network with Cubic Topology



Diego Fasoli, Anna Cattani, and Stefano Panzeri

Abstract We study analytically the changes of dynamics of a firing-rate network model with cubic topology. The present study is performed by extending to this sparse network a formalism we previously developed for the bifurcation analysis of fully-connected circuits. In particular we prove that, unlike the fully-connected model, in the cubic network the neural activity may undergo spontaneous symmetry-breaking even if the network is composed exclusively of excitatory neurons. Moreover, while in the fully-connected topology the symmetry-breaking occurs through pitchfork bifurcations, in the excitatory cubic network it occurs through complex branching-point bifurcations with five branches. These results lead to the conclusion that the sparseness of the synaptic connections may increase the complexity of dynamics compared to dense networks.

1 Introduction

The brain is often described as a complex system organized at multiple spatial scales. At the coarsest level, it can be divided into macroscopic circuits spanning several areas and containing millions of neurons, while at the finest level of organization, basic computations are performed at the single cell level by neurons.

* Diego Fasoli and Anna Cattani, these authors contributed equally to this work.

D. Fasoli (✉)

Laboratory of Neural Computation, Center for Neuroscience and Cognitive Systems@UniTn,
Istituto Italiano di Tecnologia, 38068 Rovereto, Italy

Center for Brain and Cognition, Computational Neuroscience Group, Universitat Pompeu Fabra,
08002 Barcelona, Spain

e-mail: diego.fasoli@iit.it

A. Cattani (✉) · S. Panzeri

Laboratory of Neural Computation, Center for Neuroscience and Cognitive Systems@UniTn,
Istituto Italiano di Tecnologia, 38068 Rovereto, Italy
e-mail: anna.cattani@iit.it; stefano.panzeri@iit.it

However, there is an intermediate, mesoscopic level of organization between the macroscopic and microscopic one: neurons are organized in microcircuits, whose size can vary from several thousands of cells as in cortical columns, to a few tens of cells as in micro-columns [15]. This mesoscopic level of investigation has been the object in recent years of considerable attention, both from the theoretical [2, 11, 20] and experimental [5, 8] point of view.

While previous mathematical descriptions of the dynamics of neural networks focused either on large [3, 4, 10] or on small [6, 13, 16] networks, in our recent work [9] we introduced a mathematical formalism based on bifurcation theory for studying the changes of dynamics of a fully-connected firing-rate network model of arbitrary size. The choice of the activation function of the model, which represents the sigmoidal relation between the membrane potentials and the firing rates of each neuron in the network, does not affect qualitatively the structure of the bifurcation diagram [19]. In particular, the techniques introduced in [9] rely on an algebraic activation function. Interestingly, the use of this function, combined with simple mathematical manipulations of the eigenvalues of the network, allows the study of special bifurcations that would be analytically intractable otherwise. This formalism promises to be of interest to study the dynamics of networks with arbitrary size and with a range of different topologies. To test the capabilities of this formalism, in the present article we extend the results in [9] by relaxing the all-to-all connectivity hypothesis. In particular, we consider a fully homogeneous single-population network with cubic topology. We also suppose that the network is purely excitatory, because in this case the dynamical differences between the cubic topology and the fully-connected one are stronger.

Similarly to [9], whenever the analytical bifurcation analysis is not possible, we combined it with a numerical study of the bifurcations. The numerical analysis is performed through the MatCont Matlab toolbox [7], which is built up the mathematical theory of bifurcations described in [14, 17]. We found a mathematical description of several bifurcations that occur when varying the network parameters. In particular, we found that the purely excitatory cubic network under consideration may undergo the formation of heterogeneous neuronal activity through spontaneous symmetry-breaking due to sparseness of the synaptic connections. On the contrary, in fully-connected networks this phenomenon occurs only if inhibitory populations are present [9]. In the current paper we found that the phenomenon of symmetry-breaking that occurs in the cubic network is characterized by a larger degree of complexity than that of inhibitory fully-connected networks. Indeed, while in fully-connected networks with inhibitory populations symmetry-breaking occurs through pitchfork bifurcations [9], where the number of stationary solutions changes from one to three, in the excitatory cubic network symmetry-breaking occurs through a mathematically more complex branching-point bifurcation, where the number of solutions changes from one to five. This is a consequence of the sparseness of the synaptic connections.

This article is organized as follows. In Sect. 2 we describe the neural model we use. Then, in Sect. 3 we apply our formalism to the cubic network and we describe analytically most of its (local) bifurcations. To conclude, in Sect. 4 we discuss our findings and the advantages of our mathematical formalism.

2 Materials and Methods

We consider a widely used firing-rate model to describe the dynamics of single neurons [9, 12, 18]:

$$\frac{dV_i(t)}{dt} = -\frac{1}{\tau} V_i(t) + \frac{1}{M} \sum_{j=0}^{N-1} J_{ij} \mathcal{A}(V_j(t)) + I(t), \quad i = 0, \dots, N-1. \quad (1)$$

In Eq. (1), N , τ and I represent the size of the network, the membrane time constant and the stimulus to each neuron, respectively, while V_i is the membrane potential of the i th neuron. Moreover, J_{ij} is the weight of the synaptic connection from the j th (presynaptic) neuron to the i th (postsynaptic) neuron, while $\mathcal{A}(\cdot)$ is an activation function which converts the membrane potential V into the corresponding firing rate $v = \mathcal{A}(V)$. In particular, our formalism is based on the choice of the following algebraic activation function:

$$\mathcal{A}(V) = \frac{v^{\max}}{2} \left[1 + \frac{\frac{\Lambda}{2} (V - V^T)}{\sqrt{1 + \frac{\Lambda^2}{4} (V - V^T)^2}} \right], \quad (2)$$

where v^{\max} , V^T and Λ are the maximum firing rate, the threshold and the slope parameter, respectively. Moreover, in Eq.(1), M is a normalization factor that represents the number of incoming connections to each neuron.

The network we consider in this article is composed of eight neurons, which are connected according to the cubic topology shown in Fig. 1. Moreover, we suppose

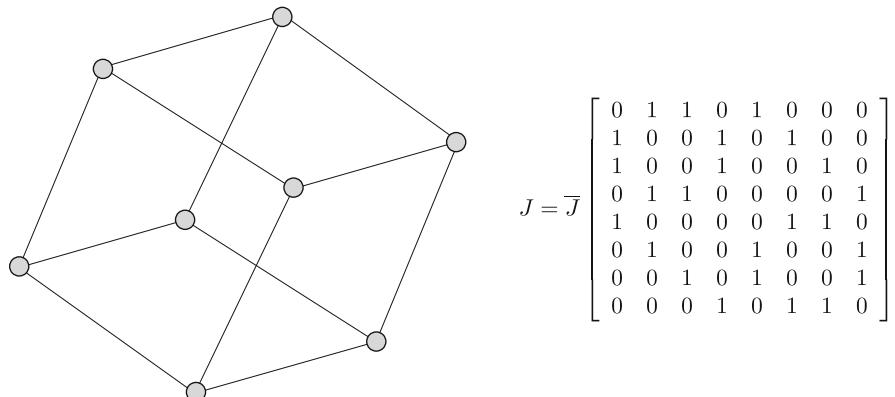


Fig. 1 Network topology. The left panel shows the graph of the cubic network that we study. The right panel shows the corresponding synaptic connectivity matrix, where $\bar{J} > 0$ is a free parameter that represents the overall strength of the excitatory synaptic connections

Table 1 Network parameters

Network size	$N = 8$
Norm. factor	$M = 3$
Membr. time const.	$\tau = 1$
Synaptic weights	$\bar{J} = 20$
Activation function	$v^{\max} = 1, \Lambda = V^T = 2$

This table reports the values of the parameters introduced in Eqs. (1)–(2) and in Fig. 1

that the network is composed of a fully homogeneous population. For this reason, the parameters in Eqs. (1)–(2) do not depend on the neuronal index i (see Table 1).

3 Results

In this work we prove that spontaneous symmetry-breaking can occur also in purely excitatory networks, as a consequence of the sparseness of the synaptic connections. Generally, neural networks undergo spontaneous symmetry-breaking at their branching-point bifurcations. The peculiarity of the cubic model discussed here is the formation of several secondary branches of stationary solutions that emanate from the primary branch at these bifurcation points. For this reason, the results in this section are organized as follows. First, in Sect. 3.1 we compute the primary branch and we shortly introduce the analytical formulas of the eigenvalues of the cubic network. Then, in Sect. 3.2 we use these formulas to characterize the formation of the secondary branches.

3.1 Primary Branch and Eigenvalues

3.1.1 Stationary Solutions

The stationary solutions of the network are obtained by setting $\frac{dV_i(t)}{dt} = 0$ in Eq. (1). Due to the symmetry of the neural equations on the primary branch, the membrane potentials are homogeneous (see the left cubic topology of Fig. 2), so that the stationary membrane potential μ satisfies the following equation:

$$-\frac{1}{\tau}\mu + \bar{J}\mathcal{A}(\mu) + I = 0. \quad (3)$$

After some algebra, Eq. (3) can be transformed into a fourth degree polynomial, which can be solved analytically in order to get an exact expression of μ as a function of I . Alternatively, we can plot the primary branch by observing that Eq. (3)

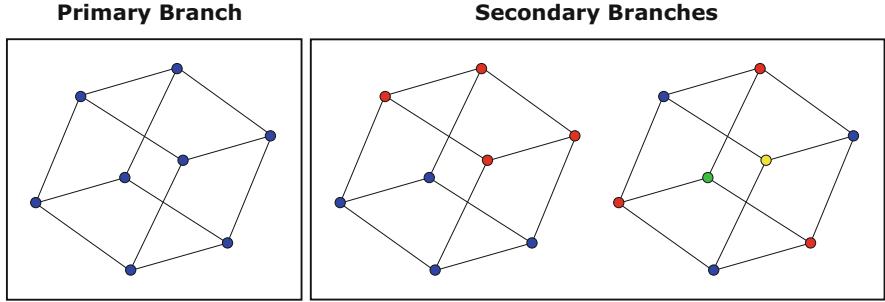


Fig. 2 Forms of spontaneous symmetry-breaking in the cubic network. On the primary branch, the network symmetry is not broken and the membrane potentials (as given by Eq.(3)) are homogeneous (left). On the secondary branches, neuronal symmetry can be broken in two different ways. In the first case (middle), the neurons split into two groups A and B (red and blue), whose membrane potentials are given by Eq.(11). In the second case (right), the neurons split into four groups α , β , γ and δ (red, blue, green and yellow, respectively), whose membrane potentials are given by Eq.(15)

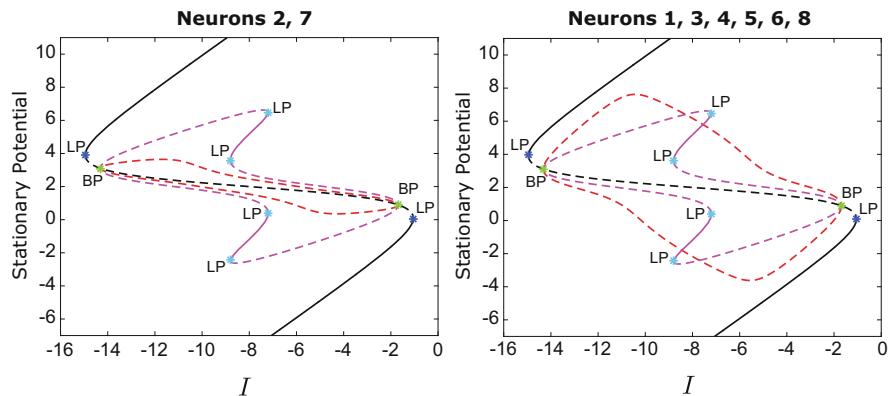


Fig. 3 Codimension one bifurcation diagrams of the cubic network. This figure shows the bifurcation structure in the $I - V$ plane, obtained for $\bar{J} = 20$. In both the panels, the stable/unstable primary stationary curve is described by plain/dashed black curves, while the secondary ones are described by plain/dashed violet and red curves. Colored dots describe the codimension one bifurcations LP and BP. Each panel describes the diagram of a subgroup of excitatory neurons

provides a relation $I = I(\mu)$, which can be plotted on the domain $\mu \in (-\infty, +\infty)$ (see the black curve in Fig. 3).

The eigenvalues of the Jacobian matrix of the cubic network on the primary branch are:

$$\lambda_{0,1}(\mu) = -\frac{1}{\tau} \pm \bar{J}\mathcal{A}'(\mu), \quad \lambda_{2,3}(\mu) = -\frac{1}{\tau} \pm \frac{\bar{J}}{3}\mathcal{A}'(\mu), \quad (4)$$

where $\lambda_{0,1}(\mu)$ and $\lambda_{2,3}(\mu)$ have multiplicity 1 and 3, respectively. As shown in the following sections, Eq.(4) will allow us to derive analytical expressions of the bifurcations.

3.1.2 Limit-Point Bifurcations

Limit-point (also known as *saddle-node*) bifurcations (LP for short) are codimension one bifurcations describing the formation of multiple stationary solutions. For this reason, they satisfy the equation $\frac{dI}{d\mu}\Big|_{\mu_{LP}} = 0$, where μ_{LP} is the stationary membrane potential at the limit-point bifurcations. According to Eq.(3), this relation can be written as follows:

$$\frac{1}{\tau} - \bar{J}\mathcal{A}'(\mu_{LP}) = 0. \quad (5)$$

Therefore, these bifurcations are determined by the condition $\lambda_0(\mu_{LP}) = 0$ (see Eq.(4)). By inverting Eq.(5) we get:

$$\mu_{LP} = \mu_{LP}(\bar{J}) = V^T \pm \frac{2}{\Lambda} \sqrt[3]{\sqrt{\left(\frac{\tau\bar{J}\nu^{\max}\Lambda}{4}\right)^2 - 1}}, \quad (6)$$

for $\bar{J} \geq \frac{4}{\tau\nu^{\max}\Lambda}$. This expression of μ_{LP} can be replaced into Eq.(3), providing:

$$I_{LP} = \frac{1}{\tau}\mu_{LP}(\bar{J}) - \bar{J}\mathcal{A}(\mu_{LP}(\bar{J})). \quad (7)$$

For \bar{J} fixed, Eqs.(6) and (7) provide the coordinates of two limit-point bifurcations in the codimension one bifurcation diagrams (i.e. in the $I - V$ plane, see the blue dots in Fig.3), while Eq.(7) describes the limit-point curves in the codimension two bifurcation diagram for varying \bar{J} (i.e. the $\bar{J} - I$ plane, see the blue line in Fig.4).

3.1.3 Cusp Bifurcation

The network undergoes a cusp bifurcation (CP for short) when the two limit-point bifurcations of the primary branch merge in a single point. Thus, CP is a codimension two bifurcation, and according to Eq.(6) it occurs when:

$$\bar{J} = \bar{J}_{CP} \stackrel{\text{def}}{=} \frac{4}{\tau\nu^{\max}\Lambda}. \quad (8)$$

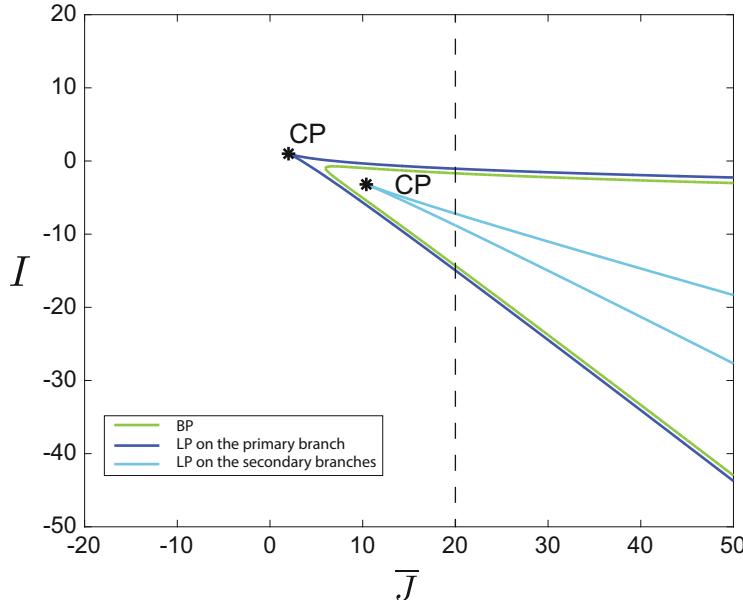


Fig. 4 Codimension two bifurcation diagram of the cubic network. This figure represents the bifurcation structure in the $\bar{J} - I$ plane. The dashed line corresponds to the value of the synaptic weight ($\bar{J} = 20$) that we chose for plotting Fig. 3

Finally, by replacing \bar{J}_{CP} into Eq. (7), we get:

$$I_{\text{CP}} = \frac{V^T}{\tau} - \frac{2}{\tau \Lambda}. \quad (9)$$

The $(\bar{J}_{\text{CP}}, I_{\text{CP}})$ -pair represents the coordinates of the cusp bifurcation in the codimension two bifurcation diagram of the network (see Fig. 4).

3.1.4 Branching-Point Bifurcations

Branching points (BP for short) are another kind of codimension one bifurcations, which increase considerably the complexity of the network dynamics. At these bifurcations the network undergoes the formation of multiple stationary solutions (secondary branches) that emanate from the primary branch. This phenomenon corresponds to a spontaneous symmetry-breaking of the neural activity, since we observe the formation of heterogeneous membrane potentials even if the neural equation (1) does not contain any term that breaks explicitly the symmetry.

According to Eq. (4), for $\bar{J} > 0$ the branching-point bifurcations can be described only by the condition $\lambda_2(\mu_{\text{BP}}) = 0$, where μ_{BP} is the stationary membrane potential

at the branching-point bifurcations (indeed, the eigenvalues λ_1 and λ_3 affect the bifurcation structure only for $\bar{J} < 0$). By inverting this formula, we get:

$$\mu_{\text{BP}} = \mu_{\text{BP}}(\bar{J}) = V^T \pm \frac{2}{\Lambda} \sqrt[3]{\sqrt{\left(\frac{\tau v^{\max} \Lambda \bar{J}}{12}\right)^2 - 1}}$$

$$I_{\text{BP}} = \frac{1}{\tau} \mu_{\text{BP}}(\bar{J}) - \bar{J} \mathcal{A}(\mu_{\text{BP}}(\bar{J})), \quad (10)$$

for $\bar{J} \geq \frac{12}{\tau v^{\max} \Lambda}$. For \bar{J} fixed, Eq.(10) provides the coordinates of two branching-point bifurcations in the codimension one bifurcation diagrams (see the green dots in Fig. 3), while for varying \bar{J} the second equation describes the branching-point curve in the codimension two bifurcation diagram (see the green line in Fig. 4).

3.2 Secondary Branches

In this section we extend our analysis to describe the secondary branches that emanate from the branching-point bifurcations.

3.2.1 Stationary Solutions

We found that, at the branching-point bifurcations, symmetry can be broken in two different ways.

In the first case, we observe the formation of two unstable groups A and B , characterized by distinct membrane potentials (see the middle cubic topology of Fig. 2). According to Eq. (1), the stationary potentials $\mu_{A,B}$ in the two groups satisfy the following system of equations:

$$\begin{cases} -\frac{1}{\tau} \mu_A + \frac{2\bar{J}}{3} \mathcal{A}(\mu_A) + \frac{\bar{J}}{3} \mathcal{A}(\mu_B) + I = 0 \\ -\frac{1}{\tau} \mu_B + \frac{2\bar{J}}{3} \mathcal{A}(\mu_B) + \frac{\bar{J}}{3} \mathcal{A}(\mu_A) + I = 0. \end{cases} \quad (11)$$

By subtracting one equation from the other and after some algebra, we get the following fourth degree polynomial:

$$a\mu_B^4 + b\mu_B^3 + c\mu_B^2 + d\mu_B + e = 0, \quad (12)$$

where:

$$\begin{aligned}
a &= \frac{\Lambda^2}{4\tau^2}, \\
b &= -\frac{\Lambda^2}{2\tau} \left(\phi + \frac{V^T}{\tau} \right), \\
c &= \frac{\Lambda^2}{4} \left[\phi^2 + \left(\frac{V^T}{\tau} \right)^2 + \frac{4}{\tau} V^T \phi \right] + \frac{1}{\tau^2} - \xi, \\
d &= -\frac{\Lambda^2}{2} \phi V^T \left(\frac{V^T}{\tau} + \phi \right) - \frac{2}{\tau} \phi + 2\xi V^T, \\
e &= \left(\frac{\Lambda}{2} \phi V^T \right)^2 + \phi^2 - \xi (V^T)^2, \\
\phi &= \frac{1}{\tau} \mu_A - \frac{\bar{J}}{3} \mathcal{A}(\mu_A) + \frac{\nu^{\max} \bar{J}}{6}, \\
\xi &= \left(\frac{\nu^{\max} \Lambda \bar{J}}{12} \right)^2.
\end{aligned} \tag{13}$$

By following [9], it is possible to prove that this equation admits one solution for $\bar{J} \leq \frac{12}{\tau \nu^{\max} \Lambda}$, and up to three solutions otherwise. For $\bar{J} > \frac{12}{\tau \nu^{\max} \Lambda}$, the new solutions, which represent the secondary branches of Eq. (1), emanate from the primary branch at the potential μ_{BP} , as given by Eq. (10). Equation (12) can be solved exactly, providing an analytical relation $\mu_B = \mu_B(\mu_A, \bar{J})$. By replacing this formula into the first equation of the system (11), we get:

$$I = \frac{1}{\tau} \mu_A - \frac{2\bar{J}}{3} \mathcal{A}(\mu_A) - \frac{\bar{J}}{3} \mathcal{A}(\mu_B(\mu_A, \bar{J})), \tag{14}$$

which can be plotted as a function of μ_A for \bar{J} fixed (see the violet curves in Fig. 3).

However, the symmetry can be broken also in a second, more complex way. In this case, we observe the formation of four unstable groups α, β, γ and δ (see the right cubic topology of Fig. 2), whose membrane potentials satisfy the following system of equations:

$$\begin{cases} -\frac{1}{\tau} \mu_\alpha + \frac{2\bar{J}}{3} \mathcal{A}(\mu_\beta) + \frac{\bar{J}}{3} \mathcal{A}(\mu_\gamma) + I = 0 \\ -\frac{1}{\tau} \mu_\beta + \frac{2\bar{J}}{3} \mathcal{A}(\mu_\alpha) + \frac{\bar{J}}{3} \mathcal{A}(\mu_\delta) + I = 0 \\ -\frac{1}{\tau} \mu_\gamma + \bar{J} \mathcal{A}(\mu_\alpha) + I = 0 \\ -\frac{1}{\tau} \mu_\delta + \bar{J} \mathcal{A}(\mu_\beta) + I = 0. \end{cases} \tag{15}$$

This system can be solved only numerically, and provides the red curves in Fig. 3.

3.2.2 Limit-Point and Cusp Bifurcations

For sufficiently strong \bar{J} , we observe the formation of limit-point bifurcations on the two violet branches. In the codimension one bifurcation diagrams, these bifurcations are represented by the light-blue dots in Fig. 3, while in the codimension two bifurcation diagram by the light-blue curve in Fig. 4. Again, these bifurcations can be calculated only numerically, due to the complexity of the corresponding equations. Interestingly, they cause a change of stability of the secondary branches, leading to the formation of stable solutions of neuronal activity. Moreover, we observe the formation of a new cusp bifurcation on the curve in the codimension two bifurcation diagram, which is analytically intractable.

4 Discussion

In this article we extended the formalism introduced first in [9], by performing the bifurcation analysis of a firing-rate network model with fully homogeneous cubic topology. Our approach allowed us to study analytically most of the changes of dynamics caused by variations of the network parameters, in particular the external stimulus and the strength of the synaptic connections between neurons.

The changes of dynamics occur at the bifurcation points of the network. Through the codimension one and codimension two bifurcation diagrams, we mathematically characterized the limit-point and branching-point bifurcations, which correspond to the formation of multistable solutions of neuronal activity.

The most remarkable result is the formation, for sufficiently strong excitatory connections, of a branching-point bifurcation where the stationary solutions switch from one to five branches. This result proves that spontaneous symmetry-breaking may occur in purely excitatory networks, provided their synaptic connections are sufficiently sparse (compare with [9], where we found that in fully-connected networks the branching-point bifurcations occur only if inhibitory populations are present).

For the sake of example, in this article we applied our formalism to a single-population network with cubic topology. However, it is important to stress that our approach is general enough to be extended to multi-population networks with more complex topologies and arbitrary size, providing an analytical description of most of their (local) bifurcations. In particular, we expect the formation of even more complex branching-point bifurcations than those observed in the cubic network, with a number of secondary branches that depends on the sparseness of the network and on the symmetries of its synaptic connections.

Finally, the findings of this article lead to the conclusion that our formalism may be used for integrating the bifurcation analysis performed by the numerical continuation toolboxes, in order to provide a more complete characterization

of the changes of dynamics that occur in highly symmetric systems. Indeed, standard numerical continuation toolboxes may fail when applied to highly symmetric systems due to the potential formation of multiple branches or to the existence of degenerate eigenvalues [1]. This was confirmed by the fact that, during the numerical analysis that we performed for validating our analytical results, the MatCont Matlab toolbox detected only the red secondary branches of Fig. 3.

Acknowledgements This research was supported by the Autonomous Province of Trento, Call “Grandi Progetti 2012,” project “Characterizing and improving brain mechanisms of attention—ATTEND”.

The funders had no role in study design, data collection and analysis, decision to publish, interpretation of results, or preparation of the manuscript.

References

1. Ashwin, P., Coombes, S., Nicks, R.: Mathematical frameworks for oscillatory network dynamics in neuroscience. *J. Math. Neurosci.* **6**, 2 (2016)
2. Bohlard, J.W., et al.: A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLoS Comput. Biol.* **5**, e1000334 (2009)
3. Bressloff, P.C.: Stochastic neural field theory and the system-size expansion. *SIAM J. Appl. Math.* **70**(5), 1488–1521 (2009)
4. Buice, M.A., Chow, C.C.: Dynamic finite size effects in spiking neural networks. *PLoS Comput. Biol.* **9**(1), e1002872 (2013)
5. Buzsáki, G., Anastassiou, C.A., Koch, C.: The origin of extracellular fields and currents – EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012)
6. Das, P.K., Schieve, W.C.: A bifurcation analysis of the four dimensional generalized Hopfield neural network. *Physica D* **88**, 14–28 (1995)
7. Dhooge, A., Govaerts, W., Kuznetsov, Y.A.: MATCONT: a MATLAB package for numerical bifurcation analysis of ODEs. *ACM Trans. Math. Softw.* **29**(2), 141–164 (2003)
8. Einevoll, G.T., Kayser, C., Logothetis, N.K., Panzeri, S.: Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat. Rev. Neurosci.* **14**, 770–785 (2013)
9. Fasoli, D., Cattani, A., Panzeri, S.: The complexity of dynamics in small neural circuits. *PLoS Comput. Biol.* **12**(8), e1004992 (2016)
10. Faugeras, O., MacLaurin, J.: Asymptotic description of neural networks with correlated synaptic weights. *Entropy* **17**(7), 4701–4743 (2015)
11. Freeman, W.J.: *Neurodynamics: An Exploration in Mesoscopic Brain Dynamics*. Springer, London (2000)
12. Hopfield, J.J.: Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci.* **81**(10), 3088–3092 (1984)
13. Kaslik, E., Balint, S.: Bifurcation analysis for a two-dimensional delayed discrete-time Hopfield neural network. *Chaos Soliton Fract.* **34**, 1245–1253 (2007)
14. Kuznetsov, Y.A.: *Elements of Applied Bifurcation Theory*, vol. 112. Springer, New York (1998)
15. Mountcastle, V.B.: The columnar organization of the neocortex. *Brain* **120**, 701–722 (1997)
16. Pasemann, F.: Complex dynamics and the structure of small neural networks. *Netw. Comput. Neural Syst.* **13**, 195–216 (2002)
17. Strogatz, S.H.: *Nonlinear Dynamics and Chaos*. Sarat Book House, Kolkata (1994)

18. Touboul, J., Hermann, G., Faugeras, O.: Noise-induced behaviors in neural mean field dynamics. *SIAM J. Appl. Dyn. Syst.* **11**(1), 49–81 (2012)
19. Wilson, H.R., Cowan, J.D.: Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**(1), 1–24 (1972)
20. Wright, J.J., Rennie, C.J., Lees, G.J., Robinson, P.A., Bourke, P.D., Chapman, C.L., Gordon, E., Rowe, D.L.: Simulated electrocortical activity at microscopic, mesoscopic, and global scales. *Neuropsychopharmacology* **28**(Suppl. 1), S80–S93 (2003)

Simultaneous Jumps in Interacting Particle Systems: From Neuronal Networks to a General Framework



Luisa Andreis, Paolo Dai Pra, and Markus Fischer

Abstract This paper comes as a survey of the links between recent works on particle systems with simultaneous jumps and the neuroscience literature. We consider systems of N weakly interacting diffusions with jumps, having the peculiar feature that the jump of one component may induce simultaneous jumps of all others. While models belonging to this class have been proposed for the dynamics of neuronal systems, and their limiting ($N \rightarrow +\infty$) behavior has been studied for some special cases, recently a study of propagation of chaos and of the corresponding McKean-Vlasov equation has appeared in a general framework. Here we justify the link between this approach and the neuronal models.

1 Introduction

Neurons are supposed to spread information by means of electrical impulses, called action potentials or *spikes*. A single neuron has its own membrane potential that varies due to external stimuli, to interactions with other neurons and to its own dynamic. When a neuron spikes its membrane potential is rapidly reset to a resting state and, at the same time, other neurons in the network receive an excitatory or inhibitory influence. Recently, models describing networks of spiking neurons by means of the mean-field approach, typical of statistical mechanics, have become widespread in neuroscience. Due to their peculiarities, sometimes these models are raising questions that have their own interest outside of the direct brain modelling. In particular, some recent works on piecewise-deterministic Markov processes for the evolution of neurons membrane potential have displayed the interesting

L. Andreis (✉)

WIAS-Weierstrass institute for applied analysis and stochastics, Mohrenstr. 39, 10117 Berlin, Germany

e-mail: luisa.andreis@wias-berlin.de

P. Dai Pra · M. Fischer

Dipartimento di Matematica, Università degli Studi di Padova, via Trieste 63, 35121 Padova, Italy
e-mail: daipra@math.unipd.it; fischer@math.unipd.it

feature of simultaneous jumps. These models represent the spike of a neuron as a discontinuity in the evolution of its membrane potential, that at the same time induces *collateral discontinuities* in the membrane potential of all the other neurons. As customary in mean field models, these *collateral discontinuities* are rescaled by the factor $\frac{1}{N}$, where N is the size of the system, typical of mean-field models. In the limit, the *collateral jumps* collapse into an additional non-linear drift term but the spike component is preserved. Since this seems to be rather new in mean-field modelling, in [1] the authors give a general description of this class of models, proving propagation of chaos by means of coupling methods under different sets of assumptions. With the aim of giving to specialists a general and flexible class of models with simultaneous jumps, in this paper we summarize the approach and we compare it with the neuroscience models that have inspired the study.

2 Mean Field Models in Neuroscience

The mean-field approach in neuroscience consists in describing large populations of neurons of the same types by means of the behavior of a so-called “typical neuron”. The large number of neurons and of connections between them make indeed reasonable to describe some parts of the brain (as the visual cortex), even though they are finite-size networks, as the infinite-size limit of a system of particles in mean-field interactions, i.e. where the graph of interactions is complete. This approach origins in statistical mechanics, from the seminal work of Kac [9], in which the author builds a microscopic system of interacting Markov processes, representing the molecules of a rarefied gas, to justify the macroscopic description given by the spatially homogeneous Boltzmann equation. The link between microscopic and macroscopic level is given by the *propagation of chaos*, see the well-known reference from Sznitmann [11]. Propagation of chaos basically says that, when the size of the system grows to infinity, the particles tends to de-correlate, despite their interaction. As observed by Galves and Löcherbach in [7], it is hard to find a systematical overview on the biological justification and experimental confirmation of propagation of chaos in the brain behavior, although the goodness of this approach seems to be validated in Baladron et al. [2]. There the authors cite experimental results in [4], where de-correlation of neuronal firing in visual cortex is observed.

2.1 Neuroscience Models with Simultaneous Jumps

Models with Poisson spikes account for the intrinsic randomness of spikes describing them by means of inhomogeneous Poisson processes with rate depending on the membrane potential. In this framework, the membrane potential is modelled by a piecewise deterministic Markov process and the interaction occurs through simultaneous jumps. When a neuron spikes, randomly according to its rate, it is reset to zero and it makes the membrane potential of every other neuron increase

by a small quantity, depending on the synaptic weight between the two neurons. In this way, the jumps in the network are simultaneous and, even if some of them are of the order $\frac{1}{N}$, they may cause problems when letting the size N of the network going to infinity. The literature on mean field models with jumps is less rich than the one on continuous models, nevertheless in some recent papers [3, 5, 7, 10] the authors prove propagation of chaos for models in this class. These toy models have been proposed, based on a previous network model by Galves and Löcherbach [6], in the two linked papers [3, 5]. At the same time, another group of authors, Robert and Touboul [10], has proposed a similar neuronal network study. These models describe the membrane potentials of neurons as quantities on the positive real line. Let us summarize the main peculiarities of the model by describing the finite size network of neurons. Let $N \geq 1$ be a fixed finite number of neurons in an homogeneous network (i.e. where the neurons are all of the same type), we associate to each neuron an index $i = 1, \dots, N$ and we describe the membrane potential of the network with the stochastic process $U^N(t) = (U_1^N(t), \dots, U_N^N(t)) \in \mathbb{R}_+^N$ for every $t \geq 0$, where $U_i^N(t)$ is the membrane potential of the i -th neuron at time t . The membrane potential of a neuron exponentially decays towards the resting state (here it is 0) due to the *leak current*, a continuous flow of potential. Therefore neuron i has a drift proportional to

$$-U_i^N(t) .$$

Then the neurons interact by means of *electrical synapses* and, through the *gap-junction channels*, they constantly share their potential. This pushes the system towards the average potential value, that means that the i -th neuron has also a drift proportional to

$$\sum_{j=1}^N \frac{U_j^N(t)}{N} - U_i^N(t) .$$

Finally, *chemical synapses* cause fast-events, the spikes. A neuron spikes randomly according to a state dependent rate

$$\lambda(U_i^N(t)) \geq 0 .$$

If $\lambda(0) = 0$, it is supposed that there is no external stimuli, while a positive value in 0 means that the neuron can spike even when it is at resting state, due to some external input. When neuron i spikes, its membrane potential is reset at 0 by a jump of amplitude $-U_i^N(t^-)$. Simultaneously, the non-spiking neurons receive an additional discrete influence, they increase their potential by a quantity depending on a stochastic *synaptic efficacy*. That results in a jump of amplitude

$$\frac{W_{i,j}}{N}$$

of the membrane potential $U_j^N(t^-)$ when the i -th neuron spikes and this happens simultaneously for all $j \neq i$. The above description corresponds to a

piecewise-deterministic Markov evolution for the process U^N , that is solution of the following system of SDEs. For all $i = 1, \dots, N$

$$\begin{aligned} dU_i^N(t) &= -\alpha U_i^N(t)dt - \beta \left(U_i^N(t) - \sum_{j=1}^N \frac{U_j^N(t)}{N} \right) dt \\ &\quad - U_i^N(t^-) \int_0^\infty \mathbb{1}_{[0, \lambda(U_i^N(t^-))]}(u) \mathcal{N}^i(du, dt) \\ &\quad + \sum_{j \neq i} \frac{W_{j,i}}{N} \int_0^\infty \mathbb{1}_{[0, \lambda(U_j^N(t^-))]}(u) \mathcal{N}^j(du, dt), \end{aligned} \quad (1)$$

where $\{\mathcal{N}^i\}_{i=1,\dots,N}$ is a family of independent Poisson processes with characteristic measure $dudt$. In the papers [3, 5], the authors study the case with $\alpha = 0$ and $W_{i,j} \equiv 1$ for all $i, j = 1, \dots, N$; while in [10] the authors study the case of $\beta = 0$ and the synaptic weights $W_{i,j}$ are i.i.d. positive bounded random variables. It is clear that the interactions here are all of mean-field type, but while the one due to *electrical synapses* is classical, the one given by *chemical synapses* is rather peculiar. Indeed these simultaneous jumps, one of which will remain in the limit, while the others collapse in a continuous term because of the rescaling of the order $\frac{1}{N}$, seem to be new in the mean-field models framework. In the aforementioned papers, the authors succeed in proving propagation of chaos under super-linear hypothesis on the rate function λ .

3 A General Mean Field Model with Simultaneous Jumps

In this section, we describe the general framework for particle systems with simultaneous jumps introduced in [1] and we summarize the results of that paper. The idea comes from the desire to understand if the peculiarity of the simultaneous jumps can create problems in the proof of propagation of chaos in situations different from the one described above, for example in presence of a Brownian component.

3.1 The Microscopic Dynamics

Let $X^N = (X_1^N, \dots, X_N^N) \in \mathbb{R}^{d \times N}$ be the spatial positions of N different particles moving in \mathbb{R}^d . We introduce the corresponding *empirical measure*

$$\mu_X^N \doteq \frac{1}{N} \sum_{i=1}^N \delta_{X_i^N}.$$

When the time variable appears explicitly in $X^N(t)$, we write $\mu_X^N(t)$ to indicate the time dependence of the empirical measure. Note that $\mu_X^N(t)$ is an element of $\mathcal{M}(\mathbb{R}^d)$, the set of probability measures on the Borel subsets of \mathbb{R}^d . In the following, the dependence on μ_X^N of the coefficients encodes the mean-field interaction, that is indeed an interaction via the empirical measure.

The particle positions $X^N(t)$ evolve as a jump diffusion process in $\mathbb{R}^{d \times N}$ with the following specifications for the i -th particle.

- The drift coefficient: $F(X_i^N(t), \mu_X^N(t))$ for some function $F : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ common to all particles.
- The diffusion coefficient: $\sigma(X_i^N(t), \mu_X^N(t))$ for $\sigma : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times d_1}$, again the same for all particles.
- The main jump rate: particle i performs a *main jump* with rate $\lambda(X_i^N(t), \mu_X^N(t))$, for a positive function $\lambda : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \rightarrow [0, \infty)$. With this rate, the i -th particle performs a *main jump* and simultaneously it induces in all the other particles a *collateral jump*.
- The main jump amplitude: particle i perform a main jump that is a random variable

$$\psi(X_i^N(t), \mu_X^N(t), h_i^N),$$

for a function $\psi : \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \times [0, 1] \rightarrow \mathbb{R}^d$. Here h^N is a random variable with values in $[0, 1]^N$ and its distribution is given by the symmetric measure ν_N .

- The collateral jump amplitude: the i -th particle is induced to jump by *main jumps* of every other particle. The amplitude of these *collateral jumps* is given by the function $\Theta : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \times [0, 1]^2 \rightarrow \mathbb{R}^d$. When the j -th particle jumps (this occurs with rate $\lambda(X_j^N(t), \mu_X^N(t))$, of course) the i -th particle performs a jump of amplitude

$$\frac{\Theta(X_j^N(t), X_i^N(t), \mu_X^N(t), h_j^N, h_i^N)}{N},$$

where h_i^N and h_j^N are components of a random vector h^N , with distribution ν_N .

The idea is to encode classical mean-field interaction in the dependence of the coefficients on the empirical measure. For instance, the drift coefficient due to *electrical synapses* in (1) would be written as

$$F(X_i^N, \mu_X^N) = \int_{\mathbb{R}^d} y \mu_X^N(dy) - X_i^N.$$

Additionally, there is the interaction given by the simultaneous jumps. To be as general as possible, the *main jump* of the i -th particle (that in (1) corresponds to the spike of the i -th neuron) depends not only on the position of the jumping particle, but also on the empirical measure and on a random component. This random

component is the i -th coordinate of the random vector h^N , whose distribution is given by the measure ν_N . Notice that, in this way, we do not require the randomness of the main jump i to be independent from the others, but only that the measure ν_N is symmetric. In this framework, the *collateral jumps* depends on the position of both the main jumping particle and the affected particle (the one forced to take the collateral jump), on the empirical measure and on two random parameters, components of the random vector h^N .

Rigorously, let us consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$ satisfying the usual hypotheses, rich enough to carry an independent family $(B_i, \mathcal{N}^i)_{i \in \mathbb{N}}$ of d -dimensional Brownian motions B_i and Poisson random measures \mathcal{N}^i with characteristic measure $dtduv(dh)$. Here v is a symmetric probability measure on $[0, 1]^{\mathbb{N}}$ such that, for every $N \geq 1$, the previous mentioned measure ν_N coincides with the projection of v on the first N coordinates. We will construct X^N as the solution of the following SDE:

$$\begin{aligned} dX_i^N(t) = & F(X_i^N(t), \mu_X^N(t))dt + \sigma(X_i^N(t), \mu_X^N(t))dB_t^i \\ & + \int_{[0, \infty) \times [0, 1]^{\mathbb{N}}} \psi(X_i^N(t^-), \mu_X^N(t^-), h_i) \mathbb{1}_{(0, \lambda(X_i^N(t^-), \mu_X^N(t^-)))}(u) \mathcal{N}^i(dt, du, dh) \\ & + \frac{1}{N} \sum_{j \neq i} \int_{[0, \infty) \times [0, 1]^{\mathbb{N}}} \Theta(X_j^N(t^-), X_i^N(t^-), \mu_X^N(t^-), h_j, h_i) \\ & \quad \mathbb{1}_{(0, \lambda(X_j^N(t^-), \mu_X^N(t^-)))}(u) \mathcal{N}^j(dt, du, dh), \end{aligned} \quad (2)$$

$i = 1, \dots, N$. The existence and uniqueness of a solution starting from a vector of initial conditions $(X_1^N(0), \dots, X_N^N(0))$ depends obviously on the assumptions on the coefficients, and we will specify sufficient conditions in the following sections.

Let us briefly justify the form in which we write the Poisson integrals in (2). In stochastic analysis literature the non-compensated jump component is often represented by a measure that does not directly describe the behavior of the system. Here, we want to highlight the role of the jumps and of their rates, as in the neuroscience works [5, 10]. Therefore we describe a diffusion process that at each position $X^N \in \mathbb{R}^{d \times N}$ has a certain jump rate, given by the sum of the jump rates of all particles $\sum_{i=1}^N \lambda(X_i^N, \mu^N)$, and a set of possible jumps, represented by the set of random dN -dimensional vectors $\Delta_i^N(X^N, \mu^N, h^N)$ for all $i = 1, \dots, N$ such that

$$\Delta_i^N(X^N, \mu^N, h^N)_j := \begin{cases} \psi(X_i^N(t), \mu_X^N(t), h_i^N) & \text{for } j = i; \\ \frac{\Theta(X_i^N(t), X_j^N(t), \mu_X^N(t), h_i^N, h_j^N)}{N} & \text{for } j \neq i. \end{cases}$$

One of the aims of [1] is to give results without uniform boundedness assumptions on the jump rate, the function λ . That seems to be a natural request for application purposes, but in the literature for mean-field models with jumps the study of models with unbounded jump rate seems to be very rare, an example can be found in Graham, [8]. Moreover the Poisson random measures appearing in Eq.(2) have characteristic measure defined on $[0, \infty)^2 \times [0, 1]^{\mathbb{N}}$. The system could equivalently be stated in terms of Poisson random measures with characteristic measures defined on $[0, \infty)^2 \times [0, 1]^N$ (namely, $l \times l \times \nu_N$). The reason for our seemingly unnatural choice is that it prepares for the coupling argument used to establish the propagation of chaos with a rate.

3.2 The Macroscopic Process

Suppose the solution X^N of (2) exists, and that its initial condition has a permutation invariant distribution. Fix an arbitrary component i , and assume the process X_i^N has a limit in distribution; by symmetry, the law of the limit does not depend on i , so we denote by X the limit process. To identify, at a heuristic level, its law, we make the further assumption that a law of large numbers holds, i.e. $\mu_X^N(t)$ converges, as $N \rightarrow +\infty$, to the law μ_t of $X(t)$. Letting $N \rightarrow +\infty$ in (2) we deduce, at a purely formal level, that the limit process $X(t)$ has the law of the solution of the *McKean-Vlasov SDE*:

$$\begin{aligned} dX(t) = & \left(F(X(t), \mu_t) + \left\langle \mu_t, \lambda(\cdot, \mu_t) \int_{[0,1]^2} \Theta(\cdot, X(t^-), \mu_t, h_1, h_2) v_2(dh_1, dh_2) \right\rangle \right) dt \\ & + \sigma(X(t), \mu_t) dB_t + \int_{[0, \infty) \times [0, 1]^{\mathbb{N}}} \psi(X(t^-), \mu_s, h_1) \\ & \mathbb{1}_{(0, \lambda(X(t^-), \mu_s))}(u) \mathcal{N}(dt, du, dh). \end{aligned} \quad (3)$$

Here, B is a d_1 -dimensional Brownian motion and \mathcal{N} an independent Poisson random measure with characteristic measure $dtduv(dh)$ on $[0, \infty)^2 \times [0, 1]^{\mathbb{N}}$ as above. By $\langle \cdot, \cdot \rangle$ we indicate the integral of a function on its domain with respect to a certain measure; thus, $\langle \mu, \phi \rangle = \int_{\mathbb{R}^d} \phi(y) \mu(dy)$. Note that (3) is not a standard SDE since the law μ_t of the solution appears as an argument of its coefficients. It is often referred to as McKean-Vlasov SDE, as it is customary to call McKean-Vlasov equation the partial differential equation solved by the law μ_t , namely, in the weak form,

$$\langle \mu_t, \phi \rangle - \langle \mu_0, \phi \rangle = \int_0^t \langle \mu_s, \mathcal{L}(\mu_s) \phi \rangle ds,$$

where

$$\begin{aligned} \mathcal{L}(\mu_t)\phi(x) &\doteq F(x, \mu_t)\partial\phi(x) + \frac{1}{2}\sum_{j,k=1}^d a(x, \mu_t)_{jk}\partial^2\phi(x)_{jk} \\ &+ \left\langle \mu_t, \lambda(\cdot, \mu_t) \int_{[0,1]^2} \Theta(\cdot, x, \mu_t, h_1, h_2) v_2(dh_1, dh_2) \right\rangle \partial\phi(x) \\ &+ \lambda(x, \mu_t) \int_{[0,1]} (\phi(x + \psi(x, \mu_t, h_1)) - \phi(x)) v_1(dh_1). \end{aligned}$$

3.3 Assumptions on Coefficients

We intentionally presented the particle system and the nonlinear process in full generality, without specifying any condition on coefficients, but those are of course needed to prove well-posedness of the SDEs and the desired propagation of chaos. In this section we briefly present the notation and some conditions under which propagation of chaos is proven in [1]. First of all, let $\mathcal{M}^1(\mathbb{R}^d)$ be the space of probability on \mathbb{R}^d with finite first moment:

$$\mathcal{M}^1(\mathbb{R}^d) = \{\mu \in \mathcal{M}(\mathbb{R}^d) : \int \|x\| \mu(dx) < +\infty\}.$$

This space is equipped with the W_1 Wasserstein metric:

$$\rho(\mu, \nu) \doteq \inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\| \pi(dx, dy); \pi \text{ has marginals } \mu \text{ and } \nu \right\}.$$

We also consider the following subset of $\mathcal{M}(\mathbf{D}([0, T], \mathbb{R}^d))$, the set of the probability measures on $\mathbf{D}([0, T], \mathbb{R}^d)$:

$$\mathcal{M}^1(\mathbf{D}([0, T], \mathbb{R}^d)) \doteq \left\{ \alpha \in \mathcal{M}(\mathbf{D}([0, T], \mathbb{R}^d)) : \int_{\mathbf{D}} \sup_{t \in [0, T]} \|x(t)\| \alpha(dx) < +\infty \right\},$$

and provide it with the metric

$$\rho_T(\alpha, \beta) \doteq \inf \left\{ \int_{\mathbf{D} \times \mathbf{D}} \sup_{t \in [0, T]} \|x(t) - y(t)\| P(dx, dy); \text{ where } P \text{ has marginals } \alpha \text{ and } \beta \right\}.$$

In the following we state one set of conditions on the coefficients of system (2) and the corresponding nonlinear SDE (3). It recalls in its form classical Lipschitz-type

assumptions for nonlinear SDE with a super linear drift term and it depicts a quite general framework, in view of possible modelling purposes.

Assumption 1

- The drift coefficient $F: \mathbb{R}^d \times \mathcal{M}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is of the form

$$F(x, \alpha) = -\nabla U(x) + b(x, \alpha),$$

for all $x \in \mathbb{R}^d$ and all $\alpha \in \mathcal{M}(\mathbb{R}^d)$, where U is convex and \mathcal{C}^2 . The function b is assumed to be globally Lipschitz in both variables, i.e. there exists $L > 0$ such that for all $x, y \in \mathbb{R}^d$, all $\alpha, \gamma \in \mathcal{M}^1(\mathbb{R}^d)$,

$$\|b(x, \alpha) - b(y, \gamma)\| \leq L(\|x - y\| + \rho(\alpha, \gamma)).$$

- The classical global Lipschitz assumption on σ : $\exists \hat{L} > 0$ such that, for all $x, y \in \mathbb{R}^d$, all $\alpha, \gamma \in \mathcal{M}^1(\mathbb{R}^d)$,

$$\|\sigma(x, \alpha) - \sigma(y, \gamma)\| \leq \hat{L}(\|x - y\| + \rho(\alpha, \gamma)).$$

- The integrability condition: for all $N \in \mathbb{N}$, for all $\mathbf{x} \in \mathbb{R}^{d \times N}$

$$\sup_{i \in \{1, \dots, N\}} \sup_{\alpha \in \mathcal{M}^1(\mathbb{R}^d)} \lambda(x_i, \alpha) \int_0^T \int_{[0,1]^N} \|\Delta_i^N(\mathbf{x}, \alpha, h^N)\| v_N(dh^N) dt < \infty.$$

- The L^1 -Lipschitz assumption on the jump coefficients: $\exists \bar{L} > 0$ such that, for all $x, y \in \mathbb{R}^d$, all $\alpha, \gamma \in \mathcal{M}^1(\mathbb{R}^d)$,

$$\begin{aligned} & \int_{[0, \infty) \times [0,1]} \|\psi(x, \alpha, h) \mathbb{1}_{(0, \lambda(x, \alpha))}(u) - \psi(y, \gamma, h) \\ & \quad \mathbb{1}_{(0, \lambda(y, \gamma))}(u)\| du v_1(dh) \leq \bar{L}(\|x - y\| + \rho(\alpha, \gamma)) \end{aligned}$$

and

$$\begin{aligned} & \left\| \langle \alpha, \lambda(\cdot, \alpha) \int_{[0,1]^2} \Theta(\cdot, x, \alpha, h_1, h_2) v_2(dh_1, dh_2) \rangle \right. \\ & \quad \left. - \langle \gamma, \lambda(\cdot, \gamma) \int_{[0,1]^2} \Theta(\cdot, y, \gamma, h_1, h_2) v_2(dh_1, dh_2) \rangle \right\| \\ & \leq \bar{L}(\|x - y\| + \rho(\alpha, \gamma)). \end{aligned}$$

Under Assumption 1 the system of SDE and the nonlinear SDE in the previous sections are well-posed for all square integrable initial conditions.

3.4 Propagation of Chaos and Rate of Convergence

We already mentioned that the link between the microscopic dynamics (2) and the macroscopic limit (3) is explained by the phenomenon of propagation of chaos. To describe it, let us introduce the notion of chaoticity in W_1 Wasserstein distance.

Definition 1 Let $X^N = (X_1^N, X_2^N, \dots, X_N^N)$ be a sequence of random vectors with components $X_i^N \in \mathbb{R}^d$ (resp. $X_i^N \in \mathbf{D}([0, T], \mathbb{R}^d)$). For $\mu \in \mathcal{M}^1(\mathbb{R}^d)$ (resp. $\mu \in \mathcal{M}^1(\mathbf{D}([0, T], \mathbb{R}^d))$), we say that X^N is μ -chaotic in W_1 if its distribution is permutation invariant and, for each $k \in \mathbb{N}$, the law of the vector $(X_1^N, X_2^N, \dots, X_k^N)$ converges to $\mu^{\otimes k}$ with respect to the metric ρ (resp. ρ_T).

