

<a href="#">Report on modelling LFP and optical imaging signals</a>
---

Report Version: 1

Report Preparation: 2

Classification: PU

Contract Start Date: 01/09/2005

Duration: 5 Years

Project Coordinator: Karlheinz Meier (Heidelberg)

Partners: ENSEIRB Bordeaux, CNRS (Gif-sur-Yvette, Marseille), U Debrecen, TU Dresden, U Freiburg, TU Graz, U Heidelberg, EPFL Lausanne, Funetics S.a.r.l., U London, U Plymouth, INRIA, KTH Stockholm



**Project funded by the European  
Community under the “Information Society  
Technologies” Programme**

<b>DELIVERABLES TABLE</b>
---------------------------

<b>Project Number:</b> <i>FP6-2004-IST-FETPI 15879</i>
--

**Project Acronym:** *FACETS*

<b>Title:</b> <i>Fast Analog Computing with Emergent Transient States</i>
---

Del. No.	Revision	Title	Type <sup>1</sup>	Classification <sup>2</sup>	Due Date	Issue Date
5-5		Report on modelling LFP and optical imaging signals	R	Pub	54	54

<sup>1</sup> R: Report; D: Demonstrator; S: Software; W: Workshop; O: Other – Specify in footnote

<sup>2</sup> Int.: Internal circulation within project (and Commission Project Officer + reviewers if requested)

Rest.: Restricted circulation list (specify in footnote) and Commission SO + reviewers only

IST: Circulation within IST Programme participants

FP5: Circulation within Framework Programme participants

Pub.: Public document

**DELIVERABLE SUMMARY SHEET**

Project Number: FP6-2004-IST-FETPI 15879

Project Acronym: FACETS

Title: Fast Analog Computing with Emergent Transient States

Deliverable N°: 5-5

Due date: 54

Delivery Date: 54

Short description:

This document reports the modelling results obtained in the FACETS project in the field of macroscopic modelling. This theme emerged during the FACETS project, due to (1) the availability of experimental data on macroscopic measurements of neuronal activity, such as local field potentials (LFPs) and optical imaging (VSD) signals; (2) the inadequacy of traditional models to handle and model such data. This report contains, under the form of published manuscripts, a number of theoretical approaches to obtain macroscopic models adequate to deal with LFP and VSD measurements.

Partners owning: 6a, 6b

Partners contributed: 6a, 6b, 14, 8, 3

Made available to: (Private)

The modeling of LFP and optical imaging signals is a theme that emerged during the FACETS project, as the different approaches outlined here were not initially planned in the proposal. The purpose of the present report is to give an overview of this approach and outline possible perspectives for future modeling approaches.

## **MODELLING LOCAL FIELD POTENTIALS**

Local field potentials (LFPs) are routinely measured in electrophysiological experiments, and consist of recordings of the extracellular electric potential, in general using high-impedance electrodes. Contrary to intracellular potentials, whose origin is well known today, the origin of extracellular signals is still debated. Many electrical events in the network, such as synaptic currents in dendrites, action potentials or glial potentials all generate extracellular currents and thus potentially contribute to different extents in the genesis of the LFP. Because the extracellular space is shared by many cells, the combined electrical activity of a large number of neurons is believed to be at the basis of the LFP. There is also evidence that the electric signals undergo low-pass filtering when flowing in extracellular space, as well as attenuate with distance, and thus the LFP can not be simply predicted by the simple summation of electric events but needs to be handled by appropriate models.

Because LFP signals constitute an important signal recorded experimentally, and because it is not easy to incorporate into network models, we have studied theoretically the the genesis of LFP signals. It is also important to relate the LFP with cellular activity, which revealed that both signals are consistent only if the signal is subject to a filtering of  $1/f$  type (Bedard et al., 2006). This type of filter was later found to be consistent with the effect of ionic diffusion (Bedard and Destexhe, 2009). Further, we used a simple kernel based method to transform spiking activity of a network model to macroscopic signals (such as EEG/LFP). Depending on the parameters a cortical network model can exhibit a range of activity states but asynchronous irregular (AI) and synchronous-irregular (SI) activity state are deemed physiological. Consistent with this, network activity in AI resulted in a biologically realistic macroscopic signals corresponding to awake state and SI state generated macroscopic signal that resembled more with the pathological activity (Meier et al 2007).

## **MODELLING OPTICAL IMAGING SIGNALS**

Similar to LFPs, optical imaging using voltage-sensitive dyes (VSD) provides important insight into network dynamics as it is currently the only technique that allows to visualize a large population activity at both high spatial and temporal resolution. It is based on voltage-sensitive dyes (VSDs), which binds to the membrane and transform variations in the membrane potential into optical fluorescence. The emitted fluorescence is linearly correlated with changes in membrane potential per unit of membrane area of all the stained surfaces (Grinvald and Hildesheim 2004), meaning all neuronal cells present in the cortex, excitatory and inhibitory cells, all processes (soma, dendrites, axons) but also all non-neuronal cells, such as glial cells. The measured signal is therefore a multiplexed signal that combines all these components. Therefore, although the underlying mechanism of the VSDI is well understood (Grinvald and Hildesheim 2004), the recorded signal remains very complex and it is difficult to isolate the relative contributions of such a large amount of intermingled components (Chemla and Chavane 2009). The identification of these various signal



components is, in fact, possible for in vitro optical imaging "depth" recordings, as in the hippocampal slice, as was demonstrated by Plenz and Aertsen (1993) by adapting the LFP-based Current-Source-Density (CSD) framework (Nicholson and Freeman, 1975) to the 2D real-time optical imaging profiles. However, for in vivo optical imaging "surface" recordings, such identification has, until now, not been possible.

We have proposed a biological cortical column model, at an intermediate mesoscopic scale, in order to better understand and interpret biological sources of voltage-sensitive dye imaging signal (VSD signal). To perform a quantitative analysis of the relative contributions to the VSD signal, a detailed compartmental model was developed at a scale corresponding to one pixel of optical imaging. The generated model was used to solve the VSD direct problem, i.e. generate a VSD signal, given the neural substrate parameters and activities. Here, we confirm and quantify the fact that the VSD signal is the result of an average from multiple components. Not surprisingly, the compartment contributing most to the signal is the upper layer dendrites of excitatory neurons. However, our model suggests that inhibitory cells, spiking activity and deep layers contributions are also significant and, more unexpected, are dynamically modulated with time and response strength (Chemla & Chavane, in revision).

We have further evaluated mesoscopic neural field models as a tool to simulate VSD signals in slices from ferret visual cortex. These signals reveal complex spatiotemporal patterns in response to localised pulsed electrical stimulation. Given proper stimulating electrode models, close matches between experiments and simulations are possible. Results suggest specific excitatory-inhibitory circuits shaping the local temporal dynamics as well as certain non-linearities and horizontal spread along long-range synaptic pathways within layer 2/3 shaping long-range spatial and temporal properties of the response patterns (Symes and Wennekers, 2009).

## **MACROSCOPIC MODELS**

These two approaches revealed that the macroscopic measurements are not simply deduced from the microscopic data (cellular activity, connectivity, etc). A specific type of modeling must be used to correctly capture the macroscopic aspects of neuronal activity, as visible in LFP and VSD signals. In FACETS, this type of modeling was explored along several aspects.

First, neural mass models were designed, where the variables are the population activities. A new mathematical mean-field approach was developed by INRIA to describe the populations activity in the limit of an infinite number of neurons (Faugeras, Touboul and Cessac 2009). Each neuron is described by an equation that relates the time variation of its membrane potential to its mean firing rate, and to the mean firing rate of its neighbours. The synaptic weights are assumed to be randomly distributed. The description also takes into account random currents. The mean-field description is a set of equations (one per population) bearing on the probability distribution of the mean average membrane potentials conditionally to the law of the synaptic weights and the Wiener processes that appear in the random currents. These equations are quite interesting, non-Markovian, and they provide a link between the microscopic, activity-based description, and the mesoscopic formalism of neural masses.

The INRIA partner also continued the mathematical analysis of neural field models. They are spatial arrangements of neural masses that provide a macroscopic, activity-based description, of cortical areas, e.g. visual areas. In (Faugeras, Veltz, Grimbort 2009) the

existence, uniqueness and stability of the solutions of these multi-population neural field equations was studied in detail. In particular the case of the stationary, i.e. independent of time, solutions is important because it qualifies as models of the so-called persistent states whose existence has been revealed experimentally.

Second, a macroscopic formalism was derived for networks of spiking neurons in activity states with balanced excitation-inhibition, such as AI states. In this case, mean firing rates are not sufficient, because the mean excitation and inhibition are balanced and it is the fluctuations of excitation and inhibition that drive the firing of the neurons. Thus, in this case, a mean-field formalism must be derived based on first and second-order statistics (mean and covariances). This was done for networks of current-based and conductance-based neurons using a Master Equation formalism (El Boustani and Destexhe, 2009). This formalism successfully describes the macroscopic activity of such networks which can be thought of modeling one “pixel” of the VSD imaging data. The generalization to 2D frames of pixels necessitates to link macroscopic and microscopic connectivity, thereby providing estimates of the “effective connectivity” (El Boustani et al., 2009). A macroscopic model was also developed for spike-frequency adapting neurons in the high-conductance state (Muller et al., 2007). The latter approach is particularly relevant for describing the ensemble dynamics and of neurons in AI states, as well as their response to stimuli.

## **PERSPECTIVES FOR FUTURE WORK**

This exploration of the macroscopic aspects of neuronal activity is a fascinating area which is motivated by following points. (1) The information is evidently coded in large neuronal ensembles rather than in single neurons, but the recordings of such large ensembles (LFP and VSD data) are not trivial to interpret; (2) these signals are widely available today but are not yet incorporated in models because appropriate formalisms are not necessarily available to link macroscopic and microscopic scales. We therefore conclude that it is essential to develop such macroscopic formalisms to better interpret the experimental data, extract relevant information and better constrain the models using data from multiple scales. The use of data from multiple scales is an essential aspect for future research in experimental and computational neuroscience. In this respect, a promising approach is to derive macroscopic (or mean-field) models of the “microscopic” spiking models in FACETS, with a goal to relate the two scales.

## Does the $1/f$ Frequency Scaling of Brain Signals Reflect Self-Organized Critical States?

C. Bédard,<sup>1,2</sup> H. Kröger,<sup>2</sup> and A. Destexhe<sup>1,\*</sup>

<sup>1</sup>*Integrative and Computational Neuroscience Unit (UNIC), CNRS, Gif-sur-Yvette, France*

<sup>2</sup>*Department of Physics, Université Laval, Québec, Canada*

(Received 27 March 2006; published 13 September 2006)

Many complex systems display self-organized critical states characterized by  $1/f$  frequency scaling of power spectra. Global variables such as the electroencephalogram, scale as  $1/f$ , which could be the sign of self-organized critical states in neuronal activity. By analyzing simultaneous recordings of global and neuronal activities, we confirm the  $1/f$  scaling of global variables for selected frequency bands, but show that neuronal activity is not consistent with critical states. We propose a model of  $1/f$  scaling which does not rely on critical states, and which is testable experimentally.

DOI: [10.1103/PhysRevLett.97.118102](https://doi.org/10.1103/PhysRevLett.97.118102)

PACS numbers: 87.19.Nn, 02.50.Ey, 05.40.-a, 89.75.Da

Self-organized critical states are found for many complex systems in nature, from earthquakes to avalanches [1,2]. Such systems are characterized by scale invariance, which is usually identified as a power-law distribution of variables such as event duration or the waiting time between events.  $1/f$  noise is usually considered as a footprint of such systems [1].  $1/f$  frequency scaling is interesting, because it betrays long-lasting correlations in the system, similar to the behavior near critical points.

Several lines of evidence point to the existence of such critical states in brain activity. Global variables, such as the electroencephalogram (EEG) and magnetoencephalogram, display frequency scaling close to  $1/f$  [3,4]. EEG analysis [5] and avalanche analysis of local field potentials (LFPs) recorded *in vitro* [6] provided clear evidence for self-organized critical states with power-law distributions. There is also evidence for critical states from the power-law scaling of interspike interval (ISI) distributions computed from retinal, visual thalamus and primary visual cortex neurons [7]. In addition, model networks of neurons indicate that critical states may be associated with frequency scaling consistent with experiments [8]. However, these are independent evidences from different preparations and the link between  $1/f$  frequency scaling of global variables and the existence of critical states in neural activity has not been firmly established. Moreover,  $1/f$  spectra are not necessarily associated with critical states [9], so it is not clear if the intact and functioning brain operates in a way similar to critical states.

To attempt answering these questions, we first investigated if  $1/f$  frequency scaling is present in global variables recorded close to the underlying neuronal current sources *in vivo*. We analyzed cortical LFPs which were recorded within cerebral cortex using bipolar extracellular high-impedance microelectrodes [10]. Bipolar LFP recordings sample relatively localized populations of neurons, as these signals can be very different for electrodes separated by 1 mm apart [10]. This stands in contrast with the EEG, which samples much larger populations of neurons [11] and is recorded from the surface of the scalp using

millimeter-scale electrodes. LFPs are subject to much less filtering compared to EEG, because EEG signals must diffuse through various media, such as cerebrospinal fluid, dura matter, cranium, muscle, and skin. Thus, finding  $1/f$  frequency scaling of bipolar LFPs would be a much stronger evidence that this scaling reflects neuronal activities, as these signals are directly recorded from within the neuronal tissue. Moreover, in order to distinguish state-dependent scaling properties, we have compared recordings during wakefulness and slow-wave sleep in the same experiments.

Bipolar LFPs from cat parietal association cortex show the classic landmarks of EEG signals in these states [12]; namely, during waking, LFPs are of low amplitude and very irregular (Fig. 1, top trace), and are dominated by beta frequencies (around 20 Hz). This pattern is also called “desynchronized” activity, and is typically seen during aroused states in the human EEG [11]. During slow-wave sleep, LFPs display high-amplitude slow-wave activity (Fig. 1, middle trace), similar to the “delta waves” of human sleep EEG [11]. The power spectral density (PSD) calculated from these LFPs typically shows a broadband structure. During wakefulness, the PSD shows two different scaling regions, according to the frequency band. For low frequencies (between 1 and 20 Hz), the PSD scales approximately as  $1/f$ , whereas for higher frequencies (between 20 and 65 Hz), the scaling is approximately of  $1/f^3$  (Fig. 1, black PSD). During slow-wave sleep, the additional power at slow frequencies masks the  $1/f$  scaling, but the same  $1/f^3$  scaling is present in the high-frequency band (Fig. 1, gray PSD). The same behavior was observed for other electrodes in the same experiment, and in three other animals (not shown). Thus, these results confirm that the  $1/f$  frequency scaling reported in the EEG [3] is also present in bipolar LFPs from cat association cortex, but only during waking and for specific frequency bands.

To investigate whether this  $1/f$  scaling is associated with self-organized critical states, we first analyzed the ISI distributions from neurons recorded in cat parietal

cortex. Unit activity was recorded simultaneously with LFPs at 8 locations separated by 1 mm [10]. The distribution of ISIs was computed for individual neurons, and was represented in log-linear scale (Fig. 2; log-log scale in insets). For both wakefulness and slow-wave sleep [Fig. 2(a) and 2(b)], the distributions showed no evidence for power-law behavior. During waking, the ISI distributions were close to exponentially distributed ISIs, as generated by Poisson stochastic process with same statistics as the neurons analyzed (Fig. 2, Poisson). For 22 neurons recorded during the wake state, the Pearson coefficient was of  $0.91 \pm 0.13$  for exponential distribution fits, and of  $0.86 \pm 0.16$  for power-law distribution fits. Taking only the subset of 7 neurons with more than 2000 spikes, the fit was nearly perfect for exponential distributions (Pearson coefficient of  $0.999 \pm 0.001$ ). However, during slow-wave sleep, there was a marked difference between the experimental ISI and the corresponding Poisson process

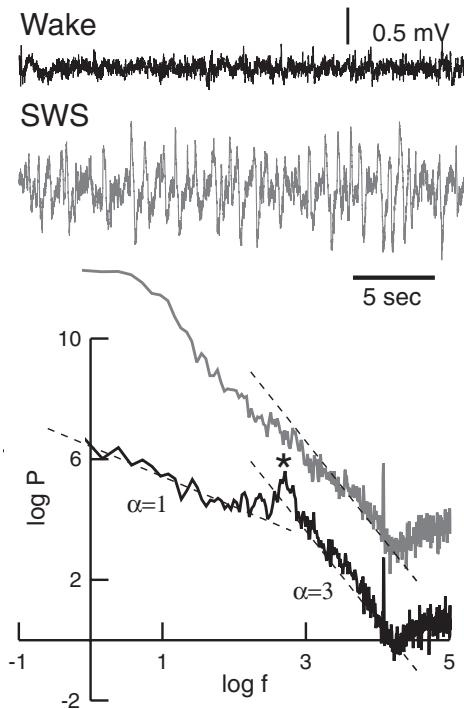


FIG. 1. Frequency scaling of local field potentials from cat parietal cortex. Top traces: LFPs recorded in cat parietal cortex during wake and slow-wave sleep (SWS) states. Bottom: Power spectral density of LFPs, calculated from 55 sec sampled at 300 Hz (150 Hz 4th-order low-pass filter), and represented in log-log scale (dashed lines represent  $1/f^\alpha$  scaling). During waking (black), the frequency band below 20 Hz scales approximately as  $1/f$  (\* shows the peak at 20 Hz beta frequency), whereas the frequency band between 20 and 65 Hz scales approximately as  $1/f^3$ . During slow-wave sleep (gray; displaced upwards), the power in the slow frequency band is increased, and the  $1/f$  scaling is no longer visible, but the  $1/f^3$  scaling at high frequencies remains unaffected. PSDs were calculated over successive epochs of 32 sec, which were averaged over a total period of 200 sec for Wake and 500 sec for SWS.

[Fig. 2(b)]. In this state, neurons tended to produce long periods of silences, which are related to EEG slow waves [10,12], and which is visible as a prominent tail of the distribution for large ISIs. This tail was well fit by a Poisson process of low rate [Fig. 2(b), dashed line].

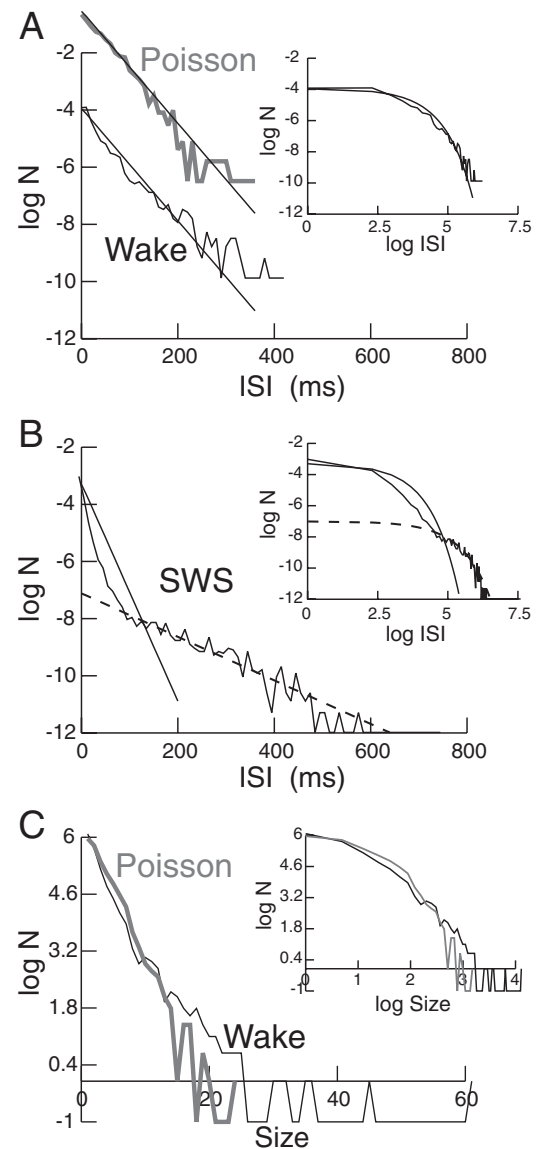


FIG. 2. Absence of power-law distributions in neuronal activity. The logarithm of the distribution of ISI during waking (Wake, (a), 1951 spikes) and slow-wave sleep (SWS, (b), 15997 spikes) is plotted as a function of ISI length, or log ISI length (insets). A Poisson process of the same rate and statistics is displayed in (a) (Poisson; gray curve displaced upwards for clarity). The exponential ISI distribution predicted by Poisson processes of equivalent rates is shown as straight lines (smooth curve in inset). The dotted line in (b) indicates a Poisson process with lower rate which fits the tail of the ISI distribution in SWS. (c) Avalanche analysis realized by taking into account the statistics from all simultaneously recorded cells in Wake. The distribution of avalanche sizes scales exponentially (black curves), similar to the same analysis performed on a Poisson process with same statistics (gray curves).

To further check for criticality, we have performed an avalanche analysis by taking into account the collective information from the multisite recordings. We used the same method as for Ref. [6], which amounts to detecting clusters of contiguous events separated by silences, by binning the system in time windows of 1 to 16 ms [6]. As there was no evidence for any recognizable event in LFPs which could be taken as avalanche (see Fig. 1), we used the spike times among the ensemble of simultaneously recorded neurons. The distribution of avalanche size does not follow power-law scaling [Fig. 2(c), black], but is closer to an exponential distribution as predicted by Poisson processes [Fig. 2(c), gray]. This analysis therefore confirms the absence of avalanche dynamics in this system [13].

To explain the  $1/f$  scaling of LFPs, we attempted to reconstruct LFPs from unit activity. Unit activity is displayed in Fig. 3 (top) for the same experiment as that of Fig. 1. Because LFPs are generated primarily by synaptic currents in neurons [11,14], and because synaptic currents are very well modeled by simple exponential relaxation processes [15], we modeled the synaptic current from the following convolution [16]:

$$C(t) = \int_{-\infty}^{\infty} D(t') \exp[-(t-t')/\tau_s] dt', \quad (1)$$

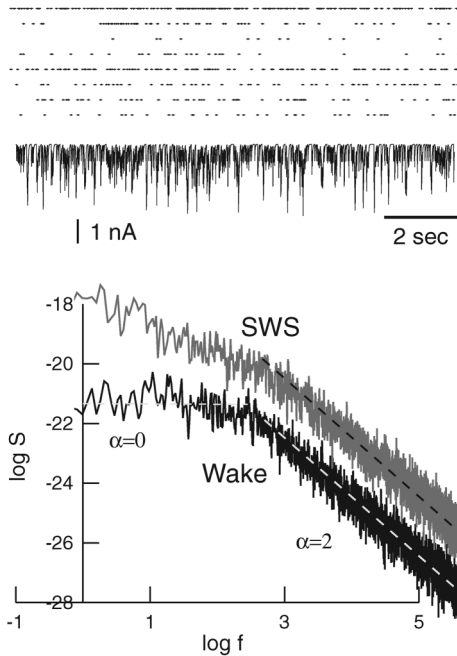


FIG. 3. Frequency scaling of synaptic currents reconstructed from spike times. Top traces: raster plot of spiking times of 8 multiunit recordings in cat cortex during wakefulness (same experiment as in Fig. 1; data from Ref. [10]). Middle trace: total synaptic current obtained by convolving the spike times with exponential relaxation processes ( $\tau_s = 10$  ms). Bottom: PSD of synaptic currents for wake (black) and slow-wave sleep (SWS; PSD in gray displaced upwards for clarity); dashed lines represent  $1/f^\alpha$  scaling.

where  $C(t)$  is the synaptic current and  $D(t)$  is the “drive” signal which consisted in the experimentally-recorded spike trains. The PSD of the synaptic current is then given by

$$S(\omega) = |C(\omega)|^2 = \frac{|D(\omega)|^2}{1 + \omega^2 \tau_s^2}. \quad (2)$$

The PSD of synaptic currents reconstructed from experimentally recorded spikes showed an approximate Lorentzian behavior ( $1/f^2$  scaling) during wakefulness (Fig. 3, Wake), as expected from the exponential nature of synaptic events. During slow-wave sleep, there was more power for slow frequencies, but the  $1/f^2$  scaling at high frequencies was still present (Fig. 3, SWS). The Lorentzian form of the PSD in Fig. 3 (Wake) shows that in the waking state,  $|D(\omega)|^2$  is approximately constant, therefore the drive  $D(t)$  is statistically equivalent to a white noise process, consistent with the apparent Poisson statistics of spikes identified in Fig. 2 (see also Refs. [17] for similar findings in awake monkeys). During slow-wave sleep, however, the deviation from the Lorentzian suggests that  $D(t)$  is a stochastic process statistically different from white noise, and contains in addition increased power at low frequencies, also consistent with the analysis of Fig. 2.

This model, however, does not yield PSD consistent with the  $1/f$  and  $1/f^3$  scaling of LFPs shown in Fig. 1. Interestingly, the scaling of this model is in  $1/f^0$  or  $1/f^2$  for the same frequency bands that displayed  $1/f$  or  $1/f^3$  in LFPs, respectively. Using a similar convolution equation to model the PSD of LFPs

$$\text{LFP}(t) = \int_{-\infty}^{\infty} C(t') F(t-t') dt', \quad (3)$$

where  $C(t)$  is the synaptic current source and  $F(t)$  is a function representing a filter. As above, the PSD is given by

$$P(\omega) = |\text{LFP}(\omega)|^2 = |C(\omega)|^2 |F(\omega)|^2. \quad (4)$$

In this model, the frequency scaling of the PSD of both wakefulness and slow-wave sleep LFPs in Fig. 1 can be explained by assuming that the filter scales as  $1/f$ , or equivalently that  $|F(\omega)|^2 \sim 1/\omega$ . In other words, this model can explain qualitatively the  $1/f$  and  $1/f^3$  scaling of LFPs under the condition that neuronal current sources are subject to an  $1/f$  filter. Such a filter is most likely due to the filtering of extracellular currents through the tissue, before it reaches the electrode [18].

Finally, we provide an intuitive justification for this predicted  $1/f$  filter, as well as possible ways to test it experimentally. The  $1/f$  filtering of extracellular media can be justified intuitively by considering the complex structure of such media, and, in particular, its spatial irregularity. Extracellular space consists of a complex arrangement of cellular processes of various size and irregular shape, while the extracellular fluid represents only a few percent of the available space [19]. The effect of a current source in such media will be a combination of resistive



effects, due to the flow of current in the conductive fluids, and capacitive effects, due to the high density of membranes (for a theoretical treatment see Refs. [18]). Such a complex arrangement of resistors and capacitors with random values is known to produce an  $1/f$  filter, as found for inhomogeneous materials [20]. Although such materials are different from the structure of biological media, it is plausible that similar considerations may explain the  $1/f$  filtering predicted here. Linear arrangements of  $RC$  circuits with random values ( $RC$  line) also generate  $1/f$  noise [21]. Superposition of a large number of exponential relaxation processes with different relaxation rates can also generate  $1/f$  scaling [22,23]. Understanding of the  $1/f$  filtering by extracellular media based on plausible biophysical models is presently under investigation. The predicted  $1/f$  filter could also be tested experimentally by injecting white noise currents (of amplitude comparable to neuronal current sources) in extracellular space, and measuring the resulting field potential at some distance from the injection site. This measured LFP should scale as  $1/f$ .

In conclusion, we have shown that the PSD of bipolar LFPs from cat parietal cortex displays several scaling regions, as  $1/f$  or  $1/f^3$  depending on the frequency band and behavioral state. By analyzing neuronal unit activity from the same experiments, we did not see evidence that this  $1/f$  scaling is associated with critical states. Neither ISI distributions nor avalanche size distributions display power-law scaling, but are rather consistent with Poisson processes. We have provided an alternative explanation for  $1/f$  frequency scaling which does not rely on critical states, but rather stems from the filtering properties of extracellular media. We have given an intuitive explanation for a possible physical origin of such  $1/f$  filtering, as well as a way to test it experimentally. These results may appear to contradict previous evidence for critical states *in vitro* [6] or in the early visual system *in vivo* [7]. However, the absence of critical states reported here may instead reflect fundamental differences between association cortex and other structures more directly related to sensory inputs. Future work should clarify why different structures show different scaling, and what implications it may have for brain dynamics and coding.

The experimental data analyzed in this article were obtained with Drs. Diego Contreras and Mircea Steriade, and were published previously [10]. We are grateful for support by NSERC Canada (H. K.), CNRS, the European Commission (FET program), and the HFSP program (A. D.).

---

\*Corresponding author.

Email address: destexhe@iaf.cnrs-gif.fr

- [1] H.J. Jensen, *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems* (Cambridge University Press, Cambridge, England, 1998).

- [2] P. Bak, *How Nature Works* (Springer-Verlag, New York, 1996).
- [3] W.S. Pritchard, *Int. J. Neurosci.* **66**, 119 (1992); W.J. Freeman, L.J. Rogers, M.D. Holmes, and D.L. Silbergeld, *J. Neurosci. Methods* **95**, 111 (2000); P.A. Robinson, C.J. Rennie, J.J. Wright, H. Bahramali, E. Gordon, and D.L. Rowe, *Phys. Rev. E* **63**, 021903 (2001).
- [4] E. Novikov, A. Novikov, D. Shannahoff-Khalsa, B. Schwartz, and J. Wright, *Phys. Rev. E* **56**, R2387 (1997).
- [5] K. Linkenkaer-Hansen, V.V. Nikouline, J.M. Palva, and R.J. Ilmoniemi, *J. Neurosci.* **21**, 1370 (2001); W.J. Freeman, M.D. Holmes, G.A. West, and S. Vanhatalo, *Clin. Neurophysiol.* **117**, 1228 (2006).
- [6] J. Beggs and D. Plenz, *J. Neurosci.* **23**, 11 167 (2003).
- [7] M.C. Teich, C. Heneghan, S.B. Lowen, T. Ozaki, and E. Kaplan, *J. Opt. Soc. Am. A* **14**, 529 (1997); A.R.R. Papa and L. da Silva, *Theory Biosci.* **116**, 321 (1997).
- [8] L. de Arcangelis, C. Perrone-Capano, and H.J. Herrmann, *Phys. Rev. Lett.* **96**, 028107 (2006).
- [9] P. De Los Rios, and Y.C. Zhang, *Phys. Rev. Lett.* **82**, 472 (1999); T. Giesinger, *Biol. Rev. Camb. Philos. Soc.* **76**, 161 (2001).
- [10] A. Destexhe, D. Contreras, and M. Steriade, *J. Neurosci.* **19**, 4595 (1999).
- [11] *Electroencephalography*, edited by E. Niedermeyer and F. Lopes da Silva (Williams & Wilkins, Baltimore, MD, 1998).
- [12] M. Steriade, *Neuronal Substrates of Sleep and Epilepsy* (Cambridge University Press, Cambridge, UK, 2003).
- [13] A complete demonstration of the absence of critical states would require investigating other dynamical features, such as long-lived metastable states; see M. Paczuski, P. Bak, and S. Maslov, *Phys. Rev. E* **53**, 414 (1996).
- [14] P.L. Nunez, *Electric Fields in the Brain: The Neurophysics of EEG* (Oxford University Press, Oxford, UK, 1981).
- [15] A. Destexhe, Z. Mainen, and T.J. Sejnowski, in *Methods in Neuronal Modeling*, edited by C. Koch and I. Segev (MIT Press, Cambridge, MA, 1998), pp. 1–26.
- [16] Nearly identical results were obtained using a biophysical model based on a passive membrane equation where experimentally recorded spikes triggered conductance-based synaptic inputs.
- [17] W.R. Softky and C. Koch, *J. Neurosci.* **13**, 334 (1993); W. Bair, C. Koch, W. Newsome, and K. Britten, *J. Neurosci.* **14**, 2870 (1994).
- [18] C. Bedard, H. Kröger, and A. Destexhe, *Biophys. J.* **86**, 1829 (2004); *Phys. Rev. E* **73**, 051911 (2006).
- [19] A. Peters, S.L. Palay, and H.F. Webster, *The Fine Structure of the Nervous System* (Oxford University Press, Oxford, 1991).
- [20] D.P. Almond and B. Vainas, *J. Phys. Condens. Matter* **13**, L361 (2001).
- [21] A.A. Verveen and L.J. DeFelice, *Prog. Biophys. Molec. Biol.* **28**, 189 (1974).
- [22] J. Bernamont, *Ann. Phys. (Paris)* **7**, 71 (1937); R.B. Anderson, *Mem. Cognition* **29**, 1061 (2001).
- [23] This mechanism is, however, not plausible, because it would require a uniform distribution of decay times of synaptic currents across a large range of values, whereas only a few receptor types are present in biological tissue, with well-defined decay times.

# Macroscopic Models of Local Field Potentials and the Apparent $1/f$ Noise in Brain Activity

Claude Bédard and Alain Destexhe\*

Integrative and Computational Neuroscience Unit (UNIC), Centre National de la Recherche Scientifique, Gif-sur-Yvette, France

**ABSTRACT** The power spectrum of local field potentials (LFPs) has been reported to scale as the inverse of the frequency, but the origin of this  $1/f$  noise is at present unclear. Macroscopic measurements in cortical tissue demonstrated that electric conductivity (as well as permittivity) is frequency-dependent, while other measurements failed to evidence any dependence on frequency. In this article, we propose a model of the genesis of LFPs that accounts for the above data and contradictions. Starting from first principles (Maxwell equations), we introduce a macroscopic formalism in which macroscopic measurements are naturally incorporated, and also examine different physical causes for the frequency dependence. We suggest that ionic diffusion primes over electric field effects, and is responsible for the frequency dependence. This explains the contradictory observations, and also reproduces the  $1/f$  power spectral structure of LFPs, as well as more complex frequency scaling. Finally, we suggest a measurement method to reveal the frequency dependence of current propagation in biological tissue, and which could be used to directly test the predictions of this formalism.

## INTRODUCTION

Macroscopic measurements of brain activity, such as the electroencephalogram (EEG), magnetoencephalogram or local field potentials (LFPs), display  $\sim 1/f$  frequency scaling in their power spectra (1–4). The origin of such  $1/f$  noise is at present unclear. The  $1/f$  spectra can result from self-organized critical phenomena (5), suggesting that neuronal activity may be working according to such states (6). Alternatively, the  $1/f$  scaling may be due to filtering properties of the currents through extracellular media (2). The latter hypothesis, however, was resting on indirect evidence, and still needs to be examined theoretically, which is one of the motivations of this article.

A continuum model (7) of LFPs incorporated the inhomogeneities of the extracellular medium into continuous spatial variations of conductivity ( $\sigma$ ) and permittivity ( $\epsilon$ ) parameters. This model reproduced a form of low-pass frequency filtering in some conditions, while considering the extracellular medium as locally neutral with  $\sigma$  and  $\epsilon$  parameters independent of frequency. This model was not entirely correct, however, because macroscopic measurements in cortex revealed a frequency dependence of electrical parameters (8). We will show here that it is possible to keep the same model structure by including plausible causes for the frequency dependence.

In a polarization model (9) of LFPs, the variations of conductivity and permittivity were considered by explicitly taking into account the presence of various cellular processes in the extracellular space around the current source. In particular, it was found that the phenomenon of surface polarization was fundamental to explain the frequency dependence

of LFPs. The continuum model (7) incorporated this effect phenomenologically through continuous variations of  $\sigma$  and  $\epsilon$ . In the polarization model, the extracellular medium is reactive in the sense that it reacts to the electric field by polarization effects. It is also locally nonneutral, which enables one to take into account the noninstantaneous character of polarization, which is at the origin of frequency dependence according to this model (9).

In this article, we propose a diffusion-polarization model that synthesizes these previous approaches and which takes into account both microscopic and macroscopic measurements. This model includes ionic diffusion, which we will show has a determinant influence on frequency filtering properties. The model also includes electric polarization, which also influences the frequency-dependent electric properties of the tissue. We show that taking into account ionic diffusion and electric polarization allows us to quantitatively account for the macroscopic measurements of electric conductivity in cortical tissue according to the experiments of Gabriel et al. (8).

However, recent measurements of Logothetis et al. (10) evidenced that the frequency dependence of cortical tissue was negligible, therefore in contradiction with the measurements of Gabriel et al. (8). We show here that the diffusion-polarization model can be consistent with both types of experiments, and thus may reconcile this contradiction. We will also examine whether this model can also explain the  $1/f$  frequency scaling observed in LFP or EEG power spectra. Finally, we consider possible ways for experimental test of the predictions of this model.

The final goal of this approach is to obtain a model of local field potentials, which is consistent with both macroscopic measurements of conductivity and permittivity, and the microscopic features of the structure of the extracellular space around the current sources.

Submitted August 28, 2008, and accepted for publication December 31, 2008.

\*Correspondence: [destexhe@unic.cnrs-gif.fr](mailto:destexhe@unic.cnrs-gif.fr)

Editor: Arthur Sherman.

© 2009 by the Biophysical Society  
0006-3495/09/04/2589/15 \$2.00

doi: 10.1016/j.bpj.2008.12.3951

## MATERIALS AND METHODS

Numerical Simulation of Macroscopic Measurements (see Numerical Simulation of Macroscopic Measurements, below) describes the impedance of the extracellular medium based on

$$Z(r_1, \omega) = \frac{1}{4\pi} \int_{r_1}^{\infty} dr' \frac{1}{r'^2} \frac{1}{\sigma^M(r', \omega) + i\omega\epsilon^M(r', \omega)}. \quad (1)$$

This equation gives the  $\omega$ -frequency component of the impedance at point  $r_1$  in extracellular space, in spherical symmetry (see (7) and Eq. 11 for details).

To evaluate this equation, we use MATLAB (The MathWorks, Natick, MA), which computes the Riemann sum,

$$Z(r_1, \omega) = \frac{1}{4\pi} \sum_{r_1}^N \frac{\Delta r'}{r'^2} \frac{1}{\sigma^M(r', \omega) + i\omega\epsilon^M(r', \omega)}, \quad (2)$$

where  $\Delta r'$  is the integration step (1  $\mu\text{m}$ ) and  $N$  is determined for a slice of 1 mm.

We also use the parametric model of Gabriel et al. (8) to simulate the frequency dependence of electrical parameters  $\sigma$  and  $\epsilon$  of the extracellular fluid from gray matter (at a temperature of 37°C). This model is valid for frequencies included in the range of 10 Hz to  $4 \times 10^8$  Hz (11). According to this model, the absolute complex and macroscopic permittivity and conductivity (measured between 10 and  $10^{10}$  Hz) in cortical gray matter is given by the Cole-Cole parametric model (12),

$$\epsilon_{\omega}^* = -i \frac{\sigma}{\omega} = \epsilon_{\infty} \epsilon_0 + \epsilon_0 \sum_{n=1}^4 \frac{\Delta \epsilon_n}{1 + (i\omega\tau_n)^{1-\alpha_n}} - i \frac{\sigma}{\omega}, \quad (3)$$

where the sum runs over four polarization modes  $n$ ,  $\epsilon_0 = 8.85 \times 10^{-12}$  F/m is vacuum permittivity,  $\epsilon_{\infty} = 4.0$  is the permittivity relative to  $f = 10^{10}$  Hz,  $\sigma = 0.02$  S/m is the static conductivity at  $f = 0$  Hz according to the chosen parametric model. The parameters under the sum of Eq. 3 are given in Table 1.

## RESULTS

We start by outlining a macroscopic model with frequency-dependent electrical parameters (Macroscopic Model of Local Field Potentials), and we discuss the main physical causes for this frequency dependence (Physical Causes for Frequency-Dependent Electrical Parameters). We then constrain the model to macroscopic measurements of electrical parameters, and provide numerical simulations to test the model and reproduce the experimental observations (Numerical Simulation of Macroscopic Measurements). Finally, we propose a possible way to test the model experimentally (Measurement of Frequency Dependence).

### Macroscopic model of local field potentials

In this section, we derive the equations governing the time evolution of the extracellular potential. We follow a formalism

**TABLE 1** Parameter values for the parametric model of Gabriel et al. (8) (see Eq. 3)

No.	$\Delta \epsilon_n$	$\tau_n$ (s)	$\alpha_n$
1	$4.50 \times 10^1$	$7.96 \times 10^{-12}$	0.10
2	$4.00 \times 10^2$	$15.92 \times 10^{-9}$	0.15
3	$2.0 \times 10^5$	$106.1 \times 10^{-6}$	0.22
4	$4.5 \times 10^7$	$5.305 \times 10^{-3}$	0.00

similar to the one developed previously (7), except that we reformulate the model macroscopically, to allow the electrical parameters (the conductivity  $\sigma$  and permittivity  $\epsilon$ ) to depend on frequency, as demonstrated by macroscopic measurements (8,11,13). The physical causes of this macroscopic frequency dependence will be examined in Physical Causes for Frequency-Dependent Electrical Parameters.

### General formalism

We begin by deriving a general equation for the electrical potential when the electrical parameters are frequency-dependent. We start from Maxwell equations, taking the first and the divergence of the fourth Maxwell equation in a medium with constant magnetic permeability, giving

$$\begin{aligned} \nabla \cdot \vec{D} &= \rho^{\text{free}} \\ \nabla \cdot \vec{j} + \frac{\partial \rho^{\text{free}}}{\partial t} &= 0, \end{aligned} \quad (4)$$

where  $\vec{D}$ ,  $\vec{j}$  and  $\rho^{\text{free}}$  are, respectively, the electric displacement, current density, and charge density in the medium surrounding the sources.

Moreover, in a linear medium the equations linking the electric field  $\vec{E}$  with electric displacement  $\vec{D}$ , and with current density  $\vec{j}$ , gives

$$\vec{D}(\vec{x}, t) = \int_{-\infty}^{\infty} \epsilon(\vec{x}, \tau) \vec{E}(\vec{x}, t - \tau) d\tau \quad (5)$$

and

$$\vec{j}(\vec{x}, t) = \int_{-\infty}^{\infty} \sigma(\vec{x}, \tau) \vec{E}(\vec{x}, t - \tau) d\tau. \quad (6)$$

The Fourier transforms of these equations are respectively  $\vec{D}_{\omega} = \epsilon_{\omega} \vec{E}_{\omega}$  and  $\vec{j}_{\omega} = \sigma_{\omega} \vec{E}_{\omega}$ , where we allow  $\sigma$  and  $\epsilon$  to depend on frequency.

Given the limited precision of measurements, we can consider  $\nabla \times \vec{E} \approx 0$  for frequencies  $< 1000$  Hz. Thus, we can assume that  $\vec{E} = -\nabla V$ , such that the complex Fourier transform of the expressions in Eq. 4 can be written as

$$\begin{aligned} \nabla \cdot (\epsilon_{\omega}(\vec{x}) \nabla V_{\omega}) &= -\rho_{\omega}^{\text{free}} \\ \nabla \cdot (\sigma_{\omega}(\vec{x}) \nabla V_{\omega}) &= i\omega \rho_{\omega}^{\text{free}}. \end{aligned}$$

Consequently, we have

$$\begin{aligned} \nabla \cdot ((\sigma_{\omega} + i\omega\epsilon_{\omega}) \nabla V_{\omega}) &= \nabla(\sigma_{\omega} + i\omega\epsilon_{\omega}) \cdot \nabla V_{\omega} \\ &+ (\sigma_{\omega} + i\omega\epsilon_{\omega}) \nabla^2 V_{\omega} = 0. \end{aligned} \quad (7)$$

Compared to previous derivations (see Eq. 49 in (7)), this equation is a more general form in which the electrical parameters can be dependent on frequency.

### Macroscopic model

In principle, it is sufficient to solve Eq. 7 in the extracellular medium to obtain the frequency dependence of LFPs. However, in practice, this equation cannot be solved because the structure of the medium is too complex to properly define



the limit conditions. The associated values of electric parameters must be specified for every point of space and for each frequency, which represents a considerable difficulty. One way to solve this problem is to consider a macroscopic or mean-field approach. This approach is justified here by the fact that the values measured experimentally are averaged values, which precision depends on the measurement technique. Because our goal is to simulate those measured values, we will use a macroscopic model, in which we take spatial averages of Eq. 7, and make a continuous approximation for the spatial variations of these average values. This type of approximation can be found in the classic theory of electromagnetism (15).

To this end, we define macroscopic electric parameters,  $\varepsilon^M$  and  $\sigma^M$ , as

$$\varepsilon_\omega^M(\vec{x}) = \langle \varepsilon_\omega(\vec{x}) \rangle_V = f(\vec{x}, \omega)$$

and

$$\sigma_\omega^M(\vec{x}) = \langle \sigma_\omega(\vec{x}) \rangle_V = g(\vec{x}, \omega),$$

where  $V$  is the volume over which the spatial average is taken. We assume that  $V$  is  $\sim \mu\text{m}^3$ , and is thus much smaller than the cortical volume, so that the mean values will be dependent of the position in cortex.

Because the average values of electric parameters are statistically independent of the mean value of the electric field, we have

$$\begin{aligned} \langle \vec{j}^{\text{total}} \rangle_V(\vec{x}, t) &= \int_{-\infty}^{\infty} \sigma^M(\tau) \langle \vec{E} \rangle_V(\vec{x}, t - \tau) d\tau \\ &+ \int_{-\infty}^{\infty} \varepsilon^M(\tau) \frac{\partial \langle \vec{E} \rangle_V}{\partial t}(\vec{x}, t - \tau) d\tau, \end{aligned}$$

where the first term on the right-hand side represents the dissipative contribution, and the second term represents the reactive contribution (reaction from the medium). Here, all physical effects, such as diffusion, resistive and capacitive phenomena, are integrated into the frequency dependence of  $\sigma^M$  and  $\varepsilon^M$ . We will examine this frequency dependence more quantitatively in Physical Causes for Frequency-Dependent Electrical Parameters.

The complex Fourier transform of  $\langle \vec{j}^{\text{total}} \rangle_V(\vec{x}, t)$  then becomes

$$\langle \vec{j}_\omega^{\text{total}} \rangle_V = (\sigma_\omega^M + i\omega\varepsilon_\omega^M) \langle \vec{E}_\omega \rangle_V = \sigma_z^M \langle \vec{E}_\omega \rangle_V, \quad (8)$$

where  $\sigma_z^M$  is the complex conductivity. We can also assume

$$\sigma_z^M = i\omega\varepsilon_z^M, \quad (9)$$

such that

$$\begin{aligned} \nabla \cdot \langle \vec{j}_\omega^{\text{total}} \rangle_V &= \nabla \cdot (\sigma_z^M \langle \vec{E}_\omega \rangle_V) \\ &= \nabla \cdot (i\omega\varepsilon_z^M \langle \vec{E}_\omega \rangle_V) = 0. \end{aligned} \quad (10)$$

Because  $\sigma_z^M = (\sigma_\omega^M + i\omega\varepsilon_\omega^M)$  and  $\vec{E}_\omega = -\nabla \langle V_\omega \rangle$ , the expressions above (Eqs. 10) can also be written in the form

$$\nabla \cdot \left( (\sigma_\omega^M + i\omega\varepsilon_\omega^M) \nabla \langle V_\omega \rangle_V \right) = 0. \quad (11)$$

We note that starting from the continuum model (7), where only spatial variations were considered, and generalizing this model by including frequency-dependent electric parameters, gives the same mathematical form as the original model (compare with Eq. 49 in (7)). This form invariance will allow us to introduce, in Physical Causes for Frequency-Dependent Electrical Parameters, the surface polarization phenomena by including an ad hoc frequency dependence in  $\sigma_\omega^M$  and  $\varepsilon_\omega^M$ . The physical causes of this macroscopic frequency dependence is that the cortical medium is microscopically nonneutral (although the cortical tissue is macroscopically neutral). Such a local nonneutrality was already postulated in a previous model of surface polarization (9). This situation cannot be accounted for by Eq. 7 if  $\sigma_\omega^M$  and  $\varepsilon_\omega^M$  are frequency-independent (in which case  $\rho_\omega = 0$  when  $\nabla V_\omega = 0$ ). Thus, including the frequency dependence of these parameters enables the model to capture a much broader range of physical phenomena.

Finally, a fundamental point is that the frequency dependences of the electrical parameters  $\sigma_\omega^M$  and  $\varepsilon_\omega^M$  cannot take arbitrary values, but are related to each other by the Kramers-Kronig relations (17–19)

$$\Delta\varepsilon^M(\omega) = \frac{2}{\pi} \int_0^\infty \frac{\sigma^M(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (12)$$

and

$$\sigma^M(\omega) = \sigma^M(0) - \frac{2\omega^2}{\pi} \int_0^\infty \frac{\Delta\varepsilon^M(\omega')}{\omega'^2 - \omega^2} d\omega', \quad (13)$$

where principal value integrals are used. These equations are valid for any linear medium (i.e., when Eqs. 5 and 6 are linear). These relations will turn out to be critical to relate the model to experiments, as we will see below.

Note that, contrary to frequency dependence, the spatial dependences of  $\sigma_\omega^M$  and  $\varepsilon_\omega^M$  are independent of each other, because these dependences are related to the spatial distribution of elements within the extracellular medium.

### Simplified geometry for macroscopic parameters

To obtain an expression for the extracellular potential, we need to solve Eq. 11, which is possible analytically only if we consider a simplified geometry of the source and surrounding medium. The first simplification is to consider the source as monopolar. The choice of a monopolar source does not intrinsically reduce the validity of the results because multipolar configurations can be composed from the arrangement of a finite number of monopoles (20). In particular, if the physical nature of the extracellular medium determines a frequency dependence for a monopolar source, it will also do so for multipolar configurations. A second

simplification will be to consider that the current source is spherical and that the potential is uniform on its surface. This simplification will enable us to calculate exact expressions for the extracellular potential and should not affect the results on frequency dependence. A third simplification is to consider the extracellular medium as isotropic. This assumption is certainly valid within a macroscopic approach, and justified by the fact that the neuropil of cerebral cortex is made of a quasirandom arrangement of cellular processes of very diverse size (21). This simplified geometry will allow us to determine how the physical nature of the extracellular medium can determine a frequency dependence of the LFPs, independently of other factors (such as more realistic geometry, propagating potentials along dendrites, etc.).

Thus, considering a spherical source embedded in an isotropic medium with frequency-dependent electrical parameters, combining with Eq. 11, we have

$$\frac{d^2\langle V_\omega \rangle|_V}{dr^2} + \frac{2}{r} \frac{d\langle V_\omega \rangle|_V}{dr} + \frac{1}{(\sigma_\omega + i\omega\varepsilon_\omega)} \frac{d(\sigma_\omega + i\omega\varepsilon_\omega)}{dr} \frac{d\langle V_\omega \rangle|_V}{dr} = 0. \quad (14)$$

Integrating this equation gives the following relation between two points  $r_1$  and  $r_2$  in the extracellular space,

$$r_1^2 \frac{d\langle V_\omega \rangle|_V}{dr}(r_1) [\sigma_\omega(r_1) + i\omega\varepsilon_\omega(r_1)] = r_2^2 \frac{d\langle V_\omega \rangle|_V}{dr}(r_2) [\sigma_\omega(r_2) + i\omega\varepsilon_\omega(r_2)]. \quad (15)$$

Assuming that the extracellular potential vanishes at large distances ( $\langle V_\omega \rangle = 0$ ), we find

$$\langle V_\omega \rangle|_V(r_1) = \frac{I_\omega}{4\pi} \int_{r_1}^{\infty} dr' \frac{1}{r'^2} \frac{1}{\sigma_\omega(r') + i\omega\varepsilon_\omega(r')}. \quad (16)$$

This equation is analogous to a similar expression derived previously (Eq. 25 in (7)), but more general. The two formalisms are related by

$$\langle \vec{j}_\omega \rangle|_V = \sigma_z \langle \vec{E}_\omega \rangle|_V = (\sigma_\omega^M + i\varepsilon_\omega^M) \langle \vec{E}_\omega \rangle|_V$$

instead of  $\vec{j}_\omega = \sigma^M \vec{E}_\omega$  (see Eq. 4 in (7)). This difference is because the conductivity here depends on frequency.

In the following, we will use the simplified notations  $\vec{j}_\omega$ ,  $\vec{E}_\omega$  and  $V_\omega$  instead of  $\langle \vec{j}_\omega \rangle|_V$ ,  $\langle \vec{E}_\omega \rangle|_V$  and  $\langle V_\omega \rangle|_V$ , respectively.

Using the relation  $V_\omega = Z_\omega I_\omega$ , the impedance  $Z_\omega$  is given by

$$Z_\omega(r) = \frac{1}{4\pi} \int_r^{\infty} dr' \frac{1}{r'^2} \frac{1}{\sigma_\omega^M(r') [1 + i\omega\tau_\omega(r')]}, \quad (17)$$

where  $\tau_\omega(r') = \frac{\varepsilon_\omega^M}{\sigma_\omega^M}$  and  $r$  is the distance between the center of the source and the position defined by  $\vec{r}$ .

## Physical causes for frequency-dependent electrical parameters

In the following, we successively consider two different cases of extracellular medium: first, nonreactive media, in which the current passively flows into the medium; and second, reactive media, in which some properties (such as charge distribution) may change after current flow. For each medium, we will consider two types of physical phenomena—the current produced by the electric field, and the current produced by ionic diffusion, as schematized in Fig. 1).

### Nonreactive media with electric fields (Model N)

Nonreactive media ( $\frac{\omega\varepsilon_\omega^M}{\sigma_\omega^M} \ll 1$ ,  $\sigma_\omega^M = \sigma^M$  and  $\varepsilon_\omega^M = \varepsilon^M$ ) are equivalent to resistive media, in which the resistance (or equivalently, the conductivity) does not change after the flow of current. The simplest type of such configuration consists of a resistive medium (such as a homogeneous conductive fluid) in which current sources solely interact via their electric field. Applying Eq. 17 to this configuration is equivalent to model the extracellular potential by Coulomb's law,

$$V_\omega(\vec{r}) = \frac{1}{4\pi\sigma^M} \cdot \frac{I_\omega}{r}, \quad (18)$$

where  $V_\omega(\vec{r})$  is the extracellular potential at a position defined by  $\vec{r}$  in extracellular space, and  $r$  is the absolute distance between  $\vec{r}$  and the center of the current source. Here, the conductivity ( $\sigma_\omega^M(r) = \sigma^M$ ) is independent of space and frequency, and thus, this model is not compatible with macroscopic measurements of frequency dependence (8,11,13).

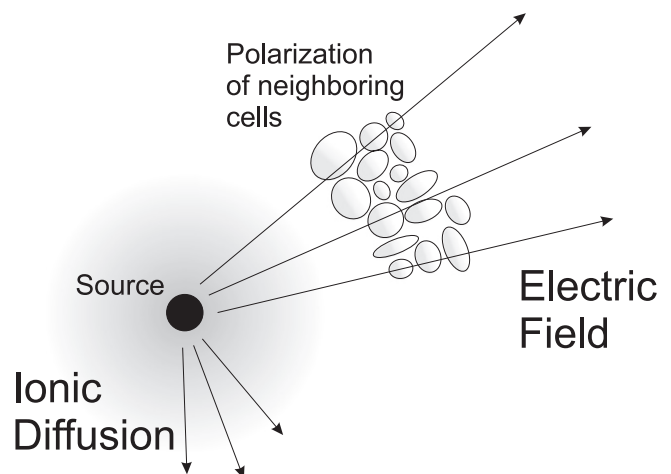


FIGURE 1 Illustration of the two main physical phenomena involved in the genesis of local field potentials. A given current source produces an electric field, which will tend to polarize the charged membranes around the source, as schematized on the top. The flow of ions across the membrane of the source will also involve ionic diffusion to reequilibrate the concentrations. This diffusion of ions will also be responsible for inducing currents in extracellular space. These two phenomena influence the frequency filtering and the genesis of LFP signals, as explored in this article.

It is, however, the most frequently used model to calculate extracellular field potentials (22). This model will be referred to as Model N in the following.

### Nonreactive media with ionic diffusion (Model D)

Because current sources are ionic currents, there is flow of ions inside or outside of the membrane, and another physical phenomena underlying current flow is ionic diffusion. Let us consider a resistive medium such as a homogeneous extracellular conductive fluid with electric parameters

$$\sigma_z^m = \sigma_\omega^m(r) \left(1 + i \frac{\omega \varepsilon_\omega^m(r)}{\sigma_\omega^m(r)}\right) = \sigma^m \left(1 + i \frac{\omega \varepsilon^m}{\sigma^m}\right) \approx \sigma^m,$$

in which the ionic diffusion coefficient is  $D$ . When the extracellular current is exclusively due to ionic diffusion, the current density depends on frequency as  $\sqrt{\omega}$  (see Appendices). A resistive medium behaves as if it had a resistivity equal to  $a(1 + b/\sqrt{\omega})$ , where  $b$  is complex. The parameter  $a$  is the resistivity for very high frequencies, and reflects the fact that the effect of ionic diffusion becomes negligible compared to calorific dissipation (Ohm's law) for very high frequencies. When ionic diffusion is dominant compared to electric field effects, the real part of  $b$  is much larger than  $a$ .

The frequency dependence of conductivity will be given by

$$\sigma_\omega^M = \frac{\sigma^m \sqrt{\omega}}{\sqrt{\omega} + k}. \quad (19)$$

Applying Eq. 17 to this configuration gives the following expression for the electric potential as a function of distance:

$$V_\omega(\vec{r}) = \frac{1}{4\pi\sigma_z^M} \cdot \frac{I_\omega}{r} = \frac{\sqrt{\omega} + k}{\sqrt{\omega}} \cdot \frac{1}{4\pi\sigma^m} \cdot \frac{I_\omega}{r}. \quad (20)$$

This expression shows that, in a nonreactive medium, when the extracellular current is dominated by ionic diffusion (compared to that directly produced by the electric field), then the conductivity will be frequency-dependent and will scale as  $\sqrt{\omega}$ . This model will be referred to as Model D in the following. Note that, if the electric field primes over ionic diffusion, then we have the situation described by Model N above.

### Reactive media with electric fields (Model P)

In reality, extracellular media contain different charge densities, for example due to the fact that cells have a nonzero membrane potential by maintaining differences of ionic concentrations between the inside and outside of the cell. Such charge densities will necessarily be influenced by the electric field or by ionic diffusion. As above, we first consider the case with only electric field effects and will consider next the influence of diffusion and the two phenomena taken together.

Electric polarization is a prominent type of reaction of the extracellular medium to the electric field. In particular, the ionic charges accumulated over the surface of cells will migrate and polarize the cell under the action of the electric field. It was shown previously in a theoretical study that this surface polarization phenomena can have important effects on the propagation of local field potentials (9). If a charged membrane is placed inside an electric field  $\vec{E}_0^S$ , there is production of a secondary electric field  $\vec{E}_\omega^S$  given by (see Eq. 31 in (9))

$$\vec{E}_\omega^S = \frac{\vec{E}_0^S}{1 + i\omega\tau_M}. \quad (21)$$

This expression is the frequency-domain representation of the effect of the inertia of charge movement associated with surface polarization, reflecting the fact that the polarization does not occur instantaneously but requires a certain time to set up. This frequency dependence of the secondary electric field was derived in Bédard et al. (9) for a situation where the current was exclusively produced by electric field. The parameter  $\tau_M$  is the characteristic time for charge movement (Maxwell-Wagner time) and equals  $\varepsilon^{\text{memb}}/\sigma^{\text{memb}}$ , where  $\varepsilon^{\text{memb}}$  and  $\sigma^{\text{memb}}$  are, respectively, the absolute (tangential) permittivity and conductivity of the membrane surface, respectively, and are in general very different from the permittivity and the conductivity of the extracellular fluid.

Let us now determine for zero-frequency the amplitude of the secondary field  $\vec{E}_0^S$  produced between two cells embedded in a given electric field. First, we assume that it is always possible to trace a continuous path which links two arbitrary points in the extracellular fluid (see Fig. 2 B). Consequently, the domain defined by extracellular fluid is said to be linearly connex. In this case, the electric potential arising from a current source is necessarily continuous in the extracellular fluid. Second, in a first approximation, we can consider that the cellular processes surrounding sources are

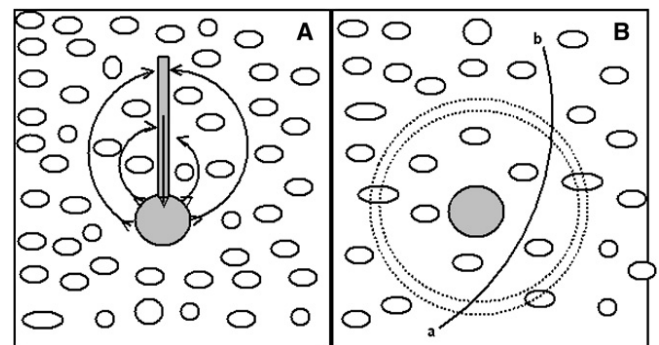


FIGURE 2 Monopole and dipole arrangements of current sources. (A) Scheme of the extracellular medium containing a quasidipole (shaded) representing a pyramidal neuron, with soma and apical dendrite arranged vertically. (B) Illustration of one of the monopoles of the dipole. The extracellular space is represented by cellular processes of various size (circles) embedded in a conductive fluid. The dashed lines represent equipotential surfaces. The  $\widehat{ab}$  line illustrates the fact that the extracellular fluid is linearly connex.

arranged randomly (by opposition to being regularly structured) and their distribution is therefore approximately isotropic. Consequently, the field produced by a given source in such a medium will also be approximately isotropic. Also consequent to this quasirandom arrangement, the equipotential surfaces around a spherical source will necessarily cut the cellular processes around the source (Fig. 2 B).

Suppose that at time  $t = 0$ , an excess of charge appears at a given point in extracellular space, then a static electric field is immediately produced. At this time, currents begins to flow in extracellular fluid, as well as inside the different cellular processes surrounding the source. These cells begin to polarize, with  $\tau_{MW}$  as the characteristic polarization time. Asymptotically, the system will reach an equilibrium where the polarization will neutralize the electric field, such that there is no electric field inside the cells (zero current). Now suppose that, in the asymptotic regime, there would still be a current flowing in between cells (in the extracellular fluid); then we have two possibilities. First, the equipotential surfaces are discontinuous, or they cut the membrane surfaces (as illustrated in Fig. 2 B). The first possibility is impossible because it would imply an infinite electric field. The second possibility is also impossible, because cells are isopotential due to polarization. Therefore, we can say that, asymptotically, there is no current flowing in extracellular fluid at  $f = 0$ , and necessarily this is equivalent to a dielectric medium. In other words, a passive inhomogeneous medium with randomly distributed passive cells is a perfect dielectric at zero frequency. In this case, the conductivity must tend to 0 when frequency tends to 0. Thus, in the following, we assume that  $\vec{E}_\omega^S = -\vec{E}_0^P$  where  $\vec{E}_0^P$  is the field produced by the source.

It follows that the expression for the current density in extracellular space as a function of the electric field is given by

$$\begin{aligned} \vec{j}_\omega &= \sigma_z^M \vec{E}_\omega^P = \sigma^m \cdot \left(1 + i \frac{\omega \varepsilon^m}{\sigma^m}\right) \cdot (\vec{E}^P + \vec{E}^S) \\ &= \sigma^m \cdot \left(1 + i \frac{\omega \varepsilon^m}{\sigma^m}\right) \cdot \frac{i\omega\tau_M}{1 + i\omega\tau_M} \cdot \vec{E}_\omega^P. \end{aligned} \quad (22)$$

In addition, for cerebral cortical tissue, we have  $1 + \frac{\omega \varepsilon^m}{\sigma^m} \approx 1$  for frequencies  $>10$  Hz and  $<1000$  Hz (see (8)). Thus, an excellent approximation of the conductivity can be written as

$$\sigma_z^M = \sigma^m \cdot \frac{i\omega\tau_M}{1 + i\omega\tau_M}. \quad (22)$$

Applying Eq. 17 gives

$$V_\omega(\vec{r}) = \frac{1}{4\pi\sigma_z^M} \cdot \frac{I_\omega}{r} = \frac{i\omega\tau_M}{1 + i\omega\tau_M} \cdot \frac{1}{4\pi\sigma^m} \cdot \frac{I_\omega}{r}. \quad (23)$$

This model describes the effect of polarization in reaction to the source electric field, and will be referred to as Model P in the following.

## Reactive media with electric field and ionic diffusion (Model DP)

The propagation of current in the medium is dominated by ionic diffusion currents or by currents produced by the electric field, according to the values of  $k$  and  $k_1$  with respect to  $\sqrt{\omega}$ . The values of  $k$  and  $k_1$  are, respectively, inversely proportional to the square root of the global ionic diffusion coefficient in the extracellular fluid, and of membrane surface (see Appendices ).

We apply the reasoning based on the connex topology of the cortical medium (see above) to deduce the order of magnitude of the induced field for zero frequency  $\vec{E}_0^S$ ,

$$\vec{E}_\omega^S = \frac{\vec{E}_0^P}{1 + i\sqrt{\omega}\tau}, \quad (24)$$

where

$$\tau = (\sqrt{\omega} + k_1)\tau_M = \sqrt{\omega} \frac{\varepsilon^{\text{memb}}}{\sigma^{\text{memb}}},$$

because the tangential conductivity on membrane surface is given by

$$\sigma_\omega^{\text{memb}} = \frac{\sigma^{\text{Memb}} \sqrt{\omega}}{\sqrt{\omega} + k_1},$$

when the current is dominated by either electric field or ionic diffusion (see Eq. 19).

It follows that the expression for the current density in extracellular space as a function of the electric field is given by

$$\begin{aligned} \vec{j}_\omega &= \sigma_z^M \vec{E}_\omega^P = \frac{\sigma^m \sqrt{\omega}}{\sqrt{\omega} + k} \cdot \left(1 + i \frac{\omega \varepsilon^m}{\sigma^m}\right) \cdot (\vec{E}^P + \vec{E}^S) \\ &\approx \frac{\sigma^m \sqrt{\omega}}{\sqrt{\omega} + k} \cdot \frac{i\sqrt{\omega}\tau}{1 + i\sqrt{\omega}\tau} \cdot \vec{E}_\omega^P, \end{aligned}$$

because  $1 + i \frac{\omega \varepsilon^m}{\sigma^m} \approx 1$  in cortical tissue for frequencies  $>10$  Hz and  $<1000$  Hz (see (8)).

Thus, we have the expression for the complex conductivity of the extracellular medium,

$$\sigma_z^M = \sigma_\omega^M + i\omega\varepsilon_\omega^M = \frac{\sigma^m \sqrt{\omega}}{\sqrt{\omega} + k} \cdot \frac{i\sqrt{\omega}\tau}{1 + i\sqrt{\omega}\tau}, \quad (25)$$

where  $\tau = (\sqrt{\omega} + k_1)\tau_M$ .

Thus, we have obtained a unique expression (Eq. 25) for the apparent conductivity in extracellular space outside of the source, and its frequency dependence due to differential Ohm's law, electric polarization phenomena, and ionic diffusion. These phenomena are responsible for an apparent frequency-dependence of the electric parameters, which will be compared to the frequency dependence observed in macroscopic measurements of conductivity (see [Numerical Simulation of Macroscopic Measurements](#), below).



Finally, Eqs. 17 and 25 imply that the macroscopic impedance of a homogeneous spherical shell of width  $R_2 - R_1$  is given by

$$Z_\omega = \frac{1}{4\pi} \int_{R_1}^{R_2} \frac{1}{r'^2} \frac{dr'}{\sigma_\omega^M + i\omega\varepsilon_\omega^M} = \frac{R_2 - R_1}{4\pi R_1 R_2} \cdot \frac{1}{\sigma_\omega^M + i\omega\varepsilon_\omega^M}. \quad (26)$$

In the following, this model will be referred as the diffusion-polarization or DP model, and we will use the above expressions (Eqs. 25 and 26) to simulate experimental measurements.

### Numerical simulation of macroscopic measurements

*Experiments of Gabriel et al. (8,11,13)*

We first consider the experiments of Gabriel et al. (8,11,13), who measured the frequency dependence of electrical parameters for a large number of biological tissues. In these experiments, the biological tissue was placed in between two capacitor plates, which were used to measure the capacitance and leak current using the relation  $I_\omega = YV_\omega$ , imposing the same current amplitude at all frequencies. Because the admittance value is proportional to  $\sigma_\omega^M + i\omega\varepsilon_\omega^M$ , measuring the admittance provides direct information about  $\sigma_\omega^M$  and  $\varepsilon_\omega^M$ .

To stay coherent with the formalism developed above, we will assume that the capacitor has a spherical geometry. The exact geometry of the capacitor should, in principle, have no influence on the frequency dependence of the admittance, because the geometry will only affect the proportionality constant between  $\sigma_z$  and  $Y_\omega$ . In the case of a spherical capacitor, by applying Eq. 26, we obtain

$$Y_\omega = \frac{1}{R} + i\omega C = 4\pi \frac{R_1 R_2}{R_2 - R_1} [\sigma_\omega^M + i\omega\varepsilon_\omega^M] = 4\pi \frac{R_1 R_2}{R_2 - R_1} \sigma_z^M. \quad (27)$$

We also take into account the fact that the resistive part is always greater than the reactive part for low frequencies (<1000 Hz), which is expressed by

$$\omega\varepsilon_\omega^M / \sigma_\omega^M \ll 1.$$

This relation can be verified, for example, from the measurement of Gabriel et al. (8), where it is true for the whole frequency band investigated experimentally (between 10 and  $10^{10}$  Hz).

The real part of  $\sigma_\omega^M = \sigma_z$  then takes the form

$$\sigma_\omega^M \approx \frac{\sigma^M \sqrt{\omega}}{\sqrt{\omega} + k} \cdot \frac{\omega\tau^2}{\omega\tau^2 + 1}, \quad (28)$$

where  $\tau = (\sqrt{\omega} + k_1)\tau_M$ .

By substituting this value of  $\tau$ , the inverse of the conductivity (the resistivity) is given by

$$\begin{aligned} \frac{1}{\sigma_\omega^M} &\approx \frac{1}{\sigma^M} \cdot \left(1 + \frac{k}{\sqrt{\omega}}\right) \cdot \left(1 + \frac{1}{\omega\tau_M^2(\sqrt{\omega} + k_1)^2}\right) \\ &= \frac{1}{\sigma^M} \cdot \left[1 + \frac{k}{\sqrt{\omega}} + \left(\frac{1}{\omega\tau_M^2} + \frac{k}{\omega^{3/2}\tau_M^2}\right) \right. \\ &\quad \left. \times \left(\frac{1}{\omega + 2k_1\sqrt{\omega} + k_1^2}\right)\right]. \end{aligned} \quad (29)$$

Finally, to reproduce the experiments of Gabriel et al., we assume  $k_1 \gg \sqrt{\omega}$ . By developing in series the last term (in parentheses) of Eq. 29, we obtain

$$\begin{aligned} \frac{1}{\sigma_\omega^M} &\approx \bar{K}_0 + \frac{\bar{K}_1}{\omega^{1/2}} + \frac{\bar{K}_2}{\omega} + \frac{\bar{K}_3}{\omega^{3/2}} \\ &= K_0 + \frac{K_1}{f^{1/2}} + \frac{K_2}{f} + \frac{K_3}{f^{3/2}}. \end{aligned} \quad (30)$$

Equation 30 corresponds to the conductivity  $\sigma^M$ , as measured in the experimental conditions of the experiments of Gabriel et al. (the permittivity  $\varepsilon^M$  is obtained by applying Kramers-Kronig relations). Fig. 3 shows that this expression for the conductivity can explain the measurements in the frequency range of 10–1000 Hz, which are relevant for LFPs. To obtain this agreement, we had to assume in Eq. 25 a relatively low Maxwell-Wagner time of  $\sim 0.15$  s ( $f_c = 1/(2\pi\tau_M)$  between 1 Hz and 10 Hz),  $k_1 > \sqrt{\omega} > k$  (for frequencies <100 Hz).

This value of Maxwell-Wagner time is necessary to explain Gabriel's experiments, and may seem very large at first sight. However, the Maxwell-Wagner time is not limited by physical constraints, because we have by definition  $\tau_{MW} = \frac{\varepsilon_\omega}{\sigma_\omega}$ . In principle, the value of  $\sigma_\omega$  can be very small, approaching zero, while  $\varepsilon_\omega$  can take very large values. For example, taking the measurements of Gabriel et al. in aqueous solutions of NaCl and in gray matter (8), gives values of  $\tau_{MW}$  comprised between 1 ms and 100 ms for frequencies at  $\sim 10$  Hz.

Thus, the model predicts that in the experiments of Gabriel et al., the transformation of electric current carried by electrons to ionic current in the biological medium necessarily implies an accumulation of ions at the plates of the capacitor. This ion accumulation will in general depend on frequency, because the conductivity and permittivity of the biological medium are frequency-dependent. This will create a concentration gradient across the biological medium, which will cause a ionic diffusion current opposite to the electric current. This ionic current will allow a greater resulting current because surface polarization is opposite to the electric field. Fig. 3 shows that such conditions give frequency-dependent macroscopic parameters consistent with the measurements of Gabriel et al.

The parameter choices to obtain this agreement can be justified qualitatively because the ionic diffusion constant on cellular surfaces is probably much smaller than in the extracellular fluid, such that  $k_1 \gg k$ . This implies the existence of

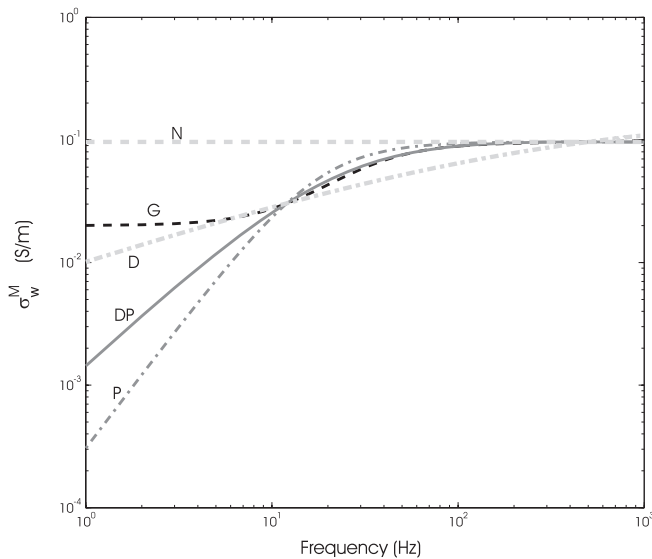


FIGURE 3 Models of macroscopic extracellular conductivity compared to experimental measurements in cerebral cortex. The experimental data (labeled *G*) show the real part of the conductivity measured in cortical tissue by the experiments of Gabriel et al. (8). The curve labeled *E* represents the macroscopic conductivity calculated according to the effects of electric field in a nonreactive medium. The curve labeled *D* is the macroscopic conductivity due to ionic diffusion in a nonreactive medium. The curve labeled *P* shows the macroscopic conductivity calculated from a reactive medium with electric-field effects (polarization phenomena). The curve labeled *DP* shows the macroscopic conductivity in the full model, combining the effects of electric polarization and ionic diffusion. Every model was fit to the experimental data by using a least-square procedure, and the best fit is shown. The *DP* model's conductivity is given by Eq. 30 with  $K_0 = 10.84$ ,  $K_1 = -19.29$ ,  $K_2 = 180.35$ , and  $K_3 = 52.56$ . The experimental data (*G*) is the parametric Cole-Cole model (12), which was fit to the experimental measurements of Gabriel et al. (8). This fit is in agreement with experimental measurements for frequencies  $>10$  Hz. No experimental measurements exist for frequencies  $<10$  Hz, and the different curves show different predictions from the phenomenological model of Cole-Cole and these models.

a frequency band  $B_f$  for which  $\sqrt{\omega}$  is negligible with respect to  $k_1$ , but not with respect to  $k$  because these constants are inversely proportional to the square root of their respective diffusion coefficients. Thus, the approximation that we suggest here is that this band  $B_f$  finishes at  $\sim 100$  Hz in the experimental conditions of Gabriel et al. It is important to note that this parameter choice is entirely dependent on the ratio between ionic diffusion current and the current produced by the electric field, and thus will depend on the particular experimental conditions.

It is interesting to note that this model and the phenomenological Cole-Cole model (12) predict different behaviors of the conductivity for low frequencies ( $<10$  Hz). In this model, the conductivity tends to zero when frequency tends to zero, while in the Cole-Cole extrapolation, it tends to a constant value (13). The main difference between these models is that the Cole-Cole model is phenomenological and has never been deduced from physical principles for

low frequencies, unlike this model, which is entirely deduced from well-defined physical phenomena.

#### Logothetis et al. (10) measurements

We next consider the experiments of Logothetis et al. (10), which reported a resistive medium, in contrast with the experiments of Gabriel et al. In these experiments, four electrodes were aligned and spaced by 3 mm in monkey cortex. The first and last electrodes were used to inject current, while the two intermediate electrodes were used to measure the extracellular voltage. The voltage was measured at different frequencies and current intensities.

One important point in this experimental setup is that the intensity of the current was such that the voltage at the extreme (injecting) electrodes saturates. One of the consequences of this saturation was to limit ionic diffusion effects, as discussed in Logothetis et al. (10). This voltage saturation will diminish the concentration difference near the source (we would have an amplification if this was not the case). It follows that, in the experiments of Logothetis et al., the ratio between diffusion current and electric field current is very small. Thus, in this case, we use values of parameters  $k_i$  very different from those assumed above to reproduce the experiments of Gabriel et al., in particular  $k_1 \ll \sqrt{\omega}$ . As we will see in Discussion, this situation may be different from the physiological conditions.

Nevertheless, the large distance between electrodes suggests that the relation between current and voltage is linear because the current density is roughly proportional to the inverse of squared distance to the source. Consequently, we can suppose that in the experiments of Logothetis et al., the ionic gradient is negligible, which prevents ionic diffusion currents. Thus, in this experiment, most—if not all—of the extracellular current is due to electric-field effects.

In this situation, the conductivity (Eq. 25) becomes

$$\sigma_\omega^M \approx \sigma^m \cdot \frac{(\omega\tau_M)^2}{1 + (\omega\tau_M)^2}, \quad (31)$$

which is similar to the Model P above.

Moreover, taking the same Maxwell-Wagner time  $\tau_M$  as above for the experiments of Gabriel et al. (which corresponds to a cutoff frequency of 1 Hz), we have for frequencies  $>10$  Hz,

$$\frac{(\omega\tau_M)^2}{1 + (\omega\tau_M)^2} \approx 1,$$

similar to a resistive medium.

Thus, in the experiments of Logothetis et al., the saturation phenomenon entrains current propagation in the biological medium as if the medium was quasiresistive for frequencies  $>10$  Hz. This constitutes a possible explanation of the contrasting results in the measurements of Logothetis et al. and Gabriel et al.

*Frequency dependence of the power spectral density of extracellular potentials*

The third type of experimental observation is the fact that the power spectral density (PSD) of LFPs or EEG signals displays  $1/f$  frequency scaling (1–4). To examine whether this  $1/f$  scaling can be accounted for by this formalism, we consider a spherical current source embedded in a continuous macroscopic medium. We also assume that the PSD of the current source is a Lorentzian, which could derive, for example, from randomly occurring exponentially decaying postsynaptic currents (2) (see Fig. 4).

To simulate this situation, we used the diffusion-polarization model with ionic diffusion and electric field effects in a reactive medium. We have estimated above that surface polarization phenomena have a cutoff frequency of  $\sim 1$  Hz, and will not play a role above that frequency. So, if we focus on the PSD of extracellular potentials in the frequency range  $> 1$  Hz, we can consider only the effect of ionic diffusion (in agreement with the experiments of Gabriel et al.; see above).