Then, let μ be a probability measure on \mathbb{R}^d . We assume that the sequence of the distributions of $X^N(0)$ is μ -chaotic. Fix an arbitrary time horizon $T > 0$, and denote by $X^N[0, T] = (X^N(t))_{t \in [0, T]}$ the random path of the microscopic process up to time T . It is said that *propagation of chaos* holds if the distribution of $X^N[0, T]$ is itself Q -chaotic for some probability measure Q on the Skorohod space of càdlàg functions $\mathbf{D}([0, T], \mathbb{R}^d)$. Notice that this property is equivalent to the statement: the sequence of the laws of the empirical measures $\{\mu_X^N\}_{N \in \mathbb{N}}$ for the solutions of (2), converges to the law of the solution of (3) in the W_1 Wasserstein distance, see [11].

Notice that, in addition to the aforementioned assumptions, for the proof of propagation of chaos, we will need the following square integrability condition on the amplitude of the collateral jumps.

Assumption 2

$$\int_0^T \int_{[0, \infty) \times [0, 1]^N} \|\Theta(x, y, \alpha, h_1, h_2) \mathbb{1}_{(0, \lambda(x, \alpha))}(u)\|^2 du v_2(dh) dt < \infty,$$

for all $x, y \in \mathbb{R}^d$ and all $\alpha \in \mathcal{M}^1(\mathbb{R}^d)$.

First, we aim to quantify the role of the simultaneous jumps in the propagation of chaos, i.e. to understand the rate of convergence to zero of that distance. Therefore we introduce an *intermediate process* $Y^N = (Y^N(t))_{t \in [0, T]}$ with values in $\mathbb{R}^{d \times N}$. This Markov process Y^N can be given as the solution of the SDE

$$\begin{aligned} dY_i^N(t) &= F(Y_i^N(t), \mu_Y^N(t))dt + \sigma(Y_i^N(t), \mu_i^N(t))dB_t^i \\ &+ \int_{[0, \infty) \times [0, 1]^N} \psi(Y_i^N(t^-), \mu_Y^N(t^-), h) \mathbb{1}_{(0, \lambda(Y_i^N(t^-), \mu_i^N(t^-)))]}(u) \mathcal{N}^i(dt, du, dh) \\ &+ \frac{1}{N} \sum_{j=1}^N \lambda(Y_j^N(t^-), \mu_Y^N(t^-)) \\ &\quad \int_{[0, 1]^2} \Theta(Y_j^N(t^-), Y_i^N(t^-), \mu_Y^N(t^-), h_1, h_2) v_2(dh_1, dh_2) dt, \end{aligned} \tag{4}$$

$i = 1, \dots, N$, where again B^i are independent d -dimensional Brownian motions and \mathcal{N}^i are independent Poisson random measures with characteristic measure $dtduv(dh)$. It is immediate to see that the process Y^N differs from the original process X^N in the jump terms; indeed, here the *collateral jumps* have been absorbed by a new drift term, while the amplitude of the remaining jumps affects only one component at a time. It is possible to couple the processes X^N and Y^N and to show that the L^1 -convergence to zero of $X_1^N[0, T] - Y_1^N[0, T]$ is of the order $\frac{1}{\sqrt{N}}$.

Theorem 1 *Grant Assumptions 1 and 2. Let X^N and Y^N be the solutions of (2) and (4). We assume the two processes are driven by the same Brownian motions and Poisson random measures, and start from the same square integrable and permutation invariant initial condition. Then there exists a constant $C_T > 0$ such that, for each fixed $i \in \mathbb{N}$, for all $N \geq 1$*

$$\mathbf{E} \left[\sup_{t \in [0, T]} \|X_i^N(t) - Y_i^N(t)\| \right] \leq \frac{C_T}{\sqrt{N}}.$$

Then, propagation of chaos for X^N follows after showing that the law of Y^N is Q -chaotic, but since Y^N has no simultaneous jumps, this can be obtained along the lines of the classical approach.

Theorem 2 *Grant Assumptions 1 and 2. Let μ_0 be a probability measure on \mathbb{R}^d such that $\int \|x\|^2 \mu_0(dx) < +\infty$. For $N \in \mathbb{N}$, let X^N be a solution of Eq. (2) in $[0, T]$. Assume that $X^N(0) = (X_1^N(0), \dots, X_N^N(0))$, $N \in \mathbb{N}$, form a sequence of square integrable random vectors that is μ_0 -chaotic in W_1 . Let μ be the law of the solution of Eq. (3) in $[0, T]$ with initial law $\mathbf{P} \circ X(0)^{-1} = \mu_0$. Then X^N is μ -chaotic in W_1 .*

As a conclusion, let us underline that the model presented here aims to be as flexible as possible, in order to encode in the same general framework a great number of mean-field models that are usually treated with more ad hoc procedures. For example the presence of a nonlinear drift term, a diffusion coefficient and an unbounded jump rate allows the formal definition of mean-field conductance-based models as the ones in [2] with discontinuities and gives a rate of convergence to the macroscopic equation, which is particularly useful in simulations. Of course this procedure is strictly related to the *mean-field* form of the interaction and cannot be extended, for example, to networks where the neurons have been observed to be strongly coupled.

References

1. Andreis, L., Dai Pra, P., Fischer, M.: McKean-Vlasov limit for interacting systems with simultaneous jumps. (2016, Preprint). [arXiv:1704.01052](https://arxiv.org/abs/1704.01052)
2. Baladron, J., Fasoli, D., Faugeras, O., Touboul, J.: Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. J. Math. Neurosci. (2012). <https://doi.org/10.1186/2190-8567-2-10>

3. De Masi, A., Galves, A., Löcherbach, E., Presutti, E.: Hydrodynamic limit for interacting neurons. *J. Stat. Phys.* **158**, 866–902 (2014)
4. Ecker, A.S., Berens, P., Keliris, G.A., Bethge, M., Logothetis, N.K., Tolias, A.S. Decorrelated neuronal firing in cortical microcircuits. *Science* **327**, 584–587 (2010)
5. Fournier, N., Löcherbach, E.: On a toy model of interacting neurons. *Ann. Inst. H. Poincaré Probab. Stat.* **52**, 1844–1876 (2016)
6. Galves, A., Löcherbach, E.: Infinite systems of interacting chains with memory of variable length stochastic model for biological neural nets. *J. Stat. Phys.* **151**, 896–921 (2013)
7. Galves, A., Löcherbach, E.: Modeling networks of spiking neurons as interacting processes with memory of variable length. *J-SFdS* **157**, 17–32 (2016)
8. Graham, C.: McKean-Vlasov Itô-Skorohod equations, and nonlinear diffusions with discrete jump sets. *Stoch. Proc. Appl.* **40**, 69–82 (1992)
9. Kac, M.: Foundations of kinetic theory. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3, pp. 171–197. University of California Press, Berkeley (1956)
10. Robert, P., Touboul, J.: On the dynamics of random neuronal networks. *J. Stat. Phys.* **165**, 545–584 (2016)
11. Sznitman, A.S.: Topics in propagation of chaos. In: *Ecole d’Eté de Probabilités de Saint-Flour XIX—1989*, pp. 165–251. Springer, Berlin (1991)

Neural Fields: Localised States with Piece-Wise Constant Interactions



Aytül Gökçe, Stephen Coombes, and Daniele Avitabile

Abstract Neural field models are typically cast as continuum integro-differential equations for describing the idealised coarse-grained activity of populations of interacting neurons. For smooth Mexican hat kernels, with short-range excitation and long-range inhibition, these non-local models can support various localised states in the form of spots in two-dimensional media. In recent years, there has been a growing interest in the mathematical neuroscience community in studying such models with a Heaviside firing rate non-linearity, as this often allows substantial insight into the stability of stationary solutions in terms of integrals over the kernels. Here we consider the use of piece-wise constant kernels that allow the explicit evaluation of such integrals. We use this to show that azimuthal instabilities are not possible for simple piece-wise constant Top Hat interactions, whilst they are easily realised for piece-wise constant Mexican hat interactions.

1 Introduction

Since the 1970s, there has been much progress in developing analytical and numerical techniques to understand brain function and the nervous system. The cortex, being the most complex part of this system, is interconnected by numerous cortico-cortical fibers ($\sim 10^{10}$), whose axon length ranges from 0.5 to 3 mm [17]. Due to its large surface area ($\sim 1600\text{--}4000\text{ cm}^2$ in total) with a small thickness ($\sim 3\text{ mm}$), the cortex is often regarded as a two-dimensional laminar structure [12, 17]. Neural field modelling, in both one and two dimensions, is a very well-known framework for approximating the coarse-grained dynamics of a real cortex [8]. It assumes that various classes of neuron types in different cortical layers and areas can be divided into two main sub-populations of neurons (excitatory

A. Gökçe (✉) · S. Coombes · D. Avitabile

School of Mathematical Sciences, The University of Nottingham, University Park, NG7 2RD
Nottingham, UK

e-mail: Aytul.Gokce@nottingham.ac.uk; Stephen.Coombes@nottingham.ac.uk;
Daniele.Avitabile@nottingham.ac.uk

and inhibitory), interacting with each other and themselves [1]. This modelling of neural activity traces its roots back to the seminal works of Wilson and Cowan [20, 21] and Amari [1]. Similar models motivated by these original works have been developed to understand phenomena such as epileptic seizures [22], orientation selectivity in visual cortex [3, 4], spiral waves in disinhibited neocortex [15, 16], and spatio-temporal patterns seen in large scale electroencephalogram (EEG) and magnetoencephalogram (MEG) neuroimaging studies [7, 9].

Techniques for the analysis of continuum population activity with local excitatory and distal inhibitory interactions (so-called Mexican-hat connectivity) have been developed to study spatially periodic and localised states in neural field models, particularly for smooth synaptic kernels [9, 18]. Here we concentrate on neural fields with piece-wise constant kernels, which have been far less studied. The exception to this statement being the work of Herrmann et al. [14]. The scalar two-dimensional neural field model that we consider is given by

$$\frac{\partial u(\mathbf{r}, t)}{\partial t} = -u(\mathbf{r}, t) + \int_{\mathcal{D}} w(\mathbf{r}, \mathbf{r}') F(u(\mathbf{r}', t)) d\mathbf{r}', \quad (1)$$

where \mathcal{D} is a planar domain $\mathcal{D} \subseteq \mathbb{R}^2$. The variable u represents average synaptic activity at position $\mathbf{r} \in \mathcal{D}$ at a time $t \in \mathbb{R}^+$. The non-linear function F represents the firing rate of the tissue and is taken to be a Heaviside as in the seminal work of Amari [1], such that $F(u) = H(u - h)$ for some constant threshold h . See [2] for a recent perspective by Amari on the “Heaviside World”. Here, the synaptic kernel w describes anatomical connectivity. For mathematical convenience we choose the kernel to depend only on Euclidean distance, so that $w(\mathbf{r}, \mathbf{r}') = w(|\mathbf{r} - \mathbf{r}'|)$. It is now of particular interest to consider synaptic connectivity kernels for which explicit analytical progress can be made. In the first instance, let us consider a piece-wise constant rotationally symmetric (Top hat) connectivity defined by

$$w(r) = \begin{cases} w_+ > 0, & r \leq \sigma \\ w_- < 0, & r > \sigma \end{cases}, \quad r = |\mathbf{r}|, \quad (2)$$

where w_+ and w_- are constant positive and negative synaptic strength factors respectively. An illustration of a piece-wise constant Top hat function is shown in Fig. 1.

Another natural choice is the piece-wise constant Mexican hat shape given by

$$w(r) = \begin{cases} w_+ > 0, & r \leq \sigma_1 \\ w_- < 0, & \sigma_1 < r \leq \sigma_2, \\ 0, & r > \sigma_2 \end{cases}. \quad (3)$$

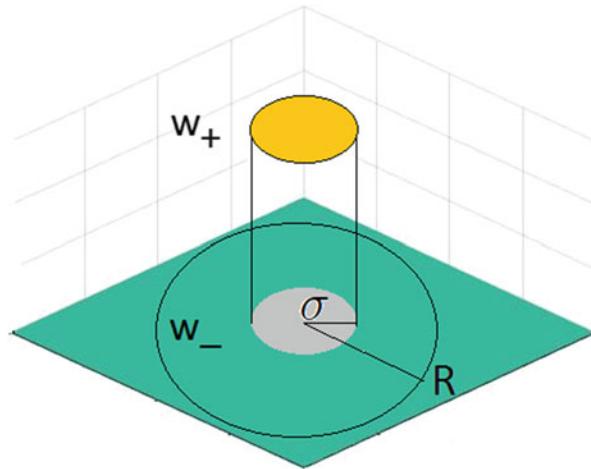


Fig. 1 An illustration of a piece-wise constant Top hat kernel. Synaptic interactions are positive up to a distance σ ($w_+ > 0$) and negative beyond this distance ($w_- < 0$)

In Sect. 2, we show how piece-wise constant caricatures of synaptic connectivity allow for simple calculations relating to localised solutions of neural fields in two-dimensions. Self-consistent equations for stationary solutions of spots are explicitly constructed in Sect. 2.1 and their stability is determined in Sect. 2.2. Here, we find that localised circular spots, obtained using piece-wise constant Mexican kernels, can destabilise to azimuthal instabilities, leading to non-circular patterns with multiple lobes. The numerical solutions beyond a predicted bifurcation point can lead to the generation of labyrinthine structures. Interestingly, a piece-wise constant Top hat kernel is more robust to azimuthal instabilities. Lastly in Sect. 3 we discuss possible extensions of the work in this article.

2 Neural Fields in Circular Geometries: Top Hat Interactions

Here we focus on the construction of rotationally symmetric (spot) solutions for neural fields with piece-wise constant Top hat interactions described by Eq. (2) and determine the stability of these solutions. Following the work of Herrmann et al. [14], we show how explicit solutions, for rotationally symmetric spots, can be easily constructed using simple geometric ideas. We further show that calculation of their stability is straightforward.

2.1 Construction

Stationary solutions $q(\mathbf{r})$ of Eq. (1) with Heaviside firing rate satisfy

$$q(\mathbf{r}) = \int_{\Omega} d\mathbf{r}' w(\mathbf{r} - \mathbf{r}'), \quad (4)$$

where $\Omega = \{\mathbf{r} \mid q(\mathbf{r}) > h\}$ represents the area over which the local field is excited. For circularly symmetric spot solutions of radius R we have that $q(\mathbf{r}) = U(r)$, where

$$U(r) = \iint_{|\mathbf{r}'| < R} d\mathbf{r}' w(|\mathbf{r} - \mathbf{r}'|), \quad U(R) = h. \quad (5)$$

For the piece-wise constant Top hat kernel given by Eq. (2), we may split the above integral as

$$U(r) = w_+ \iint_{\substack{|\mathbf{r}'| < R \\ |\mathbf{r} - \mathbf{r}'| < \sigma}} d\mathbf{r}' + w_- \iint_{\substack{|\mathbf{r}'| < R \\ |\mathbf{r} - \mathbf{r}'| > \sigma}} d\mathbf{r}'. \quad (6)$$

Introducing the area $A_+(\sigma)$ as follows

$$A_+(\sigma) = \left. \iint_{\substack{|\mathbf{r}'| < R \\ |\mathbf{r} - \mathbf{r}'| < \sigma}} d\mathbf{r}' \right|_{r=R}, \quad (7)$$

means that the self-consistent equation for a spot takes the form

$$h = (w_+ - w_-)A_+(\sigma) + w_- \pi R^2. \quad (8)$$

Here, the area $A_+(\sigma)$ can be calculated in terms of the area of overlap of two circles, one of center $(0, 0)$ and radius R , and the other of center \mathbf{r} and radius σ subject to the constraint $|\mathbf{r}| = R$. Using the results from the Appendix we find

$$A_+(\sigma) = A(R, \phi_0) + A(\sigma, \phi_1), \quad (9)$$

where $A(r, \phi) = r^2(\phi - \sin \phi)/2$ and

$$\phi_0 = 2 \cos^{-1} \left(\frac{2R^2 - \sigma^2}{2R^2} \right), \quad \phi_1 = 2 \cos^{-1} \left(\frac{\sigma}{2R} \right), \quad R > \frac{\sigma}{2}. \quad (10)$$

For example, for the special case that $R = \sigma$ and $h = 0$, it is easy to show

$$A_+(\sigma) = 2A(R, 2\pi/3) = R^2(2\pi/3 - \sqrt{3}/2). \quad (11)$$

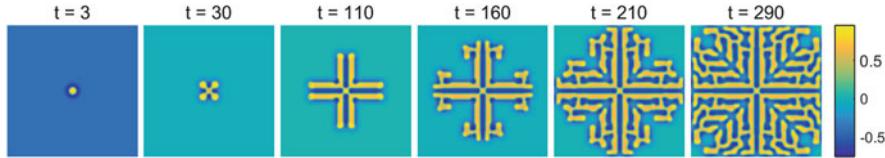


Fig. 2 Direct numerical simulations of a spreading pattern governed by the space-time model (1) with a radially symmetric piece-wise constant Mexican hat kernel on a domain of size $[-L, L] \times [-L, L]$. Parameter values: $w_+ = 0.1$, $w_- = -0.004$, $h = 0.1$, $\sigma_1 = 2$, $\sigma_2 = 10$, $L = 100$. Here, yellow and blue regions represent the excited (high activity) and quiescent states (low activity), respectively

Using Eq. (8), we obtain the ratio

$$\frac{w_+}{w_-} = 1 - \frac{\pi}{(2\pi/3 - \sqrt{3}/2)}, \quad (12)$$

which recovers the results in [14].

Another natural piece-wise constant choice for synaptic connectivity is the piece-wise constant Mexican hat shape given by (3). Using a similar argument to the one used for the piece-wise constant Top hat connectivity, we find the self consistent equation for a localised spot

$$h = (w_+ - w_-)A_+(\sigma_1) + w_-A_+(\sigma_2). \quad (13)$$

In Fig. 2, we show the results of direct numerical simulations at fixed times for a neural field with a piece-wise constant Mexican hat kernel. These simulations show the emergence of an exotic mazelike pattern with a four fold symmetry.

A plot of the theoretical spot radius as a function of the firing threshold is shown in Fig. 3. Here we see two branches of solutions, so that a wide and narrow spot can co-exist. Direct numerical simulations show, for the chosen parameters, that the wider one is stable. Moreover, the wide spot appears to be stable to deformations that change the radius according to $R \rightarrow R + \epsilon \cos m\theta$, where $\theta \in [0, 2\pi]$, $m \in \mathbb{N}$ and $|\epsilon| \ll 1$. In this setting, the simulations in Fig. 2 suggest that the spot is unstable to an azimuthal instability with $m = 4$. We shall investigate this robustness to *azimuthal instabilities* in more detail in the next sub-section. Note that, similar exotic patterns with a smooth Mexican hat connectivity have been found and discussed in [6, 8, 18].

The numerical evolution of (1) was performed using two-dimensional fast Fourier transforms to compute convolutions, discretising space into 2^9 points on a periodic domain, and evolving the resultant set of ordinary differential equations using MATLAB 2015a with a standard non-stiff *ode45* solver. For a further discussion see [8, 19].

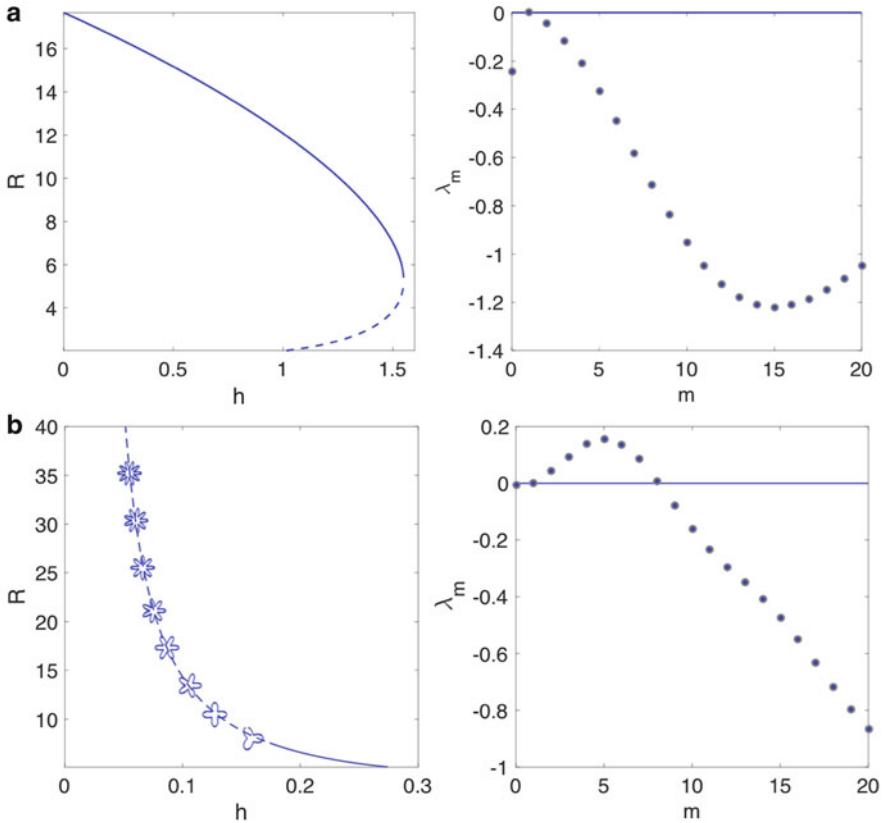


Fig. 3 Bifurcation diagrams for radius R as a function of the threshold h . **(a)** Piece-wise constant Top hat kernel, with the right panel showing the spectrum for $h = 0.1$. Parameters are $w_+ = 0.08$, $w_- = -0.002$, $\sigma = 4$ with constraint $2R > \sigma$. **(b)** Piece-wise constant Mexican hat kernel, with the right panel showing the spectrum for $h = 0.1$. Parameters are $w_+ = 0.1$, $w_- = -0.004$, $\sigma_1 = 2$, $\sigma_2 = 10$ with constraint $2R > \sigma_2 > \sigma_1$

2.2 Stability

To determine the linear stability of a spot we follow [5, 6, 8, 18] and write $u(r, t) = U(r) + e^{\lambda t} \cos(m\phi)\delta u(r)$ where $0 < \delta u \ll 1$ and $m \in \mathbb{N}$. Expanding (1) to first order gives

$$(\lambda + 1)\delta u(r) = \int_0^{2\pi} d\phi \cos(m\phi) \int_0^\infty r' dr' w(|r - r'|) H'(U(r') - h) \delta u(r'), \quad (14)$$

where we write $|\mathbf{r} - \mathbf{r}'| = \sqrt{r^2 + r'^2 - 2rr' \cos(\phi - \phi')}$ with $\mathbf{r} = (r, \phi)$ and $\mathbf{r}' = (r', \phi')$. Since H' occurs under an integral, we can formally write

$$H'(U(r) - h) = \frac{\delta(r - R)}{|U'(R)|}, \quad (15)$$

to yield

$$(\lambda + 1)\delta u(r) - \delta u(R) \frac{R}{|U'(R)|} \int_0^{2\pi} d\phi \cos(m\phi) w(|\mathbf{r} - \mathbf{r}'|) \Big|_{r'=R} = 0. \quad (16)$$

By setting $r = R$ and demanding non-trivial solutions we obtain an equation for the eigenvalues λ in the form $\mathcal{E}_m(\lambda) = 0$, $m \in \mathbb{N}$, where

$$\mathcal{E}_m(\lambda) = \lambda + 1 - \frac{R}{|U'(R)|} \int_0^{2\pi} d\phi \cos(m\phi) w(|\mathbf{r} - \mathbf{r}'|) \Big|_{r'=r=R}. \quad (17)$$

A localised spot solution will be stable if $\lambda_m < 0$ for all $m \in \mathbb{N}$, where λ_m is a zero of $\mathcal{E}_m(\lambda)$. Note that $\mathcal{E}_1(0) = 0$ is always true since a radially symmetric solution $q(r, \phi) = U(r)$ is invariant under the transformation $\phi \rightarrow \phi + \phi_0$, $\phi_0 \in [0, 2\pi]$. Once more we emphasise that the nature of constant piece-wise kernels substantially simplifies the mathematical expressions presented above. For example for the piece-wise Top hat function given by (2) it is simple to show that

$$U'(R) = \frac{\sigma(w_- - w_+)}{R} \sqrt{4R^2 - \sigma^2}, \quad (18)$$

and

$$\int_0^{2\pi} d\phi \cos(m\phi) w(|\mathbf{r} - \mathbf{r}'|) \Big|_{r'=r=R} = 2 \left(\frac{w_+ - w_-}{m} \right) \sin m\phi^*, \quad (19)$$

where ϕ^* is the smaller of the two roots of the equation $R\sqrt{2(1 - \cos\phi)} = \sigma$ for $\phi \in [0, 2\pi]$. In Fig. 3, plots of the radius R as a function of the threshold parameter h , and the spectra of selected profiles are shown for the piece-wise constant Top hat (Fig. 3a) and Mexican hat (Fig. 3b) kernels. Our linear stability analysis of the stationary circular solutions for the piece-wise constant Top hat kernel shows that lower (dashed) branch of solutions is unstable to uniform changes (expansion or contraction) of size ($m = 0$) and the upper (solid) branch is stable. Here, from Eqs. (17)–(19), it is straightforward to show that piece-wise constant Top hat kernels do not support azimuthal instabilities of spots with modes $m \geq 2$. However, the piece-wise constant Mexican hat kernel readily supports azimuthal instabilities of various modes ($m \geq 2$). These results are consistent with those previously found by several authors for smooth Mexican hat kernels, as in [5, 10, 11, 18].

For further details on pattern formation in planar neural fields with smooth kernels and a Heaviside firing rate see [5, 8, 18].

3 Discussion

In this article we have shown that the analysis of the Amari model simplifies even further for the choice of piece-wise constant synaptic kernels. These kernels allow the analysis of localised states in the form of spots and their azimuthal instabilities that can lead to spreading labyrinthine structures.

There are several possible extensions of the work presented in this article. Perhaps the most obvious extension is to expand the analysis to tackle bounded domains with imposed boundary conditions. Since the pioneering work of Amari [1], neural field models have usually been considered to describe the behaviour of tissue in infinite domains, and solutions on bounded domains have drawn very little attention. Focusing on the Dirichlet boundary conditions, the key mathematical idea to treat this has recently been formulated and discussed in [13].

Another extension would be to consider doubly periodic solutions with $u(\mathbf{r} + \mathbf{l}_{1,2}, t) = u(\mathbf{r}, t)$, for linearly independent vectors $\mathbf{l}_{1,2} \in \mathbb{R}^2$. From Eq. (4) the doubly periodic stationary solution is given by

$$U(r) = \sum_{m,n \in \mathbb{Z}} \iint_{|\mathbf{r}' + m\mathbf{l}_1 + n\mathbf{l}_2| < R} w(|\mathbf{r} - \mathbf{r}'|) d\mathbf{r}', \quad (20)$$

subject to the constraint $U(R) = h$. In Fig. 4, numerical solutions of the full neural field model show the evolution of patterns starting from doubly periodic solutions. For example in Fig. 4a, we observe destabilisation of a hexagonal tiling into an overlapping concentric circular-like pattern. In Fig. 4b the pattern destabilises to large spots that are compressed and trapped by surrounding small spots. We see a regularly deforming pattern which looks like a chequered flag in Fig. 4c, and lastly the initial conditions destabilise to a star-like pattern in Fig. 4d.

A further analysis of doubly periodic solutions will be presented elsewhere, including the use of numerical continuation methods to determine solutions and their stabilities by considering the ideas in [19].

Appendix: Circular Geometry for a Top Hat Kernel

Consider a portion of a disk whose upper boundary is an (circular) arc and whose lower boundary is a chord making a central angle $\phi_0 < \pi$, illustrated as the shaded region in Fig. 5a. The area $A = A(r_0, \phi_0)$ of the (shaded) segment is then simply given by the area of the circular sector (the entire wedge-shaped portion) minus the

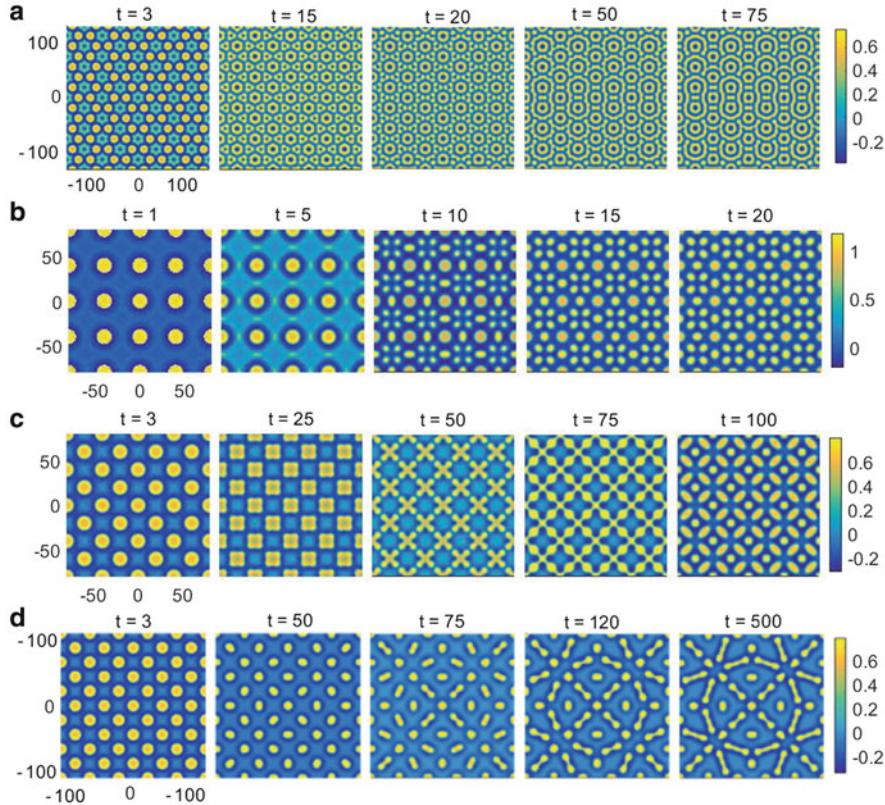


Fig. 4 Various space time simulations for a piece-wise constant Mexican hat kernel. Panels (a) and (b) are simulated with a sigmoidal firing rate function of the form $F(u) = 1/(1+e^{-\mu(u-h)})$. Panels (c) and (d) are simulated with a Heaviside firing rate. Parameters are (a) $w_+ = 0.1, w_- = -0.004, \mu = 6, \sigma_1 = 2, \sigma_2 = 10, h = 0.12, L = 150$, (b) $w_+ = 0.1, w_- = -0.004, \mu = 10, \sigma_1 = 2, \sigma_2 = 10, h = 0.23, L = 80$, (c) $w_+ = 0.1, w_- = -0.004, \sigma_1 = 2, \sigma_2 = 10, h = 0.14, L = 80$, (d) $w_+ = 0.1, w_- = -0.004, \sigma_1 = 2, \sigma_2 = 10, h = 0.283, L = 100$

area of an isosceles triangle, namely

$$A(r_0, \phi_0) = \frac{1}{2}r_0^2(\phi_0 - \sin \phi_0). \quad (21)$$

The area of the overlap of two circles, as illustrated in Fig. 5b, can be constructed as the total area of $A(r_0, \phi_0) + A(r_1, \phi_1)$. To determine the angles $\phi_{0,1}$ in terms of the centers, (x_0, y_0) and (x_1, y_1) , and radii, r_0 and r_1 , of the two circles we use the cosine formula that relates the lengths of the three sides of a triangle formed by joining the centers of the circles to a point of intersection. We denote distance between the two

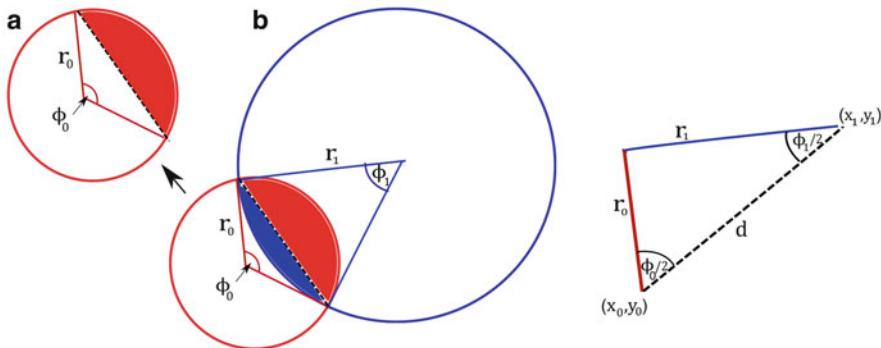


Fig. 5 The area of the total shaded segment is $r_0^2(\phi_0 - \sin \phi_0)/2$ (a). Overlap of two circles shows the area of active region (b)

centers by d where $d^2 = (x_0 - x_1)^2 + (y_0 - y_1)^2$ so that

$$r_1^2 = r_0^2 + d^2 - 2r_0d \cos(\phi_0/2), \quad r_0^2 = r_1^2 + d^2 - 2r_1d \cos(\phi_1/2). \quad (22)$$

Hence the angles are given by

$$\phi_0 = 2 \cos^{-1} \left(\frac{r_0^2 + d^2 - r_1^2}{2r_0d} \right), \quad \phi_1 = 2 \cos^{-1} \left(\frac{r_1^2 + d^2 - r_0^2}{2r_1d} \right). \quad (23)$$

References

1. Amari, S.: Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**(2), 77–87 (1977)
2. Amari, S.: Heaviside world: excitation and self-organization of neural fields. In: Neural Fields: Theory and Applications. Springer, Berlin, Heidelberg (2014)
3. Bressloff, P.C., Carroll, S.R.: Spatiotemporal dynamics of neural fields on product spaces. *SIAM J. Appl. Dyn. Syst.* **13**(4), 1620–1653 (2014)
4. Bressloff, P.C., Carroll, S.R.: Laminar neural field model of laterally propagating waves of orientation selectivity. *PLoS Comput. Biol.* **11**(10), e1004545 (2015)
5. Bressloff, P.C., Coombes, S.: Neural ‘Bubble’ dynamics revisited *Cogn. Comput.* **5**(3), 281–294 (2013)
6. Coombes, S.: Waves, bumps, and patterns in neural field theories. *Biol. Cybern.* **93**(2), 91–108 (2005)
7. Coombes, S.: Large-scale neural dynamics: simple and complex. *NeuroImage* **52**(3), 731–739 (2010)
8. Coombes, S., Schmidt, H., Bojak, I.: Interface dynamics in planar neural field models. *J. Math. Neurosci.* **2**(1), 1 (2012)
9. Coombes, S., Beim Graben, P., Potthast, R., Wright, J.: Neural Fields. Springer, Berlin (2014)
10. Coombes, S., Schmidt, H., Avitabile, D.: Spots: breathing, drifting and scattering in a neural field model. In: Neural Fields, pp. 187–211. Springer, Berlin (2014)

11. Ermentrout, G.B., Folias, S.E., Kilpatrick, Z.P.: Spatiotemporal pattern formation in neural fields with linear adaptation. In: Neural Fields, pp. 119–151. Springer, Berlin (2014)
12. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci.* **97**(20), 11050–11055 (2000)
13. Gökçe, A., Avitabile, D., Coombes, S.: The dynamics of neural fields on bounded domains: an interface approach for Dirichlet boundary conditions. *J. Math. Neurosci.* **7**(1), 12 (2017)
14. Herrmann, J. M., Schröbsdorff, H., Geisel, T.: Localized activations in a simple neural field model. *Neurocomputing* **65**, 679–684 (2005)
15. Huang, X., Troy, W.C., Yang, Q., Ma, H., Laing, C.R., Schiff, S.J., Wu, J.-Y.: Spiral waves in disinhibited mammalian neocortex. *J. Neurosci.* **24**(44), 9897–9902 (2004)
16. Laing, C.R.: Spiral waves in nonlocal equations. *SIAM J. Appl. Dyn. Syst.* **4**(3), 588–606 (2005)
17. Nunez, P.L., Srinivasan, R.: Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, Oxford (2006)
18. Owen, M.R., Laing, C.R., Coombes, S.: Bumps and rings in a two-dimensional neural field: splitting and rotational instabilities. *New J. Phys.* **9**(10), 378 (2007)
19. Rankin, J., Avitabile, D., Baladron, J., Faye, G., Lloyd, D.J.: Continuation of localized coherent structures in nonlocal neural field equations. *SIAM J. Sci. Comput.* **36**(1), B70–B93 (2014)
20. Wilson, H.R., Cowan, J.D.: Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**(1), 1 (1972)
21. Wilson, H.R., Cowan, J.D.: A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **13**(2), 55–80 (1973)
22. Zhao, X., Robinson, P.: Generalized seizures in a neural field model with bursting dynamics. *J. Comput. Neurosci.* **39**(2), 197–216 (2015)

Mathematical Models of Visual Perception Based on Cortical Architectures



Marta Favali, Giovanna Citti, and Alessandro Sarti

Abstract We present a mathematical model of figure-ground articulation, which takes into account gestalt laws and is compatible with the functional architecture of the primary visual cortex (V1) to obtain low-level object segmentation. Connectivity kernels, derived from Lie group theory, are used to describe the gestalt law of good continuation. The spectral analysis of connectivity matrix derived from these kernels allows to individuate perceptual units with the highest saliency. Some applications of this model to the problem of individuation of perceptual units in illusory figures are presented, as clear examples of problems of visual perception.

1 Introduction

The theory of the Berliner Gestalt psychology comes to the aid of explaining segmentation and grouping in illusory figures (see Fig. 1) and proposed local and global laws to describe the properties of a visual stimulus [28, 29]. Among the local laws we recall the principle of proximity, similarity and good continuation and in particular the local law of good continuation plays a central role in the problem of perceptual grouping. In order to study good continuation, Field, Hayes and Hess introduced in [12] the concept of an association field, that defines which properties the elements of the stimuli should have in order to be associated to the same perceptual unit, such as co-linearity and co-circularity. An outline of association fields is represented in Fig. 2, where the orientation depends on the orientation of the central element.

M. Favali (✉) · A. Sarti

Center of Mathematics, CNRS-EHESS, 190-198 avenue de France, Paris, France
e-mail: marta.favali@gmail.com; alessandro.sarti@ehess.fr

G. Citti

Dipartimento di Matematica, Università di Bologna, Piazza di Porta San Donato 5, Bologna, Italy
e-mail: giovanna.citti@unibo.it

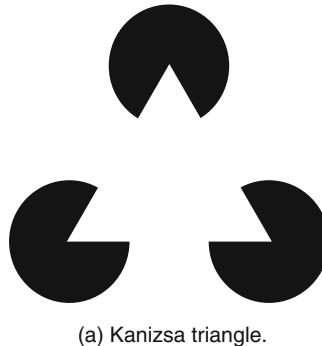


Fig. 1 Examples of illusory figures: the Kanizsa triangle

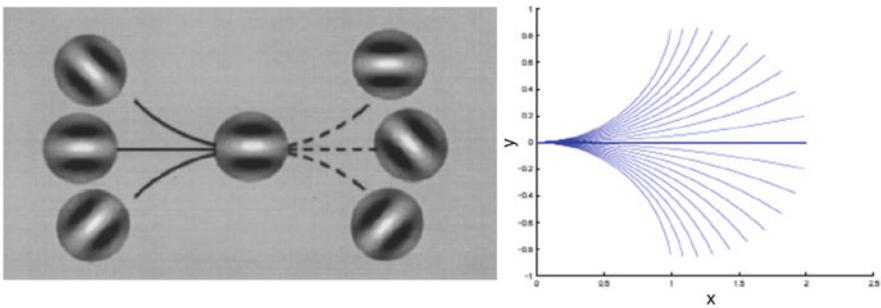


Fig. 2 On the left the association field of [12], that describe the rules of connection between patches, on the right the projection of the integral curves on the cortical plane

In [4] Bosking showed how the rules of association fields are implemented in the primary visual cortex (V1), where neurons with similar orientation are connected with long-range horizontal connectivity. A geometric model of the association fields based on the functional organization of V1 has been proposed in [5]. This geometric approach is part of the research line proposed by Koenderink and van Doorn [17], Hoffman [14], Mumford [18], Zucker [31], Sarti et al. [27], Petitot and Tondut [20] and applications to image processing can be found in [3, 7, 8, 11].

In this work we will see that the connectivity is modelled with the Fokker-Planck and the sub-Riemannian Laplacian kernel defined on the distributional lift in [5], which means associating to each point of the retina (x_1, x_2) the point $(x_1, x_2, \bar{\theta})$ in the cortex, where $\bar{\theta}$ is the maximal orientation selected by the intra-columnar connectivity (this process is called *lifting*). We will accomplish grouping with a spectral analysis inspired from the work of [25], who proved the neurophysiological plausibility of this process, and we will develop this approach to describe an algorithm for the individuation of perceptual units. See also [26] for neural grouping in the visual cortex.

2 The Mathematical Model

2.1 Lifting of the Stimulus in the Cortical Space

In this section we recall the structure of the geometry of the primary visual cortex (V1). Hubel and Wiesel [16] first discovered that the visual cortex is organized in a hypercolumnar structure, where each point corresponds to a simple cell sensitive to a stimulus positioned in (x_1, x_2) and with orientation θ . Also, it is well known that objects with different orientations can be identified by the brain even when they are partially occluded, noisy or interrupted [15]. Motivated by these findings, a new transformation was proposed [9, 10], to lift all elongated structures in 2D images to a new space of positions and orientations ($\mathbb{R}^2 \times S^1$) using elongated and oriented wavelets. By lifting the stimulus, multiple orientations per position could be detected. Using this transformation and considering only one orientation per position (the orientation with highest transformation response), the points of a curve $\gamma = (x_1, x_2)$ are lifted to new cortical curves and are described in $\mathbb{R}^2 \times S^1$: $(x_1, x_2) \rightarrow (x_1, x_2, \theta)$.

These curves have been modelled in [5] as integral curves of suitable vector fields:

$$\mathbf{X}_1 = (\cos \theta, \sin \theta, 0), \quad \mathbf{X}_2 = (0, 0, 1). \quad (1)$$

and it has been shown that are always tangent to the planes generated by the vector fields. The points of the lifted curves are connected by integral curves of these two vector fields such that:

$$\gamma : \mathbb{R} \rightarrow SE(2), \quad \gamma(s) = (x_1(s), x_2(s), \theta(s)) \quad (2)$$

$$\gamma'(s) = (k_1(s)\mathbf{X}_1 + k_2(s)\mathbf{X}_2)(\gamma(s)), \quad \gamma(0) = 0.$$

where the coefficients k_1 and k_2 represent a distance in the (x_1, x_2, θ) domain.

The metric responsible for the connectivity in the cortex is then sub-Riemannian, in particular lifted curves in the cortex will not have tangent direction along $X_3 = [X_2, X_1]$. These curves projected on the 2D cortical plane represent a good model of the association fields, as described in [5] and visualized in Fig. 2.

2.2 The Connectivity Kernels and the Affinity Matrix

In this section we present a probabilistic version of the connectivity field on the basis of Fokker Planck equations first introduced by Mumford [18] in the spatial case and interpreted as model of connectivity in Lie groups in several works such as [2, 20, 24]. The cortical connectivity can be modelled as the probability of connecting two

points in the cortex and is represented by the stochastic counterpart of the curves in Eq. (2). If we assume a deterministic component in direction X_1 (which describes the long range connectivity) and stochastic component along X_2 (the direction of intracolumnar connectivity), the equation can be written as follows:

$$(x'_1, x'_2, \theta') = \mathbf{X}_1 + N(0, \sigma^2) \mathbf{X}_2 \quad (3)$$

where $N(0, \sigma^2)$ is a normally distributed variable with zero mean and variance equal to σ^2 . This process, first described in [18], is discussed in [1, 2, 24, 30]. In particular, in [1, 2] August and Zucker introduced the Markov process that results from the Brownian motion with randomly curved paths. We denote v the probability density to find a particle at the point (x_1, x_2) considering that it started from a given location (x'_1, x'_2) and that it is moving with some known velocity. This probability density satisfies a deterministic equation known in literature as the Kolmogorov forward equation or Fokker Planck equation:

$$\partial_t v = X_1 v + \sigma^2 X_{22} v \quad (4)$$

where X_1 is the directional derivative $\cos(\theta)\partial_x + \sin(\theta)\partial_y$, $X_2 = \partial_\theta$ and $X_{22} = \partial_{\theta\theta}$ is the second order derivative. This equation has been largely used in different fields [1, 2, 7, 30]. In [24] its stationary counterpart was proposed to model the probability of co-occurrence of contours in natural images:

$$FP = X_1 + \sigma^2 X_{22}. \quad (5)$$

The fundamental solution of the Fokker Planck equation Γ_1 is:

$$X_1 \Gamma_1((x_1, x_2, \theta), (x'_1, x'_2, \theta')) + \sigma^2 X_{22} \Gamma_1((x_1, x_2, \theta), (x'_1, x'_2, \theta')) = \delta(x_1, x_2, \theta) \quad (6)$$

which is not symmetric. The connectivity kernel ω_1 obtained by symmetrization of the Fokker Planck fundamental solution is:

$$\omega_1((x_1, x_2, \theta), (x'_1, x'_2, \theta')) = \frac{1}{2} (\Gamma_1((x_1, x_2, \theta), (x'_1, x'_2, \theta')) + \Gamma_1((x'_1, x'_2, \theta'), (x_1, x_2, \theta))). \quad (7)$$

This natural operator in the differential structure of the group SE(2) is in good agreement with the connectivity found by Bosking in [4] and is interpreted as the natural propagation of the signal in the geometrical structure of the cortex (see [22]).

Let us explicitly note that stochastic propagation along the geometrical structure defined by the vector fields in Eq. (1) can be modelled by different stochastic differential equations. If we assume that intracolumnar and long range connections have comparable strength, we have to modify Eq. (3) as follows:

$$(x'_1, x'_2, \theta') = N(0, \sigma_1^2) \mathbf{X}_1 + N(0, \sigma_2^2) \mathbf{X}_2 \quad (8)$$

where $N(0, \sigma_i^2)$ are normally distributed variables with zero mean and variance equal to σ_i^2 . Since the coefficient of propagation in both directions X_1 and X_2 are stochastic, the speed of propagation in both directions is comparable. The associated probability density is the fundamental solution of the Sub-Riemannian Heat equation [3]. So the associated time independent Fokker Planck operator reduces to the Sub-Riemannian Laplacian (SRL):

$$SRL = \sigma_1^2 X_{11} + \sigma_2^2 X_{22}. \quad (9)$$

We will denote Γ_2 the fundamental solution of this operator:

$$\sigma_1^2 X_{11} \Gamma_2((x_1, x_2, \theta), (x'_1, x'_2, \theta')) + \sigma_2^2 X_{22} \Gamma_2((x_1, x_2, \theta), (x'_1, x'_2, \theta')) = \delta(x_1, x_2, \theta) \quad (10)$$

We explicitly note that the kernel Γ_2 is symmetric, by construction, since we gave the same role to the derivatives X_1 and X_2 . Hence we do not need to symmetrize it and we will use the fundamental solution as a model of the connectivity kernel:

$$\omega_2((x_1, x_2, \theta), (x'_1, x'_2, \theta')) = \Gamma_2((x_1, x_2, \theta), (x'_1, x'_2, \theta')). \quad (11)$$

An isosurface of the connectivity kernel ω_1 obtained by symmetrization of the Fokker Planck fundamental solution Eq. (7) is shown in Fig. 3a and the distribution of co-occurrence of edges in natural images (from [23]) in Fig. 3b.

Isosurface of the connectivity kernel ω_2 obtained from the fundamental solution Γ_2 of the Sub-Riemannian Laplacian equation (11) is visualized in Fig. 3c.

Starting from the connectivity kernels defined previously, it is possible to extract perceptual units from images by means of spectral analysis of suitable affinity matrices. In particular in [25] it has been shown that solutions of mean field equations of neural population are able to provide a spectral analysis of the connectivity network excited by the visual input. This corresponds precisely to the spectral analysis of the affinity matrix, that contains the connectivity between maximally firing cells. Following this model we will interpret eigenvectors of the affinity matrix as perceptual units where the salient objects in the scene will correspond to the eigenvectors with the largest eigenvalues. Its first eigenvector can be used as indicator vector for grouping problems, as shown in [19]. Since we have defined two different kernels different affinity matrices will be defined, however all kernels are real and symmetric, so that the affinity matrix is real and symmetric:

$$A_{ij} = \omega((x_i, y_i, \theta_i), (x_j, y_j, \theta_j)) \quad (12)$$

The eigenvectors of the affinity matrix are interpreted as perceptual units [25] and the eigenvectors with the highest eigenvalues are linked to the most salient objects in the scene. In this way we will assign to the eigenvalues of the affinity matrix the meaning of a saliency index of the objects.

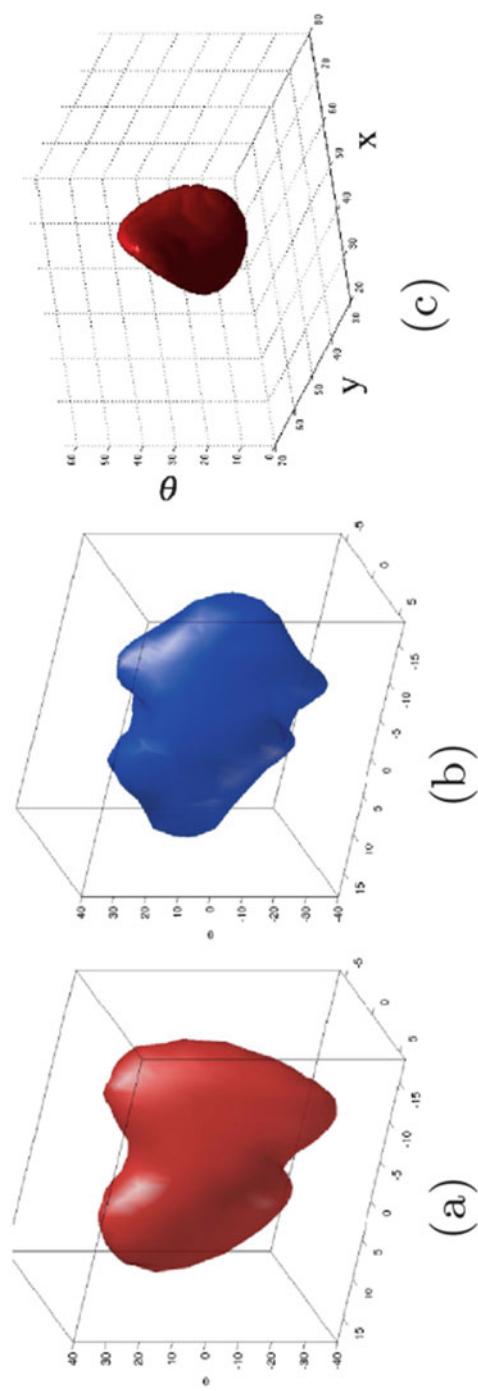


Fig. 3 (a) Isosurface of the connectivity kernel ω_1 , (b) the distribution of co-occurrence of edges in natural images [23] (c) isosurface of the connectivity kernel ω_2

2.3 Spectral Clustering and Perceptual Units

In [25] it has been shown that solutions of mean field equations of neural population are able to provide a spectral analysis of the connectivity network excited by the visual input. This corresponds precisely to the spectral analysis of the affinity matrix, that contains the connectivity between maximally firing cells. The structure of the affinity matrix is composed by blocks and the principal ones corresponds to coherent objects and its first eigenvector will have the meaning of emergent perceptual unit. To find the remaining objects in the image, the process is then repeated, the second and the following eigenvectors can be found, until the associated eigenvalue is sufficiently small. In this way only n eigenvectors are selected, with $n < N$, this procedure reduces the dimensionality of the description.

This procedure neurally reinterprets the process introduced by Perona and Freeman [19].

3 Numerical Simulations and Results

3.1 Numerical Approximations of the Kernels

In this section we numerically approximate the connectivity kernel ω_1 , defined by Eq. (7). It is possible to obtain the discrete fundamental solution Γ_1 of Eq. (6) by developing random paths from the numerical solutions of the system (3), that can be approximated by:

$$\begin{cases} x_{s+\Delta s} - x_s = \Delta s \cos(\theta) \\ y_{s+\Delta s} - y_s = \Delta s \sin(\theta), \quad s \in 0, \dots, H \\ \theta_{s+\Delta \theta} - \theta_s = \Delta s N(0, \sigma_1) \end{cases} \quad (13)$$

where H is the number of steps of the random path and $N(0, \sigma_1)$ is a generator of numbers taken from a normal distribution with mean 0 and variance σ_1 . In that way, the kernel is numerically estimated with Markov Chain Monte Carlo methods (MCMC) [21]. Various realizations n of the stochastic path will be given solving this finite difference equation n times; the estimated kernel is obtained averaging their passages over discrete volume elements. In particular, we first fix a discretization step $\Delta s = 1$ without loss of generality; then we simulate n several discrete-times random paths, assigning a value between 0 and 1 corresponding to the number of paths that passed through it, divided by n .