Thus, we can approximate the conductivity as (see Eq. 25)

$$\sigma_{\omega}^M = a\sqrt{\omega}, \quad (32)$$

where  $a$  is a constant.

It follows that the extracellular voltage around a spherical current source is given by (see Eq. 17)

$$V(r, \omega) = \frac{I_{\omega}}{4\pi a\sqrt{\omega} r} = \frac{V(r, 1)}{\sqrt{\omega}} = \frac{V(R, 1)R}{r\sqrt{\omega}}, \quad (33)$$

where  $R$  is the radius of the source.

In other words, we can say that the extracellular potential is given by the current source  $I_{\omega}$  convolved with a filter in  $1/\sqrt{\omega}$ , which is essentially due to ionic diffusion (Warburg impedance; see the literature (23–25)). A white noise current source will thus result in a PSD scaling as  $1/f$ , and can explain the experimental observations, as shown in Fig. 4. Experimentally recorded LFPs in cat parietal cortex display LFPs with frequency scaling as  $1/f$  for low frequencies, and  $1/f^3$  for high frequencies (Fig. 4, A and B). Following the same procedure as in Bédard et al. (2), we reconstructed the synaptic current source from experimentally recorded spike trains (Fig. 4, C and D). The PSD of the current source scales as a Lorentzian (Fig. 4 E) as expected from the exponential nature of synaptic currents. Calculating the LFP around the source and taking into account ionic diffusion, gives a PSD with two frequency bands, scaling in  $1/f$  for low frequencies, and  $1/f^3$  for high frequencies (Fig. 4 F). This is the frequency scaling observed experimentally for LFPs in awake cat cortex (2). We conclude that ionic diffusion is a plausible physical cause of the  $1/f$  structure of LFPs for low frequencies.

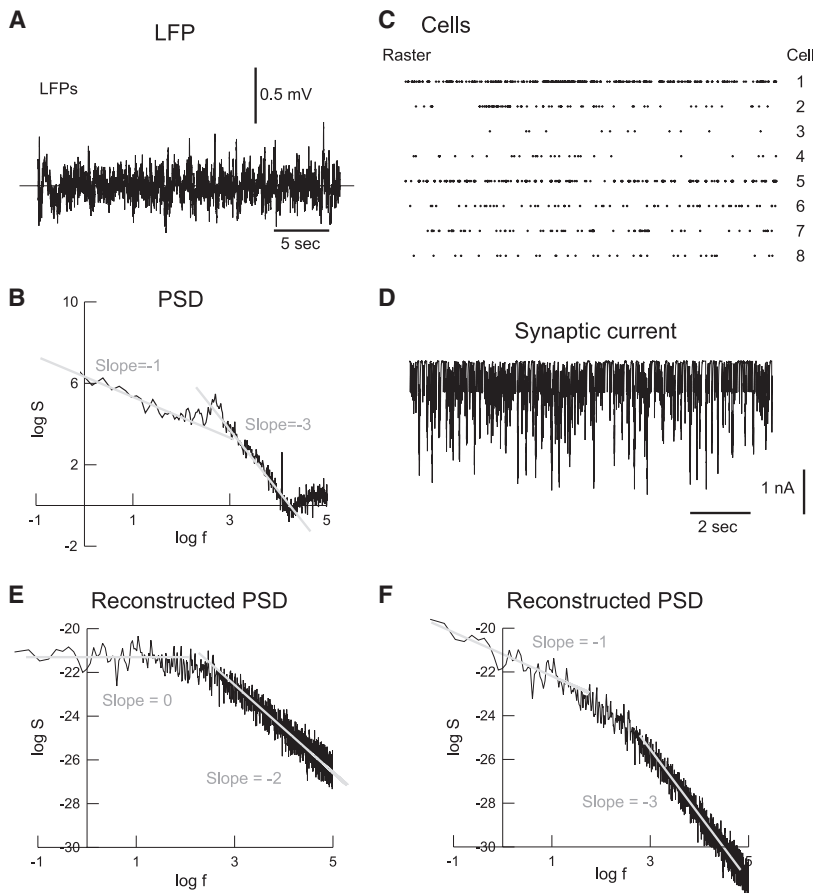


FIGURE 4 Simulation of  $1/f$  frequency scaling of LFPs during wakefulness. (A) LFP recording in the parietal cortex of an awake cat. (B) Power spectral density (PSD) of the LFP in log scale, showing two different scaling regions with a slope of  $-1$  and  $-3$ , respectively. (C) Raster of eight simultaneously-recorded neurons in the same experiment as in panel A. (D) Synaptic current calculated by convolving the spike trains in panel C with exponentials (decay time constant of 10 ms). (E) PSD calculated from the synaptic current, shown two scaling regions of slope 0 and  $-2$ , respectively. (F) PSD calculated using a model including ionic diffusion (see text for details). The scaling regions are of slope  $-1$  and  $-3$ , respectively, as in the experiments in panel B. Experimental data taken from Destexhe et al. (37); see also Bédard et al. (2) for details of the analysis in panels B–D.

Two important points must be noted. First, the diffusion-polarization model does not automatically predict  $1/f$  scaling at low frequencies, but rather implements an  $1/f$  filter, which may result in frequency scaling with larger slopes. Second, the same experimental situation may result in different frequency scaling, and this is also consistent with the diffusion-polarization model. These two points are illustrated in Fig. 5, which shows a similar analysis as Fig. 4 but during slow-wave sleep in the same experiment. The LFP is dominated by slow-wave activity (Fig. 5 A), and the different units display firing patterns characterized by concerted pauses (shaded lines in Fig. 5 B), characteristic of slow-wave sleep and which are also visible in the reconstructed synaptic current (Fig. 5 C). The PSD shows a similar scaling as  $1/f^3$

as for wakefulness, but the scaling at low frequencies is different (slope at  $\sim -2$  at low frequencies; see Fig. 5 D). The PSD reconstructed using the diffusion-polarization model displays similar features (compare with Fig. 5 E). This analysis shows that the diffusion-polarization model qualitatively accounts for different regions of frequency scaling found experimentally in different frequency bands and network states.

### Measurement of frequency dependence

In this final section, we examine a possible way to test the model experimentally. The main prediction of the model is that, in natural conditions, the extracellular current

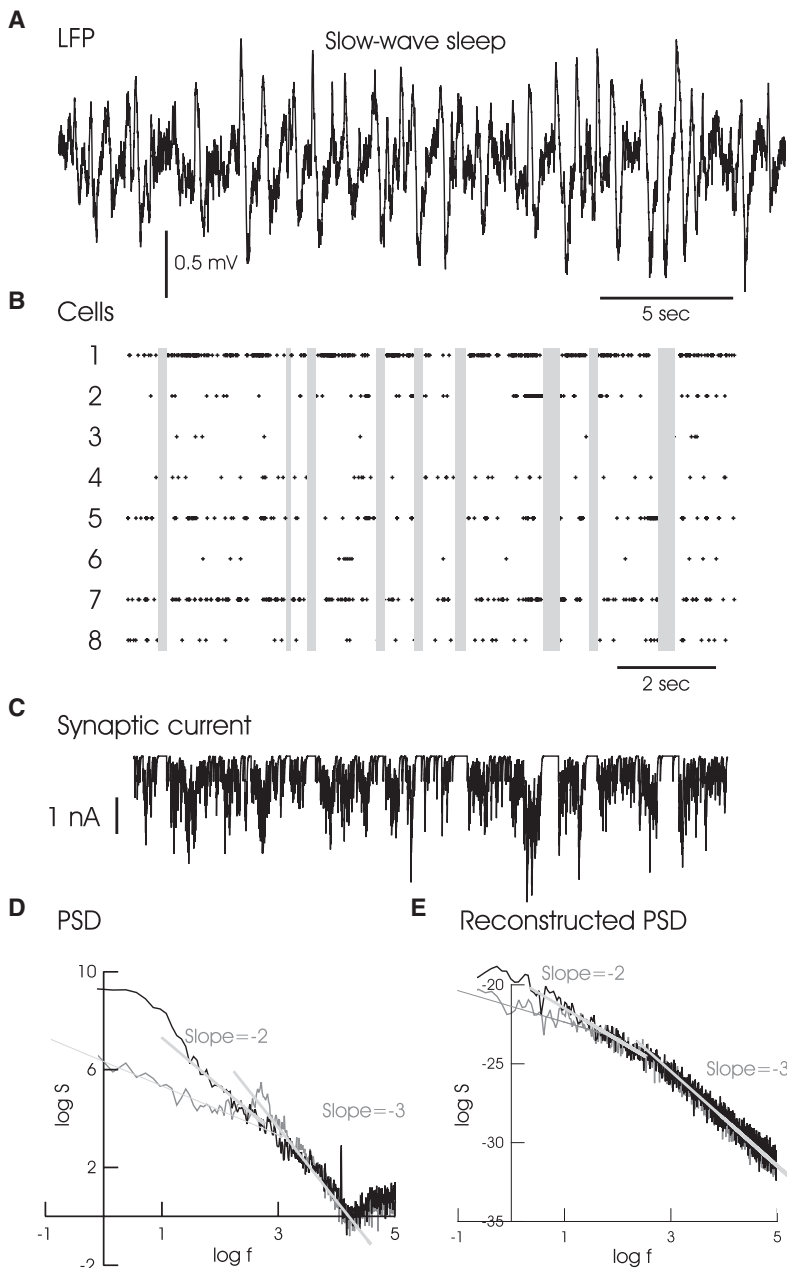


FIGURE 5 Simulation of more complex frequency scaling of LFPs during slow-wave sleep. (A) Similar LFP recording as in Fig. 4 A (same experiment), but during slow-wave sleep. (B) Raster of eight simultaneously-recorded neurons in the same experiment as in panel A. The vertical shaded lines indicate concerted pauses of firing which presumably occur during the down states. (C) Synaptic current calculated by convolving the spike trains in panel B with exponentials (decay time constant of 10 ms). (D) Power spectral density (PSD) of the LFP in log scale, showing the same scaling regions with a slope of  $-3$  at high frequencies as in wakefulness (the PSD in wake is shown in shading in the background). At low frequencies, the scaling was close to  $1/f^2$  (shaded line; the dotted line shows the  $1/f$  scaling of wakefulness). (E) PSD calculated from the synaptic current in panel C, using a model including ionic diffusion. This PSD reproduces the scaling regions of slope  $-2$  and  $-3$ , respectively (shaded lines). The low-frequency region, which was scaling as  $1/f$  in wakefulness (dotted lines), had a slope close to  $-2$ . Experimental data taken from Destexhe et al. (37).



perpendicular to the source is dominated by ionic diffusion. The experiments realized so far (8,10,11,13) used macroscopic currents that did not necessarily respect the correct current flow in the tissue.

We suggest creating more naturalistic current sources by generating ionic currents with a micropipette placed in the extracellular medium. By using periodic current injection during very short time  $\Delta t$  compared to the period (small duty cycle), we can measure, using the same electrode, the extracellular voltage  $V_{pr}$  (using a fixed reference far away from the source). If the period of the source is shorter than the relaxation time of the system, the voltage will integrate, which is due to charge accumulation.

Because  $V_{pr}$  is directly proportional to the amount of charge emitted as a function of time during  $\Delta t$  (capacitive effect of the extracellular medium), the time variation of  $V_{pr}$  is directly proportional to the ionic diffusion current. In such conditions, if the extracellular medium is purely resistive as predicted by the experiments of Logothetis et al., the relaxation time should be very small, of  $\sim 10^{-12}$  s (9), which would prevent any integration phenomena and charge accumulation for frequencies  $< 10^{12}$  Hz. If the medium has a slower relaxation due to polarization and ionic diffusion, then we should observe voltage integration and charge accumulation for physiological frequencies ( $< 1000$  Hz).

To illustrate the difference between these two situations, we consider the simplest case of a nonreactive medium (as in Model D above), in which the current can be produced by ionic diffusion or by the electric field, or by both. To calculate the time variations of ionic concentration and extracellular voltage, we consider the current density:

$$\vec{j} = D\nabla e[c] + \sigma\vec{E}. \quad (34)$$

According to the differential law for charge conservation and Poisson law, we have

$$\nabla\vec{j} + \frac{\partial\rho}{\partial t} = D\nabla^2\rho + \frac{\sigma}{\varepsilon}\rho + \frac{\partial\rho}{\partial t} = 0. \quad (35)$$

When ionic diffusion is negligible compared to Ohm's law, we have

$$\frac{\partial\rho}{\partial t} + \frac{\sigma}{\varepsilon}\rho \approx 0.$$

It follows that the charge density is given by

$$\rho = \rho_o \exp\left(-\frac{\sigma}{\varepsilon}t\right). \quad (36)$$

On the other hand, if ionic diffusion is the primary cause of current propagation in the extracellular medium, then the relaxation time should be much larger and thus, integration should be observed. When ionic diffusion is dominant, we have

$$\frac{\partial\rho}{\partial t} + D\nabla^2\rho \approx 0$$

instead of Eq. 35.

The general solution is

$$\rho = \frac{1}{\sqrt{2Dt}} \int_0^\infty \rho(r, 0) e^{-\frac{r^2}{4Dt}} dr. \quad (37)$$

The difference between the expressions above (Eqs. 36 and 37) shows that the time variation of charge density is different according to which current dominates, electric field current or ionic diffusion current. The same applies to the electric potential between the electrode and a given reference, because this potential is linked to charge density through Poisson's law. Therefore, this experiment would be crucial to clearly show which of the two currents primes for currents perpendicular to the source (this would not apply to longitudinal currents, like axial currents in dendrites). In the hypothetical case of dominant ionic diffusion, the cortex would be similar to a Warburg impedance and one can estimate the macroscopic diffusion coefficient using Eq. 37.

Thus, using a micropipette injecting periodic current pulses, it should be possible to test the capacity of the medium to create charge accumulation for physiological frequencies. If this is the case, it would constitute evidence that ionic currents are nonnegligible in the physiological situation.

## DISCUSSION

In this article, we have proposed a framework to model local field potentials, and which synthesizes previous measurements and models. This framework integrates microscopic measurements of electric parameters (conductivity  $\sigma$  and permittivity  $\varepsilon$ ) of extracellular fluids, with macroscopic measurements of those parameters ( $\sigma_\omega^M, \varepsilon_\omega^M$ ) in cortical tissue (8,10). It also integrates previous models of LFPs, such as the continuum model (7), which was based on a continuum hypothesis of electric parameters variations in extracellular space, or the polarization model (9), which explicitly considered different media (fluid and membranes) and their polarization by the current sources. This model is more general and also integrates ionic diffusion, which is predicted as a major determinant of the frequency dependence of LFPs. This diffusion-polarization model also accounts for observations of  $1/f$  frequency scaling of LFP power spectra, which is due here to ionic diffusion, and is therefore predicted to be a consequence of the genesis of the LFP signal, rather than being solely due to neuronal activity (see (2)). Finally, this work suggests that emphatic interactions between neurons can occur not only through electric fields but that ionic diffusion should also be considered in such interactions.

As discussed in Simplified Geometry for Macroscopic Parameters, this model rests on several approximations, which were necessary to obtain the analytic expressions used here. These approximations were that current sources were considered as monopolar entities (longitudinal currents

such as axial currents in dendrites were not taken into account), the current source was spherical and the extracellular medium was considered isotropic. Because multipolar effects can be reconstructed from the superposition of monopoles (20), the monopolar configuration should not affect the results on frequency dependence, as long as the extracellular current perpendicular to the source is considered. Similarly, the exact geometry of the current source should have no influence on the frequency dependence far away from the sources. However, in the immediate vicinity of the sources, the geometric nature and the synchrony of synaptic currents can have influences on the power spectrum (26). Another assumption is that the extracellular medium is isotropic, which was justified within the macroscopic framework followed here. These factors, however, will influence the exact shape of the LFPs. More quantitative models including a more sophisticated geometry of current sources and the presence of membrane excitability and action potentials should be considered (e.g., (26,27)).

The main prediction of this model is that ionic diffusion is an essential physical cause for the frequency dependence of LFPs. We have shown that the presence of ionic diffusion allows the model to account quantitatively for the macroscopic measurements of the frequency dependence of electric parameters in cortical tissue (8). Ionic diffusion is responsible for a frequency dependence of the impedance as  $1/\sqrt{\omega}$  for low frequencies ( $<1000$  Hz), which directly accounts for the observed  $1/f$  frequency scaling of LFP and EEG power spectra during wakefulness (1–4) (see Fig. 4). Note that the EEG is more complex because it depends on the diffusion of electric signals across fluids, dura matter, skull, muscles, and skin. However, this filtering is of low-pass type, and may not affect the low-frequency band, so there is a possibility that the  $1/f$  scaling of EEG and LFPs have a common origin. This model is consistent with the view that this apparent  $1/f$  noise in brain signals is not generated by self-organized features of brain activity, but is rather a consequence of the genesis of the signal and its propagation through extracellular space (2).

It is important to note that the fact that ionic diffusion may be responsible for  $1/f$  frequency scaling of LFPs is not inconsistent with other factors, which may also influence frequency scaling. For example, the statistics of network activity—and more generally network state—can affect frequency scaling. This is apparent when comparing awake and slow-wave sleep LFP recordings in the same experiment, showing that the  $1/f$  scaling is only seen in wakefulness but  $1/f^2$  scaling is instead seen during sleep (2) (see Fig. 5). In agreement with this, recent results indicate that the correlation structure of synaptic activity may influence frequency scaling at the level of the membrane potential, and that correlated network states scale with larger (more negative) exponents (28).

We also investigated ways to explain the measurements of Logothetis et al. (10), who reported that the extracellular medium was resistive and therefore did not display frequency

dependence, in contradiction with the measurements of Gabriel et al. (8). We summarize and discuss our conclusions below.

In the experiments of Gabriel et al. (8), one measures permittivity and conductivity in the medium in between two metal plates. This forms a capacitor, which (macroscopic) complex impedance is measured. This measure actually consists of two independent measurements, the real and imaginary part of the impedance. These values are used to deduce the macroscopic permittivity and macroscopic conductivity of the medium. However, at the interface between the medium and the metal plates, the flow of electrons in the metal corresponds to a flow of charges in the tissue, and a variety of phenomena can occur, which can interfere with the measurement. The accumulation of charges that occurs at the interface between the electrode and the extracellular fluid implies a polarization impedance, which depends on the interaction between ions and the metal plate. Because this accumulation of charge implies a variation of concentration, the flow of ions may involve an important component of ionic diffusion.

In the experiments of Logothetis et al., a system of four electrodes is used; the two extreme electrodes inject current in the medium, while the two electrodes in the middle are exclusively used to measure the voltage. This system is supposed to be more accurate than Gabriel's, because the electrodes that measure voltage are not subject to charge accumulation. However, the drawback of this method are nonlinear effects. The magnitude of the injected current is such that the voltage at the extreme electrodes saturates. This voltage saturation also implies saturation of concentration (capacitive effect between electrodes), which limits ionic diffusion currents. Thus, the ratio between ionic diffusion currents and the currents due to the electric field is greatly diminished relative to the experiments of experiments of Gabriel et al.

We think that natural current sources are closer to the situation of Gabriel et al. for several reasons. First, the magnitude of the currents produced by biological sources is far too low for saturation effects. Second, the flow of charges across ion channels will produce perturbations of ionic concentration, which will be reequilibrated by diffusion. The effects may not be as strong as the perturbations of concentrations induced by the experiments of Gabriel et al., but ionic diffusion should play a role in both cases. This is precisely one of the aspects that should be evaluated in further experiments.

The experiments of Logothetis et al. were done using a four-electrode setup, which neutralizes the influence of electrode impedance on voltage measurements (29,30). This system was used to perform high-precision impedance measurements, also avoiding ionic diffusion effects (10). Indeed, these experimental conditions, and the apparent resistive medium, could be reproduced by this model if ionic diffusion was neglected. This model therefore formulates the strong prediction that ionic diffusion is important, and that

any measurement technique should allow ionic diffusion to reveal the correct frequency-dependent properties of impedance and electric parameters in biological tissue.

The critical question that remains to be solved is whether, in physiological conditions, ionic diffusion plays a role as important as suggested here. We propose a simple method to test this hypothesis. The frequency dependence could be evaluated by using an extracellular electrode injecting current in conditions as close as possible to physiological conditions (a micropipette would be appropriate). By measuring the integration of the extracellular voltage after periodic current injection, one could estimate the relaxation time of the medium with respect to charge accumulation. If this relaxation time occurs at timescales relevant to neuronal currents (milliseconds) rather than the fast relaxation predicted by a purely resistive medium (picoseconds), then ionic diffusion will necessarily occur in physiological conditions, which would provide evidence in favor of this mechanism.

## APPENDIX A: ESTIMATION OF IONIC DIFFUSION CURRENT VERSUS ELECTRIC DIFFUSION IN SEA WATER AND IN CORTEX

In this Appendix, we evaluate the ratio between ionic diffusion currents and electric field currents in the extracellular space directly adjacent to the source. This ratio measures whether the ionic diffusion current perpendicular to the membrane is greater than the electric field current. We will design this ratio by the term  $r_{ie}$ .

We have in general

$$\vec{j}_{\text{Total}} = eD\nabla[C] + \sigma\nabla V, \quad (38)$$

where the first term on the right-hand side is the electric current density produced by ionic diffusion, and the second term is that produced by differential Ohm's law.

For a displacement  $\Delta r = 10$  nm in the direction across the membrane (from inside to outside), we have approximatively

$$\vec{j}_{\text{Total}} \cdot d\vec{r} \approx \vec{j}_{\text{Total}} \cdot \vec{\Delta r} = e\Delta[C]_{\Delta r=10 \text{ nm}} + \sigma_{\omega}\Delta V_{\Delta r=10 \text{ nm}}. \quad (39)$$

Suppose that we have a spherical cell of 10- $\mu$ m radius, embedded in sea water and at resting potential. The resting membrane potential is a dynamic equilibrium between inflow and outflow of charges, in which these two fluxes are equal on (temporal) average. Fluctuations of current around this average in the extracellular medium around the membrane have all characteristics of thermal noise (31) because the shot noise (see (16,32)) is zero when the current is zero on average, such that the net charge on the external side of the membrane varies around a mean value with the same characteristics as white noise (thermal noise). These fluctuations will therefore be present also at the level of the membrane potential. In this Appendix, we evaluate the order of magnitude of the electric current caused by ionic diffusion, relative to the electric field for this situation of dynamic equilibrium. First, the ratio between the membrane voltage noise and the variation of total charge concentration is given by

$$\begin{aligned} \Delta Q_{\text{tot}} &= C\Delta V_{\text{membrane}} = k_1\Delta V_{\text{membrane}} \\ &= 1.25 \times 10^{-11}\Delta V_{\text{membrane}}, \end{aligned} \quad (40)$$

because the cell's capacitance is given by  $C = 4\pi R^2 C_m = 0.04\pi R^2$  ( $C_m \approx 10^{-2}$  F/m<sup>2</sup>).

Second, mass conservation imposes

$$\Delta Q_{\text{tot}} = e^* v_{\text{eff}}^* \Delta[C]_{\text{tot}}, \quad (41)$$

where  $e = 1.69 \times 10^{-19}$  C and  $v_{\text{eff}}$  is the volume of the spherical shell containing the charges. Because the charges are not uniformly distributed inside the cell, but rather distributed within a thin spherical shell adjacent to the membrane, because the electric field developed across the membrane is very intense (of  $\sim \frac{70 \times 10^{-3}}{7 \times 10^{-9}} = \text{V/m} = 10^7$  V/m). Thus, the width of the shell is  $\sim dR \approx 10^{-4}R < 1$  nm, where the volume of the spherical shell is approximately equal to  $4\pi R^2 dR$ . In this case, we have for monovalent ions ( $|z| = 1$ )

$$\begin{aligned} \Delta Q_{\text{tot}} &= k_2 \Delta[C]_{\text{tot}} \approx 2.2 \times 10^{-38} \Delta[C]_{\text{tot}} \\ &= 2.2 \times 10^{-38} \Delta[C]_{\Delta r=10 \text{ nm}} \end{aligned} \quad (42)$$

if we assume that the variation of concentration on the adjacent border of the exterior surface of the cell is within a width of 10 nm.

In this case, we have

$$\frac{\Delta[C]_{\Delta r=10 \text{ nm}}}{\Delta V_{\text{membrane}}} = \frac{k_1}{k_2} \approx 10^{27} \text{ C/m}^3 \text{V}. \quad (43)$$

Third, the potential difference between the cell surface and 10 nm away from it, is given by

$$\begin{aligned} \Delta V_{\Delta r=10 \text{ nm}} &= \Delta V_{\text{membrane}} \\ &\quad - \frac{R\Delta V_{\text{membrane}}}{R+r} \approx 10^{-3}\Delta V_{\text{membrane}}. \end{aligned} \quad (44)$$

Consequently, the ratio between ionic diffusion current and electric diffusion current caused by thermal noise in sea water obeys

$$r_{ie} \approx \frac{eD_{\text{sea}}\Delta[C]_{\Delta r=10 \text{ nm}}\Delta t}{\sigma_{\omega}^{\text{sea}}\Delta V_{\Delta r=10 \text{ nm}}} = \frac{eD_{\text{sea}}k_1}{\sigma_{\omega}^{\text{sea}}k_2} \approx \frac{10^2}{\sigma_{\omega}^{\text{sea}}}, \quad (45)$$

where the diffusion constant of K<sup>+</sup> or Na<sup>+</sup> in sea water is  $\sim 10^{-9} \frac{\text{m}^2}{\text{s}}$ . This implies that the ratio  $r$  is  $\gg 1$  for frequencies  $< 1000$  Hz because  $\sigma_{\omega}^{\text{sea}}$  of sea water is necessarily  $< 2$  S/m.

Because tortuosity is given by  $\lambda = \sqrt{D_{\text{sea}}/D_{\text{cortex}}}$ , and is comprised between 1.6 and 2.2 (for small and large molecules, respectively) in cerebral cortex (33–35), the macroscopic diffusion constant in cortex is certainly larger than  $D_{\text{sea}}/10$ . Thus, we have

$$r_{ie}^{\text{cortex}} > \frac{10}{\sigma_{\omega}^{\text{cortex}}} \quad (46)$$

where  $\sigma_{f=100\text{Hz}}^{\text{cortex}} \approx 0.1$  S/m (see (11)).

This evaluation shows that the phenomenon of ionic diffusion is essential to determine the current field in the cortex.

Finally, we note that we did not need to evaluate the absolute magnitude of  $\Delta V$  in our evaluation. This evaluation is valid for a physical situation where we have a permanent white noise over a distance of 10 nm, independently of the intensity of this noise (which in practice will depend on many factors, such as the size of the cell, the number of ion channels, etc).

## APPENDIX B: FREQUENCY SCALING OF IONIC DIFFUSION

In this Appendix, we calculate the frequency dependence of ionic diffusion current outside of a spherical current source. We consider a constant variation of ionic concentration,  $\Delta X_i$ , on the surface of the source, and a null variation at an infinite distance (Warburg conditions).

The diffusion equation for a given ionic species is

$$\frac{\partial \Delta X_i}{\partial t} = D_i \nabla^2 \Delta X_i, \quad (47)$$

where  $\Delta X_i$  is the perturbation of concentration  $X_i$  of ion  $i$  around the steady-state value, and  $D_i$  is the associated diffusion coefficient. This diffusion coefficient depends of the ionic species considered and the structure of the medium. Because the geometry of the problem and the limit conditions respect spherical symmetry, we use spherical coordinates. In this coordinate system, we have

$$\frac{\partial \Delta X_i}{\partial t} = D_i \left[ \frac{\partial^2 \Delta X_i}{\partial r^2} + \frac{2}{r} \frac{\partial \Delta X_i}{\partial r} \right] \quad (48)$$

because  $\Delta X_i$  does not depend on  $\theta$  and of  $\Phi$  (spherical symmetry). The Fourier transform of  $\Delta X_i$  with respect to time gives

$$\frac{\partial^2 X_{i\omega}}{\partial r^2} + \frac{2}{r} \frac{\partial \Delta X_{i\omega}}{\partial r} = \frac{d^2 X_{i\omega}}{dr^2} + \frac{2}{r} \frac{d \Delta X_{i\omega}}{dr} = \frac{i\omega}{D_i} \Delta X_{i\omega}, \quad (49)$$

which general solution is given by

$$\Delta X_{i\omega} = A(\omega) \frac{e^{\sqrt{\frac{i\omega}{D_i}} r}}{r} + B(\omega) \frac{e^{-\sqrt{\frac{i\omega}{D_i}} r}}{r}. \quad (50)$$

For a variation of concentration at the source border which is independent of frequency and which satisfies the Warburg hypothesis (the variation of concentration tends to zero at an infinite distance (23,36)), we have

$$\Delta X_{i\omega}(r) = \Delta X_{i\omega}(R) \cdot \frac{R e^{-\sqrt{\frac{i\omega}{D_i}}(r-R)}}{r}, \quad (51)$$

where  $r$  is the distance between the center of the source and  $R$  is the radius of the source.

Thus, the electric current density produced by ionic diffusion is given by

$$\vec{J}_i(r) = ZeD_i \frac{\partial \Delta X_i}{\partial r} \hat{r} = -ZeD_i \left( \frac{1}{r} + \sqrt{\frac{i\omega}{D_i}} \right) \Delta X_{i\omega}(r) \hat{r}, \quad (52)$$

where  $Ze$  is the charge of ions  $i$ .

Because we can consider that the source and extracellular medium form a spherical capacitor, the voltage difference between the surface of the source and infinite distance is given by  $ZeC\Delta X_{i\omega}(R)$ , where  $C$  is the capacitance value. Thus, the electric impedance of the medium is given by

$$Z_\omega = \frac{C}{D_i \left( \frac{1}{R} + \sqrt{\frac{i\omega}{D_i}} \right)}. \quad (53)$$

For a source of radius  $R = 10 \mu\text{m}$  and a macroscopic ionic diffusion coefficient of  $\sim 10^{-11} \text{ m}^2/\text{s}$ , and for frequencies  $> 1 \text{ Hz}$ , we can approximate the impedance by

$$Z_\omega \approx \frac{C}{\sqrt{i\omega D_i}}. \quad (54)$$

The same expression for the impedance is also obtained in cylindrical coordinates or planar Cartesian coordinates (not shown).

Note that if several ionic species are present, then the superposition principle applies (Fick equations are linear) and therefore the contribution of each ion will add up. The diffusion constants for different ions are of the same order of magnitude (for  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Cl}^-$ ,  $\text{Ca}^{2+}$ ), so no particular ion would be expected to dominate.

We thank Christoph Kayser, Nikos Logothetis, Axel Oeltermann, and anonymous reviewers for useful comments on the manuscript.

This research was supported by Centre National de la Recherche Scientifique, Agence Nationale de la Recherche, and the European Community (FACETS project).

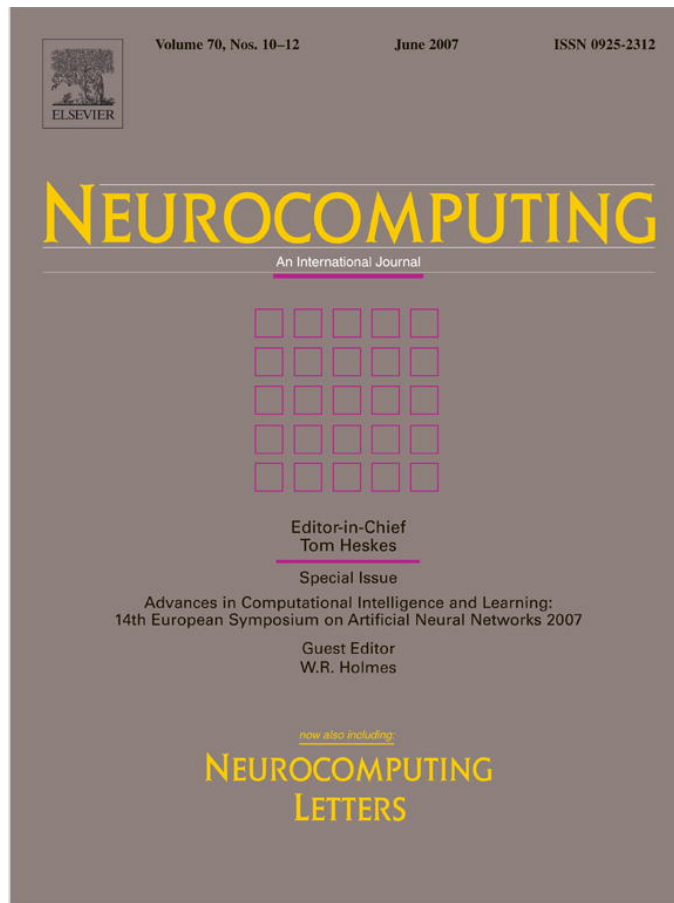
## REFERENCES

- Bhattacharya, J., and H. Petsche. 2001. Universality in the brain while listening to music. *Proc. Biol. Sci.* 268:2423–2433.
- Bédard, C., H. Kröger, and A. Destexhe. 2006. Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Phys. Rev. Lett.* 97:118102.
- Novikov, E., A. Novikov, D. Shannahoff-Khalsa, B. Schwartz, and J. Wright. 1997. Scale-similar activity in the brain. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 56:R2387–R2389.
- Pritchard, W. S. 1992. The brain in fractal time: 1/f-like power spectrum scaling of the human electroencephalogram. *Int. J. Neurosci.* 66:119–129.
- Jensen, H. J. 1998. Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems. Cambridge University Press, Cambridge, UK.
- Beggs, J., and D. Plenz. 2003. Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23:11167–11177.
- Bédard, C., H. Kröger, and A. Destexhe. 2004. Modeling extracellular field potentials and the frequency-filtering properties of extracellular space. *Biophys. J.* 86:1829–1842.
- Gabriel, S., R. W. Lau, and C. Gabriel. 1996. The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz. *Phys. Med. Biol.* 41:2251–2269.
- Bédard, C., H. Kröger, and A. Destexhe. 2006. Model of low-pass filtering of local field potentials in brain tissue. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 73:051911.
- Logothetis, N. K., C. Kayser, and A. Oeltermann. 2007. In vivo measurement of cortical impedance spectrum in monkeys: implications for signal propagation. *Neuron.* 55:809–823.
- Gabriel, S., R. W. Lau, and C. Gabriel. 1996. The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum tissues. *Phys. Med. Biol.* 41:2271–2293.
- Cole, K. S., and R. H. Cole. 1941. Dispersion and absorption in dielectrics. I. Alternating current characteristics. *J. Chem. Phys.* 9:341–351.
- Gabriel, S., R. W. Lau, and C. Gabriel. 1996. The dielectric properties of biological tissues: I. Literature survey. *Phys. Med. Biol.* 41:2231–2249.
- Berg, H. C. 1983. *Random Walks in Biology* Princeton University Press, Princeton.
- Maxwell, J. C. 1873. *A Treatise on Electricity and Magnetism.*, Vol. 1 Clarendon Press, Oxford, UK.
- Buckingham, M. J. J. 1985. *Noise in Electronic Devices and Systems* John Wiley & Sons, New York.
- Foster, K. R., and H. P. Schwan. 1989. Dielectric properties of tissues and biological materials: a critical review. *Crit. Rev. Biomed. Eng.* 17:25–104.
- Kronig, R. D. L. 1926. On the theory of dispersion of x-rays. *J. Opt. Soc. Am.* 12:547–556.
- Landau, L. D., and E. M. Lifshitz. 1984. *Electrodynamics of Continuous Media* Pergamon Press, Moscow, Russia.
- Purcell, E. M. 1984. *Electricity and Magnetism* McGraw Hill, New York.
- Braitenberg, V., and A. Shutz. 1998. *Cortex: Statistics and Geometry of Neuronal Connectivity*, 2nd Ed. Springer-Verlag, Berlin, Germany.
- Nunez, P. L., A. Novikov, and R. Srinivasan. 2006. *Electric Fields of the Brain*, 2nd Ed. Oxford University Press, Oxford, UK.
- Diard, J.-P., B. Le Gorrec, and C. Montella. 1999. Linear diffusion impedance. General expression and applications. *J. Electroanal. Chem.* 471:126–131.

24. Hooge, F. N. P. A. B. 1997. On the correlation function of  $1/f$  noise. *Physica B (Amsterdam)*. 239:223–230.
25. Hooge, F. N. 1962.  $1/f$  noise is no surface effect. *Phys. Lett.* 29A:139–140.
26. Pettersen, K. H., and G. T. Einevoll. 2008. Amplitude variability and extracellular low-pass filtering of neuronal spikes. *Biophys. J.* 94:784–802.
27. Gold, C., D. A. Henze, C. Koch, and G. Buzsaki. 2006. On the origin of the extracellular action potential waveform: a modeling study. *J. Neurophysiol.* 95:3113–3128.
28. Marre, O., S. El Boustani, P. Baudot, M. Levy, C. Monier, et al. 2007. Stimulus-dependency of spectral scaling laws in V1 synaptic activity as a read-out of the effective network topology. *Soc. Neurosci. Abstr.* 33:790.6.
29. Geddes, L. A. 1997. Historical evolution of circuit models for the electrode-electrolyte interface. *Ann. Biomed. Eng.* 25:1–14.
30. McAdams, E. T., and J. Jossinet. 1992. A physical interpretation of Schwan's limit current of linearity. *Ann. Biomed. Eng.* 20:307–319.
31. Nyquist, H. 1928. Thermal agitation of electric charge in conductors. *Phys. Rev.* 32:110–113.
32. Vasilyev, A. M. 1983. An Introduction to Statistical Physics, MIR Editions, Moscow, Russia.
33. Nicholson, C., and E. Sykova. 1998. Extracellular space structure revealed by diffusion analysis. *Trends Neurosci.* 21:207–215.
34. Nicholson, C. 2005. Factors governing diffusing molecular signals in brain extracellular space. *J. Neural Transm.* 112:29–44.
35. Rusakov, D. A., and D. M. Kullmann. 1998. Geometric and viscous components of the tortuosity of the extracellular space in the brain. *Proc. Natl. Acad. Sci. USA.* 95:8975–8980.
36. Taylor, S. R., and E. Gileadi. 1995. The physical interpretation of the Warburg impedance. *Corrosion.* 51:664–671.
37. Destexhe, A., D. Contreras, and M. Steriade. 1999. Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *J. Neurosci.* 19:4595–4608.



Provided for non-commercial research and educational use only.  
Not for reproduction or distribution or commercial use.



This article was originally published in a journal published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues that you know, and providing a copy to your institution's administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<http://www.elsevier.com/locate/permissionusematerial>



# Comparison of dynamical states of random networks with human EEG

Ralph Meier<sup>a,c,\*</sup>, Arvind Kumar<sup>a</sup>, Andreas Schulze-Bonhage<sup>b,c</sup>, Ad Aertsen<sup>a,c</sup>

<sup>a</sup>Neurobiology and Biophysics, Institute of Biology III, Albert-Ludwigs-University, Freiburg, Germany

<sup>b</sup>Center for Epilepsy, Department of Neurosurgery, University Clinics, Freiburg, Germany

<sup>c</sup>Bernstein Center for Computational Neuroscience Freiburg, Freiburg, Germany

Available online 11 November 2006

## Abstract

Existing models of EEG have mainly focused on relations to network dynamics characterized by firing rates [L. de Arcangelis, H.J. Herrmann, C. Perrone-Capano, Activity-dependent brain model explaining EEG spectra, arXiv:q-bio.NC/0411043 v1, 23 Nov 2004; D.T. Liley, D.M. Alexander, J.J. Wright, M.D. Aldous, Alpha rhythm emerges from large-scale networks of realistically coupled multicompartamental model cortical neurons, *Network* 10(1) (1999) 79–92; O. David, J.K. Friston, A neural mass model for MEG/EEG: coupling and neuronal dynamics, *NeuroImage* 20 (2003) 1743–1755]. Generally, these models assume that there exists a linear mapping between network firing rates and EEG states. However, firing rate is only one of several descriptors for network activity states. Other relevant descriptors are synchrony and irregularity of firing patterns [N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, *J. Comput. Neurosci.* 8(3) (2000) 183–208]. To develop a better understanding of the EEG we need to relate these state descriptors to EEG states. Here, we try to go beyond the firing rate based approaches described in [D.T. Liley, D.M. Alexander, J.J. Wright, M.D. Aldous, Alpha rhythm emerges from large-scale networks of realistically coupled multicompartamental model cortical neurons, *Network* 10(1) (1999) 79–92; O. David, J.K. Friston, A neural mass model for MEG/EEG: coupling and neuronal dynamics, *NeuroImage* 20 (2003) 1743–1755] and relate synchronicity and irregularity in the network to EEG states. We show that the transformation between network activity and EEG can be approximately mediated by linear kernel with the shape of an  $\alpha$ - or  $\gamma$ -function, allowing us a comparison between EEG states and network activity space. We find that the simulated EEG generated from asynchronous irregular type network activity is closely related to the human EEG recorded in the awake state, evaluated using power spectral density characteristics.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** Simulated EEG; EEG model; Human EEG; Cortical dynamics; Brain state; Asynchronous irregular activity; Parallel computing; Power spectral density

## 1. Introduction

Cortical activity can be recorded at various levels of details ranging from in vivo intracellular recording (microscopic activity) to global population activity such as LFP, ECoG, and EEG (macroscopic activity). While there is a good understanding of the origin of the microscopic activity, very little is known about the origin of the macroscopic activity. It has long been speculated that the macroscopic cortical activity is generated as a

consequence of network activity [5,11,10]. In fact, several modeling studies have been able to relate network dynamics to EEG states by assuming a linear mapping between the network firing rates and oscillations in the EEG [1,7,4]. However, network dynamics is not only characterized by firing rates, but also by synchronization in neural populations and irregularity of single-neuron firing patterns [3]. To understand how cortical background activity states generate the EEG we need to relate these state descriptors to EEG states. We show that the mapping between the population activity in the network can be approximated by a linear kernel described by either an  $\alpha$ -function or a  $\gamma$ -function. The simulated EEG (see materials and methods) corresponding to asynchronous irregular (AI) and synchronous irregular states showed a

\*Corresponding author. Neurobiology and Biophysics, Institute of Biology III, Albert-Ludwigs-University, Freiburg, Germany. Tel.: +49 761 2032864; fax: +49 761 2032860.

E-mail address: [meier@biologie.uni-freiburg.de](mailto:meier@biologie.uni-freiburg.de) (R. Meier).

good match with the human EEG—especially in theta and delta bands. Heterogeneous network simulations resembled the human background EEG even better—also in the alpha and beta bands.

## 2. Material and methods

### 2.1. Networks

We performed simulations of homogeneous and heterogeneous networks consisting of 50,000 leaky integrate and fire type neurons (80% excitatory and 20% inhibitory neurons), representing  $\approx 0.5\text{mm}^2$  slice of cortex [2]. The neurons were connected randomly with a connection probability of 0.1. In a homogeneous network all neurons had identical passive properties. To introduce heterogeneity into the network, the passive properties (membrane capacity  $C$  and conductivity at resting condition  $G_{\text{rest}}$ ) and the spiking threshold ( $V_{\text{thresh}}$ ) of the neurons were chosen from a normal distribution (mean  $\pm$  SD);  $C = 250 \pm 25$  pF,  $G_{\text{rest}} = 16,7 \pm 1.5$  nS and  $V_{\text{thresh}} = -55 \pm 5.5$  mV. The mean values of the passive parameters in the heterogeneous network were identical to the values of the passive properties in the homogeneous network. The simulations of the neuronal networks were performed using the NEST [9,12] simulation environment. The output of the spiking network simulations (spike patterns of sustained activity in dynamical networks recorded over several seconds) were then used for further analysis.

We obtained a network population signal ( $N_{\text{pop}}$ ) by binning the spikes (binwidth = 2 ms) of all neurons in the network. To characterize the dynamical states of simulated network activity, both at the level of single neurons and neuron populations, we employed the following descriptors, see [6] for details:

*Mean firing rate* of the activity was estimated as the mean spike count per second of the neurons in the network.

*Synchrony* in the network was measured by the pair wise correlations ( $\rho_{\text{net}}$ ) in the network. A population of identical independent Poisson processes yield a  $\rho_{\text{net}} = 0$ , any mutual dependence results in an increase in  $\rho_{\text{net}}$ .

*Irregularity* of individual spike trains was measured by the squared coefficient of variation of the corresponding inter-spike interval (ISI) distribution. Low values reflect more regular spiking, and a clock-like pattern yields  $CV^2 = 0$ . By contrast,  $CV^2 = 1$  indicates Poisson-type behavior (cf. Fig. 1a and b).

### 2.2. Generating simulated EEG

The spectral bandwidth of  $N_{\text{pop}}$  is much wider than the EEG signal bandwidth. To draw a comparison between  $N_{\text{pop}}$  and EEG, it is required to limit the bandwidth of  $N_{\text{pop}}$ . To achieve that we transformed  $N_{\text{pop}}$  to Sim-EEG by convolving  $N_{\text{pop}}$  with either an  $\alpha$ -function or a  $\gamma$ -function shaped kernel (cf. Fig. 1c). The parameters for the two kernels were determined in an optimization process by

maximizing the similarity of power spectral densities (PSDs) obtained from the Sim-EEG generated using one AI network state (see Section 3) and EEG data from two healthy, awake subjects who were asked to focus on a fixation point. After the optimization process, the kernel parameters were fixed. In an independent test set (both networks and EEG data) we then compared different network states (both homo- and heterogeneous) as the basis for Sim-EEG using the aforementioned kernels and comparing the resulting Sim-EEG to the human background EEG data from more than 90 subjects (cf. [8]). We used the cross correlation ( $\rho$ ) between the powerspectra of the SIM-EEG and the recorded EEG (data was kindly provided by the Center for Epilepsy, University Clinics Freiburg) to quantify the similarity between the two signals.

## 3. Results

### 3.1. Network activity dynamics

A large random network of integrate and fire neurons exhibits a continuum of activity states, depending on the intensity of external excitatory inputs ( $v_{\text{ext}}$ ), and on the recurrent inhibition/excitation balance ( $g$ ). The firing pattern of individual neurons varies between regular ( $R$ ) ( $CV_{\text{ISI}} \approx 0$ ) and irregular ( $I$ ) ( $CV_{\text{ISI}} \approx 1$ ), population activity varies between synchronous ( $S$ ) ( $\rho_{\text{net}} \approx 1$ ) and asynchronous ( $A$ ) ( $\rho_{\text{net}} \approx 0$ ). Still, the network activity state can be attributed to one of four characteristic states, viz. AI, SI, AR, or SR as a function of  $v_{\text{ext}}$  and ( $g$ ) [cf. 3,6]. Of these, it is the AI regime where network activity is considered to most closely resemble cortical spiking activity in vivo (Fig. 1a and b). Note that in the AI regime  $\rho_{\text{net}} \approx 0.002$ , this results in transient synchrony in the network [6]. The  $\rho_{\text{net}}$  can be further reduced by introducing heterogeneities in the network (data not shown). The homogeneous and heterogeneous networks, however, do not differ systematically in the repertoire of states they exhibit.

### 3.2. Sim-EEG

As the AI state resembles the ongoing activity in vivo most closely, we assumed the AI state to be the network activity underlying the EEG obtained from healthy awake human subjects and used the corresponding  $N_{\text{pop}}$  to optimize the time constant of the convolution kernels ( $\alpha$ -function and  $\gamma$ -function). The resulting optimal width of the  $\alpha$ -function was estimated to be  $\approx 40$  ms (cf. Fig. 1c). Examples for the resulting power spectra for the generated SIM-EEG based on different network states in comparison to the recorded EEG are shown in Fig. 1d. To quantify the similarity between the power spectra we chose to calculate the mean correlation coefficient for relevant frequency bands. Fig. 1e shows the correlation coefficients in the independent test set ( $\rho$ ) between PSD of Sim-EEG



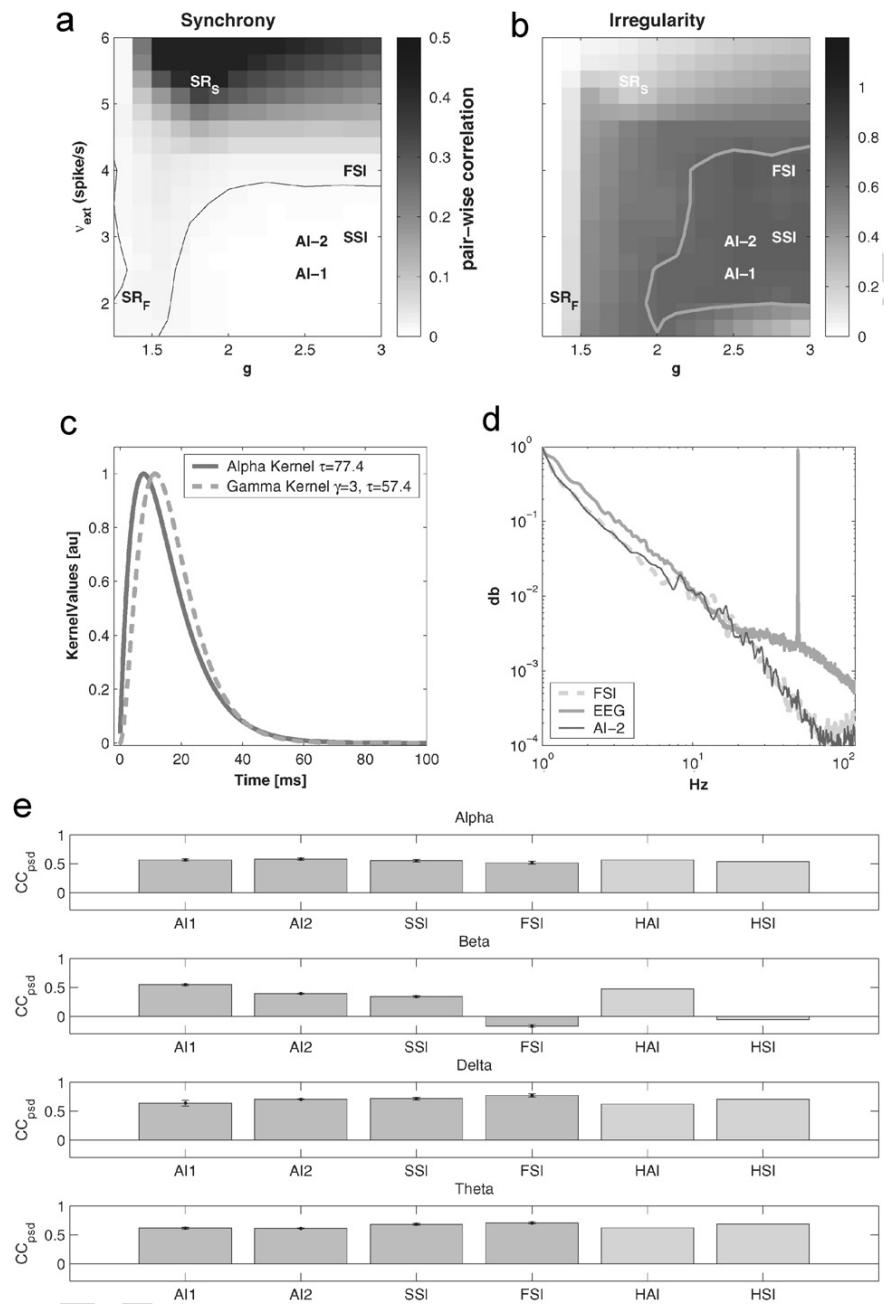


Fig. 1. Network activity states, kernels, and power spectra; correlation of SIM-EEG to EEG: (a) synchrony and (b) irregularity in the homogeneous network as a function of  $v_{ext}$  and  $g$ . The contour ( $\rho_{net} = 0.02$ ) in (a) separates synchronous states from asynchronous states, while the contour ( $CV^2 = 0.8$ ) in (b) separates regular states from irregular states. Characteristic network states (e.g. SR, AI, and SI) are indicated. (c) Visualization of estimated  $\alpha$ - and  $\gamma$ -kernel with respective parameters. (d) Exemplary power spectra for two different simulated network states (nAI and AI-SI) and human EEG. (e) Correlation coefficients ( $\rho$ ) between different characteristic bands ( $\alpha, \beta, \delta$ , and  $\theta$ ) for the PSDs of SIM-EEG obtained from homogeneous networks mapping different states (asynchronous irregular (AI-1), nearly AI (AI-2), slow synchronous irregular (SSI), and fast synchronous irregular (FSI)) and heterogeneous networks with AI and SI states (HAI and HSI) to human EEG representing normal and awake behaving activity.

(homogeneous and heterogeneous networks) and PSDs of EEG recorded in human subjects, in four different frequency bands (viz.  $\alpha$ [8–13 Hz],  $\beta$ [14–35 Hz],  $\theta$ [4–7 Hz], and  $\delta$ [ $\leq 3$  Hz]). The Sim-EEG corresponding to the AI-2 and SI states showed a good match with the human EEG ( $\rho \geq 0.6$ ) in the  $\theta$  and  $\delta$  bands, less in the  $\alpha$  and

$\beta$  bands. The Sim-EEG generated from the heterogeneous network simulations resembled the human EEG even better ( $\rho \geq 0.7$ ), now also in the  $\alpha$  and  $\beta$  frequency bands.

We observed a small mismatch between the Sim-EEG and human EEG, which could be due to an inappropriate choice of the convolution kernel. Therefore we changed the

kernel function to be a  $\gamma$ -function, and repeated the function fitting and re-evaluated the correlations of the PSDs. We found that convolution of  $N_{\text{pop}}$  with a  $\gamma$ -function kernel gave a slightly better fit, measured by correlation of PSD bands with the human EEG (data not shown).

#### 4. Discussion

Here we presented a first attempt to relate the spiking activity of cortical network, to the macroscopic activity of the brain, as captured by the scalp EEG. Generally, the models AI state very closely resembles cortical activity in vivo in awake, behaving animals. Therefore, we assumed that healthy human EEG recordings correspond to an AI state in the cortical network model.

Our comparison of Sim-EEG based on AI state networks and recorded EEG from awake humans supports this assumption. For this comparison we started with an  $\alpha$ -function shaped kernel. This choice was motivated by the fact that the network activity is low-pass filtered by the cortical tissue and the skull. The  $\alpha$ -function shaped kernel resulted in a reasonably high correlation between Sim-EEG and real EEG. However, there were also notable differences between the SIM-EEG and EEG. Therefore we used another kernel ( $\gamma$ -function), which additionally allowed us to control the rising behavior of the kernel. The  $\gamma$ -function shaped kernel indeed resulted in a higher correlation and the spectra of SIM-EEG resembled recorded human EEG slightly better. This might be due to the additional degree of freedom allowed in the kernel-estimation process.

For this study we ignored the orientation of cortical cells with respect to the recording surface electrode. We assumed that all neurons in our simulations contribute equally to the surface background EEG. From previous modeling studies it is known, that the state space of networks studied here does not change considerably when the total number of neurons is increased [6]. This might allow us, in a first approach, to consider only neurons contributing to (SIM-) EEG recordings. Moreover, since a comparison of spectral properties from both, SIM-EEG and EEG, eliminates temporal causality, potential intracortical firing patterns in neurons of different orientations cannot currently be studied by this approach. However, this basic approach, with the aim of building up a simple phenomenological connection between the background EEG and microscopic network activity, can be extended to connect the activity of layered/oriented networks to EEG—where it naturally would be important to characterize the role of parallel currents.

Currently we are investigating the potential mapping of various network activity states to clinically and behaviorally relevant EEG states. Though there is good hope for bridging the gap between network simulations and electrophysiological population activity data from human recordings, further improvements in both, network models and conversion procedures, will be needed.

#### Acknowledgments

We acknowledge stimulating discussions with Dr. Tonio Ball. This work was supported by the DFG GraKo-843 and the German Federal Ministry of Education and Research (BMBF Grant 01GQ0420 to BCCN Freiburg).

#### References

- [1] L. de Arcangelis, H.J. Herrmann, C. Perrone-Capano, Activity-dependent brain model explaining EEG spectra, arXiv:q-bio.NC/0411043 v1, 23 Nov 2004.
- [2] V. Braitenberg, A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*, second ed., Springer, Berlin, 1998.
- [3] N. Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, *J. Comput. Neurosci.* 8 (3) (2000) 183–208.
- [4] O. David, J.K. Friston, A neural mass model for MEG/EEG: coupling and neuronal dynamics, *NeuroImage* 20 (2003) 1743–1755.
- [5] W.J. Freeman, *Mass Action in the Nervous System*, Academic Press, New York, 1975.
- [6] A. Kumar, S. Schrader, S. Rotter, A. Aertsen, Dynamics of random networks of spiking neurons with conductance-based synapses, in: *Computational and Systems Neuroscience (Cosyne)*, 2005, p. 153.
- [7] D.T. Liley, D.M. Alexander, J.J. Wright, M.D. Aldous, Alpha rhythm emerges from large-scale networks of realistically coupled multicompartmental model cortical neurons, *Network* 10 (1) (1999) 79–92.
- [8] R. Meier, H. Dittrich, A. Schulze-Bonhage, A. Aertsen, Automatic detection of different seizure morphologies in surface EEG without use of prior information, in: 44th Conference of the German Section of the International Liga against Epilepsy, 2004.
- [9] A. Morrison, C. Mehring, T. Geisel, A. Aertsen, M. Diesmann, Advancing the boundaries of high-connectivity network simulation with distributed computing, *Neural Comput.* 17 (8) (2005) 1776–1801.
- [10] P.L. Nunez, *Neocortical Dynamics and Human EEG Rhythms*, Oxford University Press, New York, 1995.
- [11] A. Rotterdam, F.H. Lopes da Silva, J. van der Ende, M.A. Viergever, A.J. Hermans, A model of the spatio-temporal characteristics of the alpha rhythm, *Bull. Math. Biol.* 44 (1982).
- [12] The Neural Simulation Technology Initiative (NEST), software available at ([www.nest-initiative.org](http://www.nest-initiative.org)).



Ralph Meier was born in Germany in 1976. He obtained his Diploma in Biology at the Albert-Ludwigs-University of Freiburg in 2003. Then he obtained his Ph.D. in cooperation with the Center for Epilepsy, Freiburg and the Neurobiology & Biophysics Department at the University Freiburg, Germany in 2006. Currently he is a post-doctoral fellow at the Bernstein Center for Computational Neuroscience, Freiburg. His research is focused on understanding the dynamics of large scale neuronal networks, emergence of epileptiform activity and the development of scientific software.



Arvind Kumar was born in India in 1976. He did his M.E. (Electrical Engg.) from Birla Institute of Technology and Science, Pilani, India in 1999. After a short association with Indian Institute of Technology, Delhi, India, as a senior research fellow, he moved to the University of Freiburg, Germany, where he obtained his Ph.D. in 2006. Currently he is a post-doctoral fellow at Department of Neuroscience, Brown University Providence, USA. His research is focused on understanding the dynamics of neuronal networks and modeling of cortical activity.



**Andreas Schulze-Bonhage** was born in 1960 in Berlin, Germany. He studied medicine at the University of Münster and consecutively worked there at the Institute of Neuroanatomy and Neurophysiology. After his training as a neurologist, he joined the Epilepsy Centre at the University of Bonn as a research fellow and consultant. After visits at the Montreal Neurological Institute and of the Cleveland Clinic Foundation, he became head of the newly founded epilepsy centre at the University of

Freiburg. His research interests focus on clinical epileptology, brain imaging and EGG time series evaluation.

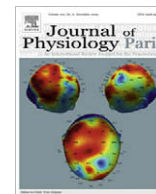


**Ad Aertsen** was born in 1948 in Holland, where he obtained his M.Sc. (University Utrecht) and Ph.D. (University Nijmegen) degrees in Physics. After associations with the University of Pennsylvania (Philadelphia), the Max-Planck-Institute for Biological Cybernetics (Tuebingen), the Hebrew University (Jerusalem), the Ruhr-University (Bochum), and the Weizmann Institute of Science (Rehovot), he is now Professor of Neurobiology and Biophysics at the Albert-Ludwigs-University in Freiburg, Germany ([www.brainworks.uni-freiburg.de](http://www.brainworks.uni-freiburg.de)) and Coordinator of the Bernstein Center for Computational Neuroscience ([www.bccn-freiburg.de](http://www.bccn-freiburg.de)). His research interests focus on the analysis and modeling of activity in biological neural networks and the associated development of neurotechnology.

Author's personal copy

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

## Journal of Physiology - Paris

journal homepage: [www.elsevier.com/locate/jphysparis](http://www.elsevier.com/locate/jphysparis)

## Voltage-sensitive dye imaging: Technique review and models

S. Chemla<sup>a,\*</sup>, F. Chavane<sup>b</sup><sup>a</sup>NeuroMathComp Team, INRIA Sophia-Antipolis, 2004 route des Lucioles, 06902 Sophia-Antipolis, France<sup>b</sup>Institut de Neurosciences Cognitives de la Mediterranee, CNRS, Aix-Marseille University, UMR6193, 31 Chemin Joseph Aiguier, 13402 Marseille, France

## ARTICLE INFO

## Article history:

Available online xxx

## Keywords:

Biophysical model  
Cortical column  
Optical imaging  
Mesoscopic scale  
Voltage-sensitive dyes

## ABSTRACT

In this review, we present the voltage-sensitive dye imaging (VSDI) method. The possibility offered for *in vivo* (and *in vitro*) brain imaging is unprecedented in terms of spatial and temporal resolution. However, the unresolved multi-component origin of the optical signal encourages us to perform a detailed analysis of the method limitation and the existing models. We propose a biophysical model at a mesoscopic scale in order to understand and interpret this signal.

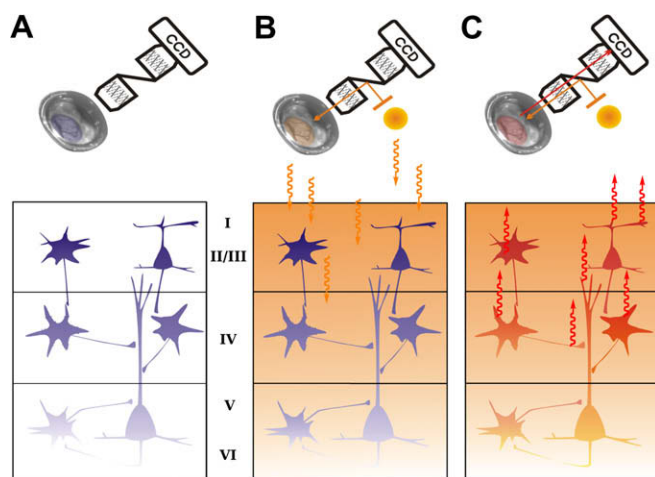
© 2009 Elsevier Ltd. All rights reserved.

## Contents

1. Introduction .....	00
2. VSDI for beginners .....	00
2.1. General principle .....	00
2.2. Optical imaging of neuronal population activity .....	00
2.2.1. General history .....	00
2.2.2. High spatial resolution for brain mapping .....	00
2.2.3. High temporal resolution unveils the dynamics of cortical processing .....	00
2.2.4. Functional connectivity reveals its dynamics .....	00
2.3. Conclusion .....	00
3. The multi-component origin of the optical signal .....	00
3.1. About the contribution from glial cells .....	00
3.2. About the contribution from excitatory versus inhibitory cells .....	00
3.3. About the contribution from the various compartments .....	00
3.4. About the contribution from cortical layers .....	00
3.5. About the contribution from thalamic versus horizontal connections .....	00
3.6. Conclusion .....	00
4. Benefits of modeling for optical signal analysis .....	00
4.1. Which scale for which model? .....	00
4.2. Mesoscopic scale: models of a cortical area .....	00
4.2.1. Extended LISSOM model .....	00
4.2.2. Neural field model of a cortical area .....	00
4.2.3. Conductance-based IAF neuronal network model .....	00
4.2.4. Linear model of the raw VSD signal .....	00
4.3. Biophysical model at the intermediate mesoscopic scale .....	00
4.3.1. Model specifications .....	00
4.3.2. Computation of the VSD signal .....	00
5. Conclusion .....	00
Acknowledgements .....	00
References .....	00

\* Corresponding author.

E-mail address: [sandrine.chemla@sophia.inria.fr](mailto:sandrine.chemla@sophia.inria.fr) (S. Chemla).



**Fig. 1.** VSDI principle in three steps. The imaging chamber allows a direct access of the primary visual cortex V1 represented as a patch of cortex with its six layers. (A) The dye, applied on the surface of the cortex, penetrates through the cortical layers of V1. (B) All neuronal and non-neuronal cells are now stained with the dye and when the cortex is illuminated, the dye molecules act as molecular transducers that transform changes in membrane potential into optical signals. (C) The fluorescent signal (red arrow) is recorded by a CCD camera.

## 1. Introduction

Optical imaging comes within the scope of new imaging techniques that allow us to visualize the functioning brain at both high spatial and temporal resolutions. Specifically, there are two techniques mostly used *in vivo* (see Grinvald et al. (1999) for a detailed review); the first is based on intrinsic signals, and the second is based on voltage-sensitive dyes (VSDs). In this review, we focus on the second technique, aiming at better understand the origin of the optical signal. Extensive reviews of VSDI have been published elsewhere (e.g. Grinvald et al., 2004; Roland, 2002). Although the underlying mechanism of this optical method is nowadays well understood, the recorded signal remains very complex and it seems difficult to isolate the contributions from its different components. This review suggests modeling as the appropriate solution. Few models of the VSD signal exist that help to understand the optical signal in terms of functional organization and dynamics of a population neural network. A closer interaction between VSDI experimentalists and modelers is desirable.

In the first part of this review, we give a general introduction to VSDI, followed by examples of applications to brain imaging. We compare *in vitro* and *in vivo* recordings obtained with VSDI in several animal studies. In a second part, we make the underlying limitations of this method explicit: what does the VSD signal measure? A question that is not completely answered in the literature. Finally, this review shows the benefit of brain activity modeling for optical signal analysis. Models of VSDI measures are reported. We both address what has already been done and what will be interesting to do in order to interpret the origins of the optical imaging signal.

## 2. VSDI for beginners

### 2.1. General principle

VSDI offers the possibility to visualize, in real time, the cortical activity of large neuronal populations with high spatial resolution (down to 20–50  $\mu\text{m}$ ) and high temporal resolution (down to the millisecond). With such resolutions, VSDI appears to be the best

technique to study the dynamics of cortical processing at neuronal population level.

This invasive technique is also called “extrinsic optical imaging” because of the use of voltage-sensitive dyes (Cohen et al., 1974; Ross et al., 1977; Waggoner and Grinvald, 1977; Gupta et al., 1981). After opening the skull and the dura mater of the animal, the dye molecules are applied on the surface of the cortex (Fig. 1A). They bind to the external surface of the membranes of all cells without interrupting their normal function and act as molecular transducers that transform changes in membrane potential into optical signals. More precisely, once excited with the appropriate wavelength (Fig. 1B), VSDs emit instantaneously an amount of fluorescent light that is function of changes in membrane potential, thus allowing for an excellent temporal resolution for neuronal activity imaging (Fig. 1C). The fluorescent signal is proportional to the membrane area of all stained elements under each measuring pixel.

“All elements” means all neuronal cells present in the cortex but also all non-neuronal cells, like glial cells (see Section 3.1 for more details). Moreover, neuronal cells include excitatory cells and inhibitory cells, whose morphology and intrinsic properties are quite different (see Salin and Bullier (1995) for a review on the different type of neurons and connections in the visual cortex). Furthermore, each cell has various compartments, including dendrites, somata and axons. The measured signal thus combine all these components, which are all likely to be stained in the same manner. The dye concentration is only depending on the depth of the cortex.

The fluorescent signal is then recorded by the camera of the optical video imaging device and displayed as dynamic sequences on computer (see Fig. 1). The submillisecond temporal resolution is reached by using ultra sensitive charge-coupled device (CCD) camera, whereas the spatial resolution is limited by optical scattering of the emitted fluorescence (Orbach and Cohen, 1983).

### 2.2. Optical imaging of neuronal population activity

#### 2.2.1. General history

The earliest optical recordings were made, at the single neuron level, both from cultured cells (Tasaki et al., 1968) and from various invertebrate preparations like ganglia of the leech (Salzberg et al., 1973), or the giant axon of the squid (Davila et al., 1973). For all other VSDI experiments, the VSD signal has a neuronal population resolution.

The VSDI method has then been used *in vitro* on brain slices, mainly in rodent and ferret. It allowed to optically record from the hippocampus (Grinvald et al., 1982), the visual cortex (Bolz et al., 1992; Albowitz and Kuhnt, 1993; Nelson and Katz, 1995; Yuste et al., 1997; Contreras and Llinas, 2001; Tucker and Katz, 2003a; Tucker and Katz, 2003b), the somatosensory cortex (Yuste et al., 1997; Antic et al., 1999; Contreras and Llinas, 2001; Petersen et al., 2001; Jin et al., 2002; Laaris and Keller, 2002; Berger et al., 2007) and from the auditory cortex (Jin et al., 2002; Kubota et al., 2006).

The salamander, largely used *in vitro* (Orbach and Cohen, 1983; Cinelli and Salzberg, 1992), was the first species also used *in vivo* for studying the olfactory system using VSDI (Orbach and Cohen, 1983), followed by the frog for the visual system (Grinvald et al., 1984), and the rodent for the somatosensory system. Indeed, initial *in vivo* studies of the somatosensory cortex have been made in anesthetized rodents, taking advantage of the thinness of the cortical dura (Orbach et al., 1985). More recently, VSDI in freely moving mice has also been performed with success (Ferezou et al., 2006).

Rodent and ferret were also used for studying the visual cortex *in vivo* (Roland et al., 2006; Lippert et al., 2007; Xu et al., 2007;



Ahmed et al., 2008). However, the main VSDI experiments on visual modality were conducted on two other mammalian species: cat and monkey (Grinvald et al., 1994; Arieli et al., 1995; Sterkin et al., 1998; Shoham et al., 1999; Sharon and Grinvald, 2002; Slovin et al., 2002; Seidemann et al., 2002; Jancke et al., 2004; Sharon et al., 2007; Benucci et al., 2007; Reynaud et al., 2007; Yang et al., 2007). Experiments on anesthetized cats are very attractive for mapping and studying the primary visual cortex, whereas monkey experiments also associate behavioral measures.

### 2.2.2. High spatial resolution for brain mapping

One domain of application of the VSDI, as other brain functional imaging, is brain mapping. Indeed, VSDI allows to build high-resolution functional maps, such as orientation or ocular-dominance maps (Shoham et al., 1999; Grinvald et al., 1999; Slovin et al., 2002; Sharon and Grinvald, 2002), as also obtained with optical imaging based on intrinsic signals (ISI) (Blasdel and Salama, 1986; Ts'o et al., 1990; Grinvald et al., 1991; Bonhoeffer and Grinvald, 1991; Hubener et al., 1997; Rubin and Katz, 1999). Comparison between the two imaging techniques (Shoham et al., 1999; Grinvald et al., 1999; Slovin et al., 2002) confirms the high spatial resolution of VSDI methodology for mapping the functional architecture of the visual cortex. However, although it is possible to do such brain mapping using VSDI, it does not take advantage of the possibility to inspect neuronal activation dynamics.

### 2.2.3. High temporal resolution unveils the dynamics of cortical processing

The main benefit of the VSDI technique is the possibility for neuroscientists to go further electrophysiological studies and low

resolution (either temporal or spatial) imaging techniques, since visualizing in real time with high spatial resolution large populations of neurons, while supplying information about cortical networks temporal dynamics. Many neuroscientists are motivated to investigate how a sensory stimulus is represented dynamically on the cortical surface in space and time (Grinvald et al., 1984; Grinvald et al., 1994; Arieli et al., 1996; Petersen et al., 2003; Civillico and Contreras, 2006). More precisely, the spatiotemporal dynamics of the response to simple stimuli, e.g. local drifting-oriented gratings or single whisker stimulation, have been visualized using VSDI on *in vivo* preparations (Cat: Sharon et al., 2007; Rodent: Petersen et al., 2003). Complex stimuli, e.g. the line motion or apparent motion illusions, have also been achieved using VSDI in the visual cortex of cats (Jancke et al., 2004) or ferrets (Ahmed et al., 2008), revealing fundamental principles of cortical processing *in vivo*. Nowadays, rapid and precise dynamic functional maps can even be obtained on behaving animals, as shown by Seidemann et al. (2002), Slovin et al. (2002) and Yang et al. (2007) on behaving monkeys, or by Ferezou et al. (2006) in freely moving mice.

These questions are conceivable thanks to the persistent development of novel dyes (Shoham et al., 1999; Grinvald et al., 2004; Kee et al., 2009). Indeed, the developed dyes allowed to monitor in real time neuronal activation both in *in vivo* and *in vitro* preparations (Arieli et al., 1996; Grinvald et al., 1999; Petersen et al., 2001; Petersen et al., 2003).

### 2.2.4. Functional connectivity reveals its dynamics

Combining the spatial and temporal advantages, an other direct application of VSDI is the possibility to study the functional con-

**Table 1**

Non-exhaustive list of publications related to VSDI, classified by experimental conditions (either *in vitro* or *in vivo*) and by species.

Conditions	Species	Related publications	Structure	Dye	$\lambda_{exc}$ (nm)	
<i>In vitro</i> (invertebrate preparations, cultured cells or brain slices)	Invertebrate (squid, skate, snail, leech)	Tasaki et al. (1968), Davila et al. (1973), Salzberg et al. (1973), Woolum and Strumwasser (1978), Gupta et al. (1981), Konnerth et al. (1987), Cinelli and Salzberg (1990), Antic and Zecevic (1995), and Zochowski et al. (2000)	Giant neurons	Styryl JPW1114 optimized for intracellular applications	540	
			Axons	JPW1114 (fluorescence)	520	
			Cerebellar parallel fibres	Pyrazo-oxonol RH482, RH155 (absorption)		
	Goldfish Salamander	Manis and Freeman (1988)	Optic tectum	Styryl RH414 (fluorescence)	540	
		Orbach and Cohen (1983) and Cinelli and Salzberg (1992)	Olfactory bulb	Merocyanine XVII optimized for absorption measurements (Ross et al., 1977; Gupta et al., 1981), RH414, RH155		
			Visual cortex	Fluorochrome Di-4-ANEPPS, RH414, Styryl RH795 (fluorescence)	500, 540	
Rodent	Grinvald et al. (1982), Bolz et al. (1992), Albowitz and Kuhnt (1993), Yuste et al. (1997), Antic et al. (1999), Petersen et al. (2001), Contreras and Llinas (2001), Laaris and Keller (2002), Jin et al. (2002), Kubota et al. (2006), Berger et al. (2007), Carlson and Coulter (2008), and Kee et al. (2009)	Barrel cortex	JPW2038, RH155, RH482, NK3630, JPW1114, RH414, RH795			
		Auditory cortex	RH795 for fluorescence, Oxonol NK3630 for absorption	520, 705		
		hippocampus	WW401	520		
	Ferret	Nelson and Katz (1995) and Tucker and Katz (2003a,b)	Visual cortex	RH461 (fluorescence)	590	
		Frog Salamander	Grinvald et al. (1984)	Visual cortex	Styryl RH414	520
			Orbach and Cohen (1983) and Kauer (1988)	Olfactory bulb	Styryl RH160 and RH414 optimized for fluorescence measurements (Grinvald et al., 1982)	510, 540
<i>In vivo</i> (anesthetized or awake)	Rodent	Orbach et al. (1985), Orbach and Van Essen (1993), Petersen et al. (2003), Derdikman et al. (2003), Civillico and Contreras (2006), Ferezou et al. (2006), Berger et al. (2007), Lippert et al. (2007), Xu et al. (2007), and Brown et al. (2009)	Barrel cortex	RH795, Oxonol RH1691, RH1692 and RH1838 optimized for <i>in vivo</i> fluorescent measurements (Shoham et al., 1999; Spors et al., 2002)	540, 630	
			Visual cortex	RH1691, RH1838	630	
			Visual cortex	RH795, RH1691	530, 630	
	Ferret	Roland et al. (2006) and Ahmed et al. (2008)	Visual cortex	RH795, RH1692	530-40, 630	
		Cat	Arieli et al. (1995), Sterkin et al. (1998), Shoham et al. (1999), Sharon and Grinvald (2002), Jancke et al. (2004), Sharon et al. (2007), and Benucci et al. (2007)	Visual cortex (area 17/18)		
	Grinvald et al. (1994), Shoham et al. (1999), Slovin et al. (2002), Seidemann et al. (2002), Reynaud et al. (2007), and Yang et al. (2007)		Visual cortex (V1/V2)	RH1691, RH1692, RH1838	630	
	Monkey	Grinvald et al. (1994), Shoham et al. (1999), Slovin et al. (2002), Seidemann et al. (2002), Reynaud et al. (2007), and Yang et al. (2007)	FEF	RH1691	630	

nectivity of neuronal populations. Yuste et al. (1997) for example, investigated the connectivity diagram of rat visual cortex using VSDI. Vertical and horizontal connections have been detected. More generally, intracortical and intercortical interactions, occurring during sensory processing (especially visual), have been largely explored using VSDI, either *in vitro* or *in vivo*: Mapping functional connections using VSDI, has been done *in vitro* in the rat visual cortex (Bolz et al., 1992; Carlson and Coulter, 2008), in the guinea pig visual cortex (Albowitz and Kuhnt, 1993) and in the ferret visual cortex (Nelson and Katz, 1995; Tucker and Katz, 2003a; Tucker and Katz, 2003b), providing not only functional, but also anatomical and physiological information on the local network. For example, Tucker and Katz (2003a) investigated with VSDI how neurons in layer 2/3 of ferret visual cortex integrate convergent horizontal connections.

Orbach and Van Essen (1993) used VSDI in the visual system of the rat *in vivo* to map striate and extrastriate pathways. Feedforward propagating waves from V1 to other cortical areas, and feedback waves from V2 to V1 have been recently reported by Xu et al. (2007), thanks to VSDI. In addition, feedback depolarization waves (from areas 21 and 19 toward areas 18 and 17) were extensively studied by Roland et al. (2006) in ferrets after staining the visual cortex with VSD.

### 2.3. Conclusion

By adding a new dimension to existing brain functional imaging techniques, VSDI directly reports the spatiotemporal dynamics of neuronal populations activity. Many VSDI studies have then been conducted in order to investigate the spatiotemporal patterns of activity occurring in different parts of the CNS, *in vitro* or *in vivo*, on several preparations or animal species. The Table 1 lists most articles presenting experimental results using VSDI techniques. The publications are first classified by the condition of the experiment, either *in vitro* or *in vivo*, and then by the experimental preparations or animal species. Additional information about dyes is available in the last columns (see Ebner and Chen (1995) for a compilation of the commonly used dyes and their properties).

## 3. The multi-component origin of the optical signal

### 3.1. About the contribution from glial cells

In general, glial cells have been neglected by neuroscientists for a long time, especially because unlike neurons, they do not carry action potentials. However, glial cells have important functions (see Cameron and Rakic (1991) for a review) and they may contribute to the VSD signal.

Glial cells are known as the “supporting cells” of the CNS and are estimated to outnumber neurons by as much as 50–1. However, their role in information representation or processing remains unresolved. Indeed, *in vitro* studies have shown increasing evidence for an active role of astrocytes in brain function. However, little is known about the behavior of astrocytes *in vivo*.

When interpreting the VSD signal, we face two conflicting viewpoints. Konnerth and Orkand (1986), Lev-Ram and Grinvald (1986), Konnerth et al. (1987), Konnerth et al. (1988) and Manis and Freeman (1988) showed that the optical signal has two components: a “fast” followed by a “slow” signal. The latter has been revealed by doing successive staining with different dyes (e.g. RH482 and RH155), since each of them may preferentially stain different neuronal membranes. The authors then present evidence that this slow signal has a glial origin.

However, Kelly and Van Essen (1974) showed that the glial responses are weak (depolarizations of only 1–7 mV in response to

visual stimuli) and have a time scale of seconds. Recent paper of Schummers et al. (2008) confirms that the astrocyte response is delayed 3–4 s from stimulus onset, which is a very slow temporal response compared to neuron response. Generally, in VSDI, only the first 1000 ms are considered, since intrinsic activity may affect the signal after this time.

We understand here that the controversy about glial contribution is directly link to the used dye (Ebner and Chen, 1995), and the time course of the optical signal generated. Thus, glial activity is very unlikely to participate significantly to the VSD signal (when considering recent fast dyes), since the amplitude of glial response is weak and its time course is very slow.

### 3.2. About the contribution from excitatory versus inhibitory cells

In the neocortex, neurons (despite their morphologic diversity) can be functionally classified in two groups: excitatory neurons, which represent about 80% of the cortical cells, and inhibitory neurons which represent about 20% of cortical cells (Douglas and Martin, 1990). Thus, it is tempting to say that the VSD signal mainly reflects the activity of excitatory neurons (Grinvald et al., 1999).

However, the VSD signal is proportional to changes in membrane potential. Thus, both excitatory and inhibitory neurons contribute positively to the VSD signal and it is hard to tease apart contributions from excitatory or inhibitory cells. An additional level of complexity arises from the fact that inhibition operates generally in a shunting “silent” mode (Borg-Graham et al., 1998). In this mode, inhibition suppresses synaptic excitation without hyperpolarizing the membrane potential.

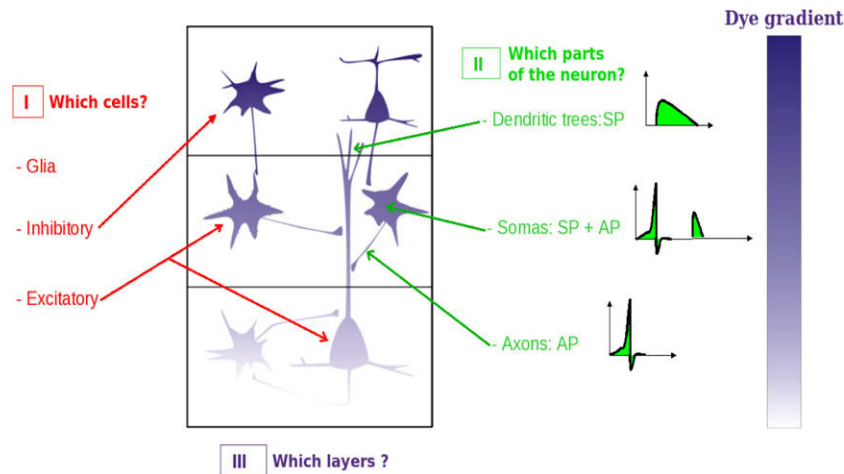
To conclude, the contribution of inhibitory cells to the VSD signal is unclear and would obviously benefit from modeling studies.

### 3.3. About the contribution from the various compartments

Neurons can be also decomposed into their main various compartments, whose surface and electrical activity are different (see Fig. 2, green part):

- (a) The soma, whose electrical activity can be either synaptic (SP for synaptic potential) or spiking (AP for action potential).
- (b) The dendrites, that integrate presynaptic AP information from others cells. The electrical activity is mainly synaptic, however, back-propagating AP could be recorded in the dendrites (see Waters et al. (2005) for a review). Dendritic surface area of mammalian neurons have been estimated by Sholl (1955a), Aitken (1955), and Young (1958) to be 10–12 times larger than cell bodies surface area, and to represents 90% of the total neuronal cell membrane (Eberwine, 2001).
- (c) The axon, which carries spiking signals from the soma to the axon terminal. Spiking activity can be recorded on this part of neuron. In contrast with dendrites, the surface area of axons represents 1% of the total neuronal cell surface (Eberwine, 2001).

In the literature, regarding the difference in membrane areas of the various neuronal components and the nature of the signal, it is commonly accepted that the optical signal, in a given pixel, mostly originates from the dendrites of cortical cells, and therefore, mainly reflects dendritic post-synaptic activity (Orbach et al., 1985; Grinvald et al., 2004). Extensive comparisons between intracellular recordings from a single neuron and VSDI also showed that the optical signal correlates closely with synaptic membrane potential changes (Petersen et al., 2003; Contreras and Llinas, 2001). However, no real quantitative analysis has been performed to date



**Fig. 2.** Contributions of the optical signal. Once neurons are stained by the VSD, every neuronal membrane contributes to the resulting fluorescent signal, but from where? and in which proportion? Answering these four questions could clarify the optical signal origins: (1) Which cells? (2) Which parts of the cell? (3) Which layers? (4) Which presynaptic origins?

and it is more correct to state that the optical signal is multi-component since the VSD signal reflects the summed intracellular membrane potential changes of all neuronal compartments at a given cortical site. The aim then, is to determine the exact contribution of each component, which remains unknown. More precisely, what is quantitatively the contribution of dendritic activity? Can spiking activity be neglected?

#### 3.4. About the contribution from cortical layers

The depth of the neocortex is about 2 mm. It is made up of six horizontal layers principally segregated by cell types and neuronal connections. The layer II mostly contains small pyramidal neurons that make strong connections with large pyramidal neurons of the layer V (Thomson and Morris, 2002).

Improved dyes, when put at the surface of the exposed cortex, can reach a depth of about 400–800  $\mu\text{m}$  from the cortical surface, which mainly corresponds to superficial layers (Grinvald et al., 1999; Petersen et al., 2003). Furthermore, measures of the distribution of dye fluorescence intensity in rat visual and barrel cortex confirm that the optical signal mostly originates from superficial layers I–III (Ferezou et al., 2006; Lippert et al., 2007). Note that Lippert et al. (2007) used a special staining procedure, i.e. keeping the dura mater intact, but dried.

However, they did not take into account the fact that the activity in superficial layers could arise from neurons in deep layers, due to their dendritic arborization. Indeed, large pyramidal neurons in layer V have apical dendrites that reach superficial layers and may contribute to the signal. Therefore, the exact contribution of each cortical layer still has to be clarified.

#### 3.5. About the contribution from thalamic versus horizontal connections

The origin of the signal can also be problematic when looking at the contribution from the different presynaptic activity origins, e.g. direct thalamic synaptic inputs, or horizontal inputs. Indeed, in response to a local stimulation, slow propagating waves can be recorded (Grinvald et al., 1994; Jancke et al., 2004; Roland et al., 2006; Xu et al., 2007; Benucci et al., 2007). We can question what is the relative contribution of all the synaptic input sources of this phenomenon, i.e. feedforward, horizontal or feedback inputs. Dedicated models could help teasing apart those various contributions.

#### 3.6. Conclusion

Fig. 2 summarizes the four main questions not completely clarified to date:

- What are the contributions of the various neurons and neuronal components to the optical signal?
- What is the ratio between spiking and synaptic activity?
- What are the respective contributions of cells from deep versus superficial layers?
- What is the origin of the synaptic input? More precisely, what are the respective contributions of thalamic, local and long-range inputs?

To answer these questions, a possibility is to develop computational models in order to reproduce and analyse VSD signals. Models of VSD signals are reported in the next and last part of this review.

### 4. Benefits of modeling for optical signal analysis

The goal of this section is to investigate the different models from the literature, used to reproduce and analyse the VSD signal. We quickly emphasize three of these models because of their scale of analysis. In the last subsection, we present, in detail, an intermediate model that would allow to answer the previous questions about the VSD signal contributions.

#### 4.1. Which scale for which model?

As previously described, the origin of the VSD signal is complex and remains to be estimated and explored. Therefore, it could be interesting to see if the activity of a computational model could be related to this signal. However, the choice of the model's scale is very important and depends on what exactly the model is designed for. We propose in the following paragraphs that the mesoscopic scale seems to be the best scale for analyzing the population VSD signal. In neuroscience, this scale is generally used to define the elementary processing unit in the brain, the cortical column. We start by defining our concept of cortical columns.

Since the 1950s, thanks to the work of Mountcastle (1957), we know that the cerebral cortex has a columnar organization. In 1960s and 1970s, Hubel and Wiesel (1962, 1965, 1977) followed



Mountcastle's discoveries by showing that ocular dominance and orientations are organized in a columnar manner in cat and monkey visual cortex. Today, the notion of cortical column becomes a large controversy since the original concept is expanding, year after year, discovery after discovery, to embrace a variety of different structures, principles and names. A 'column' now refers to cells in any vertical cluster that share the same tuning for any given receptive field attribute (see Horton and Adams (2005) for a detailed review on the cortical column concept). A novel and useful concept is to propose that each definition of cortical column depends on its type (anatomical, functional, physico-functional) and its spatial scale, as detailed in Table 2. A minicolumn or a microcolumn is an anatomical column of about 100 neurons, since its spatial scale is about 40  $\mu\text{m}$ . Next, orientation or ocular dominance columns are classified as functional columns whose the spatial scale is between 200 and 300  $\mu\text{m}$ , containing several minicolumns. An hypercolumn in V1 or a macrocolumn in the general case, then represents a physico-functional unit containing a full set of values for any given functional parameter. Its spatial scale can be up to 600  $\mu\text{m}$  and contains about  $10^4$  neurons. Finally, neural mass is a mesoscopic concept which depends on the spatial scale. When looking at a cortical area, it can be used to represent, for example, all the pyramidal neurons contained in it (about  $10^5$  neurons).

Into these definitions and in order to reproduce exactly the same signal, i.e. time course and spatial extent, it seems appropriate to construct models at a large mesoscopic scale which could represent an entire cortical area. Models from Miikkulainen et al. (2005), Grimbert et al. (2007), Rangan et al. (2005) and La Rota (2003) consider this scale, that can be view as the neurons population scale.

An other point of view is to choose a much finer scale allowing to construct a more detailed biophysical model in order to quantitatively estimate the exact contribution of the VSD signal (excitation vs. inhibition, parts of the neuron, layers participation, etc.). In optical imaging, the visual scale studied, which is about 50  $\mu\text{m}$ , corresponds to one pixel. It is still a population activity since it represents about 200 neurons, but the scale being relatively small, we will call it "intermediate mesoscopic scale". This model is detailed in the last section.

## 4.2. Mesoscopic scale: models of a cortical area

### 4.2.1. Extended LISSOM model

The Laterally Interconnected Synergetically Self-Organizing Map (LISSOM) family of models was developed by Bednar, Choe, Miikkulainen and Sirosh, at the University of Texas (Miikkulainen et al., 2005; Sirosh and Miikkulainen, 1994), as models of human visual cortex at a neural column level. It is based on the Self-Organizing Maps (SOM) algorithm (from Kohonen, 2001) used to visualize and interpret large high-dimensional data sets. When extended, the LISSOM neural network models takes into account lateral interactions (excitatory and inhibitory connections), allowing to reproduce the pinwheel organization of the primary visual cortex map, such as orientation, motion direction selectivity and ocular-dominance maps.

**Table 2**

The different types of cortical columns.

	Anatomical	Ol pixel	Functional	Physico-functional	Cortical area
Type of cortical column	Microcolumn or minicolumn	<i>Our column</i>	Orientation, ocular dominance column	Macrocolumn or hypercolumn (V1)	Neural mass
Spatial scale	40–50 $\mu\text{m}$	50–100 $\mu\text{m}$	200–300 $\mu\text{m}$	600 $\mu\text{m}$ (and more)	10 mm
Number of neurons	80–100 neurons	200 neurons	Several minicolumns	60–100 minicolumns or 10,000 neurons	100X Thousand neurons of the same type (pyr, stellate, etc.)

Sit and Miikkulainen used such a LISSOM model to represent V1 and tried to show how the activity of such a computational model of V1 can be related to the VSD signal (Sit and Miikkulainen, 2007). Indeed, with an extended LISSOM model including propagation delays in the cortical connections, they showed that the orientation tuning curve and the response dynamics of the model were similar to those measured with VSDI.

The model is a couple of two layers of neural units that represent the retina and V1. In V1, neural units account for a whole vertical column of cells. They receive input from the retina and also from neighbour columns (short-rang lateral excitatory and long-rang lateral inhibitory connections). Thus, the neuronal activity of unit  $\mathbf{r}$  in V1 writes:

$$A(\mathbf{r}, t) = \sigma(V(\mathbf{r}, t)),$$

$$V(\mathbf{r}, t) = \sum_{\rho} \gamma_{\rho} \sum_{\mathbf{r}'} W_{\rho, \mathbf{r}, \mathbf{r}'} A(\mathbf{r}', t - d(\mathbf{r}, \mathbf{r}')) + \sum_{\mathbf{s}} \chi_{\mathbf{s}} R_{\mathbf{s}, \mathbf{r}}, \quad (1)$$

where  $\sigma$  is a sigmoid activation function and the two terms are respectively the weighted sum of the lateral activations and the input activation from the retina.  $W_{\rho, \mathbf{r}, \mathbf{r}'}$  and  $R_{\mathbf{s}, \mathbf{r}}$  are respectively the synaptic weights matrix of lateral and retinal connections, and  $d(\mathbf{r}, \mathbf{r}')$  is the delay function between unit  $\mathbf{r}$  and unit  $\mathbf{r}'$ . This is thus a scalar model of the neural activity.

Then, the computation of the VSD signal is done by looking only at the subthreshold activity  $V(\mathbf{r}, t)$ , given by the weighted sum of presynaptic activity. To simplify, the authors have extended the LISSOM model with delayed lateral connections to compute the VSD signal from subthreshold signal. This is thus a scalar linear model of the VSD signal built on convolutions.

This model, based on Hebbian self-organizing mechanisms, is simple and efficient to replicate the detailed development of the primary visual cortex. It is thus very useful to study VSDI functional maps. However, this model is not specific enough to answer the previous asked questions (see Section 3.6).

### 4.2.2. Neural field model of a cortical area

Another approach, introduced by Grimbert et al. (2007) and Grimbert et al. (2008), proposes neural fields as a suitable mesoscopic models of cortical areas, in link with VSD. Neural field are continuous networks of interacting neural masses, describing the dynamics of the cortical tissue at the population level (Wilson and Cowan, 1972; Wilson et al., 1973). It could thus be applied to solve the direct problem of the VSD signal, providing the right parameters. More precisely, the authors showed that neural fields can easily integrate the biological knowledge of cortical structure, especially horizontal and vertical connectivity patterns. Hence, they proposed a biophysical formula to compute the VSD signal in terms of the activity of a field.

The classical neural field model equation is used, either written in terms of membrane potential or in terms of activity of the different neural masses present in a cortical column. For example, if  $\mathbf{r}$  represents one spatial position of the spatial domain defining the area, then the underlying cortical column is described, at time  $t$ , by either a vector  $\mathbf{V}(\mathbf{r}, t)$  or  $\mathbf{A}(\mathbf{r}, t)$ :

$$\dot{\mathbf{V}}(\mathbf{r}, t) = -\mathbf{L}\mathbf{V}(\mathbf{r}, t) + \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{S}(\mathbf{V}(\mathbf{r}', t)) d\mathbf{r}' + \mathbf{I}_{ext}(\mathbf{r}, t), \quad (2)$$

and

$$\dot{\mathbf{A}}(\mathbf{r}, t) = -\mathbf{L}\mathbf{A}(\mathbf{r}, t) + \mathbf{S}\left(\int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{A}(\mathbf{r}', t) d\mathbf{r}' + \mathbf{I}_{ext}(\mathbf{r}, t)\right) \quad (3)$$

Here,  $\mathbf{V}(\mathbf{r}, t)$  contains the average soma membrane potentials of the different neural masses present in the column (the vector's dimension then represents the number of neuronal types considered in every column).  $\mathbf{A}(\mathbf{r}, t)$  contains the average activities of the masses. For example,  $A_i$  is the potential quantity of post-synaptic potential induced by mass  $i$  on the dendrites of all its post-synaptic partners. The actual quantity depends on the strength and sign (excitatory or inhibitory) of the projections (see Grimbert et al. (2007, 2008) and Faugeras et al. (2008) for more details on the model's equations). The model include horizontal intercolumnar connections and also vertical intracolumnar connections between neural masses. The latter gives an advantage to this model compared to the previous one, since the vertical connectivity was not taken into account in the extended LISSOM model. Furthermore, extracortical connectivity is not made explicit here, though taken into account in Grimbert et al. (2007).

Hence, based on this biophysical formalism (and especially the activity-based model, which is more adapted than the voltage-based model), the authors propose a formula involving the variables and parameters of a neural field model to compute the VSD signal:

$$OI(\mathbf{r}, t) = \sum_{j=1}^N \int_{\Omega} \tilde{w}_j(\mathbf{r}, \mathbf{r}') A_j(\mathbf{r}', t) d\mathbf{r}', \quad (4)$$

where  $\tilde{w}_j(\mathbf{r}, \mathbf{r}')$  contains all the biophysical parameters accounting for a cortical area structure stained by a voltage-sensitive dye, i.e. the different layers, the number of neurons, the number of dye molecules per membrane surface unit, the attenuation coefficient of light and also the horizontal and vertical distribution patterns of intra and intercortical connectivities.

This formula is the result of many decompositions of the total optical signal, from layer level to cellular membrane level, where the signal is simply proportional to the membrane potential.

Better than the Lissom model for our considerations, this large-scale model reproduces the spatiotemporal interactions of a cortical area in response to complex stimuli, e.g. line motion illusion, and allows, on average, to answer at the mesoscopic scale some previous questions (see Section 3.6). However, improvements on parameters tuning are still needed.

#### 4.2.3. Conductance-based IAF neuronal network model

Another large-scale computational model of the primary visual cortex have been proposed by Rangan et al. (2005). The model is a two-dimensional patch of cortex, containing about  $10^6$  neurons with a preferred orientation, whose 80% are excitatory and 20% are inhibitory. The dynamics of single cell  $i$  is described by a single compartment, conductance-based, exponential integrate-and-fire equation (see Geisler et al. (2005) for more details on this neuron model). The derivation of this equation gives the membrane potential of neuron  $i$  of spatial position  $\mathbf{r}_i$ :

$$V(\mathbf{r}_i, t) = \frac{g^L V^L + (g_i^A(t) + g_i^N(t)) V^E + g_i^G(t) V^I}{g^L + g_i^A(t) + g_i^N(t) + g_i^G(t)} \quad (5)$$

where  $g^L$ ,  $g_i^A$ ,  $g_i^N$  and  $g_i^G$  are respectively leak, AMPA, NMDA and GABA conductances, and  $V^L$ ,  $V^E$  and  $V^I$  are respectively leak, excitatory and inhibitory reversal potentials.

The authors then use  $V(\mathbf{r}, t)$  to represent the VSD signal, i.e. the subthreshold dendritic activity in the superficial layers of the cor-

tex. Poisson processes are used to simulate inputs from the thalamus and background noise.

This model allows, like the previous one Grimbert et al. (2007), to reproduce the spatiotemporal activity patterns of V1, as revealed by VSDI, in response to complex stimuli, e.g. the line motion illusion. However, in comparison with Grimbert et al. (2007), no laminar structure is taken into account.

#### 4.2.4. Linear model of the raw VSD signal

With the same scale of analysis, La Rota (2003) presented an interesting linear model in order to study the neural sources of the mesoscopic VSD signal. The author chose a compromise between a detailed and a "black-box" model of the signal, by taking into account the important properties of the VSD signal and also the artefacts directly linked to its measure, in a mesoscopic, linear and additive model. The VSD signal of a cortical area can then be modeled by an intrinsic and an extrinsic components:

$$OI(t) = A(t) + \rho(t), \quad (6)$$

where  $A(t)$  represents the activity of the intrinsic component of the optical signal (i.e. the synaptic activity of the cortical area observed) and  $\rho(t)$  represents all the noise and artefacts due to the measure (e.g. hemodynamic artefact, cardiovascular and respiratory movements, instrumental noise, etc.). In this model, inputs from the thalamus are considered as background noise and thus enter in the  $\rho$  component.

The model is interesting because it both takes into account the intrinsic and the extrinsic variability of the VSD signal. The latter being supposed already removed, when analyzing the signal in the three other presented models.

### 4.3. Biophysical model at the intermediate mesoscopic scale

Since none of the previous models was specific enough to determine the different contributions of the optical signal, a biological cortical column model, at an intermediate mesoscopic scale, has also been proposed in order to better understand and interpret biological sources of VSD signals (Chemla et al., 2007). This scale corresponds to one pixel of optical imaging: about  $50 \mu\text{m}$  and the related model solves the direct VSD problem, i.e. generates a VSD signal, given the neural substrate parameters and activities. Using a detailed compartmental model allows to push the state of the art at this level. This model confirms and quantifies the fact that the VSD signal is the result of an average from multiple components.

#### 4.3.1. Model specifications

Into the above cortical columns paradigm and for our specific model, we introduced a new distinction of a cortical column (see Fig. 2, second column). The spatial scale is about  $50 \mu\text{m}$ , corresponding to one pixel of optical imaging. Given this spatial scale, the number of neurons, that has been evaluated from Binzegger et al. (2004), is about 200.

We then consider a class of models based on a cortical microcircuit (see Raizada and Grossberg (2003), Douglas and Martin (2004), and Haeusler et al. (2007) for more details on this concept), whose synaptic connections are made only between six specific populations of neurons: two populations (excitatory and inhibitory) for three main layers (2/3, 4, 5/6).

Each neuron is represented by a reduced compartmental description (see Bush and Sejnowski (1993) for more details on the reduction method) with conductance-based Hodgkin-Huxley neuron model (see Hodgkin and Huxley, 1952) in the soma and the axon. Thus, the dynamics of single cells are described by the following equation:

$$C_m \frac{dV}{dt} = I_{ext} - \sum_i g^i(V)(V - V^i) \quad (7)$$

where  $V$  is the membrane potential,  $I_{ext}$  is an external current injected into the neuron,  $C_m$  is the membrane capacitance, and where three types of current are represented: leak, potassium and sodium conductances or respectively  $G_L$ ,  $G_K$  and  $G_{Na}$ .  $G_L$  is independent of  $V$  and determines the passive properties of the cells near resting potential. The sodium and potassium conductances are responsible for the spike generation. Furthermore, a slow potassium conductance was included in the dynamics of the excitatory population to reproduce the observed adaptation of the spike trains emitted by these neurons (see Nowak et al., 2003). This feature seems to be absent in inhibitory neurons, as taken into account in this model.

Only passive dendrites were considered. Each neuron represented with seven to nine compartments. The link between compartments can then be described by Eq. (8) (Hines and Carnevale, 1997).

$$C_j \frac{dV_j}{dt} + I_{ion_j} = \sum_k \frac{V_k - V_j}{R_{jk}} \quad (8)$$

where  $V_j$  is the membrane potential in compartment  $j$ ,  $I_{ion_j}$  is the net transmembrane ionic current in compartment  $j$ ,  $C_j$  is the membrane capacitance of compartment  $j$  and  $R_{jk}$  is the axial resistance between the centers of compartment  $j$  and adjacent compartment  $k$ .

Synaptic inputs are modeled as conductance changes. Excitatory AMPA synapses are converging on soma and dendrites of each neuron, whereas inhibitory GABA synapses are only converging on soma of each neuron (Salin and Bullier, 1995). The number of synapses involved in the projections between these different neuronal types, including the afferent from the LGN, were recalculated for a 50  $\mu\text{m}$  cortical column, based on Binzegger et al. (2004) for the considered layers, while latencies have been introduced for each connection following Thomson et al. (2007).

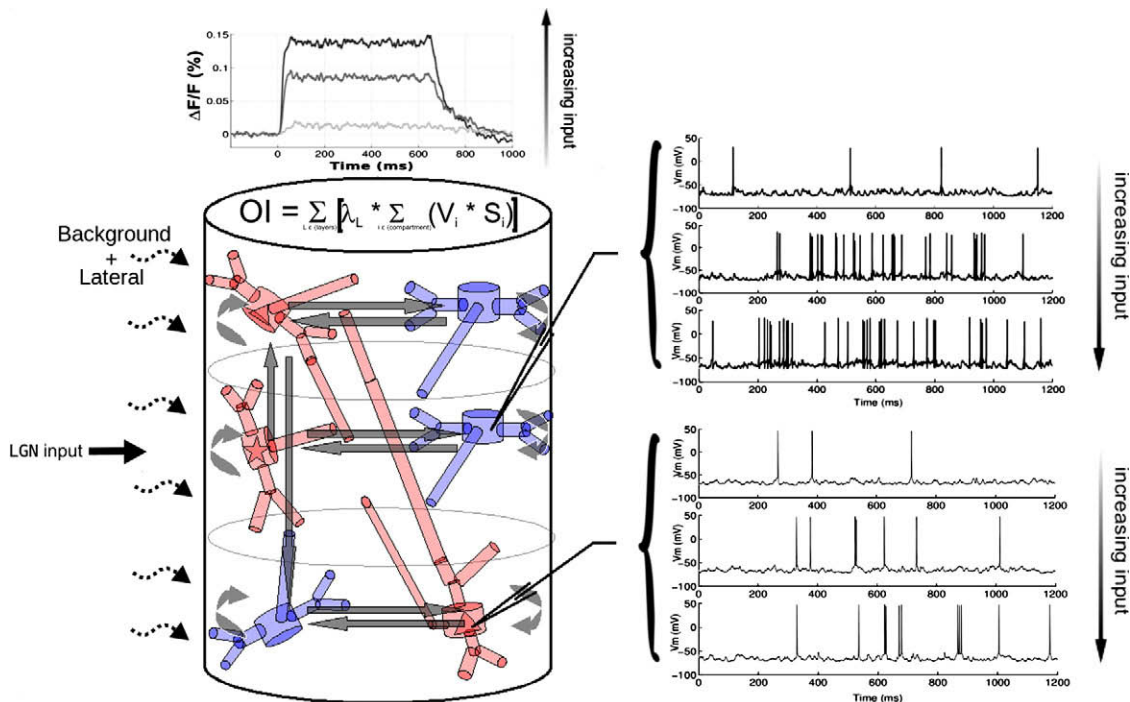
Input signals from the thalamus into the neocortex layer IV was simulated by applying random spike trains to each neuron in layer IV and random latency have been introduced for each input connection to simulate the temporal properties of geniculocortical pathway. Then we increased the frequency of the spike trains in order to represent stimulus contrast and see how the model transforms an increasing input, i.e. the contrast response function (see Albrecht et al., 1982). At this point, the column is isolated. A step further, the conditions relative to a larger network are reproduced as follows: First, "background noise" was introduced in each neuron of the column. Typically, noise can be introduced in the form of stochastic fluctuation of a current or an ionic conductance. The stochastic model of Destexhe et al. (2001), containing two fluctuating conductances, is used here, allowing us to simulate synaptic background activity similar to in vivo measurements, for a large network. Second, lateral connections between two neighboring columns are reproduced by introducing another set of random spike trains inputs whose frequency, synaptic delays and synaptic weights are adapted for fitting experimental data. Fig. 3 shows a schematic of the model, with thalamic input, background activity and lateral interactions. Examples of neuronal response have been plotted in function of increasing input or contrast.

#### 4.3.2. Computation of the VSD signal

The VSD signal is simulated using a linear integration on the membrane surface of neuronal components. Here, the use of compartmental model has a real interest. Indeed, the computation of the VSD signal, for a given layer  $L$ , is given by:

$$OI^L = \lambda^L \sum_{i \in \{\text{Compartments}\}} V_i S_i \quad (9)$$

where  $S_i$  and  $V_i$  are respectively the surface and the membrane potential of the  $i$ th compartment and  $\lambda^L$  represents the fluorescence's gradient or the illumination intensity of the dye in layer  $L$ .



**Fig. 3.** Model representation taking into account thalamic input contrast, background activity and lateral connections and offering the possibility to compute the VSD signal with a linear formula. In output, inhibitory and excitatory neuronal responses are plotted in function of increasing input or contrast (right inset). The temporal evolution of the VSD signal in response to 600 ms stimuli and in function of increasing input is also emphasized.



Thus, this model takes into account soma, axon and dendrites influences, introduces 3D geometrical properties (dendrites of large pyramidal neurons in layer 5 can reach superficial layers) and fluorescence gradient depending on depth. According to Lippert et al. (2007) and Petersen et al. (2003),  $\lambda^2 = 0.95$ ,  $\lambda^4 = 0.05$  and  $\lambda^5 = 0$ . Then, the total optical imaging signal is given by the following formula:

$$OI = \sum_{L \in \{\text{Layers}\}} OI^L \quad (10)$$

Following this framework, the VSD signal is simulated in response to known stimuli (Fig. 3, bottom right inset) and compared to experimental results (Chemla et al., 2008).

Thanks to its compartmental construction, this model can predict the different contributions of the VSD signal. It thus gives the possibility to quantitatively answer the previous asked questions: excitation vs. inhibition, spiking vs. synaptic activity and superficial vs. deep layers: The model confirms that the VSD signal mainly reflects dendritic activity (75%) of excitatory neurons (80%) in superficial layers (80%). However, these numbers are changing when increasing the level of input activity. At high level of activity, inhibitory cells, spiking activity and deep layers become non-negligible, and should be taken into account in the computation of the VSD signal. These results will be the subject of a future publication.

## 5. Conclusion

In this review, we have presented the voltage-sensitive dyes imaging (VSDI) technique in a general and elementary manner. This optical technique, thanks to its excellent spatial and temporal resolution, offers many possibilities for *in vitro* and more interestingly *in vivo* brain imaging.

However, the recorded optical signal is multi-component and its origins are still unresolved. Indeed, the contribution of each component, i.e. glial cells, excitatory cells, inhibitory cells, somas, axons, dendrites, layers, is very difficult to isolate from the others.

This review suggests modeling as the appropriate solution. We reported four existing models that try to reproduce and analyse the VSD signal. The main advantage of these models, all built at a mesoscopic scale, is the ability to compare the same signal, i.e. the signal of an entire cortical area. However, for our considerations, i.e. find the different contributions of the VSD signal, those models have not the right scale. Thus, we proposed a biophysical cortical column model, at an intermediate mesoscopic scale, in order to find the biological sources of the VSD signal. Using a such compartmental model should be of great value for doing a quantitative analysis of the different contributions of the optical signal.

## Acknowledgements

The authors are grateful to Thierry Vieville, from the CORTEX Lab., INRIA, Sophia-Antipolis, for his active participation in the model development and his help in writing this review. They also thank Francois Grimbart, at Northwestern University in the Department of Neurobiology and Physiology, for helpful discussions on the subject. Partially supported by the EC IP Project FP6015879, FACETS and the MACCAC ARC projects.

## References

Ahmed, B., Hanazawa, A., Undeman, C., Eriksson, D., Valentiniene, S., Roland, P.E., 2008. Cortical dynamics subserving visual apparent motion. *Cereb. Cortex* 18 (12), 2796–2810.  
Aitken, J.T., 1955. Observations on the larger anterior horn cells in the lumbar region of the cat's spinal cord. *J. Anat.* 89, 571.

Albowitz, B., Kuhnt, U., 1993. The contribution of intracortical connections to horizontal spread of activity in the neocortex as revealed by voltage sensitive dyes and a fast optical recording method. *Eur. J. Neurosci.* 5 (10), 1349–1359.  
Albrecht, D., Hamilton, D., 1982. Striate cortex of monkey and cat: contrast response function. *J. Neurophysiol.* 48 (1), 217–233.  
Antic, S., Zecevic, D., 1995. Optical signals from neurons with internally applied voltage-sensitive dyes. *J. Neurosci.* 15 (2), 1392–1405.  
Antic, S., Major, G., Zecevic, D., 1999. Fast optical recordings of membrane potential changes from dendrites of pyramidal neurons. *J. Neurophysiol.* 82 (3), 1615–1621.  
Arieli, A., Shoham, D., Hildesheim, R., Grinvald, A., 1995. Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. *J. Neurophysiol.* 73 (5), 2072–2093.  
Arieli, A., Sterkin, A., Grinvald, A., Auerster, A., 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273 (5283), 1868–1871.  
Benucci, A., Robert, A.F., Carandini, M., 2007. Standing waves and traveling waves distinguish two circuits in visual cortex. *Neuron* 55 (1), 103–117.  
Berger, T., Borgdorff, A., Crochet, S., Neubauer, F.B., Lefort, S., Fauvet, B., Ferezou, I., Carleton, A., Luscher, H.R., Petersen, C.C., 2007. Combined voltage and calcium epifluorescence imaging *in vitro* and *in vivo* reveals subthreshold and suprathreshold dynamics of mouse barrel cortex. *J. Neurophysiol.* 97 (5), 3751–3762.  
Binzegger, T., Douglas, R., Martin, K., 2004. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* 24 (39), 8441–8453.  
Blasdel, G.G., Salama, G., 1986. Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature* 321 (6070), 579–585.  
Bolz, J., Novak, N., Staiger, V., 1992. Formation of specific afferent connections in organotypic slice cultures from rat visual cortex cocultured with lateral geniculate nucleus. *J. Neurosci.* 12 (8), 3054–3070.  
Bonhoeffer, T., Grinvald, A., 1991. Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature* 353 (6343), 429–431.  
Borg-Graham, L., Monier, C., Fregnac, Y., 1998. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* 393, 369–373.  
Brown, C.E., Aminoltejeri, K., Erb, H., Winship, I.R., Murphy, T.H., 2009. *In vivo* voltage-sensitive dye imaging in adult mice reveals that somatosensory maps lost to stroke are replaced over weeks by new structural and functional circuits with prolonged modes of activation within both the peri-infarct zone and distant sites. *J. Neurosci.* 29 (6), 1719–1734.  
Bush, P., Sejnowski, T., 1993. Reduced compartmental models of neocortical pyramidal cells. *J. Neurosci. Methods* 46, 159–166.  
Cameron, R.S., Rakic, P., 1991. Glial cell lineage in the cerebral cortex: a review and synthesis. *Glia* 4, 124–137.  
Carlson, G.C., Coulter, D.A., 2008. *In vitro* functional imaging in brain slices using fast voltage-sensitive dye imaging combined with whole-cell patch recording. *Nat. Protocols* 3 (2), 249–255.  
Chemla, S., Chavane, F., Vieville, T., Kornprobst, P., 2007. Biophysical cortical column model for optical signal analysis. In: *Cns07*. <<http://www.cnsorg.org>>.  
Chemla, S., Vieville, T., Chavane, F., 2008. Biophysical cortical column model for optical signal analysis. In: *Areadne08*. <<http://www.aredne.org/>>.  
Cinelli, A.R., Salzberg, B.M., 1990. Multiple site optical recording of transmembrane voltage (msorty), single-unit recordings, and evoked field potentials from the olfactory bulb of skate (*Raja erinacea*). *J. Neurophysiol.* 64 (6), 1767–1790.  
Cinelli, A.R., Salzberg, B.M., 1992. Dendritic origin of late events in optical recordings from salamander olfactory bulb. *J. Neurophysiol.* 68 (3), 786–806.  
Civillico, E.F., Contreras, D., 2006. Integration of evoked responses in supragranular cortex studied with optical recordings *in vivo*. *J. Membr. Biol.* 96 (1), 336–351.  
Cohen, L., Salzberg, B.M., Davila, H.V., Ross, W.N., Landowne, D., Waggoner, A.S., Wang, C.H., 1974. Changes in axon fluorescence during activity: molecular probes of membrane potential. *J. Membr. Biol.* 19 (1), 1–36.  
Contreras, D., Llinas, R., 2001. Voltage-sensitive dye imaging of neocortical spatiotemporal dynamics to afferent activation frequency. *J. Neurosci.* 21 (23), 9403–9413.  
Davila, H.V., Salzberg, B.M., Cohen, L.B., Waggoner, A.S., 1973. A large change in axon fluorescence that provides a promising method for measuring membrane potential. *Nat. New Biol.* 241 (109), 159–160.  
Derdikman, D., Hildesheim, R., Ahissar, E., Arieli, A., Grinvald, A., 2003. Imaging spatiotemporal dynamics of surround inhibition in the barrels somatosensory cortex. *J. Neurosci.* 23 (8), 3100–3105.  
Destexhe, A., Rudolph, M., Fellous, J., Sejnowski, T., 2001. Fluctuating synaptic conductances recreate *in-vivo*-like activity in neocortical neurons. *Neuroscience* 107, 13–24.  
Douglas, R.J., Martin, K.A.C., 1990. Neocortex. In: Shepherd, G. (Ed.), *Synaptic Organization of the Brain*. Oxford University Press, New York, pp. 220–248.  
Douglas, R., Martin, K.A.C., 2004. Neuronal circuit of the neocortex. *Ann. Rev. Neurosci.* 27, 419.  
Eberwine, J., 2001. Molecular biological of axons: a turning point. *Neuron* 32 (6), 959–960.  
Ebner, T.J., Chen, G., 1995. Use of voltage-sensitive dyes and optical recordings in the central nervous system. *Prog. Neurobiol.* 46 (5), 463–506.  
Faugeras, O., Grimbart, F., Slotine, J.-J., 2008. Absolute stability and complete synchronization in a class of neural fields models. *SIAM J. Appl. Math.* 61 (1), 205–250.  
Ferezou, I., Bolea, S., Petersen, C.C.H., 2006. Visualizing the cortical representation of whisker touch: voltage-sensitive dye imaging in freely moving mice. *Neuron* 50, 617–629.

- Geisler, C., Brunel, N., Wang, X.-J., 2005. Contributions of intrinsic membrane dynamics to fast network oscillations with irregular neuronal discharges. *J. Neurophysiol.* 94, 4344–4361.
- Grimbert, F., Faugeras, O., Chavane, F., 2007. From neural fields to VSD optical imaging. In: *Cns07*. <<http://www.cnsorg.org>>.
- Grimbert, F., Faugeras, O., Chavane, F., 2008. Neural field model of vsd optical imaging signals. In: *Areadne08*. <<http://www.aredne.org/>>.
- Grinvald, A., Manker, A., Segal, M., 1982. Visualization of the spread of electrical activity in rat hippocampal slices by voltage-sensitive optical probes. *J. Physiol.* 333, 269–291.
- Grinvald, A., Anglister, L., Freeman, J.A., Hildesheim, R., Manker, A., 1984. Real time optical imaging of naturally evoked electrical activity in the intact frog brain. *Nature* 308 (5962), 848–850.
- Grinvald, A., Frostig, R.D., Siegel, R.M., Bartfeld, E., 1991. High-resolution optical imaging of functional brain architecture in the awake monkey. *Proc. Natl. Acad. Sci. USA* 88 (24), 11559–11563.
- Grinvald, A., Lieke, E., Frostig, R.D., Hildesheim, R., 1994. Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *J. Neurosci.* 14, 2545–2568.
- Grinvald, A., Shoham, D., Shmuel, A., Glaser, D., Vanzetta, I., Shtoyerman, E., Slovlin, H., Arieli, A., 1999. In-vivo optical imaging of cortical architecture and dynamics. In: Windhorst, U., Johansson, H. (Eds.), *Modern Techniques in Neuroscience Research*. Springer Verlag, pp. 893–969.
- Grinvald, A., Hildesheim, R., 2004. VSDi: a new era in functional imaging of cortical dynamics. *Nature* 5, 874–885. <[www.nature.com/reviews/neuro](http://www.nature.com/reviews/neuro)>.
- Gupta, R.G., Salzberg, B.M., Grinvald, A., Cohen, L., Kamino, K., Boyle, M.B., Waggoner, S., Wang, C.H., 1981. Improvements in optical methods for measuring rapid changes in membrane potential. *J. Membr. Biol.* 58 (2), 123–137.
- Haessler, S., Maass, W., 2007. A statistical analysis of information-processing properties of lamina-specific cortical microcircuits models. *Cereb. Cortex* 17 (January), 149–162.
- Hines, M., Carnevale, N., 1997. The neuron simulation environment. *Neural Comput.* 9, 1179–1209.
- Hodgkin, A., Huxley, A., 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.
- Horton, J., Adams, D.L., 2005. The cortical column: a structure without a function. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360 (1456), 837–862.
- Hubel, D., Wiesel, T., 1962. Receptive fields, binocular interaction and functional architecture in the cat visual cortex. *J. Physiol.* 160, 106–154.
- Hubel, D., Wiesel, T., 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* 28, 229–289.
- Hubel, D., Wiesel, T., 1977. Functional architecture of macaque monkey. *Proc. Roy. Soc., Lond. [B]*, 1–59.
- Hubener, M., Shoham, D., Grinvald, A., Bonhoeffer, T., 1997. Spatial relationships among three columnar systems in cat area 17. *J. Neurosci.* 17 (23), 9270–9284.
- Jancke, D., Chavane, F., Naaman, S., Grinvald, A., 2004. Imaging cortical correlates of illusion in early visual cortex. *Nature* 428, 423–426.
- Jin, W., Zhang, R., Wu, J., 2002. Voltage-sensitive dye imaging of population neuronal activity in cortical tissue. *J. Neurosci. Methods* 115, 13–27.
- Kauer, J.S., 1988. Real-time imaging of evoked activity in local circuits of the salamander olfactory bulb. *Nature* 331 (6152), 166–168.
- Kee, M.Z., Wuskell, J.P., Loew, L.M., Augustine, G.J., Sekino, Y., 2009. Imaging activity of neuronal populations with new long-wavelength voltage-sensitive dyes. *Brain Cell Biol.* 36 (5–6), 157–172.
- Kelly, J.P., Van Essen, D.C., 1974. Cell structure and function in the visual cortex of the cat. *J. Physiol.* 238, 515–547.
- Kohonen, T., 2001. *Self-Organizing Maps*, third extended ed. Springer.
- Konnerth, A., Orkand, R.K., 1986. Voltage-sensitive dyes measure potential changes in axons and glia of the frog optic nerve. *Neurosci. Lett.* 66 (1), 49–54.
- Konnerth, A., Obaid, A.L., Salzberg, B.M., 1987. Optical recording of electrical activity from parallel fibres and other cell types in skate cerebellar slices in vitro. *J. Physiol.* 393, 681–702.
- Konnerth, A., Orkand, P.M., Orkand, R.K., 1988. Optical recording of electrical activity from axons and glia of frog optic nerve: potentiometric dye responses and morphometrics. *Glia* 1 (3), 225–232.
- Kubota, M., Hosokawa, Y., Horikawa, J., 2006. Layer-specific short-term dynamics in network activity in the cerebral cortex. *Neuroreport* 17 (11), 1107–1110.
- Laaris, N., Keller, A., 2002. Functional independence of layer iv barrels. *J. Neurophysiol.* 87 (2), 1028–1034.
- La Rota, C., 2003. *Analyse de l'activité électrique multi-ties du cortex auditif chez le cobaye*. Ph.D. Thesis, Université Joseph Fourier, Grenoble I.
- Lev-Ram, V., Grinvald, A., 1986. Ca<sup>2+</sup>- and K<sup>+</sup>-dependent communication between central nervous system myelinated axons and oligodendrocytes revealed by voltage-sensitive dyes. *Proc. Natl. Acad. Sci. USA* 83 (17), 6651–6655.
- Lippert, M.T., Takagaki, K., Xu, W., Huang, X., Wu, J.Y., 2007. Methods for voltage-sensitive dye imaging of rat cortical activity with high signal-to-noise ratio. *J. Neurophysiol.* 98, 502–512.
- Manis, P.B., Freeman, J.A., 1988. Fluorescence recordings of electrical activity in goldfish optic tectum in vitro. *J. Neurosci.* 8 (2), 383–394.
- Miikkulainen, R., Bednar, J.A., Choe, Y., Sirosh, J., 2005. *Computational Maps in the Visual Cortex*. Springer, Berlin.
- Mountcastle, V., 1957. Modality and topographic properties of single neurons of cat's somatosensory cortex. *J. Neurophysiol.* 20, 408–434.
- Nelson, D.A., Katz, L.C., 1995. Emergence of functional circuits in ferret visual cortex visualized by optical imaging. *Neuron* 15 (1), 23–34.
- Nowak, L.G., Azouz, R., Sanchez-Vives, M.V., Gray, C., McCormick, D., 2003. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *J. Neurophysiol.* 89, 1541–1566. <<http://jn.physiology.org/cgi/content/abstract/89/3/1541>>.
- Orbach, H.S., Cohen, L.B., 1983. Optical monitoring of activity from many areas of the in vitro and in vivo salamander olfactory bulb: a new method for studying functional organization in the vertebrate central nervous system. *J. Neurosci.* 3, 2251–2262.
- Orbach, H.S., Van Essen, D.C., 1993. In vivo tracing of pathways and spatio-temporal activity patterns in rat visual cortex using voltage sensitive dyes. *Exp. Brain Res.* 94 (3), 371–392.
- Orbach, H.S., Cohen, L.B., Grinvald, A., 1985. Optical mapping of electrical activity in rat somatosensory and visual cortex. *J. Neurosci.* 5, 1886–1895.
- Petersen, C., Sakmann, B., 2001. Functional independent columns of rat somatosensory barrel cortex revealed with voltage-sensitive dye imaging. *J. Neurosci.* 21 (21), 8435–8446.
- Petersen, C., Grinvald, A., Sakmann, B., 2003. Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. *J. Neurosci.* 23 (3), 1298–1309.
- Raizada, R., Grossberg, S., 2003. Towards a theory of the laminar architecture of the cerebral cortex: computational clues from the visual system. *Cereb. Cortex* 13, 100–113.
- Rangan, A.V., Cai, D., McLaughlin, D.W., 2005. Modeling the spatiotemporal cortical activity associated with the line-motion illusion in primary visual cortex. *Proc. Natl. Acad. Sci. USA* 102 (52), 18793–18800.
- Reynaud, A., Barthelemy, F., Masson, G., Chavane, F., 2007. Input-output transformation in the visuo-oculomotor network. *Arch. Ital. Biol.*, 145.
- Roland, P.E., 2002. Dynamic depolarization fields in the cerebral cortex. *Trends Neurosci.* 25.
- Roland, P.E., Hanazawa, A., Undeman, C., Eriksson, D., Tompa, T., Nakamura, H., Valentiniene, S., Ahmed, B., 2006. Cortical feedback depolarization waves: a mechanism of top-down influence on early visual areas. *Proc. Natl. Acad. Sci. USA* 103 (33), 12586–12591.
- Ross, W.N., Salzberg, B.M., Cohen, L.B., Grinvald, A., Davila, H.V., Waggoner, A.S., Wang, C.H., 1977. Changes in absorption, fluorescence, dichroism, and birefringence in stained giant axons: optical measurements of membrane potential. *J. Membr. Biol.* 33 (1–2), 141–183.
- Rubin, B.D., Katz, L.C., 1999. Optical imaging of odorant representations in the mammalian olfactory bulb. *Neuron* 23 (3), 499–511.
- Salin, P., Bullier, J., 1995. Corticocortical connections in the visual system: structure and function. *Psychol. Bull.* 75, 107–154.
- Salzberg, B.M., Davila, H.V., Cohen, L.B., 1973. Optical recording of impulses in individual neurons of an invertebrate central nervous system. *Nature* 246, 508–509.
- Schummers, J., Yu, H., Sur, M., 2008. Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex. *Science* 320, 1638–1643.
- Seidemann, E., Arieli, A., Grinvald, A., Slovlin, H., 2002. Dynamics of depolarization and hyperpolarization in the frontal cortex and saccade goal. *Science* 295 (5556), 862–865.
- Sharon, D., Grinvald, A., 2002. Dynamics and constancy in cortical spatiotemporal patterns of orientation processing. *Science* 295 (5554), 512–515.
- Sharon, D., Jancke, D., Chavane, F., Na'aman, S., Grinvald, A., 2007. Cortical response field dynamics in cat visual cortex. *Cereb. Cortex* 17 (12), 2866–2877.
- Shoham, D., Glaser, D., Arieli, A., Kenet, T., Wijnberge, C., Toledo, Y., Hildesheim, R., Grinvald, A., 1999. Imaging cortical dynamics at high spatial and temporal resolution with novel blue voltage-sensitive dyes. *Neuron* 24 (4), 791–802.
- Sholl, D.A., 1955a. The organization of the visual cortex in the cat. *J. Anat.* 89, 33–46.
- Sirosh, J., Miikkulainen, R., 1994. Cooperative self-organization of afferent and lateral connections in cortical maps. *Biol. Cybern.* 71, 66–78.
- Sit, Y.F., Miikkulainen, R., 2007. A computational model of the signals in optical imaging with voltage-sensitive dyes. *Neurocomputing* 70 (10–12), 1853–1857.
- Slovlin, H., Arieli, A., Hildesheim, R., Grinvald, A., 2002. Long-term voltage-sensitive dye imaging reveals cortical dynamics in behaving monkeys. *J. Neurophysiol.* 88 (6), 3421–3438.
- Spors, H., Grinvald, A., 2002. Spatio-temporal dynamics of odor representations in the mammalian olfactory bulb. *Neuron* 34 (2), 301–315.
- Sterkin, A., Lampl, I., Ferster, D., Grinvald, A., Arieli, A., 1998. Realtime optical imaging in cat visual cortex exhibits high similarity to intracellular activity. *Neurosci. Lett.* 51.
- Tasaki, I., Watanabe, A., Carnay, L., 1968. Changes in fluorescence, turbidity, and birefringence associated with nerve excitation. *Proc. Natl. Acad. Sci. USA* 61, 883–888.
- Thomson, A., Lamy, C., 2007. Functional maps of neocortical local circuitry. *Front. Neurosci.* 1 (1), 19–42.
- Thomson, A., Morris, O., 2002. Selectivity in the inter-laminar connections made by neocortical neurones. *J. Neurocytol.* 31 (3–5), 239–246.
- Ts'o, D.Y., Frostig, R.D., Lieke, E.E., Grinvald, A., 1990. functional organization of primate visual cortex revealed by high resolution optical imaging. *Science* 249 (4967), 417–420.
- Tucker, T.R., Katz, L.C., 2003a. Recruitment of local inhibitory networks by horizontal connections in layer 2/3 of ferret visual cortex. *J. Neurophysiol.* 89 (1), 501–512.

- Tucker, T.R., Katz, L.C., 2003b. Spatiotemporal patterns of excitation and inhibition evoked by the horizontal network in layer 2/3 of ferret visual cortex. *J. Neurophysiol.* 89 (1), 488–500.
- Waggoner, A.S., Grinvald, A., 1977. Mechanisms of rapid optical changes of potential sensitive dyes. *Ann. NY Acad. Sci.* 30 (303), 217–241.
- Waters, J., Schaefer, A., Sakmann, B., 2005. Backpropagating action potentials in neurones: measurement, mechanisms and potential functions. *Prog. Biophys. Mol. Biol.* 87 (1), 145–170.
- Wilson, H., Cowan, J., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* 12, 1–24.
- Wilson, H., Cowan, J., 1973. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biol. Cybern.* 13 (2), 55–80.
- Woolum, J.C., Strumwasser, F., 1978. Membrane-potential-sensitive dyes for optical monitoring of activity in aplysia neurons. *J. Neurobiol.* 9 (3), 185–193.
- Xu, W., Huang, X., Takgaki, K., Wu, J., 2007. Compression and reflection of visually evoked cortical waves. *Neuron* 55 (1), 119–129.
- Yang, Z., Heeger, D.J., Seidemann, E., 2007. Rapid and precise retinotopic mapping of the visual cortex obtained by voltage-sensitive dye imaging in the behaving monkey. *J. Neurophysiol.* 98 (2), 1002–1014.
- Young, J.Z., 1958. Anatomical considerations. *EEG Clin. Neurophysiol.* 10, 9–11.
- Yuste, R., Tank, D.K., Kleinfeld, D., 1997. Functional study of the rat cortical microcircuitry with voltage-sensitive dye imaging of neocortical slices. *Cereb. Cortex* 7 (6), 546–558.
- Zochowski, M., Wachowiak, M., Falk, C.X., Cohen, L.B., Lam, Y.W., Antic, S., Zecevic, D., 2000. Imaging membrane potential with voltage-sensitive dyes. *Biol. Bull.* 198 (1), 749–762.

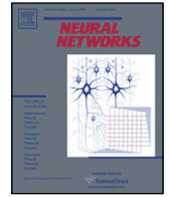




ELSEVIER

Contents lists available at ScienceDirect

## Neural Networks

journal homepage: [www.elsevier.com/locate/neunet](http://www.elsevier.com/locate/neunet)

2009 Special Issue

## Spatiotemporal dynamics in the cortical microcircuit: A modelling study of primary visual cortex layer 2/3

Andrew Symes, Thomas Wennekers\*

Centre for Theoretical and Computational Neuroscience, Portland Square Building, Room A218, University of Plymouth, PL4 8AA Plymouth, Devon, United Kingdom

## ARTICLE INFO

## Article history:

Received 3 February 2009  
 Received in revised form 20 May 2009  
 Accepted 14 July 2009

## Keywords:

Lateral connections  
 Voltage sensitive dye  
 Mean field model  
 Extracellular stimulation  
 Inhibition

## ABSTRACT

The plexus of long and short range lateral connections is a prominent feature of the layer 2/3 microcircuit in the primary visual cortex. Despite the scope for possible functionality, the interdependence of local and long range circuits is still unclear. Spatiotemporal patterns of activity appear to be shaped by the underlying connectivity architecture and strong inhibition. A modelling study has been conducted to capture population activity that has been observed in vitro using voltage sensitive dyes. The model demonstrates that the precise spatiotemporal spread of activity seen in the cortical slice results from long range connections that target specific orientation domains whilst distinct regions of suppressed activity are shown to arise from local isotropic axonal projections. Distal excitatory activity resulting from long range axons is shaped by local interneurons similarly targeted by such connections. It is shown that response latencies of distal excitation are strongly influenced by frequency dependent facilitation and low threshold characteristics of interneurons. Together, these results support hypotheses made following experimental observations in vitro and clearly illustrate the underlying mechanisms. However, predictions by the model suggest that in vivo conditions give rise to markedly different spatiotemporal activity. Furthermore, opposing data in the literature regarding inter-laminar connectivity give rise to profoundly different spatiotemporal patterns of activity in the cortex.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

Lateral connections are a prominent feature of the visual cortex and comprise the dominant synaptic input to layer 2/3 cortical cells (Binzegger, Douglas, & Martin, 2004). Such projections exhibit striking specificity in their patterns of connectivity (Bosking, Zhang, Schofield, & Fitzpatrick, 1997; Buzás et al., 2006; Gilbert & Wiesel, 1983; Kisvárdy, Tóth, Rausch, & Eysel, 1997; Malach, Amir, Harel, & Grinvald, 1993; Schmidt, Goebel, Löwel, & Singer, 1997; Sincich & Blasdel, 2001; Tanigawa, Wang, & Fujita, 2005). Within layer 2/3, glutamatergic lateral connections appear to form two distinct circuits. Local axonal projections extend in an isotropic pattern over several hundred microns and synapse on cells of all orientation tunings. Long range connections can extend many times further. Evidence also shows that these connections tend to project anisotropically in directions that are coaxial with the orientation tuning of the presynaptic cell (Bosking et al., 1997; Schmidt et al., 1997; Sincich & Blasdel, 2001). In addition, such projections target postsynaptic cells that have a similar orientation tuning to the presynaptic cell. In contrast the inhibitory lateral

circuitry of layer 2/3 appears far less extensive (Kisvárdy et al., 1997). Such a proliferation of highly specific axonal projections suggests that they play an important role in cortical behaviour. As such, horizontal connections have been associated with a number of cortical functions; non-classical receptive fields (Seriés, Lorenceau, & Frégnac, 2003); orientation tuning (Ferster & Miller, 2000); motion selectivity (Seriés, George, Lorenceau, & Frégnac, 2002). To better understand the implications of such hypotheses we consider the spatiotemporal dynamics of both the local and long range microcircuits, and their interdependence.

In vitro focal extracellular stimulation of ferret layer 2/3 slices produces very specific spatiotemporal patterns of activity (Tucker & Katz, 2003a, 2003b). Voltage sensitive dye (VSD) techniques (Fitzpatrick, 2000) allowed the authors to image activity in vitro at the population level. Clear distinctions can be made in the patterns of activity proximal or distal to the stimulus site. Local activity is extensive and diffuse whilst distal activity is more discrete and correlated with iso-orientation domains. Inhibition plays a prominent role in activity observed both locally and distally. The authors hypothesised that various mechanisms such as frequency dependent facilitation and plasticity might underlie these observations. It is proposed here that these spatiotemporal patterns of excitatory and inhibitory activity result directly from the specific underlying microcircuitry and in particular the dichotomy of local and long range connection architectures. We present a mod-

\* Corresponding author. Tel.: +44 1752 23 3593; fax: +44 1752 23 3349.  
 E-mail address: [thomas.wennekers@plymouth.ac.uk](mailto:thomas.wennekers@plymouth.ac.uk) (T. Wennekers).

elling study which clearly demonstrates such behaviour through a combination of characteristic axonal projections and electrophysiological properties of cortical cells. Furthermore, the specific underlying mechanisms responsible for the observed data are clarified. In addition, the work presented here provides additional evidence regarding which elements are activated by extracellular stimulation, and the interpretation of VSD signals. Finally, predictions were made of the expected activity in the layer 2/3 lateral microcircuitry in vivo. Driving the layer 2/3 microcircuit by extralaminar input results in profoundly different spatiotemporal patterns of activity. Distal activity is significantly attenuated whilst proximal activity is more localised and does not regenerate the same spatiotemporal patterns during repeated stimulation.

A computational model has been developed based on a mean field premise representing a patch of layer 2/3 visual cortex. Detailed patterns of connectivity specific to excitatory and inhibitory cells are represented including both local and long range paradigms. The model also captures the differences in temporal characteristics exhibited by excitatory and inhibitory cells, and their associated synapses. Propagation delays are incorporated reflecting the different synaptic pathways.

## 2. Methods

A coarse grain modelling approach has been adopted with activity represented using a mean field model that describes the time dependent activity of cortical cells. The model represents a small patch of layer 2/3 of the primary visual cortex approximately 3 mm<sup>2</sup>, using two arrays of cells, one excitatory the other inhibitory. Each set is arranged in a 51 × 51 grid, a cell at position (x, y) representing the activity at a corresponding spatial position on the cortical patch. Thus for 2601 evenly distributed locations on the cortical patch there is an excitatory and inhibitory model cell representing the average local activity of a small population of excitatory and inhibitory neurons. The time dependent activity of model cells is expressed as a series of ordinary differential equations that are numerically solved by iteration of Euler's method. Representation of both cells and synapses are influenced by Gerstner and Kistler (2002), Lumer, Edelman, and Tononi (1997) and Song, Miller, and Abbott (2000). The membrane potential,  $V(x, y)$ , of a model cell at location (x, y), is governed by

$$\tau \frac{dV(x, y, t)}{dt} = -(V(x, y, t) - V_r) - g_{ex}(V(x, y, t) - E_{ex}) - g_{in}(V(x, y, t) - E_{in}). \quad (1)$$

The resting potential of the cell is determined by  $V_r$ , and was set to -70 mV from the observation of the in vitro data that the population resting potential was close to the GABA<sub>A</sub> reversal potential (Tucker & Katz, 2003b). The passive membrane time constant,  $\tau$ , for excitatory and inhibitory cells were representative of regular and fast spiking cells in the cat and took values of 10.4 ms and 7.6 ms respectively (Nowak, Azouz, Sanchez-Vives, Gray, & McCormick, 2003). The excitatory and inhibitory synaptic inputs to a cell, representing AMPA and GABA<sub>A</sub>, are given by  $g_{ex}$  and  $g_{in}$ , respectively, with associated reversal potentials  $E_{ex}$  and  $E_{in}$ . AMPA and GABA<sub>A</sub> reversal potentials were set at 0 mV and -70 mV (Lumer et al., 1997).

As the phenomena under investigation have relatively brief temporal dynamics, only AMPA and GABA<sub>A</sub>, with their comparatively rapid kinetics, are considered. In keeping with the notion that each model cell represents a number of individual neurons, connections between model cells similarly represent a collection of neurites and synapses. Connectivity is simplified further by modelling all synaptic input of a given type, i.e., AMPA or GABA<sub>A</sub>, to a

cell by a two-stage low-pass filter of the form

$$\tau_r \frac{dg}{dt} = -g + h \quad (2)$$

$$\tau_f \frac{dh}{dt} = -h + w \sum_{x,y} \phi(d) f(V(x, y, t - \Delta t)). \quad (3)$$

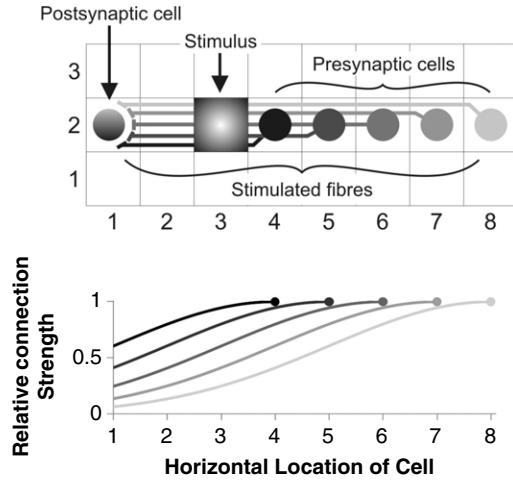
This represents essentially a difference of exponentials with  $\tau_r$  and  $\tau_f$  the rise and fall time constants for conductance changes in response to spikes. Rise and decay time constants were 0.5 ms and 2.4 ms for AMPA synapses; 1 ms and 7 ms for GABA<sub>A</sub> synapses (Lumer et al., 1997). The summation term of Eq. (3) is over all locations, (x, y), of cells that provide synaptic input. Thus for AMPA synaptic input this summation is over all excitatory cell locations, and for GABA<sub>A</sub> over all inhibitory cell locations. The distance between each of these cells and the cell that they provide synaptic input to is given by  $d$ . Synaptic efficacy is considered to decrease as the separation between pre- and postsynaptic cells increases and is reflected by the term  $\phi(d)$ .  $f(V(x, y, t - \Delta t))$  is an activation function that determines the firing rate of cell at location (x, y) at time  $t - \Delta t$  where  $t$  is the current time and  $\Delta t$  a propagation delay determined by  $d$ . Propagation delays were based on a conduction velocity of 0.2m/s in keeping with experimental data (Bringuier, Chavane, Glaeser, & Frégnac, 1999; Grinvald, Lieke, Frostig, & Hildesheim, 1994; Tucker & Katz, 2003b). The activation function is essentially the rectification model of Carandini and Ferster (2000). Spike threshold was the same for both excitatory and inhibitory cells at -54 mV (Carandini & Ferster, 2000). Gain in terms of spikes per mV had values of 2.5 spikes/mV and 10.0 spikes/mV for excitatory and inhibitory cells respectively. The value of  $w$  is determined by the type, excitatory or inhibitory, of pre- and post synaptic cells. It is a tuneable parameter that controls the relative strengths of the four connection types; excitatory to excitatory; excitatory to inhibitory; inhibitory to excitatory; and inhibitory to inhibitory. This makes the selection of gain parameters for the activation function somewhat arbitrary.

Modelling studies suggest that the number of synapses made between cells is a function of their separation (Stepanyants & Chklovskii, 2005; Stepanyants et al., 2008). It is assumed here that the number of synapses made between two cells reflects a putative connection strength. To reflect this connection strength, the efficacy of connections between cells is determined by a simple Gaussian function. Locally within layer 2/3, lateral connectivity is isotropic (Bosking et al., 1997; Buzás et al., 2006; Kisvárdy et al., 1997; Malach et al., 1993; Roerig & Kao, 1999; Schmidt et al., 1997; Sincich & Blasdel, 2001; Tanigawa et al., 2005; Tucker & Katz, 2003b), with excitatory and inhibitory model cells making projections in a radially symmetric halo (Kisvárdy et al., 1997). For local isotropic connections the connection efficacy,  $\phi(d)$ , is given by

$$\phi(d) = e^{-\frac{d^2}{2\sigma^2}}. \quad (4)$$

Data from Kisvárdy et al. (1997) suggests that the density and extent of lateral excitatory connections in cat layer 2/3 is significantly greater than that of inhibitory connections. This characteristic is captured in the model by using different values of  $\sigma$  for connections from excitatory and connections from inhibitory cells. Whilst biological data suggests excitatory connections are 2–3 times more extensive than inhibitory connections (Kisvárdy et al., 1997), results derived from modelling software (Stepanyants et al., 2008) would imply the two networks are more similar in extent. Here, connections from excitatory cells assume  $\sigma = 300 \mu\text{m}$  and those from inhibitory cells that  $\sigma = 200 \mu\text{m}$ .

From Eq. (4), the efficacy of a connection from a presynaptic cell located at (x, y) to postsynaptic cell located at (x<sub>j</sub>, y<sub>j</sub>) is



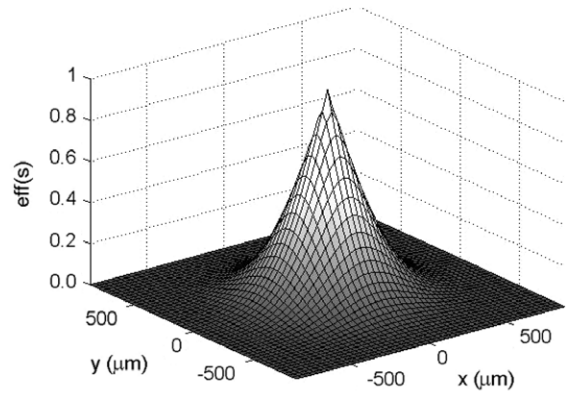
**Fig. 1.** Model of extracellular stimulation of axons. The upper panel depicts an  $8 \times 3$  grid of model cells. An extracellular stimulus is applied at the shaded square location (3, 2). The 5 solid shaded circles to the right of the stimulus site indicate presynaptic cells and the shaded lines their axonal projections; circles and lines of the same shade denote a model cell and its axonal projection. The gradient shaded circle to the left of the stimulus site at location (1, 2) marks a postsynaptic cell. Axonal projections are considered to follow a straight line from pre- to postsynaptic cell. For this postsynaptic cell only those afferent axons shown will be activated by the stimulus as no other afferent axons lie on a straight line through the stimulus site. For modelling purposes an axon passes through the stimulus site if the  $(x, y)$  location of the stimulus cell lies on the line connection the  $(x, y)$  locations of pre- and postsynaptic cells. The lower panel shows the efficacy of connections between pre- and postsynaptic cells. Each circle corresponds to the horizontal location of a presynaptic cell in the upper panel. For each circle, the line of the same shade shows the connection efficacy of projections to progressively more distant horizontal locations. The value of each line where it crosses the  $y$ -axis (i.e., at horizontal location 1) shows the efficacy of the corresponding axonal projection at the postsynaptic cell.

$\exp(-0.5((x - x_j)^2 + (y - y_j)^2)/\sigma^2)$ . Model cells are also assumed to make long range connections to a number,  $n$ , of distal patches each of which has a centre located at  $(x_i, y_i)$ ,  $i = 1, \dots, n$ . For local and long range connections, the combined efficacy for a connection from presynaptic cell at  $(x, y)$  to postsynaptic cell at  $(x_j, y_j)$  is

$$C(x, y, x_j, y_j) = \max_{i=1, \dots, n} \left( e^{-\frac{(x-x_j)^2+(y-y_j)^2}{2\sigma^2}}, a_i e^{-\frac{(x_i-x_j)^2+(y_i-y_j)^2}{2\sigma_i^2}} \right). \quad (5)$$

The parameter  $a_i$  enables weighting of distal patchy connection strengths relative to each other and the local isotropic connections. The spatial extent,  $\sigma_i$ , of each patch was identical, with value  $135 \mu\text{m}$  (Tucker & Katz, 2003b). As the current study is not specifically concerned with the relationship between long range patches and orientation tuning, their number,  $n$ , and location,  $(x_i, y_i)$  was arbitrarily chosen.

Both experimental (Nowak & Bullier, 1998a, 1998b) and theoretical evidence (McIntyre & Grill, 1999) indicate activity evoked by extracellular stimulation originates in axons rather than somata. Modelling the excitation of such fibres by extracellular stimulation is kept very simple. Consider the  $8 \times 3$  grid of model cells in the upper panel of Fig. 1 where an extracellular stimulation is applied to location (3, 2), and is indicated by the diffusely shaded square. Lateral connections are considered to project in a straight line. Thus the postsynaptic model cell at location (1, 2), indicated by the gradient shaded circle (left most circle), will receive input from stimulated fibres that lie in a straight line between itself and the stimulus location. The presynaptic cells that project these fibres (the 5 solid shaded circles to the right of the stimulus site) are seen to be those cells that lie on a straight line that passes through the stimulus site to the postsynaptic cell.



**Fig. 2.** Input from extracellular stimulation of axons. A  $1.76 \text{ mm} \times 1.76 \text{ mm}$  patch of cortex is modelled and an extracellular stimulus applied to the central location. The normalized efficacy of stimulated afferents for each location is shown. Efficacy is greatest for cells located at the stimulus site and drops off sharply for more distant cells.

From Fig. 1 the total efficacy of afferent projections to the postsynaptic cell is

$$\sum_{i=s+1}^{\infty} e^{-\frac{a_i^2}{2\sigma^2}}. \quad (6)$$

This is given the connection efficacy specified by Eq. (4). The horizontal location of the stimulus is given by  $s$  whilst  $a$  is the cortical distance between model cells. Using a continuous sum over all possible presynaptic cells, including the stimulus site, Eq. (6) can be rewritten as

$$\int_s^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx. \quad (7)$$

Here  $x$  is the distance between pre- and postsynaptic cells and  $s$  the distance between postsynaptic cell and stimulus. Normalising, the efficacy function,  $eff(s)$ , is given by

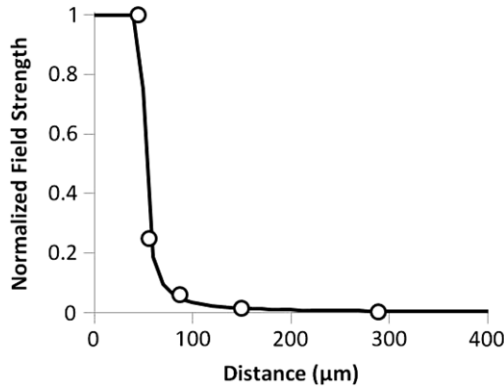
$$eff(s) = 1 - \text{erf} \left( \frac{s}{\sqrt{2}\sigma} \right). \quad (8)$$

Fig. 2 shows the efficacy of afferents for a  $1.76 \times 1.76 \text{ mm}$  cortical patch when a stimulus is applied at the centre of the grid. For postsynaptic cells at location  $(x, y)$ , the  $z$  axis gives the efficacy of stimulated afferent axons. Axons are considered to project from excitatory cells with  $\sigma = 300 \mu\text{m}$  (see Eq. (4)). This model only captures orthodromic input and not antidromic propagation. Inclusion of antidromic activation would simply lead to a fractional elevation of all cell membrane potentials. Further, this may be dwarfed by potential amplification of axonal activity by chemical synapses (Kandel, Schwartz, & Jessell, 2000).

The afferent efficacy  $eff(s)$  only considers local isotropic connections and so must be augmented to include contributions from long range anisotropic connections. However, such connections are very specific regarding pre- and postsynaptic cells. Of those anisotropic afferents that might augment  $eff(s)$ , contribution from axons of presynaptic cells at the stimulus site will dominate. Thus for a cell at location  $(x, y)$ , including anisotropic afferents, gives the new definition

$$eff(x, y, x_s, y_s) = \max_{i=1, \dots, n} \left( 1 - \text{erf} \left( \frac{\sqrt{(x-x_s)^2 + (y-y_s)^2}}{\sqrt{2}\sigma} \right), a_i e^{-\frac{\sqrt{(x-x_i)^2+(y-y_i)^2}}{\sqrt{2}\sigma_i}} \right). \quad (9)$$





**Fig. 3.** Decay in field strength. The decay in electric field strength is plotted against horizontal distance from the electrode tip. The circles show normalised data taken directly from Figure 3 of Gimsa et al. (2006). The solid line shows the fitted estimate of field strength.

Here,  $(x_s, y_s)$  denotes the location of the extracellular stimulus;  $n$  specifies the number of distal patches that cells located at  $(x_s, y_s)$  make long range anisotropic projections to;  $(x_i, y_i)$  are the corresponding centres of each patch,  $i = 1, \dots, n$ ; and  $a_i$  controls the relative weighting of each patch.

Modelling studies suggest that extracellular stimulation with bipolar electrodes produces a very localised electric field (Gimsa, Schreiber, Habel, Flehr, van Rienen, & Gimsa, 2006). These predicted electric fields were used to estimate the decay in strength with horizontal distance. Fig. 3 shows the results for a round tipped electrode with an outer pole diameter of 125 μm as used for in vitro stimulation (Tucker & Katz, 2003b) (personal communication with author).

From Fig. 3 it can be seen that the spatial extent of the electric field may cover an area representing a number of model cells, rather than just one as in Fig. 1. To maintain simplicity, each model location stimulated by the field is treated independently. Thus an addition grid of model cells is created representing excitatory afferent axons. Each of these cells is governed by

$$\tau \frac{dV(x, y, t)}{dt} = -(V(x, y, t) - V_r) + wI(x, y). \quad (10)$$

This model is similar to Eq. (1), with  $V(x, y, t)$  the axon membrane potential of axonal afferents at location  $(x, y)$ ,  $V_r$  the resting potential equal to  $-65$  mV,  $\tau$  the passive membrane time constant with value 1.5 ms (Beccroft, Alkhateeb, & Gaumont, 1994; Nowak & Bullier, 1998a) and  $I(x, y)$  the field strength at location  $(x, y)$ . The weight parameter,  $w$ , is adjusted to represent different stimulus strengths. From Eq. (10) a firing rate,  $F(x, y, t - \Delta t)$ , is determined using an activation function as in Eq. (3), with threshold  $-54$  mV and gain 10 spikes/mV. From Eqs. (9) and (10), the input to a cell at location  $(x_i, y_i)$  from afferents stimulated at location  $(x_j, y_j)$  is

$$G(x_i, y_i, x_j, y_j) = \text{eff}(x_i, y_i, x_j, y_j)F(x_j, y_j, t - \Delta t). \quad (11)$$

A second set of afferent axon cells is created to represent stimulation of afferents from GABAergic cells. Eq. (3) can then be rewritten to include input from extracellular stimulation of afferents. GABA<sub>A</sub> synapses receive input from GABAergic afferents whilst AMPA synapses receive input from glutamatergic afferents. Given a cell at location  $(p, q)$ , the general form of Eq. (3) for either type of synapse is then

$$\tau_f \frac{dh}{dt} = -h + w \sum_{x,y} \phi(d)f(V(x, y, t - \Delta t)) + w \sum_{x,y} G(p, q, x, y). \quad (12)$$

Interpreting signals from voltage sensitive dye experiments is not straightforward. The signal results from the combined activity of all membrane surfaces stained by the dye (Ebner & Chen, 1995; Grinvald et al., 1999). In terms of excitatory and inhibitory cell activity, the signal is biased by activity in the dendritic tree (Grinvald et al., 1994). To reflect the observation that the synaptic potential is slower than the current (Kandel et al., 2000) and that averaging across an area of cortex will temporally spread dendritic activity, synaptic input to a cell is low pass filtered to give the voltage sensitive dye signal

$$\tau \frac{dS(x, y)}{dt} = -S(x, y) + K(x, y). \quad (13)$$

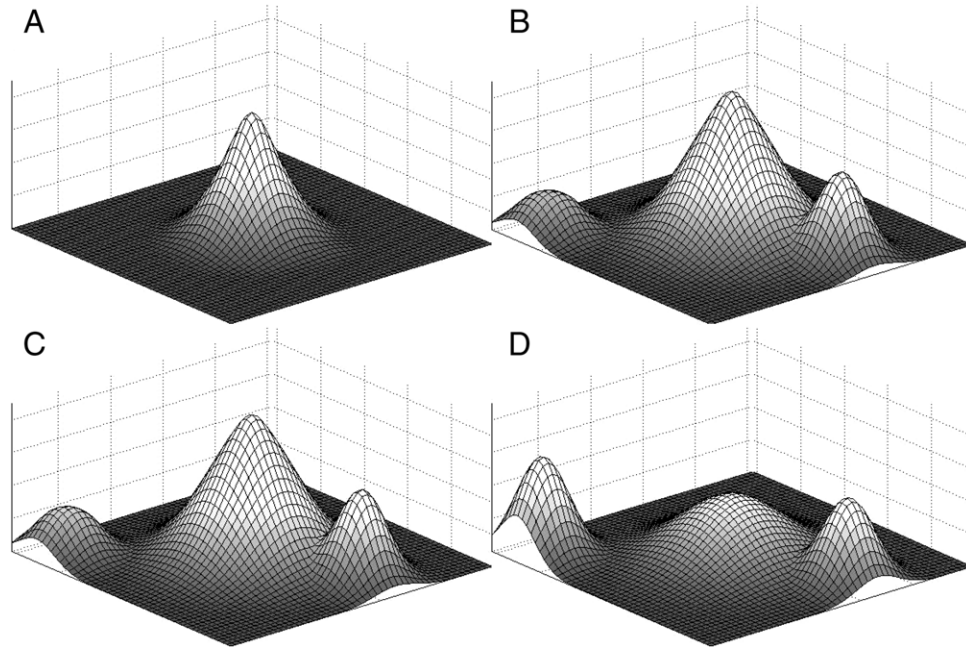
Here  $S(x, y)$  is the voltage sensitive dye signal generated by a model cell at location  $(x, y)$ .  $K(x, y)$  is the sum of synaptic inputs to the model cell located at  $(x, y)$ , i.e.,  $g_{ex}(E_{ex} - V(x, y)) + g_{in}(E_{in} - V(x, y))$  of Eq. (1), whilst  $\tau = 5$  ms. Recall that the model represents both excitatory and inhibitory cells. As such at each location  $(x, y)$  there is an inhibitory and excitatory model cell. To reflect the relative contribution of both excitatory and inhibitory cells to the voltage sensitive dye signal, the final signal at location  $(x, y)$ , is given by

$$VSD(x, y) = pS_{ex}(x, y) + (1 - p)S_{in}(x, y). \quad (14)$$

The values  $S_{ex}(x, y)$  and  $S_{in}(x, y)$  are the excitatory and inhibitory cell voltage sensitive dye signals at  $(x, y)$  as defined by Eq. (13). The relative contribution of the two cell types to the VSD signal is controlled by  $p$ . As the combined surface area of excitatory dendritic trees is expected to be significantly larger than that attributable to inhibitory dendrites  $p$  was set to 0.8. Running simulations with  $p = 0.7$  and 0.9 did not give rise to significantly different results.