This provides a distribution over the cells that, for a large value of n , gives a discrete approximation of the connectivity kernel [6]. The parameters used in the system (13) are: 1,000,000 random paths with $\sigma_1 = 0.15$. Defined d_{max} the

maximum distance between the inducers of the stimulus, the value of H is defined as: $H = \frac{1}{3}d_{max}$.

This is described in detail in [13, 25], where it is also shown that the kernel ω_1 allows to estimate in a good way the cortical connectivity. Proceeding with the same methodology the numerical evaluation of fundamental solution of the sub-Riemannian Laplacian (Eq. (9)) is obtained and the system (8) discretized:

$$\begin{cases} x_{s+\Delta s} - x_s = \Delta s N(0, \sigma_2) \cos(\theta) \\ y_{s+\Delta s} - y_s = \Delta s N(0, \sigma_2) \sin(\theta), \quad s \in 0, \dots, H \\ \theta_{s+\Delta \theta} - \theta_s = \Delta s N(0, \sigma_3) \end{cases} \quad (14)$$

where $\sigma_2 = 1.2$ and $\sigma_3 = 0.11$ is the variance in the θ direction.

These kernels will be used for the individuation of perceptual units in the image.

3.2 Emergence of Percepts

In the following experiments some numerical simulations will be performed for the detection of perceptual units in an image, using the following algorithm:

1. Define the affinity matrix $A_{i,j}$ from the connectivity kernel;
2. Solve the eigenvalue problem $A_{i,j}u_i = \lambda_i u_i$, where λ_i are in decreasing order;
3. Find and project on the segments the eigenvector u_1 associated to the largest eigenvalue.

We first consider the stimulus of Fig. 4 (left). The results of simulations with the fundamental solutions of Fokker Planck and sub-Riemannian Laplacian equations are shown. The first eigenvector is visualized in red and it corresponds to the inducers of the Kanizsa triangle in Fig. 4 (middle). In this example, after the computation of the first eigenvector of the affinity matrix, this matrix is updated removing the identified perceptual unit and then the first eigenvector of the new

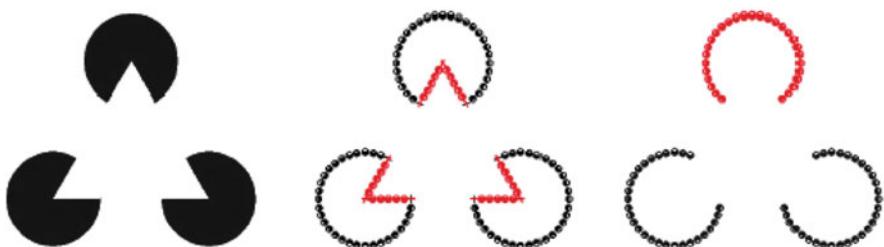


Fig. 4 From left to right: the Kanizsa triangle, the first eigenvector of the affinity matrix and the first eigenvector of the updated affinity matrix

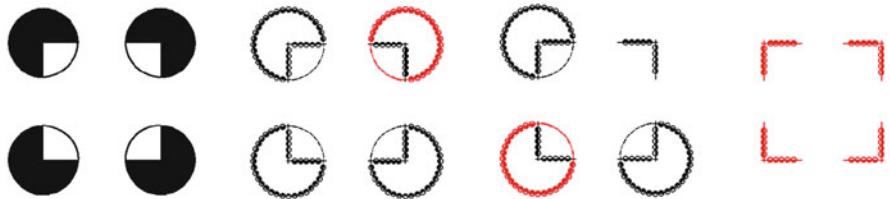


Fig. 5 From left to right: a Kanizsa square (first column) and the first eigenvector in red (second column). The affinity matrix is updated removing the segmented unit and the first eigenvector of the updated affinity matrix is shown (third column). The procedure is iterated until the last perceptual unit is segmented (last column)



Fig. 6 A stimulus containing a perceptual unit is segmented with Fokker Planck and Sub-Riemannian Laplacian equations, using the first eigenvector of the affinity matrix

matrix is computed, see Fig. 4 (right): these simulations show that circles are associated to the less salient eigenvectors. We choose here to show just one inducer in red, the other two have the same eigenvalue. In that way, the first eigenvalue can be considered as a quantitative measure of saliency, because it allows to segment the most important object in the scene and the results of simulations confirm the visual grouping.

Then we consider the stimulus of Fig. 5 (first column), where it is not possible to perceive a square and the figures become the circles. These behaviours are well represented in the simulations. We compute the connectivity kernels and the associated affinity matrix, in order to find its eigenvalues. We obtain 4 eigenvectors with almost the same eigenvalues, that represent the circles. Iterating the method, we obtain that the square is associated to the less salient eigenvector Fig. 5 (fourth column) and so is less salient than the circles.

Finally, we consider the experiment of Fig. 6 (left), similar to the ones of Field et al. [12], where a subset of elements organized in a coherent way is present out of a ground formed by a random distribution of elements. The stimulus and the results of simulations are represented with oriented segments.

The connectivity among these elements is defined as in Eqs. (6), (10). After computing the eigenvalues of the affinity matrix, the eigenvector corresponding to the highest eigenvalue is individuated and visualized in red in Fig. 6 (right).

The result shows that the stimulus is well segmented with the fundamental solutions of Fokker Planck and Sub-Riemannian Laplacian equations.

4 Conclusions

In this work we have presented a neurally based model for figure-ground segmentation using spectral methods, where segmentation has been performed by computing eigenvectors of affinity matrices.

Different connectivity kernels that are compatible with the functional architecture of the primary visual cortex have been used; we have modelled them as fundamental solution of Fokker-Planck and Sub-Riemannian Laplacian equations. We have also shown that the fundamental solutions of Fokker-Planck and Sub-Riemannian Laplacian equations are good models for the good continuation law.

With this model we have identified perceptual units of different figures, showing that this can be considered a good quantitative model for the constitution of perceptual units.

Acknowledgements The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n607643.

References

1. August, J., Zucker, S.W.: The curve indicator random field: curve organization via edge correlation. In: Perceptual Organization for Artificial Vision Systems, pp 265–288. Springer, Berlin (2000)
2. August, J., Zucker, S.W.: Sketches with curvature: the curve indicator random field and markov processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(4), 387–400 (2003)
3. Boscain, U., Duplaix, J., Gauthier, J.P., Rossi, F.: Anthropomorphic image reconstruction via hypoelliptic diffusion. *SIAM J. Control. Optim.* **50**(3), 1309–1336 (2012)
4. Bosking, W.H., Zhang, Y., Schofield, B., Fitzpatrick, D.: Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* **17**(6), 2112–2127 (1997)
5. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. *J. Math. Imaging Vision* **24**(3), 307–326 (2006)
6. Cocci, G., Barbieri, D., Citti, G., Sarti, A.: Cortical spatiotemporal dimensionality reduction for visual grouping. *Neural Comput.* **27**, 1252–1293 (2015)
7. Duits, R., Franken, E.: Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores—Part I: linear left-invariant diffusion equations on $SE(2)$. *Q. Appl. Math.* **68**(2), 255–292 (2010)
8. Duits, R., Franken, E.: Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores. Part II: nonlinear left-invariant diffusions on invertible orientation scores. *Q. Appl. Math.* **68**(2), 293–331 (2010)

9. Duits, R., Felsberg, M., Granlund, G., ter Haar Romeny, B.: Image analysis and reconstruction using a wavelet transform constructed from a reducible representation of the Euclidean motion group. *Int. J. Comput. Vis.* **72**(1), 79–102 (2007)
10. Duits, R., Duits, M., van Almsick, M., ter Haar Romeny, B.: Invertible orientation scores as an application of generalized wavelet theory. *Pattern Recognit. Image Anal.* **17**(1), 42–75 (2007)
11. Duits, R., Führ, H., Janssen, B., Bruurmijn, M., Florack, L., van Assen, H.: Evolution equations on gabor transforms and their applications. *Appl. Comput. Harmon. Anal.* **35**(3), 483–526 (2013)
12. Field, D.J., Hayes, A., Hess, R.F.: Contour integration by the human visual system: evidence for a local “association field”. *Vis. Res.* **33**(2), 173–193 (1993)
13. Higham, D.J.: An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* **43**(3), 525–546 (2001)
14. Hoffman, W.C.: The visual cortex is a contact bundle. *Appl. Math. Comput.* **32**(2), 137–167 (1989)
15. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *J. Physiol.* **160**(1), 106–154 (1962)
16. Hubel, D.H., Wiesel, T.N.: Ferrier lecture: functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **198**(1130), 1–59 (1977)
17. Koenderink, J.J., van Doorn, A.J.: Representation of local geometry in the visual system. *Biol. Cybern.* **55**(6), 367–375 (1987)
18. Mumford, D.: *Elastica and Computer Vision*. Springer, Berlin (1994)
19. Perona, P., Freeman, W.: A factorization approach to grouping. In: European Conference on Computer Vision, pp. 655–670. Springer, Berlin (1998)
20. Petitot, J., Tondut, Y.: Vers une neurogéométrie. fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques informatique et sciences humaines* **145**, 5–102 (1999)
21. Robert, C., Casella, G.: *Monte Carlo Statistical Methods*. Springer Science & Business Media, Berlin (2013)
22. Sanguinetti, G.: Invariant models of vision between phenomenology, image statistics and neurosciences. PhD Thesis. Universidad de la Republica, Montevideo, Uruguay. <https://www.colibri.udelar.edu.uy/bitstream/123456789/2902/1/San11.pdf> (2011)
23. Sanguinetti, G., Citti, G., Sarti, A.: Image completion using a diffusion driven mean curvature flow in a sub-Riemannian space. In: International Conference on Computer Vision Theory and Applications (VISAPP’08), pp. 22–25 (2008)
24. Sanguinetti, G., Citti, G., Sarti, A.: A model of natural image edge co-occurrence in the rototranslation group. *J. Vis.* **10**(14), 37 (2010)
25. Sarti, A., Citti, G.: The constitution of visual perceptual units in the functional architecture of v1. *J. Comput. Neurosci.* **38**(2), 285–300 (2015)
26. Sarti, A., Citti, G., Manfredini, M.: From neural oscillations to variational problems in the visual cortex. *J. Physiol. Paris* **97**(2), 379–385 (2003)
27. Sarti, A., Citti, G., Petitot, J.: The symplectic structure of the primary visual cortex. *Biol. Cybern.* **98**(1), 33–48 (2008)
28. Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A., Singh, M., von der Heydt, R.: A century of gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol. Bull.* **138**(6), 1172 (2012)
29. Wertheimer, M.: Laws of organization in perceptual forms. First published as *Untersuchungen zur Lehre von der Gestalt II. Psychologische Forschung* **4**, 301–350 (1923). Translation published in Ellis, W. (ed.) *A Source Book of Gestalt Psychology*, pp. 71–88. Routledge & Kegan Paul, London (1938)
30. Williams, L.R., Jacobs, D.W.: Stochastic completion fields: a neural model of illusory contour shape and salience. *Neural Comput.* **9**(4), 837–858 (1997)
31. Zucker, S.: Differential geometry from the Frenet point of view: boundary detection, stereo, texture and color. In: *Handbook of Mathematical Models in Computer Vision*, pp. 357–373. Springer, Berlin (2006)

Mathematical Models of Visual Perception for the Analysis of Geometrical Optical Illusions



Benedetta Franceschiello, Alessandro Sarti, and Giovanna Citti

Abstract In this chapter a cortical-based mathematical models for Geometrical Optical illusions (GOIs) is provided. GOIs have been of great interest due to the possibility to understand, through the effect they produce on neural connections, the behaviour of low-level visual processing. They have been defined in the XIX century by German psychologists (Oppel, Ueber geometrisch-optische Täuschungen. Jahresbericht des physikalischen Vereins zu Frankfurt am Main, pp. 37–47, 1854–1855; Hering, Beiträge zur Physiologie, von Wilhelm Engelmann, Leipzig, pp. 1–5, 1861) in terms of phenomenology of vision, as situations in which there is an awareness of a mismatch of geometrical properties between an item in object space and its associated percept (Westheimer, Vis. Res. 48(20):2128–2142, 2008). As pointed out by Eagleman (2001) the study of these systematic misperceptions combined with recent techniques for brain's activity recording provides a brilliant insight to lead new experiments on receptive fields of V1, as well as new hypothesis about the behaviour of perception. In this framework, starting from the geometrical model for the primary visual cortex introduced by Citti and Sarti in 2003, we provide an efficient mathematical model which allows to interpret these phenomena and to measure the perceived misperception based on the simulated response of simple cells in V1. The model involves image-processing techniques and allows to recover the perceived displacement by means of partial differential equations.

B. Franceschiello (✉) · A. Sarti

Center of Mathematics, CAMS, 54 Boulevard de Raspail, 75006 Paris, France
e-mail: benedetta.franceschiello@ehess.fr; alessandro.sarti@ehess.fr

G. Citti

Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy
e-mail: giovanna.citti@unibo.it

1 Introduction

Geometrical optical illusions (GOIs, see Fig. 1), discovered in the nineteenth century by German psychologists (Oppel 1854 [21], Hering, 1878, [11]) have been defined as situations in which there is an awareness of a mismatch of geometrical properties between an item in object space and its associated percept [30]. Previous mathematical models for these phenomena are presented in [7, 8, 12, 26, 26, 28], while taxonomies have been proposed in Coren e Girgus, 1978, [5]; Robinson, 1998, [23]; Wade, 1982, [27]. Here a neuro-mathematical model will be presented: it allows us to interpret at a neural level the origin of GOIs and to reproduce the arised percept for this class of phenomena. The main idea is to adopt the model of the functional geometry of V1 provided in [3] and to consider that the image stimulus will modulate the connectivity of V1 cells. When projected onto the visual space, the modulated connectivity gives rise to a Riemannian metric which is at the origin of the visual space deformation. The displacement vector field at every point of the stimulus is mathematically computed by solving a Poisson problem and the perceived image is finally reproduced see Fig. 2. The considered phenomena consist, as shown in Fig. 1, in straight lines/geometric shapes (square) over different backgrounds (in these examples a radial background).

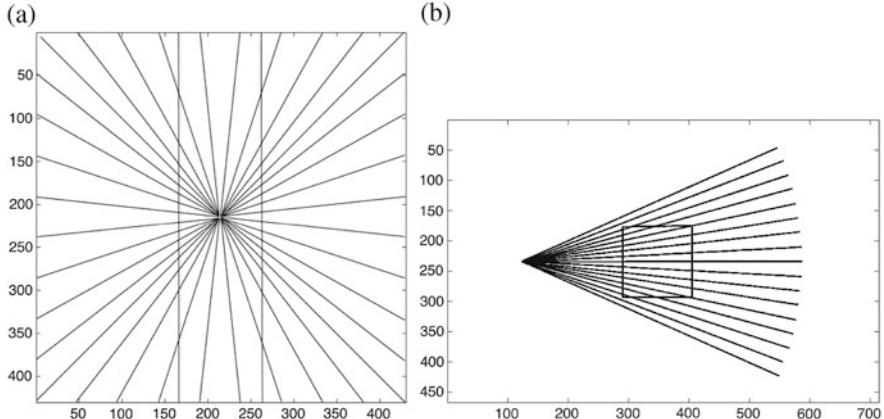


Fig. 1 Examples of geometrical optical illusions: (a) the Hering illusion and (b) Ehrenstein illusion

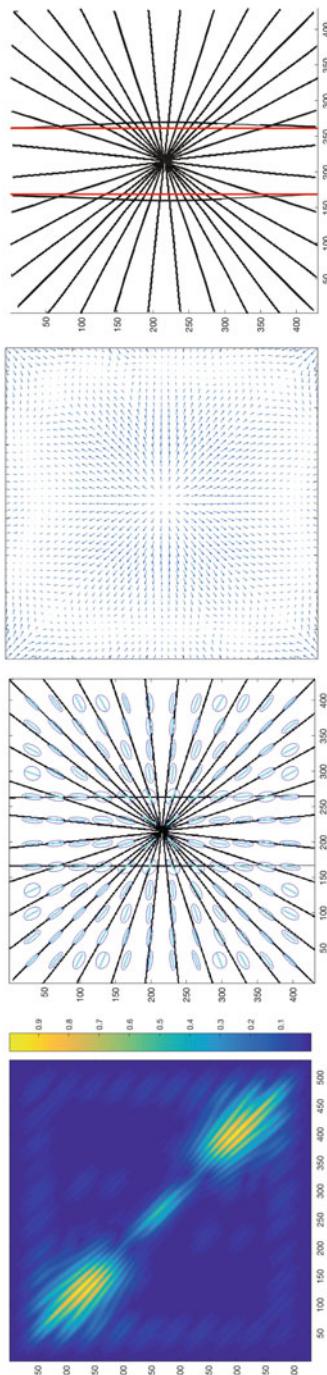


Fig. 2 From left to right: (1) the visual stimulus in Fig. 1a activates the retina whose response is described in Sects. 2.1–2.2. (2) The cortical connectivity described in Sects. 2.3–2.4 polarized by the response of simple cells induces a metric tensor on the image plane \mathbb{R}^2 , Sect. 3.1. (3) The previous metric is responsible for the perceived deformation, modelled in Sect. 3.2.1, whose displacement vector fields are represented in this image. (4) The recovered vector fields, as explained in Sect. 3.2.2, are applied to the initial stimulus to represent the perceived image

2 The Mathematical Model: Neurogeometry of the Primary Visual Cortex

2.1 The Set of Simple Cells Receptive Profiles

The primary visual cortex is the first part of the brain processing the visual signal coming from the retina. The receptive field (RF) of a cortical neuron is the portion of the retina which the neuron reacts to, and the receptive profile (RP) $\psi(\xi)$ is the function that models the activation of a cortical neuron when a stimulus is applied to a point $\xi = (\xi_1, \xi_2)$ of the retinal plane. Simple cells of visual cortices V1 and V2 are sensitive to position and orientation. The orientation selectivity, the spatial and temporal frequency of cells in V2 differs little from the one in V1 [18]. Receptive fields in V2 are larger from those in V1 [16, 18]. Considering a basic geometric model, the set of simple cells RPs can be obtained via translations of vector $x = (x_1, x_2)$ and rotation of angle θ from a unique mother profile $\psi_0(\xi)$. Receptive profiles have been modelled as oriented filters first by Daugman (1985) [6], Jones and Palmer (1987) [14]: they showed that Gabor filters were a good approximation for receptive profiles of simple cells in V1/V2. In our contribution both odd and even parts will be considered, see Fig. 3. A good expression for the mother Gabor filter is:

$$\psi_0(\xi) = \psi_0(\xi_1, \xi_2) = \frac{1}{4\pi\sigma^2} e^{\frac{-(\xi_1^2 + \xi_2^2)}{2\sigma^2}} e^{\frac{2i\bar{b}\xi_2}{\sigma}}, \quad (1)$$

where $\bar{b} = 0.56$ is the ratio between σ and the spatial wavelength of the cosine factor. Translations and rotations can be expressed as:

$$A_{(x_1, x_2, \theta)}(\xi) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}. \quad (2)$$

Hence a general RP can be expressed as:

$$\psi_{(x_1, x_2, \theta)}(\xi_1, \xi_2) = \psi_0(A_{(x_1, x_2, \theta)}^{-1}(\xi_1, \xi_2)).$$

2.2 Output of Receptive Profiles

When a visual stimulus I of intensity $I(x_1, x_2) : M \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+$, activates the retina (\mathbb{R}^2), the neurons whose RFs intersect M spike and their spike frequencies $O(x_1, x_2, \theta)$ can be modeled as:

$$O(x_1, x_2, \theta) = \int_M I(\xi_1, \xi_2) \psi_{(x_1, x_2, \theta)}(\xi_1, \xi_2) d\xi_1 d\xi_2. \quad (3)$$

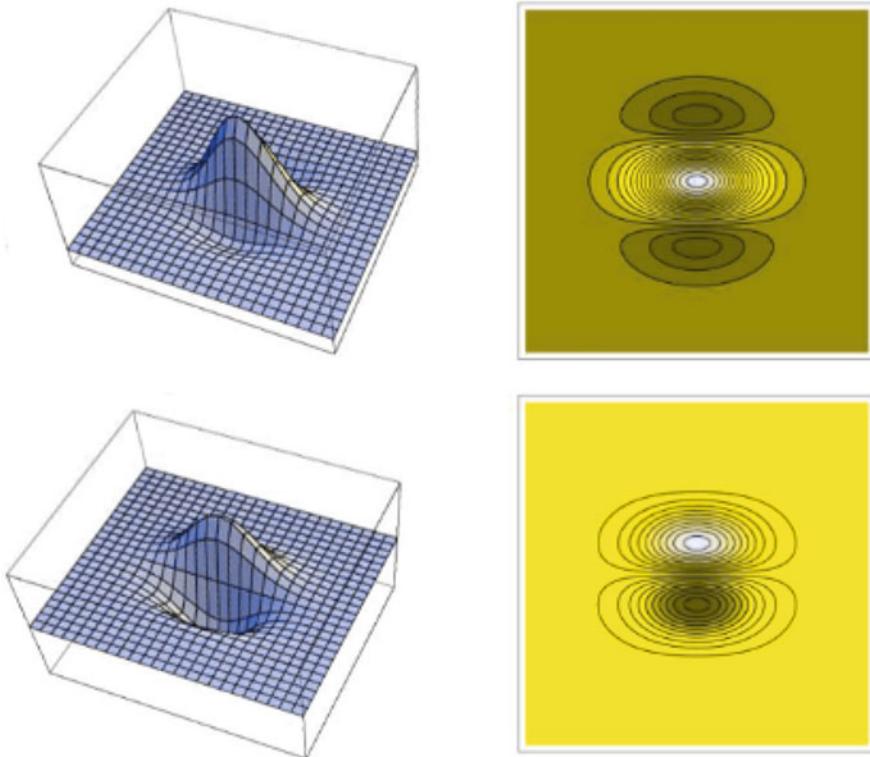


Fig. 3 Even ($e^{-(\xi_1^2 + \xi_2^2)} \cos(2\xi_2)$) and odd part ($e^{-(\xi_1^2 + \xi_2^2)} \sin(2\xi_2)$) of ψ_0

$I \in L^1(\mathbb{R}^2)$. In the right hand side of the equation the integral of the signal with the real and imaginary part of the Gabor filter is expressed. We will assign to the norm of the output the usual meaning of energy:

$$E(x_1, x_2, \theta) = \|O(x_1, x_2, \theta)\|, \quad (4)$$

where $\|\cdot\|$ denotes the complex modulus of the output O in Eq. (3), evaluated at the fixed value of σ (Fig. 4).

2.3 Hypercolumnar Structure

The hypercolumnar structure, discovered by the neuro-physiologists Hubel and Wiesel in the 1960s [13], organizes the cells of V1/V2 in columns (called hypercolumns) covering a small part of the visual field $M \subset \mathbb{R}^2$ and corresponding to parameters such as orientation, scale, direction of movement, color, for a fixed retinal position (x_1, x_2) .

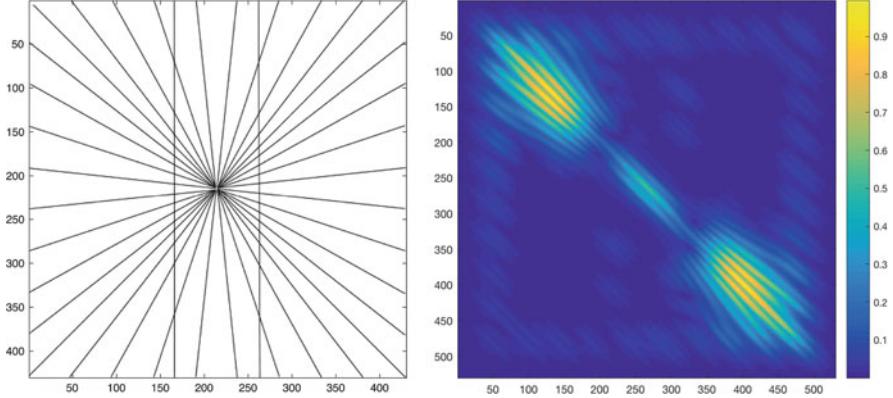


Fig. 4 Left: initial stimulus. Right: energy response for $\theta = \pi/4$

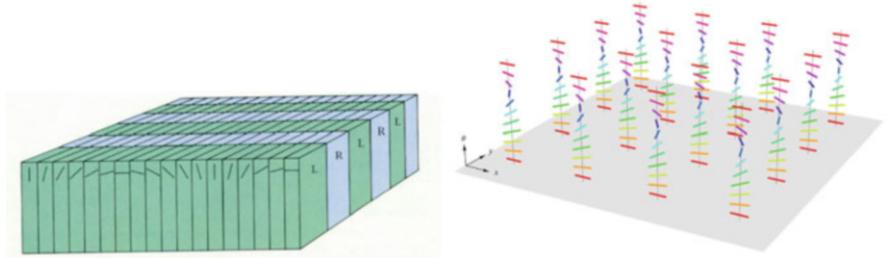


Fig. 5 Top: representation of hypercolumnar structure, for the orientation parameter [22]. Bottom: for each position of the retina (x_1, x_2) we have the set of all possible orientations

Over each retinal point we will consider a whole hypercolumn of cells, each one sensitive to a specific instance of orientation (Fig. 5). Hence for each position (x_1, x_2) of the retina $M \subset \mathbb{R}^2$ we associate a whole set of filters $RP_{(x_1, x_2)} = \{\psi_{(x_1, x_2, \theta)} : \theta \in S^1\}$. This expression associates to each point of the proximal stimulus in \mathbb{R}^2 all possible feature orientations into the space of features S^1 , and defines a fiber over each point $\{\theta \in S^1\}$.

2.4 Cortical Connectivity

The orientation selectivity between simple cells belonging to the same hypercolumn selects the most probable response from the energy of receptive profiles. Horizontal connections are long ranged and connect cells of approximately *the same orientation*. Since the connectivity between cells is defined on the tangent bundle, in [3] the whole space of features (x_1, x_2, θ) is described in terms of a 3-dimensional fiber

bundle, whose generators are X_1, X_3 for the base:

$$X_1 = \cos \theta \frac{\partial}{\partial x_1} + \sin \theta \frac{\partial}{\partial x_2}, \quad X_3 = -\sin \theta \frac{\partial}{\partial x_1} + \cos \theta \frac{\partial}{\partial x_2}.$$

and $X_2 = \frac{\partial}{\partial \theta}$ for the fiber. These vector fields generate the tangent bundle of $\mathbb{R}^2 \times S^1$. Since horizontal connectivity is very anisotropic, the three generators are weighted by a strongly anisotropic metric. Citti and Sarti in [3] proposed to endow the $\mathbb{R}^2 \times S^1$ group to model the long range connectivity of the primary visual cortex V1 with a sub-Riemannian metric g_{ij} . The inverse (responsible for the connectivity in the cortex) is:

$$g^{ij}(x_1, x_2, \theta) = \begin{pmatrix} \cos^2 \theta & \sin \theta \cos \theta & 0 \\ \sin \theta \cos \theta & \sin^2 \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (5)$$

with $i, j = 1, 2, 3$. Cortical curves in V1 will be a linear combination of vector fields X_1 and X_2 , the generators of the 2-dimensional horizontal space, while they will have a vanishing component along $X_3 = -[X_1, X_2]$. In the present paper we consider the local polarization of the horizontal connectivity, well estimated by the energy in Eq. (4).

3 The Neuro-Mathematical Model for GOIs

3.1 Output of Simple Cells and Connectivity Metric

We will now define a connectivity metric tensor on the image plane \mathbb{R}^2 starting from the connectivity metric $(g^{ij})_{i,j=1,2,3}$ defined in Eq. (5) and the output of simple cells. It will be a modified structure tensor [1, 2, 9, 29] which combines odd and even Gabor filters because we look for a technique able to correctly measure θ for contours and lines. The proposed tensor is biologically based, since it encodes the action of Gabor filters and the cortical connectivity. We consider simple cells at fixed value of σ , depending on position and orientation. For each point $(x_1, x_2, \theta) \in \mathbb{R}^2 \times S^1$, we restrict the connectivity tensor $(g^{ij}(x_1, x_2, \theta))_{i,j=1,2,3}$ to the \mathbb{R}^2 plane generated by $\{\partial_{x_1}, \partial_{x_2}\}$, subset of the tangent space to $\mathbb{R}^2 \times S^1$ at the point (x_1, x_2, θ) , and obtain the tensor:

$$h^{ij}(x_1, x_2, \theta) = \begin{pmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{pmatrix}. \quad (6)$$

Each point of the hypercolumn is weighted by the energy of simple cells normalized over the whole set of hypercolumn responses. The connectivity tensor restricted to the \mathbb{R}^2 plane and modulated by the output of simple cells will become:

$$\frac{E(x_1, x_2, \theta)}{\int_0^\pi E(x_1, x_2, \theta) d\theta} h^{ij}(x_1, x_2, \theta) \quad (7)$$

This last expression corresponds to a connectivity polarized by the normalized energy (4) of simple cells at points (x_1, x_2, θ) . The overall inverse of the metric tensor arising from the action within the hypercolumn over each retinal point (x_1, x_2) is obtained summing up along θ the previous modulated metric in Eq.(7):

$$\mathbf{p}^{-1}(x_1, x_2) = \eta^{-1} \frac{\int_0^\pi E(x_1, x_2, \theta) h^{ij}(x_1, x_2, \theta) d\theta}{\int_0^\pi E(x_1, x_2, \theta) d\theta}, \quad (8)$$

where η^{-1} is a normalization constant obtained as the L^∞ norm of the inverse of the metric determinant which appears in Eq.(8). This tensor will have principal eigenvector along the orientation $\bar{\theta}$ (cyan vector), corresponding to the maximum energy within the hypercolumn. A visualization of \mathbf{p}^{-1} is given in Fig.6b where principal and second eigenvectors correspond to first and second semi-axes of the ellipses, while lengths correspond to the magnitude of eigenvalues. Along parts of the stimulus strongly oriented, ellipses are elongated. As

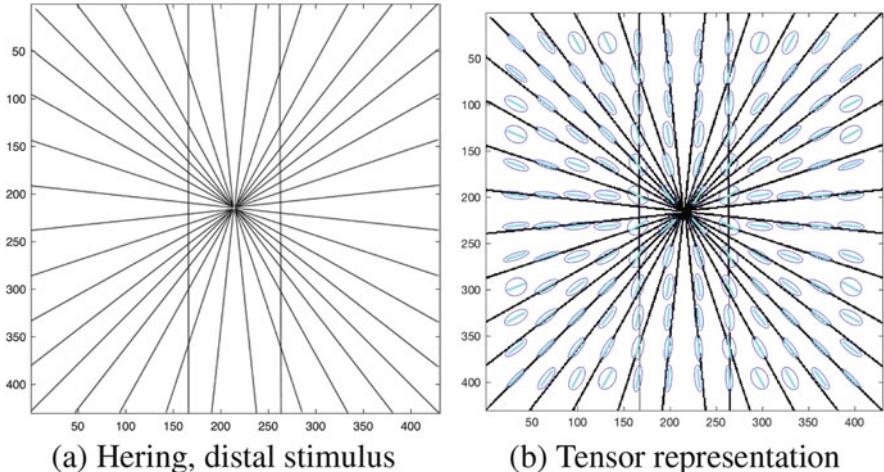


Fig. 6 (a) Proximal stimulus (Hering illusion). (b) Representation of \mathbf{p}^{-1} (blue)

far as we move toward crossing points, ellipses lost their elongated form and become rounded, in dependence with the orientation response of lines at crossing points.

3.2 From Metric Tensor Field to Image Distortion

In the previous section we described the response of the cortex in the presence of a visual stimulus:

1. the distal stimulus [10, 17] is projected onto the cortex by means of activity of simple cells,
2. the joint action of the short and long range connectivity induces a Riemannian tensor \mathbf{p}^{-1} on the \mathbb{R}^2 retinal plane.

Finally a third mechanism takes place, able to construct the perceived stimulus from the cortical activation. With this mechanism the image distortion which induces the metric tensor \mathbf{p} (inverse of \mathbf{p}^{-1}) is estimated. Here we propose to apply infinitesimal strain theory (see [19, 20]) and to identify its inverse \mathbf{p} with the strain tensor to compute the deformation. Once the displacement vector field is applied to the distal stimulus, we obtain a distorted image which models the proximal one. In this way we justify the mechanism at the basis of geometrical optical illusions.

3.2.1 Strain Tensor: Displacement Vector Field

We think at the deformation induced by a geometrical optical illusion as an isometry between the \mathbb{R}^2 plane equipped with the metric \mathbf{p} and the \mathbb{R}^2 plane with the Euclidean metric \mathbf{Id} , $\Phi : (\mathbb{R}^2, \mathbf{p}) \rightarrow (\mathbb{R}^2, \mathbf{Id})$. From the mathematical point of view this means that we look for the change of variable which induces the new metric, i.e.

$$\left(\frac{\partial \Phi^k}{\partial x_i} \right) Id_{kl} \left(\frac{\partial \Phi^l}{\partial x_j} \right) = p_{ij}(\mathbf{x}), \quad (9)$$

(see Jost [15]), where $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$, p_{ij} and Id_{kl} indicate the components of tensor \mathbf{p} and the identity \mathbf{Id} respectively and we use Einstein summation convention. Using this expression, we obtain:

$$\mathbf{p}(\mathbf{x}) = (\nabla \Phi)^T (\nabla \Phi). \quad (10)$$

Let us notice \mathbf{p}^{-1} corresponds to Φ^{-1} , the map representing the process which builds the modulated connectivity we discussed before. In strain theory \mathbf{p} satisfying Eq. (10) is called *right Cauchy-Green tensor* associated to the deformation Φ , which from the physical point of view is a map $\Phi : \bar{\Omega} \rightarrow \mathbb{R}^2$ associating the points

of the closure of a bounded open set $\Omega \subset \mathbb{R}^2$ (initial configuration of a body) to $\Phi(\Omega) \subset \mathbb{R}^2$ (deformed configuration), [19, 20]. It is possible to introduce the displacement as a map $\bar{u}(x_1, x_2) = \Phi(x_1, x_2) - (x_1, x_2)$, where $(x_1, x_2) \in \mathbb{R}^2$. It follows

$$\nabla \bar{u} = \nabla \Phi - \mathbf{Id}. \quad (11)$$

We can now express the right Cauchy-Green tensor in terms of displacement:

$$\begin{aligned} \mathbf{p} = p_{ij}(\mathbf{x}) &= (\nabla \Phi)^T (\nabla \Phi) = (\nabla \bar{u} + \mathbf{Id})^T (\nabla \bar{u} + \mathbf{Id}) \\ &= (\nabla \bar{u})^T (\nabla \bar{u}) + (\nabla \bar{u}) + (\nabla \bar{u})^T + \mathbf{Id}. \end{aligned} \quad (12)$$

The concept of strain is used to evaluate how much a given displacement differs locally from a rigid body displacement. The *linearized right Cauchy-Green tensor* has the following form:

$$G(\bar{u}) \approx (\nabla \bar{u}) + (\nabla \bar{u})^T. \quad (13)$$

Here we give the expression in components of $\epsilon(\bar{u}) = \frac{1}{2}G(\bar{u})$ (the so called Green-Lagrangian strain tensor):

$$\epsilon_{ij}(\bar{u}) = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{1}{2}\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right) \\ \frac{1}{2}\left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\right) & \frac{\partial u_2}{\partial x_2} \end{pmatrix}, \quad (14)$$

where $\bar{u} = (u_1, u_2)$. Expressing ϵ_{ij} in terms of the metric $(p_{ij})_{i,j}$ with whom the initial configuration of the considered body was equipped we obtain:

$$\epsilon_{ij}(\bar{u}) \approx \frac{1}{2}((p_{ij})_{ij} - \mathbf{Id}), \quad (15)$$

and in its matrix form:

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{1}{2}\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right) \\ \frac{1}{2}\left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\right) & \frac{\partial u_2}{\partial x_2} \end{pmatrix}. \quad (16)$$

3.2.2 Poisson Problems: Displacement

Starting from Eq. (16) we obtain a system of equations with this form:

$$\begin{cases} p_{11} - 1 = & \frac{\partial u_1}{\partial x_1} \\ p_{22} - 1 = & \frac{\partial u_2}{\partial x_2} \\ p_{12} = p_{21} = & \frac{1}{2}\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right) \end{cases} \quad (17)$$

Differentiating, substituting and imposing Neumann boundary conditions to system (17) we end up with the following differential system:

$$\left\{ \begin{array}{ll} \Delta u_1 = \frac{\partial}{\partial x_1} p_{11} + 2 \frac{\partial}{\partial x_2} p_{12} - \frac{\partial}{\partial x_1} p_{22} & \text{in } M \\ \Delta u_2 = \frac{\partial}{\partial x_2} p_{22} + 2 \frac{\partial}{\partial x_1} p_{12} - \frac{\partial}{\partial x_2} p_{11} & \\ \frac{\partial}{\partial \mathbf{n}} u_1 = 0 & \text{in } \partial M \\ \frac{\partial}{\partial \mathbf{n}} u_2 = 0 & \end{array} \right. \quad (18)$$

where M is an open subset of \mathbb{R}^2 and ∂M is Lipschitz continuous, with normal defined almost everywhere. Solutions of Eq. (18) are well defined up to an additive constant, which is recovered imposing $u(0, 0) = v(0, 0) = 0$ for symmetry reasons, where $(0, 0)$ is the center of our initial domain M . Let us explicitly note that tensor \mathbf{p} is obtained after convolution of Gabor filters, so that it is differentiable, allowing to write the system. Hence we solve Eq. (18), recovering the displacement field $\bar{u}(x_1, x_2)$.

4 Numerical Simulations and Results

The inverse of tensor expressed in Eq. (8) is computed discretizing θ as a vector of 32 values equally spaced in the interval $[0, \pi]$. The parameter σ varies in dependence of the image resolution and is set in concordance with the stimulus processed. It is taken quite large in all examples in such a way to obtain a smooth tensor field covering all points of the image, in accordance with the experimental fact that these phenomena take place in V1/V2. η has been chosen for all the examples as $\eta = 2 \times 10^{-2}$. Equation (18) is approximated with a central finite difference scheme and it is solved with a classical PDE linear solver.

4.1 Perceived Deformation in GOIs

Hering illusion (see [11]) is presented in Fig. 7 (top left). Straight lines are presented in front of a radial background, so that the lines appears as if they were bowed outwards. We first convolve the distal stimulus with the entire bank of Gabor filters, $\sigma = 6.72$ pixels. Following the process, we compute \mathbf{p}^{-1} using Eq. (8), we solve Eq. (18) obtaining the perceived displacement $\bar{u} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. Once it is applied to the initial stimulus, the proximal stimulus is recovered. The result of computation is

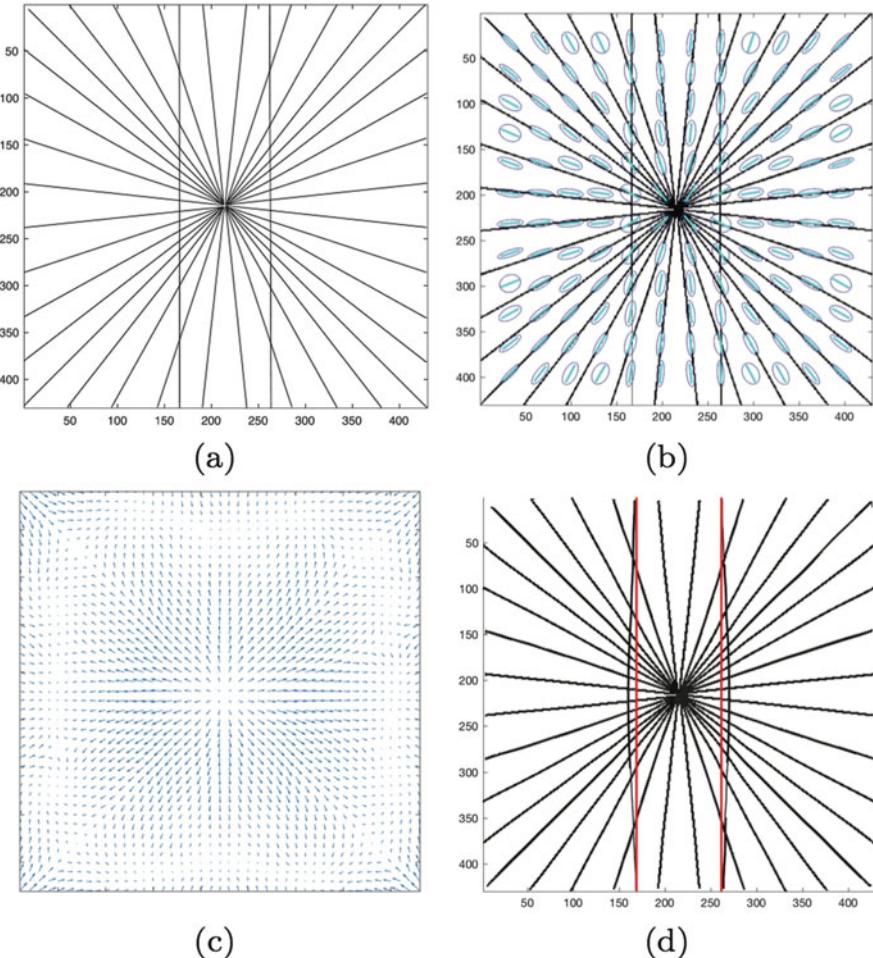


Fig. 7 (a) Hering illusion. (b) Representation of \mathbf{p}^{-1} , Eq. (7). (c) Computed displacement field $\bar{u} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. (d) Displacement applied to the image. In black we represent the proximal stimulus as displaced points of the distal stimulus: $(x_1, x_2) + \bar{u}(x_1, x_2)$. In red we give two straight lines as reference

shown in Fig. 7 (bottom right). The distorted image folds the straight lines (in black) against the straight lines (in red) of the original stimulus. In Fig. 8 a modified Hering illusion, rotated and cutted, is presented and processed through the same algorithm. Then in Fig. 9 we process Ehrenstein illusion, in which the top and bottom edges of the square appear to be unparallel due to the presence of the crossing lines in the background.

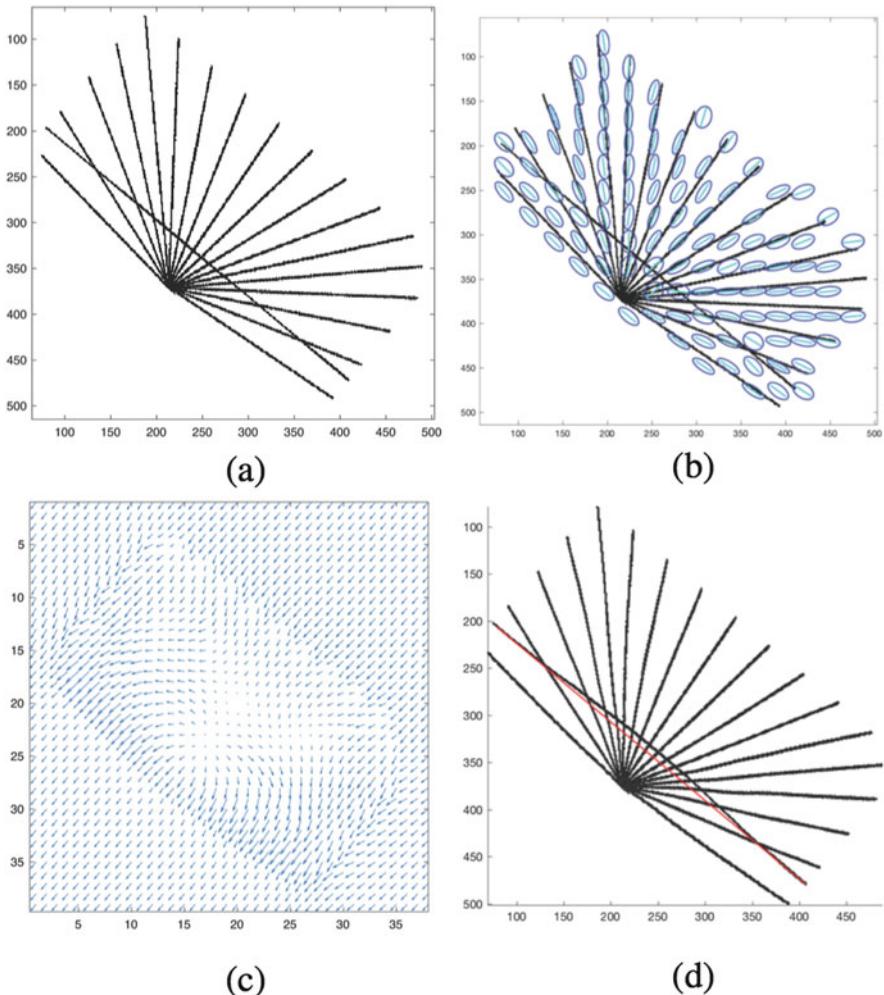


Fig. 8 Here we show the same processing as in Fig. 7 for a modified Hering illusion

5 Conclusion and Future Works

We presented a neuro-mathematical model based on the functional architecture of V1/V2 to explain and simulate perceptual distortion due to geometrical-optical illusions and to embed geometrical context. The perceptual distortion is due to the Riemannian metric induced on the image plane by the connectivity activated by the image stimulus. The presented results are qualitatively in good agreement with human perception. In the future this work could be extended to functional

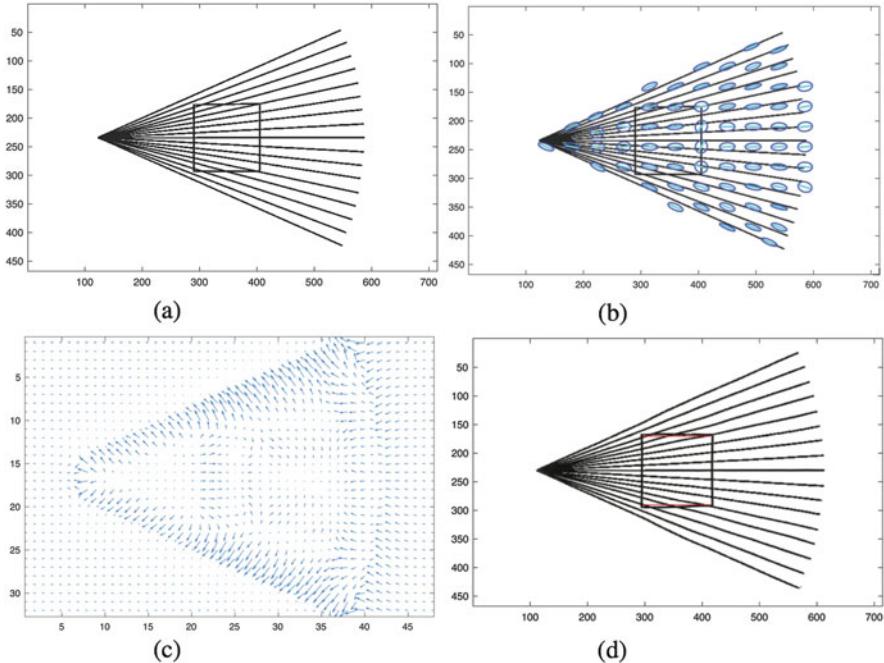


Fig. 9 Here we show the same processing as in Fig. 7 for Ehrenstein illusion

architectures involving the feature of scale, starting from models provided by Sarti et al. in [24, 25]. This will allow to provide a model for scale illusions, such as the Delboüf, see [4].

Acknowledgements The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Unions Seventh Framework Programme FP7/2007-2013/ under REA grant agreement n607643.

References

1. Bigun, J.: Optimal orientation detection of linear symmetry. In: Proceedings of the IEEE-First International Conference on Computer Vision, London, 8–11 June 1987, pp. 433–438. <http://www.diva-portal.org/smash/get/diva2:691493/FULLTEXT01.pdf>
2. Brox, T., Weickert, J., Burgeth, B., Mrázek, P.: Nonlinear structure tensors. *Image Vis. Comput.* **24**(1), 41–55 (2006)
3. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. *J. Math. Imaging Vision* **24**(3), 307–326 (2006)
4. Colman, A.M.: *A Dictionary of Psychology*. Oxford University Press, Oxford (2015)
5. Coren, S., Girkus, J.S.: *Seeing is Deceiving: The Psychology of Visual Illusions*. Hillsdale: Lawrence Erlbaum (1978)

6. Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* **2**(7), 1160–1169 (1985)
7. Ehm, W., Wackermann, J.: Modeling geometric-optical illusions: a variational approach. *J. Math. Psychol.* **56**(6), 404–416 (2012)
8. Fermüller, C., Malm, H.: Uncertainty in visual processes predicts geometrical optical illusions. *Vis. Res.* **44**(7), 727–749 (2004)
9. Förstner, W., Gülich, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: Proceedings of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, pp. 281–305 (1987)
10. Gibson, J.J.: The concept of the stimulus in psychology. *Am. Psychol.* **15**(11), 694 (1960)
11. Hering, H.E.: Beiträge zur Physiologie, pp. 1–5. von Wilhelm Engelmann, Leipzig (1861)
12. Hoffman, W.C.: Visual illusions of angle as an application of lie transformation groups. *SIAM Rev.* **13**(2), 169–184 (1971)
13. Hubel, D.H., Wiesel, T.N.: Ferrier lecture: functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. Ser. B* **198**(1130), 1–59 (1977)
14. Jones, J.P., Palmer, L.A.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58**(6), 1233–1258 (1987)
15. Jost, J.: Riemannian Geometry and Geometric Analysis. Springer Science & Business Media, Berlin (2008)
16. Kennedy, H., Martin, K., Orban, G., Whitteridge, D.: Receptive field properties of neurones in visual area 1 and visual area 2 in the baboon. *Neuroscience* **14**(2), 405–415 (1985)
17. Koffka, K.: Principles of Gestalt Psychology, vol. 44. Routledge, London (2013)
18. Levitt, J.B., Kiper, D.C., Movshon, J.A.: Receptive fields and functional architecture of macaque v2. *J. Neurophysiol.* **71**(6), 2517–2542 (1994)
19. Lubliner, J.: Plasticity Theory. Courier Corporation, North Chelmsford, MA (2008)
20. Marsden, J.E., Hughes, T.J.: Mathematical Foundations of Elasticity. Courier Corporation, North Chelmsford, MA (1994)
21. Oppel, J.J.: Ueber geometrisch-optische Täuschungen. Jahresbericht des physikalischen Vereins zu Frankfurt am Main **1854–1855**, 37–47 (1855)
22. Petitot, J.: Neurogéométrie de la vision. Edition de l'école polytechnique (2008). ISBN: 978-2-7302-1507-7
23. Robinson, J.O.: The Psychology of Visual Illusion. Courier Corporation, North Chelmsford, MA (2013)
24. Sarti, A., Citti, G., Petitot, J.: The symplectic structure of the primary visual cortex. *Biol. Cybern.* **98**(1), 33–48 (2008)
25. Sarti, A., Citti, G., Petitot, J.: Functional geometry of the horizontal connectivity in the primary visual cortex. *J. Physiol. Paris* **103**(1), 37–45 (2009)
26. Smith, D.A.: A descriptive model for perception of optical illusions. *J. Math. Psychol.* **17**(1), 64–85 (1978)
27. Wade, N.: The Art and Science of Visual Illusions. Routledge Kegan & Paul (1982)
28. Walker, E.H.: A mathematical theory of optical illusions and figural aftereffects. *Percept. Psychophys.* **13**(3), 467–486 (1973)
29. Weickert, J.: Anisotropic Diffusion in Image Processing, vol. 1. Teubner, Stuttgart (1998)
30. Westheimer, G.: Illusions in the spatial sense of the eye: geometrical-optical illusions and the neural representation of space. *Vis. Res.* **48**(20), 2128–2142 (2008)

Exergaming for Autonomous Rehabilitation



Nunzio Alberto Borghese

Abstract Novel gaming devices, like the Nintendo balance board and the Microsoft Kinect, have opened the possibility to do autonomous rehabilitation at home, where patients are guided by adequate video-games that are named therapeutic exer-games. To ensure safety and effectiveness, it has become soon evident that some form of supervision through technology itself is required. This has led to the development of a three-layers architecture designed inside the EC funded REWIRE project constituted of a hospital station (HS), a networking station (NS), and a patient station (PS). In this paper we define the challenges for the Patient Station and how these could be addressed. The PS is installed at patient's home to guide him/her in the rehabilitation exercises through video-games designed to train a specific function. The PS is constituted of a classical game engine, on top of which several functions, specifically devoted to autonomous rehabilitation at home, have been added: configuration and real-time adaptation of the level of difficulty of the exer-games, monitoring the healthiness of motion pattern used by the patient, capability of interfacing different tracking device and log motion data. Moreover, a virtual therapist avatar has been fully integrated to maximize compliance, along with randomization of assets and background music.