### 3. Results

A mean field computational model representing the extensive lateral microcircuitry of primary visual cortex layer 2/3 has been constructed, including both local and long range lateral connectivity (Bosking et al., 1997; Buzás et al., 2006; Gilbert & Wiesel, 1983; Grinvald et al., 1994; Hirsch & Gilbert, 1991; Kisvárdy et al., 1997; Malach et al., 1993; Schmidt et al., 1997; Sincich & Blasdel, 2001; Stepanyants et al., 2008; Tanigawa et al., 2005; Tucker & Katz, 2003b). This model has subsequently been used to determine the mechanisms that underlie specific patterns of spatiotemporal activity observed in vitro in layer 2/3 of ferret primary visual cortex (Tucker & Katz, 2003b). These authors present data largely at the population level in the form of optical signals from voltage sensitive dyes (Ebner & Chen, 1995; Fitzpatrick, 2000). A qualitative rather than quantitative comparison is only possible due to a number of uncertainties including; precisely what voltage sensitive dye (VSD) signals reflect (Grinvald et al., 1999); the exact stimulus parameters associated with observed data; and which neural elements are stimulated (McIntyre & Grill, 1999; Nowak & Bullier, 1998a, 1998b). In addition to the VSD data, simultaneous single cell intracellular recordings were made. Such combined recording techniques have demonstrated that optical signals from voltage sensitive dyes at the population level serve as a good predictor of subthreshold membrane activity observed in individual cells (Grinvald et al., 1999; Petersen, Grinvald, & Sakmann, 2003; Tucker & Katz, 2003b). Whilst the in vitro data revealed a number of different characteristic behaviours this study focuses on three; the spatiotemporal spread of activity; local inhibition; and reduction in time to peak of distal activity patches.



**Fig. 4.** Model spatiotemporal activity. Four weak focal pulses of an extracellular stimulus at 100 Hz were modelled at the centre of the grid. The plots A, B, C and D show the induced activity 3, 15, 25 and 40 ms after the initial pulse. As with the results of Tucker and Katz, an initial diffuse zone of activity centred on the stimulation site is followed by more distal patches of activity that emerge in panel B and are fully established in panel D. Lighter shades indicate higher activity.

### 3.1. Spatiotemporal activity

Tucker and Katz observed that in vitro 4 weak focal extracellular pulses at 100 Hz elicited a diffuse spreading zone of activity centred on the stimulus site. A number of discrete distal zones of activity, termed "optical clusters", were also produced. In the model, weak focal stimulation was applied to the central position of a  $51 \times 51$  grid of cells representing a patch of layer 2/3 cortex  $\sim 3 \text{ mm}^2$ . A distinct spatiotemporal pattern of activity similar to the in vitro data was produced and can be seen in Fig. 4. A large, spreading zone of activity is centred on the stimulus site, with two distal zones of activity beginning to emerge approximately 5 ms after the initial stimulus. These distal zones become more fully established by 40 ms as the diffuse zone is decaying. Activation spread follows specific excitatory pathways, but propagation delays induce typical temporal response dispersal. The emergence of distinct distal regions of activity is mediated by the precise targeting of long range connections. The spatial characteristics of a local diffuse zone and distal optical clusters are regenerated after each stimulus pulse. The diffuse zone 15 ms after the initial pulse has a full width at half maximum (FWHM) of approximately  $527 \mu\text{m}$ , whilst the mean FWHM for the optical clusters is  $342 \mu\text{m}$ . A closer inspection revealed that all cell activity was subthreshold. Hence the VSD signal shown in Fig. 4 results entirely from dendritic activity driven by excitation of lateral axons. The model makes two significant assumptions regarding extracellular stimulation. The first of these is that axons, rather than somata are activated by extracellular stimulation. The second is the efficacy with which this activity is propagated laterally. To test their validity, these two assumptions were tested with additional simulations. Testing the first assumption simply required activation of somata by the stimulus. The second required employing a different distribution for the efficacy of laterally propagated activity. In this case it was decided to use a Gaussian function. In the first validation simulation, activation of somata by extracellular stimulation was achieved by replacing Eq. (1) with

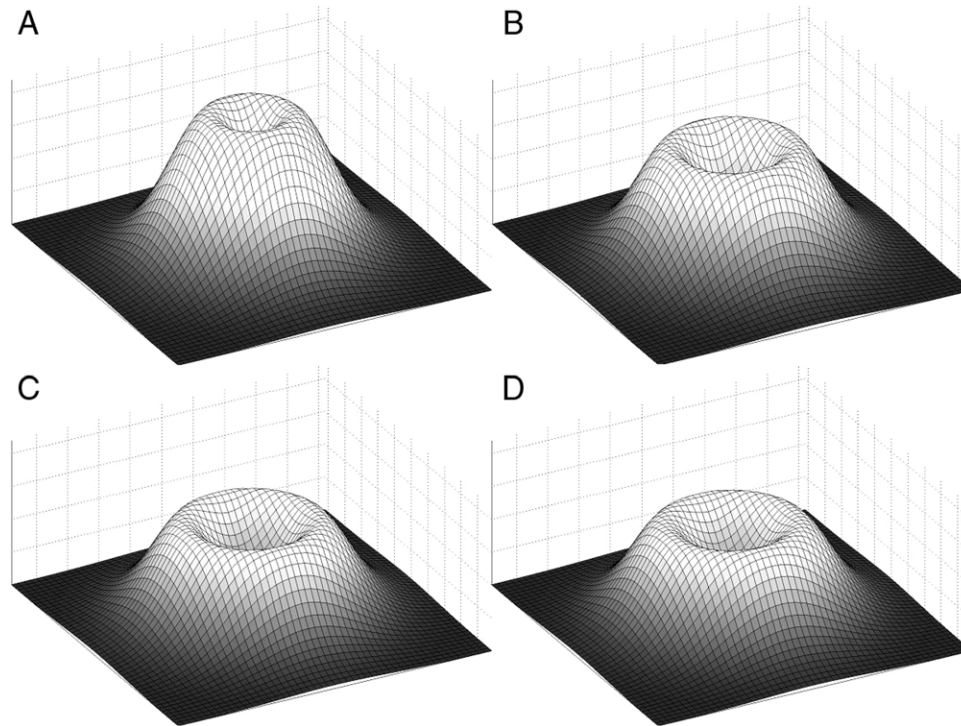
$$\tau \frac{dV(x, y, t)}{dt} = -(V(x, y, t) - V_r) - g_{ex}(V(x, y, t) - E_{ex}) - g_{in}(V(x, y, t) - E_{in}) + I(x, y). \quad (15)$$

The term  $I(x, y)$  denotes the field strength at location  $(x, y)$ . Furthermore, synaptic equation (3) rather than (12) is used. Adopting this stimulation protocol resulted in anomalous patterns of activity not observed in vitro, consisting of a deep crater situated at the electrode position that appeared immediately following stimulation (data not shown). This reduction in VSD signal was not due to inhibition but resulted from elevated membrane potentials of cells at this location. When the stimulus only activates axons, membrane potentials remain very low and similar for neighbouring cells. When only somata are activated, the membrane potential of such cells,  $V_a$ , must be suprathreshold in order to observe laterally propagating activity. However, as the stimulus is extremely localised (see Fig. 3) the membrane potential of neighbouring cells,  $V_n$ , is initially much lower and near the resting potential. Immediately following a stimulus pulse AMPA dominates synaptic activity. Since  $V_a \gg V_n$  then  $g_{ex}(E_{ex} - V_a) \ll g_{ex}(E_{ex} - V_n)$  and so from Eqs. (13) and (14) the model voltage sensitive dye signal for cells neighbouring the stimulus site is much greater than that of cells activated by the stimulus. This result supports the proposition that axons rather than somata are excited by extracellular stimulation (McIntyre & Grill, 1999; Nowak & Bullier, 1998a, 1998b). For the second validation experiment, simulations were also conducted that replaced Eq. (8) by a Gaussian function as in Eq. (4). Under these conditions, a combination of the less peaked Gaussian function and propagation velocity resulted in a noticeable expanding ring of activity following stimulus pulses which gave the appearance of an inhibitory region (data not shown). This disparity of the validation results with the in vitro data supports the model for extracellular stimulation adopted here.

### 3.2. Local inhibition

A distinctive characteristic of the optical signals recorded by Tucker and Katz was a region of suppression centred on the stimulus site. Following each stimulus pulse activity in this inhibitory region was observed to deepen and expand. In some cases this appeared to form a ring around the stimulus site (Figure 7 in Tucker and Katz (2003b)) whilst in another instance





**Fig. 5.** Expanding region of inhibition. Four focal pulses of an extracellular stimulus at 100 Hz were modelled at the centre of the grid. The plots A, B, C and D show the induced activity 10, 20, 30 and 40 ms after the initial pulse. An expanding and deepening region of inhibition is centred on the stimulus site. Lighter shades indicate higher activity.

activity at the stimulus site was also suppressed forming more of a crater than a ring (their Fig. 5<sup>1</sup>). A similar crater region of suppression was also observed in the simulations. Increasing the pulse strength of stimuli from that used to generate Fig. 4 resulted in pronounced inhibition as seen in Fig. 5. Note that the long range connections have been excluded from this simulation in order to focus on the local inhibition. Since the distal patches lay beyond the suppression region they did not affect it in any way and could safely be excluded from this particular simulation. The diameter of the inhibited region expanded from 414  $\mu\text{m}$  at 10 ms after the first stimulus (panel A of Fig. 5) to 621  $\mu\text{m}$  at 40 ms after the first stimulus (panel D of Fig. 5). This region of inhibition was still expanding when the simulation stopped, 45 ms after the initial pulse, by which time it had reached a diameter of 759  $\mu\text{m}$ . Inspection of cell activity revealed that suprathreshold activity occurred in the inhibitory population at the stimulus site. However, this only occurred following the second stimulus pulse and thus the inhibition observed in panel A of Fig. 5 is independent of suprathreshold interneuron activity. No such suprathreshold activity was observed in the excitatory population. In order to determine the contribution of suprathreshold inhibitory cell activity to the inhibition of Fig. 5, the gain term for the inhibitory cell activation function (Eq. (3)) was changed from 10 spikes/mV to 0 spikes/mV. Even with no contribution from inhibitory cells, a similar pattern of inhibition was observed to that of Fig. 5. For the same stimulus, between 10 ms and 40 ms following the first pulse the inhibitory region expanded from 414  $\mu\text{m}$  to 552  $\mu\text{m}$ , and reached a maximum of 690  $\mu\text{m}$  when the simulation stopped at 45 ms. This corresponds to a reduction in the extent of the inhibitory region of 69  $\mu\text{m}$  at both 40 ms and 45 ms, or 11% and 9% respectively. Hence the contribution

of suprathreshold interneuron activity to the observed inhibitory region is very small. Rather, the strong inhibition shown in Fig. 5 is a consequence of direct stimulation of GABAergic axons and the resultant suppression of excitatory cells they target.

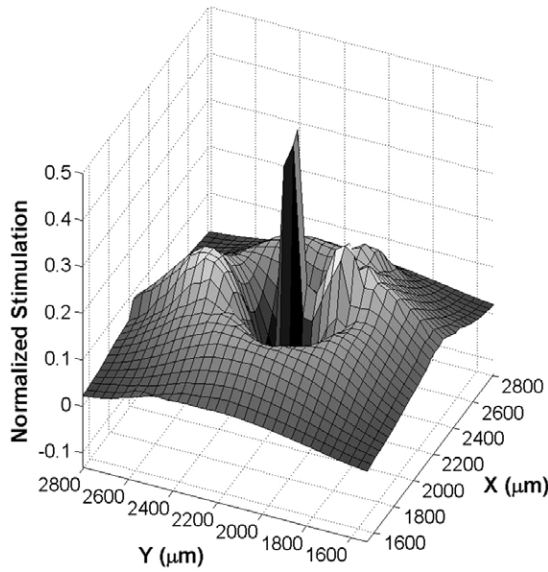
All activity in Fig. 5 is above the baseline signal observed at rest before stimulation. This is in contrast to in vitro results where VSD signals were observed to fall below resting values on occasions. The model VSD results remained positive largely due to; the proximity of the resting potential to the GABA<sub>A</sub> reversal potential; and inhibitory to excitatory connection weights. Elevating both parameters resulted in negative VSD signals in the inhibitory region (data not shown).

A search of the parameter space specifying  $w$  in Eq. (3) did not produce a ring of inhibition as observed by Tucker and Katz (2003b) and illustrated in their Fig. 7. However, the search was not exhaustive and employing an optimisation technique such as a genetic algorithm might prove more successful. As an alternative hypothesis to retuning the  $w$  parameter it was hypothesised that the ring of inhibition may be an artefact of stimulated long range connections. Recall from Eq. (9) that stimulated long range fibres only contribute to the spatiotemporal activity of distal patches and not activity close to the stimulus site. To more accurately determine the contribution of long range fibres during extracellular stimulation the model of bouton density proposed by Buzás et al. (2006) was considered. Their model predicts the bouton density that the axonal projections of a presynaptic cell will give rise to at a given cortical location. For simplicity, it is assumed here that the bouton density at a given cortical location is analogous to connection strength of axonal projections to postsynaptic cells at the same location. The bouton density model has two components, a long range orientation tuned term and a local untuned term. For the long range connections investigated here, only the tuned component is considered which is defined by

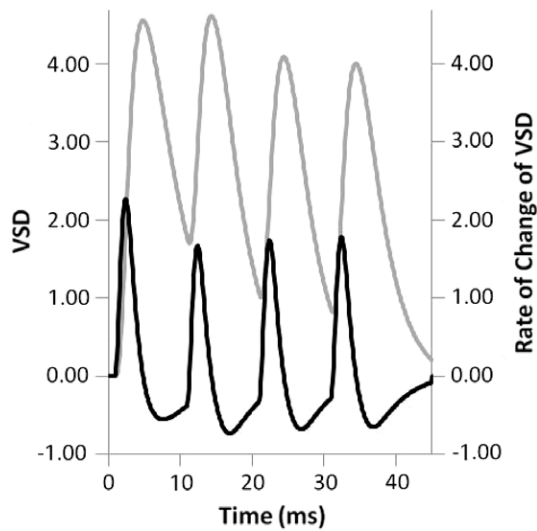
$$G(x, y, \sigma)V(\phi, \kappa, \mu). \quad (16)$$

<sup>1</sup> This is more apparent in the complete set of time series data pertaining to this figure as kindly supplied by the author in the form of an animated movie.





**Fig. 7.** Combined stimulated afferents from local and long range connections. The stimulation of local excitatory, local inhibitory and long range excitatory afferents is determined for a focal extracellular stimulation applied to the centre of a  $100 \times 100$  grid of model cells. The stimulation of local inhibitory afferents is subtracted from the sum of local and long range stimulated afferents. A distinct ring of inhibition is observed in a more diffuse isotropic region of stimulation. The central  $30 \times 30$  cells have been shown for clarity.



**Fig. 8.** VSD signal and first derivative near stimulus site. The light grey line presents the change in VSD signal with time for a point  $207 \mu\text{m}$  from the stimulus site. The corresponding rate of change of this signal with time is indicated by the darker line.

3.3. Extra-laminar stimulation

The spatiotemporal patterns of activity presented thus far have resulted from extracellular stimulation of horizontal axons. It is unlikely that activity in vivo will originate exclusively from such processes. Rather activity observed in the horizontal microcircuitry of layer 2/3 will more likely be driven by spiking activity originating from other layers. To this end further studies were conducted in order to determine whether the same spatiotemporal patterns of activity presented in Fig. 4, and in particular in Fig. 5, would be expected in vivo. Maps generated by the neurogeometry model (Stepanyants et al., 2008) were used to estimate the number of synapses made between layer 4 and layer 2/3 cells. A Gaussian estimate of each map was determined from FWHM values. The resultant Gaussian parameter values used were  $\sigma \approx 152 \mu\text{m}$ ,

$134 \mu\text{m}$ ,  $115 \mu\text{m}$ ,  $101 \mu\text{m}$  for excitatory to excitatory; excitatory to inhibitory; inhibitory to excitatory; and inhibitory to inhibitory connections. Thus Eq. (9) was replaced by

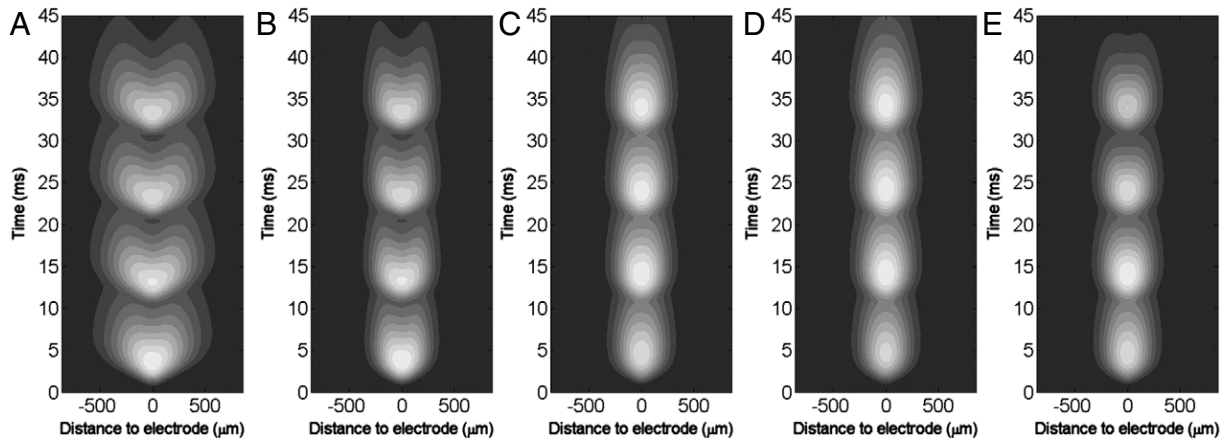
$$eff(x, y, x_s, y_s) = e^{-\frac{(x-x_s)^2 + (y-y_s)^2}{2\sigma^2}} \tag{21}$$

The first point of note is that the spatial extent of vertical projection from layer 4 to layer 2/3 is smaller than that of lateral connections in layer 2/3. The second point of note is that unlike Eq. (9), Eq. (21) does not incorporate long range patchy connections. As a consequence only disynaptic routes exist from stimulus to distal patches. For the very strong stimulus values presented in Figs. 10 and 11, long range connections were not included as such stimuli may induce suprathreshold activity in distal sites thus long range connections for model cells were not specified. When long range connections were included for these stimulus strengths only weak subthreshold activity was observed in distal patches under parameter regimes 1–4 and thus the inclusion or exclusion of long range connections is irrelevant with regard to local activity. Higher, suprathreshold, activity was observed in distal patches during strong direct stimulation of layer 2/3. However, these simulations did not incorporate facilitation of distal interneurons as discussed in the following section. Such a mechanism would have significantly attenuated any activity in distal patches of excitatory cells.

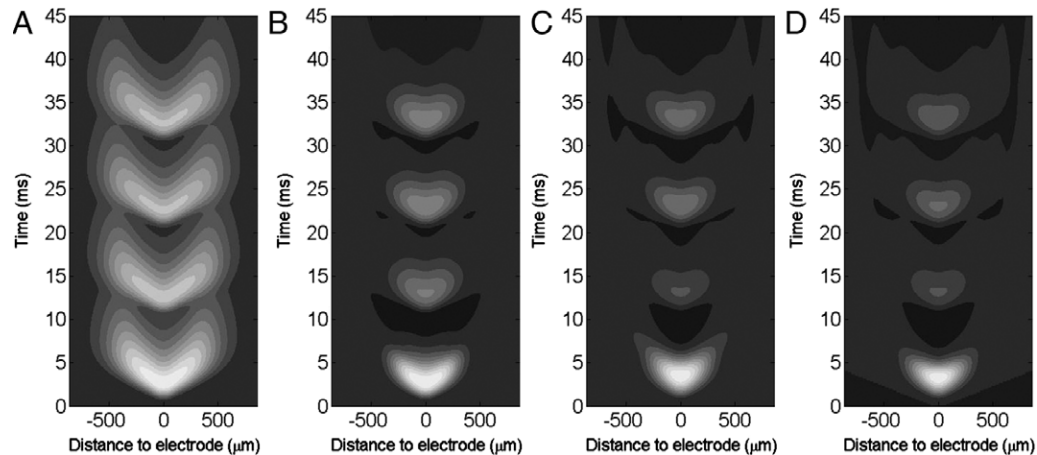
Four parameter regimes were considered. Under the first, the axonal inputs from layer 4 were weighted the same as the lateral connections within layer 2/3, i.e., the same  $w$  parameter values used in Eq. (3) were applied to axonal inputs from layer 4. The second parameter regime made the unbiased assumption that all inputs from stimulation of layer 4 projections to layer 2/3 were weighted the same. The value was selected such that extracellular stimulation could produce suprathreshold activity in cells. Data from the neurogeometry model (Stepanyants et al., 2008) suggests weighting the extra-laminar inputs roughly in the ratio 21:3:3:2 for excitatory to excitatory; excitatory to inhibitory; inhibitory to excitatory and inhibitory to inhibitory connections. This ratio was thus used to determine weights for the third regime. Other modelling studies (Binzegger et al., 2004) suggest that for these connection types the ratio of total number of synapses is 5.4:4.9:1.6:1.7. Suppose that the volume,  $V$ , below the surface defined by the connection weight of Eq. (4) ( $V = 2\pi\sigma^2$ ) is proportional to the total number of synapses of that connection type, e.g. excitatory to excitatory. Using the values of  $\sigma$  for connections from layer 4 to layer 2/3 gives weight ratios of  $\frac{5.4}{2\pi 30^2} : \frac{4.9}{2\pi 30^2} : \frac{1.6}{2\pi 28^2} : \frac{1.7}{2\pi 30^2}$  or approximately 15:22:8:8. These values were used to determine weights for the fourth regime. The  $w$  value for excitatory to excitatory connections within layer 2/3 was used for excitatory to excitatory axonal inputs from layer 4 for each parameter regime. The connection weights of all other layer 4 axonal inputs were scaled to conform to the specified weight ratio.

Stimulus strengths up to that used for Fig. 5 result in similar spatiotemporal patterns of activity under all four parameter regimes. Each parameter regime gives rise to more compact local activity than observed when layer 2/3 is stimulated directly. For the latter case the FWHM of local activity 13 ms after the initial pulse is  $515 \mu\text{m}$ . The FWHM observed 13 ms after the initial pulse for parameter regimes 1–4, is  $410, 340, 344$  and  $344 \mu\text{m}$  respectively. All local activity in layer 2/3 was subthreshold under all parameter regimes and thus no distal activity was observed. Examples of the observed activity can be seen in Fig. 9. The vertical axis represents simulation time, the horizontal axis is radial distance from the centre of the stimulus site, with lighter shades indicating higher activity. Hence, activity is a cross-section of the two dimensional VSD signal. The particular cross-section is





**Fig. 9.** Comparison of similarities between parameter regimes for extra-laminar input. For each panel the vertical axis is the simulation time; and the horizontal axis is the distance from the centre of the stimulus site with positive and negative distances indicating radial directions separated by 180 degree. The specific radial directions are unimportant due to the radial symmetry of the VSD signals about the stimulus site provided that optical clusters were not bisected. Brighter areas denote increased activity. Grey scales are specific to each panel, however zero activity is represented by the same shade in order to compare positive and negative regions. Panels B to E are from parameter regimes 1–4 whilst panel A shows activity when layer 2/3 is directly stimulated for comparison. The same stimulus as used for Fig. 5 was used for each panel.



**Fig. 10.** Comparison of similarities between parameter regimes for strong extra-laminar input. Panels B to D show the activity generated by parameter regimes 1, 2 and 4 as a result of strong extra-laminar stimulation. For comparison, panel A shows activity when layer 2/3 is directly stimulated by the same strength stimulus. The grey scale conventions of Fig. 9 are adopted here. Direct stimulation of layer 2/3 results in similar activity being observed for weaker stimuli. Activity for parameter regimes 1, 2 and 4 are similar to each other and display significant attenuation of the optical signal for stimulus pulses 2–4. Activity below that observed at rest is indicated by the darkest regions and is only observed in panels B, C and D.

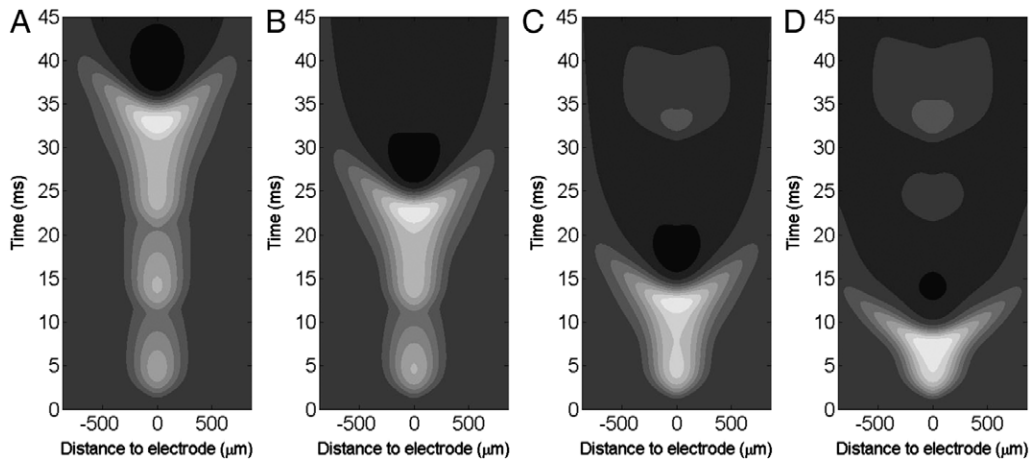
arbitrary as activity at a given instant has radial symmetry about the stimulus site provided an optical cluster is not bisected. Grey scales are relative to the range of values of each panel and thus quantitative comparisons should not be made between panels. However in all panels the same shade indicates zero activity, thus positive and negative regions can be distinguished. Panels B to E show the results from regimes 1–4 whilst panel A shows the activity observed during direct stimulation of layer 2/3 for comparison. The flattening of each ovoid region of activity and concomitant concaved upper edge indicates the appearance of an inhibitory region as observed in Fig. 5. From Fig. 9 it can be seen that inhibitory regions are present under regimes 1 and 4 (panels B and D) but are all but absent from regimes 2 and 3 (panels C and E).

Whilst Fig. 9 shows little difference in the activity observed under each parameter regime, a significant divergence in behaviour became apparent as stimulus strength was increased further. For stronger stimuli, local activity resulting from direct stimulation of layer 2/3 remains qualitatively similar to data recorded during weaker stimulation. This can be seen in panel A of Fig. 10 where the strongest stimulus produced activity characteristic of that seen in panel A of Fig. 9. Panels B, C and D of Fig. 10 show activity for parameter regimes 1, 2 and 4 using the same stimulus strength. Interestingly, under regime 1, which uses the same weights as in layer 2/3,

increasing stimulus strength results in a significant attenuation of activity generated by stimulus pulses 2–4. Also of note are the regions where activity fell below that observed at rest (indicated by the darkest areas). Fig. 10 shows that parameter regimes 1, 2 and 4 result in qualitatively similar results under strong stimulation.

The most startling results are observed under regime 3. Panels A to D of Fig. 11 show activity observed for increasing stimulus strengths with panel D generated using the same stimulus strength as that for Fig. 10. As noted previously, inhibition was not initially observed under regime 3. Here it can be seen that inhibition is initially manifest as the abolition of activity following the final stimulus pulse. Stronger stimuli abolish activity increasingly earlier in the pulse train. Associated with the appearance of inhibition is a broadening of the local activity. From panel A of Fig. 11, the FWHM 33 ms after the first pulse is 616  $\mu\text{m}$  compared with 354  $\mu\text{m}$  from panel D of Fig. 9.

The VSD signal generated by the model was not always an accurate predictor of the concomitant membrane potential of model cells. Negligible VSD signals were observed in model cells when their corresponding membrane potentials deviated significantly from the resting potential. This is presumably because excitatory and inhibitory dendritic inputs cancel. To test the relationship between VSD and membrane potential the correlation coefficient was

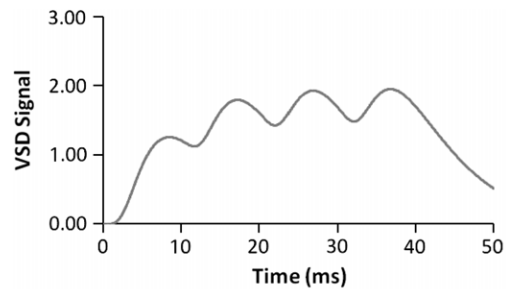


**Fig. 11.** Activity under parameter regime 3 for strong extra-laminar input. Panels A to D show the activity generated by parameter regime 3 as a result of increasing the strength of extra-laminar stimulation above that used for Fig. 10. The grey scale conventions of Fig. 9 are adopted here. The strongest stimulus strength, used for panel D, is the same as that used in Fig. 10. Under increasing stimulus strength inhibition becomes more apparent, attenuating the optical signal at earlier points in the pulse train. Regions of activity below that observed at rest are indicated by the dark areas. The region of local activity is also seen to broaden significantly in comparison with panel D of Fig. 9.

calculated between the two data sets  $VSD(x, y, t)$  and  $M(x, y, t)$ , where  $VSD(x, y, t)$  is the VSD signal at location  $(x, y)$  and time  $t$ ; and  $M(x, y, t)$  is the excitatory membrane potential at location  $(x, y)$  and time  $t$ . As model cells often displayed little or no activity, correlation coefficients were calculated for  $(x, y)$  values in the range  $x = x_l, \dots, x_u$  and  $y = y_l, \dots, y_u$ . The value of  $t$  ranged over the entire duration of the simulation. Correlation coefficients were calculated for each parameter regime and for direct stimulation of layer 2/3 using stimulus strengths up to that used for Fig. 5. All correlation coefficients irrespective of parameter regime or range of  $(x, y)$  locations were very high, and lay in the range 0.81 to 0.99. Thus, whilst discrepancies were observed, in general the VSD signal is a good predictor of the excitatory membrane potential as observed in vitro (Tucker & Katz, 2003a, 2003b) and in vivo (Petersen et al., 2003).

### 3.4. Reduction in latency of distal activity

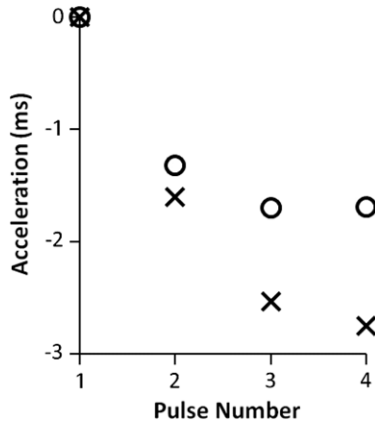
In vitro, the latency between extracellular stimulation and concomitant peak activity at distal patches decreased with successive pulses in a train of four at 100 Hz (Tucker & Katz, 2003b). The previous section has demonstrated that the activity of excitatory cells local to the stimulus site is always subthreshold. As excitatory cells are responsible for the long range patchy connectivity observed in layer 2/3 a simplified model was used to investigate the latency phenomenon. The model consisted of a number of axons driven by the extracellular stimulus and a single postsynaptic excitatory cell, and postsynaptic inhibitory cell. As the axons driven by the stimulus were separated by less than 70  $\mu\text{m}$ , and only a single postsynaptic location was considered, synaptic efficacy as a function of pre- and postsynaptic cell separation,  $\phi(d)$ , was set to 1. Using the same parameters as specified in the methodology, the temporal pattern of activity at the distal site is shown in Fig. 12. Comparing the time of each peak in Fig. 12 with the time of the preceding stimulus pulse demonstrates a reduction with successive pulses, referred to as “acceleration” (Tucker & Katz, 2003b). For a range of weak stimulus strengths the mean reduction after each pulse is shown in Fig. 13 by circles. For comparison in vitro data is shown by crosses (taken directly from Figure 8 in Tucker and Katz (2003b)). Whilst the model data does show a reduction in latency it is clearly smaller than that observed in vitro. Investigation of the model revealed that a combination of interneuron membrane properties and synaptic time constants prohibited the rapid inhibitory effects observed in vitro.



**Fig. 12.** Temporal response profile at a distal site. The activity recorded at the site of a distal patch is shown for the model.

A number of mechanisms might account for the discrepancy in latency reduction. Those initially considered in the model were spiking threshold; the temporal dynamics of synapses; and variation in conduction velocity. Each of these is now considered in turn. In vitro evidence suggests that lateral connections more readily induce a suprathreshold response in interneurons than excitatory cells (Hirsch & Gilbert, 1991). The relative depolarisation observed in some interneurons (Jonas, Bischofberger, Fricker, & Miles, 2004) would produce a similar behaviour in the model. Thus the resting potential of inhibitory cells was increased by 6 mV. Considering the second mechanism, the time course of synapses, experimental data shows significant variation. AMPA mediated postsynaptic currents can have considerably smaller time constants in hippocampal interneurons than their counterpart in excitatory cells (Jonas et al., 2004). In the rat, GABA<sub>A</sub> synapses show slightly smaller time constants than those used here (Szabadics, Tamás, & Soltesz, 2007) whilst in the mouse they show a large variation with a minimum decay time constants of 1.6 ms (Nusser, Naylor, & Mody, 2001). Furthermore the rise time and FWHM of postsynaptic potential can vary with postsynaptic cell type (Thomson, 1997). However, even significant reduction in the time constants of AMPA and GABA<sub>A</sub> synapses was insufficient to account for the reduction in latency observed in vitro. Even so, the rise and fall time constants of AMPA synapse on interneurons were reduced to 0.5 ms and 5 ms, whilst for GABA<sub>A</sub> synapses rise and fall time constants were reduced to 0.25 ms and 5 ms. The last of the three mechanisms, variation in conduction velocity is now considered. Any variation in the conduction velocity would be expected to disperse postsynaptic activity from a population of cells, as represented by a single model cell, over time. Furthermore, the extent of this dispersal would





**Fig. 13.** Reduction in time to peak. The latency between a stimulus pulse and the corresponding peak activity (acceleration) is shown for a distal patch of activity. Successive stimulus pulses lead to a reduction in the latency. For comparison, model results are shown by circles and in vitro data by crosses. Model values represent the mean latency observed over a range of weak stimulus strengths.

be amplified with increasing distance between pre- and postsynaptic cells. Subsequent postsynaptic temporal integration of such distributed activity may result in a latency reduction between successive stimulus pulses. This is of particular interest given the quoted conduction velocity  $0.24 \pm 0.2$  m/s (Tucker & Katz, 2003b) which suggests a large standard deviation of 0.2 m/s. Others also suggest a large variation in conduction velocity (Grinvald et al., 1994) and propagation delay (Hirsch & Gilbert, 1991). To capture this variation, conduction velocity, in  $\mu\text{m/ms}$ , was assumed to be normally distributed  $N(\mu_v, \sigma_v)$  where  $\mu_v$  is mean velocity and  $\sigma_v$  the standard deviation of the velocity. Maximum and minimum velocities,  $v_{\text{max}}$  and  $v_{\text{min}}$ , were constrained by

$$v_{\text{max}} = \mu_v + \sqrt{-2 \ln(T)} \quad (22)$$

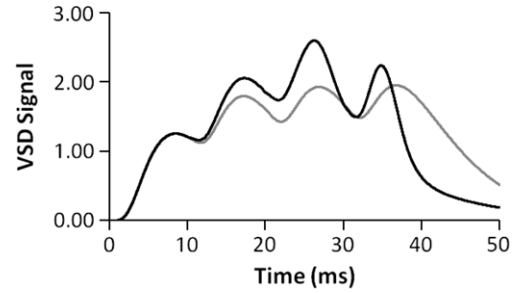
$$v_{\text{min}} = \max(1, \mu_v - \sqrt{-2 \ln(T)}). \quad (23)$$

The threshold value  $T$  was set to 0.1. As a result the activation function,  $f(V(x, y, t - \Delta t))$ , of Eq. (3) was replaced by

$$\sum_{v=v_{\text{min}}}^{v_{\text{max}}} \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{1}{2}\left(\frac{v-\mu_v}{\sigma_v}\right)^2} f(V(x, y, t - d/v)). \quad (24)$$

As with Eq. (3),  $d$  is the distance between pre- and postsynaptic cells; and  $V(x, y)$  is the membrane potential of presynaptic cell at location  $(x, y)$ . Conduction velocity variation did indeed reduce the latency of distal activity, to the extent that a standard deviation of 0.095 m/s was sufficient alone to account for the average latency reduction observed in vitro (data not shown). However, in vitro, weak stimuli result in a maximum latency reduction of  $\approx 1.4$  ms. As such stimuli are not expected to produce inhibition it is reasonable to assume that the maximum reduction in latency is attributable to variation in conduction velocity is 1.4 ms. The model was adjusted such that the maximum latency attributable to conduction velocity variation was also  $\approx 1.4$  ms.

Depolarising inhibitory cells, reduction of synaptic time constants and inclusion of conduction velocity variation did reduce the latency of distal responses. However they were insufficient to account for the reduction in distal latency observed in vitro. From Fig. 13 it can be seen that the rate at which latency reduces is higher in vitro than observed in the model. This discrepancy remained after the introduction of the aforementioned mechanisms. The relatively constant acceleration values produced by the model for stimulus pulses 2–4 results from the observation that each of these stimulus pulses regenerated the same activity in distal inhibitory cells. Consequently the degree of inhibition of distal excitatory cells was constant following stimulus pulses 2–4. To produce



**Fig. 14.** Temporal response profile of distal site incorporating synaptic facilitation. The activity recorded at the site of a distal patch is shown for the model when facilitation is present and absent. The lighter grey line shows distal activity with no facilitation and is taken from Fig. 12. The darker line indicates activity at a distal site when EPSP facilitation of AMPA synapses with interneurons is incorporated. The result is a marked decrease in the latency at the distal site as seen in the sharpening of activity peaks.

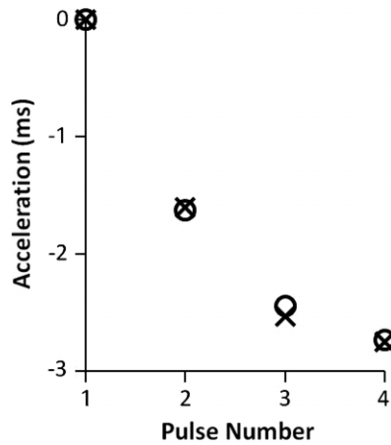
the required inhibition, the efficacy of a stimulus in exciting distal interneurons must increase with successive pulses. Trains of presynaptic spikes from pyramidal cells have been shown to produce facilitation in excitatory postsynaptic potentials (EPSPs) of interneurons (Thomson, 1997). In a short series of EPSPs this facilitation is observed to increase with successive EPSPs. A simple facilitation model was incorporated into the synapses made by excitatory cells with distal interneurons. Based on the experimental observations of Thomson (1997), the AMPA input to interneurons was weighted by

$$h = 10^{0.22p-0.22}. \quad (25)$$

Here  $h$  is the facilitation weighting, and  $p$  is the extracellular pulse number. This facilitation term resulted in a progressive increase in interneuron activity with successive stimulus pulses and a corresponding decrease in the distal latency. An example of the resultant distal VSD signal is shown in Fig. 14.

For low stimulus strengths where inhibition was subthreshold the latency reduction between the first and last stimulus pulse was  $\approx 1.3$  ms similar to the figure of 1.4 ms observed in vitro. For the model this value increased with stimulus strength up to moderate strengths after which it decreased. Again, this behaviour mirrors that reported for in vitro results. The mean latency between successive stimuli pulses was calculated over a range of strengths and is presented in Fig. 15 along with the data recorded in vitro. The mean latencies for successive stimuli pulses agree with in vitro data. For some stimulus strengths the latency did not monotonically decrease with successive pulses, rather the greatest decrease was observed following the second stimulus pulse.

To further verify the proposed mechanisms underlying activity at distal patches, the model was compared with in vitro data obtained under a stimulus protocol consisting of a single pulse of varying strength. For a single pulse qualitatively termed weak, moderate and strong, in vitro intracellular and optical recordings yielded distinct inhibitory signatures. Intracellular in vitro recordings showing inhibition in distal pyramidal cells were enhanced by depolarising the recorded cells through current injection. Such recordings showed that weak stimuli did not evoke inhibition in distal pyramidal cells, whilst progressively stronger stimuli induced greater inhibition. Strong stimuli induced significant hyperpolarisation of the cell below its depolarised resting potential. Regardless of stimulus strength, the optical signal did not fall below that observed at rest. However, evidence of inhibition was observed in the optical signal which developed a sharper peak with increasing stimulus strength. These in vitro observations in both intracellular and optical data can be seen in Fig. 11A–C. For weak, moderate and strong single pulse stimuli the model produced similar results which are presented in Fig. 16.



**Fig. 15.** Reduction in latency with synaptic facilitation. The mean latency over a range of weak stimulus strengths is shown for a distal patch of activity. Successive stimulus pulses lead to a significant reduction in the latency. For comparison, model results are shown by circles and in vitro data by crosses.

The upper three panels from left to right show model voltage sensitive dye signals for weak, moderate and strong single pulse stimuli. As with the in vitro data the model results did not fall below that observed at rest, whilst increasing stimulus strength produced a more peaked response. The lower three panels from left to right show the membrane potential of a distal excitatory cell which was been depolarised by 5 mV from a resting potential of  $-70$  mV. Again the response developed a sharper peak with increased stimulus strength as a result of stronger inhibition, a feature also observed in vitro. The model also demonstrates that a strong stimulus induced significant hyperpolarising inhibition which was manifest as membrane potentials below  $-65$  mV.

The qualitative observations of Fig. 16 can be quantified to some extent by the time to peak and FWHM. Both of these measures for intracellular and optical recording made in vitro are summarised in Table 1. For comparison, corresponding data produced by the model are also presented. The model time to peak data for both intracellular and optical recordings are representative of that observed in vitro. The only significant discrepancy between model and in vitro data is in the FWHM for weak stimuli. For weak stimuli the model produces FWHM values lower than that recorded in

**Table 1**

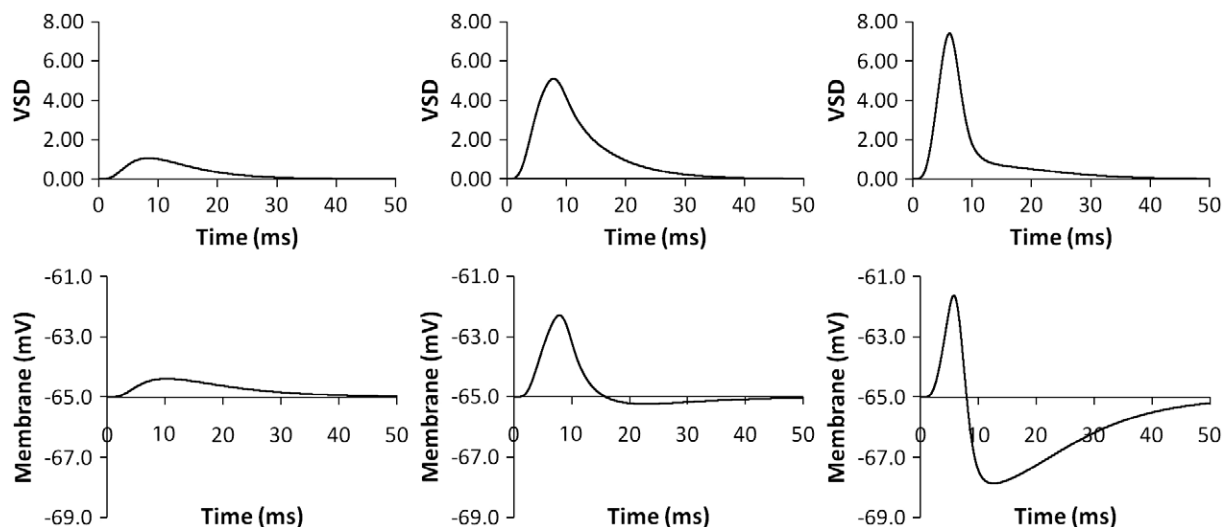
Comparison of model and in vitro results for single stimulus pulse. The time to peak and FWHM times (in ms) for voltage sensitive dye and membrane potential signals generated by the model are shown. For comparison with the model, in vitro figures taken directly from published data are also presented.

		Model	In vitro
VSD	Time to peak	8.41 ms $\pm$ 6.11 ms	9.5 ms $\pm$ 4.3 ms
	FWHM	12.67 ms $\pm$ 4.55 ms	30 ms $\pm$ 5 ms
Membrane potential	Time to peak	10.24 ms $\pm$ 5.73 ms	7 ms $\pm$ 1.9 ms
	FWHM	17.87 ms $\pm$ 3.27 ms	24 ms $\pm$ 1.5 ms

vitro. In the case of optical data this may be due to the observation that in vitro optical signals appear to persist for longer than in the model (c.f., Fig. 14 presented here with figures 8A and 9D in Tucker and Katz (2003b)). Since the optical signal is less peaked for weak stimuli, the slower decay observed in vitro may increase the corresponding FWHM. The FWHM for intracellular recordings generated by weak stimuli in vitro is also longer than the model. This may result from cellular differences in the model where cells may have had smaller membrane time constants than was the case in vitro. A closer agreement between model and in vitro intracellular FWHM data may reduce the FWHM discrepancies observed for optical data.

#### 4. Discussion

The vast majority of efferent and afferent connections in layer 2/3 of the primary visual cortex are intralaminar (Binzegger et al., 2004). Furthermore, an apparent dichotomy shows two distinct architectures; one comprising local diffuse connectivity; the other more specific, consisting of long range patchy connections (Bosking et al., 1997; Buzás et al., 2006; Gilbert & Wiesel, 1983; Kisvárdy et al., 1997; Malach et al., 1993; Schmidt et al., 1997; Sincich & Blasdel, 2001; Tanigawa et al., 2005). The specific functionality of this architecture is still unclear but elucidation is essential to understanding cortical processing of visual stimuli. In vitro focal extracellular stimulation applied to layer 2/3 slices from ferret primary visual cortex reveal population activity consistent with this architecture (Tucker & Katz, 2003b). Population activity was imaged using voltage sensitive dyes and revealed local diffuse activity and discrete patches of activity at distal locations. Superficially



**Fig. 16.** Temporal activity generated by a single stimulus pulse. The upper row shows the voltage sensitive dye signal generated by a single weak, moderate and strong stimulus pulse. The increasing stimulus strength not only increased the signal amplitude but also reduced the full width at half maximum (FWHM). The lower row shows the corresponding membrane potentials after depolarisation by 5 mV to increase the resting potential and accentuate inhibition. As with the VSD signal increasing stimulus strength increased amplitude. The underlying inhibition responsible for the decreases in FWHM was weakly present for a moderate stimulus, and fully established for a strong stimulus, c.f., Fig. 11A–C.

these data are consistent with the connectivity architecture. However, interpretation of optical signals from VSD is not trivial (Ebner & Chen, 1995; Grinvald et al., 1999). Furthermore, connectivity patterns alone are insufficient to fully describe the observed population activity. Computational modelling the VSD data offers a unique opportunity to determine more precisely the physiological properties and synaptic activity that underlies the behaviour observed in large populations of cortical cells. In addition such models offer a powerful tool to predict how such in vitro data relates to in vivo behaviour of the cortical microcircuitry.

Fundamental to the development of such models was determining an appropriate input representation, or more precisely a proper model of extracellular stimulation. Two simplifying assumptions were made regarding the model of extracellular stimulation. First that the path of axonal projections follows the shortest distance between pre and postsynaptic cells. From even the most cursory inspection of the cortex it is readily apparent that the path taken by connections between cells is anything but straight. The second simplification is that an extracellular field applied to a length of axon can be treated as a number of independent contiguous segments. A more accurate compartmental model capturing the circuitous path taken by axons and the effects of large extracellular electric fields may reveal discrepancies with the model proposed here. However, for the spatial resolution of the data modelled the simplifications adopted gave good results. At this granularity the shape of the connectivity surface (Fig. 2) dominates the spatiotemporal patterns of activity observed.

A specific feature of the spatiotemporal activity is pronounced suppression of activity local to the stimulus site. A distinct region of inhibition centred on the stimulus site is observed to deepen and expand with successive extracellular pulses during repetitive stimulation. However it should be noted that such spatiotemporal activity does not appear to be entirely stereotypical with various forms of local inhibition observed such as a partial ring (figures 4 and 7 in Tucker and Katz (2003b)), and complete suppression at the stimulus site (figures 5 and 10 in Tucker and Katz (2003b)). Tucker and Katz consider the role of increased efficacy of excitation and inhibition during high frequency stimulation and synaptic plasticity. Here we show that inhibition increases and expands almost entirely as a result of temporal integration by inhibitory synapses through lateral propagation. The temporal properties of synaptic components lead to increased inhibitory activity with repetitive stimulation at 100 Hz. This expanding and increasing region of inhibitory activity is manifest as an expanding and deepening suppression of excitation. The model presented here demonstrates that the specific balance of excitation and inhibition in combination with cellular and synaptic characteristics can combine to produce the suppressive behaviour observed in vitro.

The model shows that the ring and crater of inhibition observed in vitro are attributed to extracellular stimulation of different axonal pathways. The crater of inhibition emerges as a result of strong activation of local diffuse axonal projections. Stimulation of long range circuitry elevates the signal at the stimulus site giving rise to an inhibitory signature that manifests itself more as a ring. It is proposed that the appearance of both inhibitory patterns results from variation in the number of patches formed by long range axons. Inspection of long range patchy connectivity observed in the cortex (Bosking et al., 1997; Buzás et al., 2006; Kisvárdy et al., 1997; Sincich & Blasdel, 2001; Tanigawa et al., 2005) appears less uniform than suggested by the model of Buzás et al. (2006). Indeed, inspection of their Fig. 5 reveals that their model predicts more patches than the anatomical data show. In New World monkeys between 8 and 18 patches are observed ( $11.7 \pm 3.4$ , mean  $\pm$  standard deviation,  $n = 11$ ) (Sincich & Blasdel, 2001). In the Macaque the number of patches was between 5 and 21 ( $12 \pm 5.6$ ,  $n = 9$ ) (Tanigawa et al., 2005), however these patches were not

correlated with orientation tuning. Variation in the number and location of such patches will significantly influence the number of long range afferents to a given cell and the degree to which such afferents are stimulated. Consequently the combined stimulation of afferents as depicted in Fig. 7 can be expected to show features that range from the characteristic ring, as shown, to a crater, dependent on variations in the cortical microcircuitry.

The latency between stimulus pulse and corresponding peak in distal activity was observed in vitro to reduce with successive pulses. A similar characteristic was also observed in the model. The initial reduction in this latency, following the first pulse, was  $\approx 74\%$  of that observed in vitro. However, subsequent reductions in model results were significantly less than experimental data. This model result is largely attributed to temporal integration characteristics at the distal site. As such it is in opposition to the proposal of Tucker and Katz that the reduction in latency results from inhibitory activity at the distal site. Investigation of the model shows that at the distal site inhibition lags behind excitation due to the disynaptic temporal integration of the former, i.e., long range excitation of inhibitory cells and GABAergic synapses from inhibitory to excitatory cells at the distal patches. By comparison, excitation only suffers from temporal integration of long range excitation. As a consequence, the effective rise rate of inhibitory input to a distal excitatory cell is much slower than that of excitation. Further, inhibitory cell activity reaches a maximum early in the stimulus pulse train. Accordingly inhibition of excitatory cells is similar following successive pulses and does not result in a discernable reduction in latency. The introduction of conduction velocity variation provided another source of temporal integration that also reduced the latency at distal sites. Interestingly it was possible to produce latency reductions greater than that observed in vitro simply by increasing the temporal dispersal of postsynaptic activity. In general, this has clear implications when interpreting data as behaviour that initially appears inhibitory might equally be ascribed to temporal integration. It is particularly relevant when considering the compound signal of voltage sensitive dyes which confounds both excitatory and inhibitory activity. Whilst a number of sources of the latency reduction has been investigated, long range inhibitory connections have not been considered. Despite their paucity, GABAergic cells do make long range axonal projections, which in some cases are myelinated (Payne & Peters, 2001; Somogyi, Kisvárdy, Martin, & Whitteridge, 1983). However, stimulation of long range excitatory connections are observed to elicit spiking activity in interneurons (Hirsch & Gilbert, 1991). Therefore it is entirely possible that both types of long range connections contribute to the observed reduction in latency.

The model was also used to predict the activity expected from extracellular stimulation of innervating axons from other cortical layers. This was considered important as it offers a more accurate representation of layer 2/3 in vivo behaviour. The results of this extra-laminar stimulation demonstrate specific differences from activity elicited by intra-laminar stimulation. A number of parameter regimes was used to weight innervating axonal connections from layer 4. These included the parameters used for interlaminar connections within layer 2/3; assuming unbiased weighting; and weighting based on data from different modelling studies (Binzegger et al., 2004; Stepanyants et al., 2008). In all cases for stimulus strength up to that used to model the in vitro data only a diffuse local region of activity was observed with no distal patches. Local inhibition was not observed under all parameter regimes. For stronger stimulus strength direct activation of layer 2/3 produced characteristically similar results to those observed under weaker stimulation. However, profound changes were observed in the activity of all other parameter regimes. These included significant attenuation of local activity for



successive stimulus pulses and more extensive lateral propagation of activity. Such predictions suggest that in vivo, the contribution of diffuse and distal activity to cortical functioning is significantly less than suggested by in vitro data. This has serious implications for predicting in vivo behaviour on the basis of activity evoked in vitro by extracellular stimulation. These results also demonstrate that current predictions of intra-laminar connectivity (Binzegger et al., 2004; Stepanyants et al., 2008) lead to profoundly different spatiotemporal patterns of activity. As such this observation highlights the need for more empirical data on cortical connectivity at the population level.

## Acknowledgment

Many thanks to Thomas Tucker for additional data and its interpretation. Thanks to Zoltán Kisvárday for supplying orientation maps of cat area 18. This project was supported by funding under the Sixth Research Framework Programme of the European Union under the grant no. FP6-2004-IST-FETPI 15879 (FACETS).

Andy Symes died in July 2008, briefly after finishing the work reported here which was part of his Ph.D. project. This publication aims at posthumously honouring his outstanding work as a young scientist.

## References

- Beecroft, S. I., Alkhateeb, R. P., & Gaumond, R. P. (1994). Nerve fiber stimulation threshold for multiphase stimuli. In J. Masi, R. Davis, M. Fox, & R. Peura (Eds.), *Bioengineering: 20th annual northeast conference* (pp. 67–68). New York, City: IEEE Press.
- Binzegger, T., Douglas, R. J., & Martin, K. A. C. (2004). A quantitative map of the circuit of cat primary visual cortex. *The Journal of Neuroscience*, 24, 8441–8453.
- Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, 17, 2112–2127.
- Bringuier, V., Chavane, F., Glaeser, L., & Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science*, 283, 695–699.
- Buzás, P., Kovács, K., Ferecskó, A. S., Budd, J. M. L., Eysel, U. T., & Kisvárday, Z. F. (2006). Model-based analysis of excitatory lateral connections in the visual cortex. *The Journal of Comparative Neurology*, 499, 861–881.
- Carandini, M., & Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *The Journal of Neuroscience*, 20, 470–484.
- Ebner, T. J., & Chen, G. (1995). Use of voltage-sensitive dyes and optical recordings in the central nervous system. *Progress in Neurobiology*, 46, 463–506.
- Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Review of Neuroscience*, 23, 441–471.
- Fitzpatrick, D. (2000). Cortical imaging: Capturing the moment. *Current Biology*, 10, R187–R190.
- Gerstner, W., & Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity*. New York: Cambridge University Press.
- Gilbert, C. D., & Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *The Journal of Neuroscience*, 3, 1116–1133.
- Gimsa, U., Schreiber, U., Habel, B., Flehr, J., van Rienen, U., & Gimsa, J. (2006). Matching geometry and stimulation parameters of electrodes for deep brain stimulation experiments? Numerical considerations. *Journal of Neuroscience Methods*, 150, 212–227.
- Grinvald, A., Lieke, E. E., Frostig, R. D., & Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *The Journal of Neuroscience*, 14, 2545–2568.
- Grinvald, A., Shoham, D., Schmuell, A., Glaser, D., Vanzetta, I., Shtoyerman, E., et al. (1999). In-vivo optical imaging of cortical architecture and dynamics. In H. Johansson, & U. Windhorst (Eds.), *Modern techniques in neuroscience research* (pp. 893–969). Springer Verlag.
- Hirsch, J. A., & Gilbert, C. D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. *The Journal of Neuroscience*, 11, 1800–1809.
- Jonas, P., Bischofberger, J., Fricker, D., & Miles, R. (2004). Interneuron diversity series: Fast in, fast out – temporal and spatial signal processing in hippocampal interneurons. *Trends in Neuroscience*, 27, 30–40.
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science*. New York: McGraw-Hill.
- Kisvárday, Z. F., Tóth, É., Rausch, M., & Eysel, U. T. (1997). Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cerebral Cortex*, 7, 605–618.
- Lumer, E. D., Edelman, G. M., & Tononi, G. (1997). Neural dynamics in a model of the thalamocortical system. I. Layers, loops and the emergence of fast synchronous rhythms. *Cerebral Cortex*, 7, 207–227.
- Malach, R., Amir, Y., Harel, M., & Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 10460–10473.
- McIntyre, C. C., & Grill, W. M. (1999). Excitation of central nervous system neurons by nonuniform electric fields. *Biophysical Journal*, 76, 878–888.
- Nowak, L. G., Azouz, R., Sanchez-Vives, M. V., Gray, C. M., & McCormick, D. A. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *Journal of Neurophysiology*, 89, 1541–1566.
- Nowak, L. G., & Bullier, J. (1998a). Axons, but not cell bodies, are activated by electrical stimulation in cortical gray matter I. Evidence from chronaxie measurements. *Experimental Brain Research*, 118, 477–488.
- Nowak, L. G., & Bullier, J. (1998b). Axons, but not cell bodies, are activated by electrical stimulation in cortical gray matter II. Evidence from selective inactivation of cell bodies and axon initial segments. *Experimental Brain Research*, 118, 489–500.
- Nusser, Z., Naylor, D., & Mody, I. (2001). Synapse-specific contribution of the variation of transmitter concentration to the decay of inhibitory postsynaptic currents. *Biophysical Journal*, 80, 1251–1261.
- Payne, B. R., & Peters, A. (2001). *The cat primary visual cortex*. San Diego: Academic Press Inc.
- Petersen, C. C. H., Grinvald, A., & Sakmann, B. (2003). Spatiotemporal dynamics of sensory responses in layer 2/3 of rat barrel cortex measured in vivo by voltage-sensitive dye imaging combined with whole-cell voltage recordings and neuron reconstructions. *The Journal of Neuroscience*, 23, 1298–1309.
- Roerig, B., & Kao, J. P. Y. (1999). Organization of intracortical circuits in relation to direction preference maps in ferret visual cortex. *The Journal of Neuroscience*, 19, 1–5.
- Schmidt, K. E., Goebel, R., Löwel, S., & Singer, W. (1997). The perceptual grouping criterion of colinearity is reflected by anisotropies of connections in the primary visual cortex. *The European Journal of Neuroscience*, 9, 1083–1089.
- Seriés, P., George, S., Lorenceau, J., & Frégnac, Y. (2002). Orientation dependent modulation of apparent speed: A model based on the dynamics of feed-forward and horizontal connectivity in V1 cortex. *Vision Research*, 42, 2781–2797.
- Seriés, P., Lorenceau, J., & Frégnac, Y. (2003). The “silent” surround of V1 receptive fields: Theory and experiments. *Journal of Physiology, Paris*, 97, 453–474.
- Sincich, L. C., & Blasdel, G. G. (2001). Oriented axon projections in primary visual cortex of the monkey. *The Journal of Neuroscience*, 21, 4416–4426.
- Somogyi, P., Kisvárday, Z. F., Martin, K. A. C., & Whitteridge, D. (1983). Synaptic connections of morphologically identified and physiologically characterized large basket cells in the striate cortex of cat. *Neuroscience*, 10, 261–294.
- Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3, 919–926.
- Stepanyants, A., & Chklovskii, D. B. (2005). Neurogeometry and potential synaptic connectivity. *Trends in Neuroscience*, 28, 387–394.
- Stepanyants, A., Hirsch, J. A., Martinez, L. M., Kisvárday, Z. F., Ferecskó, A. S., & Chklovskii, D. B. (2008). Local potential connectivity in cat primary visual cortex. *Cerebral Cortex*, 18, 13–28.
- Szabados, J., Tamás, G., & Soltesz, I. (2007). Different transmitter transients underlie presynaptic cell type specificity of GABA<sub>A</sub>, slow and GABA<sub>A</sub>, fast. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 14831–14836.
- Tanigawa, H., Wang, Q., & Fujita, I. (2005). Organization of horizontal axons in the inferior temporal cortex and primary visual cortex of the macaque monkey. *Cerebral Cortex*, 15, 1887–1899.
- Thomson, A. M. (1997). Activity-dependent properties of synaptic transmission at two classes of connections made by rat neocortical pyramidal axons in vitro. *Journal of Physiology*, 502, 131–147.
- Tucker, T. R., & Katz, L. C. (2003a). Recruitment of local inhibitory networks by horizontal connections in layer 2/3 of ferret visual cortex. *The Journal of Neurophysiology*, 89, 501–512.
- Tucker, T. R., & Katz, L. C. (2003b). Spatiotemporal patterns of excitation and inhibition evoked by the horizontal network in layer 2/3 of ferret visual cortex. *The Journal of Neurophysiology*, 89, 488–500.



# A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs

Olivier Faugeras<sup>1\*</sup>, Jonathan Touboul<sup>1</sup> and Bruno Cessac<sup>1,2,3</sup>

<sup>1</sup> NeuroMathComp Laboratory, INRIA/ENS, France

<sup>2</sup> Laboratoire Jean-Alexandre Dieudonné, Nice, France

<sup>3</sup> Université de Nice, Nice, France

## Edited by:

David Hansel, University of Paris, France

## Reviewed by:

Claude Meunier, René Descartes

University, France

Germán Mato, CONICET - Centro Atómico Bariloche - Instituto Balseiro, Argentina

## \*Correspondence:

Olivier Faugeras, NeuroMathComp Laboratory, INRIA/ENS, 2004 Route des Lucioles, 06902 Sophia-Antipolis, France.

e-mail: olivier.faugeras@sophia.inria.fr

We deal with the problem of bridging the gap between two scales in neuronal modeling. At the first (microscopic) scale, neurons are considered individually and their behavior described by stochastic differential equations that govern the time variations of their membrane potentials. They are coupled by synaptic connections acting on their resulting activity, a nonlinear function of their membrane potential. At the second (mesoscopic) scale, interacting populations of neurons are described individually by similar equations. The equations describing the dynamical and the stationary mean-field behaviors are considered as functional equations on a set of stochastic processes. Using this new point of view allows us to prove that these equations are well-posed on any finite time interval and to provide a constructive method for effectively computing their unique solution. This method is proved to converge to the unique solution and we characterize its complexity and convergence rate. We also provide partial results for the stationary problem on infinite time intervals. These results shed some new light on such neural mass models as the one of Jansen and Rit (1995): their dynamics appears as a coarse approximation of the much richer dynamics that emerges from our analysis. Our numerical experiments confirm that the framework we propose and the numerical methods we derive from it provide a new and powerful tool for the exploration of neural behaviors at different scales.

**Keywords:** mean-field analysis, stochastic processes, stochastic differential equations, stochastic networks, stochastic functional equations, random connectivities, multi-populations networks, neural mass models

## INTRODUCTION

Modeling neural activity at scales integrating the effect of thousands of neurons is of central importance for several reasons. First, most imaging techniques are not able to measure individual neuron activity (“microscopic” scale), but are instead measuring mesoscopic effects resulting from the activity of several hundreds to several hundreds of thousands of neurons. Second, anatomical data recorded in the cortex reveal the existence of structures, such as the cortical columns, with a diameter of about 50  $\mu\text{m}$  to 1 mm, containing of the order of 100–100000 neurons belonging to a few different species. These columns have specific functions. For example, in the visual cortex V1, they respond to preferential orientations of bar-shaped visual stimuli. In this case, information processing does not occur at the scale of individual neurons but rather corresponds to an activity integrating the collective dynamics of many interacting neurons and resulting in a mesoscopic signal. The description of this collective dynamics requires models which are different from individual neurons models. In particular, if the accurate description of one neuron requires “ $m$ ” parameters (such as sodium, potassium, calcium conductances, membrane capacitance, etc...), it is not necessarily true that an accurate mesoscopic description of an assembly of  $N$  neurons requires  $Nm$  parameters. Indeed, when  $N$  is large enough averaging effects appear, and the collective dynamics is well described by an effective mean-field, summarizing the effect of the interactions of a neuron with the other neurons, and depending on a few effective

control parameters. This vision, inherited from statistical physics requires that the space scale be large enough to include a large number of microscopic components (here neurons) and small enough so that the region considered is homogeneous. This is in effect for instance the case of cortical columns.

However, obtaining the evolution equations of the effective mean-field from microscopic dynamics is far from being evident. In simple physical models this can be achieved via the law of large numbers and the central limit theorem, provided that time correlations decrease sufficiently fast. This type of approach has been generalized to such fields as quantum field theory or non equilibrium statistical mechanics. To the best of our knowledge, the idea of applying mean-field methods to neural networks dates back to Amari (Amari, 1972; Amari et al., 1977). In his approach, the author uses an assumption that he called the “local chaos hypothesis”, reminiscent of Boltzmann’s “molecular chaos hypothesis”, that postulates the vanishing of individual correlations between neurons, when the number  $N$  of neurons tends to infinity. Later on, Sompolinsky et al. (1998) used a dynamic mean-field approach to conjecture the existence of chaos in an homogeneous neural network with random independent synaptic weights. This approach was formerly developed by Sompolinsky and colleagues for spinglasses (Crisanti and Sompolinsky, 1987a,b; Sompolinsky and Zippelius, 1982), where complex effects such as aging or coexistence of a diverging number of metastable states, renders the mean-field analysis delicate in the long time limit (Houghton et al., 1983).



On the opposite, these effects do not appear in the neural network considered in Sompolinsky et al. (1998) because the synaptic weights are independent (Cessac, 1995) (and especially non symmetric, in opposition to spin-glasses). In this case, the Amari approach and the dynamic mean-field approach lead to the same mean-field equations. Later on, the mean-field equations derived by Sompolinsky and Zippelius (1982) for spin-glasses were rigorously obtained by Ben-Arous and Guionnet (Ben-Arous and Guionnet, 1995, 1997; Guionnet, 1997). The application of their method to a discrete time version of the neural network considered in Sompolinsky et al. (1998) and in Molgedey et al. (1992) was done by Moynot and Samuelides (2002).

Mean-field methods are often used in the neural network community but there are only a few rigorous results using the dynamic mean-field method. The main advantage of dynamic mean-field techniques is that they allow one to consider neural networks where synaptic weights are random (and independent). The mean-field approach allows one to state general and generic results about the dynamics as a function of the statistical parameters controlling the probability distribution of the synaptic weights (Samuelides and Cessac, 2007). It does not only provide the evolution of the mean activity of the network but, because it is an equation on the law of the mean-field, it also provides information on the fluctuations around the mean and their correlations. These correlations are of crucial importance as revealed in the paper by Sompolinsky et al. (1998). Indeed, in their work, the analysis of correlations allows them to discriminate between two distinct regimes: a dynamics with a stable fixed point and a chaotic dynamics, while the mean is identically 0 in the two regimes.

However, this approach has also several drawbacks explaining why it is so seldom used. First, this method uses a generating function approach that requires heavy computations and some “art” for obtaining the mean-field equations. Second, it is hard to generalize to models including several populations. Finally, dynamic mean-field equations are usually supposed to characterize in fine a stationary process. It is then natural to search for stationary solutions. This considerably simplifies the dynamic mean-field equations by reducing them to a set of differential equations (see Section “Numerical Experiments”) but the price to pay is the unavoidable occurrence in the equations of a non free parameter, the initial condition, that can only be characterized through the investigation of the nonstationary case.

Hence it is not clear whether such a stationary solution exists, and, if it is the case, how to characterize it. To the best of our knowledge, this difficult question has only been investigated for neural networks in one paper by Crisanti et al. (1990).

Different alternative approaches have been used to get a mean-field description of a given neural network and to find its solutions. In the neuroscience community, a static mean-field study of multi-population network activity was developed by Treves (1993). This author did not consider external inputs but incorporated dynamical synaptic currents and adaptation effects. His analysis was completed in Abbott and Van Vreeswijk (1993), where the authors considered a unique population of nonlinear oscillators subject to a noisy input current. They proved, using a stationary Fokker–Planck formalism, the stability of an asynchronous state in the network. Later on, Gerstner (1995) built a new approach

to characterize the mean-field dynamics for the Spike Response Model, via the introduction of suitable kernels propagating the collective activity of a neural population in time.

Brunel and Hakim (1999) considered a network composed of integrate-and-fire neurons connected with constant synaptic weights. In the case of sparse connectivity, stationarity, and considering a regime where individual neurons emit spikes at low rate, they were able to study analytically the dynamics of the network and to show that the network exhibited a sharp transition between a stationary regime and a regime of fast collective oscillations weakly synchronized. Their approach was based on a perturbative analysis of the Fokker–Planck equation. A similar formalism was used in Mattia and Del Giudice (2002) which, when complemented with self-consistency equations, resulted in the dynamical description of the mean-field equations of the network, and was extended to a multi-population network.

Finally, Chizhov and Graham (2007) have recently proposed a new method based on a population density approach allowing to characterize the mesoscopic behavior of neuron populations in conductance-based models. We shortly discuss their approach and compare it to ours in Section “Discussion”.

In the present paper, we investigate the problem of deriving the equations of evolution of neural masses at mesoscopic scales from neurons dynamics, using a new and rigorous approach based on stochastic analysis.

The article is organized as follows. In Section “Mean-Field Equations for Multi-Populations Neural Network Models” we derive from first principles the equations relating the membrane potential of each of a set of neurons as function of the external injected current and noise and of the shapes and intensities of the postsynaptic potentials in the case where these shapes depend only on the postsynaptic neuron (the so-called voltage-based model). Assuming that the shapes of the postsynaptic potentials can be described by linear (possibly time-dependent) differential equations we express the dynamics of the neurons as a set of stochastic differential equations. Assuming that the synaptic connectivities between neurons satisfy statistical relationship only depending on the population they belong to, we obtain the mean-field equations summarizing the interactions of the  $P$  populations in the limit where the number of neurons tend to infinity. These equations can be derived in several ways, either heuristically following the lines of Amari (Amari, 1972; Amari et al., 1977), Sompolinsky (Crisanti et al., 1990; Sompolinsky et al., 1998), and Cessac (Cessac, 1995; Samuelides and Cessac, 2007), or rigorously as in the work of Ben-Arous and Guionnet (Ben-Arous and Guionnet, 1995, 1997; Guionnet, 1997). The purpose of this article is not the derivation of these mean-field equations but to prove that they are well-posed and to provide an algorithm for computing their solution. Before we do this we provide the reader with two important examples of such mean-field equations. The first example is what we call the simple model, a straightforward generalization of the case studied by Amari and Sompolinsky. The second example is a neuronal assembly model, or neural mass model, as introduced by Freeman (1975) and exemplified in Jansen and Rit’s (1995) cortical column model.

In Section “Existence and Uniqueness of Solutions in Finite Time” we consider the problem of solutions over a finite time

interval  $[t_0, T]$ . We prove, under some mild assumptions, the existence and uniqueness of a solution of the dynamic mean-field equations given an initial condition at time  $t_0$ . The proof consists in showing that a nonlinear equation defined on the set of multidimensional Gaussian random processes defined on  $[t_0, T]$  has a fixed point. We extend this proof in Section “Existence and Uniqueness of Stationary Solutions” to the case of stationary solutions over the time interval  $[-\infty, T]$  for the simple model. Both proofs are constructive and provide an algorithm for computing numerically the solutions of the mean-field equations.

We then study in Section “Numerical Experiments” the complexity and the convergence rate of this algorithm and put it to good use: We first compare our numerical results to the theoretical results of Sompolinsky and colleagues (Crisanti et al., 1990; Sompolinsky et al., 1998). We then provide an example of numerical experiments in the case of two populations of neurons where the role of the mean-field fluctuations is emphasized.

Along the paper we introduce several constants. To help the reader we have collected in **Table 1** of Appendix D, the most important ones and the place where they are defined in the text.

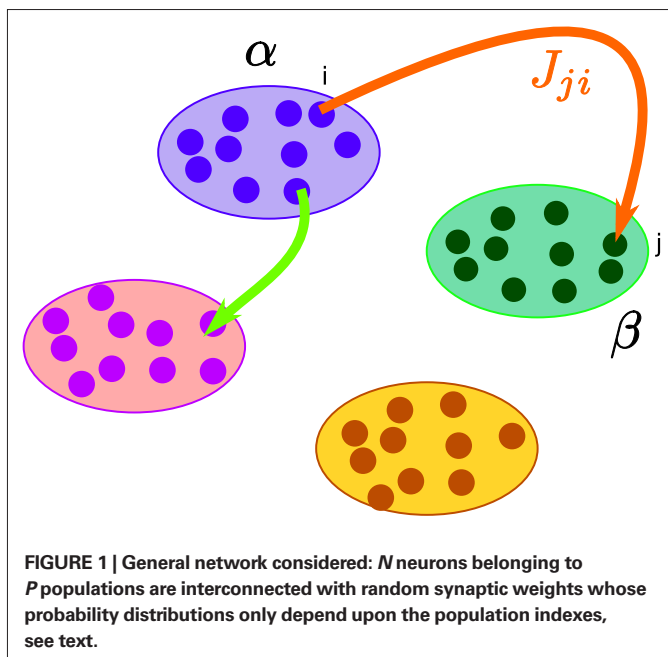
### MEAN-FIELD EQUATIONS FOR MULTI-POPULATIONS NEURAL NETWORK MODELS

In this section we introduce the classical neural mass models and compute the related mean-field equations they satisfy in the limit of an infinite number of neurons.

#### THE GENERAL MODEL

##### General framework

We consider a network composed of  $N$  neurons indexed by  $i \in \{1, \dots, N\}$  belonging to  $P$  populations indexed by  $\alpha \in \{1, \dots, P\}$  such as those shown in **Figure 1**. Let  $N_\alpha$  be the number of neurons in population  $\alpha$ . We have  $N = \sum_{\alpha=1}^P N_\alpha$ . We define the population which the neuron  $i$ ,  $i = 1, \dots, N$  belongs to.



**Definition 1.** The function  $p: \{1, \dots, N\} \rightarrow \{1, \dots, P\}$  associates to each neuron  $i \in \{1, \dots, N\}$ , the population  $\alpha = p(i) \in \{1, \dots, P\}$ , it belongs to.

We consider that each neuron  $i$  is described by its membrane potential  $V_i(t)$ , and the related instantaneous firing rate is deduced from it through a relation of the form  $v_i(t) = S_i(V_i(t))$  (Dayan and Abbott, 2001; Gerstner and Kistler, 2002), where  $S_i$  is a sigmoidal function.

A single action potential from neuron  $j$  generates a postsynaptic potential  $\text{PSP}_{ij}(u)$  on the postsynaptic neuron  $i$ , where  $u$  is the time elapsed after the spike is received. We neglect the delays due to the distance traveled down the axon by the spikes.

Assuming that the postsynaptic potentials sum linearly, the average membrane potential of neuron  $i$  is

$$V_i(t) = \sum_{j,k,t_k>t_0} \text{PSP}_{ij}(t-t_k) + V_i(t_0),$$

where the sum is taken over the arrival times of the spikes produced by the neurons  $j$  after some reference time  $t_0$ . The number of spikes arriving between  $t$  and  $t + dt$  is  $v_j(t)dt$ . Therefore we have

$$\begin{aligned} V_i(t) &= \sum_j \int_{t_0}^t \text{PSP}_{ij}(t-s)v_j(s) ds + V_i(t_0) \\ &= \sum_j \int_{t_0}^t \text{PSP}_{ij}(t-s)S_j(V_j(s)) ds + V_i(t_0), \end{aligned} \tag{1}$$

or, equivalently

$$v_i(t) = S_i \left( \sum_j \int_{t_0}^t \text{PSP}_{ij}(t-s)v_j(s) ds + V_i(t_0) \right). \tag{2}$$

The  $\text{PSP}_{ij}$ s can depend on several variables in order to account for instance for adaptation or learning.

We now make the simplifying assumption that the shape of the postsynaptic potential  $\text{PSP}_{ij}$  only depends on the postsynaptic population, which corresponds to the voltage-based models in Ermentrout’s (1998) classification.

**The voltage-based model.** The assumption, made in Hopfield (1984), is that the postsynaptic potential has the same shape no matter which presynaptic population caused it, the sign and amplitude may vary though. This leads to the relation

$$\text{PSP}_{ij}(t) = \bar{J}_{ij}g_i(t).$$

$g_i$  represents the unweighted shape (called a  $g$ -shape) of the postsynaptic potentials and  $\bar{J}_{ij}$  is the strength of the postsynaptic potentials elicited by neuron  $j$  on neuron  $i$ . At this stage of the discussion, these weights are supposed to be deterministic. This is reflected in the notation  $\bar{J}_{ij}$  which indicates an average value<sup>1</sup>. From Eq. 1 we have

$$V_i(t) = \int_{t_0}^t g_i(t-s) \left( \sum_j \bar{J}_{ij}v_j(s) \right) ds + V_i(t_0).$$

<sup>1</sup>When we come to the mean-field equations they will be modeled as random variables.

So far we have only considered the synaptic inputs to the neurons. We enrich our model by assuming that the neuron  $i$  receives also an external current density composed of a deterministic part, noted  $I_i(t)$ , and a stochastic part, noted  $n_i(t)$ , so that

$$V_i(t) = \int_{t_0}^t g_i(t-s) \left( \sum_j \bar{J}_{ij} V_j(s) + I_i(s) + n_i(s) \right) ds + V_i(t_0). \quad (3)$$

We assume, and this is essential for deriving the mean-field equations below, that all indexed quantities depend only upon the  $P$  populations of neurons (see Definition 1), i.e.,

$$\begin{aligned} g_i(t) &\stackrel{\text{def}}{=} g_{p(i)}(t) & \bar{J}_{ij} &\stackrel{\text{def}}{=} \bar{J}_{p(i)p(j)} & I_i(t) &\stackrel{\text{def}}{=} I_{p(i)}(t) \\ n_i(t) &\stackrel{\text{def}}{=} n_{p(i)}(t) & S_j(\cdot) &= S_{p(j)}(\cdot), \end{aligned} \quad (4)$$

where  $x \sim y$  indicates that the two random variables  $x$  and  $y$  have the same probability distribution. In other words, all neurons in the same population are described by identical equations (in law).

The  $g$ -shapes describe the shape of the postsynaptic potentials and can reasonably well be approximated by smooth functions.

In detail we assume that  $g_\alpha$ ,  $\alpha = 1, \dots, P$  is the Green function of a linear differential equation of order  $k$ , i.e., satisfies

$$\sum_{l=0}^k b_{l\alpha}(t) \frac{d^l g_\alpha}{dt^l}(t) = \delta(t), \quad (5)$$

where  $\delta(t)$  is the Dirac delta function.

The functions  $b_{l\alpha}(t)$ ,  $l = 0, \dots, k$ ,  $\alpha = 1, \dots, P$ , are assumed to be continuous. We also assume for simplicity that

$$b_{k\alpha}(t) \equiv c_\alpha \neq 0, \quad (6)$$

for all  $t \in \mathbb{R}$ ,  $\alpha = 1, \dots, P$ . We note  $D_\alpha^k$  the corresponding differential operator:

$$D_\alpha^k g_\alpha(t) \stackrel{\text{def}}{=} \sum_{l=0}^k b_{l\alpha}(t) \frac{d^l g_\alpha}{dt^l}(t) = \delta(t) \quad (7)$$

Applying  $D_\alpha^k$  to both sides of Eq. 3, using Eq. 7 and the fact that  $v_j(s) = S_j(V_j(s))$ , we obtain a  $k$ th-order differential equation for  $V_i$

$$D_i^k V_i(t) = \sum_{j=1}^N \bar{J}_{ij} S_j(V_j(t)) + I_i(t) + n_i(t). \quad (8)$$

With a slight abuse of notation, we split the sum with respect to  $j$  into  $P$  sums:

$$D_i^k V_i(t) = \sum_{\beta=1}^P \sum_{j=1}^{N_\beta} \bar{J}_{ij} S_j(V_j(t)) + I_i(t) + n_i(t)$$

We classically turn the  $k$ th-order differential Eq. 8 into a  $k$ -dimensional system of coupled first-order differential equations (we divided both sides of the last equation by  $c_p$ , see Eq. 6):

$$\begin{aligned} dV_{li}(t) &= V_{l+1i}(t) dt \quad l = 0, \dots, k-2 \\ dV_{k-1i}(t) &= \left( -\sum_{l=0}^{k-1} b_{lp(i)}(t) V_{li}(t) + \sum_j \bar{J}_{ij} S_{p(j)}(V_j(t)) + I_{p(i)}(t) + n_i(t) \right) dt \end{aligned} \quad (9)$$

A well-known example of  $g$ -shapes, see Section “Example II: The model of Jansen and Rit” below or Gerstner and Kistler (2002), is

$$g(t) = Ke^{-t/\tau} Y(t), \quad (10)$$

where  $Y(t)$  is the Heaviside function. This is an exponentially decaying postsynaptic potential corresponding to

$$k=1 \quad b_1(t) = \frac{1}{K} \quad \text{and} \quad b_0(t) = \frac{1}{K\tau}$$

in Eq. 5.

Another well-known example is

$$g(t) = Kte^{-t/\tau} Y(t). \quad (11)$$

This is a somewhat smoother function corresponding to

$$k=2 \quad b_2(t) = \frac{1}{K} \quad b_1(t) = \frac{2}{\tau} \quad \text{and} \quad b_0(t) = \frac{1}{\tau^2}$$

in Eq. 5.

**The dynamics.** We modify the Eq. 9 by perturbing the first  $k-1$  equations with Brownian noise and assuming that  $n_i(t)$  is white noise. This has the effect that the quantities that appear in Eq. 9 are not anymore the derivatives up to order  $k-1$  of  $V_i$ . This becomes true again only in the limit where the added Brownian noise is null. This may seem artificial at first glance but (1) it is a technical assumption that is necessary in the proofs of the well-posedness of the mean-field equations, see Assumption 1 below, and (2) it generates a rich class of external stochastic input, as shown below. With this in mind, the Eq. 9 now read

$$\begin{aligned} dV_{li}(t) &= V_{l+1i}(t) dt + f_{li}(t) dW_{li}(t) \quad l = 0, \dots, k-2 \\ dV_{k-1i}(t) &= \left( -\sum_{l=0}^{k-1} b_{lp(i)}(t) V_{li}(t) + \sum_j \bar{J}_{ij} S_{p(j)}(V_j(t)) + I_{p(i)}(t) \right) dt \\ &\quad + f_{k-1i}(t) dW_{k-1i}(t) \end{aligned} \quad (12)$$

$W_{li}(t)$ ,  $l = 0, \dots, k-1$ ,  $i = 1, \dots, N$ , are  $kN$  independent standard Brownian processes. Because we want the neurons in the same class to be essentially identical we also assume that the functions  $f_{li}(t)$  that control the amount of noise on each derivative satisfy

$$f_{li}(t) = f_{lp(i)}(t), \quad l = 0, \dots, k-1, \quad i = 1, \dots, N$$

Note that in the limit  $f_{l\alpha}(t) = 0$  for  $l = 0, \dots, k-1$  and  $\alpha = 1, \dots, P$ , the components  $V_{li}(t)$  of the vector  $\tilde{\mathbf{V}}_i(t)$  are the derivatives of the membrane potential  $V_p$  for  $l = 0, \dots, k-1$  and the Eq. 12 turn into Eq. 9. The system of differential Eq. 12 implies that the class of admissible external stochastic input  $n_i(t)$  to the neuron  $i$  are Brownian noise integrated through the filter of the synapse, i.e., involving the  $l$ th primitives of the Brownian motion for  $l \leq k$ .

We now introduce the  $k-1$   $N$ -dimensional vectors  $V_l(t) = [V_{1l}, \dots, V_{Nl}]^T$ ,  $l = 1, \dots, k-1$  of the  $l$ th-order derivative (in the limit of  $f_{lp(i)}(t) = 0$ ) of  $\mathbf{V}(t)$ , and concatenate them with  $\mathbf{V}(t)$  into the  $Nk$ -dimensional vector

$$\tilde{\mathbf{V}}(t) = \begin{bmatrix} \mathbf{V}(t) \\ \mathbf{V}_1(t) \\ \vdots \\ \mathbf{V}_{k-1}(t) \end{bmatrix}. \quad (13)$$

The  $N$ -neurons network is described by the  $Nk$ -dimensional vector  $\tilde{\mathbf{V}}(t)$ . By definition the  $l$ th  $N$ -dimensional component  $\tilde{V}_l$  of  $\tilde{\mathbf{V}}$  is equal to  $V_l$ . In the limit  $f_{i\alpha}(t) = 0$  we have

$$\tilde{V}_l = V_l = \frac{d^l \mathbf{V}}{dt^l} \quad l = 0, \dots, k-1, \quad \text{with} \quad \tilde{\mathbf{V}}_0 = \mathbf{V}$$

We next write the equations governing the time variation of the  $kN$ -dimensional sub-vectors of  $\tilde{\mathbf{V}}(t)$ , i.e., the derivatives of order  $0, \dots, k-1$  of  $V(t)$ . These are vector versions of Eq. 12. We write

$$d\tilde{\mathbf{V}}_l(t) = \tilde{\mathbf{V}}_{l+1}(t)dt + \mathbf{F}_l(t) \cdot d\mathbf{W}_l(t) \quad l = 0, \dots, k-2. \quad (14)$$

$\mathbf{F}_l(t)$  is the  $N \times N$  diagonal matrix

$$\text{diag} \left( \underbrace{f_{i1}(t), \dots, f_{i1}(t)}_{N_1}, \dots, \underbrace{f_{iP}(t), \dots, f_{iP}(t)}_{N_P} \right),$$

where  $f_{i\alpha}(t)$ ,  $\alpha = 1, \dots, P$  is repeated  $N_\alpha$  times, and the  $W_l(t)$ ,  $l = 0, \dots, k-2$ , are  $k-1$   $N$ -dimensional independent standard Brownian processes.

The equation governing the  $(k-1)$ th differential of the membrane potential has a linear part determined by the differential operators  $D_\alpha^k$ ,  $\alpha = 1, \dots, P$  and accounts for the external inputs (deterministic and stochastic) and the activity of the neighbors. We note  $\mathcal{L}(t)$  the  $N \times Nk$  matrix describing the relation between the neurons membrane potentials and their derivatives up to the order  $k-1$  and the  $(k-1)$ th derivative of  $\mathbf{V}$ . This matrix is defined as the concatenation of the  $kN \times N$  diagonal matrices

$$\mathbf{B}_l(t) = \text{diag} \left( \underbrace{b_{i1}(t), \dots, b_{i1}(t)}_{N_1}, \dots, \underbrace{b_{iP}(t), \dots, b_{iP}(t)}_{N_P} \right)$$

for  $l = 0, \dots, k-1$ :

$$\mathcal{L}(t) = [\mathbf{B}_0(t), \dots, \mathbf{B}_{k-1}(t)]$$

We have:

$$d\tilde{\mathbf{V}}_{k-1}(t) = \left( -\mathcal{L}(t) \cdot \tilde{\mathbf{V}}(t) + (\bar{\mathbf{J}} \cdot S(\tilde{\mathbf{V}}_0(t))) + \mathbf{I}(t) \right) dt + \mathbf{F}_{k-1}(t) \cdot d\mathbf{W}_{k-1}(t), \quad (15)$$

where  $\mathbf{W}_{k-1}(t)$  is an  $N$ -dimensional standard Brownian process independent of  $\mathbf{W}_l(t)$ ,  $l = 0, \dots, k-2$ . The coordinates of the  $N$ -dimensional vector  $\mathbf{I}(t)$  are the external deterministic input currents,

$$\mathbf{I}(t) = \left[ \underbrace{I_1(t), \dots, I_1(t)}_{N_1}, \dots, \underbrace{I_P(t), \dots, I_P(t)}_{N_P} \right]^T,$$

$\bar{\mathbf{J}}$  the  $N \times N$  matrix of the weights  $\bar{J}_{ij}$  which are equal to  $\bar{J}_{p(i)p(j)}$  (see Eq. 4), and  $S$  is a mapping from  $\mathbb{R}^N$  to  $\mathbb{R}^N$  such that

$$S(\mathbf{V})_i = S_{p(i)}(V_i) \text{ for } i = 1, \dots, N. \quad (16)$$

We define

$$\mathbf{L}(t) = \begin{bmatrix} 0_{N \times N} & \text{Id}_N & \cdots & 0_{N \times N} \\ 0_{N \times N} & 0_{N \times N} & \ddots & 0_{N \times N} \\ \vdots & \vdots & & \text{Id}_N \\ \mathbf{B}_0(t) & \mathbf{B}_1(t) & \cdots & \mathbf{B}_{k-1}(t) \end{bmatrix},$$

where  $\text{Id}_N$  is the  $N \times N$  identity matrix and  $0_{N \times N}$  the  $N \times N$  null matrix. We also define the two  $kN$ -dimensional vectors:

$$\tilde{\mathbf{U}}_t = \begin{bmatrix} 0_N \\ \vdots \\ 0_N \\ \bar{\mathbf{J}} \cdot S(\tilde{\mathbf{V}}_0(t)) \end{bmatrix} = \begin{bmatrix} 0_N \\ \vdots \\ 0_N \\ \bar{\mathbf{J}} \cdot S(\mathbf{V}(t)) \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{I}}_t = \begin{bmatrix} 0_N \\ \vdots \\ 0_N \\ \mathbf{I}(t) \end{bmatrix},$$

where  $0_N$  is the  $N$ -dimensional null vector.

Combining Eqs. 14 and 15 the full equation satisfied by  $\tilde{\mathbf{V}}$  can be written:

$$d\tilde{\mathbf{V}}(t) = \left( -\mathbf{L}(t)\tilde{\mathbf{V}}(t) + \tilde{\mathbf{U}}_t + \tilde{\mathbf{I}}_t \right) dt + \mathbf{F}(t) \cdot d\mathbf{W}_t, \quad (17)$$

where the  $kN \times kN$  matrix  $\mathbf{F}(t)$  is equal to  $\text{diag}(\mathbf{F}_0, \dots, \mathbf{F}_{k-1})$  and  $\mathbf{W}_t$  is an  $kN$ -dimensional standard Brownian process.

### THE MEAN-FIELD EQUATIONS

One of the central goals of this paper is to analyze what happens when we let the total number  $N$  of neurons grow to infinity. Can we “summarize” the  $kN$  equations (Eq. 17) with a smaller number of equations that would account for the populations activity? We show that the answer to this question is yes and that the populations activity can indeed be represented by  $P$  stochastic differential equations of order  $k$ . Despite the fact that their solutions are Gaussian processes, these equations turn out to be quite complicated because these processes are non-Markovian.

We assume that the proportions of neurons in each population are nontrivial, i.e.:

$$\lim_{N \rightarrow \infty} \frac{N_\alpha}{N} = n_\alpha \in (0,1) \forall \alpha \in \{1, \dots, P\}, \quad \text{and} \quad \sum_\alpha n_\alpha = 1.$$

If it were not the case the corresponding population would not affect the global behavior of the system, would not contribute to the mean-field equation, and could be neglected.

### General derivation of the mean-field equation

When investigating the structure of such mesoscopic neural assemblies as cortical columns, experimentalists are able to provide the average value  $\bar{J}_{ij}$  of the synaptic efficacy  $J_{ij}$  of neural population  $j$  to population  $i$ . These values are obviously subject to some uncertainty which can be modeled as Gaussian random variables. We also impose that the distribution of the  $J_{ij}$ s depends only on the population pair  $\alpha = p(i)$ ,  $\beta = p(j)$ , and on the total number of neurons  $N_\beta$  of population  $\beta$ :

$$J_{ij} \sim \mathcal{N} \left( \frac{\bar{J}_{\alpha\beta}}{N_\beta}, \frac{\sigma_{\alpha\beta}}{\sqrt{N_\beta}} \right). \quad (18)$$

We also make the additional assumption that the  $J_{ij}$ s are independent. This is a reasonable assumption as far as modeling cortical columns from experimental data is concerned. Indeed, it is already difficult for experimentalists to provide the average value of the synaptic strength  $\bar{J}_{\alpha\beta}$  from population  $\beta$  to population  $\alpha$  and to estimate the corresponding error bars ( $\sigma_{\alpha\beta}$ ), but measuring synaptic efficacies correlations in a large assembly of neurons seems currently out of reach. Though, it is known that synaptic weights are indeed correlated (e.g., via synaptic plasticity mechanisms), these correlations are built by dynamics via a complex interwoven evolution between neurons and synapses dynamics and postulating



the form of synaptic weights correlations requires, on theoretical grounds, a detailed investigation of the whole history of neurons–synapses dynamics.

Let us now discuss the scaling form of the probability distribution (Eq. 18) of the  $J_{ij}$ 's, namely the division by  $N_\beta$  for the mean and variance of the Gaussian distribution. This scaling ensures that the “local interaction field”  $\sum_{j=1}^{N_\beta} J_{ij} S(V_j(t))$  summarizing the effects of the neurons in population  $\beta$  on neuron  $i$ , has a mean and variance which do not depend on  $N_\beta$  and is only controlled by the phenomenological parameters  $\bar{J}_{\alpha\beta}, \sigma_{\alpha\beta}$ .

We are interested in the limit law when  $N \rightarrow \infty$  of the  $N$ -dimensional vector  $\mathbf{V}$  defined in Eq. 3 under the joint law of the connectivities and the Brownian motions, which we call the mean-field limit. This law can be described by a set of  $P$  equations, the mean-field equations. As mentioned in the introduction these equations can be derived in several ways, either heuristically as in the work of Amari (Amari, 1972; Amari et al., 1977), Sompolinsky (Crisanti et al., 1990; Sompolinsky et al., 1998), and Cessac (Cessac, 1995; Samuelides and Cessac, 2007), or rigorously as in the work of Ben-Arous and Guionnet (Ben-Arous and Guionnet, 1995, 1997; Guionnet, 1997). We derive them here in a pedestrian way, prove that they are well-posed, and provide an algorithm for computing their solution.

The effective description of the network population by population is possible because the neurons in each population are interchangeable, i.e., have the same probability distribution under the joint law of the multidimensional Brownian motion and the connectivity weights. This is the case because of the relations (Eqs. 4 and 16) which imply the form of Eq. 17.

**The mean ideas of dynamic mean-field equations.** Before diving into the mathematical developments let us comment briefly what are the basic ideas and conclusions of the mean-field approach. Following Eq. 8, the evolution of the membrane potential of some neuron  $i$  in population  $\alpha$  is given by:

$$\sum_{l=0}^k b_{l\alpha}(t) \frac{d^l V_i(t)}{dt^l} = \sum_{j=1}^N J_{ij} S_j(V_j(t)) + I_\alpha(t) + n_i(t), p(i) = \alpha. \quad (19)$$

Using the assumption that  $S_j, I_j, n_j$  depend only on neuron population, this gives:

$$\sum_{l=0}^k b_{l\alpha}(t) \frac{d^l V_i(t)}{dt^l} = \sum_{\beta=1}^P \eta_{i\beta}(V(t)) + I_\alpha(t) + n_i(t), i \in \alpha, \quad (20)$$

where we have introduced the local interaction field  $\eta_{i\beta}(V(t)) = \sum_{j=1}^{N_\beta} J_{ij} S_\beta(V_j(t))$ , summarizing the effects of neurons in population  $\beta$  on neuron  $i$  and whose probability distribution only depends on the pre- and postsynaptic populations  $\alpha$  and  $\beta$ .

In the simplest situation where the  $J_{ij}$ 's have no fluctuations ( $\sigma_{\alpha\beta} = 0$ ) this field reads  $\eta_{i\beta}(V(t)) = \bar{J}_{\alpha\beta} \Phi_\beta(V(t))$ . The term  $\Phi_\beta(V(t)) = \frac{1}{N_\beta} \sum_{j=1}^{N_\beta} S_\beta(V_j(t))$  is the frequency rate of neurons in population  $\beta$ , averaged over this population. Introducing in the same way the average membrane potential in population  $\beta$ ,  $V_\beta(t) = \frac{1}{N_\beta} \sum_{j=1}^{N_\beta} V_j(t)$ , one obtains:

$$\sum_{l=0}^k b_{l\alpha}(t) \frac{d^l V_\alpha(t)}{dt^l} = \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \Phi_\beta(V(t)) + I_\alpha(t) + n_\alpha(t). \quad (21)$$

This equation resembles very much Eq. 19 if one makes the following reasoning: “Since  $\Phi_\beta(V(t))$  is the frequency rate of neurons in population  $\beta$ , averaged over this population, and since, for one neuron, the frequency rate is  $v_i(t) = S_i(V_i(t))$  let us write  $\Phi_\beta(V(t)) = S_\beta(V_\beta(t))$ ”. This leads to:

$$\sum_{l=0}^k b_{l\alpha}(t) \frac{d^l V_\alpha(t)}{dt^l} = \sum_{\beta=1}^P \bar{J}_{\alpha\beta} S_\beta(V_\beta(t)) + I_\alpha(t) + n_\alpha(t), \quad (22)$$

which has exactly the same form as Eq. 19 but at the level of a neuron population. Equations such as (22), which are obtained via a very strong assumption:

$$\frac{1}{N_\beta} \sum_{j=1}^{N_\beta} S_\beta(V_j(t)) = S_\beta\left(\frac{1}{N_\beta} \sum_{j=1}^{N_\beta} V_j(t)\right), \quad (23)$$

are typically those obtained by Jansen and Rit (1995). Surprisingly, they are correct and can be rigorously derived, as discussed below, provided  $\sigma_{\alpha\beta} = 0$ .

However, they cannot remain true, as soon as the synaptic weights fluctuate. Indeed, the transition from Eqs. 19 to 22 corresponds to a projection from a  $NP$ -dimensional space to a  $P$ -dimensional one, which holds because the  $NP \times NP$  dimensional synaptic weights matrix has in fact only  $P$  linearly independent rows. This does not hold anymore if the  $J_{ij}$ 's are random and the synaptic weights matrix has generically full rank. Moreover, the effects of the nonlinear dynamics on the synaptic weights variations about their mean, is not small even if the  $\sigma_{\alpha\beta}$ 's are and the real trajectories of Eq. 19 can depart strongly from the trajectories of Eq. 22. This is the main message of this paper.

To finish this qualitative description, let us say in a few words what happens to the mean-field equations when  $\sigma_{\alpha\beta} \neq 0$ . We show below that the local interaction fields  $\eta_{i\alpha\beta}(V(t))$  becomes, in the limit  $N_\beta \rightarrow \infty$ , a time-dependent Gaussian field  $U_{\alpha\beta}(t)$ . One of the main results is that this field is non-Markovian, i.e., it integrates the whole history, via the synaptic responses  $g$  which are convolution products. Despite the fact that the evolution equation for the membrane potential averaged over a population writes in a very simple form:

$$\sum_{l=0}^k b_{l\alpha}(t) \frac{d^l V_\alpha(t)}{dt^l} = \sum_{\beta=1}^P U_{\alpha\beta}(t) + I_\alpha(t) + n_\alpha(t), \quad (24)$$

it hides a real difficulty, since  $U_{\alpha\beta}(t)$  depends on the whole past. Therefore, *the introduction of synaptic weights variability leads to a drastic change in neural mass models, as we now develop.*

**The Mean-Field equations.** We note  $C([t_0, T], \mathbb{R}^P)$  (respectively  $C((-\infty, T], \mathbb{R}^P)$ ) the set of continuous functions from the real interval  $[t_0, T]$  (respectively  $(-\infty, T]$ ) to  $\mathbb{R}^P$ . By assigning a probability to subsets of such functions, a continuous stochastic process  $X$  defines a positive measure of unit mass on  $C([t_0, T], \mathbb{R}^P)$  (respectively  $C((-\infty, T], \mathbb{R}^P)$ ). This set of positive measures of unit mass is noted  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^P))$  (respectively  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P))$ ).

We now define a process of particular importance for describing the limit process: the effective interaction process.

**Definition 2.** (Effective interaction process). Let  $X \in \mathcal{M}_1^+(C([t_0, T], \mathbb{R}^P))$  (respectively  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P))$ ) be a given



Gaussian stochastic process. The effective interaction term is the Gaussian process  $\mathbf{U}^X \in \mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{P \times P}))$ , (respectively  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^{P \times P}))$ ) defined by:

$$\begin{cases} \mathbb{E}[U_{\alpha\beta}^X(t)] = \bar{J}_{\alpha\beta} m_{\beta}^X(t) \\ \text{Cov}(U_{\alpha\beta}^X(t), U_{\gamma\delta}^X(s)) = \begin{cases} \alpha_{\alpha\beta}^2 \Delta_{\beta}^X(t, s) & \text{if } \alpha = \gamma \text{ and } \beta = \delta, \\ 0 & \text{otherwise} \end{cases} \end{cases} \quad (25)$$

where

$$m_{\beta}^X(t) \stackrel{\text{def}}{=} \mathbb{E}\left[S_{\beta}(X_{\beta}(t))\right],$$

and

$$\Delta_{\beta}^X(t, s) \stackrel{\text{def}}{=} \mathbb{E}[S_{\beta}(X_{\beta}(t)) S_{\beta}(X_{\beta}(s))]$$

In order to construct the solution of the mean-field equations (see Section ‘‘Existence and Uniqueness of Solutions in Finite Time’’) we will need more explicit expressions for  $m_{\beta}^X(t)$  and  $\Delta_{\beta}^X(t, s)$  which we obtain in the next proposition.

**Proposition 1.** Let  $\mu(t) = \mathbb{E}[X_t]$  be the mean of the process  $X$  and  $C(t, s) = \mathbb{E}[(X_t - \mu(t))(X_s - \mu(s))^T]$  be its covariance matrix. The vectors  $\mathbf{m}^X(t)$  and  $\Delta^X(t, s)$  that appear in the definition of the effective interaction process  $\mathbf{U}^X$  are defined by the following expressions:

$$m_{\beta}^X(t) = \int_{\mathbb{R}} S_{\beta}(x \sqrt{C_{\beta\beta}(t, t)} + \mu_{\beta}(t)) D\mathbf{x}, \quad (26)$$

and

$$\begin{aligned} \Delta_{\beta}^X(t, s) = \int_{\mathbb{R}^2} & \left( \frac{\sqrt{C_{\beta\beta}(t, t) C_{\beta\beta}(s, s) - C_{\beta\beta}(t, s)^2}}{\sqrt{C_{\beta\beta}(t, t)}} x \right. \\ & \left. + \frac{C_{\beta\beta}(t, s)}{\sqrt{C_{\beta\beta}(t, t)}} y + \mu_{\beta}(s) \right) S_{\beta}(y \sqrt{C_{\beta\beta}(t, t)} + \mu_{\beta}(t)) D\mathbf{x} D\mathbf{y}, \end{aligned} \quad (27)$$

where

$$D\mathbf{x} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

is the probability density of a 0-mean, unit variance, Gaussian variable.

*Proof.* The results follow immediately by a change of variable from the fact that  $X_{\beta}(t)$  is a univariate Gaussian random variable of mean  $\mu_{\beta}(t)$  and variance  $C_{\beta\beta}(t, t)$  and the pair  $(X_{\beta}(t), X_{\beta}(s))$  is bivariate Gaussian random variable with mean  $(\mu_{\beta}(t), \mu_{\beta}(s))$  and covariance matrix

$$\begin{bmatrix} C_{\beta\beta}(t, t) & C_{\beta\beta}(t, s) \\ C_{\beta\beta}(t, s) & C_{\beta\beta}(s, s) \end{bmatrix} \quad \square$$

Choose  $P$  neurons  $i_1, \dots, i_p$ , one in each population (neuron  $i_{\alpha}$  belongs to the population  $\alpha$ ). We define the  $kP$ -dimensional vector  $\tilde{\mathcal{V}}^{(N)}(t)$  by choosing, in each of the  $k$   $N$ -dimensional components  $\tilde{\mathcal{V}}_l(t), l = 0, \dots, k-1$ , of the vector  $\tilde{\mathcal{V}}(t)$  defined in Eq. 13 the coordinates of indexes  $i_1, \dots, i_p$ . Then it can be shown, using either a

heuristic argument or large deviations techniques (see Appendix A), that the sequence of  $kP$ -dimensional processes  $(\tilde{\mathcal{V}}_{t \geq t_0}^{(N)})_{N \geq 1}$  converges in law to the process  $\tilde{\mathcal{V}}(t) = [\mathcal{V}(t)^T, \mathcal{V}_1(t)^T, \dots, \mathcal{V}_{k-1}(t)^T]^T$  solution of the following mean-field equation:

$$d\tilde{\mathcal{V}}(t) = (-\mathbf{L}(t)\tilde{\mathcal{V}}(t) + \tilde{\mathbf{U}}_t^{\mathcal{V}} + \tilde{\mathbf{I}}(t))dt + \mathbf{F}(t) \cdot d\mathbf{W}_t. \quad (28)$$

$\mathbf{L}$  is the  $P_k \times P_k$  matrix

$$\mathbf{L}(t) = \begin{bmatrix} \mathbf{0}_{P \times P} & \mathbf{Id}_p & \cdots & \mathbf{0}_{P \times P} \\ \mathbf{0}_{P \times P} & \mathbf{0}_{P \times P} & \ddots & \mathbf{0}_{P \times P} \\ \vdots & \vdots & & \mathbf{Id}_p \\ \mathbf{B}_0(t) & \mathbf{B}_1(t) & \cdots & \mathbf{B}_{k-1}(t) \end{bmatrix}$$

The  $P \times P$  matrixes  $\mathbf{B}_l(t), l = 0, \dots, k-1$  are, with a slight abuse of notations, equal to  $\text{diag}(b_{l1}(t), \dots, b_{lp}(t))$ .  $(\mathbf{W}_t)_{t \geq t_0}$  is a  $kP$ -dimensional standard Brownian process.  $\tilde{\mathbf{U}}^{\mathcal{V}}$  has the law of the  $P$ -dimensional effective interaction vector associated to the vector  $\mathcal{V}$  (first  $P$ -dimensional component of  $\tilde{\mathcal{V}}$ ) and is statistically independent of the external noise  $(\mathbf{W}_t)_{t \geq t_0}$  and of the initial condition  $\tilde{\mathcal{V}}(t_0)$  (when  $t_0 > -\infty$ ):

$$\tilde{\mathbf{U}}_t^{\mathcal{V}} = \begin{bmatrix} \mathbf{0}_P \\ \vdots \\ \mathbf{0}_P \\ \mathbf{U}_t^{\mathcal{V}} \cdot \mathbf{1} \end{bmatrix} \quad \tilde{\mathbf{I}}(t) = \begin{bmatrix} \mathbf{0}_P \\ \vdots \\ \mathbf{0}_P \\ \mathbf{I}(t) \end{bmatrix} \quad \mathbf{F}(t) = \text{diag}(\mathbf{F}_0(t), \dots, \mathbf{F}_{k-1}(t)).$$

We have used for the matrixes  $\mathbf{F}_l(t), l = 0, \dots, k-1$  the same abuse of notations as for the matrixes  $\mathbf{B}_l(t)$ , i.e.,  $\mathbf{F}_l(t) = \text{diag}(f_{l1}(t), \dots, f_{lp}(t))$  for  $l = 0, \dots, k-1$ .  $\mathbf{I}(t)$  is the  $P$ -dimensional external current  $[I_1(t), \dots, I_p(t)]^T$ .

The process  $(\mathbf{U}_t^{\mathcal{V}})_{t \geq t_0}$  is a  $P \times P$ -dimensional process and is applied, as a matrix, to the  $P$ -dimensional vector  $\mathbf{1}$  with all coordinates equal to 1, resulting in the  $P$ -dimensional vector  $\mathbf{U}_t^{\mathcal{V}} \cdot \mathbf{1}$  whose mean and covariance function can be readily obtained from Definition 2:

$$\mathbb{E}[\mathbf{U}_t^{\mathcal{V}} \cdot \mathbf{1}] = \sum_{\beta=1}^P \bar{J}_{\alpha\beta} m_{\beta}^{\mathcal{V}}(t), \quad m_{\beta}^{\mathcal{V}}(t) = \mathbb{E}\left[S_{\beta}(\mathcal{V}_{\beta}(t))\right] \quad (29)$$

and

$$\text{Cov}\left(\left(\mathbf{U}_t^{\mathcal{V}} \cdot \mathbf{1}\right)_{\alpha} \left(\mathbf{U}_s^{\mathcal{V}} \cdot \mathbf{1}\right)_{\gamma}\right) = \begin{cases} \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \Delta_{\beta}^{\mathcal{V}}(t, s) & \text{if } \alpha = \gamma \\ 0 & \text{otherwise} \end{cases} \quad (30)$$

We have of course

$$\Delta_{\beta}^{\mathcal{V}}(t, s) = \mathbb{E}\left[S_{\beta}(\mathcal{V}_{\beta}(t)) S_{\beta}(\mathcal{V}_{\beta}(s))\right]$$

Equations (28) are formally very similar to Eq. 17 but there are some very important differences. The first ones are of dimension  $kP$  whereas the second are of dimension  $kN$  which grows arbitrarily large when  $N \rightarrow \infty$ . The interaction term of the second,  $\bar{\mathbf{J}} \cdot \mathbf{S}(\mathbf{V}(t))$ , is simply the synaptic weight matrix applied to the activities of the  $N$  neurons at time  $t$ . The interaction term of the first equation,  $\tilde{\mathbf{U}}_t^{\mathcal{V}}$ , though innocuous looking, is in fact quite complex (see Eqs. 29 and 30). In fact the stochastic process  $\tilde{\mathbf{U}}_t^{\mathcal{V}}$ , putative solution of Eq. 28, is in general non-Markovian.

To proceed further we formally integrate the equation using the flow, or resolvent, of the Eq. 28, noted  $\Phi_L(t, t_0)$  (see Appendix B), and we obtain, since we assumed  $\mathbf{L}$  continuous, an implicit representation of  $\tilde{\mathcal{V}}(t)$ :

$$\begin{aligned} \tilde{\mathcal{V}}(t) &= \Phi_L(t, t_0) \tilde{\mathcal{V}}(t_0) + \int_{t_0}^t \Phi_L(t, s) \cdot (\tilde{\mathbf{U}}_s^\vee + \tilde{\mathbf{I}}(s)) ds \\ &\quad + \int_{t_0}^t \Phi_L(t, s) \cdot \mathbf{F}(s) \cdot d\mathbf{W}_s, \end{aligned} \tag{31}$$

We now introduce for future reference a simpler model which is quite frequently used in the description on neural networks and has been formally analyzed by Sompolinsky and colleagues (Crisanti et al., 1990; Sompolinsky et al., 1998) in the case of one population ( $P = 1$ ).

**Example I: The Simple Model**

In the Simple Model, each neuron membrane potential decreases exponentially to its rest value if it receives no input, with a time constant  $\tau_\alpha$  depending only on the population. In other words, we assume that the g-shape describing the shape of the PSPs is Eq. 10, with  $K = 1$  for simplicity. The noise is modeled by an independent Brownian process per neuron whose standard deviation is the same for all neurons belonging to a given population.

Hence the dynamics of a given neuron  $i$  from population  $\alpha$  of the network reads:

$$\begin{aligned} dV_i(t) &= \left[ -\frac{V_i(t)}{\tau_{p(i)}} + \sum_{\beta=1}^P \sum_{j=1}^{N_\beta} J_{ij} S_{p(j)}(V_j(t)) + I_{p(i)}(t) \right] dt \\ &\quad + f_{p(i)} dW_i(t). \end{aligned} \tag{32}$$

This is a special case of Eq. 12 where  $k = 1, b_{0\alpha}(t) = 1/\tau_\alpha, b_{1\alpha}(t) = 1$  for  $\alpha = 1, \dots, P$ . The corresponding mean-field equation reads:

$$\begin{aligned} d\mathcal{V}_\alpha(t) &= \left( -\frac{\mathcal{V}_\alpha(t)}{\tau_\alpha} + \sum_{\beta=1}^P U_{\alpha\beta}^\vee(t) + I_\alpha(t) \right) dt + f_\alpha dW_\alpha(t), \\ \forall \alpha \in \{1, \dots, P\}, \end{aligned} \tag{33}$$

where the processes  $(W_\alpha(t))_{t \geq t_0}$  are independent standard Brownian motions,  $\mathbf{U}^\vee(t) = (U_{\alpha\beta}^\vee(t); \alpha, \beta \in \{1, \dots, P\})$ , is the effective interaction term, see Definition 2. This is a special case of Eq. 28 with  $\mathbf{L} = \text{diag}(\frac{1}{\tau_1}, \dots, \frac{1}{\tau_P})$ , and  $\mathbf{F} = \text{diag}(f_1, \dots, f_P)$ .

Taking the expected value of both sides of Eq. 33 and using we obtain Eq. 26 that the mean  $\mu_\alpha(t)$  of  $\mathcal{V}_\alpha(t)$  satisfies the differential equation

$$\frac{d\mu_\alpha(t)}{dt} = -\frac{\mu_\alpha(t)}{\tau_\alpha} + \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \int_{\mathbb{R}} S_\beta(x\sqrt{C_{\beta\beta}(t,t)} + \mu_\beta(t)) Dx + I_\alpha(t),$$

If  $C_{\beta\beta}(t, t)$  vanishes for all  $t \geq t_0$  this equation reduces to:

$$\frac{d\mu_\alpha(t)}{dt} = -\frac{\mu_\alpha(t)}{\tau_\alpha} + \sum_{\beta=1}^P \bar{J}_{\alpha\beta} S_\beta(\mu_\beta(t)) + I_\alpha(t),$$

which is precisely the “naive” mean-field equation (Eq. 22) obtained with the assumption (Eq. 23). We see that Eq. 22 are indeed correct, provided that  $C_{\beta\beta}(t, t) = 0, \forall t \geq t_0$ .

Equation 33 can be formally integrated implicitly and we obtain the following integral representation of the process  $\mathcal{V}_\alpha(t)$ :

$$\begin{aligned} \mathcal{V}_\alpha(t) &= e^{-\frac{(t-t_0)}{\tau_\alpha}} \mathcal{V}_\alpha(t_0) + \int_{t_0}^t e^{-\frac{(t-s)}{\tau_\alpha}} \left( \sum_{\beta=1}^P U_{\alpha\beta}^\vee(s) + I_\alpha(s) \right) ds \\ &\quad + f_\alpha \int_{t_0}^t e^{-\frac{(t-s)}{\tau_\alpha}} dW_\alpha(s) \end{aligned} \tag{34}$$

where  $t_0$  is the initial time. It is an implicit equation on the probability distribution of  $\mathcal{V}(t)$ , a special case of (Eq. 31), with  $\Phi_L(t, t_0) = \text{diag}(e^{-(t-t_0)/\tau_1}, \dots, e^{-(t-t_0)/\tau_P})$ .

The variance  $C_{\alpha\alpha}(t, t)$  of  $\mathcal{V}_\alpha(t)$  can easily be obtained from Eq. 34. It reads

$$\begin{aligned} C_{\alpha\alpha}(t, t) &= e^{-\frac{2t}{\tau_\alpha}} \left[ C_{\alpha\alpha}(t_0, t_0) + \frac{\tau_\alpha f_\alpha^2}{2} \left( e^{\frac{2t}{\tau_\alpha}} - 1 \right) \right. \\ &\quad \left. + \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \int_{t_0}^t \int_{t_0}^t e^{-\frac{(u+v)}{\tau_\alpha}} \Delta_\beta(u, v) dudv \right], \end{aligned}$$

where  $\Delta_\beta(u, v)$  is given by Eq. 27.

If  $\sigma_{\alpha\beta} = 0$  and if  $s_\alpha = 0$  then  $C_{\alpha\alpha}(t, t) = 0, \forall t \geq t_0$  is a solution of this equation. Thus, mean-field equations for the simple model reduce to the naive mean-field Eq. 22 in this case. This conclusion extends as well to all models of synaptic responses, ruled by Eq. 5.

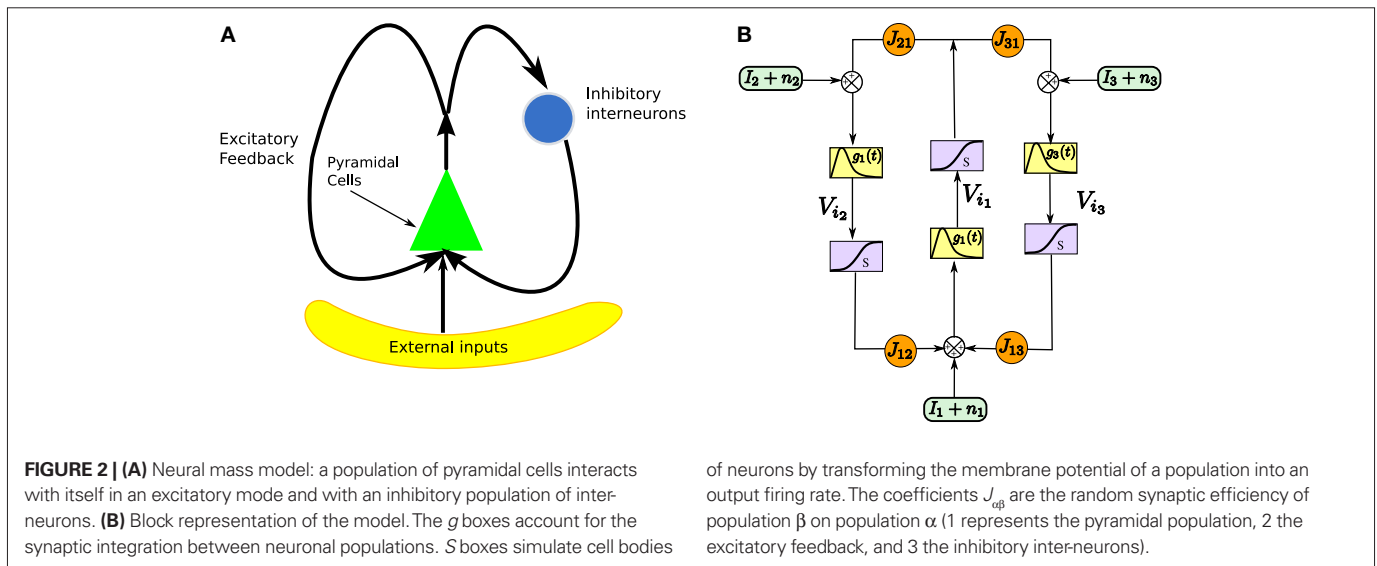
However, the equation of  $C_{\alpha\alpha}(t, t)$  shows that, in the general case, in order to solve the differential equation for  $\mu_\alpha(t)$ , we need to know the whole past of the process  $\mathcal{V}$ . This exemplifies a previous statement on the non-Markovian nature of the solution of the mean-field equations.

**Example II: The model of Jansen and Rit**

One of the motivations of this study is to characterize the global behavior of an assembly of neurons in particular to get a better understanding of recordings of cortical signals like EEG or MEG. One of the classical models of neural masses is Jansen and Rit’s mass model (Jansen and Rit, 1995), in short the JR model (see **Figure 2**).

The model features a population of pyramidal neurons that receives inhibitory inputs from local inter-neurons, excitatory feed-backs, and excitatory inputs from neighboring cortical units and sub-cortical structures such as the thalamus. The excitatory input is represented by an external firing rate that has a deterministic part  $I_1(t)$  accounting for specific activity of other cortical units and a stochastic part  $n_1(t)$  accounting for a non specific background activity. We formally consider that the excitatory feedback of the pyramidal neurons is a new neural population, making the number  $P$  of populations equal to 3. We also represent the external inputs to the other two populations by the sum of a deterministic part  $I_j(t)$  and a stochastic part  $n_j(t), j = 2, 3$ , see **Figure 2**.

In the model introduced originally by Jansen and Rit, the connectivity weights were assumed to be constant, i.e., equal to their mean value. Nevertheless, there exists a variability of these coefficients, and as we show in the sequel, the effect of the connectivity variability impacts the solution at the level of the neural mass. Statistical properties of the connectivities have been studied in details for instance in (Braitenberg and Schüz, 1998).



We consider a network of  $N$  neurons,  $N_\alpha$ ,  $\alpha = 1, 2, 3$  belonging to population  $\alpha$ . We index by 1 (respectively 2, and 3) the pyramidal (respectively excitatory feedback, inhibitory interneuron) populations. We choose in each population a particular neuron indexed by  $i_\alpha$ ,  $\alpha = 1, 2, 3$ . The evolution equations of the network can be written for instance in terms of the potentials  $V_{i1}$ ,  $V_{i2}$  and  $V_{i3}$  labeled in **Figure 2** and these equations read:

$$\begin{cases} V_{i1} = g_1 * \left( \sum_{j=1}^{N_2} J_{i1j} S(V_j) + \sum_{j=1}^{N_3} J_{i1j} S(V_j) + I_1 + n_1 \right) \\ V_{i2} = g_2 * \left( \sum_{j=1}^{N_1} J_{i2j} S(V_j) + I_2 + n_2 \right) \\ V_{i3} = g_3 * \left( \sum_{j=1}^{N_1} J_{i3j} S(V_j) + I_3 + n_3 \right) \end{cases}$$

In the mean-field limit, denoting by  $\mathcal{V}_\alpha$ ,  $\alpha = 1, 2, 3$  the average membrane potential of each class, we obtain the following equations:

$$\begin{cases} \mathcal{V}_1 = g_1 * (U_{12}^\nu + U_{13}^\nu + I_1 + n_1) \\ \mathcal{V}_2 = g_2 * (U_{21}^\nu + I_2 + n_2) \\ \mathcal{V}_3 = g_3 * (U_{31}^\nu + I_3 + n_3) \end{cases} \quad (35)$$

where  $\mathbf{U}^\nu = (U_{\alpha\beta}^\nu)_{\alpha,\beta=1,2,3}$  is the effective interaction process associated with this problem, i.e., a Gaussian process of mean:

$$\begin{cases} \mathbb{E}[U_{12}^\nu] = \bar{J}_{12} \mathbb{E}[S(\mathcal{V}_2)] \\ \mathbb{E}[U_{13}^\nu] = \bar{J}_{13} \mathbb{E}[S(\mathcal{V}_3)] \\ \mathbb{E}[U_{21}^\nu] = \bar{J}_{21} \mathbb{E}[S(\mathcal{V}_1)] \\ \mathbb{E}[U_{31}^\nu] = \bar{J}_{31} \mathbb{E}[S(\mathcal{V}_1)], \end{cases}$$

All other mean values correspond to the non-interacting populations and are equal to 0. The covariance matrix can be deduced from Eq. 25:

of neurons by transforming the membrane potential of a population into an output firing rate. The coefficients  $J_{\alpha\beta}$  are the random synaptic efficiency of population  $\beta$  on population  $\alpha$  (1 represents the pyramidal population, 2 the excitatory feedback, and 3 the inhibitory interneurons).

$$\text{Cov}(U_{\alpha\beta}^\nu(t), U_{\gamma\delta}^\nu(s)) = \begin{cases} \sigma_{\alpha\beta}^2 \Delta_\beta^\nu(t, s) & \text{if } \alpha = \beta \text{ and } \gamma = \delta \\ 0 & \text{otherwise} \end{cases}$$

where

$$\Delta_\beta^\nu(t, s) = \mathbb{E}[S(\mathcal{V}_\alpha(t))S(\mathcal{V}_\beta(s))]$$

This model is a voltage-based model in the sense of Ermentrout (1998). Let us now instantiate the synaptic dynamics and compare the mean-field equations with Jansen's population equations<sup>2</sup> (sometimes improperly called also mean-field equations).

The simplest model of synaptic integration is a first-order integration, which yields exponentially decaying postsynaptic potentials:

$$g(t) = \begin{cases} Ke^{-\frac{t}{\tau}} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Note that this is exactly Eq. 10. The corresponding  $g$ -shape satisfies the following first-order differential equation

$$\dot{g}(t) = -\frac{1}{\tau}g(t) + K\delta(t),$$

In this equation  $\tau$  is the time constant of the synaptic integration and  $K$  the synaptic efficiency. The coefficients  $K$  and  $\tau$  are the same for the pyramidal and the excitatory feedback population (characteristic of the pyramidal neurons and defining the  $g$ -shape  $g_1$ ), and different for the inhibitory population (defining the  $g$ -shape  $g_3$ ). In the pyramidal or excitatory (respectively the inhibitory) case we have  $K = K_1$ ,  $\tau = \tau_1$  (respectively  $K = K_3$ ,  $\tau = \tau_3$ ). Finally, the sigmoid functions  $S$  is given by

$$S(v) = \frac{v_{\max}}{1 + e^{r(v_0 - v)}},$$

where  $v_{\max}$  is the maximum firing rate, and  $v_0$  is a voltage reference.

<sup>2</sup>We have modified the original model which is not voltage-based.

With this synaptic dynamics we obtain the first-order Jansen and Rit's equation:

$$\begin{cases} \frac{d\mathcal{V}_1}{dt} = -\frac{1}{\tau_1}\mathcal{V}_1 + K_1(U_{12}^\mathcal{V} + U_{13}^\mathcal{V} + I_1 + n_1) \\ \frac{d\mathcal{V}_2}{dt} = -\frac{1}{\tau_1}\mathcal{V}_2 + K_1(U_{21}^\mathcal{V} + I_2 + n_2) \\ \frac{d\mathcal{V}_3}{dt} = -\frac{1}{\tau_3}\mathcal{V}_3 + K_3(U_{31}^\mathcal{V} + I_3 + n_3) \end{cases} \quad (36)$$

The "original" Jansen and Rit's equation (Grimbert and Faugeras, 2006; Jansen and Rit, 1995) amount considering only the mean of the process  $\mathcal{V}$  and assuming that  $\mathbb{E}[S_i(\mathcal{V}_j)] = S_i(\mathbb{E}[\mathcal{V}_j])$  for  $i, j \in \{1, 2, 3\}$ , i.e., that the expectation commutes with the sigmoidal function  $S$ . This is a very strong assumption, and that the fluctuations of the solutions of the mean-field equation around the mean imply that the sigmoid cannot be considered as linear in the general case.

A higher order model was introduced by van Rotterdam et al. (1982) to better account for the synaptic integration and to better reproduce the characteristics of real postsynaptic potentials. In this model the g-shapes satisfy a second-order differential equation:

$$g(t) = \begin{cases} Kte^{-\frac{t}{\tau}} & t \geq 0, \\ 0 & t < 0 \end{cases}$$

We recognize the g-shape defined by Eq. 11 solution of the second-order differential equation  $\ddot{y}(t) + \frac{2}{\tau}\dot{y}(t) + \frac{1}{\tau^2}y(t) = K\delta(t)$ . With this type of synaptic integration, we obtain the following mean-field equations:

$$\begin{cases} \frac{d^2\mathcal{V}_1}{dt^2} = -\frac{2}{\tau_1}\frac{d\mathcal{V}_1}{dt} - \frac{1}{\tau_1^2}\mathcal{V}_1 + K_1(U_{12}^\mathcal{V} + U_{13}^\mathcal{V} + I_1 + n_1) \\ \frac{d^2\mathcal{V}_2}{dt^2} = -\frac{2}{\tau_1}\frac{d\mathcal{V}_2}{dt} - \frac{1}{\tau_1^2}\mathcal{V}_2 + K_1(U_{21}^\mathcal{V} + I_2 + n_2) \\ \frac{d^2\mathcal{V}_3}{dt^2} = -\frac{2}{\tau_3}\frac{d\mathcal{V}_3}{dt} - \frac{1}{\tau_3^2}\mathcal{V}_3 + K_3(U_{31}^\mathcal{V} + I_3 + n_3) \end{cases} \quad (37)$$

Here again, going from the mean-field Eq. 37 to the original Jansen and Rit's neural mass model consists in studying the equation of the mean of the process given by Eq. 37 and commuting the sigmoidal function with the expectation.

Note that the introduction of higher order synaptic integrations results in richer behaviors. For instance, Grimbert and Faugeras (2006) showed that some bifurcations can appear in the second-order JR model giving rise to epileptic like oscillations and alpha activity, that do not appear in the first-order model.

### EXISTENCE AND UNIQUENESS OF SOLUTIONS IN FINITE TIME

The mean-field equation (Eq. 31) is an implicit equation of the stochastic process  $(V(t))_{t \geq t_0}$ . We prove in this section that under some mild assumptions this implicit equation has a unique solution. These assumptions are the following.

### Assumption 1.

- (a) The matrix  $L(t)$  is  $C^0$  and satisfies  $\|L(t)\| \leq k_L$  for all  $t$  in  $[t_0, T]$ , for some matrix norm  $\|\cdot\|$  and some strictly positive constant  $k_L$ .
- (b) The matrix  $F(t)$  has all its singular values lowerbounded (respectively upperbounded) by the strictly positive constant<sup>3</sup>  $\lambda_{\min}^\Gamma$  (respectively  $\lambda_{\max}^\Gamma$ ) for all  $t$  in  $[t_0, T]$ .
- (c) The deterministic external input vector  $I(t)$  is bounded and we have  $\|I(t)\|_\infty \leq I_{\max}$  for all  $t$  in  $[t_0, T]$  and some strictly positive constant  $I_{\max}$ .

This solution is the fixed point in the set  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$  of  $kP$ -dimensional processes of an equation that we will define from the mean-field equations. We will construct a sequence of Gaussian processes and prove that it converges in distribution toward this fixed point.

We first recall some results on the convergence of random variables and stochastic processes.

### CONVERGENCE OF GAUSSIAN PROCESSES

We recall the following result from Bogachev (1998) which formalizes the intuition that a sequence of Gaussian processes converges toward a Gaussian process if and only if the means and covariance functions converge. In fact in order for this to be true, it is only necessary to add one more condition, namely that the corresponding sequence of measures (elements of  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$ ) do not have "any mass at infinity". This property is called uniform tightness (Billingsley, 1999). More precisely we have

**Definition 3.** (Uniform tightness). Let  $\{X_n\}_{n=1}^\infty$  be a sequence of  $kP$ -dimensional processes defined on  $[t_0, T]$  and  $P_n$  be the associated elements of  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$ . The sequence  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$  is called uniformly tight if and only if for all  $\varepsilon > 0$  there exists a compact set  $K$  of  $C([t_0, T], \mathbb{R}^{kP})$  such that  $P_n(K) > 1 - \varepsilon, n \geq 1$ .

**Theorem 1.** Let  $\{X_n\}_{n=1}^\infty$  be a sequence of  $kP$ -dimensional Gaussian processes defined on  $[t_0, T]$  or on an unbounded interval<sup>4</sup> of  $\mathbb{R}$ . The sequence converges to a Gaussian process  $X$  if and only if the following three conditions are satisfied:

- The sequence  $\{X_n\}_{n=1}^\infty$  is uniformly tight.
- The sequence  $\mu^n(t)$  of the mean functions converges for the uniform norm.
- The sequence  $C^n$  of the covariance operators converges for the uniform norm.

We now, as advertised, define such a sequence of Gaussian processes.

Let us fix  $Z_0$ , a  $kP$ -dimensional Gaussian random variable, independent of the Brownian and of the process  $((X)_t)_{t \in [t_0, T]}$ .

**Definition 4.** Let  $X$  be an element of  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$  and  $\mathcal{F}_k$  be the function  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP})) \rightarrow \mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$  such that

<sup>3</sup>We note  $\Gamma(t)$  the matrix  $F(t)F(t)^T$ .

<sup>4</sup>In Bogachev (1998; Chapter 3.8), the property is stated whenever the mean and covariance are defined on a separable Hilbert space.



$$\mathcal{F}_k(X)_t = \Phi_L(t, t_0) \cdot Z_0 + \int_{t_0}^t \Phi_L(t, s) \cdot (\tilde{U}_s^X + \tilde{I}(s)) ds + \int_{t_0}^t \Phi_L(t, s) \cdot F(s) dW_s$$

where  $\tilde{U}_s^X$  and  $\tilde{I}(s)$  are defined<sup>5</sup> in Section “Mean-Field Equations for Multi-Populations Neural Network Models”.

Note that, by Definition 2 the random process  $(\mathcal{F}_k(X))_{t \in [t_0, T]}$ ,  $k \geq 1$  is the sum of a deterministic function (defined by the external current) and three independent random processes defined by  $Z_0$ , the interaction between neurons, and the external noise. These three processes being Gaussian processes, so is  $(\mathcal{F}_k(X))_{t \in [t_0, T]}$ . Also note that  $(\mathcal{F}_k(X))_{t_0} = Z_0$ . It should be clear that a solution  $\mathcal{V}$  of the mean-field equation (Eq. 31) satisfies  $\mathcal{V}(t_0) = Z_0$  and is a fixed point of  $\mathcal{F}_k$ , i.e.,  $\mathcal{F}_k(\mathcal{V})t = \mathcal{V}(t)$ .

Let  $X$  be a given stochastic process of  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$  such that  $X_{t_0} = Z_0$  (hence  $X_{t_0}$  is independent of the Brownian). We define the sequence of Gaussian processes  $\{X_n\}_{n=0}^\infty \in \mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$  by:

$$\begin{cases} X_0 = X \\ X_{n+1} = \mathcal{F}_k(X_n) = \mathcal{F}_k^{(n)}(X_0). \quad n \geq 0, \quad \mathcal{F}_k^{(0)} = \text{Id} \end{cases} \quad (38)$$

In the remaining of this section we show that the sequence of processes  $\{\mathcal{F}_k^{(n)}(X)\}_{n=0}^\infty$  converges in distribution toward the unique fixed-point  $Y$  of  $\mathcal{F}_k$  which is also the unique solution of the mean-field equation (Eq. 31).

### EXISTENCE AND UNIQUENESS OF A SOLUTION FOR THE MEAN-FIELD EQUATIONS

The following upper and lower bounds are used in the sequel.

**Lemma 1.** Consider the Gaussian process  $((U_t^X \cdot \mathbf{1})_{t \in [t_0, T]})$ .  $U^X$  is defined in Sections “The Mean-Field Equations” and “Introduction” is the  $P$ -dimensional vector with all coordinates equal to 1. We have

$$\|E[U_t^X \cdot \mathbf{1}]\|_\infty \leq \mu \stackrel{\text{def}}{=} \max_\alpha \sum_\beta |\bar{J}_{\alpha\beta}| \|S_\beta\|_\infty \quad (39)$$

for all  $t_0 \leq t \leq T$ . The maximum eigenvalue of its covariance matrix is upperbounded by  $\sigma_{\max}^2 \stackrel{\text{def}}{=} \max_\alpha \sum_\beta \sigma_{\alpha\beta}^2 \|S_\beta\|_\infty^2$  where  $\|S_\beta\|_\infty$  is the supremum of the absolute value of  $S_\beta$ . We also note  $\sigma_{\min}^2 \stackrel{\text{def}}{=} \min_{\alpha\beta} \sigma_{\alpha\beta}^2$ .

*Proof.* The proof is straightforward from Definition 4.  $\square$

The proof of existence and uniqueness of solution, and of the convergence of the sequence (Eq. 38) is in two main steps. We first prove that the sequence of Gaussian processes  $\{\mathcal{F}_k^{(n)}(X)\}_{n=0}^\infty$ ,  $k \geq 1$  is uniformly tight by proving that it satisfies Kolmogorov’s criterion for tightness. This takes care of condition 1 in Theorem 1. We then prove that the sequences of the mean functions and covariance operators are Cauchy sequences for the uniform norms, taking care of conditions 2 and 3.

#### Uniform tightness

We first recall the following theorem due to Kolmogorov (Kushner, 1984, Chapter 4.1).

<sup>5</sup>For simplicity we abuse notations and identify  $\tilde{X}$  and  $X$ .

**Theorem 2.** (Kolmogorov’s criterion for tightness). Let  $\{X_n\}_{n=1}^\infty$  be a sequence of  $kP$ -dimensional processes defined on  $[t_0, T]$ . If there exist  $\alpha, \beta, C > 0$  such that

$$E[\|X_n(t) - X_n(s)\|^\beta] \leq C |t - s|^{\alpha + \beta} \quad \forall s, t \in [t_0, T] \quad n \geq 1,$$

then the sequence is uniformly tight.

Using this theorem we prove that the sequence  $\{\mathcal{F}_k^{(n)}(X)\}_{n=0}^\infty$ ,  $k \geq 1$  satisfies Kolmogorov’s criterion for  $\beta = 4$  and  $\alpha \geq 1$ . The reason for choosing  $\beta = 4$  is that, heuristically,  $dW \approx (dt)^{1/2}$ . Therefore in order to upperbound  $E[\|X_n(t) - X_n(s)\|^\beta]$  by a power of  $|t - s| \geq 2$  (hence strictly larger than 1) we need to raise  $\|X_n(t) - X_n(s)\|$  to a power at least equal to 4. The proof itself is technical and uses standard inequalities (Cauchy–Schwarz’s and Jensen’s), properties of Gaussian integrals, elementary properties of the stochastic integral, and Lemma 1. It also uses the fact that the input current is bounded, i.e., that  $\max_{\alpha=1, \dots, P} \sup_{t \in [t_0, T]} |I_\alpha(t)| \leq I_{\max} < \infty$  this is Assumption (c) in 1.

**Theorem 3.** The sequence of processes  $\{\mathcal{F}_k^{(n)}(X)\}_{n=0}^\infty$ ,  $k \geq 1$  is uniformly tight.

*Proof.* We do the proof for  $k = 1$ , the case  $k > 1$  is similar. If we assume that  $n \geq 1$  and  $s < t$  we can rewrite the difference  $\mathcal{F}_1^{(n)}(X)_t - \mathcal{F}_1^{(n)}(X)_s$  as follows, using property (i) in Proposition B.1 in Appendix B.

$$\begin{aligned} \mathcal{F}_1^{(n)}(X)_t - \mathcal{F}_1^{(n)}(X)_s &= (\Phi_L(t, t_0) - \Phi_L(s, t_0)) X_{t_0} \\ &+ (\Phi_L(t, s) - \text{Id}) \int_{t_0}^s \Phi_L(s, u) U_u^{\mathcal{F}_1^{(n-1)}(X)} \cdot \mathbf{1} du \\ &+ \int_s^t \Phi_L(t, u) U_u^{\mathcal{F}_1^{(n-1)}(X)} \cdot \mathbf{1} du + (\Phi_L(t, s) - \text{Id}) \int_{t_0}^s \Phi_L(s, u) F(u) dW_u \\ &+ \int_s^t \Phi_L(t, u) F(u) dW_u + (\Phi_L(t, s) - \text{Id}) \int_{t_0}^s \Phi_L(s, u) \mathbf{I}(u) du \\ &+ \int_s^t \Phi_L(t, u) \mathbf{I}(u) du \end{aligned}$$

The righthand side is the sum of seven terms and therefore (Cauchy–Schwarz inequality):

$$\begin{aligned} \frac{1}{7} \|\mathcal{F}_1^{(n)}(X)_t - \mathcal{F}_1^{(n)}(X)_s\|^2 &\leq \|\Phi_L(t, t_0) - \Phi_L(s, t_0)\|^2 \|X_{t_0}\|^2 \\ &+ (s - t_0) \|\Phi_L(t, s) - \text{Id}\|^2 \int_{t_0}^s \|\Phi_L(s, u)\|^2 \|U_u^{\mathcal{F}_1^{(n-1)}(X)} \cdot \mathbf{1}\|^2 du \\ &+ (t - s) \int_s^t \|\Phi_L(t, u)\|^2 \|U_u^{\mathcal{F}_1^{(n-1)}(X)} \cdot \mathbf{1}\|^2 du \\ &+ \|\Phi_L(t, s) - \text{Id}\|^2 \left\| \int_{t_0}^s \Phi_L(s, u) F(u) dW_u \right\|^2 + \left\| \int_s^t \Phi_L(t, u) F(u) dW_u \right\|^2 \\ &+ (s - t_0)^2 \|\Phi_L(t, s) - \text{Id}\|^2 I_{\max}^2 \sup_{u \in [t_0, s]} \|\Phi_L(s, u)\|^2 \\ &+ (t - s) I_{\max}^2 \sup_{u \in [s, t]} \|\Phi_L(t, u)\|^2. \end{aligned}$$

Because  $\|\Phi_L(t, t_0) - \Phi_L(s, t_0)\| \leq |t - s| \|\mathbf{L}\|$  we see that all terms in the righthand side of the inequality but the second one involving the Brownian motion are of the order of  $(t - s)^2$ . We raise again both sides to the second power, use the Cauchy–Schwarz inequality, and take the expected value:

$$\begin{aligned} \frac{1}{7^3} \mathbb{E} \left[ \left\| \mathcal{F}_1^{(n)}(X)_t - \mathcal{F}_1^{(n)}(X)_s \right\|^4 \right] &\leq \|\Phi_L(t, t_0) - \Phi_L(s, t_0)\|^4 \mathbb{E} \left[ \|X_{t_0}\|^4 \right] \\ &+ (s - t_0)^3 \|\Phi_L(t, s) - Id\|^4 \int_{t_0}^s \|\Phi_L(s, u)\|^4 \mathbb{E} \left[ \|\mathbf{U}_u^{\mathcal{F}_1^{(n-1)}(X)} \cdot \mathbf{1}\|^4 \right] du \\ &+ (t - s)^3 \int_s^t \|\Phi_L(t, u)\|^4 \mathbb{E} \left[ \|\mathbf{U}_u^{\mathcal{F}_1^{(n-1)}(X)} \cdot \mathbf{1}\|^4 \right] du \\ &+ \|\Phi_L(t, s) - Id\|^4 \mathbb{E} \left[ \left\| \int_{t_0}^s \Phi_L(s, u) F(u) d\mathbf{W}_u \right\|^4 \right] \\ &+ \mathbb{E} \left[ \left\| \int_s^t \Phi_L(t, u) F(u) d\mathbf{W}_u \right\|^4 \right] \\ &+ (s - t_0)^4 \|\Phi_L(t, s) - Id\|^4 I_{\max}^4 \sup_{u \in [t_0, s]} \|\Phi_L(s, u)\|^4 \\ &+ (t - s)^4 I_{\max}^4 \sup_{u \in [s, t]} \|\Phi_L(t, u)\|^4. \end{aligned} \tag{40}$$

Remember that  $\mathbf{U}_u^{\mathcal{F}_1^{(n-1)}(X)} \cdot \mathbf{1}$  is a  $P$ -dimensional diagonal Gaussian process, noted  $\mathbf{Y}_u$  in the sequel, therefore:

$$\mathbb{E} \|\mathbf{Y}_u\|^4 = \sum_{\alpha} \mathbb{E} [Y_{\alpha}(u)^4] + \sum_{\alpha_1 \neq \alpha_2} \mathbb{E} [Y_{\alpha_1}^2(u)] \mathbb{E} [Y_{\alpha_2}^2(u)].$$

The second-order moments are upperbounded by some regular function of  $\mu$  and  $\sigma_{\max}$  (defined in Lemma 1) and, because of the properties of Gaussian integrals, so are the fourth-order moments.

We now define  $\mathbf{B}(u) = \Phi_L(s, u)F(u)$  and evaluate  $\mathbb{E} \left[ \left\| \int_{t_0}^s \mathbf{B}(u) d\mathbf{W}_u \right\|^4 \right]$ . We have

$$\begin{aligned} \mathbb{E} \left[ \left\| \int_{t_0}^s \mathbf{B}(u) d\mathbf{W}_u \right\|^4 \right] &= \mathbb{E} \left[ \left( \left\| \int_{t_0}^s \mathbf{B}(u) d\mathbf{W}_u \right\|^2 \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \sum_{i=1}^P \left( \sum_{j=1}^P \int_{t_0}^s B_{ij}(u) dW_u^j \right) \left( \sum_{k=1}^P \int_{t_0}^s B_{ik}(u) dW_u^k \right) \right)^2 \right] \\ &= \sum_{i_1, j_1, j_2, k_1, k_2} \mathbb{E} \left[ \int_{t_0}^s B_{i_1 j_1}(u) dW_u^{j_1} \int_{t_0}^s B_{i_1 k_1}(u) dW_u^{k_1} \right. \\ &\quad \left. \times \int_{t_0}^s B_{i_2 j_2}(u) dW_u^{j_2} \int_{t_0}^s B_{i_2 k_2}(u) dW_u^{k_2} \right]. \end{aligned}$$

Because  $W_u^i$  is by construction independent of  $W_u^j$  if  $i \neq j$  and  $\mathbb{E} \left[ \int_{t_0}^s B_{ij}(u) dW_u^j \right] = 0$  for all  $i, j$  (property of the It integral), the last term is the sum of only three types of terms:

1. If  $j_1 = k_1 = j_2 = k_2$  we define

$$T_1 = \sum_{i_1, i_2} \mathbb{E} \left[ \left( \int_{t_0}^s B_{i_1 j}(u) dW_u^j \right)^2 \left( \int_{t_0}^s B_{i_2 j}(u) dW_u^j \right)^2 \right],$$

and, using Cauchy–Schwarz:

$$T_1 \leq \sum_{i_1, i_2} \mathbb{E} \left[ \left( \int_{t_0}^s B_{i_1 j}(u) dW_u^j \right)^4 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left( \int_{t_0}^s B_{i_2 j}(u) dW_u^j \right)^4 \right]^{\frac{1}{2}}$$

2. If  $j_1 = k_1$  and  $j_2 = k_2$  but  $1 \neq j_2$  we define

$$T_2 = \sum_{i_1, i_2, j_1 \neq j_2} \mathbb{E} \left[ \left( \int_{t_0}^s B_{i_1 j_1}(u) dW_u^{j_1} \right)^2 \left( \int_{t_0}^s B_{i_2 j_2}(u) dW_u^{j_2} \right)^2 \right],$$

which is equal, because of the independence of  $W_u^{j_1}$  and  $W_u^{j_2}$  to

$$\sum_{i_1, i_2, j_1 \neq j_2} \mathbb{E} \left[ \left( \int_{t_0}^s B_{i_1 j_1}(u) dW_u^{j_1} \right)^2 \right] \mathbb{E} \left[ \left( \int_{t_0}^s B_{i_2 j_2}(u) dW_u^{j_2} \right)^2 \right].$$

3. Finally, if  $j_1 = j_2$  and  $k_1 = k_2$  but  $j_1 \neq k_1$  we define

$$\begin{aligned} T_3 = \sum_{i_1, i_2, j_1 \neq k_1} \mathbb{E} \left[ \int_{t_0}^s B_{i_1 j_1}(u) dW_u^{j_1} \int_{t_0}^s B_{i_2 j_1}(u) dW_u^{j_1} \right. \\ \left. \times \int_{t_0}^s B_{i_1 k_1}(u) dW_u^{k_1} \int_{t_0}^s B_{i_2 k_1}(u) dW_u^{k_1} \right], \end{aligned}$$

which is equal, because of the independence of  $W_u^{j_1}$  and  $W_u^{k_1}$  to

$$\begin{aligned} \mathbb{E} \left[ \int_{t_0}^s B_{i_1 j_1}(u) dW_u^{j_1} \int_{t_0}^s B_{i_2 j_1}(u) dW_u^{j_1} \right] \\ \mathbb{E} \left[ \int_{t_0}^s B_{i_1 k_1}(u) dW_u^{k_1} \int_{t_0}^s B_{i_2 k_1}(u) dW_u^{k_1} \right], \end{aligned}$$

because of the properties of the stochastic integral,  $\int_{t_0}^s B_{i_1 j}(u) dW_u^j = \mathcal{N}(0, (\int_{t_0}^s B_{ij}^2(u) du)^{1/2})$  hence, because of the properties of the Gaussian integrals

$$\mathbb{E} \left[ \left( \int_{t_0}^s B_{ij}(u) dW_u^j \right)^4 \right] = k \left( \int_{t_0}^s B_{ij}^2(u) du \right)^2,$$

for some positive constant  $k$ . This takes care of the terms of the form  $T_1$ . Next we have

$$\mathbb{E} \left[ \left( \int_{t_0}^s B_{i_1 j_1}(u) dW_u^{j_1} \right)^2 \right] = \int_{t_0}^s B_{i_1 j_1}^2(u) du,$$

which takes care of the terms of the form  $T_2$ . Finally we have, because of the properties of the It integral

$$\mathbb{E} \left[ \int_{t_0}^s B_{i_1 j_1}(u) dW_u^{j_1} \int_{t_0}^s B_{i_2 j_1}(u) dW_u^{j_1} \right] = \int_{t_0}^s B_{i_1 j_1}(u) B_{i_2 j_1}(u) du,$$

which takes care of the terms of the form  $T_3$ .

This shows that the term  $\mathbb{E} \left[ \left\| \int_{t_0}^t \Phi_L(t, u) F(u) d\mathbf{W}_u \right\|^4 \right]$  in Eq. 40 is of the order of  $(t - s)^{1+a}$  where  $a \geq 1$ . Therefore we have

$$\mathbb{E} \left[ \left\| \mathcal{F}_1^{(n)}(X)_t - \mathcal{F}_1^{(n)}(X)_s \right\|^4 \right] \leq C |t - s|^{1+a}, \quad a \geq 1$$

for all  $s, t$  in  $[t_0, T]$ , where  $C$  is a constant independent of  $t, s$ . According to Kolmogorov criterion for tightness, the sequence of processes  $\{\mathcal{F}_1^{(n)}(X)\}_{n=0}^\infty$  is uniformly tight.

The proof for  $\mathcal{F}_k, k > 1$  is similar. □

**The mean and covariance sequences are Cauchy sequences**

Let us note  $\mu^n(t)$  [respectively  $C^n(t, s)$ ] the mean (respectively the covariance matrix) function of  $X_n = \mathcal{F}_k(X_{n-1}), n \geq 1$ . We have:

$$\begin{aligned} \mu^n(t) &= \Phi_L(t, t_0) \mu^{Z_0} + \int_{t_0}^t \Phi_L(t, u) \left( \mathbb{E}[\tilde{\mathbf{U}}_u^{X_n}] + \tilde{\mathbf{I}}(u) \right) du = \Phi_L(t, t_0) \mu^{Z_0} \\ &+ \int_{t_0}^t \Phi_L(t, u) \left( \left[ \begin{array}{c} 0_P^T, \dots, 0_P^T, \left[ \sum_{\beta} \tilde{J}_{\alpha\beta} m_{\beta}^{X_n}(u) \right]_{\alpha=1, \dots, P} \end{array} \right]^T + \tilde{\mathbf{I}}(u) \right) du, \end{aligned} \quad (41)$$

where  $m_{\beta}^{X_n}(u)$  is given by Eq. 26. Similarly we have

$$\begin{aligned} C^{n+1}(t, s) &= \Phi_L(t, t_0) \Sigma^{Z_0} \Phi_L(s, t_0)^T \\ &+ \int_{t_0}^{t \wedge s} \Phi_L(t, u) \mathbf{F}(u) \mathbf{F}(u)^T \Phi_L(s, u)^T du \\ &+ \int_{t_0}^t \int_{t_0}^s \Phi_L(t, u) \text{Cov}(\tilde{\mathbf{U}}_u^{X_n}, \tilde{\mathbf{U}}_v^{X_n}) \Phi_L(s, v)^T du dv \end{aligned} \quad (42)$$

Note that the  $kP \times kP$  covariance matrix  $\text{Cov}(\tilde{\mathbf{U}}_u^{X_n}, \tilde{\mathbf{U}}_v^{X_n})$  has only one nonzero  $P \times P$  block:

$$\text{Cov}(\tilde{\mathbf{U}}_u^{X_n}, \tilde{\mathbf{U}}_v^{X_n})_{kk} = \text{Cov}(\mathbf{U}_u^{X_n} \cdot \mathbf{1}, \mathbf{U}_v^{X_n} \cdot \mathbf{1}), \quad (43)$$

According to Definition 2 we have

$$\text{Cov}(\mathbf{U}_u^{X_n} \cdot \mathbf{1}, \mathbf{U}_v^{X_n} \cdot \mathbf{1}) = \text{diag} \left( \sum_{\beta} \sigma_{\alpha\beta}^2 \Delta_{\beta}^{X_n}(u, v) \right),$$

where  $\Delta_{\beta}^{X_n}(u, v)$  is given by Eq. 27 and  $Dx$  is defined in Proposition 1.

In order to prove our main result, that the two sequences of functions  $(\mu^n)$  and  $(C^n)$  are uniformly convergent, we require the following four lemmas that we state without proofs, the proofs being found in Appendixes E–H. The first lemma gives a uniform (i.e., independent of  $n \geq 2$  and  $\alpha = 1, \dots, kP$ ) strictly positive lowerbound for  $C_{\alpha\alpha}^n(t, t)$ . In what follows we use the following notation: Let  $C$  be a symmetric positive definite matrix, we note  $\lambda_{\min}^C$  its smallest eigenvalue.

**Lemma 2.** The following upperbounds are valid for all  $n \geq 1$  and all  $s, t \in [t_0, T]$ .

$$\|\mu^n(t)\|_{\infty} \leq e^{k_L(T-t_0)} \left[ \|\mathbb{E}[Z_0]\|_{\infty} + (\mu + I_{\max})(T - t_0) \right] \stackrel{\text{def}}{=} \mu_{\max},$$

$$\|C^n(t, s)\|_{\infty} \leq e^{(k_L - k_{Lr})(T-t_0)}$$

$$\left[ \rho(\Sigma^{Z_0}) + \lambda_{\max}^{\Gamma}(T - t_0) + \sigma_{\max}^2(T - t_0)^2 \right] \stackrel{\text{def}}{=} \Sigma_{\max},$$

where  $\mu$  and  $\sigma_{\max}$  are defined in Lemma 1,  $\lambda_{\max}^{\Gamma}$  is defined in Assumption 1.

**Lemma 3.** For all  $t \in [t_0, T]$  all  $\alpha = 1, \dots, kP$ , and  $n \geq 1$ , we have

$$C_{\alpha\alpha}^n(t, t) \geq \lambda_{\min} \lambda_{\min}^{\Sigma^{Z_0}} \stackrel{\text{def}}{=} k_0 > 0,$$

where  $\lambda_{\min}$  is the smallest singular value of the positive symmetric definite matrix  $\Phi_L(t, t_0) \Phi_L(t, t_0)^T$  for  $t \in [t_0, T]$  and  $\lambda_{\min}^{\Sigma^{Z_0}}$  is the smallest eigenvalue of the positive symmetric definite covariance matrix  $\Sigma^{Z_0}$ .

The second lemma also gives a uniform lowerbound for the expression  $C_{\alpha\alpha}^n(s, s)C_{\alpha\alpha}^n(t, t) - C_{\alpha\alpha}^n(t, s)^2$  which appears in the definition of  $C^{n+1}$  through Eqs. 43 and 27. The crucial point is that this function is  $O(|t - s|)$  which is central in the proof of Lemma 5.

**Lemma 4.** For all  $\alpha = 1, \dots, kP$  and  $n \geq 1$  the quantity  $C_{\alpha\alpha}^n(s, s)C_{\alpha\alpha}^n(t, t) - C_{\alpha\alpha}^n(t, s)^2$  is lowerbounded by the positive symmetric function:

$$\theta(s, t) \stackrel{\text{def}}{=} |t - s| \lambda_{\min}^2 \lambda_{\min}^{\Sigma^{Z_0}} \lambda_{\min}^{\Gamma},$$

where  $\lambda_{\min}^{\Gamma}$  is the strictly positive lower bound, introduced in Assumption 1, on the singular values of the matrix  $\mathbf{F}(u)$  for  $u \in [t_0, T]$ .

The third lemma shows that an integral that appears in the proof of the uniform convergence of the sequences of functions  $(\mu^n)$  and  $(C^n)$  is upperbounded by the  $n$ th term of a convergent series.

**Lemma 5.** The  $2n$ -dimensional integral

$$\begin{aligned} I_n &= \int_{[t_0, T]^{2n}} \rho_1(u_1, v_1) \left( \int_{[t_0, u_1, v_1]^2} \dots \left( \int_{[t_0, u_{n-2}, v_{n-2}]^2} \rho_{n-1}(u_{n-1}, v_{n-1}) \right. \right. \\ &\quad \left. \left. \times \left( \int_{[t_0, u_{n-1}, v_{n-1}]^2} \rho_n(u_n, v_n) du_n dv_n \right) du_{n-1} dv_{n-1} \right) \dots \right) du_1 dv_1, \end{aligned}$$

where the functions  $\rho_i(u_i, v_i), i = 1, \dots, n$  are either equal to 1 or to  $1/\sqrt{\theta(u_i, v_i)}$  (the function  $\theta$  is defined in Lemma 4), is upperbounded by  $k^n/(n - 1)!$  for some positive constant  $k$ .

With these lemmas in hand we prove Proposition 3. The proof is technical but its idea is very simple. We find upperbounds for the matrix infinite norm of  $C^{n+1}(t, s) - C^n(t, s)$  and the infinite norm of  $\mu^{n+1}(t) - \mu^n(t)$  by applying the mean value Theorem and Lemmas 3 and 4 to these norms. These upperbounds involve integrals of the infinite norms of  $C^n(t, s) - C^{n-1}(t, s)$  and  $\mu^n(t) - \mu^{n-1}(t)$  and, through Lemma 4, one over the square root of the function  $\theta$ . Proceeding recursively and using Lemma 5, one easily shows that the infinite norms of  $C^{n+1} - C^n$  and  $\mu^{n+1} - \mu^n$  are upperbounded by the  $n$ th term of a convergent series from which it follows that the two sequences of functions are Cauchy sequences, hence convergent.

**Proposition 3.** The sequences of covariance matrix functions  $C^n(t, s)$  and of mean functions  $\mu^n(t), s, t$  in  $[t_0, T]$  are Cauchy sequences for the uniform norms.