1 Introduction

Once discharged from hospital, stroke patients have to continue exercising to further improve their condition or simply not to lose the benefits gained by hospital treatment, in which they underwent intensive personalized exercising aimed to stimulate lost functionalities such that recovering can occur. However, it has been shown that patients can also recover long after the stroke provided that they exercise intensively, where intensity is the product of time by challenge level [1].

N.A. Borghese (✉)

Department of Computer Science, Laboratory of Applied Intelligent Systems, Università degli Studi di Milano, Milano, Italy
e-mail: Alberto.borghese@unimi.it

In the traditional rehabilitation setting, this requires daily one-to-one rehabilitation sessions with a therapist but it involves a high burden to patients and caregivers that have to organize their lives to accommodate it and high costs that are mostly in charge on the patient's family.

The development of low tracking reliable devices for the video-games market, is offering a great opportunity to change this landscape. Devices such as the Nintendo balance board or the Microsoft Kinect have shown to be able to track human motion in real-time with an accuracy adequate for rehabilitation [2, 3]. This capability has been quickly seized by researchers in the field who have started exploring the integration of these devices with adequate video-games to guide patients to exercise. A large number of papers have been published on this topic and have all shown how boring and repetitive exercises could be turned into an amusing and entertaining experience [4-7]. However, several concerns have been raised on safety as several dropouts have been registered: patients are left alone with the exer-games with no therapist on his/her side and maladaptation and wrong movements can easily occur.

We show here how game technology can be pushed one step forward integrating proper artificial intelligence real-time algorithms to provide safety while exercising. This has been accomplished introducing a virtual therapist who supervises patient's activity inside the games.

2 Methodology

Our system is based on a hierarchical approach studied and developed in the framework of the EC funded REWIRE project (<http://www.rewire-project.eu>, [8], Fig. 1), that involves patients, hospitals or rehabilitation centers and health providers.

The central node is the Hospital Station (HS). This is a web-based application that allows clinicians to plan patients' exercising activity and review their performance. Each clinician selects a mix of exercises adequate for the patient. For posture rehabilitation, for instance, these will be a mix of coordination, force and endurance exercises. For each exercise one or more video-games can be selected and personalize to the current patient's ability. Such games are usually called exer-games to stress that they are videogames whose primary goal is not entertainment but exercising. Such exergames are designed in tight collaboration with therapists [9] in order to provide a progression in the therapy. For instance, posture rehabilitation starts with simple exer-games that require that the patient stands still with his/her weight well centered on the support. This is a fundamental pre-requisite to regain a good control of posture. Afterwards the patient starts exercises that require moving the body weight, first shifting the center of mass without moving the legs, then through lateral, frontal or oblique steps to arrive raising legs.

Personalization is particularly critical, as each patient's residual ability needs to be matched to achieve a proper amount of challenge: an exer-game that is too easy would become soon boring, a too difficult exer-game would be too challenging and pose the patient at risk. Such degree of difficulty is regulated by a few parameters



Fig. 1 The three layers of a platform that enables patients to do rehabilitation autonomously at home: the Patient Station is deployed at patient's home and it guides him/her in the exer-games. It receives the list of exercises, with constraints and level of difficulty matching the patient ability from the Hospital Station, and it returns results of the exercises. Each Hospital Station provides the Networking Station with the results of its patients thus providing optimization of the treatment

that control speed/accuracy/amplitude of the movements required to complete the games. They are matched to assets in the game. For instance to increase the amplitude of motion elicited from the patient, in the game “fruit catcher” reported in Fig. 2, apples can be spawn from the tree more spaced apart.

Another important element for rehabilitation is monitoring. This is implemented here through a knowledge based fuzzy system. Clinicians can define constraints on the movement, like for instance the maximum angle/height/force that should occur for a certain type of exercise. Typical examples of these constraints for posture rehabilitation are: equality of pressure exerted by the two feet on the ground, excessive bending of the spine/knee during some movements. Such conditions should be avoided as they increase the possibility of strain and bony and cartilage damage that manifests as pain [10].

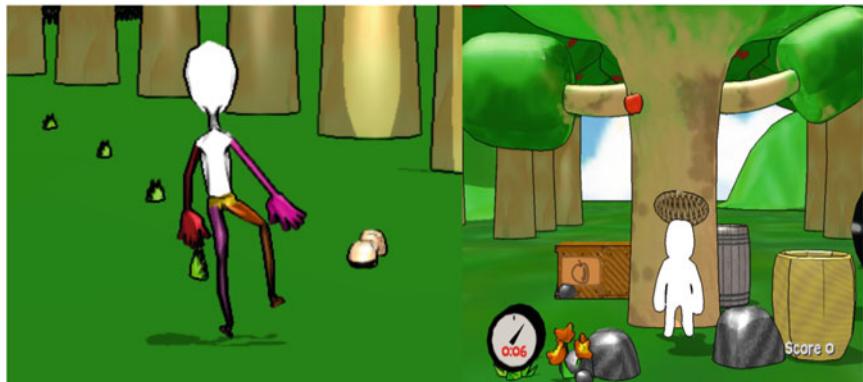


Fig. 2 Two examples of exer-games: on the left “Raising legs” the patient has to raise the foot to avoid a worm passing by and on the right, “Fruit Catcher” he/she has to catch virtual fruit falling from the top

A graphical interface has been developed to help therapists in defining such constraints for each exercise [11]. It allows defining the specific constraints on the movement and the degree of difficulty while patients are still exercising inside the hospital so that therapists can have a direct glimpse on how patients use their residual abilities in the specific exercises. The constraints defined through the graphical interface are automatically transformed into fuzzy rules [12]. For each constrained parameter a minimum and maximum value is defined according to therapist evaluation that is mapped automatically on a danger level from no danger to most dangerous. The range between minimum and maximum is fuzzyfied into four classes: zero, mild, medium, serious. An alarm level is associated to each of these classes: the more the patient gets close to the rule boundary, the higher becomes the alarm level.

The mix of exercises defined by the clinician in the hospital is then sent to the Patient Station (PS), installed at patient’s home. It is preferentially constituted of a PC connected with a TV and with adequate tracking devices (e.g. Kinect + Balance Board for posture rehabilitation). When the patient is ready to exercise, he/she switches on the host PC and launches the rehabilitation application. This shows him/her on the home TV the menu of the exercises. From this onwards, the patient can ignore the PC and interacts with the system through gestures or voice with the TV screen, making the interaction more natural.

The patient sees himself/herself as an avatar on the TV screen that has to complete a given goal (e.g. letting a worm pass below its foot by raising the foot or catching a virtual apple falling from a tree top by shifting the body laterally, Fig. 2).

The PS measures in real-time the motion of the patient’s body, and the pressure exerted on the ground when a balance board is used, and logs these quantitative data for further analysis. It also verifies with these data the fuzzy rules in real-time to

detect possible deviations from normal movement. In case the movement of one or more body segments does not obey the requirements specified by the therapist inside the hospital, the patient is warned as his/her exercise is becoming more dangerous healthy for him/her.

For immediate feed-back, a specific color coding scheme [11] is used: the color of the avatar segment (or segments) that are moving uncorrectly is changed according to the degree of deviation from normal behavior. When the movement is correct, the avatar has a white or green color. As soon as a segment enters inside a region in which the movement is not correct anymore, the associated segments change color from green to yellow, orange and red according to the degree of uncorrectness. For instance, if the patient is tilting his back during an exergame, the avatar's back would change the color from green to yellow, orange and then to red in the most severe cases. The PS computes a general alarm from the single alarm raised, and it transmits it to an avatar. In case the alarm is too large, the Virtual Therapist Avatar (VTA) pops up and advises the patient on how to correct his/her motion. In extreme cases, it stops the exer-game sending a warning to the hospital as it evaluates that continuing to exercise would be too risky for the patient.

To maximize the compliance, the same VTA pops up on the TV screen also in other phases: it welcomes friendly the patient at the beginning and it shows to the patient his/her results and progression at the end of the session. Moreover, it encourages the patient in between the different exer-games or whenever performances decreases.

We explicitly remark that a game engine for rehabilitation should be able to interface different devices upon the exercise requirements, as rehabilitation of different body districts requires the use of different tracking devices. For posture rehabilitation, we have used Microsoft Kinect as master device as it provides whole body tracking, and we have combined it with a balance board: either a Nintendo Wii or a Tyromotion Timo plate. We have also interfaced a Sony PS3 camera, a Novint Falcon haptic device, Moticon insoles and a PA10 robotics arm to track also arm movements. To this aim, we have incorporated in our system a middleware layer, named IDRA (Input Devices for Rehabilitation Abstraction layer), designed to match all the data streams required by each game to those provided by the devices. IDRA works also to avoid conflicts between multiple devices. We have called the resulting system IGER (Intelligent Game Engine for Rehabilitation).

IGER is complemented with Natural User Interfaces (NUI) that allow the user to interact with the game and with the menus through speech or gestures, forgetting of the PC for the entire duration of the rehabilitation session as mouse or keyboard are not needed. We have used here the functionalities of the Microsoft Kinect SDK to provide gesture interfacing. It has been implemented in combination with Google translator to obtain text to speech conversion from different languages of good quality to provide the VTA with speech capability to provide feed-back to the patient.

Other elements that contribute to increase the compliance are music, that can be played during the exercises, and assets randomization. Music, in particular, can be chosen by the patient according to his/her preferences and associated audio files

are inserted into a pre-defined folder. IGER will play the music files in random order for the entire exer-game duration. Assets randomization contributes also to introduce variability: each time the patient sees a different scene and different targets (e.g. apples, pears, peaches) perceiving therefore a different challenge and making exercising more entertaining (Fig. 3).

Upon completion of a session, the motion data, the successful (or failure) interaction with the game elements, the violated constraints and their severity level through time are returned to the HS for being evaluated by the clinicians. The trend in terms of accuracy, speed and amplitude is automatically computed over all the exercises and it is shown versus time to enable clinicians to evaluate patient's progression and possibly schedule follow-up visits. The HS therefore closes a therapeutic loop thus making the patient always supervised.

At the top level a Networking Station, at the Health Provider site, gathers all the data and trends from all hospitals and it evaluates the quality and effectiveness of the intervention of the different hospitals. Such data can be of great value for data mining algorithms.



Fig. 3 Assets randomization in one of the exer-games "Fruit catcher": different types of trees and fruits are defined. IGER picks up randomly one of these each time the game is started, thus producing game variability

3 Discussion

At-home a real therapist is not available to help the patient. IGER tries to replace his/her functionalities partially through a Virtual Therapist that supports the functionalities specifically developed to support rehabilitation. These in IGER are: games configurability, real-time adaptation and monitoring, feed-back and data log. Without these additional functionalities the effectiveness and safety of any exer-game for autonomous rehabilitation at home can be severely questioned. We explicitly remark that these functionalities cannot be developed as stand-alone but they should work in synergy with the exer-games. They also need to be conceived from the start of the design of the exer-game [13]: for instance, the exer-game must include a total body avatar if it mandatory to monitor the spine. Moreover, constraints violation modulates real-time adaptation: when violation is too large it inhibits adaptation to occur.

Adaptation is implemented at two levels. First, it is available as a configuration tool at the hospital site to define, for each exer-game, the proper level of difficulty among a predefined set. For instance in *Fruit Catcher* where lateral steps are required to catch fruits falling from a tree, the levels of difficulty are defined by the width of the interval from which fruits are spawn: the larger the range the more difficult the game. The difficulty level of the exergame can then be modified in real-time. Several solutions have been proposed to achieve a proper level of challenge for the patient during the exercises [11, 14–18]. In IGER, we have implemented the approach described in Pirovano et al. [11] that is based on a Bayesian framework: the therapist provides an a-priori value of the parameters according to his/her experience in the hospital with the patient. Such parameter is then modified in real-time in batch mode (after several trials) according to the success rate in the game in order to achieve a certain rate (e.g. 80% of success). The a-priori term avoids that the level becomes too challenging for the patient favoring maladaptation.

The Virtual Therapist is embodied into an avatar that accompanies the patient during his rehabilitation sessions throughout the life of the application, and advises him/her. Its main functional role is to provide feed-back on constraints violation explaining how to execute correctly a movement. But it has also a very important role in increasing the compliance: it explains how to navigate the interfaces and how to play the exer-games, it introduces different options, it congratulates the patient upon his/her achievements and motivates him/her when performances are decreasing. Some video-games for fitness use a similar approach; for instance the Wii Fit suite (<http://wiifit.com/>) uses a cartoon-like animated balance board to provide feed-back to the patient; more recently operating systems are adding such avatar functionalities to help browsing and interfacing with the PC (e.g. Cortana by Microsoft). We have explored different possibilities to implement the VTA of IGER: videos of a real-therapist (realistic approach) or a cartoon unrelated to real world (e.g. a piggy animal), but most preferred avatar was a cartoon of a therapist, that we have named Hannah (Fig. 4).



Fig. 4 Hannah, Virtual Therapist avatar of the IGER system

Several exer-games have been developed with IGER to address posture and balance, as well as arm rehabilitation according to specifications set by the therapist (for a list of exer-games see [11]). Each exer-game has a double requirement: the goal of the movement (what to do?) and the quality of movement (how to do?). Both these two requirements have to be taken care in game design and have been associated to primary and secondary goals of an exercise [13]: a therapeutic exer-game should be endowed with both these two aspects.

IGER does not require to use PC mouse or keyboard throughout the rehabilitation session as it fully implements a NUI paradigm. The PC could be integrated inside the TV in the next future as soon as Smart TV will become powerful enough to run a game engine. This would make the system fully integrated inside home appliances.

4 Conclusion

The platform described here allows integrating autonomous rehabilitation inside a clinical loop. In particular the Patient Station described here fully exploits the possibility offered by Artificial Intelligence and State-of-the-art Gaming technology to guide patients at home in their prescribed exercises in a safe and reliable way.

Preliminary usability results [19] show that the system is well received by patients who have used it at home during a 3-months period. More recent results related to a clinical study [20] show the efficacy of the IGER treatment.

References

1. Kwakkel, G.: Impact of intensity of practice after stroke: issues for consideration. *Disabil. Rehabil.* **28**(13–14), 823–830 (2006)
2. Clark, R., Pua, Y., Fortin, K., Ritchie, C., Webster, K.E., Denehy, L., Bryant, A.L.: Validity of the Microsoft Kinect for assessment of postural control. *Gait Posture.* **36**, 372–377 (2012)
3. Clark, R., Bryant, A.L., Pua, Y., McCrory, P., Bennell, K., Hunt, M.: Validity reliability of Nintendo Wii Balance Board for assessment of standing balance. *Gait Posture.* **31**, 307–310 (2010)
4. Lauterbach, S.A., Foreman, M.H., Engsberg, J.R.: Computer games as therapy for persons with stroke. *Games Health.* **2**(1), 311–318 (2013)
5. Kripic, A., Savanovic, A., Cikajlo, I.: Telerehabilitation: remote multimedia-supported assistance and mobile monitoring of balance training outcomes can facilitate the clinical staff's effort. *Int. J. Rehabil. Res.* **36**(2), 162–1731 (2013)
6. Llorens, R., Noè, E., Colomer, C., Aicaniz, M.: Effectiveness, usability and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: a randomized control trial. *Arch. Phys. Med. Rehabil.* **96**(3), 418–425 (2015)
7. Lohse, K.R., Hilderman, C.G.E., Cheung, K.L., Tatla, S., Van der Loos, H.F.M.: Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. *PLoS One.* **9**(3), e93318 (2014). <https://doi.org/10.1371/journal.pone.0093318>
8. Borghese, N.A., Murray, D., Paraschiv-Ionescu, A., de Bruin, E.D., Bulgheroni, M., Steblin, A., Luft, A., Parra, C.: *Rehabilitation at home: a comprehensive technological approach*. In: Ma, M., Jain, L., Anderson, P. (eds.) *Augmented Reality and Serious Games for Healthcare 1*, Intelligent Systems Reference Library, vol. 68, pp. 289–319. Springer, Berlin (2014)
9. Borghese, N.A., Pirovano, M., Lanzi, P.L., Wuest, S., de Bruin, E.D.: Computational Intelligence and Game Design for effective home-based stroke at Home Rehabilitation. *Games Health J.* **2**(2), 81–88 (2013)
10. Prosperini, L., Fortuna, D., Gianni, C., et al.: Home based Balance training using the Wii Balance Board: A cross-over pilot study i Multiple sclerosis. *Neurorehabil. Neural Repair.* **27**(6), 516–525 (2013)
11. Pirovano, M., Mainetti, R., Baud-Bovy, G., Lanzi, P.L., Borghese, N.A.: IGER - intelligent game engine for rehabilitation. *IEEE Trans. on CIAIG.* **8**(1), 43–55 (2016)
12. Kosko, B.: *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs (1991)
13. Pirovano, M., Surer, E., Mainetti, R., Lanzi, P.L., Borghese, N.A.: Exergaming and rehabilitation: a methodology for the design of effective and safe therapeutic exergames. *Entertain. Comput.* **14**, 55–65 (2016)

14. Yannakakis, G., Hallam, J.: Real-time game adaptation for optimizing player satisfaction. *IEEE Trans. Comput. Intell. AI Games.* **1**(2), 121–128 (2009)
15. Lopes, R., Bidarra, R.: Adaptivity Challenges in Games and Simulations: A Survey. *IEEE Trans. Comput. Intell. AI Games.* **3**(2), 85–98 (2011)
16. Zimmerli, L., Krewer, C., Gassert, R., et al.: Validation of a mechanism to balance exercise difficulty in robot-assisted upper-extremity rehabilitation after stroke. *J. Neuroeng. Rehabil.* **9**, 6–13 (2012)
17. Cameirão, M.S., Badia, S.B., Oller, E.D., Verschure, P.: Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: methodology, design, psychometrics, usability and validation. *J. Neuroeng. Rehabil.* **3**, 7–48 (2010)
18. Colombo, R., Pisano, F., Mazzone, A., et al.: Design strategies to improve patient motivation during robot-aided rehabilitation. *J. Neuroeng. Rehabil.* **4**(1), 3–12 (2007)
19. Held, J.P., Ferrer, B., Mainetti, R., Steblin, A., Hertler, B., Cerny, J., Moreno-Conde, A., Dueñas, A., Pajaro, M., Parra, C., Vargiu, E., Zarco, M.J., Barrera, M., Echevarria, C., Jódar-Sánchez, F., Luft, A.R., Borghese, N.A.: Autonomous rehabilitation at home based on virtual rehabilitation: safety, usability and compliance, a pilot study. *Eur. J. Phys. Rehabil.* (in press)
20. Ferrer, B., Dueñas, A., Moreno-Conde, A., Jódar-Sánchez, F., Zarco Periñan, M.J., Echevarria, C., Pajaro, M., Parra-Calderón, C.L., Alberto Borghese, N.: REWIRE project: Clinical assessment of a virtual reality home based rehabilitation programme (In preparation)

E-Infrastructures for Neuroscientists: The GAAIN and neuGRID Examples



Daniele Orlandi, Alberto Redolfi, Jérôme Revillard, David Manset,
Stefan Teipel, and Giovanni B. Frisoni

Abstract Multisite-distributed e-infrastructures can foster computational resources, share brain imaging data, and facilitate data analysis in the study of neurodegenerative diseases. The Global Alzheimer's Association Interactive Network (GAAIN—<https://www.gaan.org>) is a platform federating worldwide datasets of the brain. The European side of the GAAIN initiative (EU-GAAIN) exposes five datasets (4051 subjects from E-ADNI/PharmaCOG, I-ADNI, ARWIBO, EDSD, OASIS) currently hosted in the distributed neuGRID e-infrastructure (www.neugrid4you.eu). EU-GAAIN offers a huge amount of morphological and functional scans, surrogate imaging biomarker values (i.e. 4653 cortical and subcortical volumes), cognitive assessments and biochemical markers (i.e.: 286 CSF values, 235 APOE genotype). GAAIN platform provides 24 datasets (almost half million subjects) that can be explored through a user-friendly interface called “GAAIN Interrogator”. Through the GAAIN web-portal, queries results are displayed in graphs and summary tables providing information to understand trends and discover new evidences. The main objective of the GAAIN project is to build a technology for understanding the underlying mechanism of Alzheimer's disease and other forms of dementia through a data-driven approach. All the GAAIN data mapped in a common data schema are federated through the Data

D. Orlandi (✉) • A. Redolfi

Laboratory of Alzheimer's Neuroimaging and Epidemiology—LANE, IRCCS Institute—The Saint John of God Clinical Research Centre, Brescia, Italy

e-mail: daniele.orlandi.psy@gmail.com; aredolfi@fatebenefratelli.eu

J. Revillard • D. Manset

Gnúbla France, Imp Pres d'en Bas, Argonay, France

S. Teipel

DZNE, German Center for Neurodegenerative Diseases, Rostock, Germany

G.B. Frisoni

Laboratory of Alzheimer's Neuroimaging and Epidemiology—LANE, IRCCS Institute—The Saint John of God Clinical Research Centre, Brescia, Italy

Memory Clinic and LANVIE—Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Geneva, Switzerland

Partner Clients (DPC) and a distributed pipeline enactment service that allows the execution of analyses relying on distributed Docker machines. Research efficiency can be increased if neuroscientists will access secure federated platforms from a single point of access. GAAIN initiative is funded by the Alzheimer’s Association and by National Institutes of Health grants.

1 Introduction

Magnetic resonance imaging (MRI) is a technique for digitalizing in vivo brain morphology, connectivity and functionality [1]. Thanks to its non-invasive characteristic, it has been extensively employed for describing changes in normal ageing [2] and measuring Alzheimer’s disease progression [3, 4]. Currently, there are evidences that cross sectional and longitudinal MRI studies can facilitate the creation of predictive models able to identify subjects at risk of cognitive decline before appearance of clinical symptoms [5]. Nevertheless, even if predictive model’s build a specific cohort can reach good accuracy percentage in recognize subject’s diagnosis, this task is still unsteady when algorithms are tested on unseen data [6, 7]. On the other hand, a large number of cases is an optimal groundwork for evaluate and compare the power of prediction of different models. Collection of data from a broad range of different sources helps creating generalizable diagnostic and prediction models that more accurately reflect the variability of data and cohort characteristics from routine data than data collected at single sites. Since prospective large-scale data collection requires consistent funding to assemble, store and maintain a cohort, data exchange could be considered as an alternative solution to carry out research with reduced costs.

Exchange of raw and meta data is already a common practice in several branches of science, as genomics [8] and proteomics [9]. In neuroimaging, a growing number of initiatives provide external investigators with access to data of healthy people [10, 11] and people with diseases [12] to. One of the first pioneer projects is called *Alzheimer’s Disease Neuroimaging Initiative—ADNI*. The project started in 2004 with the aim to find more sensitive and accurate methods to detect Alzheimer’s disease at earlier stages and mark its progress through biomarkers. The private–public partnership is now in its fourth phase (i.e.: ADNI1, ADNI-GO, ADNI2, ADNI3) and counts more than 1000 subjects longitudinally followed from 55 different centers [13]. The availability of *ADNI* data has generated a huge interest within the scientific community helping to define biomarker trajectories in Alzheimer’s disease (AD). Other remarkable data sharing initiatives aimed to facilitate large-scale dataset access are the *1000 Functional Connectomes’ Project—1000FCP* [14] and the *Human Connectome Project—HCP* [15]. These projects released public datasets for mapping the brain functional connectome. Even if these renowned projects have demonstrated the advantages of data sharing for the scientific community, the majority of data owners are reluctant to share their data with other researchers due to fear of privacy issues and technical challenges to enable secure data sharing. For this reason, a data federation approach is more appealing for researchers [16, 17].

Indeed, while *data sharing* implies copy or transfer of raw data, *data federation* make available only metadata and derived information, leaving to coordinators the complete control on their own data. Noteworthy examples of data federation projects are the *European Medical Information Framework—EMIF* [18], the *Human Brain Project—HBP* [19] and the *Global Alzheimer’s Association Interactive Network—GAAIN* [20]. These projects are at different stages and they are respectively focused on data coming from health care (*EMIF* and *HBP*) and research studies (*GAAIN*). The effort carried out by these initiatives well responds to the growing need of data integration raised by the scientific community [21].

This paper describes two interlinked e-platforms (*NeuGRID* and *GAAIN*) devoted to dementia and dealing with data sharing and federation concepts. In the following section, the e-infrastructures are extensively described providing an overview of their respective data management approaches and services.

2 Methods

NeuGRID and *GAAIN* are virtual laboratories where researchers can find datasets coming from different sources. Even though the projects adopt different technical solutions, both initiatives embrace the considerable demand to test experimental hypothesis on large-scale data.

The first platform we want to describe is *NeuGRID* (<https://neugrid4you.eu>), an e-infrastructure for clinicians and neuroscientists that serves as a virtual laboratory where users can find data and run analyses in a secured user-friendly environment [22].

NeuGRID architecture (Fig. 1) is based on the grid technology and can rely on 20 TB of physical space and 6000 cores placed in three main dedicated nodes (Hôpitaux Universitaires de Genève—HUG; Fatebenefratelli—FBF; CEA—Neurospin), and European Grid Infrastructure (EGI) public resources. This technical framework currently hosts 17 different datasets (Table 1) with a total of more than 120,000 multimodal images. All the facilities are freely accessible by registered users, in particular, datasets are available with three different levels of restrictions (open, facilitated, restricted access) after signing related data access policies.

As far as data ownership is concerned, principal investigators (PI) can entrust their study to *NeuGRID* and safeguard their own data on the basis of unambiguous terms of use. The different levels of access rights are:

- OPEN ACCESS: these data can be downloaded and analyzed from *neuGRID* for free. As soon as users sign the N4U form there are no further restrictions or formalities.
- FACILITATED ACCESS: data analysis in *neuGRID* is allowed only after having obtained an account from *ADNI* and permission for the intended use. *ADNI* will provide an account to anyone requesting it, except in some circumstances. The *N4U Specific Support Centre (SSC)* will facilitate data exploitation by uploading data into a private user area within *neuGRID*.

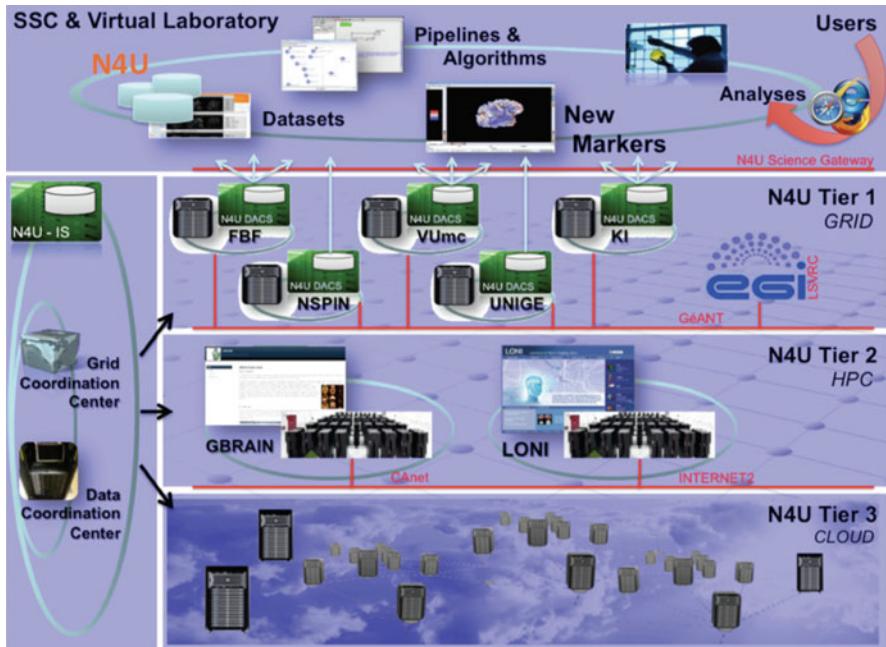


Fig. 1 The platform is mainly divided in two sections, in the upper part is shown the User Level (UL). The UL is made from a Specific Support Centre (SSC) and the Virtual Laboratory (VL). The SSC grant support and provides training guides to users while in the VL users will find data, algorithms, computational resources, statistical tools, and appropriate help in a user-friendly environment. Under the UL is represented the second main part of the platform, the Distributed Computing Infrastructure (DCI). The DCI is composed of three Tiers and is built up to respond at the different needs of users. The tier 1 represents the core of the infrastructure. Tier 2 connects the Grid environment to international initiatives (e.g.: LONI, CBRAIN) while the tier 3 attaches cloud solutions. The three tiers are orchestrated by the Grid Coordination Centre (GCC) and the Data Coordination Centre (DCC). Reproduced with permission from [22]

- **RESTRICTED ACCESS:** data access is restricted exclusively to co-workers or collaborators of the Consortium. Authorization to access these data for external users must be obtained from the Consortium, such as *E-ADNI/PharmaCog*, by submitting an application and eventually signing an ad-hoc scientific agreement.

In order to facilitate data intelligibility, each database is described with two main *comma-separated values (CSV)* files, *Clinical Variable* and *Data Dictionary*. The *Clinical Variable* file gathers all the attributes that characterize the study, e.g. demographic, neuropsychological, diagnosis, biological, genetic data and study related scores. Scores and values are indexed in the *Data Dictionary* file where variables are divided into specific categories (i.e. variable name, category, unit of measure, data type, value range, element description, value range description) and explained thanks to a detailed description. File structure for both *Clinical Variable* and *Data Dictionary* is harmonized for all the databases. That allows

Table 1 List of the available datasets in the NeuGRID and GAAIN platforms

COHORT	NeuGRID	GAAIN	Subjects	Diagnosis	Mild Cognitive Impairment	Alzheimer's Disease	Psychiatric Disease	Imaging	Genotype	CSF
1000-FCP	✓		1263	✓				✓	✓	✓
ABIDE	✓		1112	✓				✓	✓	
ADHD-200	✓		964	✓				✓	✓	
ADNI	✓		1387	✓	✓			✓	✓	
AIBL	✓		1413	✓	✓	✓		✓	✓	
ARWIBO	✓		2619	✓	✓	✓		✓	✓	
BIOCARD	✓		348	✓	✓	✓		✓	✓	
CAMD	✓		6500	✓	✓	✓		✓	✓	
CLSA	✓		21,237	✓	✓	✓		✓	✓	
COBRE	✓		148	✓				✓	✓	
DART		✓	938	✓				✓	✓	
DIAN		✓	342	✓				✓	✓	
EDSD	✓		467	✓	✓	✓		✓	✓	
FBIRN	✓		253	✓				✓	✓	
Framingham Heart Study		✓	15,117	✓	✓	✓			✓	
BNA	✓		202,375	✓	✓	✓		✓	✓	
Fundació ACE	✓		12,126	✓	✓	✓		✓	✓	
Health and Retirement Study		✓	38,183	✓					✓	
HELIAD		✓	1225	✓	✓	✓		✓	✓	
H2H	✓		185	✓	✓	✓			✓	

(continued)

Table 1 (continued)

COHORT	NeuGRID	GAAIN	Subjects	Diagnosis	Mild Cognitive Impairment	Alzheimer's Disease	Psychiatric Disease	Imaging	Genotype	CSF
INDI	✓		415	✓	✓	✓	✓	✓		
Italian ADNI	✓	✓	262	✓	✓	✓	✓	✓		
IAADC	✓	✓	990	✓	✓	✓	✓	✓		
LMRR	✓	✓	318	✓	✓	✓	✓	✓		
MAGNIMS	✓		127	✓	✓	✓	✓	✓		
MIRIAD	✓		69	✓	✓	✓	✓	✓		
NACC	✓	✓	32,364	✓	✓	✓	✓	✓		
NIAGADS	✓		24,975	✓	✓	✓	✓	✓		
NUSDAST	✓		362	✓		✓		✓		
OASIS	✓	✓	565	✓	✓	✓	✓	✓		
E-ADNI/ PharmaCog Volunteers	✓		30	✓						
E-ADNI/ PharmaCog Patients	✓	✓		147	✓			✓	✓	✓
Penn INDD	✓		4424	✓	✓	✓	✓	✓	✓	
Wisconsin Longitudinal Study	✓		19,095	✓	✓	✓	✓			
WRAP	✓		1541	✓	✓	✓		✓		✓

exploring, comparing and understanding attributes faster than a random information disposition (e.g. clinical information gathered in uncoupled files not harmonized between them) for all the databases. In addition, more than 20 pipelines are already available to compute surrogate imaging biomarkers. Some of the most commonly used pipelines for processing and analyzing structural and functional images as well as statistical programs and brain atlases are already installed. Applications for single case analyses are also provided, such *AdaBoost* [23], a machine learning meta-algorithm applied to hippocampal segmentation and implemented with the *Harmonized Hippocampal Protocol (HarP)* [24]. All *NeuGRID* tools are summarized in Table 2 under three different neuroscience communities, namely, (i) including neurodegenerative (NS-NDD), (ii) white matter (NS-WMD) and (iii) psychiatric diseases (NS-PSY). Moreover, *NeuGRID* administrators care about users request uploading/updating datasets and algorithms as well as offer assisted and on-line e-training for facilitating infrastructure usage. As proof of concept, a *Data Challenge* [25] was carried out with excellent results. It has been shown that 1100 images could be analyzed in *NeuGRID* much faster than if the analysis had taken place in a small laboratory with low computational power (10 weeks vs. 5 years). Due to exponential increase of computing power over the last years, the pure computational advantage will have decreased, but the advantages of an integrated environment providing data and established pipelines in a user-friendly environment remain.

The second project is the Global Alzheimer's Association Interactive Network (GAAIN—<http://www.gaan.org/>), a collaborative partnership among the Laboratory of Neuro Imaging (LONI) at the University of Southern California (USC), the National Centre for Alzheimer's Disease Research and Care at the University of Geneva, the Laboratory of Alzheimer's Neuroimaging and Epidemiology of Brescia (LANE) and Gnúbila France. *GAAIN* is funded by the Alzheimer's Association. *GAAIN* is a project of data federation with the aim to gather different kinds of information (i.e. genetic, phenotypic, imaging) coming from worldwide dementia studies in a global virtual laboratory [20].

This innovative project is based on a sophisticated interactive network (Fig. 2) collating raw data through the *Data Partner Client (DPC)*. The *DPC* is a Java jar file compatible with every operative system (e.g. windows, ubuntu) and have function to bridge the data partner site and the *GAAIN* central servers [16]. Through it, data owners can choose the online/offline status of their contribution in the global cohort and connect/disconnect their own center from the central server at any time (Fig. 2). When online, metadata are automatically shared feeding the global cohort displayed in the *Scoreboard* (Fig. 3) and in the *Interrogator* (Fig. 4). In contrast, users cannot visualize any data from an offline center. *GAAIN* architecture allows PIs to (1) be the unique data keepers, (2) decide to show data only when they want, (3) share only metadata instead of the whole dataset. In order

Table 2 List of the available software in the NeuGRID platform

Software	NS-NDD	NS-WMD	NS-PSV	References
FREESURFER	✓	✓		http://surfer.nmr.mgh.harvard.edu/fswiki/
FREESURFER ReconAll Cross Sectional	✓	✓		http://surfer.nmr.mgh.harvard.edu/fswiki/
FREESURFER ReconAll Longitudinal	✓	✓		http://surfer.nmr.mgh.harvard.edu/fswiki/
FREESURFER Tracula	✓	✓	✓	http://surfer.nmr.mgh.harvard.edu/fswiki/
MAPS-HBSI	✓	✓	✓	https://www.ncbi.nlm.nih.gov/pubmed/20230901
FSL	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
FSL_TBSS—N2	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
FSL FIRST	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
FSL SIENA	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
FSL SIENAX	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
FSL Melodic	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
FSL ProtrackX2	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
FSL fpirit	✓	✓	✓	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
MIPAV	✓	✓	✓	https://mipav.cit.nih.gov/
MNI LIBRARIES			✓	http://packages.bic.mni.mcgill.ca/
SPM-LST			✓	http://www.applied-statistics.de/lst.html
CASCADE			✓	http://ki.se/en/mys/cascade
kNN-Tissue Type Prior				http://packages.bic.mni.mcgill.ca/
Lesion TOADS			✓	http://packages.bic.mni.mcgill.ca/
AdaBoost	✓	✓	✓	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2624575/
CLASP-CIVET	✓	✓	✓	https://www.ncbi.nlm.nih.gov/pubmed/15896981
BRAINVISA	✓	✓	✓	http://brainvisa.info/web/index.html
R	✓	✓	✓	https://www.r-project.org/
OCTAVE	✓	✓	✓	http://www.gnu.org/software/octave/

ITK—VTK	✓	✓	✓	http://www.vtk.org/
SPM—VBM	✓	✓	✓	http://www.fil.ion.ucl.ac.uk/spm/
AFNI—SUMA	✓	✓	✓	https://afni.nimh.nih.gov/afni/suma/
ImageJ	✓	✓	✓	https://imagej.nih.gov/ij/
ITK-SNAP	✓	✓	✓	http://www.itksnap.org/pmwiki/pmwiki.php
3D-Slicer	✓	✓	✓	https://www.slicer.org/

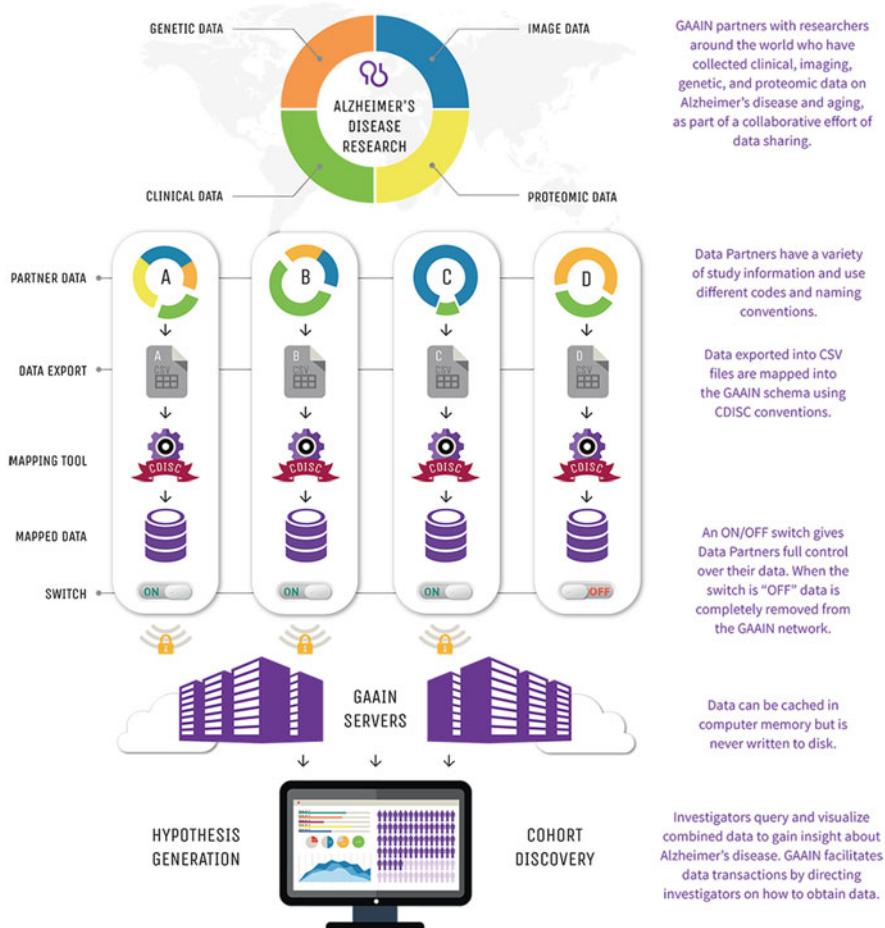


Fig. 2 GAAIN architecture is composed by a web portal, a central server and multiple Data Partner Clients (DPC). Every single center organizes scores in a CSV file and load it in the respective DPC application. Dataset information are automatically mapped in a common data schema and graphically represented in the GAAIN web page. A switch system embedded in the DCP, grant to PIs the freedom to disconnect data from the network

to aggregate disparate information in a harmonized data schema, the *US-GAAIN* has developed an automatic tool for data mapping, the *GAAIN Entity Mapper* [26, 27]. Metadata are accessible through three different interactive interfaces, namely: The *Cohort Scout*, the *Scoreboard*, the *Interrogator*. The first two sections are publicly accessible while the last one requires a personal account. Currently, each database is characterized by 24 variables stackable in four main categories: Imaging, Clinic, Proteomic, Genetic.

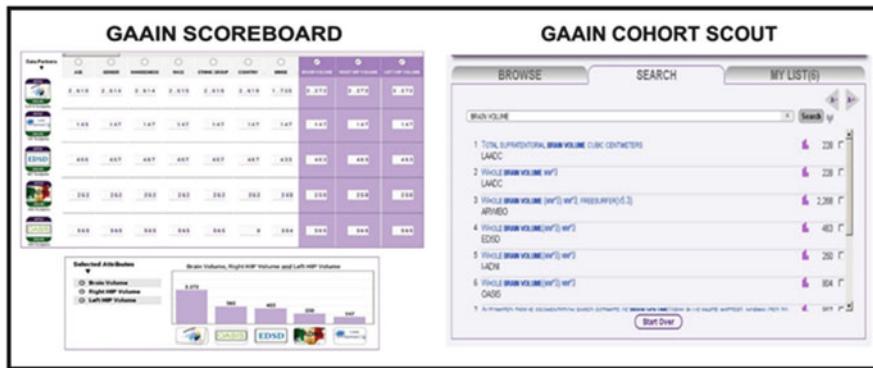


Fig. 3 On the left side: GAAIN Scoreboard selection of the datasets actually exposed in the GAAIN portal and hosted by the NeuGRID platform. On the top row of the page are reported 24 variables, while datasets logos are represented in the first left-hand column. The match between variable and dataset raw is a number showing the frequency of appearance. When a specific variable is selected (violet column), in the column chart, datasets are ordered by their contribution (i.e. brain volumes). On the right side: GAAIN Cohort Scout is a catalogue containing a detailed description of all the variables held in a specific dataset

At the bottom of the *GAAIN scoreboard* (Fig. 3), there is a column chart. This represents the selected variables and organizes columns in order of magnitude. The *GAAIN Scoreboard* allows researchers to estimate type and amount of information held in each dataset. Through the search bar of the *GAAIN Cohort Scout*, it is possible to insert a single keyword and launch a cross studies research without combining every dataset. The output is a match between the keyword and the contents of each dataset.

The *GAAIN Interrogator* (Fig. 4) is a user-friendly interface able to create online graphical visualization of data. On the left side, the cohort definition menu allows users to define a specific query using filters and editing attribute value. As soon as the cohort definition is refined, graphs automatically update showing data trends. The cohort can be further reshaped by selecting and editing the variables range at the extremes of the axes. In the upper part of the page logos and the number of subjects meeting the criteria defined in the cohort definition menu are displayed. When researchers find interesting results, by clicking on *apply*, they can automatically send a data request to the *Data Partner*. Through this service, users can trigger image and statistical analysis on a remote server (i.e. *NeuGRID* servers) moving pipelines instead of images. This function is still under development and it is exploiting cutting edge IT solutions through the Docker Machine technology.

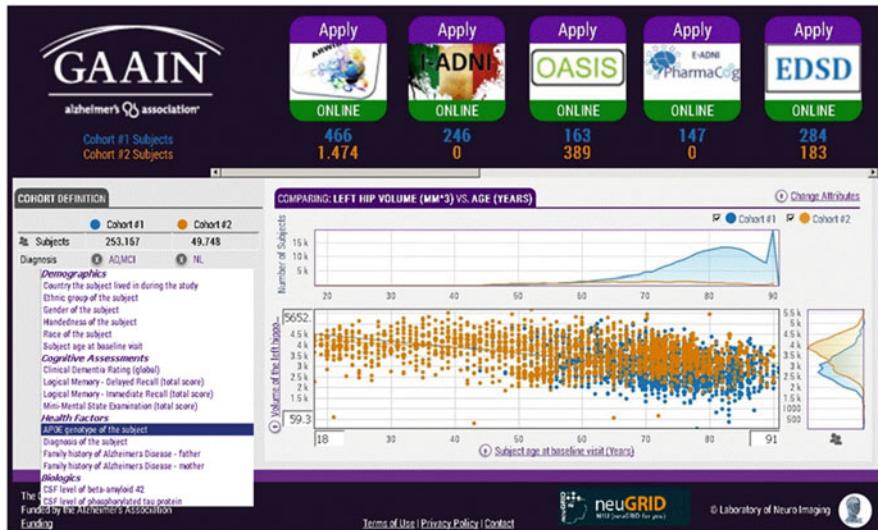


Fig. 4 GAAIN Interrogator is a dynamic interface that allows complex queries across five data domains (Demographic, Healthy Factors, Cognitive assessment, Biologic, Imaging). A graphic and intuitive representation is displayed in the center of the screen. In the upper part of the image are displayed the logos and the number of subjects that meet the criteria defined in the cohort definition menu (left side of the image). Through the cohort definition menu, two groups with different diagnosis (#1 AD, MCI and 2# NL) are compared. As the axes reference two variables were selected, Hippocampal left volume and Subjects age the baseline. At this point, a specific cohort is created, when researchers find some results interesting and would like to get them, clicking on apply a data request is automatically sent to the data ownership

3 Results

So far, 24 Data Partners are collaborating to the *GAAIN* project making available metadata through the web portal (Table 1). Thanks to particular platform architecture, populations belonging to the different Data Partners are merged in a unique cohort. Currently, data of 388,958 subjects are available in *GAAIN*. The EU-GAAIN, through the *NeuGRID* platform is contributing with the federation of five datasets (OASIS, I-ADNI, E-ADNI/PharmaCOG, ARWIBO, EDSD) feeding the *GAAIN* cohort with a significant number of scans (T13D, T2FLAIR, T2FSE, DTI, rsfMRI, 18F[FDG]), surrogate imaging biomarkers (4653 cortical/subcortical volumes), cognitive assessments (Logical-Memory, MMSE, CDR), and biochemical markers (286 CSF values, 235 APOE genotype). The robustness of the *GAAIN* platform has been recently tested investigating the influence of APOE on gender in developing AD on a population of almost 58,000 subjects [28]. Contrary to how is supported from studies involving a lower number of participants, the research shows that APOE $\epsilon 3/\epsilon 4$ genotype carriers have the same probability to develop AD

independently from which gender they belong, albeit women have increased risk at younger ages.

Moreover, *NeuGRID* represents the pillar of the *GAAIN* project for the European area (*EU-GAAIN*), offering space of storage for those PIs willing to make their data available to the scientific community.

4 Discussion

We have illustrated two examples of virtual laboratories, *NeuGRID* and *GAAIN*, and described their services. Both platforms aim to help researchers in finding and accessing dementia related datasets through different IT architectures and data management systems. In particular, the *NeuGRID* platform relies on a European grid network. The project promotes a data sharing approach while the structural aspect of the infrastructure permits to host and redistribute duplicated datasets. Differently, *GAAIN* portal pursues a data federation strategy based on connection of independent repositories. Moreover, the *DPC* applications allow coordinators to keep control of data, granting at the same time a distributed pipeline execution system. In this framework, the *GAAIN* architecture encourages coordinators to contribute to a global cohort that today has reached half a million of subjects.

The *NeuGRID* and *GAAIN* infrastructures operate separately from each other. A virtualized solution based on a *Docker Machine* technology is under development with the aim to connect both platforms. The growing synergy between portals could lead to crucial benefits for researchers, as maximizing research participant's contribution, empowering new questions and reproducibly, testing new analysis methods/workflow or clinically relevant and generalizable imaging markers.

In the last years, data integration's process has reshaped the size of studies from few dozens to thousands of subjects showing trends that could not be considered in a small-scale study. Despite advantages that data exchange can bring in discovering underlying mechanism of the brain (i.e. shorten the time for finding new effective treatments for dementia), for several reasons, plenty datasets are not still shared or reused. Today, the scientific community offers practical solutions in order to pool medical archives and boost brain comprehension. The process of data aggregation will accelerate the identification of new neuroimaging biomarkers for Alzheimer's Disease and shed new light on neurodegenerative pathologies.

Acknowledgments This work was supported by the Global Alzheimer's Association Interactive Network (GAAIN) initiative of the Alzheimer's Association (grant 003278) and by National Institutes of Health grants 5P41 EB015922-16 and 1U54EB020406-01. The investigators would like to acknowledge Elia Sbeiti for his helpful IT assistance and Margherita Mauri for having polished the English language of the manuscript.