*Proof.* We have

$$\begin{aligned} C^{n+1}(t, s) - C^n(t, s) &= \int_{t_0}^t \int_{t_0}^s \Phi_L(t, u) \left( \text{Cov}(\tilde{\mathbf{U}}_u^{X_n}, \tilde{\mathbf{U}}_v^{X_n}) \right. \\ &\quad \left. - \text{Cov}(\tilde{\mathbf{U}}_u^{X_{n-1}}, \tilde{\mathbf{U}}_v^{X_{n-1}}) \right) \Phi_L(s, v)^T du dv. \end{aligned}$$

We take the infinite matrix norm of both sides of this equality and use the upperbounds  $\|\Phi_L(t, u)\|_{\infty} \leq e^{\|L\|_{\infty}(T-t_0)} = k_L$  and  $\|\Phi_L(t, u)^T\|_{\infty} \leq e^{\|L^T\|_{\infty}(T-t_0)} = k_{L^T}$  (see Appendix B) to obtain<sup>6</sup>

<sup>6</sup>The notation  $\|\cdot\|$  is introduced in Appendix C.

$$\begin{aligned} \|C^{n+1}(t,s) - C^n(t,s)\|_\infty &\leq k_L k_{L'} \int_{t_0}^t \int_{t_0}^s \left\| \text{Cov}(\tilde{U}_u^{X_n}, \tilde{U}_v^{X_n}) \right. \\ &\quad \left. - \text{Cov}(\tilde{U}_u^{X_{n-1}}, \tilde{U}_v^{X_{n-1}}) \right\|_\infty^v du dv \\ &= k_L k_{L'} \int_{t_0}^t \int_{t_0}^s \left\| \text{Cov}(U_u^{X_n} \cdot \mathbf{1}, U_v^{X_n} \cdot \mathbf{1}) \right. \\ &\quad \left. - \text{Cov}(U_u^{X_{n-1}} \cdot \mathbf{1}, U_v^{X_{n-1}} \cdot \mathbf{1}) \right\|_\infty^v du dv. \end{aligned} \tag{44}$$

According to Eq. 27 we are led to consider the difference  $A_n - A_{n-1}$ , where:

$$\begin{aligned} A_n &\stackrel{\text{def}}{=} S_\beta \left( \frac{\sqrt{C_{\beta\beta}^n(u,u)C_{\beta\beta}^n(v,v) - C_{\beta\beta}^n(u,v)^2}}{\sqrt{C_{\beta\beta}^n(u,u)}} x + \frac{C_{\beta\beta}^n(u,v)}{\sqrt{C_{\beta\beta}^n(u,u)}} y + \mu_\beta^n(v) \right) \\ &\quad S_\beta \left( y \sqrt{C_{\beta\beta}^n(u,u)} + \mu_\beta^n(u) \right) \\ &\stackrel{\text{def}}{=} S_\beta \left[ \mathcal{P}_\beta^n(u,v)x + \mathcal{S}_\beta^n(u,v)y + \mu_\beta^n(v) \right] S_\beta \left[ \mathcal{T}_\beta^n(u)y + \mu_\beta^n(u) \right]. \end{aligned}$$

We write next:

$$\begin{aligned} A_n - A_{n-1} &= S_\beta \left[ \mathcal{P}_\beta^n(u,v)x + \mathcal{S}_\beta^n(u,v)y + \mu_\beta^n(v) \right] \left( S_\beta \left[ \mathcal{T}_\beta^n(u)y + \mu_\beta^n(u) \right] \right. \\ &\quad \left. - S_\beta \left[ \mathcal{T}_\beta^{n-1}(u)y + \mu_\beta^{n-1}(u) \right] \right) \\ &\quad + S_\beta \left[ \mathcal{T}_\beta^{n-1}(u)y + \mu_\beta^{n-1}(u) \right] \\ &\quad \times \left( S_\beta \left[ \mathcal{P}_\beta^n(u,v)x + \mathcal{S}_\beta^n(u,v)y + \mu_\beta^n(v) \right] \right. \\ &\quad \left. - S_\beta \left[ \mathcal{P}_\beta^{n-1}(u,v)x + \mathcal{S}_\beta^{n-1}(u,v)y + \mu_\beta^{n-1}(v) \right] \right). \end{aligned}$$

The mean value theorem yields:

$$\begin{aligned} |A_n - A_{n-1}| &\leq \|S_\beta\|_\infty \|S'_\beta\|_\infty \left( |x| |\mathcal{P}_\beta^n(u,v) - \mathcal{P}_\beta^{n-1}(u,v)| \right. \\ &\quad \left. + |y| |\mathcal{S}_\beta^n(u,v) - \mathcal{S}_\beta^{n-1}(u,v)| + |\mu_\beta^n(v) - \mu_\beta^{n-1}(v)| \right) \\ &\quad + |y| |\mathcal{T}_\beta^n(u) - \mathcal{T}_\beta^{n-1}(u)| + |\mu_\beta^n(u) - \mu_\beta^{n-1}(u)|. \end{aligned}$$

Using the fact that  $\int_{-\infty}^\infty |x| D_x = \sqrt{2/\pi}$ , we obtain:

$$\begin{aligned} \|C^{n+1}(t,s) - C^n(t,s)\|_\infty &\leq k_L k_{L'} k_C \left( \sqrt{\frac{2}{\pi}} \int_{t_0}^t \int_{t_0}^s \|\mathcal{P}^n(u,v) - \mathcal{P}^{n-1}(u,v)\|_\infty dudv \right. \\ &\quad \left. + \sqrt{\frac{2}{\pi}} \int_{t_0}^t \int_{t_0}^s \|\mathcal{S}^n(u,v) - \mathcal{S}^{n-1}(u,v)\|_\infty dudv \right. \\ &\quad \left. + (t - t_0) \int_{t_0}^s \|\mu^n(v) - \mu^{n-1}(v)\|_\infty dv \right. \\ &\quad \left. + (s - t_0) \int_{t_0}^t \|\mu^n(u) - \mu^{n-1}(u)\|_\infty du \right. \\ &\quad \left. + \sqrt{\frac{2}{\pi}} (s - t_0) \int_{t_0}^t \|T^n(u) - T^{n-1}(u)\|_\infty du \right), \end{aligned}$$

where the constants  $k_L$  and  $k_{L'}$  are defined in Appendix B and

$$k_C \stackrel{\text{def}}{=} \max_\alpha \sum_\beta \sigma_{\alpha\beta}^2 \|S_\beta\|_\infty \|S'_\beta\|_\infty. \tag{45}$$

A similar process applied to the mean values yields:

$$\begin{aligned} \|\mu^{n+1}(t) - \mu^n(t)\|_\infty &\leq k_L \mu \left( \int_{t_0}^t \|T^n(u) - T^{n-1}(u)\|_\infty du \right. \\ &\quad \left. + \int_{t_0}^t \|\mu^n(u) - \mu^{n-1}(u)\|_\infty du \right), \end{aligned}$$

where  $\mu$  is defined in Lemma 1. We now use the mean value Theorem and Lemmas 3 and 4 to find upperbounds for  $\|\mathcal{P}^n(u,v) - \mathcal{P}^{n-1}(u,v)\|_\infty$ ,  $\|\mathcal{S}^n(u,v) - \mathcal{S}^{n-1}(u,v)\|_\infty$  and  $\|T^n(u) - T^{n-1}(u)\|_\infty$ . We have

$$\begin{aligned} |\mathcal{T}_\beta^n(u) - \mathcal{T}_\beta^{n-1}(u)| &= \left| \sqrt{C_{\beta\beta}^n(u,u)} - \sqrt{C_{\beta\beta}^{n-1}(u,u)} \right| \\ &\leq \frac{1}{2\sqrt{k_0}} |C_{\beta\beta}^n(u,u) - C_{\beta\beta}^{n-1}(u,u)|, \end{aligned}$$

where  $k_0$  is defined in Lemma 3. Hence:

$$\|T^n(u) - T^{n-1}(u)\|_\infty \leq \frac{1}{2\sqrt{k_0}} \|C^n(u,u) - C^{n-1}(u,u)\|_\infty.$$

Along the same lines we can show easily that:

$$\begin{aligned} \|\mathcal{S}^n(u,v) - \mathcal{S}^{n-1}(u,v)\|_\infty &\leq k \left( \|C^n(u,v) - C^{n-1}(u,v)\|_\infty \right. \\ &\quad \left. + \|C^n(u,u) - C^{n-1}(u,u)\|_\infty \right), \end{aligned}$$

and that:

$$\begin{aligned} \|\mathcal{P}^n(u,v) - \mathcal{P}^{n-1}(u,v)\|_\infty &\leq \frac{k}{\sqrt{\theta(u,v)}} \left( \|C^n(u,v) - C^{n-1}(u,v)\|_\infty \right. \\ &\quad \left. + \|C^n(u,u) - C^{n-1}(u,u)\|_\infty \right. \\ &\quad \left. + \|C^n(v,v) - C^{n-1}(v,v)\|_\infty \right), \end{aligned}$$

where  $\theta(u,v)$  is defined in Lemma 4. Grouping terms together and using the fact that all integrated functions are positive, we write:

$$\begin{aligned} \|C^{n+1}(t,s) - C^n(t,s)\|_\infty &\leq k \left( \int_{[t_0,t \vee s]^2} \frac{1}{\sqrt{\theta(u,v)}} \|C^n(u,v) - C^{n-1}(u,v)\|_\infty dudv \right. \\ &\quad \left. + \int_{[t_0,t \vee s]^2} \frac{1}{\sqrt{\theta(u,v)}} \|C^n(u,u) - C^{n-1}(u,u)\|_\infty dudv \right. \\ &\quad \left. + \int_{[t_0,t \vee s]^2} \|C^n(u,v) - C^{n-1}(u,v)\|_\infty dudv \right. \\ &\quad \left. + \int_{[t_0,t \vee s]^2} \|C^n(u,u) - C^{n-1}(u,u)\|_\infty dudv \right. \\ &\quad \left. + \int_{[t_0,t \vee s]^2} \|\mu^n(u) - \mu^{n-1}(u)\|_\infty dudv \right). \end{aligned} \tag{46}$$

Note that, because of Lemma 3, all integrals are well-defined. Regarding the mean functions, we write:

$$\begin{aligned} \|\mu^{n+1}(t) - \mu^n(t)\|_\infty &\leq k \left( \int_{[t_0,t \vee s]^2} \|C^n(u,u) - C^{n-1}(u,u)\|_\infty dudv \right. \\ &\quad \left. + \int_{[t_0,t \vee s]^2} \|\mu^n(u) - \mu^{n-1}(u)\|_\infty dudv \right). \end{aligned} \tag{47}$$



Proceeding recursively until we reach  $C^0$  and  $\mu^0$  we obtain an upperbound for  $\|C^{n+1}(t, s) - C^n(t, s)\|_\infty$  (respectively for  $\|\mu^{n+1}(t) - \mu^n(t)\|_\infty$ ) which is the sum of  $<5^n$  terms each one being the product of  $k$  raised to a power  $\leq n$ , times  $2\mu_{\max}$  or  $2\Sigma_{\max}$  (upperbounds for the norms of the mean vector and the covariance matrix defined in Lemma 2), times a  $2n$ -dimensional integral  $I_n$  given by

$$\int_{[t_0, t \vee s]^2} \rho_1(u_1, v_1) \left( \int_{[t_0, u_1 \vee v_1]} \dots \left( \int_{[t_0, u_{n-2} \vee v_{n-2}]} \rho_{n-1}(u_{n-1}, v_{n-1}) \left( \int_{[t_0, u_{n-1} \vee v_{n-1}]} \rho_n(u_n, v_n) du_n dv_n \right) du_{n-1} dv_{n-1} \right) \dots \right) du_1 dv_1,$$

where the functions  $\rho_i(u_i, v_i), i = 1, \dots, n$  are either equal to 1 or to  $1/\sqrt{\theta(u_i, v_i)}$ . According to Lemma 5, this integral is of the order of some positive constant raised to the power  $n$  divided by  $(n - 1)!$ . Hence the sum is less than some positive constant  $k$  raised to the power  $n$  divided by  $(n - 1)!$ . By taking the supremum with respect to  $t$  and  $s$  in  $[t_0, T]$  we obtain the same result for  $\|C^{n+1} - C^n\|_\infty$  (respectively for  $\|\mu^{n+1} - \mu^n\|_\infty$ ). Since the series  $\sum_{n \geq 1} \frac{k^n}{n!}$  is convergent, this implies that  $\|C^{n+p} - C^n\|_\infty$  (respectively  $\|\mu^{n+p} - \mu^n\|_\infty$ ) can be made arbitrarily small for large  $n$  and  $p$  and the sequence  $C^n$  (respectively  $\mu^n$ ) is a Cauchy sequence.  $\square$

**Existence and uniqueness of a solution of the mean-field equations**

It is now easy to prove our main result, that the mean-field equations (Eq. 31) or equivalently (Eq. 28) are well-posed, i.e., have a unique solution.

**Theorem 4.** For any nondegenerate  $kP$ -dimensional Gaussian random variable  $Z_0$ , independent of the Brownian, and any initial process  $X$  such that  $X(t_0) = Z_0$ , the map  $\mathcal{F}_k$  has a unique fixed point in  $\mathcal{M}_k^+(C([t_0, T], \mathbb{R}^{kP}))$  toward which the sequence  $\{\mathcal{F}_k^{(n)}(X)\}_{n=1}^\infty$  of Gaussian processes converges in law.

*Proof.* Since  $C([t_0, T], \mathbb{R}^{kP})$  (respectively  $C([t_0, T]^2, \mathbb{R}^{kP \times kP})$ ) is a Banach space for the uniform norm, the Cauchy sequence  $\mu^n$  (respectively  $C^n$ ) of Proposition 3 converges to an element  $\mu$  of  $C([t_0, T], \mathbb{R}^{kP})$  (respectively an element  $C$  of  $C([t_0, T]^2, \mathbb{R}^{kP \times kP})$ ). Therefore, according to Theorem 1, the sequence  $\{\mathcal{F}_k^{(n)}(X)\}_{n=0}^\infty$  of Gaussian processes converges in law toward the Gaussian process  $Y$  with mean function  $\mu$  and covariance function  $C$ . This process is clearly a fixed point of  $\mathcal{F}_k$ .

Hence we know that there exists at least one fixed point for the map  $\mathcal{F}_k$ . Assume there exist two distinct fixed points  $Y_1$  and  $Y_2$  of  $\mathcal{F}_k$  with mean functions  $\mu_i$  and covariance functions  $C_i, i = 1, 2$ , with the same initial condition. Since for all  $n \geq 1$  we have  $\mathcal{F}_k^{(n)}(Y_i) = Y_i, i = 1, 2$ , the proof of Proposition 3 shows that  $\|\mu_1^n - \mu_2^n\|_\infty$  (respectively  $\|C_1^n - C_2^n\|_\infty$ ) is upperbounded by the product of a positive number  $a_n$  (respectively  $b_n$ ) with  $\|\mu_1 - \mu_2\|_\infty$  (respectively with  $\|C_1 - C_2\|_\infty$ ). Since  $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = 0$  and  $\mu_i^n = \mu_i, i = 1, 2$  (respectively  $C_i^n = C_i, i = 1, 2$ ), this shows that  $\mu_1 = \mu_2$  and  $C_1 = C_2$ , hence the two Gaussian processes  $Y_1$  and  $Y_2$  are indistinguishable.  $\square$

**CONCLUSION**

We have proved that for any nondegenerate Gaussian initial condition  $Z_0$  there exists a unique solution of the mean-field equations. The proof of Theorem 4 is constructive, and hence provides a way for computing the solution of the mean-field equations by

iterating the map  $\mathcal{F}_k$  defined in 3.2, starting from any initial process  $X$  satisfying  $X(t_0) = Z_0$ , for instance a Gaussian process such as an Ornstein-Uhlenbeck process. We build upon these facts in Section “Numerical Experiments”.

Note that the existence and uniqueness is true whatever the initial time  $t_0$  and the final time  $T$ .

**EXISTENCE AND UNIQUENESS OF STATIONARY SOLUTIONS**

So far, we have investigated the existence and uniqueness of solutions of the mean-field equation for a given initial condition. We are now interested in investigating stationary solutions, which allow for some simplifications of the formalism.

A stationary solution is a solution whose probability distribution does not change under the flow of the equation. These solutions have been already investigated by several authors (see Brunel and Hakim, 1999; Sompolinsky et al., 1998). We propose a new framework to study and simulate these processes. Indeed we show in this section that under a certain contraction condition there exists a unique solution to the stationary problem. As in the previous section our proof is constructive and provides a way to simulate the solutions.

**Remark.** The long-time mean-field description of a network is still a great endeavor in mathematics and statistical physics. In this section we formally take the mean-field equation we obtained and let  $t_0 \rightarrow -\infty$ . This way we obtain an equation which is the limit of the mean-field equation when  $t_0 \rightarrow -\infty$ . It means that we consider first the limit  $N \rightarrow \infty$  and then  $t_0 \rightarrow -\infty$ . These two limits do not necessarily commute and there are known examples, for instance in spin-glasses, where they do not.

It is clear that in order to get stationary solutions, the stochastic system has to be autonomous. More precisely, we modify Assumption 1 as follows

**Assumption 2.**

- (a) The matrixes  $L(t)$  and  $F(t)$ , the input currents  $I(t)$  do not depend upon  $t$ .
- (b) The real parts of the eigenvalues of  $L$  are negative:
 
$$\text{Re}(\lambda) < -\lambda_L \lambda_L > 0 \tag{48}$$
 for all eigenvalues  $\lambda$  of  $L$ .
- (c) The matrix  $F$  has full rank.

Under Assumption (a) of 2, the resolvent  $\Phi_L(t, s)$  is equal to  $e^{L(t-s)}$ . Under Assumption (b) we only consider first-order system since otherwise the matrix  $L$  has eigenvalues equal to 0. We now prove the following proposition.

**Proposition 4.** Under the previous assumptions we have:

1. 
$$\begin{cases} \lim_{t_0 \rightarrow -\infty} e^{L(t-s)} = 0, \\ \int_{-\infty}^t \|e^{L(t-s)}\| ds = \int_0^\infty \|e^{Lu}\| du \stackrel{\text{def}}{=} M_L < \infty, \\ \int_{-\infty}^t \|e^{L^T(t-s)}\| ds = \int_0^\infty \|e^{L^T u}\| du \stackrel{\text{def}}{=} M_{L^T} < \infty, \end{cases}$$
2. the process  $Y_t^{t_0} = \int_{t_0}^t e^{L(t-s)} F \cdot dW_s$  is well-defined, Gaussian and stationary when  $t_0 \rightarrow -\infty$ .

*Proof.* The first point property follows from the fact that  $\text{Re}(\lambda) < -\lambda_L$  for all eigenvalues  $\lambda$  of  $\mathbf{L}$ . This assumption also implies that there exists a norm on  $\mathbb{R}^P$  such that

$$\|e^{L^t}\| \leq e^{-\lambda_L t} \forall t \geq 0,$$

and hence

$$\|e^{L^t}\|_{\infty} \leq ke^{-\lambda_L t} \forall t \geq 0, \tag{49}$$

for some positive constant  $k$ . This implies the remaining two properties.

We now address the second point of the property. The stochastic integral  $Y_t^{t_0} = \int_{t_0}^t e^{L(t-s)} \mathbf{F} \cdot d\mathbf{W}_s$  is well-defined  $\forall t \leq T$  and is Gaussian with 0-mean. Its covariance matrix reads:

$$\sum_{t_0}^{Y_t^{t_0} Y_t^{t_0}} = \int_{t_0}^{t \wedge t'} e^{L(t-s)} \mathbf{F} \mathbf{F}^T e^{L^T} (t' - s) ds.$$

Let us assume for instance that  $t' < t$  and perform the change of variable  $u = t - s$  to obtain

$$\sum_{t_0}^{Y_t^{t_0} Y_t^{t_0}} = \left( \int_{t-t'}^{t-t_0} e^{Lu} \mathbf{F} \mathbf{F}^T e^{L^T u} du \right) e^{L^T} (t' - t).$$

Under the previous assumptions this matrix integral is defined when  $t_0 \rightarrow -\infty$  (dominated convergence theorem) and we have

$$\lim_{t_0 \rightarrow -\infty} \sum_{t_0}^{Y_t^{t_0} Y_t^{t_0}} \stackrel{\text{def}}{=} \sum_{t-t'}^{Y_t^{t_0} Y_t^{t_0}} = \left( \int_{t-t'}^{+\infty} e^{Lu} \mathbf{F} \mathbf{F}^T e^{L^T u} du \right) e^{L^T} (t' - t), \tag{50}$$

which is a well-defined function of  $t' - t$ .  $\square$

The second point of Proposition 4 guarantees the existence of process

$$\mathbf{X}_0(t) = \int_{-\infty}^t e^{L(t-s)} \mathbf{F} \cdot d\mathbf{W}_s.$$

as the limit of the processes  $Y_t^{t_0}$  when  $t_0 \rightarrow -\infty$ . This process is a stationary distribution of the equation:

$$d\mathbf{X}_0(t) = \mathbf{L} \cdot \mathbf{X}_0(t) dt + \mathbf{F} \cdot d\mathbf{W}_t, \tag{51}$$

it is Gaussian, of mean  $\mathbb{E}[\mathbf{X}_0(t)] = 0$  and of covariance matrix  $\Sigma^0$  is equal to  $\sum_{t-t'}^{Y_t^{t_0} Y_t^{t_0}}$  defined by Eq. 50 and which is independent of  $t$ .

We call *long term mean-field equation* (LTMFE) the implicit equation:

$$\mathbf{V}(t) = \int_{-\infty}^t e^{L(t-s)} (\mathbf{U}_s^V \cdot \mathbf{1} + \mathbf{I}) ds + \mathbf{X}_0(t) \tag{52}$$

where  $\mathbf{X}_0$  is the stationary process defined by Eq. 51 and where  $\mathbf{U}^V(t)$  is the effective interaction process introduced previously.

We next define the long term function  $\mathcal{F}_{\text{stat}}: \mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P)) \rightarrow \mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P))$ :

$$\mathcal{F}_{\text{stat}}(\mathbf{X})_t = \int_{-\infty}^t e^{L(t-s)} (\mathbf{U}_s^X \cdot \mathbf{1} + \mathbf{I}) ds + \mathbf{X}_0(t).$$

**Proposition 5.** The function  $\mathcal{F}_{\text{stat}}$  is well-defined on  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P))$ .

*Proof.* We have already seen that the process  $\mathbf{X}_0$  is well-defined. The term  $\int_{-\infty}^t e^{L(t-s)} \mathbf{I} ds = (\int_{-\infty}^t e^{L(t-s)} ds) \mathbf{I}$  is also well-defined because of the assumptions on  $\mathbf{L}$ .

Let  $X$  be a given process in  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P))$ . To prove the proposition we just have to ensure that the Gaussian process  $\int_{-\infty}^t e^{L(t-s)} \mathbf{U}_s^X \cdot \mathbf{1} ds$  is well-defined. This results from the contraction assumption on  $\mathbf{L}$  and the fact that the functions  $S_\beta$  are bounded. We decompose this process into a “long memory” term  $\int_{-\infty}^0 e^{L(t-s)} \mathbf{U}_s^X \cdot \mathbf{1} ds$  and the interaction term from time  $t = 0$ , namely  $\int_0^t e^{L(t-s)} \mathbf{U}_s^X \cdot \mathbf{1} ds$ . This latter term is clearly well-defined. We show that the memory term is also well-defined as a Gaussian random variable.

We write this term  $e^{L^t} \int_{-\infty}^0 e^{-Ls} \mathbf{U}_s^X \cdot \mathbf{1} ds$  and consider the second factor. This random variable is Gaussian, its mean reads  $\int_0^\infty e^{Ls} \mu^{U_s^X}$  where

$$\mu^{U_s^X} = \left( \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \mathbb{E}[S_\beta(X_\beta(-s))] + I_\alpha \right)_{\alpha=1 \dots P}$$

The integral defining the mean is well-defined because of Eq. 49 and the fact that the functions  $S_\beta$  are bounded. A similar reasoning shows that the corresponding covariance matrix is well-defined. Hence the Gaussian process  $\int_{-\infty}^0 e^{-Ls} \mathbf{U}_s^X \cdot \mathbf{1} ds$  is well-defined, and hence for any process  $X \in \mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P))$ , the process  $\mathcal{F}_{\text{stat}}(X)$  is well-defined.  $\square$

We can now prove the following proposition.

**Proposition 6.** The mean vectors and the covariance matrices of the processes in the image of  $\mathcal{F}_{\text{stat}}$  are bounded.

*Proof.* Indeed, since  $\mathbb{E}[X_0(t)] = 0$ , we have:

$$\|\mathbb{E}[\mathcal{F}_{\text{stat}}(\mathbf{X})_t]\|_{\infty} = \left\| \int_{-\infty}^t e^{L(t-s)} \mu^{U_s^X} ds \right\|_{\infty} \leq M_L (\mu + \|I\|_{\infty}) \stackrel{\text{def}}{=} \mu_{LT}.$$

In a similar fashion the covariance matrices of the processes in the image of  $\mathcal{F}_{\text{stat}}$  are bounded. Indeed we have:

$$\begin{aligned} \mathbb{E}[\mathcal{F}_{\text{stat}}(\mathbf{X})_t \mathcal{F}_{\text{stat}}(\mathbf{X})_t^T] &= \Sigma^0 \\ &+ \int_{-\infty}^t \int_{-\infty}^t e^{L(t-s_1)} \text{diag} \left( \sum_{\beta} \sigma_{\alpha\beta}^2 \mathbb{E}[S_\beta(X_\beta(s_1)) S_\beta(X_\beta(s_2))] \right) e^{L^T(t-s_2)} ds_1 ds_2, \end{aligned}$$

resulting in

$$\|\mathbb{E}[\mathcal{F}_{\text{stat}}(\mathbf{X})_t \mathcal{F}_{\text{stat}}(\mathbf{X})_t^T]\|_{\infty} \leq \|\Sigma^0\|_{\infty} + k^2 \left( \frac{\sigma \max}{\lambda_L} \right)^2 \stackrel{\text{def}}{=} \Sigma_{LT}. \quad \square$$

**Lemma 6.** The set of stationary processes is invariant by  $\mathcal{F}_{\text{stat}}$ .

*Proof.* Since the processes in the image of  $\mathcal{F}_{\text{stat}}$  are Gaussian processes, one just needs to check that the mean of the process is constant in time and that its covariance matrix  $C(s, t)$  only depends on  $t - s$ .

Let  $Z$  be a stationary process and  $Y = \mathcal{F}_{\text{stat}}(Z)$ . We denote by  $\mu_{\alpha}^Z$  the mean of the process  $Z_{\alpha}(t)$  and by  $C_{\alpha}^Z(t-s)$  its covariance function. The mean of the process  $U_{\alpha\beta}^Z$  reads:

$$m_{\alpha,\beta}^Z(t) = \mathbb{E}[S_{\beta}(Z_{\beta}(t))] = \frac{1}{\sqrt{2\pi C_{\beta}^Z(0)}} \int_{\mathbb{R}} S_{\beta}(x) e^{\frac{(x-\mu_{\beta}^Z)^2}{2C_{\beta}^Z(0)}} dx$$

and hence does not depends on time. We note  $\mu^Z$  the mean vector of the stationary process  $\mathbf{U}^Z \cdot \mathbf{1}$ .

Similarly, its covariance function reads:

$$\begin{aligned} \Delta_{\alpha\beta}^Z(t,s) &= \mathbb{E}\left[S_{\beta}(Z_{\beta}(t))S_{\beta}(Z_{\beta}(s))\right] \\ &= \int_{\mathbb{R}^2} S_{\beta}(x)S_{\beta}(y) \exp\left(-\frac{1}{2}\begin{pmatrix} x-\mu_{\beta}^Z \\ y-\mu_{\beta}^Z \end{pmatrix}^T \begin{pmatrix} C_{\beta}^Z(0) & C_{\beta}^Z(t-s) \\ C_{\beta}^Z(t-s) & C_{\beta}^Z(0) \end{pmatrix} \begin{pmatrix} x-\mu_{\beta}^Z \\ y-\mu_{\beta}^Z \end{pmatrix}\right) dx dy \end{aligned}$$

which is clearly a function, noted  $\Delta_{\alpha\beta}^Z(t-s)$ , of  $t-s$ . Hence  $\mathbf{U}^Z \cdot \mathbf{1}$  is stationary and we denote by  $C^{\mu^Z}(t-s)$  its covariance function.

It follows that the mean of  $Y_t$  reads:

$$\begin{aligned} \mu^Y(t) &= \mathbb{E}\left[\mathcal{F}_{\text{stat}}(Z)_t\right] \\ &= \mathbb{E}\left[X_{\theta}(t)\right] + \mathbb{E}\left[\int_{-\infty}^t e^{\mathbf{L}(t-s)} (\mathbf{I} + \mathbf{U}_s^Z \cdot \mathbf{1}) ds\right] \\ &= \int_{-\infty}^t e^{\mathbf{L}(t-s)} (\mathbf{I} + \mathbb{E}\left[\mathbf{U}_s^Z \cdot \mathbf{1}\right]) ds \\ &= \left(\int_{-\infty}^0 e^{\mathbf{L}u} du\right) (\mathbf{I} + \mu^Z) \end{aligned}$$

since we proved that  $\mathbb{E}\left[\mathbf{U}_s^Z \cdot \mathbf{1}\right] = \mu^Z$  was not a function of  $s$ .

Similarly, we compute the covariance function and check that it can be written as a function of  $(t-s)$ . Indeed, it reads:

$$\begin{aligned} C^Y(t,s) &= \int_{-\infty}^t \int_{-\infty}^s e^{\mathbf{L}(t-u)} \text{Cov}\left(\mathbf{U}_u^Z \cdot \mathbf{1}, \mathbf{U}_v^Z \cdot \mathbf{1}\right) e^{\mathbf{L}^T(s-v)} du dv \\ &\quad + \text{Cov}\left(X_{\theta}(t), X_{\theta}(s)\right) \\ &= \int_{-\infty}^0 \int_{-\infty}^0 e^{\mathbf{L}u} C^{\mu^Z}(t-s+(u-v)) e^{\mathbf{L}^T v} du dv + \text{Cov}\left(X_{\theta}(t), X_{\theta}(s)\right) \end{aligned}$$

since the process  $X_{\theta}$  is stationary.  $C^Y(t,s)$  is clearly a function of  $t-s$ . Hence  $Y$  is a stationary process, and the proposition is proved.  $\square$

**Theorem 5.** *The sequence of processes  $\{\mathcal{F}_{\text{stat}}^{(n)}(X)\}_{n=0}^{\infty}$  is uniformly tight.*

*Proof.* The proof is essentially the same as the proof of Theorem 3, since we can write

$$\mathcal{F}_{\text{stat}}(X)_t = e^{\mathbf{L}t} \mathcal{F}_{\text{stat}}(X)_0 + \int_0^t e^{\mathbf{L}(t-s)} (\mathbf{U}_s^X \cdot \mathbf{1} + \mathbf{I}) ds + \int_0^t e^{\mathbf{L}(t-u)} \mathbf{F} d\mathbf{W}_s$$

$\mathcal{F}_{\text{stat}}(X)_t$  appears as the sum of the random variable  $\mathcal{F}_{\text{stat}}(X)_0$  and the Gaussian process defined by  $\int_0^t e^{\mathbf{L}(t-s)} (\mathbf{U}_s^X \cdot \mathbf{1} + \mathbf{I}) ds + \int_0^t e^{\mathbf{L}(t-u)} \mathbf{F} d\mathbf{W}_s$  which is equal to  $\mathcal{F}_k(X)_t$  defined in Section “Existence and Uniqueness of Solutions in Finite Time” for  $t_0 = 0$ . Therefore  $\mathcal{F}_{\text{stat}}^{(n)}(X)_t = \mathcal{F}_k^{(n)}(X)_t$  for  $t > 0$ . We have proved the uniform tightness of the sequence of processes  $\{\mathcal{F}_k^{(n)}(X)\}_{n=0}^{\infty}$  in Theorem 3. Hence, according to Kolmogorov’s criterion for tightness, we just have to prove that the sequence of Gaussian random variables:

$$\mathcal{F}_{\text{stat}}^{(n)}(X)_0 = \left\{ \int_{-\infty}^0 \Phi_L(-u) (\mathbf{U}_u^{\mathcal{F}_{\text{stat}}^{(n)}(X)} \cdot \mathbf{1} + \mathbf{I}) du + \mathbf{X}_0(0) \right\}_{n \geq 0}$$

is uniformly tight. Since it is a sequence of Gaussian random variables, it is sufficient to prove that their means are bounded and

their covariance matrices upperbounded to obtain that for any  $\varepsilon > 0$  there exists a compact  $K_{\varepsilon}$  such that for any  $n \in \mathbb{N}$ , we have  $\mathbb{P}(\mathcal{F}_{\text{stat}}^{(n)}(X)_0 \in K_{\varepsilon}) \geq 1 - \varepsilon$ . This is a consequence of Proposition 6 for the first random variable and of the definition of  $\mathbf{X}_0$  for the second. By Kolmogorov’s criterion the sequence of processes  $\{\mathcal{F}_{\text{stat}}^{(n)}(X)\}_{n=0}^{\infty}$  is uniformly tight.  $\square$

In order to apply Theorem 1 we need to prove that the sequences of covariance and mean functions are convergent. Unlike the case of  $t_0$  finite, this is not always true. Indeed, to ensure existence and uniqueness of solutions in the stationary case, the parameters of the system have to satisfy a contraction condition, and Proposition 3 extends as follows.

**Proposition 7.** *If  $\lambda_L$  defined in Eq. 48 satisfies the conditions (Eq. 53) defined in the proof, depending upon  $k_c$  (defined in Eq. 45),  $k_0$ ,  $\mu_{LT}$  and  $\Sigma_{LT}$  (defined in Proposition 6) then the sequences of covariance matrix functions  $C^n(t,s)$  and of mean functions  $\mu^n(t), s, t$  in  $[t_0, T]$  are Cauchy sequences for the uniform norms.*

*Proof.* The proof follows that of Proposition 3 with a few modifications that we indicate. In establishing the equation corresponding to Eq. 44 we use the fact that  $\|\Phi_L(t,u)\|_{\infty} \leq ke^{-\lambda_L(t-u)}$  for some positive constant  $k$  and all  $u, t, u \leq t$ . We therefore have:

$$\begin{aligned} \|C^{n+1}(t,s) - C^n(t,s)\|_{\infty} &\leq k^2 e^{-\lambda_L(t+s)} \int_{-\infty}^t \int_{-\infty}^s e^{-\lambda_L(u+v)} \|\text{Cov}\left(\mathbf{U}_u^{X_n}, \mathbf{U}_v^{X_n}\right) \\ &\quad - \text{Cov}\left(\mathbf{U}_u^{X_{n-1}}, \mathbf{U}_v^{X_{n-1}}\right)\|_{\infty}^2 du dv \end{aligned}$$

The rest of the proof proceeds the same way as in Proposition 3. Equations 46 and 47 become:

$$\begin{aligned} \|C^{n+1}(t,s) - C^n(t,s)\|_{\infty} &\leq Ke^{-\lambda_L(t+s)} \left( \int_{[-\infty, t \vee s]^2} \frac{e^{\lambda_L(u+v)}}{\sqrt{f(u,v)}} \|C^n(u,v) - C^{n-1}(u,v)\|_{\infty} dudv \right. \\ &\quad + \int_{[-\infty, t \vee s]^2} \frac{e^{\lambda_L(u+v)}}{\sqrt{f(u,v)}} \|C^n(u,u) - C^{n-1}(u,u)\|_{\infty} dudv \\ &\quad + \int_{[-\infty, t \vee s]^2} e^{\lambda_L(u+v)} \|C^n(u,v) - C^{n-1}(u,v)\|_{\infty} dudv \\ &\quad + \int_{[-\infty, t \vee s]^2} e^{\lambda_L(u+v)} \|C^n(u,u) - C^{n-1}(u,u)\|_{\infty} dudv \\ &\quad \left. + \int_{[-\infty, t \vee s]^2} e^{\lambda_L(u+v)} \|\mu^n(u) - \mu^{n-1}(u)\|_{\infty} dudv \right), \end{aligned}$$

and

$$\begin{aligned} \|\mu^{n+1}(t) - \mu^n(t)\|_{\infty} &\leq Ke^{-\lambda_L(t+s)} \left( \int_{[-\infty, t \vee s]^2} e^{\lambda_L(u+v)} \|C^n(u,u) \right. \\ &\quad \left. - C^{n-1}(u,u)\|_{\infty} dudv + \int_{[-\infty, t \vee s]^2} e^{\lambda_L(u+v)} \|\mu^n(u) - \mu^{n-1}(u)\|_{\infty} dudv \right), \end{aligned}$$

for some positive constant  $K$ , function of  $k, k_c$  (defined in Eq. 45), and  $k_0$ .

Proceeding recursively until we reach  $C^0$  and  $\mu^0$  we obtain an upperbound for  $\|C^{n+1}(t,s) - C^n(t,s)\|_{\infty}$  (respectively for  $\|\mu^{n+1}(t) - \mu^n(t)\|_{\infty}$ ) which is the sum of  $<5^n$  terms each one being

the product of  $K^n$ , times  $2 \mu_{LT}$  or  $2 \Sigma_{LT}$ , times a  $2n$ -dimensional integral  $I_n$  given by:

$$\int_{[-\infty, t \vee s]^2} \rho_1(u_1, v_1) \left( \int_{[-\infty, u_1 \vee v_1]^2} \dots \left( \int_{[-\infty, u_{n-2} \vee v_{n-2}]^2} \rho_{n-1}(u_{n-1}, v_{n-1}) \right. \right. \\ \left. \left. \times \left( \int_{[-\infty, u_{n-1} \vee v_{n-1}]^2} e^{\lambda_L(u_n + v_n)} \rho_n(u_n, v_n) du_n dv_n \right) du_{n-1} dv_{n-1} \right) \dots \right) du_1 dv_1,$$

where the functions  $\rho_i(u_i, v_i)$ ,  $i = 1, \dots, n$  are either equal to 1 or to  $1/\sqrt{\theta(u_i, v_i)}$ .

It can be shown by straightforward calculation that each sub-integral contributes at most either

$$\frac{K_0}{\lambda_L^2} \text{ if } \rho_i = 1 \text{ or } \sqrt{\frac{\pi}{2}} \frac{K_0}{\lambda_L^{\frac{3}{2}}},$$

in the other case. Hence we obtain factors of the type

$$K_0^n \left( \frac{1}{\lambda_L^2} \right)^p \left( \sqrt{\frac{\pi}{2}} \frac{K_0}{\lambda_L^{3/2}} \right)^{n-p} = \left( \sqrt{\frac{\pi}{2}} \right)^{n-p} \left( \frac{1}{\lambda_L} \right)^{\frac{(3n+p)}{2}} K_0^n,$$

where  $0 \leq p \leq n$ . If  $\lambda_L < 1$ ,  $(\lambda_L)^{(3n+p)/2} \geq \lambda_L^{2n}$  and else  $(\lambda_L)^{(3n+p)/2} \geq \lambda_L^{3n/2}$ . Since  $(\sqrt{\pi/2})^{n-p} \leq (\sqrt{\pi/2})^n$  we obtain the two conditions

$$1 > \lambda_L^2 \geq 5 \sqrt{\frac{\pi}{2}} K K_0 \text{ or } \left\{ \lambda_L^{\frac{3}{2}} > 5 \sqrt{\frac{\pi}{2}} K K_0 \text{ and } \lambda_L \geq 1 \right\} \quad (53)$$

Putting all these results together we obtain the following theorem of existence and uniqueness of solutions for the LTMFE:

**Theorem 6.** *Under the contraction conditions (Eq. 53), the function  $\mathcal{F}_{stat}$  has a unique solution in  $\mathcal{M}_1^+(C((-\infty, T], \mathbb{R}^P))$  which is stationary, and for any process  $X$ , the sequence  $\{\mathcal{F}_{stat}^{(n)}(X)\}_{n=0}^\infty$  of Gaussian processes converges in law toward the unique fixed point of the function  $\mathcal{F}_{stat}$ .*

*Proof.* The proof is essentially similar to the one of Theorem 4. Indeed, the mean and the covariance matrixes converge since they are Cauchy sequences in the complete space of continuous functions equipped with the uniform norm. Using Theorem 1, we obtain that the sequence converges to a process  $Y$  which is necessarily a fixed point of  $\mathcal{F}_{stat}$ . Hence we have existence of a fixed point for  $\mathcal{F}_{stat}$ . The uniqueness comes from the results obtained in the proof of Proposition 7. The limiting process is necessarily stationary. Indeed, let  $X$  be a stationary process. Then for any  $n \in \mathbb{N}$ , the process  $\mathcal{F}_{stat}^{(n)}(X)$  will be stationary by the virtue of Lemma 6, and hence so will be the limiting process which is the only fixed point of  $\mathcal{F}_{stat}$ .  $\square$

Hence in the stationary case, the existence and uniqueness of a solution is not always ensured. For instance if the leaks are too small (i.e., when the time constants of the decay of the membrane potentials are too long) then the sequence can diverge or have multiple fixed points.

## NUMERICAL EXPERIMENTS

### SIMULATION ALGORITHM

Beyond the mathematical results, the framework that we introduced in the previous sections gives us a strategy to compute numerically the solutions of the dynamic mean-field equations. Indeed, we proved in Section “Existence and Uniqueness of Solutions in Finite

Time” that under very moderate assumptions on the covariance matrix of the noise, the iterations of the map  $\mathcal{F}_k$  starting from any initial condition converge to the solution of the mean-field equations.

This convergence result gives us a direct way to compute numerically the solution of the mean-field equations. Since we are dealing with Gaussian processes, determining the law of the iterates of the map  $\mathcal{F}_k$  amounts to computing its mean and covariance functions. In this section we describe our numerical algorithm in the case of the Simple Model of Section “Example I: The Simple Model”.

### Computing $\mathcal{F}_k$

Let  $X$  be a  $P$ -dimensional Gaussian process of mean  $\mu^X = (\mu_\alpha^X(t))_{\alpha=1 \dots P}$  and covariance  $C^X = (C_{\alpha\beta}^X(s, t))_{\alpha, \beta \in \{1 \dots P\}}$ . We fix a time interval  $[t_0 = 0, T]$  and denote by  $Y$  the image of the process  $X$  under  $\mathcal{F}_1$ . In the case of the simple model, the covariance of  $Y$  is diagonal. Hence in this case the expressions we obtain in Section “Existence and Uniqueness of Solutions in Finite Time” simply read:

$$\mu_\alpha^Y(t) = \mu_\alpha^X(0) e^{-\frac{t}{\tau_\alpha}} + \int_0^t e^{-\frac{(t-s)}{\tau_\alpha}} \left( \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \mathbb{E}[S_\beta(X_\beta(s))] + I_\alpha(s) \right) ds \\ = \mu_\alpha^X(0) e^{-\frac{t}{\tau_\alpha}} + \int_0^t e^{-\frac{(t-s)}{\tau_\alpha}} I_\alpha(s) ds \\ + \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \int_0^t e^{-\frac{(t-s)}{\tau_\alpha}} \int_{-\infty}^{+\infty} S_\beta \left( x \sqrt{v_\beta^X(s)} + \mu_\beta^X(s) \right) Dxds,$$

where we denoted  $v_\alpha^X(s)$  the standard deviation of  $X_\alpha$  at time  $s$ , instead of  $C_{\alpha\alpha}^X(s, s)$ . Thus, knowing  $v_\alpha^X(s)$ ,  $s \in [0, t]$  we can compute  $\mu_\alpha^Y(t)$  using a standard discretization scheme of the integral, with a small time step compared with  $\tau_\alpha$  and the characteristic time of variation of the input current  $I_\alpha$ . Alternatively, we can use the fact that  $\mu_\alpha^Y$  satisfies the differential equation:

$$\frac{d\mu_\alpha^Y}{dt} = -\frac{\mu_\alpha^Y}{\tau_\alpha} + \sum_{\beta=1}^P \bar{J}_{\alpha\beta} \int_{-\infty}^{+\infty} S_\beta \left( x \sqrt{v_\beta^X(t)} + \mu_\beta^X(t) \right) Dxd + I_\alpha(t),$$

and compute the solution using a Runge–Kutta algorithm (which is faster and more accurate). Note that, when all the standard deviations of the process  $X$  are null for all time  $t \in [0, T]$ , we obtain a standard dynamical system. Nevertheless, in the general case,  $v_\beta^X(t) > 0$  for some  $\beta$ 's, and the dynamical evolution of  $\mu_\alpha^Y$  depends on the Gaussian fluctuations of the field  $X$ . These fluctuations must be computed via the complete equation of the covariance diagonal coefficient  $C_{\alpha\alpha}^Y(t, s)$ , which reads:

$$C_{\alpha\alpha}^Y(t, s) = e^{-(t+s)/\tau_\alpha} \left[ v_\alpha^X(0) + \frac{\tau_\alpha \sigma_\alpha^2}{2} (e^{2s/\tau_\alpha} - 1) \right. \\ \left. + \sum_{\beta=1}^P \sigma_{\alpha\beta}^2 \int_0^t \int_0^s e^{(u+v)/\tau_\alpha} \Delta_{\alpha\beta}^X(u, v) dudv \right],$$

where:

$$\Delta_{\alpha\beta}^X(u, v) = \int_{\mathbb{R}^2} S_\beta \left( x \frac{\sqrt{v_\beta^X(u)} v_\beta^X(v) - C_{\beta\beta}^X(u, u)^2}{\sqrt{v_\beta^X(v)}} + y \frac{C_{\beta\beta}^X(u, v)}{\sqrt{v_\beta^X(v)}} + \mu_\beta^X(u) \right) \\ \times S_\beta \left( y \sqrt{v_\beta^X(v)} + \mu_\beta^X(v) \right) DxDy.$$



Unless if we assume the stationarity of the process (see e.g., Section “The Importance of the Covariance: Simple Model, One Population”), this equation cannot be written as an ordinary differential equation. We clearly observe here the non-Markovian nature of the problem:  $C_{\alpha\alpha}^X(t, s)$  depends on the whole past of the process until time  $t \vee s$ .

This covariance can be split into the sum of two terms: the external noise contribution  $C_{\alpha\alpha}^{OU}(t, s) = e^{-(t+s)/\tau_\alpha} [\nu_\alpha^X(0) + \frac{\tau_\alpha \sigma_\alpha^2}{2} (e^{2s/\tau_\alpha} - 1)]$  and the interaction between the neurons. The external noise contribution is a simple function and can be computed directly. To compute the interactions contribution to the standard deviation we have to compute the symmetric two-variables function:

$$H_{\alpha\beta}^X(t, s) = e^{-(t+s)/\tau_\alpha} \int_0^t \int_0^s e^{(u+v)/\tau_\alpha} \Delta_{\alpha\beta}^X(u, v) du dv,$$

from which one obtains the standard deviation using the formula

$$C_{\alpha\alpha}^Y(t, s) = C_{\alpha\alpha}^{OU}(t, s) + \sum_{\beta=1}^p \sigma_{\alpha\beta}^2 H_{\alpha\beta}^X(t, s).$$

To compute the function  $H_{\alpha\beta}^X(t, s)$ , we start from  $t = 0$  and  $s = 0$ , where  $H_{\alpha\beta}^X(0, 0) = 0$ . We only compute  $H_{\alpha\beta}^X(t, s)$  for  $t > s$  because of the symmetry. It is straightforward to see that:

$$H_{\alpha\beta}^X(t + dt, s) = H_{\alpha\beta}^X(t, s) \left[ 1 - \frac{dt}{\tau_\alpha} \right] + D_{\alpha\beta}^X(t, s) dt + o(dt),$$

with

$$D_{\alpha\beta}^X(t, s) = e^{-s/\tau_\alpha} \int_0^s e^{v/\tau_\alpha} \Delta_{\alpha\beta}^X(t, v) dv.$$

Hence computing  $H_{\alpha\beta}^X(t + dt, s)$  knowing  $H_{\alpha\beta}^X(t, s)$  amounts to computing  $D_{\alpha\beta}^X(t, s)$ . Fix  $t \geq 0$ . We have  $D_{\alpha\beta}^X(t, 0) = 0$  and

$$D_{\alpha\beta}^X(t, s + ds) = D_{\alpha\beta}^X(t, s) \left( 1 - \frac{ds}{\tau_\alpha} \right) + \Delta_{\alpha\beta}^X(t, s) ds + o(ds).$$

This algorithm enables us to compute  $H_{\alpha\beta}^X(t, s)$  for  $t > s$ . We deduce  $H_{\alpha\beta}^X(t, s)$  for  $t < s$  using the symmetry of this function. Finally, to get the values of  $H_{\alpha\beta}^X(t, s)$  for  $t = s$ , we use the symmetry property of this function and get:

$$H_{\alpha\beta}^X(t + dt, t + dt) = H_{\alpha\beta}^X(t, t) \left[ 1 - \frac{2dt}{\tau_\alpha} \right] + 2D_{\alpha\beta}^X(t, t) dt + o(dt).$$

These numerical schemes provide an efficient way for computing the mean and the covariance functions of the Gaussian process  $\mathcal{F}_1(X)$  (hence its probability distribution) knowing the law of the Gaussian process  $X$ . The algorithm used to compute the solution of the mean-field equations for the general models GM1 and GMk is a straightforward generalization.

**Analysis of the algorithm**

**Convergence rate.** As proved in Theorem 4, given  $Z_0$  a nondegenerate  $kP$ -dimensional Gaussian random variable and  $X$  a Gaussian process such that  $X(0) = Z_0$ , the sequences of mean values and covariance functions computed theoretically converge uniformly toward those of the unique fixed point of the map  $\mathcal{F}_k$ . It is clear that our algorithm converges uniformly toward the real function it emulates.

Hence for a finite  $N$ , the algorithm will converge uniformly toward the mean and covariance matrix of the process  $\mathcal{F}_k^N(X)$ .

Denote by  $X_f$  the fixed point of  $\mathcal{F}_k$  in  $\mathcal{M}_1^+(C([t_0, T], \mathbb{R}^{kP}))$ , of mean  $\mu^{X_f}(t)$  and covariance matrix  $C^{X_f}(t, s)$  and by  $\widehat{\mathcal{F}}_k^N(X)$  the numerical approximation of  $\mathcal{F}_k^N(X)$  computed using the algorithm previously described, whose mean is noted  $\mu^{\widehat{\mathcal{F}}_k^N(X)}(t)$  and whose covariance matrix is noted  $C^{\widehat{\mathcal{F}}_k^N(X)}(t, s)$ . The uniform error between the simulated mean after  $N$  iterations with a time step  $dt$  and the fixed point's mean and covariance is the sum of the numerical error of the algorithm and the distance between the simulated process and the fixed point, is controlled by:

$$\left\| \mu^{\widehat{\mathcal{F}}_k^N(X)} - \mu^{X_f} \right\|_\infty + \left\| C^{\widehat{\mathcal{F}}_k^N(X)} - C^{X_f} \right\|_\infty = O((N + T)dt + R_N(k_{\max})) \quad (54)$$

where  $k_{\max} = \max(k, \tilde{k})$  and  $k$  and  $\tilde{k}$  are the constants that appear in the proof of Proposition 3 for the mean and covariance functions, and  $R_N(x)$  is the exponential remainder, i.e.,  $R_N(x) = \sum_{n=N}^\infty x^{n/n!}$ .

Indeed, we have:

$$\left\| \mu^{\widehat{\mathcal{F}}_k^N(X)} - \mu^{X_f} \right\|_\infty \leq \left\| \mu^{\widehat{\mathcal{F}}_k^N(X)} - \mu^{\mathcal{F}_k^N(X)} \right\|_\infty + \left\| \mu^{\mathcal{F}_k^N(X)} - \mu^{X_f} \right\|_\infty \quad (55)$$

The discretization algorithm used converges in  $O(dt)$ . Let us denote by  $C_1$  the convergence constant, which depends on the sharpness of the function we approximate, which can be uniformly controlled over the iterations. Iterating the numerical algorithm has the effect of propagating the errors. Using these simple remarks we can bound the first term of the righthand side of Eq. 55, i.e., the approximation error at the  $N$ th iteration:

$$\left\| \mu^{\widehat{\mathcal{F}}_k^N(X)} - \mu^{\mathcal{F}_k^N(X)} \right\|_\infty \leq C_1 N dt$$

Because the sequence of mean values is a Cauchy sequence, we can also bound the second term of the righthand side of Eq. 55:

$$\left\| \mu^{\mathcal{F}_k^N(X)} - \mu^{X_f} \right\|_\infty \leq \sum_{n=N}^\infty \left\| \mu^{\mathcal{F}_k^{n+1}(X)} - \mu^{\mathcal{F}_k^n(X)} \right\|_\infty \leq \sum_{n=N}^\infty \frac{k^n}{n!} =: R_N(k)$$

for some positive constant  $k$  introduced in the proof of Proposition 3. The remainders sequence  $(R_n(k))_{n \geq 0}$  converges fast toward 0 (an estimation of its convergence can be obtained using the fact that  $\limsup_{k \rightarrow \infty} (\frac{1}{k!})^{1/k} = 0$  by Stirling's formula).

Hence we have:

$$\left\| \mu^{\widehat{\mathcal{F}}_k^N(X)} - \mu^{X_f} \right\|_\infty \leq C_1 N dt + R_N(k) \quad (56)$$

For the covariance, the principle of the approximation is exactly the same:

$$\left\| C^{\widehat{\mathcal{F}}_k^N(X)} - C^{X_f} \right\|_\infty \leq \left\| C^{\widehat{\mathcal{F}}_k^N(X)} - C^{\mathcal{F}_k^N(X)} \right\|_\infty + \left\| C^{\mathcal{F}_k^N(X)} - C^{X_f} \right\|_\infty$$

The second term of the righthand side can be controlled using the same evaluation by  $R_N(\tilde{k})$  where  $\tilde{k}$  is the constant introduced in the proof of Proposition 3, and the first term is controlled by the rate of convergence of the approximation of the double integral, which is bounded by  $C_2(N + T) dt$  where  $C_2$  depends on the parameters of the system and the discretization algorithm used.

Hence we have:

$$\left\| C^{\tilde{x}^N}(x) - C^{X_f} \right\|_{\infty} \leq C_2(N + T - t_0)dt + R_N(\tilde{k}) \tag{57}$$

The expressions (Eqs. 56 and 57) are the sum of two terms, one of which is increasing with  $N$  and  $T$  and decreasing with  $dt$  and the other one decreasing in  $N$ . If we want to obtain an estimation with an error bounded by some  $\varepsilon > 0$ , we can for instance fix  $N$  such that  $\max(R_N(k), R_N(\tilde{k})) < \frac{\varepsilon}{2}$  and then fix the time step  $dt$  smaller than  $\min(\frac{\varepsilon}{(2C_2N)}, \frac{\varepsilon}{(2C_2(N+T-t_0))})$ .

**Complexity.** The complexity of the algorithm depends on the complexity of the computations of the integrals. The algorithm described hence has the complexity  $O(N(\frac{T}{dt})^2)$ .

**THE IMPORTANCE OF THE COVARIANCE: SIMPLE MODEL, ONE POPULATION**

As a first example and a benchmark for our numerical scheme we revisit the work of Sompolinsky et al. (1998). These authors studied the case of the simple model with one population ( $P = 1$ ), with the centered sigmoidal function  $S(x) = \tanh(gx)$ , centered connectivity weights  $\bar{J} = 0$  of standard deviation  $\sigma = 1$  and no input ( $I = 0; \Lambda = 0$ ). Note therefore that there is no “noise” in the system, which therefore does not match the nondegeneracy conditions of Proposition 2 and of Theorem 4. This issue is discussed below. In this case, the mean equals 0 for all  $t$ . Nevertheless, the Gaussian process is nontrivial as revealed by the study of the covariance  $C(t, s)$ .

**Stationary solutions**

Assuming that the solution of the mean-field equation is a stationary solution with  $C(t, s) \equiv C(t - s) = c(\tau)$ , Sompolinsky and his collaborators found that the covariance obeyed a second-order differential equation:

$$\frac{d^2C}{d\tau^2} = -\frac{\partial V_q}{\partial C} \tag{58}$$

This form corresponds to the motion of a particle in a potential well and it is easy to draw the phase portrait of the corresponding dynamical system. However, there is a difficulty. The potential  $V_q$  depends on a parameter  $q$  which is in fact precisely the covariance at  $\tau = 0$  ( $q = C(0)$ ). In the stationary case, this covariance depends on the whole solution, and hence cannot be really considered as a parameter of the system. This is one of the main difficulties in this approach: mean-field equations in the stationary regime are self-consistent.

Nevertheless, the study of the shape of  $V_q$ , considering  $q$  as a free parameter gives us some information. Indeed,  $V_q$  has the following Taylor expansion ( $V_q$  is even because  $S$  is odd):

$$V_q(C) = \frac{\lambda}{2}C^2 + \frac{\gamma}{4}C^4 + O(C^6)$$

where  $\lambda = (1 - g^2 J^2 \langle S' \rangle_q^2)$  and  $\gamma = \frac{1}{6} J^2 g^2 \langle S^{(3)} \rangle_q^2, \langle \phi \rangle_q$  being the average value of  $\phi$  under the Gaussian distribution with mean 0 and variance  $q = C(0)$ .

If  $\lambda > 0$ , i.e., when  $g^2 J^2 \langle S' \rangle_q^2 < 1$ , then the dynamical system (Eq. 58) has a unique solution  $C(t) = 0, \forall t \geq 0$ . This corresponds to a stable fixed point (i.e., a deterministic trajectory,  $\mu = 0$  with

no fluctuations) for the neural network dynamics. On the other hand, if  $g^2 J^2 \langle S' \rangle_q^2 \geq 1$ , there is a homoclinic trajectory in Eq. 58 connecting the point  $q = C^* > 0$  where  $V_q$  vanishes to the point  $C = 0$ . This solution is interpreted by the authors as a chaotic solution in the neural network. A stability analysis shows that this is the only stable<sup>7</sup> stationary solution (Sompolinsky et al., 1998).

The equation for the homoclinic solution is easily found using energy conservation and the fact that  $V_q(q) = 0$  and  $\frac{dV_q}{dC}(q) = 0$ . One finds:

$$u = \frac{dC}{dx} = -\sqrt{-V_q(C)}$$

At the fourth-order in the Taylor expansion of  $V_q$  this gives

$$C(\tau) = \frac{\sqrt{\frac{-2\lambda}{\gamma}}}{\cosh\left(\sqrt{\frac{\lambda}{2}}\tau\right)}$$

Though  $\lambda$  depends on  $q$  it can be used as a free parameter for interpolating the curve of  $C(\tau)$  obtained from numerical data.

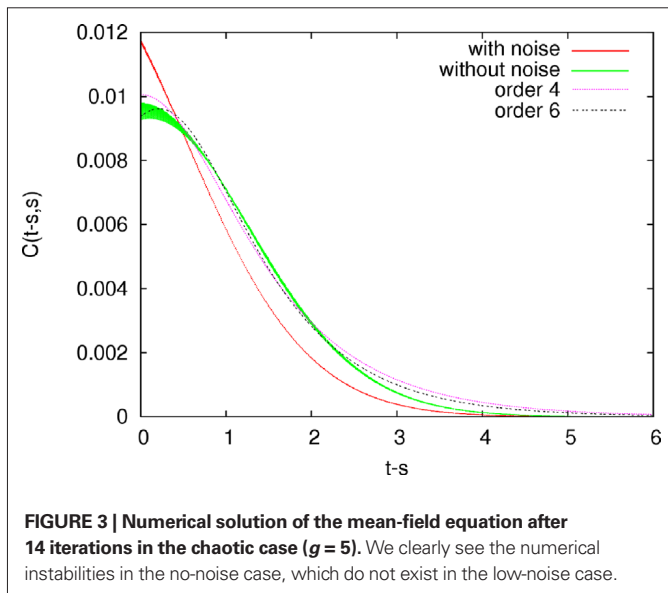
**Numerical experiments**

This case is a good benchmark for our numerical procedure since we know analytically the solutions we are searching for. We expect to find two regimes. In one case the correlation function is identically 0 in the stationary regime, for sufficiently small  $g$  values or for a sufficiently small  $q$  (trivial case). The other case corresponds to a regime where  $C(\tau) > 0$  and  $C(\tau) \rightarrow 0$  has  $\tau \rightarrow +\infty$  (“chaotic” case). This regime requires that  $g$  be sufficiently large and that  $q$  be large too. We took  $\tau_\alpha = 0.25, \sigma_{\alpha\alpha} = 1$ . For these values, the change in dynamics predicted by Sompolinsky and collaborators is  $g_c = 4$ .

In Sections “Existence and Uniqueness of Solutions in Finite Time” and “Existence and Uniqueness of Stationary Solutions” we have introduced the assumption of nondegeneracy of the noise, in order to ensure that the mean-field process was nondegenerate. However, in the present example, there is no external noise in the evolution, so we can observe the effects of relaxing this hypothesis in a situation where the results of Proposition 2 and of Theorem 4 cannot be applied. First, we observed numerically that, without external noise, the process could become degenerate [namely some eigenvalues of the covariance matrix  $C_\alpha(t, s)$  become very small and even vanish.]. This has also an incidence on the convergence of the method which presents numerical instabilities, though the iterations leads to a curve which is well fitted by the theoretical results of Sompolinsky et al. (see **Figure 3**). The instability essentially disappears if one adds a small noise. But, note that in this case, the solution does not match with Sompolinsky et al. theoretical calculation (see **Figure 3**).

Modulo this remark, we have first considered the trivial case corresponding to small  $g$  values. We took  $g = 0.5$  and  $T = 5$ . We choose as initial process the stationary Ornstein–Uhlenbeck process corresponding to the uncoupled system with  $\Lambda = 0.1$ . We drew  $\mu_\alpha(0)$  randomly from the uniform distribution in  $[-1, 1]$  and  $\nu_\alpha(0)$  randomly from the uniform distribution in  $[0, 1]$ .

<sup>7</sup>More precisely, this is the only minimum for the large deviation functional.



**FIGURE 3 | Numerical solution of the mean-field equation after 14 iterations in the chaotic case ( $g = 5$ ).** We clearly see the numerical instabilities in the no-noise case, which do not exist in the low-noise case.

Starting from this initial stationary process, we iterated the function  $\mathcal{F}_1$ . Then, during the iterations, we set  $s_\alpha = 0$  in order to match the conditions imposed by Sompolinsky and colleagues. We observe that the method converges toward the expected solution: the mean function converges to 0, while the variance  $v(t)$  decreases exponentially fast in time toward a constant value corresponding to the stationary regime. This asymptotic value decreases between two consecutive iterations, which is consistent with the theoretical expectation that  $v(t) = 0$  in the stationary regime of the trivial case. Finally, we observe that the covariance  $C(t - s, s)$  stabilizes to a curve that does not depend on  $s$  and the stationary value (large  $t - s$ ) converges to 0.

We applied the same procedure for  $g = 5$  corresponding to the “chaotic” regime. The behavior was the same for  $\mu(t)$  but was quite different for the covariance function  $C(t, s)$ . Indeed, while in the first case the stationary value of  $v(t)$  tends to 0 with the number of iterations, in the chaotic case it stabilizes to a finite value. In the same way, the covariance  $C(t - s, s)$  stabilizes to a curve that does not depend on  $s$ . The shape of this curve can be extrapolated thanks to Sompolinsky et al. results. We observe a very good agreement with the theoretical predictions with a fit  $f_4(x) = \frac{\rho}{\cosh(b(x-\delta))}$ , corresponding to the fourth expansion of  $V_q$ . Using a sixth-order expansion of  $V_q(x) = \frac{a}{2}x^2 + \frac{b}{4}x^4 + \frac{c}{6}x^6$  gives a fit

$$f_6(x) = \frac{\rho}{\cosh(\lambda(x-\delta))} \frac{1}{\sqrt{1 + K^2 - \frac{1}{\cosh^2(\lambda(x-\delta))}}},$$

where  $\rho, K, \lambda$  are explicit functions of  $a, b, c$ , we obtain a slightly better approximation.

**MEAN-FIELD EQUATIONS FOR TWO POPULATIONS WITH A NEGATIVE FEEDBACK LOOP**

Let us now present a case where the fluctuations of the Gaussian field act on the dynamics of  $\mu_\alpha(t)$  in a nontrivial way, with a behavior strongly departing from the naive mean-field picture. We consider two interacting populations where the connectivity weights are Gaussian random variables  $J_{\alpha\beta} \equiv \mathcal{N}(\bar{J}_{\alpha\beta}, \sigma_{\alpha\beta} = 1)$  for  $(\alpha, \beta) \in \{1, 2\}^2$ . We set  $S_\beta(x) = \tanh(gx)$  and  $I_\alpha = 0, s_\alpha = 0, \alpha = 1, 2$ .

**Theoretical framework**

The dynamic mean-field equation for  $\mu_\alpha(t)$  is given, in differential form, by:

$$\frac{d\mu_\alpha}{dt} = -\frac{\mu_\alpha}{\tau_\alpha} + \sum_{\beta=1}^2 \bar{J}_{\alpha\beta} \int_{-\infty}^{\infty} S(\sqrt{v_\beta(t)}x + \mu_\beta(t))Dx, \quad \alpha = 1, 2.$$

Let us denote by  $G_\alpha(\mu, v(t))$  the function in the righthand side of the equality. Since  $S$  is odd,  $\int_{-\infty}^{\infty} S(\sqrt{v_\beta(t)}x)Dx = 0$ . Therefore, we have  $G_\alpha(0, v(t)) = 0$  whatever  $v(t)$ , and hence the point  $\mu_1 = 0, \mu_2 = 0$  is always a fixed point of this equation.

Let us study the stability of this fixed point. To this purpose, we compute the partial derivatives of  $G_\alpha(\mu, v(t))$  with respect to  $\mu_\beta$  for  $(\alpha, \beta) \in \{1, 2\}^2$ . We have:

$$\frac{\partial G_\alpha}{\partial \mu_\beta}(\mu, v(t)) = -\frac{\delta_{\alpha\beta}}{\tau_\alpha} + g\bar{J}_{\alpha\beta} \int_{-\infty}^{\infty} \left(1 - \tanh^2(\sqrt{v_\beta(t)}x + \mu_\beta(t))\right)Dx,$$

and hence at the point  $\mu_1 = 0, \mu_2 = 0$ , these derivatives read:

$$\frac{\partial G_\alpha}{\partial \mu_\beta}(0, v(t)) = -\frac{\delta_{\alpha\beta}}{\tau_\alpha} + g\bar{J}_{\alpha\beta} h(v_\beta(t)),$$

$$\text{where } h(v_\beta(t)) = 1 - \int_{-\infty}^{\infty} \tanh^2(\sqrt{v_\beta(t)}x)Dx.$$

In the case  $v_\alpha(0) = 0, J = 0, s_\alpha = 0$ , implying  $v_\alpha(t) = 0, t \geq 0$ , the equation for  $\mu_\alpha$  reduces to:

$$\frac{d\mu_\alpha}{dt} = -\frac{\mu_\alpha}{\tau_\alpha} + \sum_{\beta=1}^2 \bar{J}_{\alpha\beta} S(\mu_\beta(t))$$

which is the standard Amari–Cohen–Grossberg–Hopfield system. This corresponds to the naive mean-field approach where Gaussian fluctuations are neglected. In this case the stability of the fixed point  $\mu = 0$  is given by the sign of the largest eigenvalue of the Jacobian matrix of the system that reads:

$$\begin{pmatrix} -\frac{1}{\tau_1} & 0 \\ 0 & -\frac{1}{\tau_2} \end{pmatrix} + g \begin{pmatrix} \bar{J}_{11} & \bar{J}_{12} \\ \bar{J}_{21} & \bar{J}_{22} \end{pmatrix}.$$

For the sake of simplicity we assume that the two time constants  $\tau_\alpha$  are equal and we denote this value  $\tau$ . The eigenvalues are in this case  $-\frac{1}{\tau} + g\lambda$ , where  $\lambda$  are the eigenvalues of  $\bar{J}$  and have the form:

$$\lambda_{1,2} = \frac{\bar{J}_{11} + \bar{J}_{22} \pm \sqrt{(\bar{J}_{11} - \bar{J}_{22})^2 + 4\bar{J}_{12}\bar{J}_{21}}}{2}.$$

Hence, they are complex whenever  $\bar{J}_{12}\bar{J}_{21} < -(\bar{J}_{11} - \bar{J}_{22})^2 / 4$ , corresponding to a negative feedback loop between population 1 and 2. Moreover, they have a real part only if  $\bar{J}_{11} + \bar{J}_{22}$  is nonzero (self interaction).

This opens up the possibility to have an instability of the fixed point ( $\mu = 0$ ) leading to a regime where the average value of the membrane potential oscillates. This occurs if  $\bar{J}_{11} + \bar{J}_{22} > 0$  and if  $g$  is larger than:

$$g_c = \frac{2}{\tau(\bar{J}_{11} + \bar{J}_{22})}.$$

The corresponding bifurcation is a Hopf bifurcation.

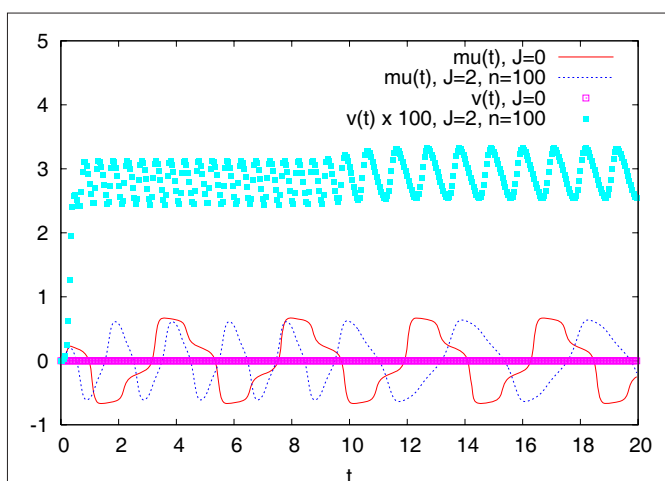
The situation is different if one takes into account the fluctuations of the Gaussian field. Indeed, in this case the stability of the fixed point  $\mu = 0$  depends on  $v(t)$ . More precisely, the real and imaginary part of the eigenvalues of  $DG(0, v(t))$  depend on  $v(t)$ . Therefore, the variations of  $v(t)$  act on the stability and oscillations period of  $v(t)$ . Though the evolution of  $\mu(t)$ ,  $v(t)$  are coupled we cannot consider this evolution as a coupled dynamical system, since  $v(t) = C(t, t)$  is determined by the mean-field equation for  $C(t, s)$  which cannot be written as an ordinary differential equation. Note that we cannot assume stationarity here, as in the previous case, since  $\mu(t)$  depends on time for sufficiently large  $g$ . This opens up the possibility of having complex dynamical regimes when  $g$  is large.

### Numerical experiments

We have considered the case  $\bar{J}_{11} = \bar{J}_{22} = 5, \tau = 0.1$  giving a Hopf bifurcation for  $g_c = 2$  when  $J = 0$  (Figure 4). The trajectory of  $\mu_1(t)$  and  $v_1(t)$  is represented in Figure 4 in the case  $g = 3$ . When  $J = 0$ ,  $\mu_1(t)$  presents regular oscillations (with non-linear effects since  $g = 3$  is larger than the critical value for the Hopf bifurcation,  $g_c = 2$ ). In this case, the solution  $v_1(t) = 0$  is stable as seen on the figure. When  $J \neq 0$  the Gaussian field has (small) fluctuations which nevertheless strongly interact with the dynamics of  $\mu_1(t)$ , leading to a regime where  $\mu_1(t)$  and  $v_1(t)$  oscillate periodically.

## DISCUSSION

The problem of bridging scales is overwhelming in general when studying complex systems and in particular in neuroscience. After many others we looked at this difficult problem from the theoretical and numerical viewpoints, hoping to get closer to its solution from relatively simple and physically/biologically plausible first principles and assumptions. One of our motivations is to better understand such phenomenological neural mass models as that of Jansen and Rit (1995).



**FIGURE 4 | Evolution of the mean  $\mu_1(t)$  and variance  $v_1(t)$  for the mean-field of population 1, for  $J = 0$  and 2, over a time window  $[0, 20]$ .  $n$  is the number of iterations of  $\mathcal{F}_1$  defined in Section “Existence and Uniqueness of Solutions in Finite Time”. This corresponds to a number of iterations for which the method has essentially converged (up to some precision). Note that  $v_1(t)$  has been magnified by a factor of 100. Though Gaussian fluctuations are small, they have a strong influence on  $\mu_1(t)$ .**

We consider several populations of neurons and start from a microscopic, i.e., individual, description of the dynamics of the membrane potential of each neuron that contains four terms.

The first one controls the intrinsic dynamics of the neuron. It is linear in this article but this assumption is not essential and could probably be safely removed if needed.

The second term is a stochastic input current, correlated or uncorrelated. The corresponding noise model is very rich, depending on the degree  $k$  of smoothness of the  $g$ -shapes. It features integrated Brownian noise up to order  $k - 1$ .

The third term is a deterministic input current, and the fourth one describes the interaction between the neurons through random connectivity coefficients that weigh the contributions of other neurons through a set of functions that are applied to their membranes potentials. The only hypothesis on these functions is that they are smooth and bounded, as well as their first-order derivative. The obvious choice of sigmoids is motivated by standard rate models ideas. Another appealing choice is a smooth approximation to a Dirac delta function thereby opening a window on the world of spiking neurons. Thus, the model presented in this paper is *more general* than the instantaneous rate model that is underlying Ermentrout’s voltage-based model (Ermentrout, 1998) even though we have not explored this avenue.

We then derive the mean-field equations and provide a constructive and new proof, under some mild assumptions, of the existence and uniqueness of a solution of these equations over finite and infinite time intervals. The key idea is to look at this mean-field description as a *global problem* on the probability distribution of the membranes potentials, unlike previous studies. Our proof provides an efficient way of computing this solution and our numerical experiments show a good agreement with previous studies. It is interesting to note that a sufficient condition for the convergence of our algorithm is related to the previously mentioned noise model. We prove that if the noise matrix  $F$  is full rank, with bounded eigenvalues, then the algorithm is in general convergent. An important fact to note is that the solutions of the mean-field equations that we construct are fundamentally *non-Markovian* eliminating the need for such approximations as the introduction of the  $q$  parameter summarizing the whole history of the non-Markovian process, see below.

In the case where the nonlinearities are chosen to be sigmoidal our results shed a new light on existing neural mass models. Indeed, as shown in Section “General Derivation of the Mean-Field Equation”, these appear as approximations of the mean-field equations where the intricate but fundamental coupling between the time variations of the mean membrane potentials and their fluctuations, as represented by the covariance functions, is neglected.

An alternative approach has been recently proposed by Chizhov and collaborators<sup>8</sup> (Chizhov and Graham, 2007; Chizhov et al., 2007). The approach of these authors consists in reducing the large number,  $N$ , of degrees of freedom of the neural assembly by constructing a probability density  $\rho$  on the phase space of neurons states in the limit  $N \rightarrow \infty$ . This is a non-rigorous approach where the evolution equations for  $\rho$  are heuristically derived. Especially, it is assumed that  $\rho$  depends on two parameters only: the current time  $t$  and the time elapsed since the last spike  $t^*$ . Under these

<sup>8</sup>We thank one of the referees for pointing out these references to us.



assumptions the initial phase space of neurons states is mapped to a two dimensional space  $t, t^*$ , while  $\rho(t, t^*)dt$  characterizes the fraction of neurons which have fired in the time interval  $[t - t^*, t - t^* + dt]$ . Therefore, this approach intrinsically holds for integrate and fire neurons models where the neuron's membrane potential history is summarized by the last spike time, when it is reset to a constant value. As noticed by these authors, this allows to circumvent the main problem in mean-field approaches for firing rate models, that we also discuss in the present paper: When using mean-field theory to characterize stationary regimes, one needs to introduce ad hoc parameters (see e.g., the parameter  $q$  introduced in Section "Stationary Solutions") summarizing the whole history of the non-Markovian process. Introducing a "history cut-off" while resetting the membrane potential to a constant value indeed removes this difficulty. Therefore, it might be interesting to compare our approach in the case of integrate-and-fire models (see above remark on the choice of the nonlinearity), to the approach of Chizhov and collaborators. This could provide some rigorous basis for their analysis and allow to elucidate the role of field fluctuations which does not appear explicitly in the probability density approach.

## CONCLUSION AND FURTHER WORK

On more general grounds, our goal is now to extend the present work in several directions.

### BIFURCATIONS ANALYSIS OF THE DYNAMIC MEAN-FIELD EQUATIONS

From the present analysis, and as shown in the simple examples of Section "Numerical Experiments", the mesoscopic dynamics of the average membrane potential of a neurons population can be really different from the classical phenomenological equations la Jansen–Rit if one includes the non-Markovian fluctuations of the interaction fields, which summarize the cumulative effects of the nonlinear interactions of a given neuron with the bulk of other neurons. Jansen–Rit equations are commonly used in the neuroscience community either to anticipate the dynamics of local field potential in relation with imaging (Optical Imaging, MEG/EEG), or to understand neurophysiological disorders such as epilepsy. Bifurcations analysis of these equations reveal dynamical regimes that can be related to experiments (Grimbert and Faugeras, 2006). They can be generalized using more accurate neural models (Wendling et al., 2005). Is there any need to generalize these equations, that we claim to be incomplete, while people commonly use them with some satisfaction? Are there new phenomena, *experimentally accessible*, that can be exhibited by the generalized mean-field equations and that do not appear in the naive ones? These are obviously important questions that we intend to address in the near future. On mathematical grounds, the goal is to make a bifurcation analysis of the map  $\mathcal{F}$  on the space of trajectories, introduced in the present paper. Do any new salient dynamical regimes appear? If such regimes exist, the goal will be, on experimental grounds, to interact with experimentalists in order to see in which conditions such a regime can be exhibited, and what are its implications on cortical columns dynamics or function.

### INVESTIGATIONS OF NONSTATIONARY REGIMES

As discussed in this paper, and as is well-known in the physicists' community (especially spin-glasses community), the dynamic

mean-field approach raises serious difficulties as far as one is trying to describe *stationary dynamics*. On technical grounds, this relies on the non-commutativity of the two limits  $N \rightarrow \infty$  and  $t \rightarrow \infty$  already discussed in Sompolinsky and Zippelius (1982). As a result, one is led to introduce *ad hoc* phenomenological parameters, depending on initial conditions, that can be determined in statistical physics models where the distribution of equilibria is known (Gibbs distribution), using sophisticated techniques such as the replica "trick" (Houghton et al., 1983). For spin-glasses it is only in the high temperature regime that a simple solution to this problem is known. This restriction also appears in the present paper, where the existence and uniqueness of a stationary solution is proved only for low values of the gain parameter  $g$  (which plays a role similar to the inverse temperature). However, we are not so much interested in stationary dynamics, since brain processes are ultimately nonstationary. Our approach, valid for any finite time  $T$ , opens up the possibility to characterize mean-field equations in transient regimes, with an analysis strategy that can moreover be easily implemented. To the best of our knowledge, this type of techniques has never been used in the statistical physics community, where iterations on space trajectories are not in the standard toolbox. Therefore, our work could allow the (numerical) investigation of cortical columns submitted to nonstationary inputs, with strong implications on neuroscience.

### EXTENSION TO A LARGER CLASS OF MODELS

A very challenging question is the application of this theory to spiking neurons models. We have briefly mentioned in Section "Discussion" that this may be possible through the use of non-sigmoidal functions in the interaction terms. This idea could be applied to the analysis of Integrate and Fire models with conductance based synapses, which constitute good models of spiking neurons. As discussed at the end of Section "Discussion", the analysis of the mean-field equations could be simplified by the fact that memory is reset after a neuron fires. There is however a need to characterize parameter space regions where neurons can take an arbitrary large time to fire for the first time (Cessac, 2008; Cessac and Viéville, 2008). This is the main obstruction in the application of our theory to this type of models.

## APPENDIX

### A. IDENTIFICATION OF THE MEAN-FIELD EQUATIONS

Ben-Arous and Guionnet studied from a mathematical point of view the problem of finding a mean-field description of large networks of spin-glasses. They obtained using different methods of stochastic analysis a weak limit of the law of a given spin and proved their independence.