References

1. Tan, X.Y., Pi, Y.L., Wang, J., Li, X.P., Zhang, L.L., Dai, W., Zhu, H., Ni, Z., Zhang, J., Wu, Y.: Morphological and functional differences between athletes and novices in cortical neuronal networks. *Front. Hum. Neurosci.* **10**, 660 (2017)
2. Fjell, A.M., Walhovd, K.B.: Structural brain changes in aging: courses, causes and cognitive consequences. *Rev. Neurosci.* **21**(3), 187–221 (2010)
3. Dubois, B., Feldman, H.H., Jacova, C., Hampel, H., Molinuevo, J.L., Blennow, K., DeKosky, S.T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G.B., Fox, N.C., Galasko, D., Habert, M.O., Jicha, G.A., Nordberg, A., Pasquier, F., Rabinovici, G., Robert, P., Rowe, C., Salloway, S., Sarazin, M., Epelbaum, S., de Souza, L.C., Vellas, B., Visser, P.J., Schneider, L., Stern, Y., Scheltens, P., Cummings, J.L.: Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol.* **13**(6), 614–629 (2014)
4. Frisoni, G.B., Fox, N.C., Jack Jr., C.R., Scheltens, P., Thompson, P.M.: The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol.* **6**(2), 67–77 (2010)
5. Den Heijer, T., van der Lijn, F., Koudstaal, P.J., Hofman, A., van der Lugt, A., Krestin, G.P., Niessen, W.J., Breteler, M.M.: A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain.* **133**(Pt 4), 1163–1172 (2010)
6. Allen Genevera, I., Amoroso, N., Anghel, C., Balagurusamy, V., Bare, C.J., Beaton, D., Bellotti, R., Bennett, D.A., Boehme, K., Boutros, P.C., Caberlotto, L., Caloian, C., Campbell, F., Chaibub Neto, E., Chang, Y.-C., Chen, B., Chen, C.-Y., Chien, T.-Y., Clark, T., Das, S., Davatzikos, C., Deng, J., Dillenberger, D., Dobson, R.J.B., Dong, Q., Doshi, J., Duma, D., Errico, R., Erus, G., Everett, E., Fardo, D.W., Friend, S.H., Fröhlich, H., Gan, J., St George-Hyslop, P., Ghosh, S.S., Glaab, E., Green, R.C., Guan, Y., Hong, M.-Y., Huang, C., Hwang, J., Ibrahim, J., Inglese, P., Jiang, Q., Katsumata, Y., Kauwe, J.S.K., Klein, A., Kong, D., Krause, R., Lalonde, E., Lauria, M., Lee, E., Lin, X., Liu, Z., Livingstone, J., Logsdon, B.A., Lovestone, S., Lyappan, A., Ma, M., Malhotra, A., Mangavite, L.M., Maxwell, T.J., Merrill, E., Nagorski, J., Namasivayam, A., Narayan, M., Naz, M., Newhouse, S.J., Norman, T.C., Nurtdinov, R.N., Oyang, Y.-J., Pawitan, Y., Peng, S., Peters, M.A., Piccolo, S.R., Praveen, P., Priami, C., Sabelnykova, V.Y., Senger, P., Shen, X., Simmons, A., Sotiras, A., Stolovitzky, G., Tangaro, S., Tateo, A., Tung, Y.-A., Tustison, N.J., Varol, E., Vradenburg, G., Weiner, M.W., Xiao, G., Xie, L., Y Xie, J.X., Yang, H., Zhan, X., Zhou, Y., Zhu, F., Zhu, H., Zhu, S., Alzheimer's Disease Neuroimaging Initiative: Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. *Alzheimers Dement.* **12**(6), 645–653 (2016)
7. Bron, E.E., Smits, M., Flier, v.d.W.M., Vrenken, H., Barkhof, F., Scheltens, P., Papma, J.M., Steketee, R.M.E., Méndez Orellana, C., Meijboom, R., Pinto, M., Meireles, J.R., Garrett, C., Bastos-Leite, A.J., Abdulkadir, A., Ronneberger, O., Amoroso, N., Bellotti, R., Cárdenas-Peña, D., Álvarez-Meza, A.M., Dolph, C.V., Iftekharuddin, K.M., Eskildsen, S.F., Coupé, P., Fonov, V.S., Franke, K., Gaser, C., Ledig, C., Guerrero, R., Tong, T., Gray, K.R., Moradi, E., Tohka, J., Routier, A., Durrleman, S., Sarica, A., Di Fatta, G., Sensi, F., Chincarini, A., Smith, G.M., Stoyanov, Z.V., Sørensen, L., Nielsen, M., Tangaro, S., Inglese, P., Wachinger, C., Reuter, M., van Swieten, J.C., Niessen, W.J., Kleina, S., Alzheimer's Disease Neuroimaging Initiative: Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADementia challenge. *Neuroimage.* **111**, 562–579 (2015)
8. Clarke, L., Fairley, S., Zheng-Badley, X., Streeter, I., Perry, E., Lowy, E., Tassé, A.M., Flück, P.: The international Genome sample resource (IGSR): a worldwide collection of genome variation incorporating the 1000 genomes project data. *Nucleic Acids Res.* **45**(D1), D854–D859 (2017)
9. Vaudel, M., Verheggen, K., Csordas, A., Ræder, H., Berven, F.S., Martens, L., Vizcaíno, J.A., Barsnes, H.: Exploring the potential of public proteomics data. *Proteomics.* **16**(2), 214–225 (2016)

10. Job, D.E., Dickie, D.A., Rodriguez, D., Robson, A., Danso, S., Pernet, C., Bastin, M.E., Boardman, J.P., Murray, A.D., Ahearn, T., Waiter, G.D., Staff, R.T., Deary, I.J., Shenkin, S.D., Wardlaw, J.M.: A brain imaging repository of normal structural MRI across the life course: brain images of normal subjects (BRAINS). *Neuroimage*. **144**(Pt B), 299–304 (2017)
11. Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A., Farrell, J.A., Bogovic, J.A., Hua, J., Chen, M., Jarso, S., Smith, S.A., Joel, S., Mori, S., Pekar, J.J., Barker, P.B., Prince, J.L., van Zijl, P.C.: Multi-parametric neuroimaging reproducibility: a 3-T resource study. *Neuroimage*. **54**(4), 2854–2866 (2011)
12. Fotenos, A.F., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L.: Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*. **64**, 1032–1039 (2005)
13. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R. Jr, Jagust, W., Morris, J.C., Petersen, R.C., Salazar, J., Saykin, A.J., Shaw, L.M., Toga, A.W., Trojanowski J.Q., Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's disease neuroimaging initiative 3: continued innovation for clinical trial improvement. *Alzheimers Dement.* (2016) (pii: S1552-5260(16)33072-2)
14. Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kotter, R., Li, S.J., Lin, C.P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.F., Zhang, H.Y., Castellanos, F.X., Milham, M.P.: Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* **107**(10), 4734–4739 (2010)
15. Van Essen, D.C., Glasser, M.F. The human connectome project: progress and prospects. *Cerebrum*. cer-10-16 (2016)
16. Neu, S.C., Crawford, K.L., Toga, A.W.: Sharing data in the global Alzheimer's association interactive network. *Neuroimage*. **124**(Pt B), 1168–1174 (2016)
17. Ritchie, C.W., Molinuevo, J.L., Truyen, L., Satlin, A., Van der Geyten, S., Lovestone, S., European Prevention of Alzheimer's Dementia (EPAD) Consortium: Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. *Lancet Psychiatry*. **3**(2), 179–186 (2016)
18. Bastião Silva, L.A., Díaz, C., van der Lei, J., Oliveira, J.L.: Architecture to summarize patient-level data across borders and countries. *Stud. Health Technol. Inform.* **216**, 687–690 (2015)
19. Amunts, K., Ebll, C., Muller, J., Telefont, M., Knoll, A., Lippert, T.: The human brain project: creating a European research infrastructure to decode the human brain. *Neuron*. **92**(3), 574–581 (2016)
20. Toga, A.W., Neu, S.C., Bhatt, P., Crawford, K.L., Ashish, N.: The global Alzheimer's Association Interactive Network. *Alzheimer's Dementia*. **12**(1), 49–54 (2016)
21. Poldrack, R.A., Gorgolewski, K.J.: Making big data open: data sharing in neuroimaging. *Nat. Neurosci.* **17**(11), 1510–1517 (2014)
22. Redolfi, A., Bosco, P., Manset, D., Frisoni, G.B., NeuGRID Consortium: Brain investigation and brain conceptualization. *Funct. Neurol.* **28**(3), 175–190 (2013)
23. Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikhshak, N., Hua, X., Toga, A.W., Jack Jr, C.R., Weiner, M.W., Thompson, P.M., Alzheimer's Disease Neuroimaging Initiative: Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage*. **43**(1), 59–68 (2008)
24. Frisoni, G.B., Jack, C.R.: HarP: the EADC-ADNI harmonized protocol for manual hippocampal segmentation. A standard of reference from a global working group. *Alzheimers Dement.* **11**(2), 107–110 (2015)

25. Redolfi, A., Manset, D., Barkhof, F., Wahlund, L.O., Glatard, T., Mangin, J.F., Frisoni, G.B., neuGRID Consortium, for the Alzheimer's Disease Neuroimaging Initiative: Head-to-head comparison of two popular cortical thickness extraction algorithms: a cross-sectional and longitudinal study. *PLoS One.* **17**(3), e0117692 (2015)
26. Ashish, N., Dewan, P., Ambite, J.L., Toga, A.W.: GEM: the GAAIN entity mapper. *Data Integr Life Sci.* **9162**, 13–27 (2015)
27. Ashish, N., Dewan, P., Toga, A.W.: The GAAIN entity mapper: an active-learning system for medical data mapping. *Front. Neuroinform.* **9**, 30 (2016)
28. Neu, S.C., Pa, J., Kukull, W., Beekly, D., Kuzma, A., Gangadharan, P., Wang, L.S., Romero, K., Arneric, S.P., Redolfi, A., Orlandi, D., Frisoni, G.B., Au, R., Devine, S., Auerbach, S., Espinosa, A., Boada, M., Ruiz, A., Johnson, S.C., Kosik, R., Wang, J.J., Hsu, W.C., Chen, Y.L., Toga, A.W.: Apolipoprotein E genotype and sex risk factors for Alzheimer disease: a meta-analysis. *JAMA Neurol.* **74**(10), 1178–1189 (2017)

Theory and Application of Nonlinear Time Series Analysis



Angelo Di Garbo

Abstract The nonlinear time series analysis is an important approach for the analysis of signals from physics, astrophysics to neurophysiology, from chemistry, geophysics to economy. In the first part of this contribution the theory underlying the nonlinear time series analysis will be described synthetically. The second part will instead be devoted to the description and application of some methods for the analyse of univariate/multivariate data sets. In particular, testing their capability to extract hidden information from these signals. Moreover, the results of the analysis of Local Field Recordings (LFP), measured from electrodes placed on the visual cortex of mice, will be discussed too.

1 Introduction

One of the first step when studying a system, of which we don't know its functioning mechanisms, is to identify some well defined observables and, in a suitable experimental setting, follow and characterize their time evolution. This is a general paradigm applicable to systems from different scientific disciplines: from physics, biology, chemistry to astronomy, from geophysics to climatology. The final result is a collection of time series, each describing the evolution over time of the values of a given observable. In principle these data contain information on the properties of the system under investigation. Thus, when analysing such time series, one of the main aim is to recover the hidden information on the basic rules governing the functioning of the system. When such task is fully accomplished, one can also try to build a mathematical model to describe and predict the evolution of the system over time.

If we are interested to access the states of the system, starting from a single time series, we should be able to reconstruct its original phase space. Generally, if our system can be described by a finite dimensional deterministic dynamical system, then it is possible to do that. This is guaranteed by specific and powerful theorems

A. Di Garbo (✉)

CNR - Institute of Biophysics, via G. Moruzzi 1, 56124 Pisa, Italy

e-mail: angelo.digarbo@pi.ibf.cnr.it

that are the theoretical background for the implementation of nonlinear times series analysis methods based on the phase space reconstruction [1, 2].

During the last years a very large number of nonlinear time series analysis methods have been proposed and an account of them can be found in [3–8]. Some of them are suitable to characterize the dynamical properties of a single signal, while others can be used to quantify the level of interdependence between two time series. In addition, yet others can also be employed to infer the coupling directionality.

Finally, it is worth mentioning, that the more recent meaning of nonlinear time series analysis also include methods not necessarily connected with phase space reconstruction. For instance the symbolic transfer entropy can be used to detect the coupling directionality, but it does not require phase space reconstruction.

In this contribution we will describe some well known nonlinear techniques for the analysis of time series. Several examples of application, either to artificial time series or physiological recordings, will be discussed too. To better appreciate the theoretical basis of the nonlinear time series analysis methods, a preliminary step will be that of introducing and briefly discussing some general properties of dynamical system theory. Indeed, as will be evident in the next sections, many nonlinear times series analysis methods are heavily based (directly or indirectly) on these basic properties of dynamical system theory. Summarizing, the main goals of the present contribution are the following: (a) to illustrate synthetically some basic notions of the dynamical system theory underlying the phase space reconstruction end embedding of time series; (b) then to show how to implement some nonlinear time series methods when working with artificial or neurophysiological data. Based on the above discussion, it is useful to organize the paper as follows: Sect. 2 will be devoted to a short introduction to dynamical system theory, Sect. 3 will be dedicated to the presentation and formulation of the embedding theorem, Sect. 4 will describe some well known methods to determine the phase space reconstruction parameters (lag time and embedding dimensions) starting from real data, in Sect. 5 a specific nonlinear prediction method will be introduced; Sect. 6 will present some methods to quantify particular geometrical and dynamical properties of attractors; lastly, in Sects. 7 and 8 some nonlinear methods to study bivariate data sets will be discussed; then the main conclusions of this contribution will be reported in Sect. 9. Moreover, Sects. 4, 5, 7, and 8 will report several results of the application of the above techniques to artificial and neurophysiological signals.

2 Dynamical Systems

2.1 Attractors

An autonomous and continuous time deterministic dynamical system can be expressed formally as a set of coupled differential equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \quad (1)$$

where $\mathbf{x} \in R^n$ and $\mathbf{x} = (x_1, x_2, \dots, x_n)$ describes the state of the system. Each \mathbf{x} is represented as a point in phase space P and during the time evolution this point describes a trajectory. The dimension of the phase space corresponds to the number of degrees of freedom of the system. In Eq. (1) the vector field $\mathbf{F}(\mathbf{x}) : R^n \rightarrow R^n$ is a mapping that is assumed to satisfy locally the Lipschitz condition. This ensures that the solution of Eq. (1) exists and is unique. Thus given an initial condition $\mathbf{x}(0) = \mathbf{x}_0$ the solution of (1), denoted here as $\mathbf{x}(t)$, will describe a well defined trajectory in the phase space. An important point is to understand the nature of the solutions asymptotically. In other words, assigned the initial condition $\mathbf{x}(0) = \mathbf{x}_0 \in P$ let us indicate with ϕ_t a one parameter family of maps (or operators) such that $\mathbf{x}(t) = \phi_t(\mathbf{x}_0)$. Then, we are interested to understand what kind of set we get as we consider the collection of all trajectories $\mathbf{x}(t) = \phi_t(\mathbf{x}_0)$, $\forall \mathbf{x}_0 \in P$, when all transient dynamics died out. If we denote such set by $\phi_t(P)$, then initially ($t = 0$) the volume of $\phi_t(P)$ is trivially that of P . If the dynamical system is conservative then the volume of the phase space is conserved and both P and $\phi_t(P)$ have the same volume. When the dynamical system is not conservative, the volume is not conserved and generally the dimension of $\phi_t(P)$ is less than that of P . In the case of a dissipative dynamical system all trajectories initiated from all possible initial conditions are attracted to a subspace of the original phase space. This subspace is called an attracting set. It is worth to introduce a more formal and precise definition of attracting set.

Definition 1 A set $A \subset P$ is called an invariant set if $\forall \mathbf{x}_0 \in A$ it is $\mathbf{x}(t) = \phi_t(\mathbf{x}_0) \in A$ for $-\infty < t < \infty$.

Let A be an invariant set and U a set such that $A \subset U$. Then, A is called an attracting set if $\forall \mathbf{x}_0 \in U$ it follows that $\lim_{t \rightarrow +\infty} \phi_t(\mathbf{x}_0) \in A$. As will be shown in the following, an attractor is a special type of attracting set. With the help of the above definitions it is possible to give an explicit description of what kind of conditions a set $A \subset P$ must satisfy in order to be an attractor.

Definition 2 A set $A \subset P$ is an attractor of the dynamical system described by Eq. (1) if:

- A is an invariant set: $\forall \mathbf{x}_0 \in A$ it is $\phi_t(\mathbf{x}_0) \in A$;
- A is an attracting set;
- for any initial conditions $\mathbf{x}_0 \in A$ the corresponding trajectory repeatedly comes close to \mathbf{x}_0 when $t \rightarrow +\infty$;
- the set A cannot be decomposed into two or more parts.

Dynamical system can have attractors constituted by a single state (point attractor) or a periodic orbit. The first attractor corresponds to a stationary solution of Eq. (1), while the second one corresponds to a dynamic attractor. Other two important attractors are quasiperiodic attractors and chaotic attractor. The set $B \subset P$ of all initial conditions \mathbf{x}_0 such that $\lim_{t \rightarrow +\infty} \phi_t(\mathbf{x}_0) \in A$ is called the basin of attraction of A .

The above concepts, discussed up to now for continuous time dynamical system, can be extended also to the case of discrete time dynamical system described by maps $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ [9].

2.2 Equivalence Class

When studying Eq. (1) it is impossible in most cases to get the explicit solution, i.e. the knowledge of the operator ϕ_t . Therefore it is important to set up specific techniques to study qualitatively the solutions of Eq. (1). Poincare' was the first who applied this point of view to the study of a dynamical systems [10]. This approach consists in the study of the geometric properties of the solutions of Eq. (1). The qualitative study of a dynamical system requires the notion of topological equivalence. As will be evident in Sect. 3, the notion of equivalence of dynamical systems plays an important role in the embedding of time series. Before proceeding it is worth to introduce some definitions.

Definition 3 Let $U \subset R^n$ be an open set and \mathbf{F} the map $\mathbf{F} : U \rightarrow R^m$. The map is of class C^k if it is k -fold continuously differentiable.

When $k = \infty$ the map is called smooth. The following definition characterizes both the invertibility and differential properties of \mathbf{F} .

Definition 4 Let \mathbf{F} an invertible C^k map, if \mathbf{F}^{-1} is also C^k then the map is a C^k -diffeomorphism.

It is useful here to introduce also the definition of the equivalence of two sets $U \subset R^n$ and $V \subset R^m$.

Definition 5 Let be $\mathbf{F} : U \rightarrow V$ with $U \subset R^n$ and $V \subset R^m$. If \mathbf{F} is a C^k -diffeomorphism then the U and V are said diffeomorphic.

Lastly, an other important definition that will be useful in the following, is that of manifold. When studying a dynamical system, as that described by Eq. (1), the corresponding asymptotic dynamic ($t \rightarrow +\infty$) can live on a subset, of the whole phase space, that is not Euclidean. A typical example is the situation in which the attractor A is a torus.

Definition 6 A set M is a m -dimensional manifold if for each $\mathbf{x}_0 \in M$ exists an open neighbourhood $W \subset M$, containing \mathbf{x}_0 , such that W is diffeomorphic to an open subset of R^m .

The above definition allows to define the differentiability in M using that of R^m . Finally, let us consider now two dynamical systems described by the following set of differential equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \quad (2)$$

$$\frac{d\mathbf{y}}{dt} = \mathbf{G}(\mathbf{y}) \quad (3)$$

where $\mathbf{x}, \mathbf{y} \in R^n$ and $\mathbf{F}(\mathbf{x}), \mathbf{G}(\mathbf{y}) : R^n \rightarrow R^n$ are C^k maps. Let be ϕ_t and ψ_t the corresponding evolution maps: i.e. $\mathbf{x}(t) = \phi_t \mathbf{x}_0$ and $\mathbf{y}(t) = \psi_t \mathbf{y}_0$. The two dynamical system are defined to be C^r ($r \leq k$) equivalent if exists a C^r diffeomorphism Φ such that $\mathbf{y} = \Phi(\mathbf{x})$. Then, generally the following relationship

$\phi_{t_1}(\mathbf{x}_0) = \psi_{t_2}(\Phi(\mathbf{x}_0))$ holds. The meaning of the above definition is that of a nonlinear change of coordinates that takes orbits of Eq. (2) into orbits of Eq. (3). The parametrization in time is different because Φ does not conserve it generally. For instance if both $\mathbf{x}(t)$ and $\mathbf{y}(t)$ describe periodic orbits their periods can be different. The above definition is very important because allows to define specific class of equivalence of dynamical systems all exhibiting the same asymptotic behaviour. In this case the qualitative study of a dynamical system can be simplified if we can identify a canonical model, of a given class of equivalent dynamical systems, of which we know its dynamical behaviour (existence of stationary state, periodic solutions or the existence of specific invariant sets).

3 Embedding of Time Series

The concept introduced in the previous section will be useful to introduce the notion of embedding and the corresponding application to time series. As it will be shown, this is achieved using the embedding theorem of Takens [1, 2]. Let us discuss first, in an intuitive way, which are the key points underlying the embedding theory. Let us consider a dynamical system described by Eq. (1) and let us consider its generic state $\mathbf{x}(t) = \phi_t(\mathbf{x}_0)$. Moreover, let us assume to perform the simultaneous measurements of some e independent observables $s_1(t), s_2(t), \dots, s_e(t)$. This measure at time t can be expressed formally as $\mathbf{s}(t) = \mathbf{V}(\mathbf{x}(t))$, where $\mathbf{s}(t) = (s_1(t), s_2(t), \dots, s_e(t))$. Then, the time evolution of the vector $\mathbf{s}(t)$ is an example of multivariate time series. The map \mathbf{V} can be thought as a nonlinear change of coordinates from the original phase space of the dynamical system to R^e . An interesting question is the following: is the information contained in $\mathbf{s}(t)$ sufficient to recover the dynamical properties associated to the dynamical system? This is a sort of inverse problem and it is very relevant in the experimental context. In most practical situations one typically performs the measurement of a single observable and thus in the following it is assumed that it is $e = 1$. When studying the dynamical system defined in Eq. (1) one is interested to study the long term behavior of the system and here we assume that all trajectories are asymptotic to an attractor $A \subset R^n$. Thus, by neglecting the transient dynamics, the properties of the systems are determined by the trajectories lying on the attractor. Let us assume here that A is a compact manifold of dimension d . This implies, for the Definition 6, that for each state $\mathbf{x}(t)$ an open neighbourhood W and an open subset of R^d exist and they are diffeomorphic. Now, let us assume that the measurements, with an assigned sampling time interval Δt_s , of the values of a single observable was performed. Then this process leads to the generation of the univariate time series $s(t_1), s(t_2), \dots, s(t_j), \dots$ where $t_j - t_{j-1} = \Delta t_s$: $s(t_j) = \mathbf{V}(\mathbf{x}(t_j))$ ($\mathbf{V} : R^n \rightarrow R$). Thus, the idea is to build, in a specific way, vectors starting from the measured values of $s(t)$, and then try to prove that the use of such vectors is sufficient to characterize the dynamical properties of the system: i.e. the set of vectors build with the values of $s(t)$ is equivalent, in the sense of Definition 5, to the set of states $\mathbf{x}(t) \in A$. The source of this idea comes from a well known

embedding theorem proved by Whitney [11]. It is worth to recall the definition of embedding.

Definition 7 Let be A a compact manifold, an embedding from A to a space U is a C^1 diffeomorphism Φ from A to $\Phi(A) \subset U$ if the derivative map $D\Phi$ (the Jacobian matrix) is one-to-one at every point $\mathbf{x} \in A$

Using the above definition the Whitney theorem asserts the following.

Theorem 1 Any compact manifold A of dimension d can be embedded in R^{2d+1} .

To proceed further let us start by defining a special set of vectors starting from the measured values of the signal $s(t)$. It is worth to remark that in the following, to simplify the notation, we will denote a generic value of time series $s(t_1), s(t_2), \dots, s(t_j), \dots$ simply by $s(t)$. But, when necessary we will recover explicitly the discrete time dependence of the values of the time series $s(t)$.

Definition 8 Let $s(t) \in R$ be a time series, then a time-delay vector (or embedding vector) is defined as $\mathbf{v}(t) = (s(t), s(t + \tau), s(t + 2\tau), \dots, s(t + (E - 1)\tau)) \in R^E$.

In the above definition $\tau = l\Delta t_s$ where l is a positive integer. Now let us assume that a map $\Omega : A \rightarrow R^E$ exists such that:

$$\mathbf{v}(t) = (s(t), s(t + \tau), s(t + 2\tau), \dots, s(t + (E - 1)\tau)) = \Omega(\mathbf{x}(t)). \quad (4)$$

It is worth to point out that there are situations for which Ω is not invertible (we will discuss this with more details below); for instance, if A is a periodic orbit of period T and $\tau = T$, then Ω^{-1} does not exist. By assuming that Ω is invertible the previous equation can be rewritten as follows

$$\mathbf{v}(t) = \Omega(\psi_t(\mathbf{x}(t_0))) = \Omega(\psi_t(\Omega^{-1}\mathbf{v}(t_0))) = \Theta_t(\mathbf{v}(t_0)), \quad (5)$$

where $\Theta_t = \Omega \psi_t \Omega^{-1}$ is the time evolution map associated to the embedding vectors in R^E . The above qualitative description, connecting the time evolution of the states in the attractor A with that of the time-delay vectors, was rigorously formulated by Takens in his well-known theorem [1].

Theorem 2 Let A be a compact C^2 manifold of dimension d , $\psi_t : A \rightarrow A$ a map and $\mathbf{V} : A \rightarrow R$. If ψ_t and \mathbf{V} are C^2 then the map $\Omega : A \rightarrow R^E$ (defined in Eq. (4)) is a diffeomorphism for almost $\tau > 0$ and $E > 2d$.

The above theorem implies that the map Ω is an embedding of the attractor of the dynamical system in the space of the time-delay embedding vectors, defined through the values of the scalar signal $s(t)$. It is worth remarking that in the above formulation of the theorem the term “for almost” should be interpreted in the sense of “genericity”: i.e. the set of all maps from A to R^E is open and dense. The meaning of the term “open” is that a small perturbation of an embedding is again an embedding, while that of “dense” is that any diffeomorphism Ω from A to R^E (regardless whether is or not an embedding) is arbitrarily close to an embedding. Some criticality of the above theorem can occur when the attractor contains periodic orbits. This was pointed out by Sauer in the generalization of Theorem 2 to attractors

with not integer dimension (like strange attractors) [2]. Let us suppose that A is a periodic orbit of period T and that we want to use the above theorem for phase space reconstruction. If we choose $\tau = T$ then the embedding vector defined in Eq. (4) collapses to the diagonal line of the embedding space and in this way the map Ω is not any more one-to-one. The same problem occurs in the case in which A is a periodic orbit of period $2T$ and $\tau = T$. To see that let us consider the following function $g(\mathbf{x}) = \mathbf{V}(\mathbf{x}(t)) - \mathbf{V}(\psi_\tau(\mathbf{x}(t))) = s(t) - s(t + \tau)$. It can be shown that this function possesses a solution: i.e. a state \mathbf{x}_0 exists such that $g(\mathbf{x}_0) = 0$. Then it will be $s(t^*) = s(t^* + \tau)$, where t^* is the time for which $\mathbf{x}(t^*) = \mathbf{x}_0$. In this case we also obtain the collapse along the diagonal line in the embedding space and Ω fails to be an embedding. For periodic orbit of period $3\tau, 4\tau, \dots$ no problems occur. The generalization of the Takens theorem allows to use this theorem in the cases in which the dimension of the attractor is not an integer (see [2] for details). In particular, the new condition for the embedding of a time series is $E > 2d_B$, where d_B is the box counting dimension of the attractor (see Sect. 6 for its definition).

Up to now we have discussed the theory without any discussion on the application of the above results to real data. For instance, the above theorems does not mention how to choose the embedding dimension or the time delay for phase space reconstruction in the case we are dealing with experimental data sets. The need to introduce specific methods in order to determine the above parameters, was clearly evident among scientists who aimed to apply the above theory to real data. In the next section we will discuss these methods together with some examples of applications of this theory.

4 Determination of Parameters for Phase Space Reconstruction

In the following section the typical approach to be applied for the reconstruction of the phase space, starting from an experimental time series, will be presented. In particular, examples of phase space reconstruction will be considered both for artificial time series and physiological signals. In particular, concerning the last type of time series, we will consider Local Field Potential (LFP) recording from visual cortex of mice. However the approach is general and can be applied to other physiological (and not physiological) recordings.

4.1 Lag Time

Let us assume to have a system of which little is known about the processes driving its dynamics. Then repeated measurements of a single observable $s(t)$ are performed with a well defined sampling intervals. The result will be the time series $s(t_1), s(t_2), \dots, s(t_j), \dots$ denoted as $\{s(j), j = 1, 2, 3, \dots, N\}$, where

$t_j - t_{j-1} = \Delta t_s$ and Δt_s is the sampling interval. The first step, for the phase space reconstruction, consists in the determination of the lag time τ . During the last years several criteria have been proposed to estimate τ [12–14]. One of the simplest method to estimate τ is based on the linear autocorrelation properties of the signal $s(i)$ defined by the autocorrelation function:

$$\rho(k) = \frac{\sum_{j=1}^{N-k} (s(i) - \bar{s})(s(i+k) - \bar{s})}{\sum_{j=1}^N (s(i) - \bar{s})^2} \quad (6)$$

where \bar{s} is the mean value of the time series and $|\rho| \leq 1$. For instance in [12] the value of the lag time was estimated as the time when the autocorrelation function of the signal decays to e^{-1} of its initial value, in another paper [14] the value of τ was determined as the time after which the autocorrelation function is not significantly different from zero. However, the above method is based on a linear approach and it is sensitive to the linear correlations alone. Therefore, a new method to estimate τ , taking into account the contribution from both the linear and nonlinear correlations, was proposed [13]. This new approach is based on the information theory and can be implemented as follows. Let $\{r(i), i = 1, 2, 3, \dots, m\}$ and $\{q(j), j = 1, 2, 3, \dots, n\}$ two finite alphabet random variables. Let us denote with $\{p_r(i), i = 1, 2, 3, \dots, m\}$, $\{p_q(j), j = 1, 2, 3, \dots, m\}$ and $\{p_{rq}(i,j), i = 1, 2, 3, \dots, m; j = 1, 2, 3, \dots, n\}$ the corresponding marginal and joint probabilities, respectively. Then, the mutual information between the two random variables is defined as

$$\psi = \sum_{i=1}^m \sum_{j=1}^n p_{rq}(i,j) \log \frac{p_{rq}(i,j)}{p_r(i)p_q(j)}. \quad (7)$$

Then, by binning the values of the time series $\{s(i), i = 1, 2, 3, \dots, N\}$ and $\{s(i+k), i = 1, 2, 3, \dots, N-k\}$ and estimating the corresponding marginal and joint probabilities, the mutual information can be computed as a function of the time delay: i.e. $\psi(k)$. In particular, this method suggests that an optimal choice for τ is the time k^* for which $\psi(k)$ gets its first local minimum. Now, as an example, we will apply the above methods to artificial time series. The first signal was a realization of an uncorrelated gaussian process of zero mean and unit standard deviation. For this signal the values of the autocorrelation function $\rho(k)$ and of $\psi(k)/\psi(0)$ are reported in the left panel of Fig. 1. As expected the values of both $\rho(k)$ and of $\psi(k)/\psi(0)$ clearly indicate that the signal does not possess any memory. The next artificial signal was generated by the Lorenz equations, a well known example of dynamical system exhibiting deterministic chaos. The corresponding equations are defined as: $dx/dt = 10(y - x)$, $dy/dt = x(28 - z) - y$, $dz/dt = xy - (8/3)z$ and they were solved by using a 4-order Runge-Kutta method with integration time step $h = 0.001$ and sampling interval $t_s = 10h$. The corresponding time series was obtained by sampling the values of the variable $x(t)$. In the right panel of Fig. 1 are reported $\rho(k)$ and $\psi(k)/\psi(0)$. The plot $\rho(k)$ suggests to use lag-time values close to $\tau = 28$ when the adopted criterion is that $\rho(k)$ decays to $1/e$ of $\rho(0)$. Instead, the mutual

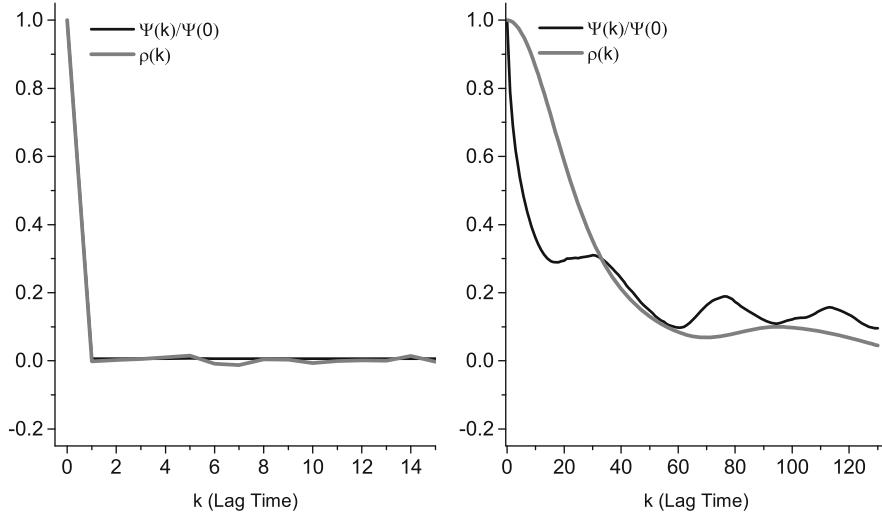


Fig. 1 Left panel: the gray (black) curve shows the values of the autocorrelation function (mutual information) for an uncorrelated Gaussian random process. Right panel: the same as for the left panel, but for a signal generated by the Lorentz equations

information indicates that the optimal value of τ , corresponding to the first minimum of the plot of $\psi(k)/\psi(0)$, is about $\tau = 18$. Lastly, the two methods to find the lag-time for phase space reconstruction were applied also to LFP recordings from the visual cortex of mouse. These signals were obtained after the injection of tetanus toxin in one of the two hemispheres of the animal. More details about the description of the experimental setting can be found in [15]. As an example in the left (right) panel of Fig. 2 are reported the corresponding plots for the LFP signal from the non injected (injected) cortex. The plot of $\rho(k)$ in the left panel indicates to choose values of τ close to 2, while those corresponding to the mutual information are close to 4. Similar results can be inferred from the data reported in the right panel.

4.2 Embedding Dimension

Having at disposal the lag time, the next step for the phase space reconstruction consists in the estimation of the value of the embedding dimension E . To this aim several methods have been proposed [3–8]. In the following the method of false nearest neighbours will be used [16]. Let us assume to have a time series $\{s(i), i = 1, 2, 3, \dots, N\}$. Then, the idea of the method is the following: let be $\mathbf{x}(t_i) = (s(t_i), s(t_i + \tau), s(t_i + 2\tau), \dots, s(t_i + (E-1)\tau))$ an embedding vector and let be $\mathbf{x}_{nn}(t_{h(i)})$ its nearest neighbour with respect to an assigned distance between vector in R^E (for instance the Euclidean distance). Here, the

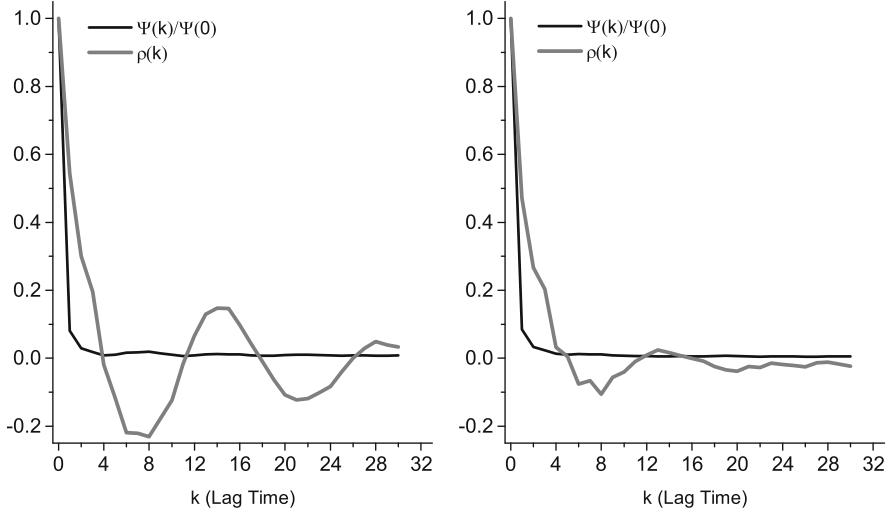


Fig. 2 Left panel: the gray (black) curve shows the values of the autocorrelation function (mutual information) for a LFP signal recorded in the non injected cortex of a mouse. Right panel: the same as for the left panel, but for a LFP signal recorded in the TeNT injected cortex of the animal

index $h(i)$ defines the position of the corresponding embedding vector in the whole set of reconstructed states. In other words, $h(i)$ identifies the index of the first component of the embedding vector $\mathbf{x}_{nn}(t_{h(i)})$: i.e. $\mathbf{x}_{nn}(t_{h(i)}) = (s(t_{h(i)}), s(t_{h(i)} + \tau), s(t_{h(i)} + 2\tau), \dots, s(t_{h(i)} + (E-1)\tau))$. Let $d_E(\mathbf{x}(i), \mathbf{x}_{nn}(h(i)))$ be their distance when the embedding dimension is E . If these two vectors are false neighbour then it is expected that, increasing the embedding dimension to $(E+1)$, the ratio of the distances will satisfy

$$d_{E+1}(\mathbf{x}(i), \mathbf{x}_{nn}(h(i)))/d_E(\mathbf{x}(i), \mathbf{x}_{nn}(h(i))) > R_{Tol} \quad (8)$$

where R_{Tol} is a threshold [16]. However, with limited amount of data, the above criterion alone may not be sufficient thus an additional criterion is added. The additional condition requires that

$$d_{E+1}(\mathbf{x}(i), \mathbf{x}_{nn}(h(i)))/R_A > A_{Tol} \quad (9)$$

where $R_A^2 = \frac{1}{N} \sum_{k=1}^N (s(k) - \bar{s})^2$ and \bar{s} is the mean value of the time series. An intuitive explanation for the introduction of this additional constraint to declare a neighbour as false is the following. Let us assume to have a false nearest neighbour $\mathbf{x}_{nn}(h(i))$ far apart from $\mathbf{x}(i)$ (in the sense that $d_E(\mathbf{x}(i), \mathbf{x}_{nn}(h(i))) \approx D_A$, where D_A is an estimation of the size of the attractor). Then, if we consider their distance in $E+1$ dimension it should be $d_{E+1}(\mathbf{x}(i), \mathbf{x}_{nn}(h(i))) > D_A$ and this leads to the condition defined in Eq.(9). In other words, if the signal is high dimensional (for instance noise) then going up of one dimension pulls the corresponding reconstructed vectors

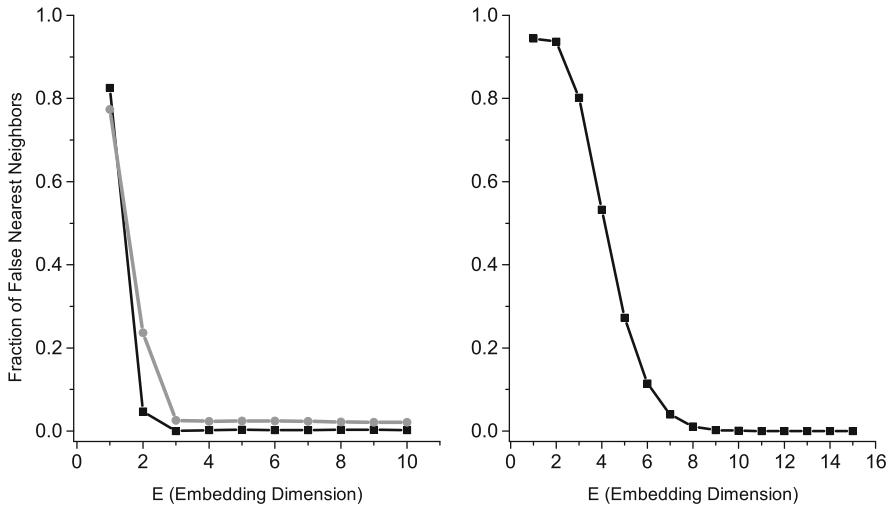


Fig. 3 Left panel: values of false nearest neighbours for a signal generated by the Lorenz equations; the gray (black) curve corresponds to a lag-time $\tau = 28$ ($\tau = 8$) obtained by using the autocorrelation function (mutual information). Right panel: values of the fraction of false nearest neighbours for an uncorrelated Gaussian random process

to the edges of the attractor. In conclusion, a pair of vectors in the embedding space will be defined false nearest neighbours if both criteria are satisfied. As an example in the left panel of Fig. 3 the fraction of the false nearest neighbours for the time series generated by the sampling of the $x(t)$ variable of the Lorenz equations is reported. The black curve refers to a values of the lag-time $\tau = 18$ obtained from the first minimum of the mutual information, while the other curve was obtained with $\tau = 28$ and corresponding to the value for which the autocorrelation decayed of e^{-1} (see right panel of Fig. 1).

Lastly, an explicit example of reconstruction of a chaotic attractor is reported in Fig. 4. On the left panel is shown the projection of the plane $(x(t), y(t))$ of the attractor of the Lorenz equations, while on the right the reconstructed projection is plotted. Their shapes are different, but as expected from the theory, a nonlinear transformation of coordinates exists linking them in a one-to-one way. Therefore, in principle, the invariant properties of the two attractors will be the same as, for instance, the whole spectrum of Lyapunov exponents.

5 Nonlinear Predictability

Among the methods of nonlinear time series analysis one of the most relevant is the nonlinear prediction method proposed in [17]. This method can be employed to estimate the level of determinism (or predictability) of a signal. When a

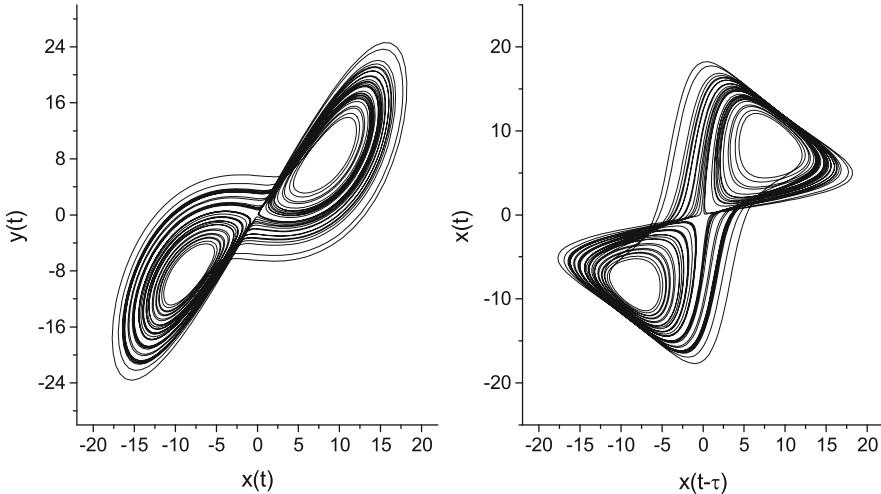


Fig. 4 Example of phase space reconstruction by using a time signal generated by the Lorenz equations. Left panel: true projection in the (x, y) plane of the attractor of the Lorenz equation. Right panel: the same as for the left panel, but using the reconstructed attractor

time series exhibits complex fluctuations this could be an indication that these random changes of the measured quantity could arise from a deterministic chaotic dynamics. Therefore, within this frame, a nonlinear prediction method is employed to quantify the level of nonlinear predictability of an assigned time series. This method is based on the phase space reconstruction and its implementation can be done as follows. Let $\{s(i), i = 1, 2, 3, \dots, N\}$ be a signal and suppose that the lag time τ and the embedding dimension E have been estimated by using the methods described in the previous section. Let us consider the generic state $\mathbf{x}(i) = [s(i), s(i - \tau), s(i - 2\tau), \dots, s(i - (E - 1)\tau)] \in R^E$ ($i = 1 + (E - 1)\tau, \dots, N$) in the reconstructed phase space. Let us now search for the first M_{nn} nearest neighbours of $\mathbf{x}(i)$ and let us denote them as $\mathbf{x}_{nn}(h_j(i))$ ($j = 1, 2, 3, \dots, M_{nn}$). Then, the above embedding vectors can be used to predict, Δ -steps in the future, the time evolution of the state $\mathbf{x}(i)$ [17]. More precisely, the nonlinear prediction is performed by using the following expression

$$\hat{\mathbf{x}}(i + \Delta) = \frac{1}{M_{nn}} \sum_{j=1}^{M_{nn}} \mathbf{x}_{nn}(h_j(i) + \Delta). \quad (10)$$

In particular, using the definition of the embedding vector introduced previously, the prediction of the $s(i + \Delta)$ value of the time series is given by $\hat{s}(i + \Delta) = \frac{1}{M_{nn}} \sum_{j=1}^{M_{nn}} s(h_j(i) + \Delta)$. Then, the overall level of determinism of the time series $s(i)$

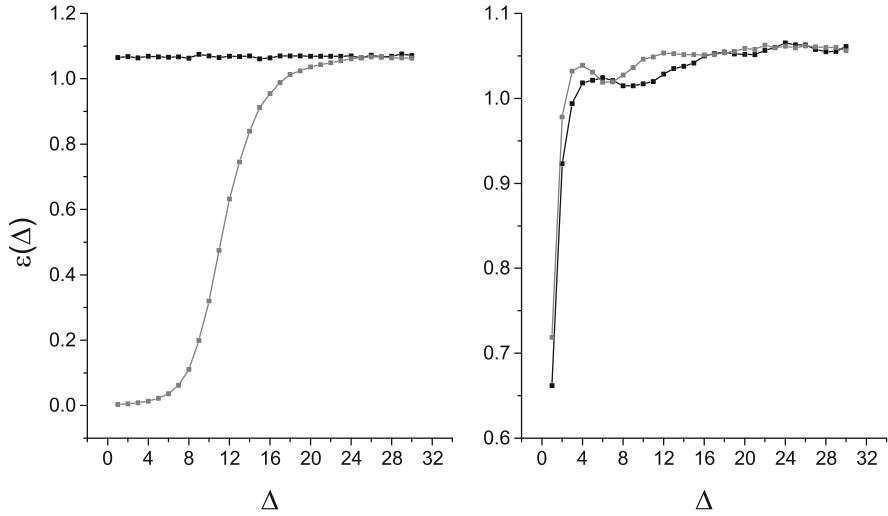


Fig. 5 Examples of nonlinear predictability. Left panel: the gray (black) curve show the values of the normalized nonlinear prediction error for a signal generated with the Henon map (uncorrelated Gaussian white noise). Right panel: the gray (black) curve show the values of the normalized nonlinear prediction error for the LFP recorded in the TeNT injected (not injected) cortex of the animal

can be quantified by using the normalized nonlinear prediction error:

$$\epsilon(\Delta) = \sqrt{\frac{\sum_{j=1+(E-1)\tau}^N (\tilde{s}(i+\Delta) - s(i+\Delta))^2}{\sum_{j=1+(E-1)\tau}^N (\bar{s} - s(i+\Delta))^2}} \quad (11)$$

where \bar{s} is the mean value of the time series. Values of $\epsilon(\Delta)$ close to zero means that the time series is well predictable Δ -steps in the future, while the quality of the prediction deteriorate as $\epsilon(\Delta)$ gets values close to 1 [18]. In Fig. 5 are reported some examples of nonlinear predictability either for artificial data and for LFP recordings. In the left panel is reported the normalized prediction errors for an uncorrelated gaussian noise (gray squares) and for the Henon map (black squares). The noise was generated by using a pseudorandom number generators, while the signal for the Henon map was generated by taking the value of x_n of the corresponding recursive map: $x_{n+1} = 1.4 - x_n^2 + 0.3y_n$, $y_{n+1} = x_n$. This map exhibit deterministic chaos and the dimension of the corresponding chaotic attractor is about 1.22. The values of $\epsilon(\Delta)$ for the gaussian noise are close to 1, meaning that the time series does not possess any deterministic character. The corresponding values for the Henon data show, correctly, that the system is driven by chaotic deterministic rules. Indeed the system is well predictable only for a short time scale (of the order of few $1/\lambda$ where λ is the maximum Lyapunov exponent) and becomes completely unpredictable for longer period of times. In the right panel the values of $\epsilon(\Delta)$ for LFP data are

reported. The black (gray) squares quantify the level of predictability for the LFP from the TeNT non injected (injected) cortex of a mice. These data show that the presence of the TeNT in the cortex where the recording was made, promotes a reduction of the determinism level (or predictability) of the corresponding signal. It is worth to remark that linear correlated gaussian stochastic processes could be predictable for a short time scale (of the order of the decorrelation time). To distinguish between linear and nonlinear predictability the method of surrogate data can be employed and additional details can be found in [19, 20].

6 Geometrical and Dynamical Characterization of Attractors

Let us assume that the time evolution of a dynamical system, after that all transient died out, is on an attractor. This could be a stationary state $\mathbf{x}(t) = \mathbf{x}_0$, a periodic or a quasiperiodic orbit or a chaotic attractor. Then, it is of interest, for modelling purposes, to characterize the dimension of these objects. For instance if the dynamics is a stationary state the corresponding topological dimension of the attractor is 0. If the attractor is a periodic orbit or a two dimensional torus the corresponding dimensions are 1 and 2, respectively. However chaotic attractors are complex geometrical objects typically characterized by fractal properties. In these case a geometrical characterization of the dimension of the attractor can be done by using a capacity (or box-counting) dimension of the attractor. This can be done as follows. Let us consider an attractor $A \subset R^n$ and cover it with cubes of edge ϵ . Let $N(\epsilon)$ be the total number of such cubes. Then, the capacity dimension of the attractor A is defined as

$$D_c = - \lim_{\epsilon \rightarrow 0} \frac{\ln N(\epsilon)}{\ln \epsilon} \quad (12)$$

obviously when the above limit exist. When the attractor is covered by geometrical objects of different size and shape then we get a measure called Hausdorff dimension (that not necessarily is coincident with the capacity dimension). However, from a practical point of view such measure is very difficult to implement especially when dealing with experimental data sets.

The above definitions are based on the idea of covering a set and then by searching for a scaling property as the edge of cubes goes to zero. But, such approach does not take into account the dynamical properties associated to the orbits, that are the skeleton of the attractor A . Thus, to consider the contribution on the estimation of the dimension coming from the dynamics, it is possible to define the generalized correlation dimension:

$$D_q = \frac{1}{q-1} \lim_{\epsilon \rightarrow 0} \frac{\ln \sum_{j=1}^N P_j^q(\epsilon)}{\ln \epsilon} \quad (13)$$

where $q > 0$ and $P_j^q(\epsilon)$ is defined as follows. Let us assume to have a finite number of state $\mathbf{x}(t_i)$ sampling the whole attractor A . Then, let us cover A with a set of N cubes (or spheres) of edge (radius) ϵ and count, for the j th cube (sphere), the number of states $\mathbf{x}(t_i)$ falling inside it. Let us denote this quantity by $n_j(\epsilon)$. Then, the probability of finding a state in the j th cube is defined as $P_j(\epsilon) = n_j(\epsilon)/N$. This last quantity is then employed to estimate the value of D_q as defined in Eq. (13). For $q = 2$ the corresponding value of D_q is equivalent to the correlation dimension of the attractor introduced by Grassberger and Procaccia [21]. This last quantity was largely employed in the analysis of time series from neuroscience, physics etc. (see [3–8] and references herein).

Deterministic chaotic systems exhibit a sensitive dependence on the initial condition: i.e. let \mathbf{x}_0 and $\mathbf{x}_0 + \delta$ ($\delta \ll 1$) be two initial conditions at $t = 0$, then in the linear approximation the average separation of the corresponding orbits is given by $\Delta\mathbf{x}(t) \approx \delta e^{\lambda_{max} t}$. In the previous expression the quantity λ_{max} is positive for chaotic systems and is called the maximum Lyapunov exponent [22]. For a n -dimensional continuous time dynamical system there are n Lyapunov exponent and, if the dynamics exhibits deterministic chaos, then almost one of the above exponent will be positive. Similar concepts can be introduced for maps. Then, this way of quantifying the instability of the orbits of a dynamical system was also applied to the study of time series. In particular, given a time series containing a sufficient large number of points it is possible to estimate the corresponding Lyapunov exponents [23]. However in many situations one is interested to estimate the maximum Lyapunov exponent and an efficient algorithm to do that was proposed in [24]. The above approach to quantify the complexity of a time series was largely employed in several contexts: form neurobiology to physics.

7 Multivariate Time Series: Quantifying the Level of Interdependence

Today, the availability of powerful scientific instrumentations permits to do the simultaneous measure of more than one observable. Thus, in many fields of scientific researches, the data collected during experiments are typically multivariate. This is particularly true in neuroscience where we can measure, at the same time, dozens of variables (for instance EEG signals or LFP multi array recordings). One important point in such cases is to employ some quantitative techniques for the estimation of the interdependence levels between all time series. In the following section some of the most employed methods that can be adopted to quantify the coupling level between a pair of signals will be discussed.

7.1 Cross Correlation

Let $x(t_i), y(t_i)$ ($i = 1, 2, \dots, N$) be two discrete signals, then the cross correlation coefficient at lag L [3, 4] is defined as:

$$\rho_{xy}(L) = \frac{1}{N} \frac{\sum_{i=0}^{N-1} (x(t_{i+L}) - \bar{x})(y(t_i) - \bar{y})}{\sigma_x \sigma_y}, \quad (14)$$

where \bar{x} (\bar{y}), σ_x (σ_y) are the mean and the standard deviation of each signal, respectively.

According to [25, 26], the linear level of coupling between the two time series can be estimated as:

$$\rho_{xy} = \frac{1}{L+1} \sum_{i=-L}^L | \rho_{xy}(L) | . \quad (15)$$

When dealing with time series with large number of points, for which nonstationarities could bias the value of the cross correlation, then a windowing of the signals is a good strategy for their analysis. More precisely, the signals are divided into N_w windows, each containing N_p data points, then the corresponding $\rho_{xy}(l)$ is estimated on the l th pair of windows. Lastly, the mean value of the set $\bar{\rho}_{xy}(l)$ ($l = 1, N_w$)

$$\bar{\rho}_{xy} = \frac{1}{N_w} \sum_{l=1}^{N_w} \bar{\rho}_{xy}(l). \quad (16)$$

is adopted as a measure of the coupling level between the two signals $x(t_i), y(t_i)$ ($i = 1, 2, \dots, N$).

7.2 Mutual Information

As well known, the cross correlation is a measure sensitive to the linear correlation alone. To quantify also the contribution coming from the nonlinear correlation then it is better to use the mutual information. This measure was encountered in Sect. 4, where it was employed for the determination of the lag-time. Let $x(t_i), y(t_i)$ ($i = 1, 2, \dots, N$) be two discrete signals, then the corresponding mutual information at time lag L can be estimated by binning the data:

$$\psi_{xy}(L) = \sum_{ij} p_{xy}(i,j) \log \frac{p_{xy}(i,j)}{p_x(i)p_y(j)}, \quad (17)$$

where $p_x(i)$ ($p_y(j)$) is the probability to find the values $x(t_{i+L})$ (($y(t_i)$) in the i th (j th) bin and $p_{xy}(i,j)$ is the corresponding joint probability [3, 4]. For long time series the problem arising from nonstationarity can be treated as for the cross correlation measure of coupling: by partitioning the signals and by averaging over the corresponding windows.

7.3 Spearman Rank Coefficient

A nonparametric method to quantify the correlation between two time series $x(t_i), y(t_i)$ ($i = 1, 2, \dots, N$) is the Spearman Rank Coefficient. Let R_i be the rank of $x(t_i)$ among the other x 's and S_i be the rank of $y(t_i)$ among the other y 's (with ties being assigned the appropriate midrank, see [27]). Then, the Spearman rank-order correlation coefficient is defined to be the linear correlation coefficient between the ranks of the two signals:

$$r_{xy} = \frac{\sum_{i=0}^N (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=0}^N (R_i - \bar{R})^2} \sqrt{\sum_{i=0}^N (S_i - \bar{S})^2}}. \quad (18)$$

In the following will be shown some examples of application of these methods to artificial data and LFP recordings. Artificial data were generated by using two different models. The first model permits of generating two gaussian random variables with an assigned level of cross correlation (here the corresponding cross correlation coefficient was 0.6). The second model consists of a pair of unidirectional coupled Henon map with a coupling coefficient of amplitude 0.5. In this case the first Henon map drives the second one and in this way the signals generated by the two maps share a common level of dynamical information. The LFP recording are those discussed in the previous section. In particular here we are interested to investigate if the presence of the TeNT toxin affects the cross talking between the two cortices. This problem was recently studied in [28] and it was shown that the presence of the toxin modulates the neural communication between opposite hemispheres. In the left panel of Fig. 6 are reported the values of the coupling between the two correlated gaussian random variables using different methods. In the right panel are shown the corresponding results for the case of coupled Henon maps. In both panels the lag for the cross correlation (see Eq. (15)) was $L = 2$. These data show that, apart from the different amplitude of the employed coupling measures, all adopted methods are able to correctly detect and quantify the level of interdependence between the time series.

Instead in Fig. 7 the results for neurophysiological recordings are reported. In particular the data refer to the measures of coupling between the LFP signals, recorded in the two opposite hemispheres, in the presence/absence of TeNT toxin. All employed methods predict that the presence of the toxin promotes reduction of the crosstalk between the two cortices with respect to the control conditions.

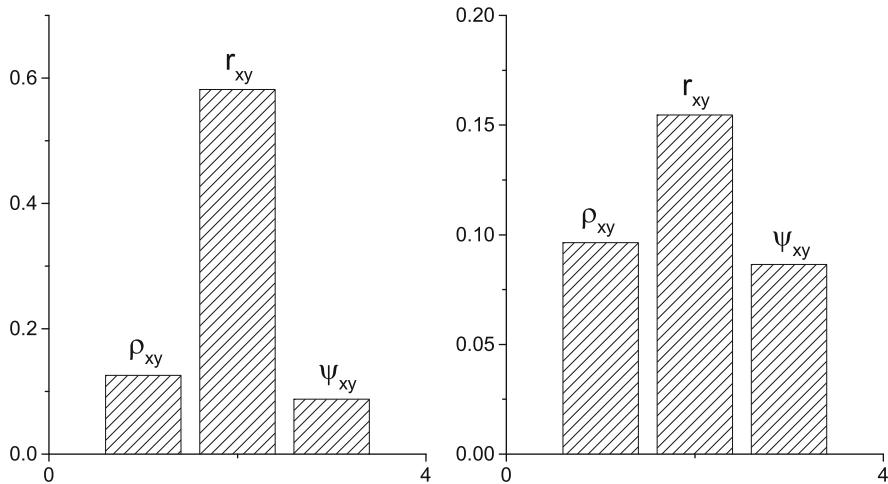


Fig. 6 Measures of the coupling level between signals. Left panel: for a pair of correlated gaussian random variables; the value of averaged cross correlation was abstained with $L = 2$; the value of the cross correlation coefficient between the two gaussian random variables (at lag 0) was 0.6. The same as for the left panel, but with signal generated by a unidirectional coupled Henon maps (with coupling amplitude equal to 0.5)

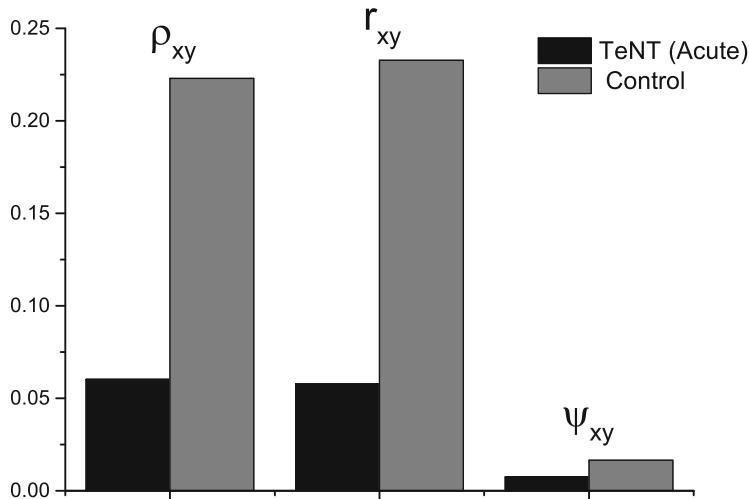


Fig. 7 Comparison of the values of the coupling level between LFP signals in the presence of TeNT toxin with those corresponding to the control

There are several methods, besides those discussed here, that can be employed to characterize the level of interdependence between time series. We don't discuss them here and detailed descriptions can be found in the literature [4, 8]. However, before concluding this section, we briefly describe another approach for inferring

the coupling level between two signals that employ the information content of the phases of the two time series. More precisely, for each signal $s(i)$ ($i = 1, 2, \dots, N$) a definition of its phase $\phi(i)$ in the i th point is introduced and then the two sets of phases are employed to quantify the coupling of the corresponding signals.

7.4 Slope Phase Coherence

This method was proposed for the analysis of neurophysiological recordings, like LFP or EEG [26]. The idea behind is that an LFP (or EEG) signal can be thought as an implicit measure of the local collective electrical activity of thousands of neurons [29]. Thus, it is reasonable that the speed of variation of the signal in time (i.e. its derivative) should contain information about the amount of synaptic intercommunication among neurons. Therefore, if $x(t_i), y(t_i)$ ($i = 1, 2, \dots, N$) are two signals, the algorithm to implement this method is the following: (1) both signals are normalized to zero mean and unit standard deviation; (2) the derivatives of both signals in the i th point, $D_i(x)$ and $D_i(y)$, are estimated by using a quadratic polynomial interpolation; (3) lastly, the phase of each signal are defined as $\phi_i(x) = \arctan(D_i(x))$ and $\phi_i(y) = \arctan(D_i(y))$ ($i = 1, 2, \dots, N$). Then, the level of coupling between the two time series is quantified by the slope phase coherence measure:

$$C_{xy} = \left| \frac{1}{N} \sum_{j=1}^N \exp i [\phi_j(x) - \phi_j(y)] \right|. \quad (19)$$

For uncorrelated signals the value of the above quantity is close to zero, while it approaches 1 when $x(t_i) = y(t_i)$, ($i = 1, 2, \dots, N$). Some examples of application of the above measure of coupling are reported in Fig. 8. In particular all time series were partitioned in half-overlapping windows, each containing $N = 5000$ data points, on which the values of the slope phase coherence were estimated. The time evolution of the values of C_{xy} either for artificial and neurophysiological signals are plotted in Fig. 8. The data in the left panel indicate that the presence of the TeNT toxin leads to a reduction of the coupling between the LFP recordings in agreement to those reported in Fig. 7. In the right panel the results obtained with artificial data are reported, showing that this coupling measure, as expected, correctly detects the level of interdependence between signals.

8 Measures of Coupling Directionality

In many experimental situations understanding the mechanism underlying the functioning of a system requires the detection not only of the interdependence level between the measured signals, but also the directionality of the corresponding

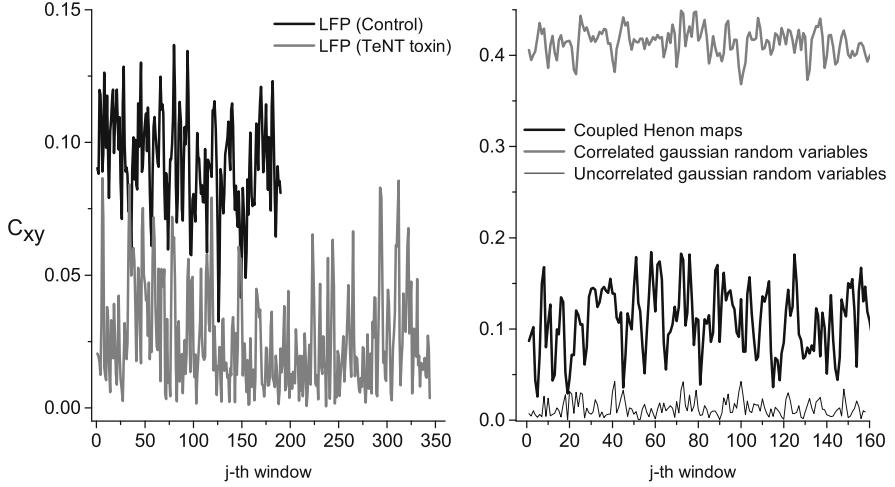


Fig. 8 Left panel: time evolution of the values of the Slope Phase Coherence between LFP signals in the presence and absence of the TeNT toxin. Right panel: the same as for the left panel, but for artificial time series

coupling. To this end during the last decades several methods have been proposed and a detailed discussion of them can be found in [4, 8]. Here we decided to present two of them largely employed in the analysis of neurophysiological recordings.

8.1 Granger Causality

In 1969 C. Granger suggested a criterion for testing the causality between time series [30]. The Granger-causality (GC) test searches for a conditional dependence between two signals. The most common implementation of GC is by linear auto-regressive modelling of time series. Let us show how the GC test is implemented starting from two discrete signals $x(t_i), y(t_i)$ ($i = 1, 2, \dots, N$). Let us assume that each time series can be modelled with an auto-regressive model

$$x(t_i) = \sum_{j=1}^p a_j^x x(t_{i-j}) + \varepsilon_i^x, \quad (20)$$

$$y(t_i) = \sum_{j=1}^p a_j^y y(t_{i-j}) + \varepsilon_i^y, \quad (21)$$

and that

$$x(t_i) = \sum_{j=1}^p a_j^{xy} x(t_{i-j}) + \sum_{j=1}^p b_j^{xy} y(t_{i-j}) + \varepsilon_i^{xy}, \quad (22)$$

$$y(t_i) = \sum_{j=1}^p a_j^{yx} y(t_{i-j}) + \sum_{j=1}^p b_j^{yx} x(t_{i-j}) + \varepsilon_i^{yx}, \quad (23)$$

where p is the order of the models and $\varepsilon_i^x, \varepsilon_i^y, \varepsilon_i^{xy}, \varepsilon_i^{yx}$ are random variables (or prediction errors). The prediction errors are assumed to be independently and identically distributed (i.i.d.) and serially uncorrelated. Then, the GC can be quantified as the ratio

$$R_G = \frac{r_{xy}}{r_{yx}}, \quad (24)$$

where $r_{xy} = \frac{\sigma_{xy}}{\sigma_x}$, $r_{yx} = \frac{\sigma_{yx}}{\sigma_y}$, and $\sigma_{xy}, \sigma_{yx}, \sigma_x, \sigma_y$ are the variances of the corresponding prediction errors. If Y drives X then $R_G < 1$, while $R_G > 1$ if X drives Y . The calculation of R_G in Eq. (24) from the data is carried out in two steps, as it was done in a previous paper [31]. First an appropriate model order (i.e. the number of model parameters) is determined to achieve the best model fit avoiding, at the same time, an over-fitting of the data (for instance by using the Akaike criterium). Then the parameters of the auto-regressive models are chosen in order to minimize the prediction errors. To this aim the ordinary least squares technique is employed.

8.2 Symbolic Transfer Entropy

The GC measure of coupling directionality assumes a linear model of the data, but this can be inappropriate when analysing data from complex system were the nonlinearity of the underlying processes can play a key role for shaping dynamical properties of the measured signals. Thus in these case it is useful to employ also a nonlinear approach to investigate coupling directionality. The symbolic transfer entropy is a nonlinear method to assess the coupling directionality between two signals $x(t_i), y(t_i)$ ($i = 1, 2, \dots, N$) [32]. Let E be the embedding dimension, τ the lag time, and $X(t_i) = [x(t_i), x(t_i + \tau), x(t_i + 2\tau), \dots, x(t_i + (E - 1)\tau)]$ the reconstructed state. Then from $X(t_i)$ let us define the new vector $\bar{X}(t_i) = [x(t_i + (m_{i1} - 1)\tau), x(t_i + (m_{i2} - 1)\tau), \dots, x(t_i + (m_{iE} - 1)\tau)]$, whose components are rearranged in increasing order. In the case of components with the same amplitude the corresponding ordering can be performed in agreement to the following rule: if $x(t_i + (m_{il} - 1)\tau) = x(t_i + (m_{ik} - 1)\tau)$, then the rearrangement is defined according to whether the value of m_{il} is $<$ (or $>$) of m_{ik} . Finally, from $\bar{X}(t_i)$ let us define the discrete symbolic vector $s_{x,i} = (m_{i1}, m_{i2}, m_{i3}, \dots, m_{iE})$, representing one of the $E!$ permutations of the first E natural numbers. The same can be done for the time series $y(t_i)$ ($i = 1, 2, \dots, N$) and then, according to [32], the symbolic transfer entropy from $Y \rightarrow X$ can be defined as

$$STE_{YX} = \sum p(s_{x,i+\delta}, s_{x,i}, s_{y,i}) \log \frac{p(s_{x,i+\delta}|s_{x,i}, s_{y,i})}{p(s_{x,i+\delta}|s_{x,i})}, \quad (25)$$

where the sum is over all possible symbols. In the above expression δ is a time step, the logarithm is with base 2, $p(s_{x,i+\delta}, s_{x,i}, s_{y,i})$ is the joint probability of occurrence

of the corresponding symbols and $p(s_{x,i+\delta}|s_{x,i}, s_{y,i})$, $p(s_{x,i+\delta}|s_{x,i})$ are conditional probabilities. By changing $s_{x,i}$ with $s_{y,i}$ and vice versa, it is possible to define STE_{XY} , quantifying the coupling level from $X \rightarrow Y$. Then, analogously to what done for the Granger causality test, the new measure of coupling directionality can be defined as

$$R_{STE} = \frac{STE_{XY}}{STE_{YX}}, \quad (26)$$

where $R_{STE} < 1$ if Y drives X and $R_g > 1$ if X drives Y . A recent example of application of the above method to neurophysiological data can be found, for instance, in [28].

9 Conclusions

In this paper we shown that there is an increasing interest to use nonlinear time series methods for the analysis of data sets. Today the term nonlinear time series analysis is employed to denote, in a wider meaning, all those methods that are intrinsically able to extract from the signals the linear and nonlinear information on the processes driving the system functioning. As discussed in the first part of this contribution, the nonlinear methods were developed starting from the theory of dynamical systems. In particular we shown that starting from a univariate time series it is possible, in principle, to reconstruct the phase space of the original system and thus to characterize its dynamical properties. Then, we demonstrated which are, in practical contexts, the main steps to achieve this goal. The reconstruction of the phase space techniques were applied to artificial data and to neurophysiological recordings, in particular LFP recorded in the visual cortex of mice in two conditions: in the presence or absence of the tetanus toxin TeNT. Then, we introduced methods for the linear and nonlinear analysis of bivariate data sets. Moreover we presented also methods able to detect the direction of the coupling between bivariate time series. The main conclusion of the present contribution is that the combined approach of both linear and nonlinear methods to the study of neurophysiological recordings allows to obtain very important information on the mechanisms driving the system functioning. This approach could be very useful to set up an appropriate mathematical model to describe and predict the time evolution of the system under study.

Acknowledgements I wish to express my thanks and gratitude to thank Dr. Eleonora Vannini end Dr. Matteo Caleo, from the Institute of Neuroscience of the National Research Council of Italy, for providing the neurophysiological recordings analysed in the present work.

References

1. Takens, F.: Detecting strange attractors in turbulence. *Lect. Notes Math.* **898**, 366–381 (1981)
2. Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. *J. Stat. Phys.* **65**, 579–616 (1991)
3. Abarbanel, H.D.I.: *Analysis of Observed Chaotic Data*. Springer, New York (1991)
4. Kantz, K., Schreiber, T.: *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge (2003)
5. Pikovsky, A., Rosenblum, M., Kurths, J.: *Synchronization. A Universal Concept in Nonlinear Sciences*. Cambridge University Press, Cambridge (2001)
6. Sprott, J.C.: *Chaos and Time-Series Analysis*. Oxford University Press, Oxford (2003)
7. Small, M.: *Applied Nonlinear Time Series Analysis*. World Scientific, Singapore (2005)
8. Bezruchko, B.P., Smirnov, D.A.: *Extracting Knowledge from Time Series*. Springer, Berlin (2010)
9. Guckenheimer, J., Holmes, P.: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, New York (1983)
10. Poincaré, H.J.: Sur le problème des trois corps et les équations de la dynamique. *Acta Math.* **13**, 1–270 (1890)
11. Whitney, H.: The self-intersections of a smooth n-manifold in 2n-space. *Ann. Math.* **45**, 220–246 (1944)
12. Zeng, X., Eykholt, R., Pielke, R.A.: Estimating the Lyapunov-exponent spectrum from short time series of low precision. *Phys. Rev. Lett.* **66**, 3229–3232 (1991)
13. Fraser, A., Swinney, H.: Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**, 1134–1140 (1986)
14. Schiff, S.J., Chang, T.: Differentiation of linearly correlated noise from chaos in a biologic system using surrogate data. *Biol. Cybern.* **67**, 387–393 (1992)
15. Vallone, F., Vannini, E., Cintio, A., Caleo, M., Di Garbo, A.: Time evolution of interhemispheric coupling in a model of focal neocortical epilepsy in mice. *Phys. Rev. E* **94**, 032409 (2016)
16. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403–3411 (1992)
17. Sugihara, G., May, R.M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**, 734–741 (1990)
18. Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845–848 (1987)
19. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D.: Testing for nonlinearity in time series: the method of surrogate data. *Phys. D* **58**, 77–94 (1992)
20. Schreiber, T., Schmitz, A.: Surrogate time series. *Phys. D* **142**, 346–382 (2000)
21. Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. *Phys. D* **9**, 189–208 (1983)
22. Eckmann, J.P., Ruelle, D.: Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985)
23. Wolf, A., Swift, J.B., Swinney, H.L., Vastano, J.A.: Determining Lyapunov exponents from a time series. *Phys. D* **16D**, 285–317 (1985)
24. Rosenstein, M.T., Collins, J.J., De Luca, C.J.: A practical method for calculating largest Lyapunov exponents from small data sets. *Phys. D* **65**, 117–134 (1993)
25. Mormann, F., Andrzejak, R.G., Kreuz, T., Rieke, C., David, P., Elger, C.E., Lehnertz, K.: Automated detection of a preseizure state based on a decrease in synchronization in intracranial electroencephalogram recordings from epilepsy patients. *Phys. Rev. E* (2003). <https://doi.org/10.1103/PhysRevE.67.021912>
26. Di Garbo, A., Mainardi, M., Chillemi, S., Maffei, L., Caleo, M.: Environmental enrichment modulates cortico-cortical interactions in the mouse. *PLoS ONE* (2011). <https://doi.org/10.1371/journal.pone.0025285>

27. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in Fortran. Cambridge University Press, Cambridge (1992)
28. Vallone, F., Vannini, E., Cintio, A., Caleo, M., Di Garbo, A.: Time evolution of interhemispheric coupling in a model of focal neocortical epilepsy in mice. *Phys. Rev. E* (2016). <https://doi.org/10.1103/PhysRevE.94.032409>
29. Buzsaki, G., Anastassiou, C.A., Koch, C.: The origin of extracellular fields and currents – EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012)
30. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969)
31. Antonucci, F., Di Garbo, A., Novelli, E., Manno, I., Sartucci, F., Bozzi, Y., Caleo, M.: Botulinum neurotoxin E (BoNT/E) reduces CA1 neuron loss and granule cell dispersion, with no effects on chronic seizures, in a mouse model of temporal lobe epilepsy. *Exp. Neurol.* **210**, 388–401 (2008)
32. Staniek, M., Lehnertz, K.: Symbolic transfer entropy. *Phys. Rev. Lett.* (2008). <https://doi.org/10.1103/PhysRevLett.100.158101>

Measures of Spike Train Synchrony and Directionality



Eero Satuvuori, Irene Malvestio, and Thomas Kreuz

Abstract Measures of spike train synchrony have become important tools in both experimental and theoretical neuroscience. Three time-resolved measures called the ISI-distance, the SPIKE-distance, and SPIKE-synchronization have already been successfully applied in many different contexts. These measures are time scale independent, since they consider all time scales as equally important. However, in real data one is typically less interested in the smallest time scales and a more adaptive approach is needed. Therefore, in the first part of this Chapter we describe recently introduced generalizations of the three measures, that gradually disregard differences in smaller time-scales. Besides similarity, another very relevant property of spike trains is the temporal order of spikes. In the second part of this chapter we address this property and describe a very recently proposed algorithm, which quantifies the directionality within a set of spike train. This multivariate approach sorts multiple spike trains from leader to follower and quantifies the consistency of the propagation patterns. Finally, all measures described in this chapter are freely available for download.

E. Satuvuori (✉)

Department of Physics and Astronomy, University of Florence, Sesto Fiorentino, Italy

Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

e-mail: eero.satuvuori@unifi.it

I. Malvestio

Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

Department of Physics and Astronomy, University of Florence, Sesto Fiorentino, Italy
e-mail: irene.malvestio@upf.edu

T. Kreuz

Institute for Complex Systems, CNR, Sesto Fiorentino, Italy
e-mail: thomas.kreuz@cnr.it

1 Introduction

The brain can be considered as a huge network of spiking neurons. It is typically assumed that only the spikes, and not the shape of the action potential nor the background activity, convey the information processed within this network [11]. Sequences of consecutive spikes are called spike trains. Measures of spike train synchrony are estimators of the similarity between two or more spike trains, which are important tools for many applications in neuroscience. Among others, they allow to test the performance of neuronal models [10], they can be used to quantify the reliability of neuronal responses upon repeated presentations of a stimulus [18], and they help in the understanding of neural networks and neural coding [28].

Over the years many different methods have been developed in order to quantify spike train synchrony. They can be divided in two classes: time-scale dependent and time-scale independent methods. The two most known time-scale dependent methods are the Victor-Purpura distance [29] and the van Rossum distance [26]. They describe spike train (dis)similarity based on a user-given time-scale to which the measures are mainly sensitive to. Time scale independent methods have been developed more recently. In particular, the ISI-distance [13], the SPIKE-distance [14, 15] and SPIKE-synchronization [16] are parameter-free distances, with the capability of discerning similarity across different spatial scales. All of these measures are time-resolved, so they are able to analyze the time dependence of spike train similarity.

One problematic aspect of time-scale independent methods is that they consider all time-scales as equally important. However, in real data one typically is not interested in the very small time scales. Especially in the presence of bursts (multiple spikes emitted in rapid succession), a more adaptive approach that gradually disregards differences in smaller time-scales is needed. Thus, in the first part of this chapter we describe the recently developed adaptive extensions of these three parameter-free distances: A-ISI-distance, A-SPIKE-distance and A-SPIKE-synchronization [24].

All of these similarity measures are symmetric and in consequence invariant to changes in the order of spike trains. However, often information about directionality is needed, in particular in the study of propagation phenomena. For example, in epilepsy studies, the analysis of the varying similarity patterns of simultaneously recorded ensembles of neurons can lead to a better understanding of the mechanisms of seizure generation, propagation, and termination [4, 25].

In the second part of this chapter we address the question: Which are the neurons that tend to fire first, and which are the ones that tend to fire last? We present SPIKE-Order [17], a recently developed algorithm which is able to discern propagation pattern in neuronal data. It is a multivariate approach which allows to sort multiple spike trains from leader to follower and to quantify the consistency of the temporal leader-follower relationships. We close this chapter by describing some applications of the methods presented.

2 Measures of Spike Train Synchrony

Two of the most well known spike train distances, the Victor-Purpura [29] and the van Rossum distance [26], are time-scale dependent. One drawback of these methods is the fixed time-scale, since it sets a boundary between rate and time coding for the whole recording. In the presence of bursts, where multiple spikes are emitted in rapid succession, there are usually many time-scales in the data and this is difficult to detect when using a measure that is sensitive to only one time-scale at a time [6].

The problem of having to choose one time-scale has been eliminated in the time-scale independent ISI-distance [13], SPIKE-distance [14, 15] and SPIKE-synchronization [16], since these methods always adapt to the local firing rate. The ISI-distance and the SPIKE-distance are time resolved, time-scale free measures of dissimilarity between two or more spike trains. The ISI-distance is a measure of rate dissimilarity. It uses the interspike intervals (ISIs) to estimate local firing rate of spike trains and measures time-resolved differences between them. The SPIKE-distance, on the other hand, compares spike time accuracy between the spike trains and uses local firing rates to adapt to the time-scale. SPIKE-synchronization is also time-scale free and is a discrete time resolved measure of similarity based on ISI derived coincidence windows that determine if two spikes in a spike train set are coincident or not.

The ISI-distance, SPIKE-distance, and SPIKE-synchronization are looking at all time-scales at the same time. However, in real data not all time-scales are equally important, and this can lead to spuriously high values of dissimilarity when looking only at the local information. Many sequences of discrete events contain different time-scales. For example, in neuronal recordings besides regular spiking one often finds bursts, i.e., rapid successions of many spikes. The A-ISI-distance, A-SPIKE-distance and A-SPIKE-synchronization [24] are generalized versions of previously published methods the ISI-distance [12], SPIKE-distance [14] and SPIKE-synchronization [16]. The generalized measures also contain a notion of global context that discriminates between relative importance of differences in the global scale. This is done by means of a normalization based on a minimum relevant time-scale (MRTS). They start to gradually ignore differences between spike trains for interspike intervals (ISIs) that are smaller than the MRTS. The generalization provided by the MRTS is implemented with the threshold parameter thr , which is then applied in a different way to each of the measures. The threshold is used to determine if a difference between the spike trains should be assessed in a local or in a global context. This threshold is used for all three measures, but the way it is applied varies. The extended methods fall back to the original definitions when $thr = 0$ and we refer to this whenever we talk of the original methods. In this case even the smallest time-scales matter and all differences are assessed in relation to their local context only.

Throughout this Section we denote the number of spike trains by N , indices of spike trains by n and m , spike indexes by i and j and the number of spikes in spike train n by M_n . The spike times of spike train n are denoted by $\{t_i^{(n)}\}$ with $i = 1 \dots M_n$.

2.1 Adaptive ISI-Distance

The A-ISI-distance [24] measures the instantaneous rate difference between spike trains (see Fig. 1a). It relies on a time-resolved profile, meaning that in a first step

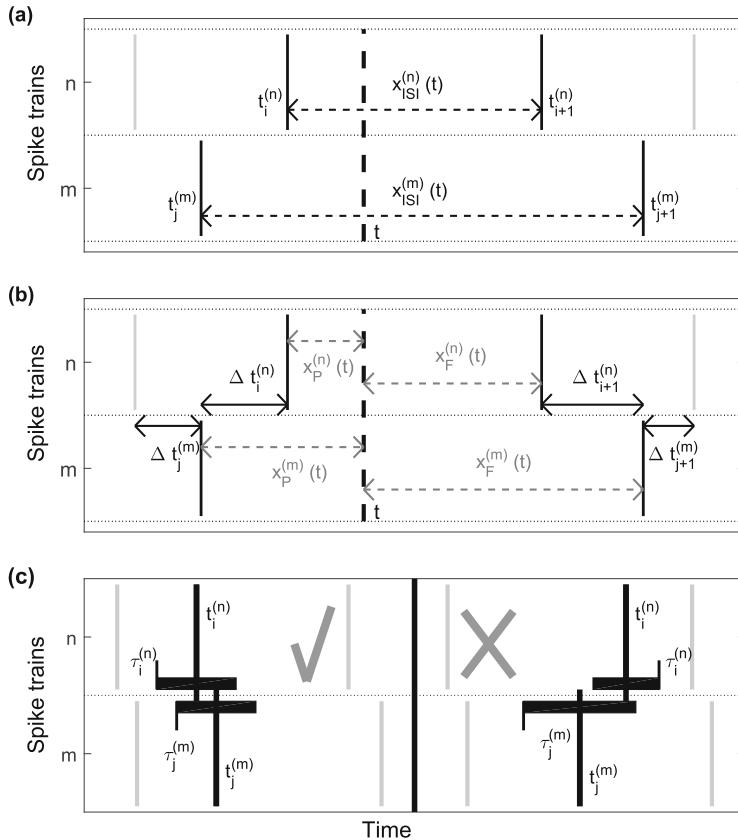


Fig. 1 Schematic drawing for all three measures. (a) Illustration of the variables used to define the **A-ISI-distance**. All measures use the instantaneous interspike interval $x_{ISI}^{(n)}(t)$ to adapt to the local firing rate. (b) Additional variables used for the **A-SPIKE-distance**. (c) Coincidence criterion for the **A-SPIKE-synchronization**. The coincidence window of each spike is derived from the two surrounding interspike intervals. For simplicity the $thr = 0$ case is shown. Here we illustrate two different examples. The two spikes on the left side are considered coincident since both lie in each other's coincidence windows. On the right there is no coincidence because the spike from the second spike train is outside of the coincidence window from the spike of the first spike train

a dissimilarity value is assigned to each time instant. To obtain this profile, we first assign to each time instant t the times of the previous spike and the following spike

$$t_p^{(n)}(t) = \max\{t_i^{(n)} | t_i^{(n)} \leq t\} \quad \text{for } t_1^{(n)} \leq t \leq t_{M_n}^{(n)} \quad (1)$$

$$t_f^{(n)}(t) = \min\{t_i^{(n)} | t_i^{(n)} > t\} \quad \text{for } t_1^{(n)} \leq t \leq t_{M_n}^{(n)}. \quad (2)$$

From this for each spike train n an instantaneous ISI can be calculated as

$$x_{\text{ISI}}^{(n)}(t) = t_f^{(n)}(t) - t_p^{(n)}(t). \quad (3)$$

The A-ISI-profile is defined as a normalized instantaneous ratio in ISIs:

$$I_{n,m}^A(t) = \frac{|x_{\text{ISI}}^{(n)}(t) - x_{\text{ISI}}^{(m)}(t)|}{\max\{x_{\text{ISI}}^{(n)}(t), x_{\text{ISI}}^{(m)}(t), thr\}}. \quad (4)$$

For the A-ISI-distance the MRTS is defined so that when the ISI of both spike trains are smaller than a threshold value thr , the threshold value is used instead. The multivariate A-ISI-profile is obtained by averaging over all pairwise A-ISI-profiles:

$$I^A(t) = \frac{1}{N(N-1)/2} \sum_{n=1}^{N-1} \sum_{m=n+1}^N I_{n,m}^A(t). \quad (5)$$

This is a non-continuous piecewise constant profile and integrating over time gives the A-ISI-distance:

$$D_I^A = \frac{1}{t_e - t_s} \int_{t_s}^{t_e} I^A(t) dt. \quad (6)$$

Here t_s and t_e are the start and end times of the recording respectively. If thr is set to zero, the method falls back to the ISI-distance [12].

Figure 2a shows an artificial spike train dataset together with the corresponding A-ISI-profile in Fig. 2b. The A-ISI-profile for the example dataset shows high dissimilarity for the left side of the raster plot, where noise is high. When the noise is decreased and rates become more similar in the right side, the dissimilarity profile goes down. The overall ISI-distance is the mean value of the profile.

2.2 Adaptive SPIKE-Distance

The A-SPIKE-distance [24] measures the accuracy of spike times between spike trains relative to local firing rates (see Fig. 1b). In order to assess the accuracy of

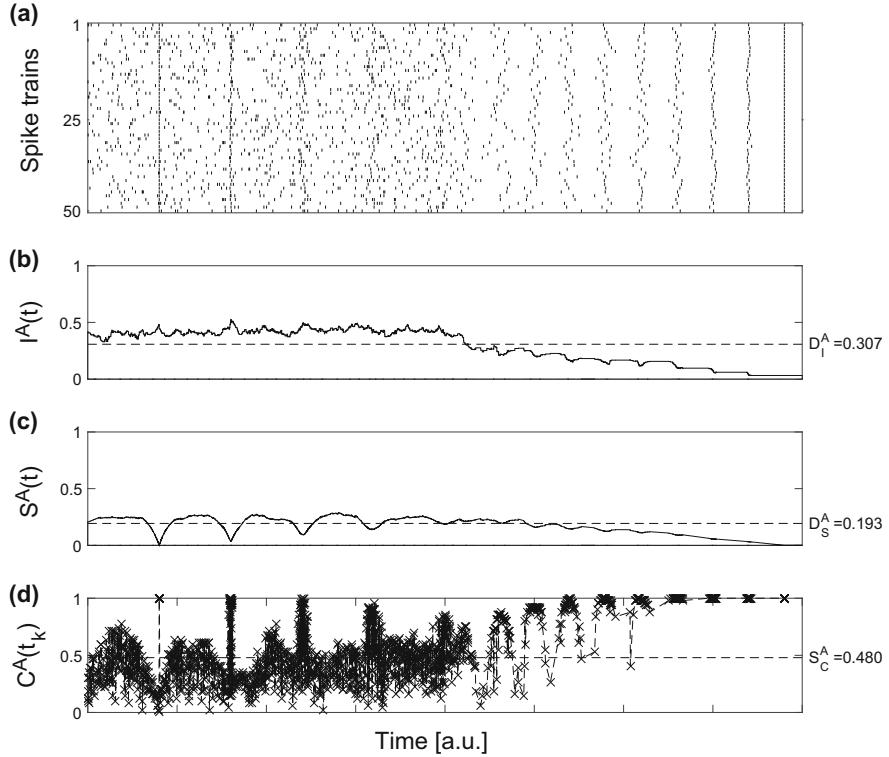


Fig. 2 Profiles of A-ISI-distance (a), A-SPIKE-distance (b) and A-SPIKE-synchronization (c) for an artificial example dataset of 50 spike trains with population events with different jitters and decreasing noise over time

spike events, each spike is assigned a distance to its nearest neighbour in the other spike train:

$$\Delta t_i^{(n)} = \min_j (|t_i^{(n)} - t_j^{(m)}|). \quad (7)$$

The distances are interpolated between spikes using for all times t the time differences to the previous and to the following spikes $x_P^{(n)}(t)$ and $x_F^{(n)}(t)$:

$$x_P^{(n)}(t) = t - t_i^{(n)} \quad \text{for} \quad t_i^{(n)} \leq t \leq t_{i+1}^{(n)} \quad (8)$$

$$x_F^{(n)}(t) = t_{i+1}^{(n)} - t \quad \text{for} \quad t_i^{(n)} \leq t \leq t_{i+1}^{(n)}. \quad (9)$$

These equations provide time-resolved quantities needed to define time-resolved dissimilarity profile from discrete values the same way as Eqs. (1) and (2) provide them for A-ISI-distance. The weighted spike time difference for a spike train is then

calculated as an interpolation from one difference to the next by

$$S_n(t) = \frac{\Delta t_i^{(n)}(t)x_F^{(n)}(t) + \Delta t_{i+1}^{(n)}(t)x_P^{(n)}(t)}{x_{\text{ISI}}^{(n)}(t)} \quad , \quad t_i^{(n)} \leq t \leq t_{i+1}^{(n)}. \quad (10)$$

This continuous function is analogous to term $x_{\text{ISI}}^{(n)}$ for the ISI-distance, except that it is piecewise linear instead of piecewise constant. The pairwise A-SPIKE-distance profile is calculated by temporally averaging the weighted spike time differences, normalizing to the local firing rate average and, finally, weighting each profile by the instantaneous firing rates of the two spike trains:

$$S_{m,n}^A(t) = \frac{S_n x_{\text{ISI}}^m(t) + S_m x_{\text{ISI}}^n(t)}{2\langle x_{\text{ISI}}^{n,m}(t) \rangle \max\{\langle x_{\text{ISI}}^{n,m}(t) \rangle, thr\}}, \quad (11)$$

where $\langle x_{\text{ISI}}^{n,m}(t) \rangle$ is the mean over the two instantaneous ISIs. MRTS is defined by using a threshold, that replaces the denominator of weighting to spike time differences if the mean is smaller than the thr . This profile is analogous to the pairwise A-ISI-profile $I_{n,m}^A(t)$, but again it is piecewise linear, not piecewise constant. Unlike $S_n(t)$ it is not continuous, typically it exhibits instantaneous jumps at the times of the spikes. The multivariate A-SPIKE-profile is obtained the same way as the multivariate A-ISI-profile, by averaging over all pairwise profiles:

$$S^A(t) = \frac{1}{N(N-1)/2} \sum_{n=1}^{N-1} \sum_{m=n+1}^N S_{m,n}^A(t). \quad (12)$$

The final A-SPIKE-distance is calculated as the time integral over the multivariate A-SPIKE-profile the same way as the A-ISI-distance:

$$D_S^A = \frac{1}{t_e - t_s} \int_{t_s}^{t_e} S^A(t) dt. \quad (13)$$

The effect of applying the threshold can be seen in Fig. 3. With $thr = 0$ the method falls back to the regular SPIKE-distance [14]. The A-SPIKE-profile for the artificial test dataset in Fig. 2c goes to zero when the spikes in all spike trains appear exactly at the same time.

2.3 Adaptive SPIKE-Synchronization

A-SPIKE-synchronization [24] quantifies how many of the possible coincidences in a dataset are actually coincidences (Fig. 1c). While the A-ISI-distance and the A-SPIKE-distance are measures of dissimilarity which obtain low values for similar

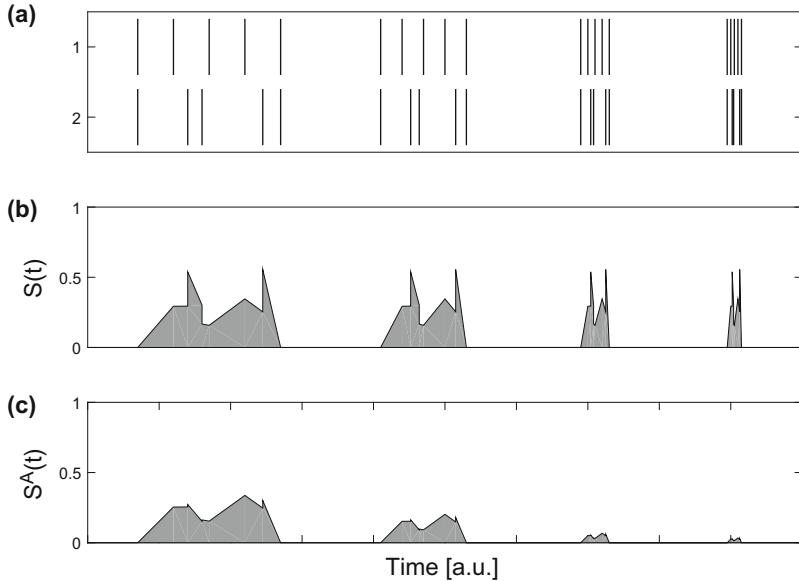


Fig. 3 An example spike train pair and its SPIKE-distance and A-SPIKE-distance profiles. **(a)** Two spike trains having four events with five spikes per event in each spike train. The sequence of spikes in all four events is the same but the event is increasingly compressed. The only thing that changes is the time-scale. From a global perspective the first event consists of non-synchronous individual spikes, while the last event consists of coincident bursts. The two events in the middle are intermediates. **(b)** The SPIKE-distance $S(t)$ looks only at the local context and has the same profile shape for all events. **(c)** The A-SPIKE-distance considers also the global context and judges the first event like the SPIKE-distance as being dissimilar, but scales down the small spike time differences in the burst and considers the coincident burst as very similar

spike trains, A-SPIKE-synchronization measures similarity. If all the spikes are coincident with a spike in all the other spike trains, the value will be one. In contrast, if none of the spikes are coincident, it will be zero.

The original SPIKE-synchronization [16] is parameter- and time-scale-free, since it uses the adaptive coincidence detection which was first proposed for the measure Event synchronization [22]. The coincidence window, i.e., the time lag below which two spikes from two different spike trains, $t_i^{(n)}$ and $t_j^{(m)}$, are considered to be coincident, is adapted to the local firing rate. Spikes are coincident only if they both lie in each others coincidence windows. A-SPIKE-synchronization is a generalized version of the SPIKE-synchronization. The MRTS is used to decide if the window is determined locally or if the global context should be taken into account.

As a first step, we define the ISI before and after the spike as

$$x_{iP}^{(n)} = \lim_{t \rightarrow t_i^-} x_{\text{ISI}}^{(n)}(t) \quad (14)$$

$$x_{iF}^{(n)} = \lim_{t \rightarrow t_i^+} x_{\text{ISI}}^{(n)}(t). \quad (15)$$

The coincidence window for spike i of spike train n is defined by determining a minimum coincidence window size for a spike as half of the ISIs adjacent to the spike and allowing asymmetric coincidence windows based on MRTS. This is done by using thr instead of the minimum, if it is smaller. Since the threshold value is based on ISIs and the coincidence window spans both sides of the spike, only half of the threshold spans each side. For the A-ISI- and the A-SPIKE-distance the changes induced by the threshold appear gradually, but for A-SPIKE-synchronization it is a sudden change from a non-coincidence to coincidence for a spike. Therefore, due to the binary nature of A-SPIKE-synchronization, the threshold is additionally divided by two. The coincidence window is not allowed to overlap with a coincidence window of another spike and is thus limited to half the ISI even if the threshold is larger. The base of the window is defined by the two adjacent ISIs:

$$\tau_i^{(n)} = \frac{1}{2} \min\{x_{iP}^{(n)}, x_{iF}^{(n)}\}. \quad (16)$$

The coincidence window of a spike is then defined in an asymmetric form by using the coincidence window part before and after the spike as

$$\tau_{iP}^{(n)} = \min\{\max(\frac{1}{4}thr, \tau_i^{(n)}), \frac{1}{2}x_{iP}^{(n)}\} \quad (17)$$

$$\tau_{iF}^{(n)} = \min\{\max(\frac{1}{4}thr, \tau_i^{(n)}), \frac{1}{2}x_{iF}^{(n)}\}. \quad (18)$$

The combined coincidence window for spikes i and j is then defined as

$$\tau_{ij}^{(n,m)} = \begin{cases} \min\{\tau_{iF}^{(n)}, \tau_{jP}^{(m)}\} & \text{if } t_i \leq t_j \\ \min\{\tau_{iP}^{(n)}, \tau_{jF}^{(m)}\} & \text{otherwise} \end{cases}. \quad (19)$$

The coincidence criterion can be quantified by means of a coincidence indicator

$$C_i^{(n,m)} = \begin{cases} 1 & \text{if } \min_j \{|t_i^{(n)} - t_j^{(m)}|\} < \tau_{ij}^{(n,m)} \\ 0 & \text{otherwise} \end{cases}. \quad (20)$$

This definition ensures that each spike can only be coincident with at most one spike in the other spike train. The coincidence criterion assigns either a one or a zero to each spike depending on whether it is part of a coincidence or not. For each spike

of every spike train, a normalized coincidence counter

$$C_i^{(n)} = \frac{1}{N-1} \sum_{m \neq n} C_i^{(n,m)} \quad (21)$$

is obtained by averaging over all $N - 1$ bivariate coincidence indicators involving the spike i in spike train n .

This way we have defined a coincidence indicator for each individual spike in the spike trains. In order to obtain one combined similarity profile, we pool the spikes of the spike trains as well as their coincidence indicators by introducing one overall spike index k and defining

$$M = \sum_{n=1}^N M_n. \quad (22)$$

This yields one unified set of coincidence indicators C_k :

$$\{C_k\} = \bigcup_n \{C_i^{(n)}\}. \quad (23)$$

From this discrete set of coincidence indicators C_k the A-SPIKE-synchronization profile $C^A(t_k)$ is obtained by $C^A(t_k) = C_k$.

Finally, A-SPIKE-synchronization is defined as the average value of this profile

$$S_C^A = \frac{1}{M} \sum_{k=1}^M C^A(t_k). \quad (24)$$

It is important to note that since A-SPIKE-synchronization is a measure of similarity, reducing differences below threshold adds coincidences and thus the value obtained increases. In Fig. 4 we illustrate how the asymmetric coincidence windows of A-SPIKE-synchronization allow for a better coverage of burst events which makes it easier to match spikes when compared to SPIKE-synchronization ($thr = 0$) [16].

As can be seen in Fig. 2d, the A-SPIKE-synchronization profile is discrete and only defined at spike times. A dotted line between the points is added as visual aid. The profile gets higher values the more coincidences are found for each spike in other spike trains.

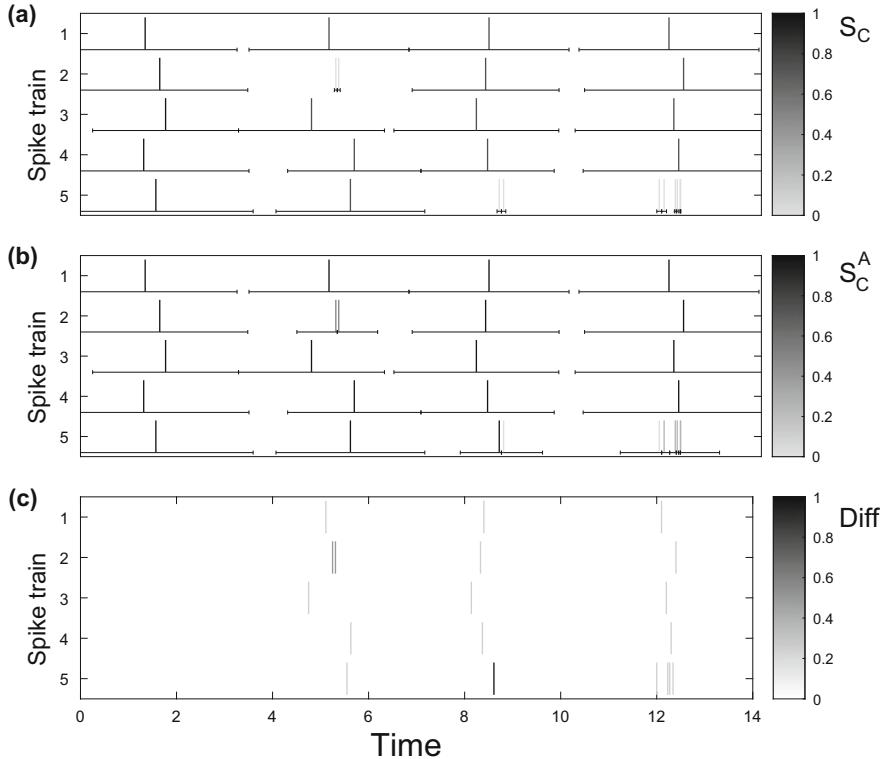


Fig. 4 SPIKE-synchronization, A-SPIKE-synchronization and their difference illustrated using five spike trains with four simple events. Without the correction (a) in case of spike doublets (first and second event) or bursts (third event) the small interspike intervals result in an unreasonably high demand for spike timing accuracy. With the adaptive correction (b) for all these cases the likelihood increases that at least one of the spikes is part of a coincidence. On the other hand, if there are no doublets or bursts (last event), nothing changes (best seen in c). Note that for better visibility the colour scales differ, we use grey-black in a and b and white-black in c

2.4 Selecting the Threshold Value

In some cases spikes that occur less than a second apart might be considered more simultaneous than those taking place within minutes, and in applications like meteorological systems, weeks instead of months. Setting the minimum relevant time-scale might not be a simple task. If no information of the system producing the spikes is available, all one can do to estimate an appropriate threshold value is to look at the ISIs.

There are two criteria that a threshold value extracted from the data has to fulfil. First of all it needs to adapt to changes in spike count so that adding more spikes gives shorter threshold. Additionally we want the threshold to adapt to changes in the ISI-distribution when the spike count is fixed. The more pronounced bursts are

found in the data, the more likely any differences within aligned bursts are not as important as their placement. Thus, we want our threshold to get longer if the spikes are packed together. To do so, all the ISIs in all spike trains are pooled and the threshold is determined from the pooled ISI distribution.

One should not just take a value based on ISI-distribution that counts the interspike intervals, as the mean does, but weight them by their length, which is equivalent to taking the average of the second moments of ISIs. Doing this reduces the importance of very short ISIs even if they are statistically much more common. In order to obtain a value with the right dimension, the square root of the average must be taken:

$$thr = \sqrt{\langle (L_{\text{ISI}})^2 \rangle} = \sqrt{\frac{\sum_{n=1}^N a_n L_{\text{ISI}}^n}{\sum_{n=1}^N a_n}}. \quad (25)$$

Here we denoted a single ISI length in the pooled distribution as L_{ISI}^n and the number of ISI with length L_{ISI}^n as a_n . It is important to note, however, that this is only an estimate based on different time-scales found in the data. The selected MRTS is not an indicator of a time-scale of the system that produced the spikes.

As an example of how the threshold works we apply the threshold to Gamma $\Gamma(k, x)$ distribution. Since the kurtosis of the distribution is proportional to $1/k$, for small k the distribution contains large number of small ISI and few long ones. This is the property the threshold is tracking. The mean of a gamma distribution is k/x and the second moment $(k+1)k/x$. thus the ratio of the threshold and the mean ISI is $thr/\langle L_{\text{ISI}} \rangle = \sqrt{x(k+1)/k}$. From the formula we can see that for small k , where the distribution is more skewed, the ratio between the mean ISI and the threshold increases. This means that mainly the rare and large inter-burst ISIs are taken into account.

The threshold value determines the outcome of the adaptive methods. However, the threshold is not a hard set limiter neglecting everything below the threshold, but rather the point from which on differences are considered in a global instead of a local context.

3 Measures of Spike Train Directionality

Often a set of spike trains exhibits well-defined patterns of spatio-temporal propagation where some prominent feature first appears at a specific location and then spreads to other areas until potentially becoming a global event. If a set of spike trains exhibits perfectly consistent repetitions of the same global propagation pattern, this can be called a *synfire pattern*. For any spike train set exhibiting propagation patterns the questions arises naturally whether these patterns show any consistency, i.e., to what extent do the spike trains resemble a synfire pattern, are

there spike trains that consistently lead global events and are there other spike trains that invariably follow these leaders?

In the second part of this chapter we describe a framework consisting of two directional measures (*SPIKE-Order* and *Spike Train Order*) that allows to define a value termed *Synfire Indicator* which quantifies the consistency of the leader-follower relationships [17]. This Synfire Indicator attains its maximal value of 1 for a perfect synfire pattern in which all neurons fire repeatedly in a consistent order from leader to follower (Fig. 5a).

The same framework also allows to sort multiple spike trains from leader to follower, as illustrated in Fig. 5b, c. This is meant purely in the sense of temporal sequence. Whereas Fig. 5b shows an artificially created but rather realistic spike train set, in Fig. 5c the same spike trains have been sorted to become as close as possible to a synfire pattern. Now the spike trains that tend to fire first are on top whereas spike trains with predominantly trailing spikes are at the bottom.

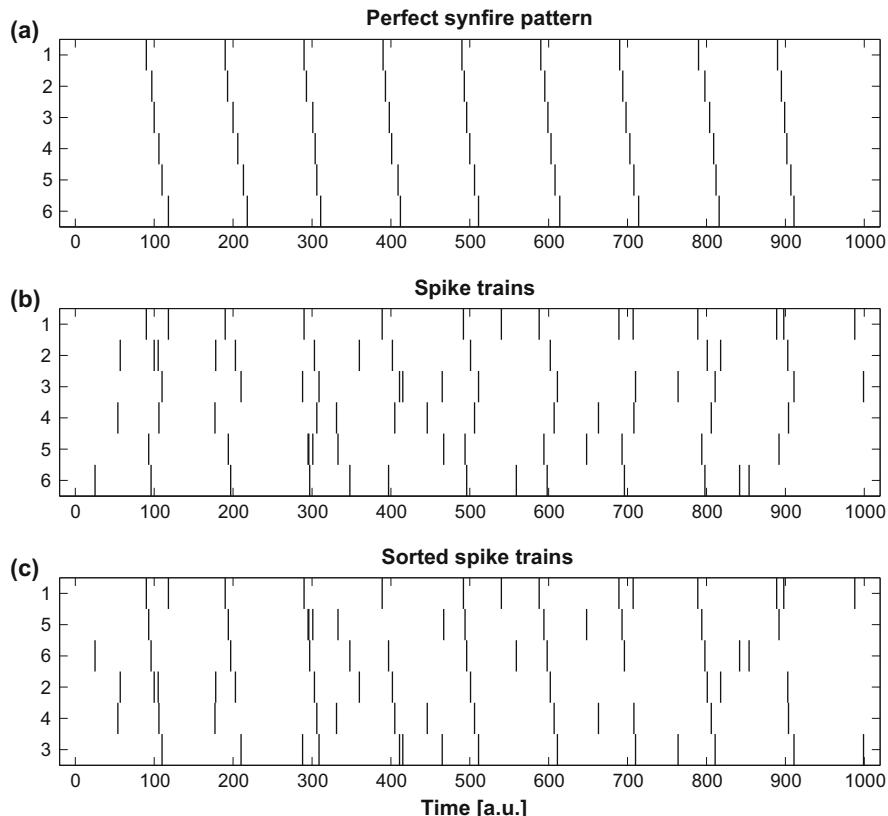


Fig. 5 Motivation for SPIKE-order and Spike Train Order. (a) Perfect Synfire pattern. (b) Unsorted set of spike trains. (c) The same spike trains as in (b) but now sorted from leader to follower

Analyzing leader-follower relationships in a spike train set requires a criterion that determines which spikes should be compared against each other. What is needed is a match maker, a method which pairs spikes in such a way that each spike is matched with at most one spike in each of the other spike trains. This match maker already exists. It is the adaptive coincidence detection first used as the fundamental ingredient for the bivariate measure *SPIKE-synchronization* [16] (see Sect. 2.3).