Our equations do not directly fit in their study: indeed, the spin intrinsic dynamics is nonlinear while the interaction is linear, and everything is done in dimension one. Nevertheless, their proof extends to our case which is somehow more simple. For instance in the case of the Simple Model with one population, we can readily adapt their proof in our case. More precisely, let  $P = 1$ , the equation of the network reads:

$$\tau dV_t^j = \left( -V_t^j + \sum_{i=1}^N J_{ij} S(V_t^i) \right) dt + \sigma dW_t^j$$

In this case, we define for  $X \in \mathcal{M}_1^+(C([t_0, T], \mathbb{R}))$  the effective interaction term ( $U_t^X$ ) which is the effective interaction process defined in Section “The Mean-Field Equations”, i.e., the Gaussian process of mean  $\bar{J}_{\text{eff}} \mathbb{E}[S(X_t)]$  and of covariance:  $\text{Cov}(U_t^X, U_s^X) =: \sigma_{\text{eff}}^2 \mathbb{E}[S(X_t)S(X_s)]$ .

Let us note  $\mathcal{P}$  the law of the membrane potential when there is no interaction (it is an Ornstein–Uhlenbeck process), and the empirical measure  $\hat{V}^N = \frac{1}{N} \sum_{i=1}^N \delta_{V_i}$ . We can prove that under the probability distribution averaged over the connectivities, see below, the empirical measure satisfies a large deviation principle with good rate function  $H$  defined as in Guionnet (1997). Using this large deviation result, we can prove annealed and quenched tightness of the empirical measure, and finally its convergence toward the unique process where the good rate function  $H$  achieves its unique minimum, which is defined by the property of having a density with respect to  $\mathcal{P}$  and whose density satisfies the implicit equation:

$$Q \ll \mathcal{P} \quad \frac{dQ}{d\mathcal{P}} = \varepsilon \left[ \exp \left\{ \int_0^T U_t^Q dW_t - \frac{1}{2} \int_0^T (U_t^Q)^2 dt \right\} \right] \quad (59)$$

where  $\varepsilon$  denotes the expectation over the effective interaction process  $U^Q$ .

We can also prove following the steps of Ben-Arous and Guionnet (1997) that there exists a unique solution to this equation, and that this solution satisfies the nonlinear non-Markovian stochastic differential equation:

$$\begin{cases} \tau dV_t = -V_t dt + dB_t \\ dB_t = dW_t + \int_0^t dB_s \varepsilon \left[ U_s^Q U_t^Q \frac{\exp\left\{-\frac{1}{2} \int_0^t (U_u^Q)^2 du\right\}}{\varepsilon \left[ \exp\left\{-\frac{1}{2} \int_0^t (U_u^Q)^2 du\right\} \right]} \right] \\ \text{Law of } (V) = Q, \quad \text{law of } (V_0) = Z_0 \end{cases} \quad (60)$$

which can also be written as our mean-field equation, averaged on the connectivities (see Ben-Arous and Guionnet, 1995). More precisely, let  $V^L$  be the law of the solution of the equation:

$$\begin{cases} \tau dV_t = -V_t dt + dW_t + U_t^V dt \\ \text{Law of } V_0 = Z_0 \end{cases},$$

which is exactly Eq. 33. They prove that  $V$  satisfies the nonlinear equation:

$$V \stackrel{\varepsilon}{=} \varepsilon(V^L)$$

This result is probably extendable to the multi-population case using the multidimensional Girsanov’s theorem, but the corresponding mathematical developments are out of the scope of this paper.

### B. THE RESOLVENT

In this appendix we introduce and give some useful properties of the resolvent  $\Phi_L$  of a homogeneous differential equation:

$$\frac{dx}{dt} = \mathbf{L}(t)x(t) \quad x(t_0) = x_0 \in \mathbb{R}^p, \quad (61)$$

where  $\mathbf{L}: [t_0, T] \rightarrow \mathcal{M}_{p \times p}$  (or  $(-\infty, T] \rightarrow \mathcal{M}_{p \times p}$ ) is  $C^0$ .

**Definition B.1.** The resolvent of Eq. 61 is defined as the unique solution of the linear equation:

$$\begin{cases} \frac{d\Phi_L(t, t_0)}{dt} = \mathbf{L}(t)\Phi_L(t, t_0) \\ \Phi_L(t_0, t_0) = \text{Id}_p \end{cases} \quad (62)$$

where  $\text{Id}_p$  is the  $P \times P$  identity matrix.

**Proposition B.1.** The resolvent satisfies the following properties:

- (i)  $\Phi_L(t+s, t_0) = \Phi_L(t+s, t) \cdot \Phi_L(t, t_0)$
- (ii)  $\Phi_L(t, t_0)$  is invertible of inverse  $\Phi_L(t_0, t)$  which satisfies:

$$\begin{cases} \frac{d\Phi_L(t_0, t)}{dt} = -\Phi_L(t_0, t)\mathbf{L}(t) \\ \Phi_L(t_0, t_0) = \text{Id}_{p \times p} \end{cases} \quad (63)$$

- (iii) Let  $\| \cdot \|$  be a norm on  $\mathcal{M}_{p \times p}$  and assume that  $\| \mathbf{L}(t) \| \leq k_L$  on  $[t_0, T]$ . Then we have:

$$\| \Phi_L(t, t_0) \| \leq e^{k_L |t-t_0|} \quad \forall t \in [t_0, T] \quad (64)$$

Similarly, if  $\| \mathbf{L}^T(t) \| \leq k_{L^T}$  on  $[t_0, T]$  we have:

$$\| \Phi_L^T(t, t_0) \| \leq e^{k_{L^T} |t-t_0|} \quad \forall t \in [t_0, T] \quad (65)$$

- (iv) We have

$$\det \Phi_L(t, t_0) = \exp \int_{t_0}^t \text{Tr} \mathbf{L}(s) ds$$

*Proof.* The properties (i) and (ii) are directly linked with the property of group of the flow of a reversible ODE. (iii) is an application of Grunwald’s lemma. (iv) is obtained by a first-order Taylor series expansion.  $\square$

**Theorem B.2 (Solution of an inhomogeneous linear SDE).** The solution of the inhomogeneous linear stochastic differential equation:

$$\begin{cases} dX_t = (\mathbf{L}(t)X(t) + \mathbf{I}(t))dt + \mathbf{F}(s)d\mathbf{W}_s \\ X_{t_0} = X_0 \end{cases} \quad (66)$$

can be written using the resolvent:

$$X_t = \Phi_L(t, t_0)X_0 + \int_{t_0}^t \Phi_L(t, s)\mathbf{I}(s)ds + \int_{t_0}^t \Phi_L(s, t)\mathbf{F}(s)d\mathbf{W}_s \quad (67)$$

*Proof.* Pathwise (strong) uniqueness of solution directly comes from the results on the SDE with Lipschitz coefficients (see e.g., Karatzas and Shreve, 1991, Theorem 2.5 of Chapter 5). It is clear that  $X_{t_0} = X_0$ . We use Itô’s formula for the product of two stochastic processes to prove that the process Eq. 67 is solution of Eq. 66:

$$\begin{aligned} dX_t &= \left( \mathbf{L}(t)\Phi_L(t, t_0)X_0 + \Phi_L(t, t)\mathbf{I}(t) + \int_{t_0}^t \mathbf{L}(t)\Phi_L(t, s)\mathbf{I}(s)ds \right) dt \\ &\quad + \Phi_L(t, t)\mathbf{F}(t)d\mathbf{W}_t + \int_{t_0}^t \mathbf{L}(t)\Phi_L(s, t)\mathbf{F}(s)d\mathbf{W}_s dt \\ &= \left( \mathbf{L}(t) \left[ \Phi_L(t, t_0)X_0 + \int_{t_0}^t \Phi_L(s, t)\mathbf{I}(s)ds \right] \right) dt \end{aligned}$$

$$\begin{aligned}
 & + \int_{t_0}^t \Phi_L(s, t) F(s) dW_s \Big] + \mathbf{I}(t) \Big) dt + F(t) dW_t \\
 & = (\mathbf{L}(t)X(t) + \mathbf{I}(t))dt + F(t)dW_t
 \end{aligned}$$

Hence the theorem is proved.  $\square$

**C. MATRIX NORMS**

In this section we recall some definitions on matrix and vector norms. Let  $\mathcal{M}_{n \times n}$  be the set of  $n \times n$  real matrices. It is a vector space of dimension  $n^2$  and the usual  $L^p$  norms  $1 \leq p \leq \infty$  can be defined. Given  $\mathbf{L} \in \mathcal{M}_{n \times n}$ , we note  $\|\mathbf{L}\|_p$  the corresponding norm. Given a vector norm, noted  $\|\cdot\|$ , on  $\mathbb{R}^n$  the induced norm, noted  $\|\cdot\|$ , on  $\mathcal{M}_{n \times n}$  is defined as

$$\|\mathbf{L}\| = \sup_{x \in \mathbb{R}^n, \|x\| \leq 1} \frac{\|\mathbf{L}x\|}{\|x\|}$$

Since  $\mathcal{M}_{n \times n}$  is finite dimensional all norms are equivalent. In this article we use the following norms

- (i)  $\|\mathbf{L}\|_\infty = \max_i \sum_{j=1}^n |L_{ij}|$ .
- (ii)  $\|\mathbf{L}\|_\infty^v = \max_{i,j} |L_{ij}|$
- (iii)  $\|\mathbf{L}\|_2 = \sup_{x \in \mathbb{R}^n, \|x\|_2 \leq 1} \frac{\|\mathbf{L}x\|_2}{\|x\|_2}$ . This so-called *spectral* norm is equal to the square root of the largest singular value of  $\mathbf{L}$  which is the largest eigenvalue of the positive matrix  $\mathbf{L}^T \mathbf{L}$ . If  $\mathbf{L}$  is positive definite this is its largest eigenvalue which is also called its spectral radius, noted  $\rho(\mathbf{L})$ .

**D. IMPORTANT CONSTANTS**

**Table 1** summarizes some notations which are introduced in the article and used in several places.

**E. PROOF OF LEMMA 2**

**Lemma E.1.** The following upperbounds are valid for all  $n \geq 1$  and all  $s, t \in [t_0, T]$ .

$$\|\mu^n(t)\|_\infty \leq e^{k_L(T-t_0)} \left[ \|\mathbb{E}[Z_0]\|_\infty + (\mu + I_{\max})(T-t_0) \right] \stackrel{\text{def}}{=} \mu_{\max},$$

$$\begin{aligned}
 \|C^n(t, s)\|_\infty & \leq e^{(k_L+k_T)(T-t_0)} \left[ \rho(\Sigma^{Z_0}) + \lambda_{\max}^\Gamma (T-t_0) \right. \\
 & \left. + \sigma_{\max}^2 (T-t_0)^2 \right] \stackrel{\text{def}}{=} \Sigma_{\max},
 \end{aligned}$$

**Table 1 | Some important quantities defined in the article.**

Constant	Expression	Defined in
$\mu$	$\max_\alpha \sum_\beta  \bar{J}_{\alpha\beta}  \ S_\beta\ _\infty$	Lemma 1, Eq. 39
$\sigma_{\max}^2$	$\max_\alpha \sum_\beta \sigma_{\alpha\beta}^2 \ S_\beta\ _\infty^2$	Lemma 1
$\sigma_{\min}$	$\min_{\alpha,\beta} \sigma_{\alpha\beta}^2$	Lemma 1
$\mu_{\max}$	$e^{k_L(T-t_0)} \left[ \ \mathbb{E}[Z_0]\ _\infty + (\mu + I_{\max})(T-t_0) \right]$	Lemma 2
$\Sigma_{\max}$	$e^{(k_L+k_T)(T-t_0)} \left[ \rho(\Sigma^{Z_0}) + \lambda_{\max}^\Gamma (T-t_0) + \sigma_{\max}^2 (T-t_0)^2 \right]$	Lemma 2
$k_0$	$\lambda_{\min} \lambda_{\min}^{\Sigma^{Z_0}}$	Lemma 3
$K$	$\lambda_{\min} \sqrt{\lambda_{\min}^{\Sigma^{Z_0}} \lambda_{\min}^\Gamma (T-t_0)}$	Proof of Lemma 5
$k_C$	$\max_\alpha \sum_\beta \sigma_{\alpha\beta}^2 \ S_\beta\ _\infty \ S_\beta\ $	Proposition 3, Eq. 45
$\lambda_L$		Eq. 48

where  $\mu$  and  $\sigma_{\max}$  are defined in Lemma 1,  $\lambda_{\max}^\Gamma$  is defined in Assumption 1.

*Proof.* The first inequality follows from taking the infinite norm of both sides of Eq. 41 and using Assumption (a) in 1 and Eq. 64, Lemma 1, and Assumption (c) in 1.

The second inequality follows from taking the infinite norm of both sides of Eq. 42 and using Assumption (a) in 1 and Eqs. 64 and 65, Lemma 1, and Assumption (b) in 1.  $\square$

**F. PROOF OF LEMMA 3**

**Lemma F.1.** For all  $t \in [t_0, T]$  all  $\alpha = 1, \dots, kP$ , and  $n \geq 1$ , we have

$$C_{\alpha\alpha}^n(t, t) \geq \lambda_{\min} \lambda_{\min}^{\Sigma^{Z_0}} \stackrel{\text{def}}{=} k_0 > 0,$$

where  $\lambda_{\min}$  is the smallest singular value of the symmetric positive definite matrix  $\Phi_L(t, t_0) \Phi_L(t, t_0)^T$  for  $t \in [t_0, T]$  and  $\lambda_{\min}^{\Sigma^{Z_0}}$  is the smallest eigenvalue of the symmetric positive definite covariance matrix  $\Sigma^{Z_0}$ .

*Proof.*  $C_{\alpha\alpha}^n(t, t)$  is larger than  $(\Phi_L(t, t_0) \Sigma^{Z_0} \Phi_L(t, t_0)^T)_{\alpha\alpha}$  which is larger than the smallest eigenvalue of the matrix  $\Phi_L(t, t_0) \Sigma^{Z_0} \Phi_L(t, t_0)^T$ . This smallest eigenvalue is equal to

$$\begin{aligned}
 & \min_{\|x\| \leq 1} \frac{x^T \Phi_L(t, t_0) \Sigma^{Z_0} \Phi_L(t, t_0)^T x}{x^T x} \\
 & = \min_{\|x\| \leq 1} \left( \frac{x^T \Phi_L(t, t_0) \Sigma^{Z_0} \Phi_L(t, t_0)^T x}{x^T \Phi_L(t, t_0) \Phi_L(t, t_0)^T x} \frac{x^T \Phi_L(t, t_0) \Phi_L(t, t_0)^T x}{x^T x} \right) \\
 & \geq \min_{\|x\| \leq 1} \frac{x^T \Phi_L(t, t_0) \Sigma^{Z_0} \Phi_L(t, t_0)^T x}{x^T \Phi_L(t, t_0) \Phi_L(t, t_0)^T x} \min_{\|x\| \leq 1} \frac{x^T \Phi_L(t, t_0) \Phi_L(t, t_0)^T x}{x^T x}.
 \end{aligned}$$

In the last expression the first term is larger than the smallest eigenvalue  $\lambda_{\min}^{\Sigma^{Z_0}}$  of the matrix  $\Sigma^{Z_0}$  which is positive definite since we have assumed the Gaussian random variable  $Z_0$  nondegenerate. The second term is equal to the smallest singular value  $\lambda_{\min}$  of the matrix  $\Phi_L(t, t_0)$  which is also strictly positive since  $\Phi_L(t, t_0)$  is invertible for all  $t \in [t_0, T]$ , see Appendix B.  $\square$

**G. PROOF OF LEMMA 4**

**Lemma G.1.** For all  $\alpha = 1, \dots, kP$  and  $n \geq 1$  the quantity  $C_{\alpha\alpha}^n(s, s)C_{\alpha\alpha}^n(t, t) - C_{\alpha\alpha}^n(t, s)^2$  is lowerbounded by the positive symmetric function:

$$\theta(s, t) \stackrel{\text{def}}{=} |t-s| \lambda_{\min}^2 \lambda_{\min}^{\Sigma^{Z_0}} \lambda_{\min}^\Gamma,$$

where  $\lambda_{\min}^\Gamma$  is the strictly positive lower bound, introduced in 3.1, on the singular values of the matrix  $F(u)$  for  $u \in [t_0, T]$ .

*Proof.* We use Eq. 42 which we rewrite as follows, using the group property of the resolvent  $\Phi_L$ :

$$\begin{aligned}
 C^{n+1}(t, s) & = \Phi_L(t, t_0) \left( \Sigma^{Z_0} + \int_{t_0}^{t \wedge s} \Phi_L(t_0, u) F(u) F(u)^T \Phi_L(t_0, u)^T du \right. \\
 & \left. + \int_{t_0}^t \int_{t_0}^s \Phi_L(t_0, u) \text{Cov}(\tilde{\mathbf{U}}_u^{X_n}, \tilde{\mathbf{U}}_v^{X_n}) \Phi_L(t_0, v)^T dudv \right) \Phi_L(s, t_0)^T.
 \end{aligned}$$

We now assume  $s < t$  and introduce the following notations:

$$A(s) = \sum_{t_0}^{Z_0} + \int_{t_0}^s \Phi_L(t_0, u) F(u) F(u)^T \Phi_L(t_0, u)^T du$$

$$B(s, t) = \int_{s_0}^t \Phi_L(t_0, u) F(u) F(u)^T \Phi_L(t_0, u)^T du$$

$$a(t, s) = \int_{t_0}^t \int_{t_0}^s \Phi_L(t_0, u) \text{Cov}(\tilde{U}_u^{X_n}, \tilde{U}_v^{X_n}) \Phi_L(t_0, v)^T du dv$$

Let  $e_\alpha$ ,  $\alpha = 1, \dots, kP$ , be the unit vector of the canonical basis whose coordinates are all equal to 0 except the  $\alpha$ th one which is equal to 1. We note  $E_\alpha(t)$  the vector  $\Phi_L(t, t_0)^T e_\alpha$ . We have, dropping the index  $n$  for simplicity:

$$C_{\alpha\alpha}(t, s) = E_\alpha(t)^T (A(s) + a(t, s)) E_\alpha(s)$$

$$C_{\alpha\alpha}(s, s) = E_\alpha(s)^T (A(s) + a(s, s)) E_\alpha(s)$$

$$C_{\alpha\alpha}(t, t) = E_\alpha(t)^T (A(s) + B(s, t) + a(t, t)) E_\alpha(t).$$

Note that the last expression does not depend on  $s$ , since  $A(s) + B(s, t) = A(t)$ , which is consistent with the first equality. The reason why we introduce  $s$  in this expression is to simplify the following calculations.

The expression  $C_{\alpha\alpha}(s, s)C_{\alpha\alpha}(t, t) - C_{\alpha\alpha}(t, s)^2$  is the sum of four sub-expressions:

$$\epsilon_1(s, t) = (E_\alpha(s)^T A(s) E_\alpha(s)) (E_\alpha(t)^T A(s) E_\alpha(t)) - (E_\alpha(t)^T A(s) E_\alpha(s))^2,$$

which is  $\geq 0$  because  $A(s)$  is a covariance matrix,

$$\epsilon_2(s, t) = (E_\alpha(s)^T a(s, s) E_\alpha(s)) (E_\alpha(t)^T a(t, t) E_\alpha(t)) - (E_\alpha(t)^T a(t, s) E_\alpha(s))^2,$$

which is also  $\geq 0$  because  $a(t, s)$  is a covariance matrix function,

$$\begin{aligned} \epsilon_3(s, t) &= (E_\alpha(s)^T A(s) E_\alpha(s)) (E_\alpha(t)^T a(t, t) E_\alpha(t)) \\ &+ (E_\alpha(t)^T A(s) E_\alpha(t)) (E_\alpha(s)^T a(s, s) E_\alpha(s)) \\ &- 2(E_\alpha(t)^T A(s) E_\alpha(s)) (E_\alpha(t)^T a(t, s) E_\alpha(s)) \end{aligned}$$

Because  $a(t, s)$  is a covariance matrix function, we have

$$E_\alpha(t)^T a(t, t) E_\alpha(t) + E_\alpha(s)^T a(s, s) E_\alpha(s) - 2E_\alpha(t)^T a(t, s) E_\alpha(s) \geq 0,$$

and, as seen above,  $\epsilon_2(s, t) \geq 0$ . Because  $\epsilon_1(s, t) \geq 0$ , we also have

$$\begin{aligned} -\sqrt{E_\alpha(s)^T A(s) E_\alpha(s)} \sqrt{E_\alpha(t)^T A(s) E_\alpha(t)} &\leq E_\alpha(t)^T A(s) E_\alpha(s) \\ &\leq \sqrt{E_\alpha(s)^T A(s) E_\alpha(s)} \sqrt{E_\alpha(t)^T A(s) E_\alpha(t)}, \end{aligned}$$

and, as it can be readily verified, this implies  $\epsilon_3(s, t) \geq 0$ .

Therefore, we can lowerbound  $C_{\alpha\alpha}(s, s)C_{\alpha\alpha}(t, t) - C_{\alpha\alpha}(t, s)^2$  by the fourth subexpression:

$$\begin{aligned} C_{\alpha\alpha}(s, s)C_{\alpha\alpha}(t, t) - C_{\alpha\alpha}(t, s)^2 &\geq (E_\alpha(s)^T A(s) E_\alpha(s)) (E_\alpha(t)^T B(s, t) E_\alpha(t)) \\ &+ (E_\alpha(s)^T a(s, s) E_\alpha(s)) (E_\alpha(t)^T B(s, t) E_\alpha(t)) \\ &\geq (E_\alpha(s)^T A(s) E_\alpha(s)) (E_\alpha(t)^T B(s, t) E_\alpha(t)), \end{aligned}$$

since  $B(s, t)$  and  $a(s, s)$  are covariance matrices. We next have

$$E_\alpha(s)^T A(s) E_\alpha(s) = \frac{E_\alpha(s)^T A(s) E_\alpha(s) e_\alpha^T \Phi_L(s, t_0) \Phi_L(s, t_0)^T e_\alpha}{E_\alpha(s)^T E_\alpha(s) e_\alpha^T e_\alpha},$$

by definition of  $E_\alpha(s)$ . Therefore

$$E_\alpha(s)^T A(s) E_\alpha(s) \geq \lambda_{\min}^{A(s)} \lambda_{\min}^{\Phi_L(s, t_0) \Phi_L(s, t_0)^T} \geq \lambda_{\min}^{\Sigma^{z_0}} \lambda_{\min},$$

where  $\lambda_{\min}^C$  is the smallest eigenvalue of the symmetric positive matrix  $C$ . Similarly we have

$$E_\alpha(t)^T B(s, t) E_\alpha(t) \geq \lambda_{\min}^{B(s, t)} \lambda_{\min}.$$

Let us write  $\Gamma(u) = F(u)F(u)^T$ . We have (Assumption 1):

$$\begin{aligned} \lambda_{\min}^{B(s, t)} &= \min_{\|x\| \leq 1} \int_s^t \frac{x^T \Phi_L(t_0, u) \Gamma(u) \Phi_L(t_0, u)^T x}{x^T x} du \\ &= \min_{\|x\| \leq 1} \int_s^t \frac{x^T \Phi_L(t_0, u) \Gamma(u) \Phi_L(t_0, u)^T x}{x^T \Phi_L(t_0, u) \Phi_L(t_0, u)^T x} \frac{x^T \Phi_L(t_0, u) \Phi_L(t_0, u)^T x}{x^T x} du \\ &\geq \int_s^t \min_{\|x\| \leq 1} \left( \frac{x^T \Phi_L(t_0, u) \Gamma(u) \Phi_L(t_0, u)^T x}{x^T \Phi_L(t_0, u) \Phi_L(t_0, u)^T x} \frac{x^T \Phi_L(t_0, u) \Phi_L(t_0, u)^T x}{x^T x} \right) du \\ &\geq (t - s) \lambda_{\min} \lambda_{\min}^\Gamma. \end{aligned}$$

Combining these results, we have

$$C_{\alpha\alpha}(s, s)C_{\alpha\alpha}(t, t) - C_{\alpha\alpha}(t, s)^2 \geq |t - s| \lambda_{\min}^2 \lambda_{\min}^{\Sigma^{z_0}} \lambda_{\min}^\Gamma \quad \square$$

### H. PROOF OF LEMMA 5

**Lemma H.1.** The  $2n$ -dimensional integral

$$\begin{aligned} I_n &= \int_{[t_0, t \vee s]^2} \rho_1(u_1, v_1) \left( \int_{[t_0, u_1 \vee v_1]^2} \dots \left( \int_{[t_0, u_{n-2} \vee v_{n-2}]^2} \rho_{n-1}(u_{n-1}, v_{n-1}) \right. \right. \\ &\quad \left. \left. \left( \int_{[t_0, u_{n-1} \vee v_{n-1}]^2} \rho_n(u_n, v_n) du_n dv_n \right) du_{n-1} dv_{n-1} \right) \dots \right) du_1 dv_1, \end{aligned}$$

where the functions  $\rho_i(u_i, v_i)$ ,  $i = 1, \dots, n$  are either equal to 1 or to  $1/\sqrt{\theta(u_i, v_i)}$  (the function  $\theta$  is defined in Lemma 4), is upper-bounded by  $k^n/(n-1)!$  for some positive constant  $k$ .

*Proof.* First note that the integral is well-defined because of Lemma 4. Second, note that there exists a constant  $K$  such that

$$\frac{K}{\sqrt{\theta(u, v)}} \geq 1 \text{ for all } (u, v) \in [t_0, t \vee s]^2,$$

i.e.,  $K = \lambda_{\min} \sqrt{\lambda_{\min}^{\Sigma^{z_0}} \lambda_{\min}^\Gamma (T - t_0)}$ . Therefore the integral is upper-bounded by  $K_0^n$ , where  $K_0 = \max(1, K)$  times the integral obtained when  $\rho_i(u_i, v_i) = 1/\sqrt{|u_i - v_i|}$  for all  $i = 1, \dots, n$ . Let us then consider this situation. Without loss of generality we assume  $t_0 = 0$ . The cases  $n = 1, 2, 3$  allow one to understand the process.

$$I_1 \leq K_0 \int_{[0, t \vee s]^2} \frac{du dv}{\sqrt{|u - v|}}.$$

Let us rotate the axes by  $-\frac{\pi}{4}$  by performing the change of variables

$$\begin{aligned} u &= \frac{U + V}{\sqrt{2}}, \\ v &= \frac{V - U}{\sqrt{2}}. \end{aligned}$$

Using the symmetry of the integrand in  $s$  and  $t$  and the change of variable, the integral in the righthand side of Eq. 68 is equal to (see **Figure 5**):



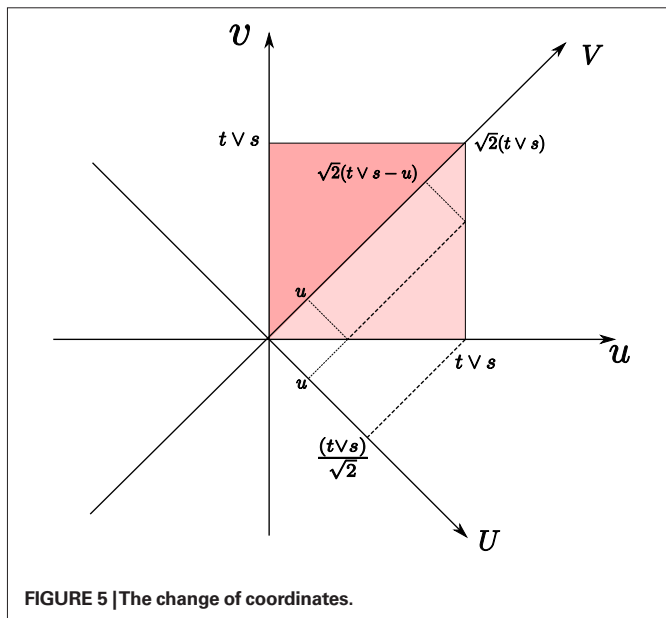


FIGURE 5 | The change of coordinates.

$$2 \frac{1}{2^{1/4}} \int_0^{(tvs)/\sqrt{2}} \int_U^{\sqrt{2}(tvs)-U} \frac{dVdU}{\sqrt{U}} = 2^{3/4} \int_0^{a/2} \frac{a-2U}{\sqrt{U}} dU = 2^{3/4} \alpha_1 a^{3/2},$$

where  $a = \sqrt{2}(tvs)$  and  $\alpha_1 = (2\sqrt{2})/3$ .

Let us now look at  $I_2$ . It is upperbounded by the factor  $K_0^2(2^{3/4})^2\alpha_1$  times the integral

$$\int_0^{a/2} \int_U^{a-U} \frac{(\sqrt{2}(u \vee v))^3}{\sqrt{U}} dUdV.$$

□

## REFERENCES

- Abbott, L. F., and Van Vreeswijk, C. A. (1993). Asynchronous states in networks of pulse-coupled neuron. *Phys. Rev. A* 48, 1483–1490.
- Amari, S.-I. (1972). Characteristics of random nets of analog neuron-like elements. *IEEE Trans. Syst. Man Cybern.* 2, 643–657.
- Amari, S.-I., Yoshida, K., and Kanatani, K.-I. (1977). A mathematical foundation for statistical neurodynamics. *SIAM J. Appl. Math.* 33, 95–126.
- Ben-Arous, G., and Guionnet, A. (1995). Large deviations for Langevin spin glass dynamics. *Probab. Theory Relat. Fields* 102, 455–509.
- Ben-Arous, G., and Guionnet, A. (1997). Symmetric Langevin spin glass dynamics. *Ann. Probab.* 25, 1367–1422.
- Billingsley, P. (1999). *Convergence of Probability Measures* (Wiley Series in Probability and Statistics). New York, NY, John Wiley and Sons, Inc.
- Bogachev, V. I. (1998). *Gaussian Measures*. Providence, RI, American Mathematical Society.
- Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*, 2nd Edn. Berlin, Springer.
- Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* 11, 1621–1671.
- Cessac, B. (1995). Increase in complexity in random neural networks. *J. Phys. I (France)* 5, 409–432.
- Cessac, B. (2008). A discrete time neural network model with spiking neurons: rigorous results on the spontaneous dynamics. *J. Math. Biol.* 56, 311–345.
- Cessac, B., and Viéville, T. (2008). On dynamics of integrate-and-fire neural networks with adaptive conductance based synapses. *Front. Neurosci.* 2, 2.
- Chizhov, A. V., and Graham, L. J. (2007). Population model of hippocampal pyramidal neurons, linking a refractory density approach to conductance-based neurons. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 75, 011924-1–011924-14.
- Chizhov, A. V., Rodrigues, S., and Terry, J. R. (2007). A comparative analysis of a firing-rate model and a conductance-based neural population model. *Phys. Lett. A* 369, 31–36.
- Crisanti, A., Sommers, H. J., and Sompolinsky, H. (1990). Chaos in neural networks: chaotic solutions.
- Crisanti, A., and Sompolinsky, H. (1987a). Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. *Phys. Rev. A* 36, 4922–4939.
- Crisanti, A., and Sompolinsky, H. (1987b). Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics. *Phys. Rev. A* 37, 4865.
- Dayan, P., and Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA, MIT Press.
- Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys.* 61, 353–430.
- Freeman, W. J. (1975). *Mass Action in the Nervous System*. New York, Academic Press.
- Gerstner, W. (1995). Time structure of the activity in neural network models. *Phys. Rev. E* 51, 738–758.
- Gerstner, W., and Kistler, W. M. (2002). Mathematical formulations of Hebbian learning. *Biol. Cybern.* 87, 404–415.
- Grimbert, F., and Faugeras, O. (2006). Bifurcation analysis of Jansen's neural mass model. *Neural Comput.* 18, 3052–3068.
- Guionnet, A. (1997). Averaged and quenched propagation of chaos for spin glass dynamics. *Probab. Theory Relat. Fields* 109, 183–215.
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA* 81, 3088–3092.
- Houghton, A., Jain, S., and Young, A. P. (1983). Role of initial conditions in the mean-field theory of spin-glass dynamics. *Phys. Rev. B Condens. Matter* 28, 2630–2637.
- Jansen, B. H., and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled

Since in the area of integration  $u \vee v = v = \frac{v-U}{\sqrt{2}}$  we are led to the product of  $2/5$  by the one-dimensional integral

$$\int_0^{a/2} \frac{(a-2U)^{5/2}}{\sqrt{U}} dUdV = \alpha_2 a^3,$$

where  $\alpha_2 = \frac{5\sqrt{2}\pi}{32}$ .

Similarly  $I_3$  is upperbounded by the product of  $K_0^3(2^{3/4})^3\alpha_1\alpha_2\frac{2}{5}\frac{2}{8}$  times the integral

$$\int_0^{a/2} \frac{(a-2U)^4}{\sqrt{U}} dUdV = \alpha_3 a^{9/2},$$

where  $\alpha_3 = \frac{128\sqrt{2}}{315}$ . One easily shows then that:

$$I_n \leq K_0^n F(2^{3/4})^n 2^n \left( \prod_{i=1}^n \alpha_i \right) \left( \frac{1}{\prod_{j=1}^n (2+3(j-1))} \right).$$

It can be verified by using a system for symbolic computation that  $0 < \alpha_i < 1$  for all  $i \geq 1$ . One also notices that

$$\prod_{j=1}^n (2+3(j-1)) \geq \frac{3^{n-1}}{2} (n-1)!,$$

therefore

$$I_n \leq K_0^n (2^{3/4})^n 2^n 3^{-(n-1)} \frac{1}{(n-1)!},$$

and this finishes the proof.

## ACKNOWLEDGEMENT

This research was partly supported by funding of the European Union under the grant no. 15879 (FACETS) and the Fondation d'Entreprise EADS. It was also supported by the MACACC ARC INRIA and the Doebelin CNRS Federation.

- cortical columns. *Biol. Cybern.* 73, 357–366.
- Karatzas, I., and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics, Vol. 113, 2nd Edn. New York, Springer-Verlag.
- Kushner, H. J. (1984). *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*. Cambridge, MA, MIT Press.
- Mattia, M., and Del Giudice, P. (2002). Population dynamics of interacting spiking neurons. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 66, 51917.
- Molgedey, L., Schuchardt, J., and Schuster, H.G. (1992). Suppressing chaos in neural networks by noise. *Phys. Rev. Lett.* 69, 3717–3719.
- Moynot, O., and Samuelides, M. (2002). Large deviations and mean-field theory for asymmetric random recurrent neural networks. *Probab. Theory Relat. Fields* 123, 41–75.
- Samuelides, M., and Cessac, B. (2007). Random recurrent neural networks. *Eur. Phys. J. Spec. Top.* 142, 7–88.
- Sompolinsky, H., Crisanti, A., and Sommers, H. J. (1998). Chaos in random neural networks. *Phys. Rev. Lett.* 61, 259–262.
- Sompolinsky, H., and Zippelius, A. (1982). Relaxational dynamics of the Edwards–Anderson model and the mean-field theory of spin-glasses. *Phys. Rev. B* 25, 6860–6875.
- Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Netw. Comput. Neural Syst.* 4, 259–284.
- van Rotterdam, A., Lopes da Silva, F. H., van den Ende, J., Viergever, M. A., and Hermans, A. J. (1982). A model of the spatial–temporal characteristics of the alpha rhythm. *Bull. Math. Biol.* 44, 283–305.
- Wendling, F., Hernandez, A., Bellanger, J.-J., Chauvel, P., and Bartolomei, F. (2005). Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG. *J. Clin. Neurophysiol.* 22, 343–356.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 04 August 2008; paper pending published: 06 October 2008; accepted: 26 January 2009; published online: 18 February 2009.

Citation: Faugeras O, Touboul J and Cessac B (2009) A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. *Front. Comput. Neurosci.* (2009) 3:1. doi: 10.3389/neuro.10.001.2009

Copyright © 2009 Faugeras, Touboul and Cessac. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution and reproduction in any medium, provided the original authors and source are credited.

## Persistent Neural States: Stationary Localized Activity Patterns in Nonlinear Continuous $n$ -Population, $q$ -Dimensional Neural Networks

**Olivier Faugeras**

*Olivier.Faugeras@sophia.inria.fr*

**Romain Veltz**

*romain.veltz@sophia.inria.fr*

**François Grimbert**

*Francois.Grimbert@sophia.inria.fr*

*INRIA/ENS/ENPC, Odyssee Team, Sophia-Antipolis 06902, France*

Neural continuum networks are an important aspect of the modeling of macroscopic parts of the cortex. Two classes of such networks are considered: voltage and activity based. In both cases, our networks contain an arbitrary number,  $n$ , of interacting neuron populations. Spatial nonsymmetric connectivity functions represent cortico-cortical, local connections, and external inputs represent nonlocal connections. Sigmoidal nonlinearities model the relationship between (average) membrane potential and activity. Departing from most of the previous work in this area, we do not assume the nonlinearity to be singular, that is, represented by the discontinuous Heaviside function. Another important difference from previous work is that we relax the assumption that the domain of definition where we study these networks is infinite, that is, equal to  $\mathbb{R}$  or  $\mathbb{R}^2$ . We explicitly consider the biologically more relevant case of a bounded subset  $\Omega$  of  $\mathbb{R}^q$ ,  $q = 1, 2, 3$ , a better model of a piece of cortex. The time behavior of these networks is described by systems of integro-differential equations. Using methods of functional analysis, we study the existence and uniqueness of a stationary (i.e., time-independent) solution of these equations in the case of a stationary input. These solutions can be seen as ‘persistent’; they are also sometimes called *bumps*. We show that under very mild assumptions on the connectivity functions and because we do not use the Heaviside function for the nonlinearities, such solutions always exist. We also give sufficient conditions on the connectivity functions for the solution to be absolutely stable, that is, independent of the initial state of the network. We then study the sensitivity of the solutions to variations of such parameters as the connectivity functions, the sigmoids, the external inputs, and, last but not least, the shape of the domain of existence  $\Omega$  of the neural continuum networks. These theoretical results are illustrated and corroborated by a large number of numerical experiments in most of the cases  $2 \leq n \leq 3$ ,  $2 \leq q \leq 3$ .

## 1 Introduction

---

We analyze the ability of neuronal continuum networks to display localized persistent activity, or *bumps*. This type of activity is related, for example, to working memory, which involves holding and processing information on the timescale of seconds. Experiments in primates have shown that there exist neurons in the prefrontal cortex that have high firing rates during the period the animal is “remembering” the spatial location of an event before using the information being remembered (Colby, Duhamel, & Goldberg, 1995; Funahashi, Bruce, & Goldman-Rakic, 1989; Miller, Erickson, & Desimone, 1996). Realistic models for this type of activity have involved spatially extended networks of coupled neural elements or neural masses and the study of spatially localized areas of high activity in these systems. A neuronal continuum network is first built from a “local” description of the dynamics of a number of interacting neuron populations where the spatial structure of the connections is neglected. This local description can be thought of as representing such a structure as a cortical column (Mountcastle, 1957, 1997; Buxhoeveden & Casanova, 2002). We call it a neural mass (Freeman, 1975). Probably the best-known neural mass model is that of Jansen and Rit (Jansen, Zouridakis, & Brandt, 1993), based on the original work of Lopes da Silva and colleagues (Lopes da Silva, Hoeks, & Zetterberg, 1974; Lopes da Silva, van Rotterdam, Barts, van Heusden, & Burr, 1976) and of Van Rotterdam and colleagues (van Rotterdam, Lopes da Silva, van den Ende, Viergever, & Hermans, 1982). A complete analysis of the bifurcation diagram of this model can be found in Grimberty and Faugeras (2006). The model has been used to simulate evoked potentials: EEG activities in normal (Jansen & Rit, 1995) and epileptic patients (Wendling, Bartolomei, Bellanger, & Chauvel, 2001; Wendling, Bellanger, Bartolomei, & Chauvel, 2000). In a similar vein, David and Friston (2003) have used an extension of this model to simulate a large variety of cerebral rhythms ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\theta$ ) in MEG/EEG simulations. Another important class of such models is the one introduced by Wilson and Cowan (1973); Hoppenstaedt & Izhikevich (1997).

These local descriptions are then assembled spatially to form the neuronal continuum network. This continuum network is meant to represent a macroscopic part of the neocortex, for example, a visual area such as V1. The spatial connections are models of cortico-cortical connections. Other, nonlocal connections with, for example, such visual areas as the lateral geniculate nucleus, or V2, are also considered. Other researchers have used several interconnected neural masses to simulate epileptogenic zones (Wendling et al., 2000, 2001; Lopes da Silva et al., 2003) or to study the connectivity between cortical areas (David, Cosmelli, & Friston, 2004). In this letter we consider a continuum of neural masses.



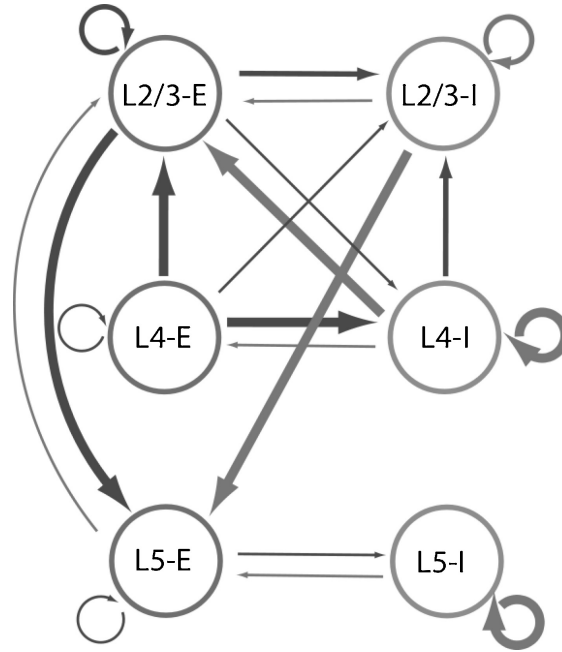


Figure 1: A model with six interacting neural populations.

## 2 The Models

---

We briefly discuss local and spatial models.

**2.1 The Local Models.** We consider  $n$  interacting populations of neurons such as those shown in Figure 1. The figure is inspired by the work of Alex Thomson (Thomson & Bannister, 2003) and Wolfgang Maass (Haeusler & Maass, 2007). It shows six interacting populations of neurons. Black indicates excitation and gray inhibition. The thickness of the arrows pertains to the strength of the interaction. In this example, the six populations are located in layers 2/3, 4, and 5 of the neocortex. Each population being described by its state (defined below), we derive from first principles the equations that describe the variations over time of the states of the different interacting populations.

Our derivation follows closely that of Ermentrout (1998). We consider that each neural population  $i$  is described by its average membrane potential  $V_i(t)$  or its average instantaneous firing rate  $v_i(t)$ , the relation between the two quantities being of the form  $v_i(t) = S_i(V_i(t))$  (Gerstner and Kistler, 2002; Dayan & Abbott, 2001), where  $S_i$  is sigmoidal and smooth. The functions  $S_i$ ,  $i = 1, \dots, n$ , satisfy the following properties introduced in the following definition:

**Definition 1.** For all  $i = 1, \dots, n$ ,  $|S_i| \leq S_{im}$  (boundedness). We note  $S_m = \max_i S_{im}$ . For all  $i = 1, \dots, n$ , the derivative  $S'_i$  of  $S_i$  is positive and bounded by

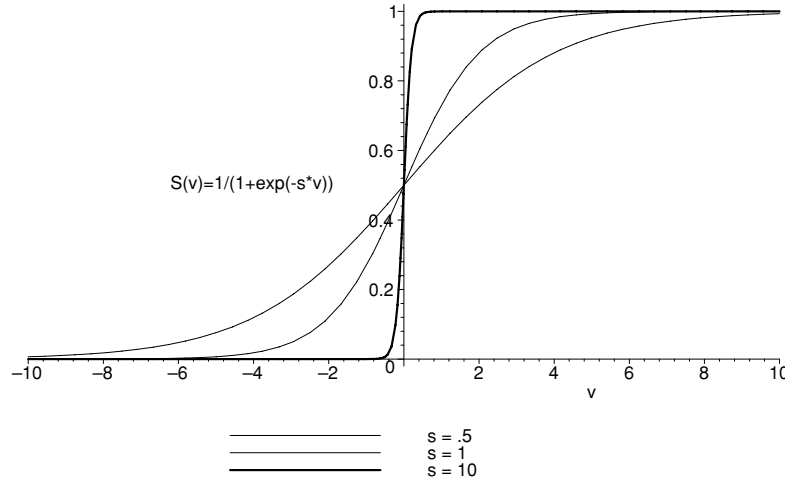


Figure 2: Three examples of sigmoid functions for different values of the parameter  $s$ . See the text.

$S'_{im} > 0$  (boundedness of the derivatives). We note  $DS_m = \max_i S'_{im}$  and  $DS_m$  the diagonal matrix  $\text{diag}(S'_{im})$ .

A typical example of a function  $S_i$  is given in equation 2.1:

$$S_i(v) = \frac{1}{1 + e^{-s_i(v-\theta_i)}}. \quad (2.1)$$

This function is symmetric with respect to the “threshold” potential  $\theta_i$  and varies between 0 and 1. The positive parameter  $s_i$  controls the slope of the  $i$ th sigmoid at  $v = \theta_i$ .

This function is shown in Figure 2 for the values of the parameters  $\theta = 0$  and  $s = 0.5, 1, 10$ . We have  $S_{im} = 1$  and  $S'_{im} = s$ . When  $s \rightarrow \infty$ ,  $S$  converges to the Heaviside function  $Y$  defined by

$$Y(v) = \begin{cases} 0 & \text{if } v < 0 \\ 1 & \text{otherwise} \end{cases}.$$

Neurons in population  $j$  are connected to neurons in population  $i$ . A single action potential from neurons in population  $j$  is seen as a postsynaptic potential  $PS P_{ij}(t - s)$  by neurons in population  $i$ , where  $s$  is the time of the spike hitting the terminal and  $t$  the time after the spike. We neglect the delays due to the distance traveled down the axon by the spikes.

Assuming that they sum linearly, the average membrane potential of population  $i$  due to action potentials of population  $j$  is

$$V_i(t) = \sum_k PS P_{ij}(t - t_k),$$

where the sum is taken over the arrival times of the spikes produced by the neurons in population  $j$ . The number of spikes arriving between  $t$  and  $t + dt$  is  $v_j(t)dt$ . Therefore, we have

$$V_i(t) = \sum_j \int_0^t P S P_{ij}(t-s)v_j(s) ds = \sum_j \int_0^t P S P_{ij}(t-s)S_j(V_j(s)) ds,$$

or, equivalently,

$$v_i(t) = S_i \left( \sum_j \int_0^t P S P_{ij}(t-s)v_j(s) ds \right). \quad (2.2)$$

The  $P S P_{ij}$  can depend on several variables in order to account for adaptation and learning, among other examples.

There are two main simplifying assumptions that appear in the literature (Ermentrout, 1998) and produce two different models.

*2.1.1 The Voltage-Based Model.* The assumption (Hopfield, 1984) is that the postsynaptic potential has the same shape no matter which presynaptic population caused it; however, the sign and amplitude may vary. This leads to the relation

$$P S P_{ij}(t) = w_{ij} P S P_i(t).$$

If  $w_{ij} > 0$ , the population  $j$  excites population  $i$ , whereas it inhibits it when  $w_{ij} < 0$ .

Finally, if we assume that  $P S P_i(t) = A_i e^{-t/\tau_i} Y(t)$ , or equivalently that

$$\tau_i \frac{d P S P_i(t)}{dt} + P S P_i(t) = A_i \delta(t), \quad (2.3)$$

we end up with the following system of ordinary differential equations,

$$\tau_i \frac{d V_i(t)}{dt} + V_i(t) = \sum_j w_{ij} S_j(V_j(t)) + I_{\text{ext}}^i(t), \quad (2.4)$$

that describes the dynamic behavior of a cortical column. We have incorporated the constant  $A_i$  in the weights  $w_{ij}$  and added an external current  $I_{\text{ext}}(t)$  to model the nonlocal connections of population  $i$ . We introduce the  $n \times n$  matrixes  $\mathbf{W}$  such that  $W_{ij} = w_{ij}/\tau_i$  and the function  $\mathbf{S}, \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $\mathbf{S}(x)$  is the vector of coordinates  $S_i(x_i)$ , if  $x = (x_1, \dots, x_n)$ . We rewrite

equation 2.4 in vector form and obtain the following system of  $n$  ordinary differential equations,

$$\dot{\mathbf{V}} = -\mathbf{L}\mathbf{V} + \mathbf{W}\mathbf{S}(\mathbf{V}) + \mathbf{I}_{\text{ext}}, \quad (2.5)$$

where  $\mathbf{L}$  is the diagonal matrix  $\mathbf{L} = \text{diag}(1/\tau_i)$ .

In terms of units, the left- and right-hand sides of this equations are in units of, say,  $\text{mV} \times \text{ms}^{-1}$ . Therefore,  $\mathbf{I}_{\text{ext}}$ , despite its name, is not a current. Note that since  $\mathbf{S}(\mathbf{V})$  is an activity, its unit is  $\text{ms}^{-1}$ , and hence  $\mathbf{W}$  is in  $\text{mV}$ .

*2.1.2 The Activity-Based Model.* The assumption is that the shape of a postsynaptic potential (PSP) depends on only the nature of the presynaptic cell, that is,

$$PSP_{ij}(t) = w_{ij} PSP_j(t).$$

As above, we suppose that  $PSP_i(t)$  satisfies the differential equation 2.3 and define the time-averaged firing rate to be

$$A_j(t) = \int_0^t PSP_j(t-s)v_j(s) ds.$$

A similar derivation yields the following set of  $n$  ordinary differential equations:

$$\tau_i \frac{dA_i(t)}{dt} + A_i(t) = S_i \left( \sum_j w_{ij} A_j(t) + I_{\text{ext}}^i(t) \right) \quad i = 1, \dots, n.$$

We include the  $\tau_i$ s in the sigmoids  $S_i$  and rewrite this in vector form:

$$\dot{\mathbf{A}} = -\mathbf{L}\mathbf{A} + \mathbf{S}(\mathbf{W}\mathbf{A} + \mathbf{I}_{\text{ext}}). \quad (2.6)$$

The units are  $\text{ms}^{-2}$  for both sides of the equation.  $\mathbf{W}$  is expressed in  $\text{mV} \times \text{ms}$ , and  $\mathbf{I}_{\text{ext}}$  is in  $\text{mV}$ .

**2.2 The Continuum Models.** We now combine these local models to form a continuum of neural masses, for example, in the case of a model of a significant part  $\Omega$  of the cortex. We consider a subset  $\Omega$  of  $\mathbb{R}^q$ ,  $q = 1, 2, 3$ , which we assume to be connected and compact (i.e., closed and bounded). We note  $|\Omega|$  its Lebesgue measure (length, area, volume). This encompasses several cases of interest.

When  $q = 1$ , we deal with one-dimensional sets of neural masses. Although this appears to be of limited biological interest, this is one of the



most widely studied cases because of its relative mathematical simplicity and because of the insights one can gain of the more realistic situations.

When  $q = 2$ , we discuss properties of two-dimensional sets of neural masses. This is perhaps more interesting from a biological point of view since  $\Omega$  can be viewed as a piece of cortex where the third dimension, its thickness, is neglected. This case has received by far less attention than the previous one, probably because of the increased mathematical difficulty. Note that we could also take into account the curvature of the cortical sheet at the cost of an increase in mathematical difficulty. This is outside the scope of this letter.

Finally  $q = 3$  allows us to discuss properties of volumes of neural masses, such as cortical sheets where their thickness is taken into account (Kandel, Schwartz, & Jessel, 2000; Chalupa & Werner, 2004).

The theoretical results presented in this letter are independent of the value of  $q$ .

We note  $\mathbf{V}(\mathbf{r}, t)$  (resp.,  $\mathbf{A}(\mathbf{r}, t)$ ) the  $n$ -dimensional state vector at the point  $\mathbf{r}$  of the continuum and at time  $t$ . We introduce the  $n \times n$  matrix function  $\mathbf{W}(\mathbf{r}, \mathbf{r}')$ , which describes how the neural mass at point  $\mathbf{r}'$  influences that at point  $\mathbf{r}$  at time  $t$ . We call  $\mathbf{W}$  the connectivity matrix function. In particular,  $\mathbf{W}(\mathbf{r}, \mathbf{r}) = \mathbf{W}$ , the matrix that appears in equations 2.5 and 2.6. More precisely,  $W_{ij}(\mathbf{r}, \mathbf{r}')$  describes how population  $j$  at point  $\mathbf{r}'$  influences population  $i$  at point  $\mathbf{r}$  at time  $t$ . Equation 2.5 can now be extended to

$$\mathbf{V}_t(\mathbf{r}, t) = -\mathbf{L}\mathbf{V}(\mathbf{r}, t) + \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}')\mathbf{S}(\mathbf{V}(\mathbf{r}', t)) d\mathbf{r}' + \mathbf{I}_{\text{ext}}(\mathbf{r}, t), \quad (2.7)$$

and equation 2.6 to

$$\mathbf{A}_t(\mathbf{r}, t) = -\mathbf{L}\mathbf{A}(\mathbf{r}, t) + \mathbf{S} \left( \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}')\mathbf{A}(\mathbf{r}', t) d\mathbf{r}' + \mathbf{I}_{\text{ext}}(\mathbf{r}, t) \right). \quad (2.8)$$

It is important to discuss again the units of the quantities involved in these equations. For equation 2.7, as for equation 2.4, the unit is  $\text{mV} \times \text{ms}^{-1}$  for both sides. Because of the spatial integration,  $\mathbf{W}$  is in  $\text{mV} \times \text{ms}^{-1} \times \text{mm}^{-q}$ , and  $q$  is the dimension of the continuum. To obtain a dimensionless equation, we normalize (i.e., divide both sides of the equation) by the Frobenius norm  $\|\mathbf{W}\|_F$  of the connectivity matrix function  $\mathbf{W}$  (see section A.1 for a definition). Equivalently, we assume that  $\|\mathbf{W}\|_F = 1$ .

We have given elsewhere (Faugeras, Grimbert, & Slotine, 2007)—but see proposition 2 below for completeness—sufficient conditions on  $\mathbf{W}$  and  $\mathbf{I}_{\text{ext}}$  for equations 2.7 and 2.8 to be well defined and studied the existence and stability of their solutions for general external currents. In this letter, we analyze in detail the case of stationary external currents (independent of

the time variable) and investigate the existence and stability of the corresponding stationary solutions of 2.7 and 2.8.

A significant amount of work has been devoted to this or closely related problems, starting perhaps with the pioneering work of Wilson and Cowan (1973). A fairly recent review of this work, and much more, can be found in Coombes (2005). Amari (1977) investigated the problem in the case  $n = q = 1$  when the sigmoid function is approximated by a Heaviside function and the connectivity function has a Mexican hat shape. He proved the existence of stable bumps in this case. His work has been extended to different firing rate and connectivity functions (Gutkin, Ermentrout, & O'Sullivan, 2000; Laing, Troy, Gutkin, & Ermentrout, 2002; Laing & Troy, 2003; Rubin & Troy, 2004; Guo & Chow, 2005a, 2005b).

The case  $n = 1, q = 2$  has been considered by several authors, including Pinto and Ermentrout (2001a, 2001b) for general firing rate functions and gaussian-like connectivity functions (Blomquist, Wyller, & Einevoll, 2005) when the firing rate functions are approximated by Heaviside functions, and more recently for translation-invariant kernels including some discussion of sigmoids in Owen, Laing, and Coombes (2007).

Extending these analyses to a two- or three-dimensional continuum is difficult because of the increase in the degrees of freedom in the choice of the connectivity function. The case  $n = 2, q = 1$  has been studied in Werner and Richter (2001) and Bressloff (2005) when the firing rate functions are approximated by Heaviside functions and the connectivity function is circularly symmetric, while the case  $n = 2, q = 2$  is mentioned as difficult in Doubrovinski (2005).

In all of these contributions, the proof of the existence of a bump solution is based on Amari's original argument (1977), which works only when  $q = 1$  and the firing rate function is approximated by a Heaviside function. Kubota and Aihara (2005); Kishimoto and Amari (1979) consider sigmoids in the case  $q = 1$  for translation-invariant symmetric connectivity functions. Solutions are usually constructed using a variant of the method of the singular perturbation construction (Pinto & Ermentrout, 2001b). Sufficient conditions for their stability are obtained by a linear stability analysis, which in general requires the use of Heaviside functions instead of sigmoids.

The approach that we describe in this letter is a significant departure from previous ones. By using simple ideas of functional analysis, we are able to:

- Prove the existence and uniqueness of a stationary solution to equations 2.7 and 2.8 for any dimensions  $n$  and  $q$ , arbitrary connectivity functions, and general firing rate functions.
- Obtain very simple conditions for the absolute stability of the solution in terms of the spectrum of the differential of the nonlinear operator that appears on the right-hand side of equations 2.7 and 2.8.

- Construct a numerical approximation as accurately as needed of the solution, when it exists, for any stationary input.
- Characterize the sensitivity of the solutions to variations of the parameters, including the shape of the domain  $\Omega$ .

To be complete, let us point out that equations of the type 2.7 and 2.8 have been studied in pure mathematics (see e.g., Hazewinkel, 2001). They are of the Hammerstein type (Hammerstein, 1930; Tricomi, 1985). This type of equation has received some recent attention (see Appell & Chen, 2006), and progress has been made toward a better understanding of their solutions. Our contributions are the articulation of the models of networks of neural masses with this type of equation, the characterization of persistent activity in these networks as fixed points of Hammerstein equations, the proof of the existence of solutions, the characterization of their stability, and the analysis of their sensitivity to variations of the parameters involved in the equations.

The rest of the letter is organized as follows. In section 3, we prove that under some mild hypotheses about the connectivity matrix, because  $\Omega$  is bounded, there always exist stationary solutions to the neural field equations, 2.7 and 2.8, for stationary inputs. In section 4, we provide sufficient conditions for these stationary solutions to be stable. Section 5 provides numerical support for the theorems of the previous two sections: because we know that there exist stable solutions, we can derive a robust and accurate numerical scheme to compute them for various values of the parameters. In section 6, we probe the sensitivity of the solutions to variations of these parameters. This is a first step in the direction of the study of their bifurcations, which will be the topic of a forthcoming paper. In section 7, we summarize our results and open some perspectives.

### 3 Existence of Stationary Solutions

---

In this section we deal with the problem of the existence of stationary solutions to equations 2.7 and 2.8 for a given stationary external current  $\mathbf{I}_{\text{ext}}$ .

As indicated in the previous section, we use functional analysis to solve this problem. Let  $\mathcal{F}$  be the set  $\mathbf{L}_n^2(\Omega)$  of square integrable functions from  $\Omega$  to  $\mathbb{R}^n$ . This is a Hilbert, hence a Banach, space for the usual inner product,

$$\langle \mathbf{V}_1, \mathbf{V}_2 \rangle = \int_{\Omega} \mathbf{V}_1(\mathbf{r})^T \overline{\mathbf{V}_2(\mathbf{r})} d\mathbf{r},$$

where  $\overline{\mathbf{V}}$  is the complex conjugate of the vector  $\mathbf{V}$ . This inner product induces the norm  $\|\mathbf{V}\|_{\mathcal{F}}^2 = \sum_{i=1, \dots, n} \int_{\Omega} |V_i(\mathbf{r})|^2 d\mathbf{r}$  (see section A.1).  $\mathcal{F}$  is the state space. Another important space is  $\mathbf{L}_{n \times n}^2(\Omega \times \Omega)$ , the space of square integrable  $n \times n$  matrices (see section A.1 for a precise definition). We

assume that the connectivity matrix functions  $\mathbf{W}(\cdot, \cdot)$  are in this space (see propositions 1 and 2 below).

We also identify  $\mathbf{L}_{n \times n}^2(\Omega \times \Omega)$  with  $\mathcal{L}(\mathcal{F})$  (the space of continuous linear operators on  $\mathcal{F}$ ) as follows. If  $\mathbf{W} \in \mathbf{L}_{n \times n}^2(\Omega \times \Omega)$ , it defines a linear mapping

$$\begin{aligned} \mathbf{W} : \mathcal{F} &\longrightarrow \mathcal{F} \quad \text{such that} \\ \mathbf{X} &\rightarrow \mathbf{W} \cdot \mathbf{X} = \int_{\Omega} \mathbf{W}(\cdot, \mathbf{r}') \mathbf{X}(\mathbf{r}') d\mathbf{r}' \end{aligned}$$

For example, this allows us to write equations 2.7 and 2.8:

$$\begin{aligned} \mathbf{V}_t &= -\mathbf{L}\mathbf{V} + \mathbf{W} \cdot \mathbf{S}(\mathbf{V}) + \mathbf{I}_{\text{ext}} \\ \mathbf{A}_t &= -\mathbf{L}\mathbf{A} + \mathbf{S}(\mathbf{W} \cdot \mathbf{A} + \mathbf{I}_{\text{ext}}). \end{aligned}$$

We first recall some results on the existence of a solution to equations 2.7 and 2.8 that will be used in the sequel.

We denote by  $J$  a closed interval of the real line containing 0. A state vector  $\mathbf{X}(\mathbf{r}, t)$  is a mapping  $\mathbf{X} : J \rightarrow \mathcal{F}$ , and equations 2.7 and 2.8 are formally recast as an initial value problem:

$$\begin{cases} \mathbf{X}'(t) = f(t, \mathbf{X}(t)) \\ \mathbf{X}(0) = \mathbf{X}_0 \end{cases}, \quad (3.1)$$

where  $\mathbf{X}_0$  is an element of  $\mathcal{F}$  and the function  $f$  from  $J \times \mathcal{F}$  is defined by the right-hand side of equation 2.7, in which case we call it  $f_v$ , or equation 2.8, in which case we call it  $f_a$ . In other words, equations 2.7 and 2.8 become differential equations defined on the Hilbert space  $\mathcal{F}$ .

We need the following two propositions, which we quote without proof (Faugeras et al., 2007):

**Proposition 1.** *If the following two hypotheses are satisfied,*

1. *The connectivity function  $\mathbf{W}$  is in  $\mathbf{L}_{n \times n}^2(\Omega \times \Omega)$  (see section A.1),*
2. *At each time instant  $t \in J$  the external current  $\mathbf{I}$  is in  $C(J; \mathcal{F})$ , the set of continuous functions from  $J$  to  $\mathcal{F}$ ,*

*then the mappings  $f_v$  and  $f_a$  are from  $J \times \mathcal{F}$  to  $\mathcal{F}$ , continuous, and Lipschitz continuous with respect to their second argument, uniformly with respect to the first.*

**Proposition 2.** *If the following two hypotheses are satisfied,*

1. *The connectivity function  $\mathbf{W}$  is in  $\mathbf{L}_{n \times n}^2(\Omega \times \Omega)$ ,*
2. *The external current  $\mathbf{I}_{\text{ext}}$  is in  $C(J; \mathcal{F})$ , the set of continuous functions from  $J$  to  $\mathcal{F}$ ,*

then for any function  $\mathbf{X}_0$  in  $\mathcal{F}$ , there is a unique solution  $\mathbf{X}$ , defined on  $\mathbb{R}$  (and not only on  $J$ ) and continuously differentiable, of the abstract initial value problem, equation 3.1, for  $f = f_v$  and  $f = f_a$ .

This proposition says that given the two hypotheses and the initial condition, there exists a unique solution to equation 2.7 or 2.8 and that this solution is in  $C^1(\mathbb{R}; \mathcal{F})$ , the set of continuously differentiable functions from  $\mathbb{R}$  to  $\mathcal{F}$ .

We now turn our attention to a special type of solution of equations 2.7 and 2.8, corresponding to stationary external currents. We call these solutions, when they exist, *stationary solutions*. The currents  $\mathbf{I}_{\text{ext}}$  are simply in  $\mathcal{F}$ .

A stationary solution of equation 2.7 or 2.8 is defined by

$$\mathbf{X} = f^L(\mathbf{X}), \quad (3.2)$$

where the function  $f^L, \mathcal{F} \rightarrow \mathcal{F}$ , is equal to  $f_v^L$  defined by

$$f_v^L(\mathbf{V})(\mathbf{r}) = \int_{\Omega} \mathbf{W}^L(\mathbf{r}, \mathbf{r}') \mathbf{S}(\mathbf{V}(\mathbf{r}')) d\mathbf{r}' + \mathbf{I}_{\text{ext}}^L(\mathbf{r}), \quad (3.3)$$

or to  $f_a^L$  defined by

$$f_a^L(\mathbf{A})(\mathbf{r}) = \mathbf{S}^L \left( \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{A}(\mathbf{r}') d\mathbf{r}' + \mathbf{I}_{\text{ext}}(\mathbf{r}) \right), \quad (3.4)$$

where  $\mathbf{W}^L = \mathbf{L}^{-1} \mathbf{W}$ ,  $\mathbf{S}^L = \mathbf{L}^{-1} \mathbf{S}$  and  $\mathbf{I}_{\text{ext}}^L = \mathbf{L}^{-1} \mathbf{I}_{\text{ext}}$ .

We now recall:

**Definition 2.** A continuous mapping  $M: \mathcal{F} \rightarrow \mathcal{F}$  (linear or nonlinear) is called compact provided that for each bounded subset  $\mathcal{B}$  of  $\mathcal{F}$ , the set  $M(\mathcal{B})$  is relatively compact; that is, its closure is compact.

We then consider the nonlinear mapping  $g_v^L: \mathcal{F} \rightarrow \mathcal{F}$ ,

$$g_v^L(\mathbf{V})(\mathbf{r}) = \int_{\Omega} \mathbf{W}^L(\mathbf{r}, \mathbf{r}') \mathbf{S}(\mathbf{V}(\mathbf{r}')) d\mathbf{r}', \quad (3.5)$$

and the linear mappings  $g_a$  and  $g_a^L$ ,

$$g_a(\mathbf{A})(\mathbf{r}) = \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{A}(\mathbf{r}') d\mathbf{r}', \quad (3.6)$$

$$g_a^L(\mathbf{A})(\mathbf{r}) = \int_{\Omega} \mathbf{W}^L(\mathbf{r}, \mathbf{r}') \mathbf{A}(\mathbf{r}') d\mathbf{r}'. \quad (3.7)$$



We have the following:

**Proposition 3.** *If  $W \in L^2_{n \times n}(\Omega \times \Omega)$ ,  $g_v^L$  and  $g_a^L$  are compact operators of  $\mathcal{F}$ .*

**Proof.** We know from proposition 1 that  $g_v^L$  is continuous and prove that for each bounded sequence  $\{\mathbf{V}_n\}_{n=1}^\infty$  of  $\mathcal{F}$ , there exists a subsequence  $\{\mathbf{V}_{n_j}\}_{j=1}^\infty$  such that  $g_v^L(\mathbf{V}_{n_j})$  is convergent in  $\mathcal{F}$ .

Because of definition 1 of  $\mathbf{S}$ , the sequence  $\{\mathbf{A}_n = \mathbf{S}(\mathbf{V}_n)\}_{n=1}^\infty$  is bounded in  $\mathcal{F}$  by  $C = S_m \sqrt{n|\Omega|} > 0$ . We prove that there exists a subsequence  $\{\mathbf{A}_{n_j}\}_{j=1}^\infty$  such that  $\{g_a^L(\mathbf{A}_{n_j}) = g_v^L(\mathbf{V}_{n_j})\}_{j=1}^\infty$  converges in  $\mathcal{F}$ .

Since  $\mathcal{F}$  is separable, its unit ball is weakly compact, and because  $\{\mathbf{A}_n\}_{n=1}^\infty$  is bounded, there exists a subsequence  $\{\mathbf{A}_{n_j}\}_{j=1}^\infty$  of  $\{\mathbf{A}_n\}_{n=1}^\infty$  that converges weakly in  $\mathcal{F}$  toward  $\mathbf{A}$ . Because of Fubini's theorem, for almost all  $\mathbf{r} \in \Omega$  (noted almost surely), the function  $\mathbf{r}' \rightarrow \mathbf{W}(\mathbf{r}, \mathbf{r}')$  is in  $\mathcal{F}$ . Therefore, almost surely,  $\mathbf{B}_{n_j} = g_a^L(\mathbf{A}_{n_j}) \rightarrow \mathbf{B}$ .

Since  $\|\mathbf{A}\|_{\mathcal{F}} \leq \liminf_{j \rightarrow \infty} \|\mathbf{A}_{n_j}\|_{\mathcal{F}} \leq C$ ,  $\mathbf{A}$  is also bounded by  $C$  in  $\mathcal{F}$ . It is easy to show that  $\|\mathbf{B}_{n_j} - \mathbf{B}\|_{\mathcal{F}}^2 \leq 2C \|\mathbf{W}\|_F$ , and we can apply Lebesgue's dominated convergence theorem to the sequence  $\mathbf{B}_{n_j}(\mathbf{r}) - \mathbf{B}(\mathbf{r})$  and conclude that  $\|\mathbf{B}_{n_j} - \mathbf{B}\|_{\mathcal{F}} \rightarrow 0$ , that is,  $g_v^L(\mathbf{V}_{n_j})$  is convergent in  $\mathcal{F}$ .

A small variation of the proof shows that  $g_a^L$  is compact.

From proposition 3 follows:

**Proposition 4.** *Under the hypotheses of proposition 3, if  $\mathbf{I}_{\text{ext}} \in \mathcal{F}$ ,  $f_v^L$  and  $f_a^L$ , are compact operators of  $\mathcal{F}$ .*

**Proof.** The operators  $\mathbf{X} \rightarrow \mathbf{I}_{\text{ext}}^L$  and  $\mathbf{X} \rightarrow \mathbf{I}_{\text{ext}}$  are clearly compact under the hypothesis  $\mathbf{I}_{\text{ext}} \in \mathcal{F}$ ; therefore,  $f_v^L$  is the sum of two compact operators, hence compact. For the same reason,  $g_a + \mathbf{I}_{\text{ext}}$  is also compact, and so is  $f_a^L = \mathbf{S}^L(g_a + \mathbf{I}_{\text{ext}})$  because  $\mathbf{S}^L$  is smooth and bounded.

We can now prove:

**Theorem 1.** *If  $W \in L^2_{n \times n}(\Omega \times \Omega)$  and  $\mathbf{I}_{\text{ext}} \in \mathcal{F}$ , there exists a stationary solution of equations 2.7 and 2.8.*

**Proof.** A stationary solution of equation 2.7 (resp. of equation 2.8) is a fixed point of  $f_v^L$  (resp.  $f_a^L$ ).

Define the set  $\mathcal{C}_v = \{\mathbf{V} \in \mathcal{F} \mid \mathbf{V} = \lambda f_v^L(\mathbf{V}) \text{ for some } 0 \leq \lambda \leq 1\}$ . Because of lemma A.2 for all  $\mathbf{V} \in \mathcal{C}_v$ , we have

$$\|\mathbf{V}\|_{\mathcal{F}} \leq \lambda(\|g_v^L(\mathbf{V})\|_{\mathcal{F}} + \|\mathbf{I}_{\text{ext}}^L\|_{\mathcal{F}}) \leq \lambda(S_m \sqrt{n|\Omega|} \|\mathbf{W}^L\|_F + \|\mathbf{I}_{\text{ext}}^L\|_{\mathcal{F}}).$$

Hence,  $\mathcal{C}_v$  is bounded.

Similarly, define the set  $\mathcal{C}_a = \{\mathbf{A} \in \mathcal{F} \mid \mathbf{A} = \lambda f_a^L(\mathbf{A}) \text{ for some } 0 \leq \lambda \leq 1\}$ . Because of lemma A.2 for all  $\mathbf{A} \in \mathcal{C}_a$ , we have  $\|\mathbf{A}\|_{\mathcal{F}} \leq \lambda S_m \sqrt{n|\Omega|}$ ; hence,  $\mathcal{C}_a$  is bounded.

The conclusion follows from Schaefer's fixed-point theorem (Evans, 1998).

Note that in the proofs of proposition 4 and theorem 1, the assumption that  $\Omega$  is bounded (i.e., that  $|\Omega|$  is finite) is essential.

#### 4 Stability of the Stationary Solutions

In this section we give a sufficient condition on the connectivity matrix  $\mathbf{W}$  to guarantee the stability of the stationary solutions to equations 2.7 and 2.8.

**4.1 The Voltage-Based Model.** We define the "corrected maximal" connectivity function  $\mathbf{W}_{cm}(\mathbf{r}, \mathbf{r}')$  by  $\mathbf{W}_{cm} = \mathbf{W}DS_m$ , where  $DS_m$  is defined in definition 1. We also define the corresponding linear operator  $h_m : \mathcal{F} \rightarrow \mathcal{F}$ ,

$$h_m(\mathbf{V})(\mathbf{r}) = \int_{\Omega} \mathbf{W}_{cm}(\mathbf{r}, \mathbf{r}')\mathbf{V}(\mathbf{r}') d\mathbf{r}',$$

which is compact according to proposition 3. Its adjoint, noted  $h_m^*$  is defined<sup>1</sup> by

$$h_m^*(\mathbf{V})(\mathbf{r}) = \int_{\Omega} \mathbf{W}_{cm}^T(\mathbf{r}', \mathbf{r})\mathbf{V}(\mathbf{r}') d\mathbf{r}'$$

and is also compact. Hence the symmetric part  $h_m^s = \frac{1}{2}(h_m + h_m^*)$ , the sum of two compact operators, is also compact. Furthermore, we have  $\langle \mathbf{V}, h_m(\mathbf{V}) \rangle = \langle \mathbf{V}, h_m^s(\mathbf{V}) \rangle$ , as can be easily verified. It is also self-adjoint since, clearly,  $h_m^s = h_m^{s*}$ .

We recall the following property of the spectrum of a compact self-adjoint operator in a Hilbert space (see, e.g., Dieudonné, 1960).

**Proposition 5.** *The spectrum of a compact, self-adjoint operator of a Hilbert space is countable and real. Each nonzero spectral value is an eigenvalue, and the dimension of the corresponding eigenspace is finite.*

We have the following:

**Theorem 2.** *A sufficient condition for the stability of a stationary solution to equation 2.7 is that all the eigenvalues of the linear compact, self-adjoint operator  $h_m^{L,s}$  be less than 1, where  $h_m^{L,s}$  is defined by*

$$h_m^{L,s}(\mathbf{x})(\mathbf{r}) = \frac{1}{2} \int_{\Omega} \mathbf{L}^{-1/2}(\mathbf{W}_{cm}^T(\mathbf{r}', \mathbf{r}) + \mathbf{W}_{cm}(\mathbf{r}, \mathbf{r}'))\mathbf{L}^{-1/2} \mathbf{x}(\mathbf{r}') d\mathbf{r}' \forall \mathbf{x} \in \mathcal{F},$$

<sup>1</sup>By definition,  $\langle \mathbf{V}_1, h_m(\mathbf{V}_2) \rangle = \langle h_m^*(\mathbf{V}_1), \mathbf{V}_2 \rangle$ , for all elements  $\mathbf{V}_1, \mathbf{V}_2$  of  $\mathcal{F}$ .

where  $h_m^{L,s}$  is the symmetric part of the linear compact operator  $h_m^L : \mathcal{F} \rightarrow \mathcal{F}$ :

$$h_m^L(x)(\mathbf{r}) = \int_{\Omega} \mathbf{L}^{-1/2} \mathbf{W}_{cm}(\mathbf{r}, \mathbf{r}') \mathbf{L}^{-1/2} x(\mathbf{r}') d\mathbf{r}'$$

**Proof.** The proof of this theorem is a generalization to the continuum case of a result obtained by Matsuoka (1992).

Let us note  $\underline{\mathbf{S}}$  the function  $(D\mathbf{S}_m)^{-1}\mathbf{S}$  and rewrite equation 2.7 for a homogeneous input  $\mathbf{I}_{ext}$  as follows:

$$\mathbf{V}_i(\mathbf{r}, t) = -\mathbf{L}\mathbf{V}(\mathbf{r}, t) + \int_{\Omega} \mathbf{W}_{cm}(\mathbf{r}, \mathbf{r}') \underline{\mathbf{S}}(\mathbf{V}(\mathbf{r}', t)) d\mathbf{r}' + \mathbf{I}_{ext}(\mathbf{r}).$$

Let  $\mathbf{U}$  be a stationary solution of equation 2.7. Also let  $\mathbf{V}$  be the unique solution of the same equation with initially some condition  $\mathbf{V}(0) = \mathbf{V}_0 \in \mathcal{F}$  (see proposition 2). We introduce the new function  $\mathbf{X} = \mathbf{V} - \mathbf{U}$ , which satisfies

$$\begin{aligned} \mathbf{X}_i(\mathbf{r}, t) &= -\mathbf{L}\mathbf{X}(\mathbf{r}, t) + \int_{\Omega} \mathbf{W}_{cm}(\mathbf{r}, \mathbf{r}') \Theta(\mathbf{X}(\mathbf{r}', t)) d\mathbf{r}' \\ &= -\mathbf{L}\mathbf{X}(\mathbf{r}, t) + h_m(\Theta(\mathbf{X}))(\mathbf{r}, t), \end{aligned}$$

where the vector  $\Theta(\mathbf{X})$  is given by  $\Theta(\mathbf{X}(\mathbf{r}, t)) = \underline{\mathbf{S}}(\mathbf{V}(\mathbf{r}, t)) - \underline{\mathbf{S}}(\mathbf{U}(\mathbf{r})) = \underline{\mathbf{S}}(\mathbf{X}(\mathbf{r}, t) + \mathbf{U}(\mathbf{r})) - \underline{\mathbf{S}}(\mathbf{U}(\mathbf{r}))$ . Consider now the functional

$$\Delta(\mathbf{X}) = \int_{\Omega} \left( \sum_{i=1}^n \int_0^{X_i(\mathbf{r}, t)} \Theta_i(z) dz \right) d\mathbf{r}.$$

We note that

$$\begin{aligned} z \leq \Theta_i(z) < 0 \quad \text{for } z < 0 \quad \text{and} \quad 0 < \Theta_i(z) \leq z \\ \text{for } z > 0, \Theta_i(0) = 0, \quad i = 1, \dots, n. \end{aligned}$$

This is because (Taylor expansion with integral remainder)

$$\Theta_i(z) = \underline{S}_i(z + U_i) - \underline{S}_i(U_i) = z \int_0^1 \underline{S}'_i(U_i + \zeta z) d\zeta,$$

and  $0 < \underline{S}'_i \leq 1$  by construction of the vector  $\underline{\mathbf{S}}$ . This implies that the functional  $\Delta(\mathbf{X})$  is strictly positive for all  $\mathbf{X} \in \mathcal{F} \neq 0$  and  $\Delta(0) = 0$ . It also implies—and this is used in the sequel—that  $z\Theta_i(z) \geq \Theta_i(z)^2$ .