3.1 SPIKE-Order and Spike Train Order

The symmetric measure SPIKE-Synchronization (introduced in Sect. 2.3) assigns to each spike of a given spike train pair a bivariate coincidence indicator. These coincidence indicators $C_i^{(n,m)}$ (Eq. (20)), which are either 0 or 1, are then averaged over spike train pairs and converted into one overall profile $C(t_k)$ normalized between 0 and 1. In exactly the same manner SPIKE-Order and Spike Train Order assign bivariate order indicators to spikes. Also these two order indicators, the asymmetric $D_i^{(n,m)}$ and the symmetric $E_i^{(n,m)}$, which both can take the values -1 , 0 , or $+1$, are averaged over spike train pairs and converted into two overall profiles $D(t_k)$ and $E(t_k)$ which are normalized between -1 and 1 . The SPIKE-Order profile $D(t_k)$ distinguishes leading and following spikes, whereas the Spike Train Order profile $E(t_k)$ provides information about the order of spike trains, i.e. it allows to sort spike trains from leaders to followers.

First of all, the symmetric coincidence indicator $C_i^{(n,m)}$ of SPIKE-Synchronization (Eq. (20)) is replaced by the asymmetric SPIKE-Order indicator

$$D_i^{(n,m)} = C_i^{(n,m)} \cdot \text{sign}(t_j^{(m)} - t_i^{(n)}), \quad (26)$$

where the index j is defined from the minimum in Eq. (20) with the threshold-value in Eqs. (17) and (18) set to $thr = 0$.

The corresponding value $D_j^{(m,n)}$ is obtained in an asymmetric manner as

$$D_j^{(m,n)} = C_j^{(m,n)} \cdot \text{sign}(t_i^{(n)} - t_j^{(m)}) = -D_i^{(n,m)}. \quad (27)$$

Therefore, this indicator assigns to each spike either a 1 or a -1 depending on whether the respective spike is leading or following a coincident spike from the other spike train. The value 0 is obtained for cases in which there is no coincident spike in the other spike train ($C_i^{(n,m)} = 0$), but also in cases in which the times of the two coincident spikes are absolutely identical ($t_j^{(m)} = t_i^{(n)}$).

The multivariate profile $D(t_k)$ obtained analogously to Eq. (23) is normalized between 1 and -1 and the extreme values are obtained if a spike is either leading ($+1$) or following (-1) coincident spikes in all other spike trains. It can be 0 either if a spike is not part of any coincidences or if it leads exactly as many spikes from other

spike trains in coincidences as it follows. From the definition in Eqs. (26) and (27) it follows immediately that C_k is an upper bound for the absolute value $|D_k|$.

In Fig. 6 we show the application of the SPIKE-Order framework to an example dataset. While the SPIKE-Order profile can be very useful for color-coding and visualizing local spike leaders and followers (Fig. 6a), it is not useful as an overall indicator of Spike Train Order. The profile is invariant under exchange of spike

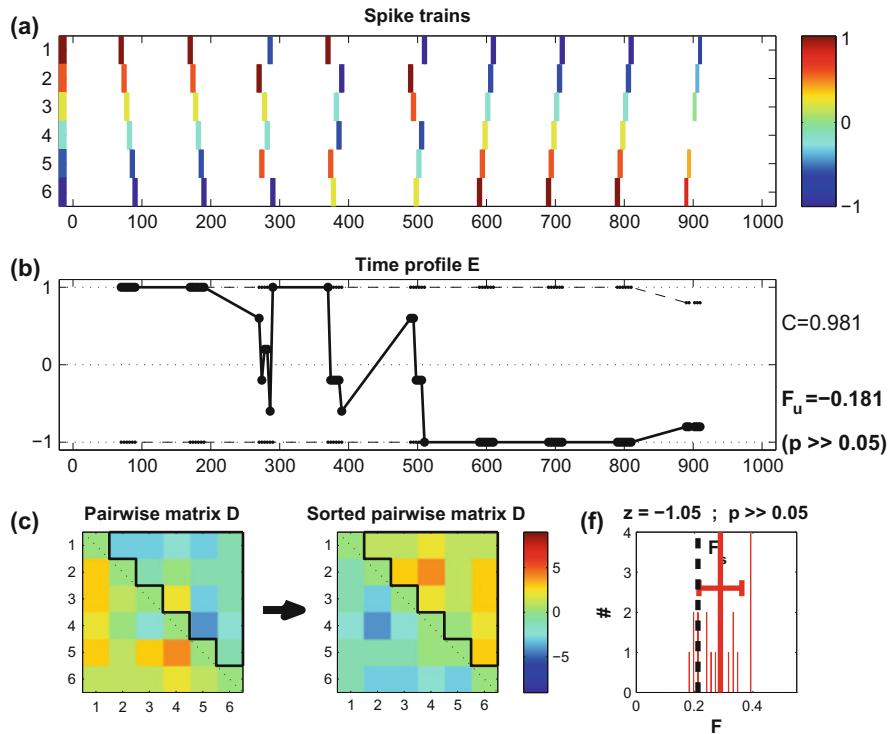


Fig. 6 Illustration of SPIKE-Order for an artificially created example dataset consisting of 6 spike trains which emit spikes in nine reliable events. For the first two events spikes fire in order, for the next three events the order is random whereas for the last four events the order is inverted. In the last event there is one spike missing. (a) Unsorted spike trains with the spikes color-coded according to the value of the SPIKE-Order $D(t_k)$. (b) Spike Train Order profile $E(t_k)$. Events with different firing order can clearly be distinguished. The SPIKE-Synchronization profile $C(t_k)$ and its mirror profile (dashed black lines) act as envelope. The Synfire Indicator F_u for the unsorted spike trains is slightly negative reflecting the dominance of the inversely ordered events. (c) Pairwise cumulative SPIKE-Order matrix D before (left) and after (right) sorting. The optimal order maximizes the upper triangular matrix $D_{(n < m)}$, marked in black. The thick black arrow in between the two matrices indicates the sorting process. (d) Spike Train Order profile $E(t_k)$ and its average values, the Synfire Indicator F_s for the sorted spike trains. (e) Sorted spike trains. (f) Statistical significance: Results of the surrogate analysis. Thick lines denote mean and standard deviation for 19 surrogates. Since the value for the original dataset (black) is not maximum, the optimally sorted spike trains do not exhibit a statistically significant synfire pattern

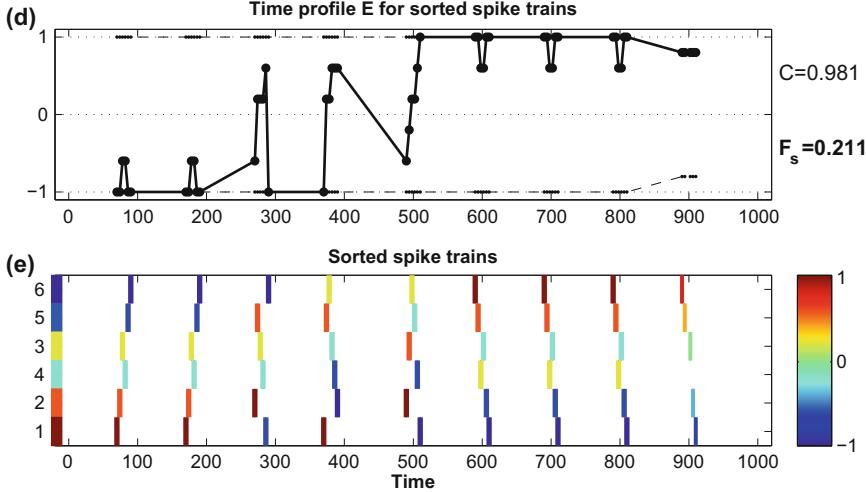


Fig. 6 (continued)

trains, i.e. it looks the same for all events no matter what the order of the firing is (in our example only the last event looks slightly different since one spike is missing). Moreover, summing over all profile values, which is equivalent to summing over all coincidences, necessarily leads to an average value of 0, since for every leading spike (+1) there has to be a following spike (-1).

So in order to quantify any kind of leader-follower information between spike trains we need a second kind of order indicator. The Spike Train Order indicator is similar to the SPIKE-Order indicator defined in Eqs. (26) and (27) but with two important differences. Both spikes are assigned the same value and this value now depends on the order of the spike trains:

$$E_i^{(n,m)} = C_i^{(n,m)} \cdot \begin{cases} \text{sign}(t_j^{(m)} - t_i^{(n)}) & \text{if } n < m \\ \text{sign}(t_i^{(n)} - t_j^{(m)}) & \text{if } n > m \end{cases} \quad (28)$$

and

$$E_j^{(m,n)} = E_i^{(n,m)}. \quad (29)$$

This symmetric indicator assigns to both spikes a +1 in case the two spikes are in the correct order, i.e. the spike from the spike train with the lower spike train index is leading the coincidence, and a -1 in the opposite case. Once more the value 0 is obtained when there is no coincident spike in the other spike train or when the two coincident spikes are absolutely identical.

The multivariate profile $E(t_k)$ Fig. 6b, again obtained similarly to Eq. (23), is also normalized between 1 and -1 and the extreme values are obtained for a coincident

event covering all spike trains with all spikes emitted in the order from first (last) to last (first) spike train, respectively (see the first two and the last four events in Fig. 6). It can be 0 either if a spike is not a part of any coincidences or if the order is such that correctly and incorrectly ordered spike train pairs cancel each other. Again, C_k is an upper bound for the absolute value of E_k .

3.2 Synfire Indicator

In contrast to the SPIKE-Order profile D_k , for the Spike Train Order profile E_k it does make sense to define an average value, which we term the Synfire Indicator:

$$F = \frac{1}{M} \sum_{k=1}^M E(t_k). \quad (30)$$

The interpretation is very intuitive. The Synfire Indicator F quantifies to what degree the spike trains in their current order resemble a perfect synfire pattern. It is normalized between 1 and -1 and attains the value 1 (-1) if the spike trains in their current order form a perfect (inverse) synfire pattern, meaning that all spikes are coincident with spikes in all other spike trains and that the order from leading (following) to following (leading) spike train remains consistent for all of these events. It is 0 either if the spike trains do not contain any coincidences at all or if among all spike trains there is a complete symmetry between leading and following spikes.

The Spike Train Order profile $E(t_k)$ for our example is shown in Fig. 6c. In this case the order of spikes within an event clearly matters. The Synfire Indicator F is slightly negative indicating that the current order of the spike trains is actually closer to an inverse synfire pattern.

Given a set of spike trains we now would like to sort the spike trains from leader to follower such that the set comes as close as possible to a synfire pattern. To do so we have to maximize the overall number of correctly ordered coincidences and this is equivalent to maximizing the Synfire Indicator F . However, it would be very difficult to achieve this maximization by means of the multivariate profile $E(t_k)$. Clearly, it is more efficient to sort the spike trains based on a pairwise analysis of the spike trains. The most intuitive way is to use the anti-symmetric cumulative SPIKE-Order matrix

$$D^{(n,m)} = \sum_i D_i^{(n,m)} \quad (31)$$

which sums up orders of coincidences from the respective pair of spike trains only and quantifies how much spike train n is leading spike train m (Fig. 6c).

Hence if $D^{(n,m)} > 0$ spike train n is leading m , while $D^{(n,m)} < 0$ means m is leading n . If the current Spike Train Order is consistent with the synfire property,

we thus expect that $D^{(n,m)} > 0$ for $n < m$ and $D^{(n,m)} < 0$ for $n > m$. Therefore, we construct the overall SPIKE-Order as

$$D_{(n < m)} = \sum_{n < m} D^{(n,m)}, \quad (32)$$

i.e. the sum over the upper right tridiagonal part of the matrix $D^{(n,m)}$.

After normalizing by the overall number of possible coincidences, we arrive at a second more practical definition of the Synfire Indicator:

$$F = \frac{2D_{(n < m)}}{(N - 1)M}. \quad (33)$$

The value is identical to the one of Eq. (30), only the temporal and the spatial summation of coincidences (i.e., over the profile and over spike train pairs) are performed in the opposite order.

Having such a quantification depending on the order of spike trains, we can introduce a new ordering in terms of the spike train index permutation $\varphi(n)$. The overall Synfire Indicator for this permutation is then denoted as F_φ . Accordingly, for the initial (**unsorted**) order of spike trains φ_u the Synfire Indicator is denoted as $F_u = F_{\varphi_u}$.

The aim of the analysis is now to find the optimal (sorted) order φ_s as the one resulting in the maximal overall Synfire Indicator $F_s = F_{\varphi_s}$:

$$\varphi_s : F_{\varphi_s} = \max_\varphi \{F_\varphi\} = F_s. \quad (34)$$

This Synfire Indicator for the sorted spike trains quantifies how close spike trains can be sorted to resemble a synfire pattern, i.e., to what extent coinciding spike pairs with correct order prevail over coinciding spike pairs with incorrect order. Unlike the Synfire Indicator for the unsorted spike trains F_u , the optimized Synfire Indicator F_s can only attain values between 0 and 1 (any order that yields a negative result could simply be reversed in order to obtain the same positive value). For a perfect synfire pattern we obtain $F_s = 1$, while sufficiently long Poisson spike trains without any synfire structure yield $F_s \approx 0$.

The complexity of the problem to find the optimal Spike Train Order is similar to the well-known travelling salesman problem [2]. For N spike trains there are $N!$ permutations φ , so for large numbers of spike trains finding the optimal Spike Train Order φ_s is a non-trivial problem and brute-force methods such as calculating the F_φ -value for all possible permutations are not feasible. Instead, we search for the optimal order using simulated annealing [7], a probabilistic technique which approximates the global optimum of a given function in a large search space. In our case this function is the Synfire Indicator F_φ (which we would like to maximize) and the search space is the permutation space of all spike trains. We start with the F_u -value from the unsorted permutation and then visit nearby permutations

using the fundamental move of exchanging two neighboring spike trains within the current permutation. All moves with positive ΔF are accepted while the likelihood of accepting moves with negative ΔF is decreased along the way according to a standard slow cooling scheme. The procedure is repeated iteratively until the order of the spike trains no longer changes or until a predefined end temperature is reached.

The sorting of the spike trains maximizes the Synfire Indicator as reflected by both the normalized sum of the upper right half of the pairwise cumulative SPIKE-Order matrix (Eq.(33), Fig. 6c) and the average value of the Spike Train Order profile $E(t_k)$ (Eq. (30), Fig. 6d). Finally, the sorted spike trains in Fig. 6e are now ordered such that the first spike trains have predominantly high values (red) and the last spike trains predominantly low values (blue) of $D(t_k)$.

The complete analysis returns results consisting of several levels of information. Time-resolved (local) information is represented in the spike-coloring and in the profiles D and E . The pairwise information in the SPIKE-order matrix reflects the leader-follower relationship between two spike trains at a time. The Synfire Indicator F characterizes the closeness of the dataset as a whole to a synfire pattern, both for the unsorted (F_u) and for the sorted (F_s) spike trains. Finally, the sorted order of the spike trains is a very important result in itself since it identifies the leading and the following spike trains.

3.3 Statistical Significance

As a last step in the analysis we evaluate the statistical significance of the optimized Synfire Indicator F_s using a set of carefully constructed spike train surrogates. The idea behind the surrogate test is to estimate the likelihood that the consistent SPIKE-Order pattern yielding a certain Synfire Indicator could have been obtained by chance. To this aim, for each surrogate we maintain the coincidence structure of the spike trains by keeping the SPIKE-Synchronization values of each individual spike constant but randomly swap the spike order in a sufficient number of coincidences. We set the number of swaps equal to the number of coincident spikes in the dataset since this way all possible spike order patterns can be reached. Only for the first surrogate we swap twice as many coincidences in order to account for transients. After each swap we take extra care that all other spike orders that are affected by the swap are updated as well. For example, if a swap changes the order between the first and the third spike in an ordered sequence of three spikes, we also swap both the order between the first and the second and the order between the second and the third spike.

For each spike train surrogate we repeat exactly the same optimization procedure in the spike train permutation space that is done for the original dataset. The original Synfire Indicator is deemed significant if it is higher than the Synfire Indicator obtained for all of the surrogate datasets (this case will be marked by two asterisks). Here we use $s = 19$ surrogates for a significance level of $p^* = 1/(s + 1) = 0.05$.

Note that in order to achieve a better sampling of the underlying null distribution a larger number of surrogates would be preferable but the chosen value of s is a compromise that takes into account the computational cost. As a second indicator we state the z-score, e.g., the deviation of the original value x from the mean μ of the surrogates in units of their standard deviation σ :

$$z = \frac{x - \mu}{\sigma}. \quad (35)$$

Results of the significance analysis for our standard example are shown in the histogram in Fig. 6f. In this case the absolute value of the z-score is smaller than one and the p -value is larger than p^* and the result is thus judged as statistically non-significant.

4 Outlook

In the first part of this chapter we describe three parameter-free and time resolved measures of spike train synchrony in their recently developed adaptive extensions, A-ISI-distance, A-SPIKE-distance and A-SPIKE-synchronization [24]. All of these measures are symmetric and so their multivariate versions are invariant to changes in the order of spike trains. Since information about directionality is very relevant, in the second part of this chapter we show an algorithm which allows to sort multiple spike trains from leader to follower. This algorithm is built on two indicators, SPIKE-Order and Spike Train Order, that define the Synfire Indicator value, which quantifies the consistency of the temporal leader-follower relationships for both the original and the optimized sorting.

Symmetric measures of spike train distances (Sect. 2) have been applied in many different contexts, not only in the field of neuroscience [10, 18, 28]. For example, they have been used in robotics [9] and prosthesis control [8]. The ISI-distance has been applied in a method for the detection of directional coupling between point processes and point processes and flows [1, 19], as an adaptation of the nonlinear technique for directional coupling detection of continuous signals [5].

Questions about leader-follower dynamics (Sect. 3) have been specifically investigated in neuroscience [21], but also in fields as wide-ranging as, e.g., climatology [3], social communication [27], and human-robot interaction [23]. SPIKE-Order has already been applied to analyze the consistency of propagation patterns in two real datasets from neuroscience (Giant Depolarized Potentials in mice slices) and climatology (El Niño sea surface temperature recordings) [17].

Finally, we would like to mention that the similarity measures A-ISI-distance, A-SPIKE-distance and A-SPIKE-synchronization, as well as SPIKE-Order, are implemented in three publicly available software packages, the Matlab-based

graphical user interface SPIKY¹ [16], cSPIKE² (Matlab command line with MEX-files), and the open-source Python library PySpike³ [20].

Acknowledgements We acknowledge funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. #642563 'Complex Oscillatory Systems: Modeling and Analysis' (COSMOS). T.K. also acknowledges support from the European Commission through Marie Curie Initial Training Network 'Neural Engineering Transformative Technologies' (NETT), project 289146. We thank Ralph G. Andrzejak, Nebojsa Bozanic, Kerstin Lenk, Mario Mulansky, and Martin Pofahl for useful discussions.

References

1. Andrzejak, R.G., Kreuz, T.: Characterizing unidirectional couplings between point processes and flows. *Europhys. Lett.* **96**, 50012 (2011)
2. Applegate, D.L., Bixby, R.E., Chvatal, V., Cook, W.J.: *The Traveling Salesman Problem: A Computational Study*. Princeton University Press, Princeton (2011)
3. Boers, N., Bookhagen, B., Barbosa, H.M.J., Marwan, N., Kurths, J., Marengo, J.A.: Prediction of extreme floods in the eastern central andes based on a complex networks approach. *Nat. Commun.* **5**, 5199 (2014)
4. Bower, M.R., Stead, M., Meyer, F.B., Marsh, W.R., Worrell, G.A.: Spatiotemporal neuronal correlates of seizure generation in focal epilepsy. *Epilepsia* **53**, 807 (2012)
5. Chicharro, D., Andrzejak, R.G.: Reliable detection of directional couplings using rank statistics. *Phys. Rev. E* **80**(2), 026217 (2009)
6. Chicharro, D., Kreuz, T., Andrzejak, R.G.: What can spike train distances tell us about the neural code? *J. Neurosci. Methods* **199**, 146–165 (2011)
7. Dowland, K.A., Thompson, J.M.: Simulated annealing. In: *Handbook of Natural Computing*, pp. 1623–1655. Springer, Berlin (2012)
8. Dura-Bernal, S., Li, K., Neymotin, S.A., Francis, J.T., Principe, J.C., Lytton, W.W.: Restoring behavior via inverse neurocontroller in a lesioned cortical spiking model driving a virtual arm. *Front. Neurosci.* **10**, 28 (2016)
9. Espinal, A., Rostro-Gonzalez, H., Carpio, M., Guerra-Hernandez, E.I., Ornelas-Rodriguez, M., Puga-Soberanes, H.J., Sotelo-Figuero, M.A., Melin, P.: Quadrupedal robot locomotion: a biologically inspired approach and its hardware implementation. *Comput. Intell. Neurosci.* **2016**, 5615618 (2016)
10. Jolivet, R., Kobayashi, R., Rauch, A., Naud, R., Shinomoto, S., Gerstner, W.: A benchmark test for a quantitative assessment of simple neuron models. *J. Neurosci. Methods* **169**, 417 (2008)
11. Kreuz, T.: Synchronization measures. In: Quiroga, R., Panzeri, S. (eds.) *Principles of Neural Coding*, p. 97. CRC Taylor and Francis, Boca Raton, FL (2013)
12. Kreuz, T., Haas, J.S., Morelli, A., Abarbanel, H.D.I., Politi, A.: Measuring spike train synchrony. *J. Neurosci. Methods* **165**, 151 (2007)
13. Kreuz, T., Mormann, F., Andrzejak, R.G., Kraskov, A., Lehnertz, K., Grassberger, P.: Measuring synchronization in coupled model systems: a comparison of different approaches. *Phys. D* **225**, 29 (2007)

¹<http://www.fi.isc.cnr.it/users/thomas.kreuz/Source-Code/SPIKY.html>.

²<http://www.fi.isc.cnr.it/users/thomas.kreuz/Source-Code/cSPIKE.html>.

³<http://mariomulansky.github.io/PySpike>.

14. Kreuz, T., Chicharro, D., Greschner, M., Andrzejak, R.G.: Time-resolved and time-scale adaptive measures of spike train synchrony. *J. Neurosci. Methods* **195**, 92 (2011)
15. Kreuz, T., Chicharro, D., Houghton, C., Andrzejak, R.G., Mormann, F.: Monitoring spike train synchrony. *J. Neurophysiol.* **109**, 1457 (2013)
16. Kreuz, T., Mulansky, M., Bozanic, N.: SPIKY: a graphical user interface for monitoring spike train synchrony. *J. Neurophysiol.* **113**, 3432 (2015)
17. Kreuz, T., Satuvuori, E., Pofahl, M., Mulansky, M.: Leaders and followers: quantifying consistency in spatio-temporal propagation patterns. *New J. Phys.* **19**, 043028 (2017)
18. Mainen, Z., Sejnowski, T.J.: Reliability of spike timing in neocortical neurons. *Science* **268**, 1503 (1995)
19. Malvestio, I., Kreuz, T., Andrzejak, RG: Robustness and versatility of a nonlinear interdependence method for directional coupling detection from spike trains. *Phys. Rev. E* **96**, 022203 (2017)
20. Mulansky, M., Kreuz, T.: Pyspike - a python library for analyzing spike train synchrony. *Softw. X* **5**, 183–189 (2016)
21. Pereda, E., Quian Quiroga, R., Bhattacharya, J.: Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.* **77**, 1 (2005)
22. Quian Quiroga, R., Kreuz, T., Grassberger, P.: Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E* **66**, 041904 (2002)
23. Rahbar, F., Anzalone, S., Varni, G., Zibetti, E., Ivaldi, S., Chetouani, M.: Predicting extraversion from non-verbal features during a face-to-face human-robot interaction. In: International Conference on Social Robotics, p. 10 (2015)
24. Satuvuori, E., Mulansky, M., Bozanic, N., Malvestio, I., Zeldenrust, F., Lenk, K., Kreuz, T.: Measures of spike train synchrony for data with multiple time-scales. *J. Neurosci. Methods* **287**, 25 (2017)
25. Truccolo, W., Donoghue, J.A., Hochberg, L.R., Eskandar, E.N., Madsen, J.R., Anderson, W.S., Brown, E.N., Halgren, E., Cash, S.S.: Single-neuron dynamics in human focal epilepsy. *Nat. Neurosci.* **14**, 635 (2011)
26. van Rossum, M.C.W.: A novel spike distance. *Neural Comput.* **13**, 751 (2001)
27. Varni, G., Volpe, G., Camurri, A.: A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. *IEEE Trans. Multimedia* **12**, 576 (2010)
28. Victor, J.D.: Spike train metrics. *Curr. Opin. Neurobiol.* **15**, 585 (2005)
29. Victor, J.D., Purpura, K.P.: Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J. Neurophysiol.* **76**, 1310 (1996)

Space-by-Time Tensor Decomposition for Single-Trial Analysis of Neural Signals



Ioannis Delis, Arno Onken, and Stefano Panzeri

Abstract How to identify the informative dimensions of large-scale neural data is an open research problem. Neural activity carries information across both time (temporal variations in neural responses) and space (differences in the activity of different neurons or brain regions). Here we review a family of analytical methods, termed space-by-time tensor decompositions, which can elucidate how the spatial and temporal dimensions of neural activity interact in order to form robust representations of neural activity in single trials. We present a set of algorithms based on non-negative matrix factorization that implement the space-by-time tensor decomposition and discuss their properties and applicability to different types of neural signals. We then propose a set of measures that can be used to assess the power of tensor decompositions and quantify their effectiveness in capturing neural information. We conclude with a demonstration of the space-by-time decomposition of real neural population spike train data.

1 Introduction

To perceive stimuli, make decisions and create behaviours, neural circuits produce coordinated and distributed patterns of neural activity that evolve over time (time-varying neural recordings) and space (different neurons or brain areas exhibit different neural responses at any time instant) [1–3]. In an experiment, neural activity is typically recorded over multiple repetitions (or trials) that may differ in the type of stimulus presented, in the task the subject is asked to perform or the subject’s response, or in other parameters collectively defining the experimental

I. Delis

Department of Biomedical Engineering, Columbia University, New York, NY, USA
e-mail: ioannis.delis@columbia.edu

A. Onken • S. Panzeri (✉)

Neural Computation Laboratory, Center for Neuroscience and Cognitive Systems @UniTn,
Istituto Italiano di Tecnologia, Rovereto, Italy
e-mail: arno.onken@iit.it; stefano.panzeri@iit.it

conditions. Hence, in order to analyse neural population activity, the data can be represented as three-dimensional arrays, or tensors, indexed by space (neurons or sensors), time (discrete time points relative to trial onset) and trials (including all experimental conditions) [4–7].

How to extract a biologically meaningful and scalable representation of such a three-dimensional tensor of neural activity remains an open problem [8]. An effective representation of neural activity should satisfy several requirements. First, it should capture the single-trial information present in neural activity patterns. Second, it should provide a compact representation of neural activity using a small number of parameters. Third, the extracted low-dimensional structures should be interpretable biologically, i.e. correspond to the constituent stereotyped neural activity patterns shared by the neural responses to different experimental conditions.

Here, we present a computational framework for the decomposition of three-dimensional tensors of neural recordings into a series of distinct ‘components’ along each of the tensor dimensions. We call this approach space-by-time tensor decomposition (see Fig. 1 for a schematic illustration). The space-by-time decomposition provides a compact representation of single trial neural activity describing where (spatial components), when (temporal components) and how much response patterns are active on each trial (scalar coefficients representing the relative contribution of each combination of spatial and temporal components). This decomposition can be generally formulated as a tensor factorization known as Tucker-2 decomposition [9, 10]. The key point of the proposed decomposition

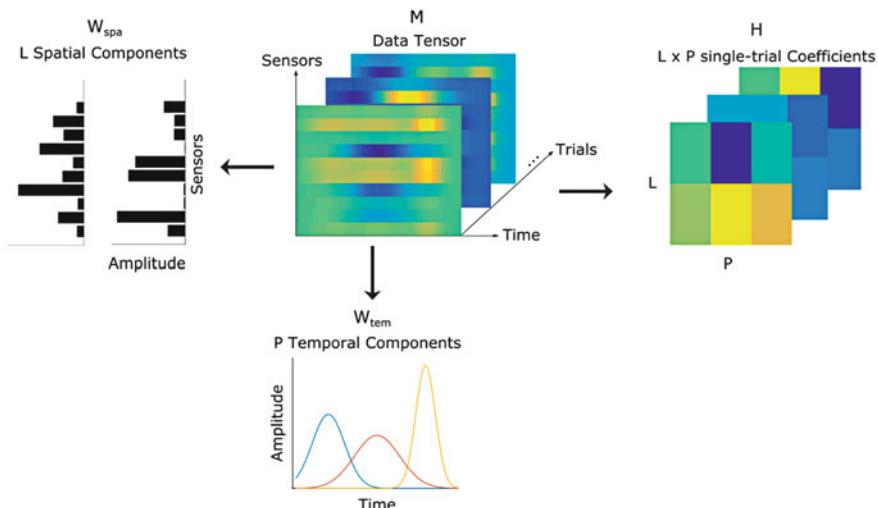


Fig. 1 Schematic illustration of the space-by-time tensor decomposition. In our formulation, a 3-dimensional data tensor \mathbf{M} (top middle) comprising the signals recorded from S sensors over T time points in N trials is decomposed into L S -dimensional spatial components (top left), P T -dimensional temporal components (bottom) and $N L \times P$ -dimensional coefficients (top right). In this example, we show $L = 2$ spatial components and $P = 3$ temporal components

resides in the possibility to combine any of the temporal components with any of the spatial ones, which leads to a low-dimensional, though flexible and functional representation of neural activity patterns.

The space-by-time decomposition can be learned using non-negative matrix factorization (NMF) [11], a dimensionality reduction algorithm that imposes non-negativity constraints on the extracted components. Compared to other possible constraints, such as orthogonality or independence, the non-negativity constraint yields several advantages [12, 13]: (a) the basis functions and coefficients of the nonnegative decomposition are, in principle, directly interpretable as firing patterns and as their strength of recruitment in single trials; (b) it generates sparse representations of the data; and (c) it can cope with non-orthogonal firing patterns such as partly overlapping ones that may be generated by neural circuits with hard-wired connectivity. These properties and the lack of any other assumptions make the space-by-time NMF decomposition a natural method for identifying low-dimensional parts-based representations of neural activity that carry the informative signal dimensions.

2 Computational Framework Formulation

In this section, we review the computational framework of the space-by-time tensor decomposition which serves for dimensionality reduction and identification of the informative dimensions of neural signals. The proposed approach aims to represent single-trial neural data as a small number of spatial and temporal components. In order to approximate the signals recorded from S sensors over T time frames, the space-by-time NMF decomposition identifies trial-independent spatial and temporal components and, for each trial, combines these components using a set of coefficients (Fig. 1). Formally, the data matrix \mathbf{M}_n with dimensions $(T \times S)$ comprising multi-sensor time-varying data during one trial n is factorized as follows:

$$\mathbf{M}_n \simeq \mathbf{W}_{\text{tem}} \mathbf{H}_n \mathbf{W}_{\text{spa}}, \forall n \in \{1, \dots, N\} \quad (1)$$

where \mathbf{W}_{tem} is a $(T \times P)$ matrix whose columns are the temporal components, \mathbf{W}_{spa} is a $(L \times S)$ matrix whose rows are the spatial components and \mathbf{H}_n is a $(P \times L)$ matrix containing the coefficients that combine each one of the P temporal components with each one of the L spatial ones (see Fig. 1 for an illustrative example of the matrices). Note that temporal and spatial components \mathbf{W}_{tem} and \mathbf{W}_{spa} are trial-independent whereas the data and coefficient matrices \mathbf{M}_n and \mathbf{H}_n depend on the trial n . The number of the temporal and spatial components (P, L respectively) are free parameters of the analyses. Below we propose two measures to select the number of dimensions in the data (see section *Assessment of the decompositions*). The product PL is the dimension of the linear space on which each single-trial activity \mathbf{M}_n is represented and dimensionality reduction is effective if $PL \ll TS$.

The tri-factorization in Eq. (1) can be rewritten as the following double summation:

$$\mathbf{m}_n(t) \simeq \sum_{i=1}^P \sum_{j=1}^L w_{\text{tem}}^i(t) h_n^{i,j} \mathbf{w}_{\text{spa}}^j, \quad (2)$$

where temporal and spatial components are now given in vector forms as $\mathbf{w}_{\text{tem}}^i$, $\mathbf{w}_{\text{spa}}^j$ respectively and $h_n^{i,j}$ is a scalar activation coefficient combining components i and j in trial n .

3 Space-by-Time Non-negative Matrix Factorization Algorithm

To learn the above decomposition, we apply a set of algorithms based on Non-negative Matrix Factorization (NMF) [14]. The objective of these algorithms is to identify the three factors defined above so as to minimize the total reconstruction error $E^2 = \sum_n E_n^2$, where E_n^2 is defined as the squared Frobenius norm of the single-trial approximation error between the original data and their space-by-time decomposition:

$$E_n^2 = \|\mathbf{M}_n - \mathbf{W}_{\text{tem}} \mathbf{H}_n \mathbf{W}_{\text{spa}}\|^2 = \sum_{i,j} \left(\mathbf{M}_n^{i,j} - (\mathbf{W}_{\text{tem}} \mathbf{H}_n \mathbf{W}_{\text{spa}})^{i,j} \right)^2. \quad (3)$$

First, we present a converging algorithm which was derived independently under the name “Non-negative Tucker-2 decomposition” in [15] and in the context of muscle synergies under the name “sample-based Non-negative Matrix Tri-Factorization” (*sNM3F*) in [16]. The algorithm decomposes non-negative data matrices \mathbf{M}_n such as those representing simultaneous recordings of firing activity of multiple neurons across time [6] or time-varying activity of several muscles [16]. The *sNM3F* algorithm factorizes the single-trial data into three non-negative matrices representing the spatial components, the temporal components and the single-trial coefficients. It takes as input the data tensor \mathbf{M} ($S \times T \times N$) including all trials, the number of spatial and temporal components to extract (L and P respectively) and the number of trials N . Then, the algorithm uses multiplicative update rules to iteratively learn the three matrices as described below. In the following, we use $(\cdot)^\top$ to denote the matrix transpose operation.

The complete space-by-time NMF algorithm takes the following form [15, 16]:

1. Initialize $\mathbf{W}_{\text{tem}}(T \times P)$, $\mathbf{H}(P \times L \times N)$, and $\mathbf{W}_{\text{spa}}(L \times S)$ with non-negative random numbers uniformly distributed between 0 and 1.

2. Given \mathbf{H} , \mathbf{W}_{tem} and the data tensor $\mathbf{M}(S \times T \times N)$, update \mathbf{W}_{spa} :

- Reshape $\mathbf{H} \rightarrow \mathbf{H}_{\text{spa}}(P \times LN)$
- Calculate $\mathbf{G}_{\text{spa}} = \mathbf{W}_{\text{tem}} \mathbf{H}_{\text{spa}}$
- Reshape $\mathbf{G}_{\text{spa}} \rightarrow \mathbf{G}_{\text{spa}}(TN \times L)$ and $\mathbf{M} \rightarrow \mathbf{M}_{\text{spa}}(TN \times S)$
- For all $i \in \{1, \dots, L\}, j \in \{1, \dots, S\}$, update \mathbf{W}_{spa} using

$$\mathbf{W}_{\text{spa}}^{i,j} \leftarrow \mathbf{W}_{\text{spa}}^{i,j} \left(\mathbf{G}_{\text{spa}}^{\top} \mathbf{M}_{\text{spa}} \right)^{i,j} / \left(\mathbf{G}_{\text{spa}}^{\top} \mathbf{G}_{\text{spa}} \mathbf{W}_{\text{spa}} \right)^{i,j} \quad (4)$$

3. Given \mathbf{H} , \mathbf{W}_{spa} and \mathbf{M} , update \mathbf{W}_{tem} :

- Reshape $\mathbf{H} \rightarrow \mathbf{H}_{\text{tem}}(PN \times L)$
- Calculate $\mathbf{G}_{\text{tem}} = \mathbf{H}_{\text{tem}} \mathbf{W}_{\text{spa}}$
- Reshape $\mathbf{G}_{\text{tem}} \rightarrow \mathbf{G}_{\text{tem}}(P \times SN)$ and $\mathbf{M}_{\text{tem}} \rightarrow \mathbf{M}_{\text{tem}}(T \times SN)$
- For all $i \in \{1, \dots, T\}, j \in \{1, \dots, P\}$, update \mathbf{W}_{tem} using

$$\mathbf{W}_{\text{tem}}^{i,j} \leftarrow \mathbf{W}_{\text{tem}}^{i,j} \left(\mathbf{M}_{\text{tem}} \mathbf{G}_{\text{tem}}^{\top} \right)^{i,j} / \left(\mathbf{W}_{\text{tem}} \mathbf{G}_{\text{tem}} \mathbf{G}_{\text{tem}}^{\top} \right)^{i,j} \quad (5)$$

4. Given \mathbf{W}_{tem} and \mathbf{W}_{spa} and \mathbf{M} , update \mathbf{H} :

- For all $n \in \{1, \dots, N\}$:
Define $\mathbf{H}_n(P \times L)$ and $\mathbf{M}_n(S \times T)$
- For all $i \in \{1, \dots, P\}, j \in \{1, \dots, L\}$:

$$\mathbf{H}_n^{i,j} \leftarrow \mathbf{H}_n^{i,j} \left(\mathbf{W}_{\text{tem}}^{\top} \mathbf{M}_n \mathbf{W}_{\text{spa}}^{\top} \right)^{i,j} / \left(\mathbf{W}_{\text{tem}}^{\top} \mathbf{W}_{\text{tem}} \mathbf{M}_n \mathbf{W}_{\text{spa}} \mathbf{W}_{\text{spa}}^{\top} \right)^{i,j} \quad (6)$$

5. If decrease in $\sum_{n=1}^N \|\mathbf{M}_n - \mathbf{W}_{\text{tem}} \mathbf{H}_n \mathbf{W}_{\text{spa}}\|^2$ is below a given tolerance, normalize \mathbf{W}_{tem} row-wise, \mathbf{W}_{spa} column-wise, rescale \mathbf{H} to make the error unchanged and stop. Otherwise, go to step 2.

The algorithm outputs the matrices \mathbf{W}_{tem} , \mathbf{W}_{spa} and \mathbf{H} as well as the total reconstruction error $E^2 = \sum_n E_n^2$.

4 Variants of the Decomposition Algorithm

In the following, we present extensions of the decomposition algorithm that incorporate different types of constraints in order to facilitate applicability of the algorithm to specific datasets and interpretability of the algorithm output.

4.1 Orthogonality Constraints

Imposing orthogonality constraints on the component matrices \mathbf{W}_{tem} and \mathbf{W}_{spa} , i.e. $\mathbf{W}_{\text{tem}}^\top \mathbf{W}_{\text{tem}} = \mathbf{I}_{P \times P}$ and $\mathbf{W}_{\text{spa}} \mathbf{W}_{\text{spa}}^\top = \mathbf{I}_{L \times L}$, where $\mathbf{I}_{K \times K}$ denotes the identity matrix of size $K \times K$, typically enhances the sparsity of the extracted components and reduces their overlap. Orthogonality can be imposed on the temporal components, the spatial components or both types of components (bi-orthogonality) and these constraints alter the corresponding update rules. Under the bi-orthogonality constraint, the new algorithm *soNM3F* can be seen as a generalization of clustering [17]. The resulting multiplicative update rules of *soNM3F* that replace Eqs. (4) and (5) in the above algorithm formulation are:

$$\mathbf{W}_{\text{spa}}^{i,j} \leftarrow \mathbf{W}_{\text{spa}}^{i,j} \left(\mathbf{G}_{\text{spa}}^\top \mathbf{M}_{\text{spa}} \right)^{i,j} / \left(\mathbf{G}_{\text{spa}}^\top \mathbf{M}_{\text{spa}} \mathbf{W}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \right)^{i,j} \quad (7)$$

$$\mathbf{W}_{\text{tem}}^{i,j} \leftarrow \mathbf{W}_{\text{tem}}^{i,j} \left(\mathbf{M}_{\text{tem}} \mathbf{G}_{\text{tem}}^\top \right)^{i,j} / \left(\mathbf{W}_{\text{tem}}^\top \mathbf{W}_{\text{tem}} \mathbf{M}_{\text{tem}} \mathbf{G}_{\text{tem}}^\top \right)^{i,j} \quad (8)$$

4.2 Discrimination Objective

A common objective of the single-trial analysis of neural signals is to discriminate between different experiment conditions. Therefore, to enhance the discrimination power of the decomposition, a discrimination objective can be added to the approximation objective of the algorithms. In this case, the new error function that the algorithm aims to minimize comprises two terms: the NMF approximation error defined in Eq. (3) $E_{\text{NMF}}^2 = \sum_n E_n^2$ and a discrimination term inspired by LDA and Fisher's criterion E_{LDA}^2 [18]:

$$E^2 = E_{\text{NMF}}^2 + E_{\text{LDA}}^2 \quad (9)$$

The latter depends on the parameters that carry information about condition differences, i.e. the single-trial coefficients \mathbf{H} of the decomposition. Specifically, we assume that the N trials can be partitioned in C classes (or conditions) K_c . We then define the within-class and between-class scatter matrices as follows [19]:

$$\mathbf{S}_w = \sum_{c=1}^C \sum_{n \in K_c} \text{vec} \left(\mathbf{H}^n - \bar{\mathbf{H}}_c \right) \text{vec} \left(\mathbf{H}^n - \bar{\mathbf{H}}_c \right)^\top \quad (10)$$

$$\mathbf{S}_b = \sum_{c=1}^C \text{vec} \left(\bar{\mathbf{H}}_c - \bar{\mathbf{H}} \right) \text{vec} \left(\bar{\mathbf{H}}_c - \bar{\mathbf{H}} \right)^\top \quad (11)$$

Where $\bar{\mathbf{H}}$ and $\bar{\mathbf{H}}_c$ represent the mean activation coefficient matrices across all samples and across samples belonging to class K_c respectively, and $\text{vec}(\cdot)$ denotes vectorization, i.e. the vertical concatenation of all columns of the argument matrix. \mathbf{S}_w represents the diffusion of sample vector coefficients around their class mean. To increase discriminability across classes, the dispersion of samples that belong to the same class around their corresponding mean should be as small as possible. A convenient metric for the dispersion of the samples is the trace of \mathbf{S}_w , which has to be minimized. \mathbf{S}_b denotes the between-class scatter matrix and defines the diffusion of the mean vectors of all classes around the global mean. To discriminate different classes, each class should be as far as possible from the other classes, thus the trace of \mathbf{S}_b should be as large as possible. We linearly combined these two discrimination metrics to obtain the discrimination objective to be minimized:

$$E_{\text{LDA}}^2 = \gamma \text{tr}(\mathbf{S}_w) - \delta \text{tr}(\mathbf{S}_b) \quad (12)$$

Where γ and δ are free tuning parameters and $\text{tr}(\cdot)$ denotes the trace operator. The negative sign of the last term indicates that $\text{tr}(\mathbf{S}_b)$ is maximized.

The resulting algorithm that minimizes the combined error of Eq. (9) is denoted *DsNM3F* [20] and requires as extra input:

- (a) the values of γ and δ trading-off approximation for discrimination and
- (b) an N -dimensional vector defining the experimental condition imposed on each trial (taking values from 1 up to C for C experimental conditions).

As the LDA-related cost acts only on the activation coefficients \mathbf{H} of the decomposition, the update rules of the temporal and spatial components are the same as for the sNM3F algorithm [Eqs. (4) and (5)]. The update rule for the activation coefficients is adapted to incorporate the task discrimination objectives and becomes as follows (supposing that trial n belongs to the group K_l (with m_l elements)):

$$\mathbf{H}_n^{i,j} \leftarrow \mathbf{H}_n^{i,j} \frac{\left(\mathbf{W}_{\text{tem}}^\top \mathbf{M}_n \mathbf{W}_{\text{spa}}^\top + \frac{\delta}{m_l} \bar{\mathbf{H}}_l + \gamma \bar{\mathbf{H}}_l + \delta \frac{C}{N} \bar{\mathbf{H}} \right)^{i,j}}{\left(\mathbf{W}_{\text{tem}}^\top \mathbf{W}_{\text{tem}} \mathbf{M}_n \mathbf{W}_{\text{spa}} \mathbf{W}_{\text{spa}}^\top + \frac{\delta}{m_l} \bar{\mathbf{H}} + \gamma \mathbf{H}^{sl} + \delta \frac{C}{N} \sum_{c=1}^C \bar{\mathbf{H}}_c \right)^{i,j}} \quad (13)$$

4.3 Application to Signed Data

As most neurobiological signals can take also negative values (e.g. M/EEG, LFP, fMRI), it is important to extend the application of the space-by-time decomposition to signed input data. Hence, we developed a set of algorithms that decompose signed data matrices into non-negative temporal and spatial components and signed single-trial coefficients. The non-negativity of the extracted components ensures that they are naturally sparse and provide low-dimensional parts-based representations of the

data, whereas the mixed-sign coefficients allow the positive or negative activations of the components to explain positive or negative signal variations.

The algorithms implementing the space-by-time decomposition on signed data, termed *ssNM3F* and *scNM3F*, are derived from the update rules of semi-NMF and cluster-NMF respectively, two algorithms that extended the application of standard NMF to matrices with negative entries [21]. The two algorithms differ in the objective functions they aim to optimize, thus they generally produce different outputs. To separate the positive and negative parts of a matrix A and obtain non-negative matrix entries, we use the following notation $A_{ik}^+ = (|A_{ik}| + A_{ik}) / 2$ and $A_{ik}^- = (|A_{ik}| - A_{ik}) / 2$. The *ssNM3F* algorithm, like *sNM3F*, minimizes the standard error function of Eq. (3) and is formulated as follows:

1. Initialize $\mathbf{W}_{\text{tem}}(T \times P)$, $\mathbf{H}(P \times LN)$, and $\mathbf{W}_{\text{spa}}(L \times S)$ with random entries
2. Given \mathbf{H} and the data tensor $\mathbf{M}(S \times T \times N)$,

- a. Reshape $\mathbf{H} \rightarrow \mathbf{H}_{\text{spa}}(P \times LN)$
- b. Calculate $\mathbf{G}_{\text{spa}} = \mathbf{W}_{\text{tem}} \mathbf{H}_{\text{spa}}$
- c. Reshape $\mathbf{G}_{\text{spa}} \rightarrow \mathbf{G}_{\text{spa}}(TN \times L)$ and $\mathbf{M} \rightarrow \mathbf{M}_{\text{spa}}(TN \times S)$
- d. For all $i \in \{1, \dots, L\}, j \in \{1, \dots, S\}$, update \mathbf{W}_{spa} using

$$\mathbf{W}_{\text{spa}}^{i,j} \leftarrow \mathbf{W}_{\text{spa}}^{i,j} \sqrt{\frac{\left[(\mathbf{M}_{\text{spa}}^\top \mathbf{G}_{\text{spa}})^+ \right]^{ij} + \left[\mathbf{W}_{\text{spa}}^\top (\mathbf{G}_{\text{spa}}^\top \mathbf{G}_{\text{spa}})^- \right]^{ij}}{\left[(\mathbf{M}_{\text{spa}}^\top \mathbf{G}_{\text{spa}})^- \right]^{ij} + \left[\mathbf{W}_{\text{spa}}^\top (\mathbf{G}_{\text{spa}}^\top \mathbf{G}_{\text{spa}})^+ \right]^{ij}}} \quad (14)$$

3. Given \mathbf{H} , \mathbf{W}_{spa} and \mathbf{M} , update \mathbf{W}_{tem} :

- a. Reshape $\mathbf{H} \rightarrow \mathbf{H}_{\text{tem}}(PN \times L)$
- b. Calculate $\mathbf{G}_{\text{tem}} = \mathbf{H}_{\text{tem}} \mathbf{W}_{\text{spa}}$
- c. Reshape $\mathbf{G}_{\text{tem}} \rightarrow \mathbf{G}_{\text{tem}}(P \times SN)$ and $\mathbf{M}_{\text{tem}} \rightarrow \mathbf{M}_{\text{tem}}(T \times SN)$
- d. For all $i \in \{1, \dots, T\}, j \in \{1, \dots, P\}$, update \mathbf{W}_{tem} using

$$\mathbf{W}_{\text{tem}}^{i,j} \leftarrow \mathbf{W}_{\text{tem}}^{i,j} \sqrt{\frac{\left[(\mathbf{M}_{\text{tem}} \mathbf{G}_{\text{tem}}^\top)^+ \right]^{ij} + \left[\mathbf{W}_{\text{tem}} (\mathbf{G}_{\text{tem}} \mathbf{G}_{\text{tem}}^\top)^- \right]^{ij}}{\left[(\mathbf{M}_{\text{tem}} \mathbf{G}_{\text{tem}}^\top)^- \right]^{ij} + \left[\mathbf{W}_{\text{tem}} (\mathbf{G}_{\text{tem}} \mathbf{G}_{\text{tem}}^\top)^+ \right]^{ij}}} \quad (15)$$

4. Given \mathbf{W}_{tem} and \mathbf{W}_{spa} :

- a. Compute pseudoinverses of $\mathbf{W}_{\text{tem}} \rightarrow \mathbf{W}_{\text{tem}}^{-1}$ and $\mathbf{W}_{\text{spa}} \rightarrow \mathbf{W}_{\text{spa}}^{-1}$
- b. For all $n \in \{1, \dots, N\}$, update \mathbf{H}_n using

$$\mathbf{H}_n \leftarrow \mathbf{W}_{\text{tem}}^{-1} \mathbf{M}_n \mathbf{W}_{\text{spa}}^{-1} \quad (16)$$

5. If decrease in approximation error $\sum_{n=1}^N \|\mathbf{M}_n - \mathbf{W}_{\text{tem}} \mathbf{H}_n \mathbf{W}_{\text{spa}}\|^2$ is below a given tolerance, stop. Otherwise, go to step 2.

The scNM3F algorithm uses the same objective [Eq. (3)] to optimize the activation coefficients \mathbf{H} , but not the components. The temporal and spatial components are updated minimizing the following approximation errors:

$$E_{\text{tem}}^2 = \left\| \mathbf{M}_{\text{tem}} - \mathbf{W}_{\text{tem}} \mathbf{W}_{\text{tem}}^\top \mathbf{M}_{\text{tem}} \right\|^2 \quad (17)$$

$$E_{\text{spa}}^2 = \left\| \mathbf{M}_{\text{spa}} - \mathbf{M}_{\text{spa}} \mathbf{W}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \right\|^2 \quad (18)$$

where \mathbf{M}_{tem} and \mathbf{M}_{spa} are reshaped versions of the input matrix \mathbf{M} with dimensions $(T \times SN)$ and $(TN \times S)$ respectively. Importantly, the components extracted by scNM3F are typically sparser and overlap less than the ones extracted by ssNM3F, which stems from these two objective functions that provide the components with a unique clustering interpretation [5, 21]. The full scNM3F algorithm takes the following form:

1. Initialize $\mathbf{W}_{\text{tem}}(T \times P)$, $\mathbf{H}(P \times LN)$, and $\mathbf{W}_{\text{spa}}(L \times S)$ with random entries
2. Given \mathbf{H} and the data tensor $\mathbf{M}(S \times T \times N)$,
 - a. Reshape $\mathbf{M} \rightarrow \mathbf{M}_{\text{spa}}(TN \times S)$
 - b. For all $i \in \{1, \dots, L\}, j \in \{1, \dots, S\}$, update \mathbf{W}_{spa} using

$$\mathbf{W}_{\text{spa}}^{i,j} \leftarrow \mathbf{W}_{\text{spa}}^{i,j} \sqrt{\frac{\left[(\mathbf{M}_{\text{spa}}^\top \mathbf{M}_{\text{spa}})^+ \mathbf{W}_{\text{spa}}^\top \right]^{i,j} + \left[\mathbf{W}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} (\mathbf{M}_{\text{spa}}^\top \mathbf{M}_{\text{spa}})^- \mathbf{W}_{\text{spa}}^\top \right]^{i,j}}{\left[(\mathbf{M}_{\text{spa}}^\top \mathbf{M}_{\text{spa}})^- \mathbf{W}_{\text{spa}}^\top \right]^{i,j} + \left[\mathbf{W}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} (\mathbf{M}_{\text{spa}}^\top \mathbf{M}_{\text{spa}})^+ \mathbf{W}_{\text{spa}}^\top \right]^{i,j}}} \quad (19)$$

- c. Reshape $\mathbf{M} \rightarrow \mathbf{M}_{\text{tem}}(T \times SN)$
- d. For all $i \in \{1, \dots, T\}, j \in \{1, \dots, P\}$, update \mathbf{W}_{tem} using

$$\mathbf{W}_{\text{tem}}^{i,j} \leftarrow \mathbf{W}_{\text{tem}}^{i,j} \sqrt{\frac{\left[(\mathbf{M}_{\text{tem}} \mathbf{M}_{\text{tem}}^\top)^+ \mathbf{W}_{\text{tem}} \right]^{i,j} + \left[\mathbf{W}_{\text{tem}} \mathbf{W}_{\text{tem}}^\top (\mathbf{M}_{\text{tem}} \mathbf{M}_{\text{tem}}^\top)^- \mathbf{W}_{\text{tem}} \right]^{i,j}}{\left[(\mathbf{M}_{\text{tem}} \mathbf{M}_{\text{tem}}^\top)^- \mathbf{W}_{\text{tem}} \right]^{i,j} + \left[\mathbf{W}_{\text{tem}} \mathbf{W}_{\text{tem}}^\top (\mathbf{M}_{\text{tem}} \mathbf{M}_{\text{tem}}^\top)^+ \mathbf{W}_{\text{tem}} \right]^{i,j}}} \quad (20)$$

3. Given \mathbf{W}_{tem} and \mathbf{W}_{spa} :

- a. Compute pseudoinverses of $\mathbf{W}_{\text{tem}} \rightarrow \mathbf{W}_{\text{tem}}^{-1}$ and $\mathbf{W}_{\text{spa}} \rightarrow \mathbf{W}_{\text{spa}}^{-1}$
- b. For all $n \in \{1, \dots, N\}$, update \mathbf{H}_n using

$$\mathbf{H}_n \leftarrow \mathbf{W}_{\text{tem}}^{-1} \mathbf{M}_n \mathbf{W}_{\text{spa}}^{-1} \quad (21)$$

4. If decrease in approximation error $\sum_{n=1}^N \|\mathbf{M}_n - \mathbf{W}_{\text{tem}} \mathbf{H}_n \mathbf{W}_{\text{spa}}\|^2$ is below a given tolerance, stop. Otherwise, go to step 2.

5 Assessment of the Decompositions

To evaluate the quality of the extracted decompositions, we use two measures: Variance Accounted For (VAF) and decoding performance (DEC).

5.1 Approximation Power of the Decomposition

To assess how well the extracted decompositions approximate the original data, we compute VAF. VAF measures the quality of data reconstruction and is defined by the total reconstruction error normalized by the total variance of the dataset [22]:

$$\text{VAF} = 1 - \frac{\sum_n \|\mathbf{M}_n - \mathbf{W}_{\text{tem}} \mathbf{H}_n \mathbf{W}_{\text{spa}}\|^2}{\sum_n \|\mathbf{M}_n - \bar{\mathbf{M}}\|^2} \quad (22)$$

The total approximation error is computed as the squared Frobenius norm ($\|\cdot\|$) of the difference between the original data and its approximation by the space-by-time decomposition and the total variance of the dataset is the squared Frobenius norm of the difference between the original data and the mean activity across all trials $\bar{\mathbf{M}}$. VAF takes values between 0 (no significant reconstruction) and 1 (perfect reconstruction). Higher values correspond to better approximations of the original data.

5.2 Discrimination Power of the Decomposition

To assess how well the extracted decompositions discriminate between the different experimental conditions imposed on each trial, we propose a single-trial decoding metric. We use the matrix \mathbf{H} of activation coefficients representing the levels of activations of spatial and temporal components in order to predict the experimental condition imposed on each trial (an N -dimensional vector taking values from 1 up to C for C experimental conditions). This analysis can be implemented using any classifier, such as linear discriminant analysis, logistic regression or support vector machines, in conjunction with a cross-validation procedure [23]. Discrimination performance is computed as the percentage of correctly decoded trials (% correct

value). Ultimately, this evaluation serves to quantify the condition-discrimination power of the identified decompositions.

5.3 Model Order Selection

The above two metrics can be used for model order selection, i.e. for choosing the best number of spatial components L and temporal components P . Selection criteria based on the VAF graph (dependence of VAF on number of components), such as the largest change of slope or an absolute threshold, have been widely used in literature when data approximation is the main criterion of goodness of the decomposition [24–27]. When the condition discrimination capacity of the decomposition is of particular importance, model selection can be performed based on the decoding performance as a function of L and P and in conjunction with a cross-validation procedure to avoid overfitting [6, 16, 28]. Practical applications of the space-by-time decomposition showed that VAF typically exhibits a steady increase, which makes the selection of dimensions difficult [16, 29]. In contrast, decoding performance usually has a clear saturation point that indicates the smallest number of components carrying most of the discriminative power of the decomposition [6, 16]. A metric combining the two above metrics ($VDM = VAF \times DEC$, $0 \leq VDM \leq 1$) has been proposed in order to determine the decomposition that achieves the best trade-off between approximation and discrimination power [20, 29].

6 Example Application: Retinal Ganglion Cells

In this section, we demonstrate an application of the space-by-time NMF decomposition to neural data. These results were previously presented in [6].

We recorded spike trains from populations of salamander retinal ganglion cells in response to 60 natural image stimuli (Fig. 2a). Our goal was to find invariant spatial and temporal components that capture the inherent structure of the responses, and trial-dependent coefficients that represent activation levels of those components. Before applying the decomposition method, we discretized neural responses by binning neural spike trains into short intervals (here 10 ms) and counting the number of spikes in each interval. Concatenating the population responses along the trial dimension, we obtained a data tensor of the neural activity. We then decomposed this tensor using space-by-time NMF. For selecting the number of spatial and temporal components, we maximized visual information (DEC) with the smallest number of parameters, where we evaluated decoding performance using a cross-validated LDA classifier (see section *Assessment of decompositions*). This procedure yielded three temporal components (Fig. 2b) and eight spatial components (Fig. 2c).

Each temporal module (Fig. 2b) describes the temporal activation profile of the population. Two of the temporal components had a non-zero amplitude only in

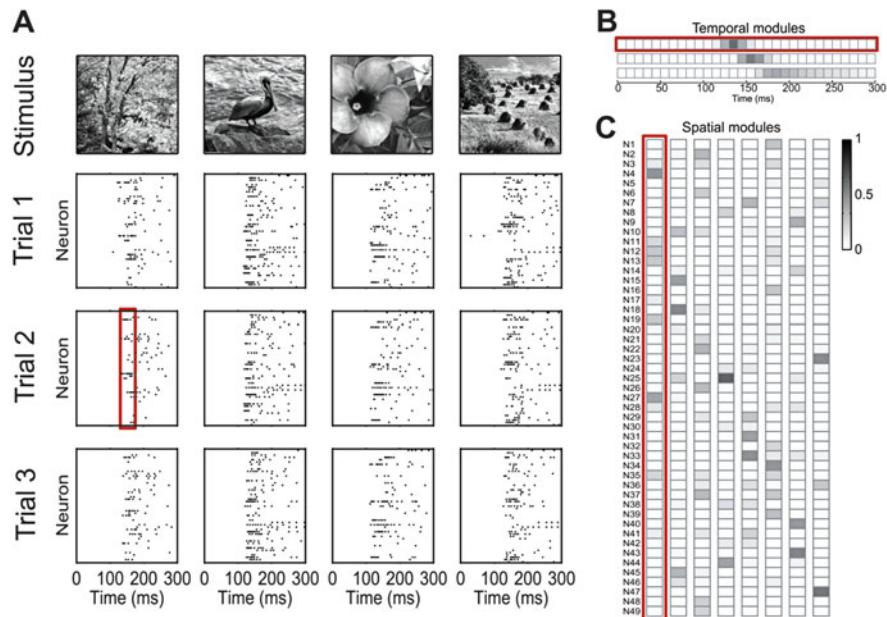


Fig. 2 Example application of space-by-time NMF decomposition to data recorded from retinal ganglion cells. **(a)** Raster plots of the spike times of 49 neurons in response to four example natural image stimuli (out of a total of 60 stimuli) over a 0–300 ms post-stimulus period. Each column shows the responses in three trials (out of a total of 15 trials per stimulus used for the decomposition) to the image shown at the top. **(b)** Three temporal components describing the temporal activation profiles of the neurons over the 0–300 ms post-stimulus period. **(c)** Eight spatial components describing which combinations of the 49 neurons (N1–N49) were activated together. Adapted with permission from [6]. Red frames illustrate a pair of a temporal **(b)** and a spatial **(c)** module that together approximate well a population response within a particular time window **(a)**

a short time region centered around two different times (140 and 160 ms post-stimulus onset, respectively) shortly after the onset of neural response to the images. These two temporal components thus described the difference in neuronal response latencies across cells and stimuli. A third module had non-zero amplitude over a wide 140 ms interval starting at approximately 160 ms post-stimulus and described the sustained part of the neural response.

Each spatial module (Fig. 2c) describes the relative degree of simultaneous firing of groups of neurons. Large amplitudes of a set of cells within a module indicate that those cells tend strongly to fire together. These cell assemblies respond in a coordinated manner to specific features in the presented images (Fig. 2a, relation between spatial components and images not shown).

Both temporal and spatial components match with firing patterns that can be spotted in the original spike trains (Fig. 2, illustrated in red frames for a particular pair of components). Together, these results highlight the ability of the space-by-time NMF decomposition to find firing patterns that occur robustly across stimuli.

7 Software Implementation

A software implementation of these algorithms is freely available online at <https://sites.google.com/site/ioannisdeliswebpage/software> or <http://stommac.eu/index.php/code>.

8 Conclusions

In this chapter, we described a set of tensor decomposition algorithms that identify the low-dimensional structure of neural signals in time and space. When neural data contain responses across multiple locations (sensors or neurons), times (relative to trial onset), and trials (including different conditions), they are naturally expressed as a third-order tensor. The methods we presented here decompose third-order tensors into a combination of trial-independent spatial patterns (describing which combinations of neurons/sensors are activated together), temporal firing patterns (describing the temporal activation profiles of the spatial patterns) and trial-dependent coefficients (describing the strength by which each combination of a spatial and temporal pattern is recruited in a given trial). We also proposed a set of measures to evaluate the effectiveness of these decompositions in both approximating the original recordings and conveying information about differences in the experimental conditions under consideration. Finally, we demonstrated an application of the proposed approach to neural signals and showed that the space-by-time decomposition yields a meaningful representation of the recorded neural activation patterns.

We suggest that this approach enables scientists to identify the informative latent dimensions in complex high-dimensional neural data and understand better the neural representations of sensory, cognitive and motor functions. One of the findings that this analysis starts to reveal is the high decoding performance, data robustness and generalization properties of tensor decompositions. The work presented here can thus contribute to establishing space-by-time decompositions as a mainstream approach in neural signal analysis. In particular, we suggest that these decompositions can be particularly effective when the aim is to (a) tease apart condition-dependent neural patterns with different functional roles, (b) reveal their spatial and temporal representations and (c) quantify their relative contribution to discrimination between experimental conditions.

In addition to revealing the dimensions of neural activity that encode different sensory, cognitive and motor functions, the proposed tools can also be used for comparing the tensor structures of different datasets recorded during different tasks or from different brain areas. Differences in tensor structure can be informative about the predominant origin of variability in neural activity, and can be used to test specific hypotheses about how neural computations may differ in different brain areas or during different behaviours [7].

In summary, it is our hope that data-analysis tools such as those described here will help to provide insights into the information processing and representation mechanisms in the nervous system.

References

- Panzeri, S., Macke, J.H., Gross, J., Kayser, C.: Neural population coding: combining insights from microscopic and mass signals. *Trends Cogn. Sci.* **19**, 162–172 (2015)
- Abeles, M., Gerstein, G.L.: Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.* **60**, 909–924 (1988)
- Quiroga, R.Q., Panzeri, S.: Extracting information from neuronal populations: information theory and decoding approaches. *Nat. Rev. Neurosci.* **10**, 173–185 (2009)
- Cong, F., Lin, Q.H., Kuang, L.D., Gong, X.F., Astikainen, P., Ristaniemi, T.: Tensor decomposition of EEG signals: a brief review. *J. Neurosci. Methods.* **248**, 59–69 (2015)
- Delis, I., Onken, A., Schyns, P.G., Panzeri, S., Philiastides, M.G.: Space-by-time decomposition for single-trial decoding of M/EEG activity. *Neuroimage.* **133**, 504–515 (2016)
- Onken, A., Liu, J.K., Karunasekara, P.P.C.R., Delis, I., Gollisch, T., Panzeri, S.: Using matrix and tensor factorizations for the single-trial analysis of population spike trains. *PLoS Comput. Biol.* **12**, e1005189 (2016)
- Seely, J.S., Kaufman, M.T., Ryu, S.I., Shenoy, K.V., Cunningham, J.P., Churchland, M.M.: Tensor analysis reveals distinct population structure that parallels the different computational roles of areas M1 and V1. *PLoS Comput. Biol.* **12**, e1005164 (2016)
- Cunningham, J.P., Yu, B.M.: Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014)
- Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500 (2009)
- Tucker, L.R.: Implications of factor analysis of three-way matrices for measurement of change. In: Harris, C.W. (ed.) *Problems in Measuring Change*, pp. 122–137. University of Wisconsin Press, Madison (1963)
- Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature.* **401**, 788–791 (1999)
- Cichocki, A., Zdunek, P., Phan, A.H., Amari, S.: *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis*. Wiley, New York (2009)
- Cichocki, A., Mandic, D.P., Phan, A.H., Caiafa, C.F., Zhou, G.X., Zhao, Q.B., et al.: Tensor decompositions for signal processing applications. *IEEE Signal Process Mag.* **32**, 145–163 (2015)
- Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Leen, T.K., Dietterich, T.G. (eds.) *Advances in Neural Information Processing Systems*, pp. 556–562. MIT Press, Cambridge (2001)
- Kim, Y.-D., Choi, S.: Nonnegative tucker decomposition. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2007, CVPR'07*, pp. 1–8 (2007)
- Delis, I., Panzeri, S., Pozzo, T., Berret, B.: A unifying model of concurrent spatial and temporal modularity in muscle activity. *J. Neurophysiol.* **111**, 675–693 (2014)
- Ding, C., Li, T., Peng, W., Park, H.: Orthogonal nonnegative matrix tri-factorizations for clustering. In: *Proceedings of 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 126–135. New York (2006)
- Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936)
- Zafeiriou, S., Tefas, A., Buciu, I., Pitas, I.: Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification. *IEEE Trans. Neural Netw.* **17**, 683–695 (2006)

20. Delis, I., Panzeri, S., Pozzo, T., Berret, B.: Task-discriminative space-by-time factorization of muscle activity. *Front. Hum. Neurosci.* **9**, 399 (2015)
21. Ding, C., Li, T., Jordan, M.I.: Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 45–55 (2010)
22. d'Avella, A., Portone, A., Fernandez, L., Lacquaniti, F.: Control of fast-reaching movements by muscle synergy combinations. *J. Neurosci.* **26**, 7791–7810 (2006)
23. Duda, O., Hart, P., Stork, D.: *Pattern Classification*. Wiley, New York (2001)
24. Tresch, M.C., Cheung, V.C., d'Avella, A.: Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. *J. Neurophysiol.* **95**, 2199–2212 (2006)
25. Cheung, V.C., d'Avella, A., Tresch, M.C., Bizzi, E.: Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. *J. Neurosci.* **25**, 6419–6434 (2005)
26. Torres-Oviedo, G., Macpherson, J.M., Ting, L.H.: Muscle synergy organization is robust across a variety of postural perturbations. *J. Neurophysiol.* **96**, 1530–1546 (2006)
27. Cattell, R.B.: The scree test for the number of factors. *Multivar. Behav. Res.* **1**, 245–276 (1966)
28. Delis, I., Berret, B., Pozzo, T., Panzeri, S.: Quantitative evaluation of muscle synergy models: a single-trial task decoding approach. *Front. Comput. Neurosci.* **7**, 8 (2013)
29. Semprini, M., Cuppone, A.V., Delis, I., Squeri, V., Panzeri, S., Konczak, J.: Biofeedback signals for robotic rehabilitation: assessment of wrist muscle activation patterns in healthy humans. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 883–892 (2016)

Inverse Modeling for MEG/EEG Data



Alberto Sorrentino and Michele Piana

Abstract We provide an overview of the state-of-the-art for mathematical methods that are used to reconstruct brain activity from neurophysiological data. After a brief introduction on the mathematics of the forward problem, we discuss standard and recently proposed regularization methods, as well as Monte Carlo techniques for Bayesian inference. We classify the inverse methods based on the underlying source model, and discuss advantages and disadvantages. Finally we describe an application to the pre-surgical evaluation of epileptic patients.

1 Introduction

Neurophysiological investigation based on modalities like electro- and magneto-encephalography (EEG [30] and MEG [17]), electrocorticography (ECoG, [33]), and stereo EEG (SEEG, [7]) is experiencing an impressive growth in both basic research and clinical applications. The reason of this development is twofold. First, neurophysiological devices are by far the functional approaches that assure the best possible time resolution (up to 1 ms); second, recent hardware developments like multi-channel whole-head EEG and MEG helmets provide a spatial resolution of around 1 cm which, although not as impressive as the time resolution, still permits to draw reliable conclusions from both a scientific and a diagnostic viewpoint. For example, among the available brain imaging modalities, EEG, MEG, and ECoG uniquely allow the realization of functional experiments able to investigate the role of brain oscillatory systems [10]. Further, EEG is for sure the gold standard in Brain Computer Interface experiments [46], while MEG and SEEG are certainly the most promising modalities for the comprehension of epileptogenic activity and for the neurophysiological interpretation of epileptic seizures [28].

A. Sorrentino (✉) · M. Piana

Dipartimento di Matematica, Università degli Studi di Genova, Via Dodecaneso 35, Genova, Italy
CNR–SPIN, Genova, Italy

e-mail: sorrentino@dima.unige.it; piana@dima.unige.it

Despite these notable potentialities, the application of neurophysiological modalities in research and, in particular, in the clinical workflow, is still limited by the fact that the interpretation of neurophysiological signals is often a very challenging task, which can be accomplished by exploiting powerful and sophisticated computational methods. Methodological problems in this framework are typically of three kinds. First, looking at the data recorded in the sensors' space, these can be encoded by means of classification algorithms based on *machine learning*, which aims at interpreting the measured times series as different clusters of data mirroring cortical activities of different natures. Second, these same experimental measurements can be modeled in the cortical source space by means of *regularization methods* that numerically solve the problem of reconstructing the neural currents from the recorded neurophysiological time series. Third, both the measured data and the data-modeling cortical sources can be interpreted in the framework of *connectivity maps* that can be constructed by designing either statistical or deterministic metrics in both the time and frequency domain.

The focus of this chapter is the source modeling of cortical sources from the knowledge of electro- or magneto-physiological data. This is a difficult mathematical issue, requiring the solution of a dynamical, highly ill-posed inverse problem. In particular, such ill-posedness implies that the neural configuration explaining the measurements is not unique (there is an infinite number of neural current distributions producing the same dataset, [9]) and this technical difficulty has inspired the adoption of many different strategies (some proposed by the inverse problems community, some others coming from the engineering framework) for the selection of the optimal neural constellation from the infinite set of possible solutions. The available algorithms are usually divided into two classes, based on the physical model used to represent the neural currents: *distributed* methods assume a continuous current distribution and solve a linear inverse problem consisting in recovering the dynamics of the local strength of the current density at each point of a computational grid introduced in the brain volume; *dipolar* methods introduce in the reconstruction procedure the information that the neural sources can be modeled as a set of a small number of point-like currents (current dipoles), whose parameters (position, orientation, strength) have to be recovered; assuming this dipolar model, the inverse problem is non linear since the measurements have a strongly non-linear dependence with respect to the unknown source locations. Distributed methods address the MEG data analysis as an image restoration problem whereby the restored map solves a constrained minimization; among these methods we recall Tikhonov regularization with L2 norm [16] and L1 norm [44] on the penalty term; various type of beamformers [3, 36], which spatially filters the signal to focus the source as a weighted combination of the measurements. On the other hand dipolar methods estimate dipoles parameters by non linear optimization techniques; among these we recall the dipole fitting procedure, the multiple signal classification (MUSIC) [26], and its evolution called recursively applied and projected MUSIC [27]; genetic algorithms [43], which find the sources with a trial-and-error procedure. However, both classes suffer from well-known shortcomings: among the distributed methods, the L2-norm estimation tends to provide solutions which

are too wide-spread from a neurophysiological viewpoint, L1-norm estimation requires an expert user interpretation of the reconstructed sources and beamformers suppress temporally correlated sources; among dipolar approaches, many non-linear optimization methods suffer from convergence problems; furthermore many methods consider a fixed (pre-determined or estimated from the data) number of sources, maintaining their position and often also orientation across time.

Recently, owing to the increase in the available computational power, Bayesian methods have become feasible [6, 24, 25, 37, 39]. They cast the inverse problem as a problem of statistical inference by means of Bayesian statistics: the unknown and the measurements are modelled as Random Variables (RV) and the solution of the inverse problem is the posterior probability density function of the unknown obtained by the Bayes theorem. Bayesian methods typically use the whole time series as input data and therefore can be considered as methods that address the dynamical source modeling problem. However, some of these approaches aim at reconstructing the source configuration at a given time point, while in others the sequential application of Bayes theorem, which requires a prior density at each time step, is mediated by the use of the Chapman-Kolmogorov equation. The main strengths of Bayesian approaches are in their notable generality (they can be applied in very general frameworks with minimal a priori assumptions) and in the fact that they can naturally account for a priori information coming from either physiology or experimental measurements provided by other modalities. On the other hand, these methods are typically demanding from a computational viewpoint and require the (often difficult) optimization of several input parameters.

The next sections will provide an overview of the state-of-the-art and some open issues concerned with the source modeling problem in neurophysiology. Specifically, Sect. 2 describes the forward modeling of data formation for the electro- and magneto-physiology problems; Sect. 3 will discuss several inversion approaches to data modeling and Sect. 4 will show an application to the case of measurements acquired during epileptic seizures. Finally, our conclusions will be offered in Sect. 5.

2 Data Formation

The mathematical framework that allows the definition of the forward problem for EEG, MEG, ECoG and SEEG is given by the Maxwell equations under the quasi-static condition, i.e.

$$\nabla \cdot E = \frac{\rho}{\epsilon_0} , \quad (1)$$

$$\nabla \times E = 0 , \quad (2)$$

$$\nabla \cdot B = 0 , \quad (3)$$

and

$$\nabla \times B = \mu_0 J , \quad (4)$$

where $E = E(r, t)$ and $B = B(r, t)$ are the electric and magnetic fields at position $r \in \mathbb{R}^3$ and time t , respectively; ϵ_0 and μ_0 are the electric and magnetic permittivities in vacuum, respectively, and $J = J(r, t)$ is the current density. Equation (2) allows the introduction of the scalar potential $V = V(r, t)$ such that

$$E = -\nabla V , \quad (5)$$

and Eq. (4) implies that

$$\nabla \cdot J = 0 . \quad (6)$$

In each point of the brain volume, the current density J is made by the superposition of two contributions, i.e.

$$J = J_p + J_v , \quad (7)$$

where the primary current J_p passing through the axons induces a passive volume current density J_v , which solves the local Ohm law

$$J_v(r, t) = \sigma(r)E(r, t) , \quad (8)$$

where $\sigma(r)$ is the tissue conductivity at r . Combining Eqs. (5)–(8) leads to

$$\nabla \cdot J_p(r, t) = -\nabla \cdot (\sigma(r)\nabla V(r, t)) , \quad (9)$$

which defines the forward model for the formation of EEG, SEEG, and ECoG signals. In fact, assuming that the conductivity is known (for example, thanks to the information provided by some anatomical medical imaging modality like Magnetic Resonance Imaging) and given a vector field $J_p = J_p(r, t)$, the numerical solution of Eq. (9) at sampled points r_1, \dots, r_N on the skull provides the EEG signal recorded by the helmet sensors placed at those points and at time t [34]. On the other hand, if the points are on either a limited portion of the cortex or a line through the brain, solving the equation provides the signal recorded by the sensors of an ECoG grid and a SEEG device, respectively.

The definition of the forward model mimicking the MEG signal requires the introduction of the vector potential $A = A(r)$ such that

$$B = -\nabla \times A , \quad (10)$$

whose existence is a consequence of Eq. (3). In order to constrain the many possible shapes of the field A satisfying (10), the Coulomb gauge condition

$$\nabla \cdot A = 0 \quad (11)$$

is imposed in the following. Under such condition, the vector potential solves the Poisson equation

$$\nabla^2 A = \mu_0 J, \quad (12)$$

whose solution can be written as

$$A(r) = \mu_0 \int_{\mathbb{R}^3} G(r, r') J(r') dr', \quad (13)$$

given the Green function

$$G(r, r') = \frac{1}{4\pi |r - r'|}. \quad (14)$$

Replacing (13)–(14) into (10) and exploiting the fact that

$$\nabla \times \frac{J(r')}{|r - r'|} = J(r') \times \frac{r - r'}{|r - r'|^3} \quad (15)$$

leads to the Biot-Savart equation

$$B(r, t) = \frac{\mu_0}{4\pi} \int_{\mathbb{R}^3} [J_p(r', t) - \sigma(r') \nabla V(r', t)] \times \frac{r - r'}{|r - r'|} dr'. \quad (16)$$

The Biot-Savart equation is combined with Eq. (9) to define the forward model for MEG. In fact, assuming once again that the conductivity is known, and given the vector field $J_p = J_p(r, t)$, the computation of $V(r, t)$ in (9) for r everywhere in the cortex and then of $B(r, t)$ in (16) at specific points r outside (but close) the scalp provides the signal recorded by the MEG sensors, placed at those points.

3 The Inverse Problem

Spatio-temporal localization of brain activity from MEG/EEG data requires to solve a so-called *inverse problem*: given the recorded time series, compute an estimate of the underlying neural currents. In a discrete setting, we will be calling the recorded data $y_{1:T} := \{y_t\}_{t=1}^T$, where y_t is a vector whose components represent the data recorded by the sensors. The spatio-temporal evolution of the neural current is

represented by $x_{1:T}$, where x_t is a vector that contains a discrete representation of the neural current; more details on such discrete representation will be given shortly. We assume that the two processes are related by

$$y_t = f(x_t) + n_t \quad t = 1, \dots, T , \quad (17)$$

where n_t is the noise at time t and the *forward model* $f(\cdot)$ is not further specified at this stage, because its actual form depends on the specific choice of the discrete representation of the neural current.

It is well known that the MEG/EEG inverse problem is ill-posed; in particular, it has no unique solution, because there exist current distributions that do not produce any magnetic field. In regularization theory, this is usually dealt with by minimizing a functional that combines the fit with the data and some penalty term on the unknown, so as to encourage some desired property of the solution (smoothness, sparsity, and so on).

$$\hat{x}_{1:T} = \arg \min \left(\|y_{1:T} - f(x_{1:T})\|_2^2 + \lambda P(x_{1:T}) \right) \quad (18)$$

In the Bayesian framework, ill-posedness is dealt with by changing the point of view: rather than trying to solve the inverse problem, one aims at quantifying the amount of information that is available on the unknown underlying neural currents. The information is coded in probability distributions: specifically, one is interested in computing or approximating the posterior distribution $p(x_{1:T}|y_{1:T})$, i.e. the probability distribution of the unknown, conditioned on the measured data; this distribution is notoriously given by Bayes theorem:

$$p(x_{1:T}|y_{1:T}) \propto p(y_{1:T}|x_{1:T})p(x_{1:T}) \quad (19)$$

where $p(y_{1:T}|x_{1:T})$ is the *likelihood* function and $p(x_{1:T})$ is the *prior* distribution.

Notably, there is a clear connection between these two approaches: under Gaussian noise assumptions, the likelihood function is

$$p(y_{1:T}|x_{1:T}) \propto \exp(-\|y_{1:T} - f(x_{1:T})\|_2^2) \quad (20)$$

and the regularized solution described by (18) can be interpreted as the maximum of the log-posterior described by (19) when the prior is of the form

$$p(x_{1:T}) \propto \exp(-P(x_{1:T})) \quad (21)$$

However, Bayesian methods typically aim at approximating/calculating the whole posterior distribution, rather than just its maximum, in order to assess the uncertainty of the estimate. As one can easily understand, the whole posterior distribution (19) lives in a very high-dimensional space, which makes it almost hopeless to compute it or approximate it. As we will see, Bayesian methods will try to simplify the problem in different ways.

3.1 Classification of Inverse Methods

The specific choice of the discrete representation of the neural current will lead to a first classification of inverse methods.

Indeed, one can assume that the neural current is a continuous vector field in the whole brain volume; for computational purposes, such continuous vector field can be discretized by choosing an appropriate grid of points $\{r_i\}_{i=1}^N$ with N large enough. With reference to the notation of the previous Section, the discrete representation of the neural current is then such that $(x_t)_i = J^p(r_i, t)$. In this case, we will refer to as the *distributed* source model, the inverse problem can be written as a linear problem

$$y_t = \sum_{i=1}^N G(r_i) \cdot (x_t)_i + n_t = \mathbf{G} x_t + n_t \quad (22)$$

where \mathbf{G} is usually called *leadfield* matrix. In this case, N is usually a relatively large number ($\sim 10,000$) of voxels.

Alternatively, one can assume that the neural current is a finite set of point-like sources, termed *current dipoles*: the number of dipoles is not known a priori, neither their locations, orientations and strengths. Here, x_t is a finite set of dipoles: $x_t = (r_t^1, q_t^1, \dots, r_t^N, q_t^N)$. In this case, we still have

$$y_t = \sum_{i=1}^{N_t} G(r_t^i) \cdot q_t^i + n_t \quad (23)$$

but this does not reduce to a linear problem, because the source locations are unknown. Typical dipole models contain a relatively small (< 10) number of dipoles.

A second type of classification concerns the way different inverse methods treat the temporal dependence of the data. Indeed, due to the high temporal resolution of MEG/EEG data (around 1,000 Hz), the data tend to be fairly smooth, and the underlying electrical currents are likely to exhibit a similar degree of smoothness in the temporal variable. In fact, most methods developed until the mid 2000s solved the inverse problem independently at different time points; this is mainly due to the limited computational resources available at that time, that made the solution of the full spatio-temporal unfeasible; we will refer to these methods as *static* inverse methods. More recently developed methods try to model the temporal dependence in order to reduce the ill-posedness further; while this remains impractical in some cases (see e.g. the network of High Performance Computers needed to run a Kalman filter [23]), it is becoming increasingly doable. We will refer to these methods as *dynamic* methods.

3.2 Methods for the Distributed Model

We start with a short review of static methods, that are still the most widely used ones thanks to their computational efficiency.

3.2.1 Minimum Norm Estimate (MNE)

The Minimum Norm Estimate [16] (MNE) corresponds to solving a simplified version of (18), where each time point is treated independently and with $P(x_t) = \|x_t\|_2^2$; in Bayesian terms, this amounts to calculating the MAP of $p(x_t|y_t)$. The advantage of this approach is that analytical calculation is feasible, thanks to the relatively small size of the matrices involved:

$$\hat{x}_t^{MNE} = (G^T G + \lambda I)^{-1} G^T y_t \quad (24)$$

On the other hand, it is well known that this choice has two main drawbacks: first, it tends to provide overly smooth estimates of the neural current; second, it tends to create a bias towards superficial sources, because superficial sources are closer to the sensors and therefore weaker sources, encouraged by the ℓ^2 norm, produce stronger signals. To overcome this last point, a *weighted* MNE variant has been proposed in [22], in which the penalty term is designed to compensate for the depth bias $P(x_t) = x_t^T R^{-1} x_t$.

To overcome the first point [44] proposes an ℓ^1 penalty term $P(x_t) = \|x_t\|_1$. A recent Bayesian alternative has been described in [5], where a Bayesian hierarchical model is used to code a priori anatomical information, and a hyper-parameter tunes the degree of sparsity of the solution.

3.2.2 Mixed Norm Estimates

In order to account for the temporal smoothness of MEG/EEG data, in the last decade several regularization methods have been proposed [14, 31, 42] that use a penalty with a mixed norm [20]. An example of these mixed norms is the $\ell_{1,2}$ norm that is defined as follows

$$\|x_{1:T}\|_{1,2} = \sum_{i=1}^N \sqrt{\sum_{t=1}^T (x_t^i)^2} \quad (25)$$

By using $P(x) = \|x\|_{1,2}$ one obtains a convex optimization problem; in this case, the analytical solution is not available, but the convexity of the functional guarantees uniqueness of the solution, and convergence results for a class of iterative algorithms with proximal operators [32]. This type of mixed norm methods provide pleasant

solutions with few active regions that exhibit a smooth temporal behaviour. On the other hand, they are slightly tricky to use, as the iterative algorithms used to compute the solution require some fine-tuning of the parameters.

3.2.3 Kalman Filtering

In a Bayesian setting, it is natural to include prior information on the temporal behaviour of the neural sources by means of a Kalman filter. Rather than computing the whole posterior distribution (19), the Kalman filter provides a mean to compute the so called *filtering* distributions, i.e. $p(x_t|y_{1:t})$, for $t = 1, \dots, T$. Indeed, by assuming that the neural currents and the data form a Hidden Markov Model [35], the two following “Bayesian filtering” equations provide a formal way to calculate the filtering distributions:

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}, \quad (26)$$

$$p(x_{t+1}|y_{1:t}) = \int p(x_{t+1}|x_t)p(x_t|y_{1:t})dx_t. \quad (27)$$

For a linear Gaussian model, all the filtering distributions are Gaussian, and are therefore completely characterized by their mean and covariance matrix. If we denote by $\bar{x}_{t|s}$ and $\Gamma_{t|s}$ the mean and covariance matrix, respectively, of $p(x_t|y_{1:s})$, then the Kalman filter is given by:

$$\bar{x}_{t|t} = \bar{x}_{t|t-1} + K_t(y_t - G\bar{x}_{t|t-1}) \quad (28)$$

$$\Gamma_{t|t} = (I - K_t G)\Gamma_{t|t-1} \quad (29)$$

where

$$K_t = \Gamma_{t|t-1} G^T (G\Gamma_{t|t-1} G^T + \Sigma)^{-1} \quad (30)$$

$$\bar{x}_{t+1|t} = F\bar{x}_{t|t} \quad (31)$$

$$\Gamma_{t+1|t} = F\Gamma_{t|t}F^T + \Delta \quad (32)$$

However, these formulas are computationally demanding, due to the size of the involved matrices; in particular, for a brain discretization containing 10,000 points, the covariance matrix of an unconstrained model is a square matrix of size $30,000 \times 30,000$. For this reason, in [24] a network of high performance computers is used to calculate the full problem; in order to reduce the computational cost, in [13, 21] the

full Kalman filter is decomposed in a large number of small-dimensional problems by considering only short-range interactions.

3.3 Methods for the Dipolar Model

Methods for estimating multiple current dipoles have to face two main difficulties: first, the number of dipoles is unknown, which translates into a model order selection problem; second, the problem is genuinely non-linear, implying that the corresponding functional to be minimized has local minima, and deterministic algorithms are doomed to failure.

3.3.1 Global Optimization Methods

For years, dipole fitting methods have relied on subjective choices of initialization and non-linear optimization of individual dipole parameters with Levenberg–Marquardt. In [18], a multi-start algorithm is proposed that tried to avoid subjectivity of the results, by several random initialization of the optimization algorithm. In [43] the authors describe the use of simulated annealing and of genetic algorithms for optimizing a regularized functional. In both cases, the number of dipoles has to be fixed in advance.

3.3.2 Bayesian Monte Carlo Methods for Static Dipoles

In the Bayesian setting, one can formally include the number of dipoles among the unknowns, and try and make inference on it by exploring variable-dimension models with reversible jumps [15]. In order to make the problem computationally tractable, it is common to assume that the number of dipoles and their locations do not change in the time window under investigation; this means that the posterior distribution of interest is

$$p(N, r^1, \dots, r^N, q_1^1, \dots, q_t^1, \dots, q_T^N, \dots, q_T^N | y_{1:T}) \quad (33)$$

Monte Carlo sampling on this space is still prohibitive but, thanks to the linear relationship between the data and the dipole moments, one can marginalize out (i.e., treat analytically) the dipole moments and do Monte Carlo sampling only for $p(N, r^1, \dots, r^N)$. In [19] the authors assume a uniform prior for the number of dipoles, and use reversible-jump Markov Chain Monte Carlo (MCMC) to approximate the posterior distribution. In [38, 41], the authors assume a Poisson prior for the number of dipoles, and use sequential Monte Carlo (SMC) samplers [11] to approximate the posterior distribution; as SMC samplers employ multiple

Markov Chains running in parallel, they are less likely to remain trapped in local maxima.

3.3.3 Bayesian Monte Carlo Methods for Dynamic Dipoles

Starting from [37], Bayesian inference for dynamic dipoles has been described in several studies [2, 8, 25, 39, 40, 45]. Like for the distributed case, the easiest thing to do is to use the Bayesian filtering recursion described by Eqs. (26)–(27). However, in this case the analytic solution is not available, because the data depend non-linearly on the dipole locations and on the number of dipoles. Therefore one can use *particle* methods [12] to approximate the filtering distributions. The idea is to propagate in time a set of Monte Carlo samples (particles) by importance sampling and Markov Chain Monte Carlo, possibly combined. As time goes by, the sample set accumulates information on the underlying sources. By their nature, filtering algorithms tend to provide a poor localization of dipolar sources at their appearance, just because there is little information accumulated from the past. For this reason, in [45] a particle *smoothing* algorithm is proposed, that approximates the conditional distributions $p(x_t|y_{1:T})$ for $t = 1, \dots, T$.

4 An Application to Epilepsy

Thanks to their outstanding temporal resolution, MEG and EEG provide a unique window into the human brain, with many applications ranging from basic neuroscience to the clinical use. Here we provide some insight into a clinical application: the pre-surgical evaluation of epileptic patients. Indeed, this is the most straightforward clinical application of the source localization problem.

Epilepsy is a chronic disease that affects about 50 million people worldwide, with 30% of cases being refractory to medication. Focal epilepsy is the most common type of epilepsy in adults [29], where epileptic seizures are generated in a relatively small area of the brain, referred to as *epileptogenic zone* (EZ). When focal epilepsy is refractory to medication, surgical ablation of the EZ is considered as a possible solution. However, localization of the EZ is often not straightforward: current clinical practice envisages the use of intra-cerebral recordings [1, 7]. The use of MEG/EEG for localization of the EZ still needs further validation, but is increasingly considered; in particular, recent studies suggest that MEG may be capable of estimating not only the location, but also the size of the EZ [4].

Here we consider, as an example, the localization of an EZ from 32-channel EEG recordings. Data come from the example BESA database (BESA, Munich): 164 spikes were recorded from an epileptic patient and averaged, using the peak of the spike as a trigger. The EEG signals were sampled at 320 Hz and filtered with a Butterworth forward high-pass filter with cut-off frequency of 5 Hz. The head model is a three-layer model including the brain, the skull and the scalp; while there is no

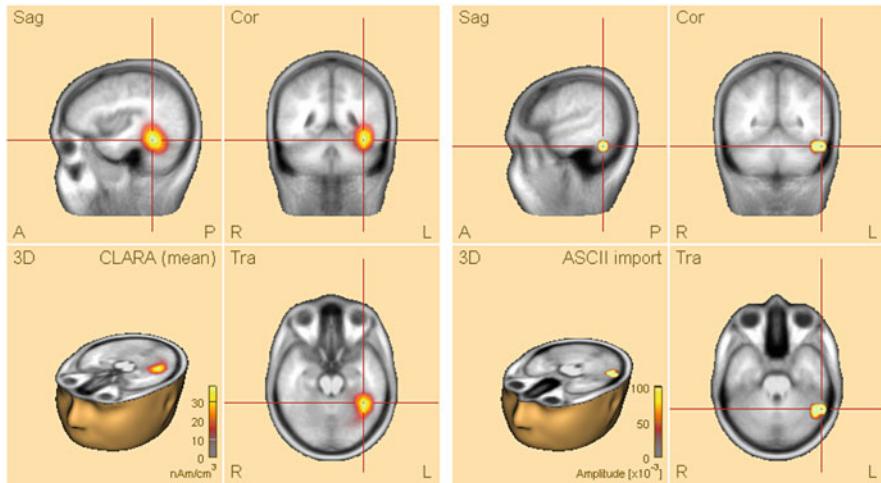


Fig. 1 Reconstructions obtained for an epileptic patient in the proximity of the spike. Left: CLARA. Right: smoothing algorithm

cerebrospinal fluid (CSF) in the model, the effect of the CSF is partly accounted for by assuming an anisotropic skull conductivity. The tangential conductivity within the skull is modeled to be three times larger than the radial conductivity across the skull.

In Fig. 1 we report the results obtained by applying two different methods: CLARA (Classical LORETA Analysis Recursively Applied), which is a distributed method with a Laplacian penalty term, and depth-weighting; a particle smoothing algorithm, which is used to approximate the smoothing distribution for a dynamic multi-dipole model. The figure reports the reconstructed map at a particular time, in correspondence with the peak of the spike. We notice that the two figures have substantially different interpretations: while the images produced by CLARA represent the estimated strength of the neural current, the images produced by the smoothing algorithm represent the posterior probability of there being a dipole at any particular location. Despite of these differences, the two methods appear to agree pleasantly on the putative location of the EZ.

5 Conclusions

Even though EEG has been around for almost a century, and MEG for almost half of it, inverse modeling of MEG/EEG data remains a challenging research field, due to the inherent ill-posedness of the problem. After the very first localization attempts in the 1980s, the search for better spatial priors in the 1990s, the inclusion of the temporal dimension in the 2000s, most likely the present decade will be

remembered for the prevalence of connectivity studies, and for the transfer to the clinical practice. Concerning connectivity studies, a lot of effort is currently spent in understanding the dynamics of the connectivity networks. On the other hand, the transfer to clinical practice is requiring a substantial validation work, using both simulations and experimental data. Finally, a fascinating future perspective concerns the use of these devices for brain computer interfaces. For sure we are only at the beginning of exciting developments in many different fields, including thought-controlled devices and improved clinical treatment of epilepsy.

References

1. Andrzejak, R.G., David, O., Gnatkovsky, V., Wendling, F., Bartolomei, F., Francione, S., Kahane, P., Schindler, K., De Curtis, M.: Localization of epileptogenic zone on pre-surgical intracranial EEG recordings: toward a validation of quantitative signal analysis approaches. *Brain Topogr.* **28**(6), 832–837 (2015)
2. Antelis, J.M., Minguez, J.: DYNAMO: concurrent dynamic multi-model source localization method for EEG and/or MEG. *J. Neurosci. Methods* **212**(1), 28–42 (2013)
3. Barnes, G.R., Hillebrand, A.: Statistical flattening of MEG beamformer images. *Hum. Brain Mapp.* **18**(1), 1–12 (2003)
4. Bouet, R., Jung, J., Delpuech, C., Ryvlin, P., Isnard, J., Guenot, M., Bertrand, O., Mauguière, F.: Towards source volume estimation of interictal spikes in focal epilepsy using magnetoencephalography. *NeuroImage* **59**(4), 3955–3966 (2012)
5. Calvetti, D., Pascarella, A., Pitolli, F., Somersalo, E., Vantaggi, B.: A hierarchical krylov–bayes iterative inverse solver for MEG with physiological preconditioning. *Inverse Probl.* **31**(12), 125005 (2015)
6. Campi, C., Pascarella, A., Sorrentino, A., Piana, M.: A rao-blackwellized particle filter for magnetoencephalography. *Inverse Probl.* **24**, 025023 (2008)
7. Cardinale, F., Cossu, M., Castana, L., Casaceli, G., Schiariti, M.P., Misericocchi, A., Fuschiello, D., Moscato, A., Caborni, C., Arnulfo, G., et al.: Stereoelectroencephalography: surgical methodology, safety, and stereotactic application accuracy in 500 procedures. *Neurosurgery* **72**(3), 353–366 (2013)
8. Chen, X., Särkkä, S., Godsill, S.: A bayesian particle filtering method for brain source localisation. *Digit. Signal Process.* **47**, 192–204 (2015)
9. Dassios, G., Fokas, A.S.: The definite non-uniqueness results for deterministic EEG and MEG data. *Inverse Probl.* **29**(6), 065012 (2013)
10. de Pasquale, F., Della Penna, S., Snyder, A.Z., Lewis, C., Mantini, D., Marzetti, L., Belardinelli, P., Ciancetta, L., Pizzella, V., Romani, G.L., et al.: Temporal dynamics of spontaneous MEG activity in brain networks. *Proc. Natl. Acad. Sci.* **107**(13), 6040–6045 (2010)
11. Del Moral, P., Doucet, A., Jasra, A.: Sequential monte carlo samplers. *J. R. Stat. Soc. Ser. B (Stat Methodol.)* **68**(3), 411–436 (2006)
12. Doucet, A., Johansen, A.M.: A tutorial on particle filtering and smoothing: fifteen years later. *Handb. nonlinear filter.* **12**(656–704), 3 (2009)
13. Galka, A., Yamashita, O., Ozaki, T., Biscay, R., Valdés-Sosa, P.: A solution to the dynamical inverse problem of EEG generation using spatiotemporal kalman filtering. *NeuroImage* **23**(2), 435–453 (2004)
14. Gramfort, A., Kowalski, M., Hämäläinen, M.: Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Phys. Med. Biol.* **57**(7), 1937 (2012)
15. Green, P.J.: Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* **82**, 711–732 (1995)

16. Hämäläinen, M.S., Ilmoniemi, R.J.: Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* **32**(1), 35–42 (1994)
17. Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V.: Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* **65**(2), 413 (1993)
18. Huang, M., Aine, C.J., Supek, S., Best, E., Ranken, D., Flynn, E.R.: Multi-start downhill simplex method for spatio-temporal source localization in magnetoencephalography. *Elec-troencephalogr. Clin. Neurophysiol./Evoked Potentials Sect.* **108**(1), 32–44 (1998)
19. Jun, S.C., George, J.S., Paré-Blagoev, J., Plis, S.M., Ranken, D.M., Schmidt, D.M., Wood, C.C.: Spatiotemporal bayesian inference dipole analysis for MEG neuroimaging data. *NeuroImage* **28**(1), 84–98 (2005)
20. Kowalski, M.: Sparse regression using mixed norms. *Appl. Comput. Harmon. Anal.* **27**(3), 303–324 (2009)
21. Lamus, C., Hämäläinen, M.S., Temereanca, S., Brown, E.N., Purdon, P.L.: A spatiotemporal dynamic distributed solution to the MEG inverse problem. *NeuroImage* **63**(2), 894–909 (2012)
22. Lin, F.-H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S.: Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage* **31**(1), 160–171 (2006)
23. Long, C.J., Purdon, P.L., Temereanca, S., Desai, N.U., Hamalainen, M., Brown, E.N.: Large scale kalman filtering solutions to the electrophysiological source localization problem—a MEG case study. In: *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pp. 4532–4535. IEEE, New York (2006)
24. Long, C.J., Purdon, P.L., Temereanca, S., Desai, N.U., Hämäläinen, M.S., Brown, E.N.: State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing. *Ann. Appl. Stat.* **5**(2B), 1207 (2011)
25. Miao, L., Zhang, J.J., Chakrabarti, C., Papandreou-Suppappola, A.: Efficient bayesian tracking of multiple sources of neural activity: algorithms and real-time FPGA implementation. *IEEE Trans. Signal Process.* **61**(3), 633–647 (2013)
26. Mosher, J.C., Leahy, R.M.: Recursive music: a framework for EEG and MEG source localization. *IEEE Trans. Biomed. Eng.* **45**(11), 1342–1354 (1998)
27. Mosher, J.C., Leahy, R.M.: Source localization using recursively applied and projected (rap) music. *IEEE Trans. Signal Process.* **47**(2), 332–340 (1999)
28. Murakami, H., Wang, Z.I., Marashly, A., Krishnan, B., Prayson, R.A., Kakisaka, Y., Mosher, J.C., Bulacio, J., Gonzalez-Martinez, J.A., Bingaman, W.E., et al.: Correlating magnetoencephalography to stereo-electroencephalography in patients undergoing epilepsy surgery. *Brain* **139**(11), 2935–2947 (2016)
29. Nguyen, D.K., Mbaefou, M.T., Nguyen, D.B., Lassonde, M.: Prevalence of nonlesional focal epilepsy in an adult epilepsy clinic. *Can. J. Neurol. Sci.* **40**(2), 198–202 (2013)
30. Niedermeyer, E., Lopes da Silva, F.H.: *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, Philadelphia (2005)
31. Ou, W., Hämäläinen, M.S., Golland, P.: A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* **44**(3), 932–946 (2009)
32. Parikh, N., Boyd, S., et al.: Proximal algorithms. *Found. Trends® Optim.* **1**(3), 127–239 (2014)
33. Pasquarella, A., Todaro, C., Clerc, M., Serre, T., Piana, M.: Source modeling of electrocorticography (ecog) data: stability analysis and spatial filtering. *J. Neurosci. Methods* **263**, 134 (2016)
34. Pursiainen, S., Sorrentino, A., Campi, C., Piana, M.: Forward simulation and inverse dipole localization with the lowest raviart-thomas elements for electroencephalography. *Inverse Probl.* **27**, 045003 (2011)
35. Rabiner, L., Juang, B.: An introduction to hidden markov models. *IEEE ASSP Mag.* **3**(1), 4–16 (1986)
36. Sekihara, K., Nagarajan, S.S., Poeppel, D., Marantz, A., Miyashita, Y.: Reconstructing spatio-temporal activities of neural sources using an MEG vector beamformer technique. *IEEE Trans. Biomed. Eng.* **48**(7), 760 (2001)

37. Somersalo, E., Voutilainen, A., Kaipio, J.P.: Non-stationary magnetoencephalography by bayesian filtering of dipole models. *Inverse Probl.* **19**(5), 1047 (2003)
38. Sommariva, S., Sorrentino, A.: Sequential monte carlo samplers for semi-linear inverse problems and application to magnetoencephalography. *Inverse Probl.* **30**(11), 114020 (2014)
39. Sorrentino, A., Parkkonen, L., Pascarella, A., Campi, C., Piana, M.: Dynamical MEG source modeling with multi-target bayesian filtering. *Hum. Brain Mapp.* **30**(6), 1911–1921 (2009)
40. Sorrentino, A., Johansen, A.M., Aston, J.A.D., Nichols, T.E., Kendall, W.S., et al.: Dynamic filtering of static dipoles in magnetoencephalography. *Ann. Appl. Stat.* **7**(2), 955–988 (2013)
41. Sorrentino, A., Luria, G., Aramini, R.: Bayesian multi-dipole modelling of a single topography in MEG by adaptive sequential monte carlo samplers. *Inverse Probl.* **30**(4), 045010 (2014)
42. Tian, T.S., Huang, J.Z., Shen, H., Li, Z.: A two-way regularization method for MEG source reconstruction. *Ann. Appl. Stat.* **6**, 1021–1046 (2012)
43. Uutela, K., Hamalainen, M., Salmelin, R.: Global optimization in the localization of neuro-magnetic sources. *IEEE Trans. Biomed. Eng.* **45**(6), 716–723 (1998)
44. Uutela, K., Hämäläinen, M., Somersalo, E.: Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* **10**(2), 173–180 (1999)
45. Vivaldi, V., Sorrentino, A.: Bayesian smoothing of dipoles in magneto-/electroencephalography. *Inverse Probl.* **32**(4), 045007 (2016)
46. Wolpaw, J.R., McFarland, D.J., Neat, G.W., Forneris, C.A.: An EEG-based brain-computer interface for cursor control. *Electroencephalogr. Clin. Neurophysiol.* **78**(3), 252–259 (1991)