The time derivative of  $\Delta$  is readily obtained:

$$\frac{d\Delta(\mathbf{X})}{dt} = \int_{\Omega} \Theta^T(\mathbf{X}(\mathbf{r}, t)) \mathbf{X}_t(\mathbf{r}, t) d\mathbf{r} = \langle \Theta(\mathbf{X}), \mathbf{X}_t \rangle.$$

We replace  $\mathbf{X}_t(\mathbf{r}, t)$  by its value in this expression to obtain

$$\frac{d\Delta(\mathbf{X})}{dt} = -\langle \Theta(\mathbf{X}), \mathbf{L}\mathbf{X} \rangle + \langle \Theta(\mathbf{X}), h_m(\Theta(\mathbf{X})) \rangle.$$

Because of a previous remark, we have

$$\mathbf{X}^T(\mathbf{r}, t) \mathbf{L} \Theta(\mathbf{X}(\mathbf{r}, t)) \geq \Theta^T(\mathbf{X}(\mathbf{r}, t)) \mathbf{L} \Theta(\mathbf{X}(\mathbf{r}, t)),$$

and this provides an upper bound for  $\frac{d\Delta(\mathbf{X})}{dt}$ :

$$\frac{d\Delta(\mathbf{X})}{dt} \leq \langle \Theta(\mathbf{X}), (-\mathbf{L} + h_m^s) \cdot \Theta(\mathbf{X}) \rangle = \langle \mathbf{L}^{1/2} \Theta(\mathbf{X}), (-\text{Id} + h_m^{L,s}) \mathbf{L}^{1/2} \Theta(\mathbf{X}) \rangle,$$

and the conclusion follows.

Note that we have the following:

**Corollary 1.** *If the condition of theorem 2 is satisfied, the homogeneous solution of equation 2.7 is unique.*

**Proof.** Indeed, the result of theorem 2 is independent of the particular stationary solution  $\mathbf{U}$  that is chosen in the proof.

**4.2 The Activity-Based Model.** We now give a sufficient condition for the stability of a solution to equation 2.8. We define the “maximal corrected” connectivity matrix function  $\mathbf{W}_{mc} = \mathbf{D}\mathbf{S}_m\mathbf{W}$  and the linear compact operator  $k_m$  from  $\mathcal{F}$  to  $\mathcal{F}$ :

$$k_m(\mathbf{x})(\mathbf{r}) = \int_{\Omega} \mathbf{W}_{mc}(\mathbf{r}, \mathbf{r}') \mathbf{x}(\mathbf{r}') d\mathbf{r}'.$$

**Theorem 3.** *A sufficient condition for the stability of a solution to equation 2.8 is that all the eigenvalues of the linear compact operator  $k_m^L$  be of a magnitude less than 1, where  $k_m^L$  is defined by*

$$k_m^L(\mathbf{x})(\mathbf{r}) = \int_{\Omega} \mathbf{L}^{-1/2} \mathbf{W}_{mc}(\mathbf{r}, \mathbf{r}') \mathbf{L}^{-1/2} \mathbf{x}(\mathbf{r}') d\mathbf{r}' \quad \forall \mathbf{x} \in \mathcal{F}.$$

**Proof.** Let  $\mathbf{U}$  be a stationary solution of equation 2.8 for a stationary external current  $\mathbf{I}_{\text{ext}}(\mathbf{r})$ . As in the proof of theorem 2, we introduce the new function

$\mathbf{X} = \mathbf{V} - \mathbf{U}$ , where  $\mathbf{V}$  is the unique solution of the same equation with initial conditions  $\mathbf{V}(0) = \mathbf{V}_0 \in \mathcal{F}$ , an element of  $C(J, \mathcal{F})$ . We have

$$\begin{aligned} \mathbf{X}_t(\mathbf{r}, t) = & -\mathbf{L}\mathbf{X}(\mathbf{r}, t) + \mathbf{S} \left( \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{V}(\mathbf{r}', t) d\mathbf{r}' + \mathbf{I}_{\text{ext}}(\mathbf{r}) \right) \\ & - \mathbf{S} \left( \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{U}(\mathbf{r}') d\mathbf{r}' + \mathbf{I}_{\text{ext}}(\mathbf{r}) \right). \end{aligned}$$

Using a first-order Taylor expansion with an integral remainder, this equation can be rewritten as

$$\begin{aligned} \mathbf{X}_t(\mathbf{r}, t) = & -\mathbf{L}\mathbf{X}(\mathbf{r}, t) + \left( \int_0^1 DS \left( \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{U}(\mathbf{r}') d\mathbf{r}' + \mathbf{I}_{\text{ext}}(\mathbf{r}) \right. \right. \\ & \left. \left. + \zeta \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{X}(\mathbf{r}', t) d\mathbf{r}' \right) d\zeta \right) \left( \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{X}(\mathbf{r}', t) d\mathbf{r}' \right). \end{aligned}$$

Consider now the functional  $\Delta(\mathbf{X}) = \frac{1}{2} \|\mathbf{X}\|_{\mathcal{F}}^2$ . Its time derivative is

$$\frac{d\Delta(\mathbf{X})}{dt} = \langle \mathbf{X}, \mathbf{X}_t \rangle.$$

We replace  $\mathbf{X}_t(\mathbf{r}, t)$  by its value in this expression to obtain

$$\frac{d\Delta(\mathbf{X})}{dt} = -\langle \mathbf{X}, \mathbf{L}\mathbf{X} \rangle + \langle \mathbf{X}, \sigma_m(\mathbf{X})k_m(\mathbf{X}) \rangle,$$

where the nonlinear operator  $\sigma_m$  is defined by

$$\begin{aligned} \sigma_m(\mathbf{X})(\mathbf{r}, t) = & \int_0^1 DS \left( \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{U}(\mathbf{r}') d\mathbf{r}' \right. \\ & \left. + \zeta \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{X}(\mathbf{r}', t) d\mathbf{r}' \right) DS_m^{-1} d\zeta, \end{aligned}$$

a diagonal matrix whose diagonal elements are between 0 and 1. We rewrite  $\frac{d\Delta(\mathbf{X})}{dt}$  in a slightly different manner, introducing the operator  $k_m^L$ :

$$\frac{d\Delta(\mathbf{X})}{dt} = -\langle \mathbf{L}^{1/2}\mathbf{X}, \mathbf{L}^{1/2}\mathbf{X} \rangle + \langle \sigma_m(\mathbf{X})\mathbf{L}^{1/2}\mathbf{X}, k_m^L(\mathbf{L}^{1/2}\mathbf{X}) \rangle.$$

From the Cauchy-Schwarz inequality and the property of  $\sigma_m(\mathbf{X})$ , we obtain

$$\begin{aligned} \left| \langle \sigma_m(\mathbf{X})\mathbf{Y}, k_m^L(\mathbf{Y}) \rangle \right| & \leq \|\sigma_m(\mathbf{X})\mathbf{Y}\|_{\mathcal{F}} \|k_m^L(\mathbf{Y})\|_{\mathcal{F}} \leq \|\mathbf{Y}\|_{\mathcal{F}} \|k_m^L(\mathbf{Y})\|_{\mathcal{F}} \\ \mathbf{Y} & = \mathbf{L}^{1/2}\mathbf{X}. \end{aligned}$$



A sufficient condition for  $\frac{d\Delta(X)}{dt}$  to be negative is therefore that  $\|k_m^L(\mathbf{Y})\|_{\mathcal{F}} < \|\mathbf{Y}\|_{\mathcal{F}}$  for all  $\mathbf{Y}$ .

## 5 Numerical Experiments

---

In this section and the next, we investigate the question of effectively (i.e., numerically) computing stationary solutions of equation 2.7, which is equivalent to computing solutions of equation 3.2. Similar results are obtained for equation 2.8.

In all of our numerical experiments, we assume the sigmoidal functions  $S_i$ ,  $i = 1, \dots, n$  introduced in definition 1 to be of the form 2.1.

**5.1 Algorithm.** We now explain how to compute a fixed point  $\mathbf{V}^f$  of equation 3.3 in which we drop for simplicity the upper index  $L$  and the lower index ext:

$$\mathbf{V}^f = \mathbf{W} \cdot \mathbf{S}(\mathbf{V}^f) + \mathbf{I} \quad (5.1)$$

The method is iterative and based on Banach's fixed-point theorem (Evans, 1998):

**Theorem 4.** *Let  $X$  be Banach space and  $M : X \rightarrow X$  a nonlinear mapping such that*

$$\forall x, y \in X, \|M(x) - M(y)\| \leq q \|x - y\|, \quad 0 < q < 1.$$

*Such a mapping is said to be contracting.  $M$  has a unique fixed point  $x^f$  and for all  $x_0 \in X$  and  $x_{p+1} = M(x_p)$  then  $(x_p)$  converges geometrically to  $x^f$ .*

Note that this method allows to computing the solution of equation 3.2 only when it admits a unique solution and  $f$  is contracting. However, it could admit more than one solution (recall it always has a solution; see theorem 1) or  $f$  could be noncontracting. Another method has to be found in these cases.

In our case,  $X = \mathcal{F} = \mathbf{L}_n^2(\Omega)$ , where  $\Omega$  is an open-bounded set of  $\mathbb{R}^n$  and  $M = f_v$ . According to lemmas 2 and 1, if  $DS_m \|\mathbf{W}\|_F < 1$ ,  $f_v$  is contracting.

Each element of the sequence  $\mathbf{V}_p$ ,  $p \geq 0$  is approximated by a piecewise constant function  $\mathbf{V}_{p,h}$ , where  $h$  is the step of the approximation, defined by a finite number of points  $\mathbf{r}_{h,j} \in \Omega$ ,  $1 \leq j \leq \lfloor \frac{1}{h} \rfloor$ . In order to avoid difficulties because  $\mathbf{V}_{p,h} \in \mathbf{L}_n^2(\Omega)$ , hence defined almost everywhere, we assume that  $\mathbf{W}$  and  $\mathbf{I}$  are smooth. It is not a restriction because every function of  $\mathbf{L}_n^2(\Omega)$  can be approximated by a smooth function. As the bump solution is smooth, as soon as  $\mathbf{W}, \mathbf{I}$  are smooth, we can use the multidimensional gaussian quadrature formula (Press, Flannery, Teukolsky, & Vetterling, 1988; Stoer

& Bulirsch, 1972) with  $N$  points (in the examples below, usually  $N = 20$ ) on each axis. In order to interpolate the values of the bump from a finite (and small) number of its values  $\mathbf{V}_n(\mathbf{r}_{h,j,Gauss})$ , we use Nyström's method (Hazewinkel, 2001), as follows:

$$\mathbf{V}_p(\mathbf{r}) = \sum_j g_j \mathbf{W}_p(\mathbf{r}, \mathbf{r}_{p,j,Gauss}) \mathbf{S}(\mathbf{V}_p(\mathbf{r}_{p,j,Gauss})) + \mathbf{I}(\mathbf{r}),$$

where the  $g_j$ s are the weights of the gaussian quadrature method and the points  $\mathbf{r}_{p,j,Gauss}$  are chosen according to the gauss quadrature formula. It is to be noted that the choice of a particular quadrature formula can make a huge difference in accuracy and computing time (see section A.2).

Having chosen the type of quadrature we solve with Banach's theorem,

$$\mathbf{V}_h^f = \mathbf{W}_h \cdot \mathbf{S}(\mathbf{V}_h^f) + \mathbf{I}_h, \quad (5.2)$$

that is, we compute the fixed point at the level of approximation defined by  $h$ .

The following theorem ensures that  $\lim_{h \rightarrow 0} \mathbf{V}_h^f = \mathbf{V}^f$ :

**Theorem 5.** *Assume that  $\lim_{h \rightarrow 0} \mathbf{W}_h = \mathbf{W}$  in  $L^2_{n \times n}(\Omega \times \Omega)$  then  $\max_{1 \leq j \leq \lfloor \frac{1}{h} \rfloor} |\mathbf{V}_h(\mathbf{r}_{h,j}) - \mathbf{V}^f(\mathbf{r}_{h,j})| = O(a_h) \xrightarrow{h} 0$  with  $a_h = \|\mathbf{W} - \mathbf{W}_h\|_F$ .*

**Proof.** The proof is an adaptation of (Krasnosel'skii, Vainikko, Zabreiko, & Stetsenko, 1972, theorem 19.5).

**5.2 Examples of Bumps.** We show four examples of the application of the previous numerical method to the computation of bumps for various values of  $n$  and  $q$ .

There are  $n$  populations ( $\mathbf{V} = [V_1, \dots, V_n]^T$ ,  $\mathbf{W} \in L^2_{n \times n}(\Omega \times \Omega)$ ), some excitatory and some inhibitory.  $\Omega = [-1, 1]^q$ . We characterize in section 6.2 how the shape of  $\Omega$  influences that of the bumps. The connectivity matrix is of the form

$$W_{ij}(\mathbf{r}, \mathbf{r}') = \alpha_{ij} \exp\left(-\frac{1}{2} \langle \mathbf{r} - \mathbf{r}', \mathbf{T}_{ij}(\mathbf{r} - \mathbf{r}') \rangle\right), \quad (5.3)$$

with  $\mathbf{T}_{ij} \in \mathcal{M}_{q \times q}$  a  $q \times q$  symmetric positive definite matrix. The weights  $\alpha_{ij}$ ,  $i, j = 1, \dots, n$  form an element  $\boldsymbol{\alpha}$  of  $\mathcal{M}_{n \times n}$ , and  $\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{bmatrix}$  is an element of  $\mathcal{M}_{nq \times nq}$ . The weights  $\boldsymbol{\alpha}$  are chosen so that  $DS_m \|\mathbf{W}\|_F < 1$ . The sign of  $\alpha_{ij}$ ,  $i \neq j$ , indicates whether population  $j$  excites or inhibits population  $i$ . The bumps are computed using the algorithm described in the previous section.

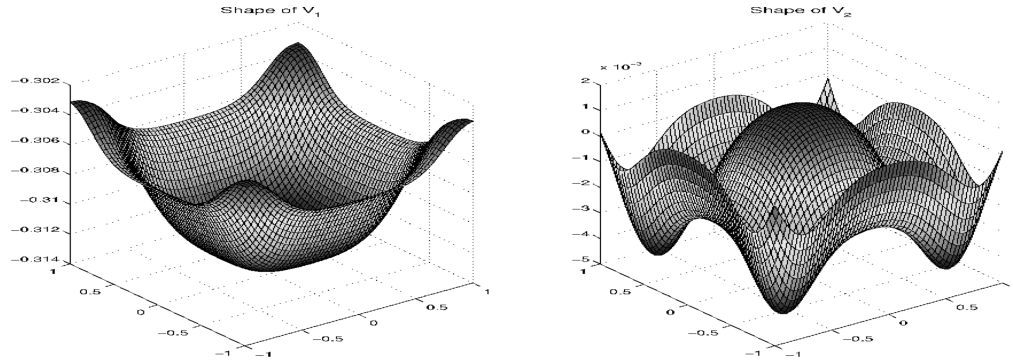


Figure 3: Example of a two-population, two-dimensional bump with constant external currents (see text).

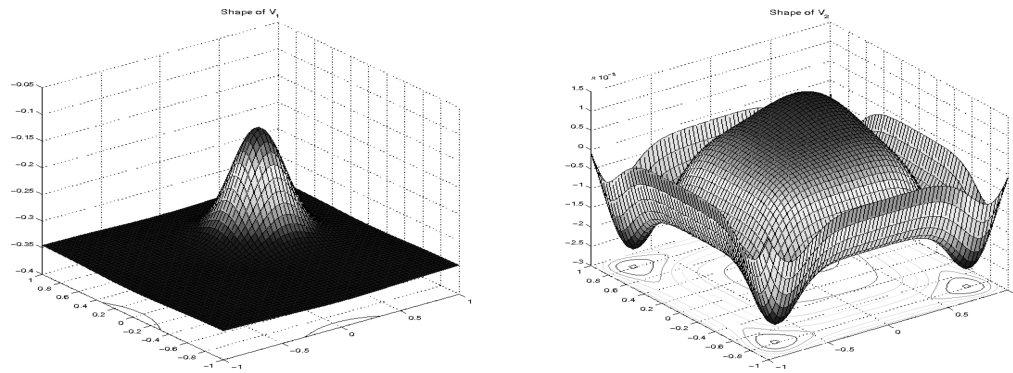


Figure 4: Example of a two-population, two-dimensional bump with gaussian-shaped external current (see text).

5.2.1 *First Example:  $n = 2, q = 2, \text{ Constant Current}$ .* Figure 3 shows an example of a bump for the following values of the parameters:

$$\alpha = \begin{bmatrix} 0.2 & -0.1 \\ 0.1 & -0.2 \end{bmatrix} \quad \mathbf{I} = [-0.3, 0]^T \quad \mathbf{T} = \begin{bmatrix} 40 & 0 & 12 & 0 \\ 0 & 40 & 0 & 12 \\ 8 & 0 & 20 & 0 \\ 0 & 8 & 0 & 20 \end{bmatrix}.$$

There is one excitatory and one inhibitory population of neural masses.

5.2.2 *Second Example:  $n = 2, q = 2, \text{ Nonconstant Current}$ .* Figure 4 shows a different example where the external current  $\mathbf{I}$  is still equal to 0 for its second coordinate and is not constant but equal to its previous value,  $-0.3$ , to which we have added a circularly symmetric 2D gaussian centered at the point of coordinates  $(0.5, 0.5)$  of the square  $\Omega$  with standard deviation 0.18 and maximum value 0.2. It is interesting to see how the shape of the

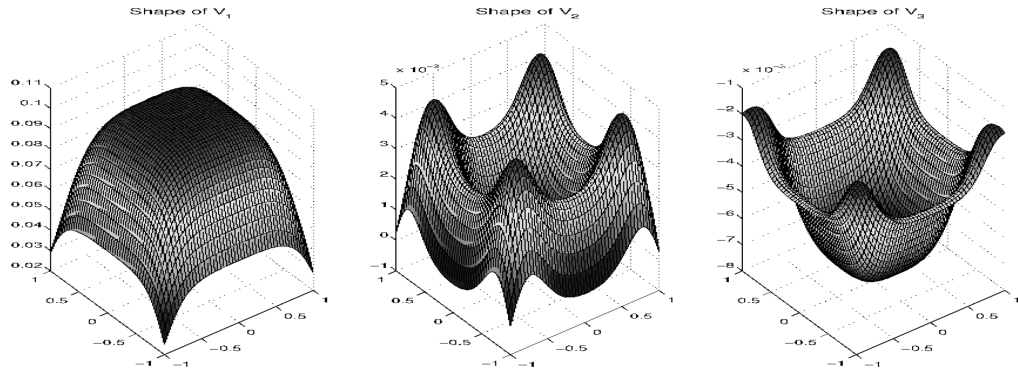


Figure 5: Example of a three-population, two-dimensional bump (see text).

previous bump is perturbed. The matrix  $\alpha$  is the same as in the first example. The matrix  $\mathbf{T}$  is equal to

$$\mathbf{T} = \begin{bmatrix} 5 & 0 & 1 & 0 \\ 0 & 5 & 0 & 1 \\ 16 & 0 & 40 & 0 \\ 0 & 16 & 0 & 40 \end{bmatrix},$$

corresponding to a spatially broader interaction for the first population and narrower for the second.

*5.2.3 Third Example:  $n = 3$ ,  $q = 2$ , Constant Current.* Figure 5 shows an example of a bump for three neural populations—two excitatory and one inhibitory—in two dimensions. We use the following values of the parameters:

$$\alpha = \begin{bmatrix} .442 & 1.12 & -0.875 \\ 0 & 0.1870 & -0.0850 \\ 0.128 & 0.703 & -0.7750 \end{bmatrix}^T \quad \mathbf{I} = [0, 0]^T$$

$$\mathbf{T} = \begin{bmatrix} 40 & 0 & 12 & 0 & 12 & 0 \\ 0 & 40 & 0 & 12 & 0 & 12 \\ 8 & 0 & 20 & 0 & 9 & 0 \\ 0 & 8 & 0 & 20 & 0 & 9 \\ 40 & 0 & 12 & 0 & 12 & 0 \\ 0 & 40 & 0 & 12 & 0 & 12 \end{bmatrix}.$$

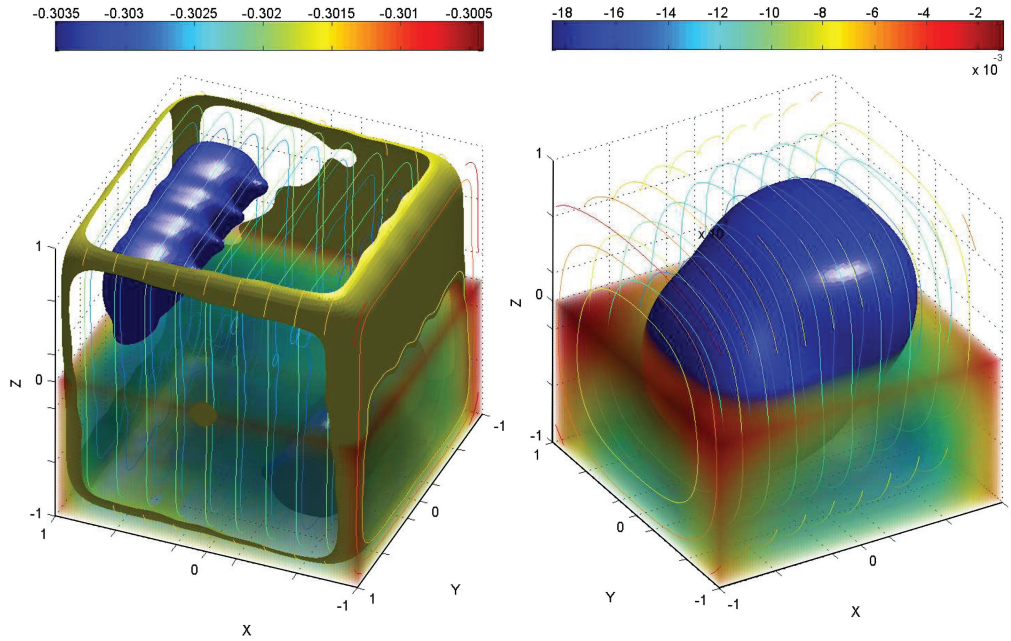


Figure 6: Example of a two-population, three-dimensional bump; isosurfaces are shown. Transparencies increase linearly from blue to red.

5.2.4 *Fourth Example:  $n = 2, q = 3$ , Constant Current.* We show an example of a three-dimensional bump for two populations of neural masses (see Figure 6). The parameters are

$$\alpha = \begin{bmatrix} 0.2 & -0.1 \\ 0.1 & -0.2 \end{bmatrix} \quad \mathbf{I} = [0, 0]^T \quad \mathbf{T} = \begin{bmatrix} 40 \text{Id}_3 & 12 \text{Id}_3 \\ 8 \text{Id}_3 & 20 \text{Id}_3 \end{bmatrix},$$

where  $\text{Id}_3$  is the  $3 \times 3$  identity matrix.

## 6 Sensitivity of the Bump to Variations of the Parameters

In this section, we characterize how the solutions of equation 3.2 vary with the parameters that appear in the equation. These parameters are of two types: first, we have a finite number of real parameters such as the external currents, the weights in the connectivity matrix  $\mathbf{W}$ , or the parameters of the sigmoids, and, second, the shape of the domain  $\Omega$ , a potentially infinite-dimensional parameter.

We focus on the voltage-based model; the analysis in the activity-based case is very similar. We start with a set of general considerations in the finite-dimensional case, which we then apply to the various cases. We then tackle the more difficult case of the dependency with respect to the shape of  $\Omega$ .



As  $f_v$  is a smooth function of the parameters  $(\mathbf{I}, \boldsymbol{\alpha}, \mathbf{S}, \dots)$ , one can show (by extending Banach's theorem) that the fixed-point  $\mathbf{V}^f$  inherits the smoothness of  $f_v$ .

**6.1 The Finite-Dimensional Parameters.** We introduce the linear operator<sup>2</sup>  $\mathbf{W} \cdot DS(\mathbf{V}^f) : \mathcal{F} \rightarrow \mathcal{F}$  such that

$$\mathbf{W} \cdot DS(\mathbf{V}^f) \cdot \mathbf{V}(\mathbf{r}) = \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}') DS(\mathbf{V}^f(\mathbf{r}')) \mathbf{V}(\mathbf{r}') d\mathbf{r}' \quad \forall \mathbf{V} \in \mathcal{F}.$$

We have the following:

**Proposition 6.** *The derivative  $\partial_{\lambda} \mathbf{V}^f$  of the fixed point  $\mathbf{V}^f$  with respect to the generic parameter  $\lambda$  satisfies the equation*

$$(\text{Id} - \mathbf{W} \cdot DS(\mathbf{V}^f)) \cdot \partial_{\lambda} \mathbf{V}^f = \mathbf{b}(\lambda, \mathbf{V}^f), \quad (6.1)$$

where  $\mathbf{b}(\lambda, \mathbf{V}^f) = (\partial_{\lambda} \mathbf{W}) \cdot \mathbf{S}(\mathbf{V}^f) + \mathbf{W} \cdot (\partial_{\lambda} \mathbf{S}(\mathbf{V}^f)) + \partial_{\lambda} \mathbf{I}$ .

**Proof.** Taking the derivative of both sides of equation 5.1 with respect to  $\lambda$ , we have

$$\partial_{\lambda} \mathbf{V}^f = \mathbf{W} \cdot DS(\mathbf{V}^f) \cdot \partial_{\lambda} \mathbf{V}^f + (\partial_{\lambda} \mathbf{W}) \cdot \mathbf{S}(\mathbf{V}^f) + \mathbf{W} \cdot (\partial_{\lambda} \mathbf{S}(\mathbf{V}^f)) + \partial_{\lambda} \mathbf{I}.$$

Hence, we obtain equation 6.1.

Note that  $\partial_{\lambda} \mathbf{S}(\mathbf{V}^f)$  is the partial derivative of the vector  $\mathbf{S}$  with respect to the scalar parameter  $\lambda$  evaluated at  $\mathbf{V} = \mathbf{V}^f$ .

Because of the assumption  $DS_m \|\mathbf{W}\|_F < 1$ , the linear operator  $\mathbf{J} = \text{Id} - \mathbf{W} \cdot DS(\mathbf{V}^f)$  is invertible with

$$\mathbf{J}^{-1} = \sum_{p=0}^{\infty} (\mathbf{W} \cdot DS(\mathbf{V}^f))^p,$$

and the series is convergent.

$\partial_{\lambda} \mathbf{V}^f$  is thus obtained from the following formula:

$$\partial_{\lambda} \mathbf{V}^f = \mathbf{J}^{-1} \mathbf{b}(\lambda, \mathbf{V}^f),$$

---

<sup>2</sup> $\mathbf{W} \cdot DS(\mathbf{V}^f)$  is the Frechet derivative of the operator  $f_v$  at the point  $\mathbf{V}^f$  of  $\mathcal{F}$ .

the right-hand side being computed by

$$\begin{cases} x_0 = \mathbf{b}(\lambda, \mathbf{V}^f) \\ x_{p+1} = x_0 + \mathbf{W} \cdot DS(\mathbf{V}^f) \cdot x_p \quad p \geq 0 \end{cases}$$

We now apply proposition 6 to the study of the sensitivity of the bump to the variations of the parameters.

*6.1.1 Sensitivity of the Bump to the Exterior Current.* When  $\lambda = I_1$ , we find:

$$\partial_{I_1} \mathbf{V}^f = \mathbf{J}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This inequality is to be understood component by component. It predicts the influence of  $I_1$  on  $\mathbf{V}^f$ . For example, with the parameters  $\alpha$  and  $\mathbf{T}$  used in Figure 3 but with an external current equal to 0, we obtain the bump shown in Figure 7 (top) with the derivatives shown at the bottom of the same figure. We also show in Figure 8 of  $V_1$  and  $V_2$  along the diagonal and the  $x$ -axis for different values of  $I_1$  close to 0. The reader can verify that the values increase with  $I_1$ , as predicted.

*6.1.2 Sensitivity of the Bump to the Weights  $\alpha$ .* For  $\lambda = \alpha_{ij}$ , one finds

$$\mathbf{J} \cdot \partial_{\alpha_{ij}} \mathbf{V}^f = \partial_{\alpha_{ij}} \mathbf{W} \cdot DS(\mathbf{V}^f).$$

We then have:

$\lambda = a$  : We find

$$\begin{aligned} \partial_a \mathbf{V}^f(\mathbf{r}) &= \mathbf{J}^{-1} \cdot \begin{bmatrix} \exp\left(-\frac{1}{2}\langle \mathbf{r} - \cdot, \mathbf{T}_{11}(\mathbf{r} - \cdot) \rangle\right) & 0 \\ 0 & 0 \end{bmatrix} \cdot DS(\mathbf{V}^f) \\ &= \mathbf{J}^{-1} \cdot \begin{bmatrix} \exp\left(-\frac{1}{2}\langle \mathbf{r} - \cdot, \mathbf{T}_{11}(\mathbf{r} - \cdot) \rangle\right) S'_1(V_1^f(\cdot)) \\ 0 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

The fixed point is an increasing function of the excitatory parameter  $a$ .

$\lambda = b$  : We find

$$\begin{aligned} \partial_b \mathbf{V}^f(\mathbf{r}) &= \mathbf{J}^{-1} \cdot \begin{bmatrix} 0 & -\exp\left(-\frac{1}{2}\langle \mathbf{r} - \cdot, \mathbf{T}_{12}(\mathbf{r} - \cdot) \rangle\right) \\ 0 & 0 \end{bmatrix} \cdot DS(\mathbf{V}^f) \\ &= \mathbf{J}^{-1} \cdot \begin{bmatrix} -\exp\left(-\frac{1}{2}\langle \mathbf{r} - \cdot, \mathbf{T}_{12}(\mathbf{r} - \cdot) \rangle\right) S'_2(V_2^f(\cdot)) \\ 0 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

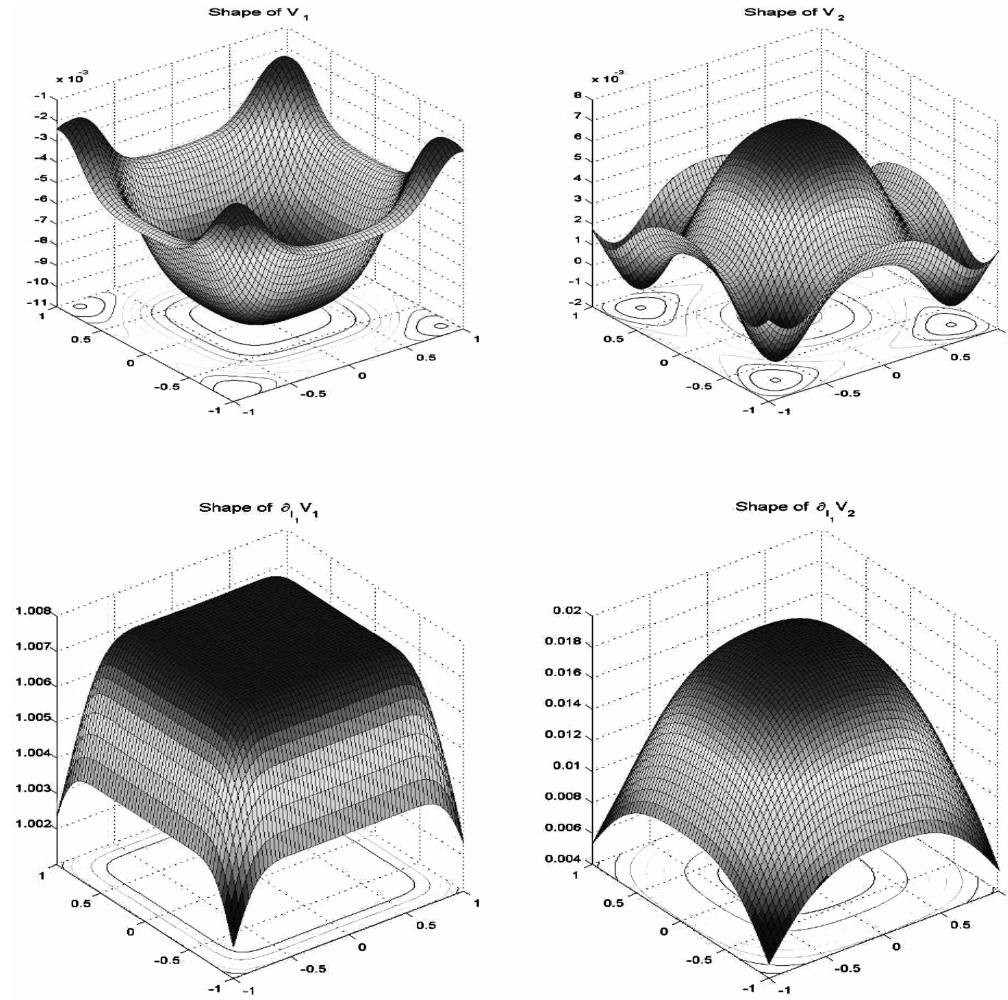


Figure 7: A bump corresponding to the following parameters:  $\alpha$  and  $\mathbf{T}$  are the same as in Figure 3,  $I = [0 \ 0]^T$  (top). Derivative of the bump with respect to the first coordinate,  $I_1$ , of the exterior current (bottom). We verify that it is positive (see text).

The fixed point is a decreasing function of the inhibitory parameter  $b$  (see Figure 9).

The other cases are similar.

*6.1.3 Sensitivity of the Bump to the Thresholds.* When  $\lambda = \theta_i$ ,  $i = 1, 2$  we have from definition 1 of  $\mathbf{S}$  and with the notations of proposition 6,

$$\mathbf{b}(\lambda, \mathbf{V}^f) = -\mathbf{W} \cdot \mathbf{DS}(\mathbf{V}^f) \cdot \mathbf{e}_i, \quad i = 1, 2,$$

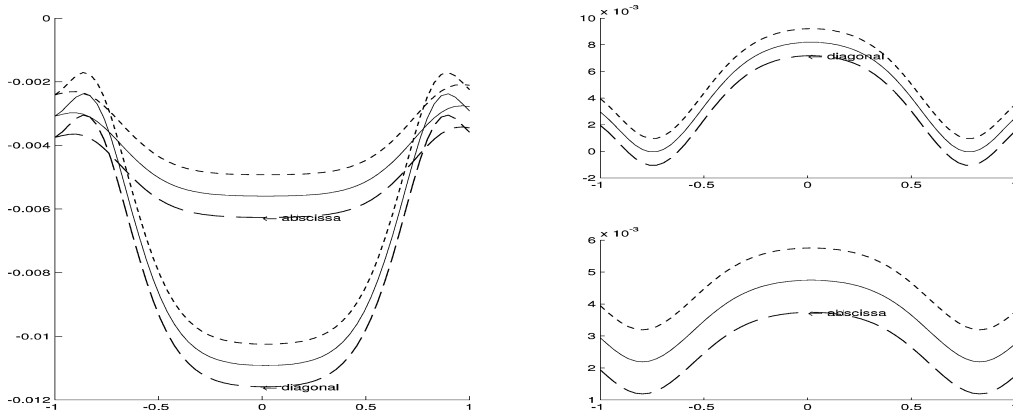


Figure 8: Cross sections of  $V_1$  (left) and  $V_2$  (right) for  $I_1 = -0.001$  (long-dashed line),  $I_1 = 0$  (continuous line) and  $I_1 = 0.001$  (dashed line).  $I_2 = 0$  in all three cases. To increase the readability of the results, we have applied an offset of 0.001 and 0.002 to the continuous and dashed curves on the right-hand side of the figure, respectively.

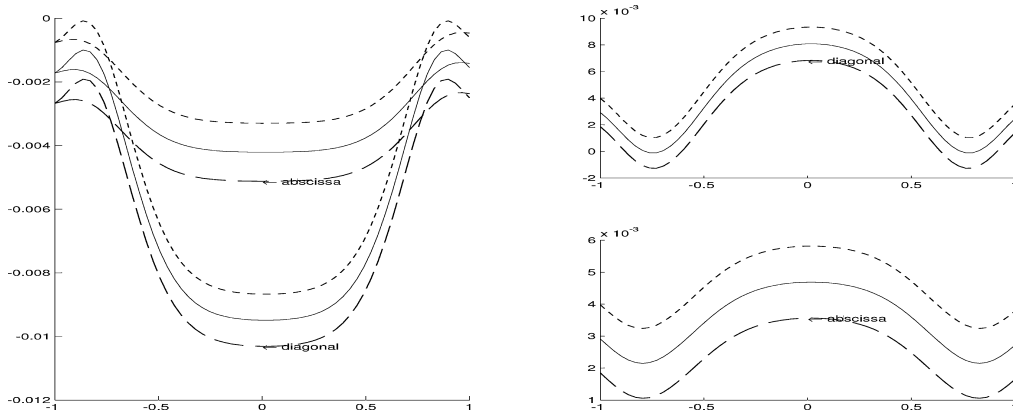


Figure 9: Cross sections of  $V_1$  (left) and  $V_2$  (right) for  $b = -0.101$  (long-dashed line),  $b = -0.1$  (continuous line), and  $b = -0.099$  (dashed line). To increase the readability of the results, we have applied an offset of 0.001 and 0.002 to all continuous and dashed curves, respectively.

where  $\mathbf{e}_1 = [1, 0]^T$ ,  $\mathbf{e}_2 = [0, 1]^T$ . We show in Figure 10 some cross sections of the bump  $\mathbf{V}^f$  obtained for the same values of the parameters as in Figure 3 and three values of the threshold vector.

6.1.4 *Sensitivity of the Bump to the Slope of the Sigmoid.* When  $\lambda = s_i$ ,  $i = 1, 2$ , we have from definition 1 of  $\mathbf{S}$  and with the notations of proposition 6,

$$\mathbf{b}(\lambda, \mathbf{V}^f) = \mathbf{W} \cdot DS(\mathbf{V}^f) \cdot \mathbf{s}(\mathbf{V}^f - \boldsymbol{\theta}),$$

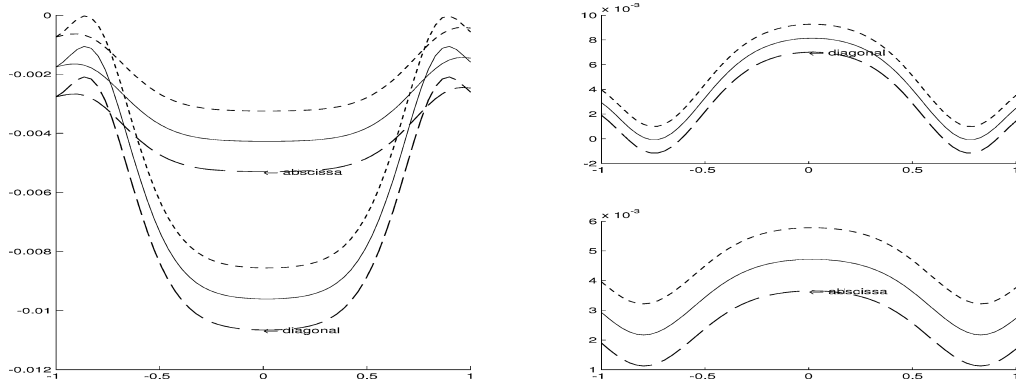


Figure 10: Cross sections of  $V_1$  (left) and  $V_2$  (right) for  $\theta = -0.101[1, 1]^T$  (long-dashed line),  $\theta = 0$  (continuous line), and  $\theta = 0.1[1, 1]^T$  (dashed line). To increase the readability of the results, we have applied an offset of 0.001 and 0.002 to all continuous and dashed curves, respectively.

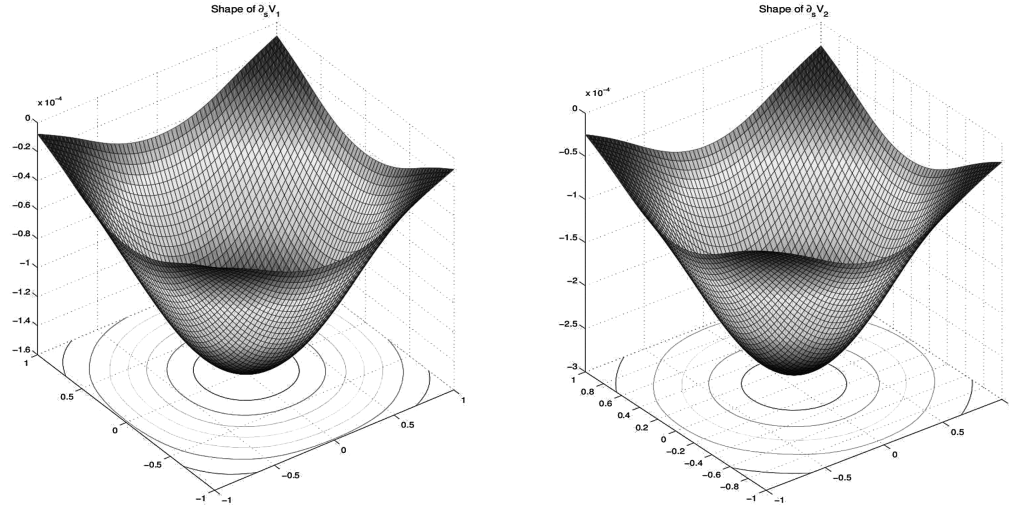


Figure 11: Plot of the derivative with respect to the slope of the sigmoids of the the bump obtained with the same parameters  $\alpha$ ,  $\mathbf{I}$ , and  $\mathbf{T}$  as in Figure 3.

where the matrix  $\mathbf{s}$  is given by

$$\mathbf{s} = \begin{bmatrix} \frac{1}{s_1} & 0 \\ 0 & \frac{1}{s_2} \end{bmatrix}.$$

Figure 11 shows the two coordinates  $\partial_s V_1^f$  and  $\partial_s V_2^f$  for  $s_1 = s_2 = s$  of the derivative of the bump  $\mathbf{V}^f$  at  $s = 1$  obtained for the same values of the other parameters as in Figure 3, except the intensity which is equal to 0.

### 6.2 Sensitivity of the Bump to Variations of the Shape of the Domain $\Omega$ .

We expect the bump to be somewhat dependent on the shape of  $\Omega$ . It would nonetheless be desirable that this dependency would not be too strong for



the modeling described in this letter to have some biological relevance. Indeed, if the bumps are metaphors of persistent states, we expect them to be relatively robust to the actual shape of the cortical part being modeled. For example, if we take  $\Omega$  to be a representation of the primary visual cortex V1 whose shape varies from individual to individual, it would come as a surprise if the shape of a bump induced by the same spatially constant stimulus were drastically different.

Technically, in order to study the dependence of  $\mathbf{V}^f$  with respect to  $\Omega$ , we need to assume that  $\Omega$  is smooth; its border  $\partial\Omega$  is a smooth curve ( $q = 2$ ) or surface ( $q = 3$ ) unlike the previous examples, where  $\Omega$  was the square  $[-1, 1]^2$ . But a difficulty arises from the fact that the set of regular domains is not a vector space—hence, the derivative of a function (the bump) with respect to a domain has to be defined with some care. The necessary background is found in section A.3.

We make explicit the fact that the connectivity function  $\mathbf{W}$  has been normalized to satisfy  $\|\mathbf{W}\|_F = 1$  by writing  $\mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega)$  where, with some abuse of notation,

$$\mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) = \mathbf{W}(\mathbf{r}, \mathbf{r}') / J(\Omega) \quad \text{with} \quad J(\Omega) = \sqrt{\int_{\Omega \times \Omega} \|\mathbf{W}(\mathbf{r}, \mathbf{r}')\|_F^2 d\mathbf{r} d\mathbf{r}'}$$

**Theorem 6.** *Let us assume that  $\Omega$  is a smooth, bounded domain of  $\mathbb{R}^q$ . If  $\mathbf{W}$  is in  $\mathbf{W}_{n \times n}^{1,2}(\Omega \times \Omega)$ ,  $\mathbf{I}_{ext}$  is in  $\mathbf{W}_n^{1,2}(\Omega)$  (see section A.1 for a definition) the material derivative (see section A.3 for a definition)  $\mathbf{V}_m^f(\mathbf{r}, \Omega)$  of the bump  $\mathbf{V}^f$  satisfies the following equation:*

$$\mathbf{V}_m^f(\mathbf{r}, \Omega, \mathbf{X}) = \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) DS(\mathbf{V}^f(\mathbf{r}', \Omega)) \mathbf{V}_m^f(\mathbf{r}', \Omega, \mathbf{X}) d\mathbf{r}' \quad (6.2)$$

$$+ \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) S(\mathbf{V}^f(\mathbf{r}', \Omega)) \text{div} \mathbf{X}(\mathbf{r}') d\mathbf{r}' \quad (6.3)$$

$$+ \int_{\Omega} D_1 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{X}(\mathbf{r}) S(\mathbf{V}^f(\mathbf{r}', \Omega)) d\mathbf{r}' \quad (6.4)$$

$$+ \int_{\Omega} D_2 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{X}(\mathbf{r}') S(\mathbf{V}^f(\mathbf{r}', \Omega)) d\mathbf{r}' \quad (6.5)$$

$$- \frac{\langle J'(\Omega), \mathbf{X} \rangle}{J(\Omega)} (\mathbf{V}^f(\mathbf{r}, \Omega) - \mathbf{I}_{ext}(\mathbf{r})) \\ + D\mathbf{I}_{ext}(\mathbf{r}) \cdot \mathbf{X}(\mathbf{r}), \quad (6.6)$$

where  $D_i$ ,  $i = 1, 2$  indicates the derivative with respect to the  $i$ th variable and  $\langle J'(\Omega), \mathbf{X} \rangle$  is the Gâteaux derivative of  $J(\Omega)$  with respect to the vector field  $\mathbf{X}$ :

$$\langle J'(\Omega), \mathbf{X} \rangle = \lim_{\tau \rightarrow 0} \frac{J(\Omega(\tau)) - J(\Omega)}{\tau},$$

where  $\mathbf{X}$  is defined in the proof below. We have

$$\begin{aligned} \langle J'(\Omega), \mathbf{X} \rangle &= \frac{1}{2J(\Omega)} \left( \int_{\Omega \times \partial\Omega} \|\mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega)\|_F^2 \langle \mathbf{X}(\mathbf{r}'), \mathbf{N}(\mathbf{r}') \rangle d\mathbf{r} d\mathbf{a}(\mathbf{r}') \right. \\ &\quad \left. + \int_{\partial\Omega \times \Omega} \|\mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega)\|_F^2 \langle \mathbf{X}(\mathbf{r}'), \mathbf{N}(\mathbf{r}') \rangle d\mathbf{a}(\mathbf{r}) d\mathbf{r}' \right), \end{aligned}$$

where  $d\mathbf{a}$  is the surface element on the smooth boundary  $\partial\Omega$  of  $\Omega$ , and  $\mathbf{N}$  is its unit inward normal vector.

**Proof.** The proof uses ideas that are developed in Delfour and Zolésio (2001) and Sokolowski and Zolésio (1992; see also section A.3). We want to compute

$$\mathbf{V}_m^f(\mathbf{r}, \Omega, \mathbf{X}) = \lim_{\tau \rightarrow 0} \frac{\mathbf{V}^f(\mathbf{r}(\tau), \Omega(\tau)) - \mathbf{V}^f(\mathbf{r}, \Omega)}{\tau}$$

from equation 3.2. As far as the computation of the derivative is concerned, only small deformations are relevant, and we consider the first-order Taylor expansion of the transformation  $T$ :

$$T(\tau, \mathbf{r}) = T(0, \mathbf{r}) + \tau \frac{\partial T}{\partial \tau}(0, \mathbf{r}) = \mathbf{r} + \tau \mathbf{X}(\mathbf{r}).$$

We define

$$\begin{aligned} \Delta &\equiv \frac{1}{\tau} \left( \int_{\Omega(\tau)} \mathbf{W}(\mathbf{r}(\tau), \mathbf{r}', \Omega(\tau)) \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega(\tau))) d\mathbf{r}' \right. \\ &\quad \left. - \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) d\mathbf{r}' \right. \\ &\quad \left. + \mathbf{I}_{\text{ext}}(\mathbf{r} + \tau \mathbf{X}(\mathbf{r})) - \mathbf{I}_{\text{ext}}(\mathbf{r}) \right). \end{aligned}$$

In the first integral, we make the change of variable  $\mathbf{r}' \rightarrow \mathbf{r}' + \tau \mathbf{X}$  and obtain

$$\begin{aligned} &\frac{1}{\tau} \int_{\Omega} \mathbf{W}(\mathbf{r} + \tau \mathbf{X}(\mathbf{r}), \mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X}) \\ &\quad \times \mathbf{S}(\mathbf{V}^f(\mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X})) |\det J_{\tau}(\mathbf{r}')| d\mathbf{r}'. \end{aligned}$$

We have

$$\det J_{\tau}(\mathbf{r}') = 1 + \tau \operatorname{div}(\mathbf{X}(\mathbf{r}')) + o(\tau).$$

Hence, for  $\tau$  sufficiently small  $\det J_\tau > 0$ . Moreover:

$$\lim_{\tau \rightarrow 0} \det J_\tau = 1 \quad \lim_{\tau \rightarrow 0} \frac{\det J_\tau(\mathbf{r}') - 1}{\tau} = \operatorname{div}(\mathbf{X}(\mathbf{r}')),$$

and

$$\begin{aligned} & \mathbf{W}(\mathbf{r} + \tau \mathbf{X}(\mathbf{r}), \mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X}) \\ &= \mathbf{W}(\mathbf{r}, \mathbf{r}') + \tau D_1 \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{X}(\mathbf{r}) + \tau D_2 \mathbf{W}(\mathbf{r}, \mathbf{r}') \mathbf{X}(\mathbf{r}') \\ & \quad - \tau \frac{\langle J'(\Omega), \mathbf{X} \rangle}{J(\Omega)} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) + o(\tau), \end{aligned}$$

where  $D_i$ ,  $i = 1, 2$  indicates the derivative with respect to the  $i$ th argument. Thus, we have

$$\begin{aligned} \tau \Delta &= \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) (\mathbf{S}(\mathbf{V}^f(\mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X})) \\ & \quad - \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega))) \det J_\tau(\mathbf{r}') d\mathbf{r}' \\ & \quad + \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) (\det J_\tau(\mathbf{r}') - 1) d\mathbf{r}' \\ & \quad + \tau \left\{ \int_{\Omega} D_1 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{X}(\mathbf{r}) \mathbf{S}(\mathbf{V}^f(\mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X})) \det J_\tau(\mathbf{r}') d\mathbf{r}' \right. \\ & \quad + \int_{\Omega} D_2 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{X}(\mathbf{r}') \mathbf{S}(\mathbf{V}^f(\mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X})) \det J_\tau(\mathbf{r}') d\mathbf{r}' \\ & \quad - \frac{\langle J'(\Omega), \mathbf{X} \rangle}{J(\Omega)} \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{S}(\mathbf{V}^f(\mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X})) \det J_\tau(\mathbf{r}') d\mathbf{r}' \\ & \quad \left. + D\mathbf{I}_{\text{ext}}(\mathbf{r}) \cdot \mathbf{X}(\mathbf{r}) \right\}. \end{aligned}$$

Because

$$\begin{aligned} & \lim_{\tau \rightarrow 0} \frac{\mathbf{S}(\mathbf{V}^f(\mathbf{r}' + \tau \mathbf{X}(\mathbf{r}'), \Omega + \tau \mathbf{X})) - \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega))}{\tau} \\ &= D\mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) \mathbf{V}_m^f(\mathbf{r}', \Omega, \mathbf{X}), \end{aligned}$$

and  $\int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) d\mathbf{r}' = \mathbf{V}^f(\mathbf{r}, \Omega) - \mathbf{I}_{\text{ext}}(\mathbf{r})$ , we obtain equation 6.2. The value of  $\langle J'(\Omega), \mathbf{X} \rangle$  is obtained from corollary 2.

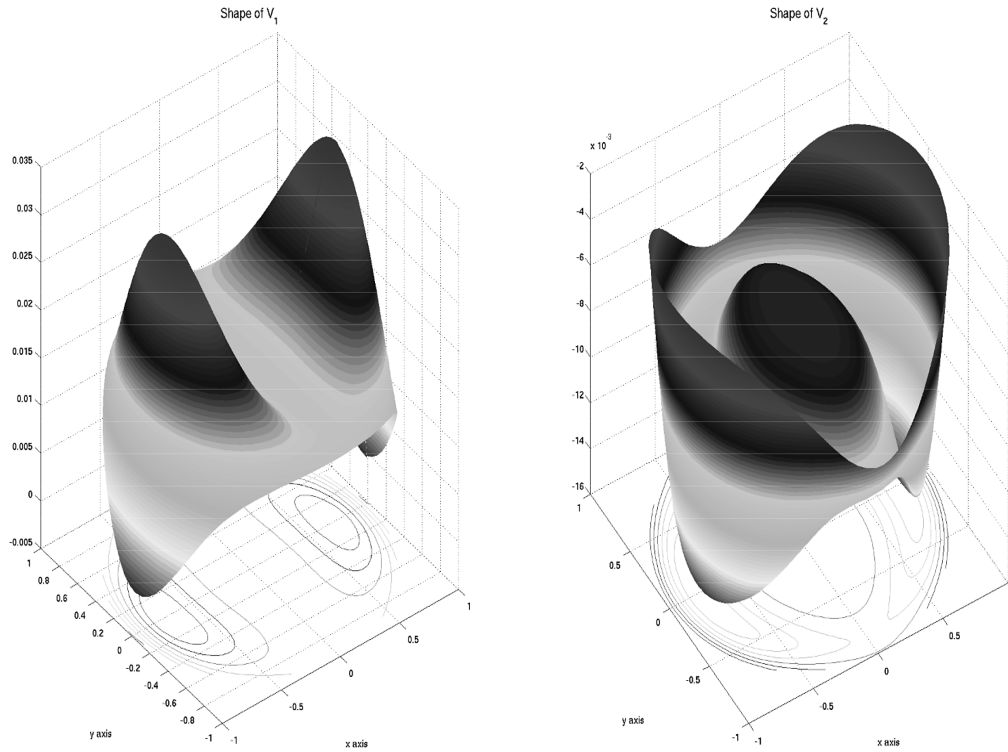


Figure 12: The unit disk and its bump  $\mathbf{V}^f$ .

Equation 6.2 is of the same form as before:

$$\begin{aligned}
 (\mathbf{J} \cdot \mathbf{V}_m^f)(\mathbf{r}, \Omega, \mathbf{X}) &= \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) \operatorname{div} \mathbf{X}(\mathbf{r}') d\mathbf{r}' \\
 &+ \int_{\Omega} D_1 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{X}(\mathbf{r}) \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) d\mathbf{r}' \\
 &+ \int_{\Omega} D_2 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) \mathbf{X}(\mathbf{r}') \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) d\mathbf{r}' \\
 &- \frac{\langle \mathbf{J}'(\Omega), \mathbf{X} \rangle}{J(\Omega)} (\mathbf{V}^f(\mathbf{r}, \Omega) - \mathbf{I}_{\text{ext}}(\mathbf{r})).
 \end{aligned}$$

This result tells us that the shape of the bump varies smoothly with respect to the shape of the domain  $\Omega$ .

*6.2.1 Numerical Application for the Domain Derivatives.* We show in Figure 12 the bump  $\mathbf{V}^f$  for  $\Omega$  equal to the unit disc  $D(0, 1)$  and in Figure 13 the one for  $\Omega$  equal to the ellipse<sup>3</sup>  $\text{Ellipse}(1.2, 1)$  of equation  $\frac{r_1^2}{a^2} + r_2^2 - 1 = 0$ .

<sup>3</sup> $\text{Ellipse}(a, b)$  represents the ellipse lying along the first axis of coordinates with semi-major axis  $a$  and semiminor axis  $b$ .

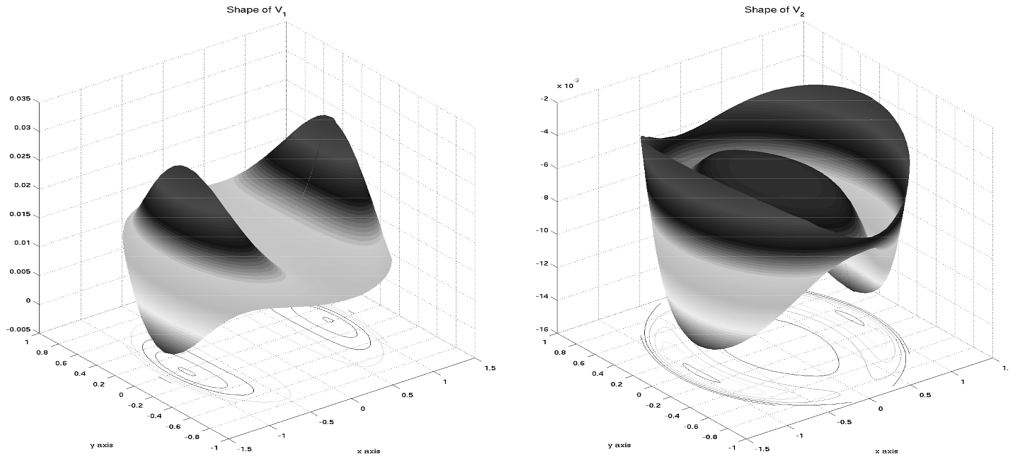


Figure 13: Bump associated with the ellipse with major axis along the  $r_1$  coordinate and the minor axis along the  $r_2$  coordinate. The ratio of the axes length is  $a = 1.2$  (see text).

The values of the weight parameters  $\alpha$  are the same as in Figure 3 and  $\mathbf{I} = [0, 0]^T$ . The matrix  $\mathbf{T}$  is equal to

$$\mathbf{T} = \begin{bmatrix} 40 & 0 & 10 & 0 \\ 0 & 10 & 0 & 12 \\ 12 & 0 & 40 & 0 \\ 0 & 40 & 0 & 40 \end{bmatrix}.$$

Note that because the diagonal elements are not equal for  $T_{11}$ ,  $T_{12}$ , and  $T_{13}$ ,  $\mathbf{W}$  is not circularly symmetric, and so is the bump in Figure 12 despite the fact that  $\Omega$  is circularly symmetric.

Finally we show in Figure 14 the two coordinates of the shape (material) derivative of the first bump in the direction of the field  $\mathbf{X}$  corresponding to the transformation

$$T(\tau, \mathbf{r}) = \mathbf{r} + \tau \begin{bmatrix} (a-1)r_1 \\ 0 \end{bmatrix}.$$

$T(1)$  transforms the disc  $D(0, 1)$  into the ellipse  $Ellipse(a, 1)$ ,  $\mathbf{X}(\mathbf{r}) = [(a-1)r_1, 0]^T$ .

Thus,  $div\mathbf{X} = (a-1)$  and, because of equation 5.3,

$$\begin{aligned} (\mathbf{J} \cdot \mathbf{V}_m^f)(\mathbf{r}, \Omega, \mathbf{X}) &= \left( a-1 - \frac{\langle J'(\Omega), \mathbf{X} \rangle}{J(\Omega)} \right) (\mathbf{V}^f - \mathbf{I}) \\ &\quad + \int_{\Omega} D_1 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) (\mathbf{X}(\mathbf{r}) - \mathbf{X}(\mathbf{r}')) \mathbf{S}(\mathbf{V}^f(\mathbf{r}', \Omega)) d\mathbf{r}', \end{aligned}$$



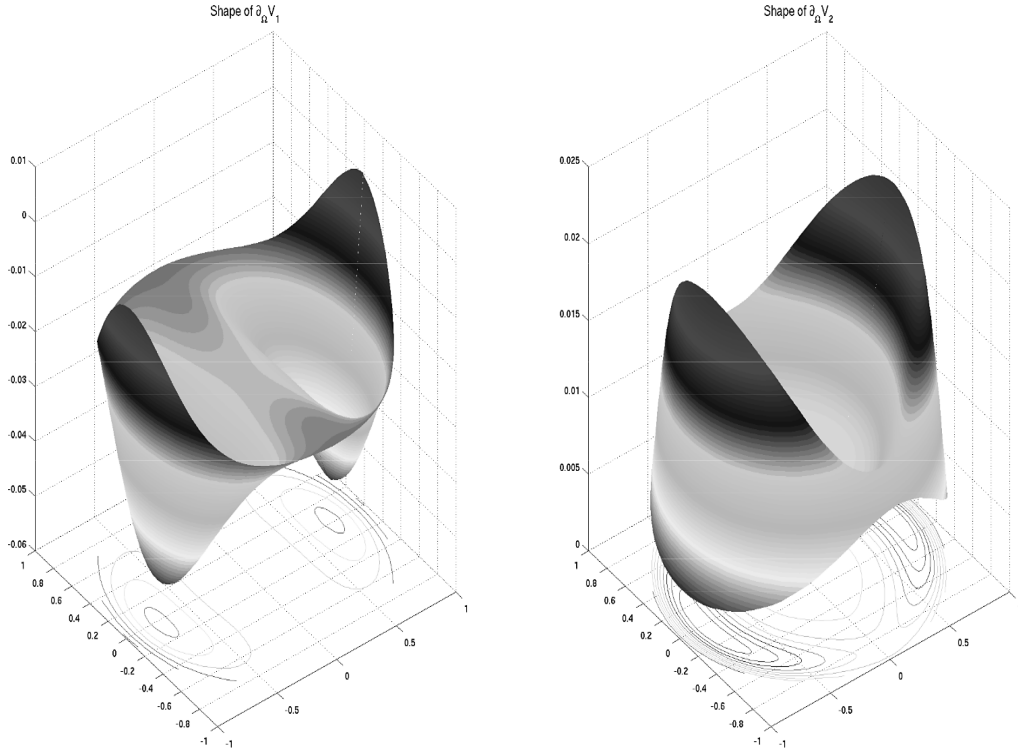


Figure 14: The shape derivative  $V_m^f$  for  $a = 1.2$ .

and

$$\langle J'(\Omega), \mathbf{X} \rangle = \frac{\int_{\Omega \times \partial\Omega} \|\mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega)\|_F^2 \langle \mathbf{X}(\mathbf{r}'), \mathbf{N}(\mathbf{r}') \rangle d\mathbf{r} d\mathbf{a}(\mathbf{r}')}{J(\Omega)}.$$

As the gaussian quadrature formula holds for a rectangular domain, we use polar coordinates to map the disk (or the ellipse) to a square. For our numerical study, we can simplify these expressions (the matrixes  $\mathbf{T}_{ij}$  are symmetric):

$$\begin{aligned} & \left[ \int_{\Omega} D_1 \mathbf{W}(\mathbf{r}, \mathbf{r}', \Omega) (\mathbf{X}(\mathbf{r}) - \mathbf{X}(\mathbf{r}')) \mathbf{S}(V^f(\mathbf{r}', \Omega)) d\mathbf{r}' \right]_i \\ &= \sum_j \int_{\Omega} (\mathbf{r} - \mathbf{r}')^T \mathbf{T}_{ij} (\mathbf{X}(\mathbf{r}) - \mathbf{X}(\mathbf{r}')) W_{ij}(\mathbf{r}, \mathbf{r}', \Omega) S_j(V_j^f(\mathbf{r}', \Omega)) d\mathbf{r}' \\ & \qquad \qquad \qquad i = 1, \dots, n. \end{aligned}$$

Thus, we can use a simple modification of the algorithm that computes  $\mathbf{W} \cdot \mathbf{S}(\mathbf{V})$  to obtain the previous expression.

$J(\Omega)$  and  $\langle J'(\Omega), \mathbf{X} \rangle$  are computed with a gauss quadrature formula. For a circle in polar coordinates,  $\mathbf{N}(\mathbf{r}') = \mathbf{r}'$ .

Let us be a bit more precise. In the case shown in Figure 12, we choose  $\mathbf{I} = 0$ . Using Banach's theorem, we compute  $\mathbf{V}_{Gauss}^f$  for  $N = 30$  and use Nyström's interpolation to compute  $\mathbf{V}_{Nys}^f$  for  $n = 100$  (for example) points on each axis.

Then, using  $\mathbf{V}_{Gauss}^f$ , we compute  $\mathbf{V}_{m,Gauss}^f$  for  $N$  points. But the equation for  $\mathbf{V}_m^f$  reads

$$\mathbf{V}_m^f = \mathbf{W}.DS(\mathbf{V}^f).\mathbf{V}_m^f + \langle J'(\Omega), \mathbf{X} \rangle.$$

Having computed a Nyström interpolation of  $n$  points for  $\mathbf{V}_m^f = \mathbf{W}.DS(\mathbf{V}^f).\mathbf{V}_m^f + \langle J'(\Omega), \mathbf{X} \rangle$ , we again use a Nyström interpolation with the last equation to compute  $\mathbf{V}_{m,Nyström}^f$  for  $n$  points on each axis.

We used this numerical method in every previous example related to the computation of derivatives.

## 7 Conclusion and Perspectives

---

We have studied two classes (voltage and activity based) of neural continuum networks in the context of modeling macroscopic parts of the cortex. In both cases, we have assumed an arbitrary number of interacting neuron populations, either excitatory or inhibitory. These populations are spatially related by nonsymmetric connectivity functions representing cortico-cortical, local connections. External inputs are also present in our models to represent nonlocal connections, for example, with other cortical areas. The relationship between (average) membrane potential and activity is described by nondegenerate sigmoidal nonlinearities, that is, not by Heaviside functions, which have often been considered instead in the literature because of their (apparent) simplicity.

The resulting nonlinear integro-differential equations are of the Hammerstein type (Hammerstein, 1930) and generalize those proposed by Wilson and Cowan (1973).

Departing from most of the previous work in this area, we relax the usual assumption that the domain of definition where we study these networks is infinite, that is, equal to  $\mathbb{R}$  or  $\mathbb{R}^2$ , and we explicitly consider the biologically much more relevant case of a bounded subset  $\Omega$  of  $\mathbb{R}^d$ ,  $d = 1, 2, 3$ , obviously a better model of a piece of cortex. The importance of this has been emphasized by Nunez (2005) for a number of years, and we fully agree with him on this.

Using methods of functional analysis, we have studied the existence and uniqueness of a stationary (i.e., time-independent) solution of these equations in the case of a stationary input. These solutions are often referred to as persistent states, or bumps, in the literature.

We have proved that under very mild assumptions on the connectivity functions, such solutions always exist (this is due in part to the fact that we

do not use Heaviside functions and mostly to the fact that we consider the cortex as a bounded set).

We have provided sufficient conditions on the connectivity functions for the solution to be absolutely stable, that is, independent of the initial state of the network. These conditions can be expressed in terms of the spectra of some functional operators, which we prove to be compact, that arise naturally from the equations describing the network activity.

We have also studied the sensitivity of the solutions to variations of such parameters as the connectivity functions, the sigmoids, the external inputs, and the shape of the domain of definition of the neural continuum networks. This last analysis is more involved than the others because of the infinite-dimensional nature of the shape parameter. An analysis of the bifurcations of the solutions when the parameters vary over large ranges requires the use of techniques of bifurcation analysis for infinite-dimensional systems and is out of the scope of this letter.

We believe, and we hope by now to have convinced the reader, that the functional analysis framework that we have used in this letter is the right one to try to answer some of the mathematical questions raised by models of connected networks of nonlinear neurons. We also believe that some of these also begin to answer biological questions since these networks models, despite the admittedly immense simplifications they are built from, are nonetheless metaphors of real neural assemblies.

## Appendix A: Notations and Background Material

---

**A.1 Matrix Norms and Spaces of Functions.** We note  $\mathcal{M}_{n \times n}$  the set of  $n \times n$  real matrixes. We consider the Frobenius norm on  $\mathcal{M}_{n \times n}$ ,

$$\|\mathbf{M}\|_F = \sqrt{\sum_{i,j=1}^n M_{ij}^2},$$

and consider the space  $\mathbf{L}_{n \times n}^2(\Omega \times \Omega)$  of the functions from  $\Omega \times \Omega$  to  $\mathcal{M}_{n \times n}$  whose Frobenius norm is in  $L^2(\Omega \times \Omega)$ . If  $\mathbf{W} \in \mathbf{L}_{n \times n}^2(\Omega \times \Omega)$ , we note  $\|\mathbf{W}\|_F^2 = \int_{\Omega \times \Omega} \|\mathbf{W}(\mathbf{r}, \mathbf{r}')\|_F^2 d\mathbf{r} d\mathbf{r}'$ . Note that this implies that each element  $W_{ij}$ ,  $i, j = 1, \dots, n$  in  $L^2(\Omega \times \Omega)$ . We note  $\mathcal{F}$  the set  $\mathbf{L}_n^2(\Omega)$  of square-integrable mappings from  $\Omega$  to  $\mathbb{R}^n$  and  $\|\mathbf{x}\|_{\mathcal{F}}$  the corresponding norm. We have the following:

**Lemma 1.** *Given  $\mathbf{x} \in \mathbf{L}_n^2(\Omega)$  and  $\mathbf{W} \in \mathbf{L}_{n \times n}^2(\Omega \times \Omega)$ , we define  $\mathbf{y}(\mathbf{r}) = \int_{\Omega} \mathbf{W}(\mathbf{r}, \mathbf{r}')\mathbf{x}(\mathbf{r}') d\mathbf{r}'$ . This integral is well defined for almost all  $\mathbf{r}$ ,  $\mathbf{y}$  is in  $\mathbf{L}_n^2(\Omega)$ , and we have*

$$\|\mathbf{y}\|_F \leq \|\mathbf{W}\|_F \|\mathbf{x}\|_F.$$

**Proof.** Since each  $W_{ij}$  is in  $L^2(\Omega \times \Omega)$ ,  $W_{ij}(\mathbf{r}, \cdot)$  is in  $L^2(\Omega)$  for almost all  $\mathbf{r}$ , thanks to Fubini's theorem. So  $W_{ij}(\mathbf{r}, \cdot)x_j(\cdot)$  is integrable for almost all  $\mathbf{r}$  from what we deduce that  $\mathbf{y}$  is well defined for almost all  $\mathbf{r}$ . Next, we have

$$|y_i(\mathbf{r})| \leq \sum_j \left| \int_{\Omega} W_{ij}(\mathbf{r}, \mathbf{r}') x_j(\mathbf{r}') d\mathbf{r}' \right|$$

and (Cauchy-Schwarz)

$$|y_i(\mathbf{r})| \leq \sum_j \left( \int_{\Omega} W_{ij}^2(\mathbf{r}, \mathbf{r}') d\mathbf{r}' \right)^{1/2} \|x_j\|_2,$$

from which it follows that (Cauchy-Schwarz again, discrete version):

$$\begin{aligned} |y_i(\mathbf{r})| &\leq \left( \sum_j \|x_j\|_2^2 \right)^{1/2} \left( \sum_j \int_{\Omega} W_{ij}^2(\mathbf{r}, \mathbf{r}') d\mathbf{r}' \right)^{1/2} \\ &= \|\mathbf{x}\|_{\mathcal{F}} \left( \sum_j \int_{\Omega} W_{ij}^2(\mathbf{r}, \mathbf{r}') d\mathbf{r}' \right)^{1/2}, \end{aligned}$$

from which it follows that  $\mathbf{y}$  is in  $L_n^2(\Omega)$  (thanks again to Fubini's theorem) and

$$\|\mathbf{y}\|_{\mathcal{F}}^2 \leq \|\mathbf{x}\|_{\mathcal{F}}^2 \sum_{i,j} \int_{\Omega \times \Omega} W_{ij}^2(\mathbf{r}, \mathbf{r}') d\mathbf{r}' d\mathbf{r} = \|\mathbf{x}\|_{\mathcal{F}}^2 \|\mathbf{W}\|_F^2.$$

We also use the following:

**Lemma 2.** For each  $\mathbf{V}$  of  $\mathcal{F}$ ,  $\mathbf{S}(\mathbf{V})$  is in  $\mathcal{F}$ ; and we have

$$\|\mathbf{S}(\mathbf{V})\|_{\mathcal{F}} \leq S_m \sqrt{n|\Omega|}.$$

For all  $\mathbf{V}_1$  and  $\mathbf{V}_2$  in  $\mathcal{F}$  we have

$$\|\mathbf{S}(\mathbf{V}_1) - \mathbf{S}(\mathbf{V}_2)\|_{\mathcal{F}} \leq DS_m \|\mathbf{V}_1 - \mathbf{V}_2\|_{\mathcal{F}},$$

where  $DS_m$  is defined in definition 1.

**Proof.** We have  $\|\mathbf{S}(\mathbf{V})\|_{\mathcal{F}}^2 = \sum_{i=1}^n \int_{\Omega} (S_i(V_i(\mathbf{r})))^2 d\mathbf{r} \leq S_m^2 n |\Omega|$ , where  $|\Omega|$  is the Lebesgue measure of  $\Omega$  (its area). Similarly,  $\|\mathbf{S}(\mathbf{V}_1) - \mathbf{S}(\mathbf{V}_2)\|_{\mathcal{F}}^2 = \sum_{i=1}^n$

$$\int_{\Omega} (S_i(V_i^1(\mathbf{r})) - S_i(V_i^2(\mathbf{r})))^2 d\mathbf{r} \leq (DS_m)^2 \sum_{i=1}^n \int_{\Omega} (V_i^1(\mathbf{r}) - V_i^2(\mathbf{r}))^2 d\mathbf{r} = (DS_m)^2 \|\mathbf{V}_1 - \mathbf{V}_2\|_{\mathcal{F}}^2.$$

In theorem 6, we use the Sobolev spaces  $\mathbf{W}_n^{1,2}(\Omega)$  and  $\mathbf{W}_{n \times n}^{1,2}(\Omega \times \Omega)$ .  $\mathbf{W}_n^{1,2}(\Omega)$  is the set of functions  $\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$  such that each component  $X_i$ ,  $i = 1, \dots, n$  is in  $W^{1,2}(\Omega)$ , the set of functions of  $L^2(\Omega)$  whose first-order derivatives exist in the weak sense and are also in  $L^2(\Omega)$  (see Evans, 1998). Similarly  $\mathbf{W}_{n \times n}^{1,2}(\Omega \times \Omega)$  is the set of functions  $\mathbf{X} : \Omega \rightarrow \mathcal{M}_{n \times n}$  such that each component  $X_{ij}$ ,  $i, j = 1, \dots, n$  is in  $W^{1,2}(\Omega)$ .

**A.2 Choice of the Quadrature Method.** We emphasize the importance of the choice of a specific quadrature formula using the following example:  $\int_{-1}^1 e^{-t} dt = e - 1/e$  where we compare a 0th-order finite elements methods with Gauss's method (the parameters of the Gauss quadrature formula are computed with a precision of  $10^{-16}$  using Newton's method):

Method	Value
Exact	2.350 402 387 287 603 ...
0th order (N = 1000)	2.351 945 ...
Finite element	
Gauss (N = 5)	2.350 402 386 46 ...

The Gauss method is far more powerful and allows us to compute bumps in 3D for an arbitrary number of populations.

**A.3 Shape Derivatives.** As it has already been pointed out, the computation of the variation of the bump with respect to the shape of the region  $\Omega$  is difficult since the set  $\mathcal{U}$  of regular domains (regular open bounded sets) of  $\mathbb{R}^q$  does not have the structure of a vector space. Variations of a domain must then be defined in some way. Let us consider a reference domain  $\Omega \in \mathcal{U}$  and the set  $\mathcal{A}$  of applications  $T : \Omega \rightarrow \mathbb{R}^q$ , which are at least as regular as homeomorphisms—one to one with  $T$  and  $T^{-1}$  one to one. In detail,

$$\mathcal{A} = \{T \text{ one to one, } T, T^{-1} \in W^{1,\infty}(\Omega, \mathbb{R}^q)\},$$

where the functional space  $W^{1,\infty}(\Omega, \mathbb{R}^q)$  is the set of mappings such that they and their first-order derivatives are in  $L^\infty(\Omega, \mathbb{R}^q)$ . In detail,

$$W^{1,\infty}(\Omega, \mathbb{R}^q) = \{T : \Omega \rightarrow \mathbb{R}^q \text{ such that } T \in L^\infty(\Omega, \mathbb{R}^q) \text{ and } \partial_i T \in L^\infty(\Omega, \mathbb{R}^q), i = 1, \dots, q\}$$

Given a shape function  $F : \mathcal{U} \rightarrow \mathbb{R}^q$ , for  $T \in \mathcal{A}$ , let us define  $\hat{F}(T) = F(T(\Omega))$ . The key point is that since  $W^{1,\infty}(\Omega, \mathbb{R}^q)$  is a Banach space, we can define the notion of a derivative with respect to the domain  $\Omega$  as:

**Definition 4.**  $F$  is Gâteaux differentiable with respect to  $\Omega$  if and only if  $\hat{F}$  is Gâteaux differentiable with respect to  $T$ .

In order to compute Gâteaux derivatives with respect to  $T$ , we introduce a family of deformations  $(T(\tau))_{\tau \geq 0}$  such that  $T(\tau) \in \mathcal{A}$  for  $\tau \geq 0$ ,  $T(0) = Id$ , and  $T(\cdot) \in C^1([0, A]; W^{1,\infty}(\Omega, \mathbb{R}^q))$ ,  $A > 0$ . From a practical point of view, there are many ways to construct such a family, the most famous one being the Hadamard deformation (Hadamard, 1968), which goes as follows.

For a point  $\mathbf{r} \in \Omega$ , we note

$$\begin{aligned} \mathbf{r}(\tau) &= T(\tau, \mathbf{r}) \quad \text{with} \quad T(0, \mathbf{r}) = \mathbf{r} \\ \Omega(\tau) &= T(\tau, \Omega) \quad \text{with} \quad T(0, \Omega) = \Omega. \end{aligned}$$

Let us now define the velocity vector field  $\mathbf{X}$  corresponding to  $T(\tau)$  as

$$\mathbf{X}(\mathbf{r}) = \frac{\partial T}{\partial \tau}(0, \mathbf{r}) \quad \forall \mathbf{r} \in \Omega.$$

From definition 4 follows:

**Definition 5.** The Gâteaux derivative of a shape function  $F(\Omega)$  in the direction of  $\mathbf{X}$ , denoted  $\langle F'(\Omega), \mathbf{X} \rangle$ , is equal to

$$\langle F'(\Omega), \mathbf{X} \rangle = \lim_{\tau \rightarrow 0} \frac{F(\Omega(\tau)) - F(\Omega)}{\tau}.$$

We also introduce:

**Definition 6.** The material derivative of a function  $f(\mathbf{r}, \Omega)$ , noted  $f_m(\mathbf{r}, \Omega, \mathbf{X})$  is defined by

$$f_m(\mathbf{r}, \Omega, \mathbf{X}) = \lim_{\tau \rightarrow 0} \frac{V(\mathbf{r}(\tau), \Omega(\tau)) - V(\mathbf{r}, \Omega)}{\tau},$$

and

**Definition 7.** The shape derivative of a function  $f(\mathbf{r}, \Omega)$ , noted  $f_s(\mathbf{r}, \Omega, \mathbf{X})$  is defined by

$$f_s(\mathbf{r}, \Omega, \mathbf{X}) = \lim_{\tau \rightarrow 0} \frac{f(\mathbf{r}, \Omega(\tau)) - f(\mathbf{r}, \Omega)}{\tau}.$$

The following theorem whose proof can be found, in Delfour and Zolésio (2001) and Sokolowski and Zolésio (1992) relates the Gâteaux derivative and the shape derivative:



**Theorem 7.** *The Gâteaux derivative of the functional  $F(\Omega) = \int_{\Omega} f(\mathbf{r}, \Omega) d\mathbf{r}$  in the direction of  $\mathbf{X}$  is given by*

$$\langle F'(\Omega), \mathbf{X} \rangle = \int_{\Omega} f_s(\mathbf{r}, \Omega, \mathbf{X}) d\mathbf{r} - \int_{\partial\Omega} f(\mathbf{r}, \Omega) \langle \mathbf{X}(\mathbf{r}), \mathbf{N}(\mathbf{r}) \rangle d\mathbf{a}(\mathbf{r}),$$

where  $\mathbf{N}$  is the unit inward normal to  $\partial\Omega$  and  $d\mathbf{a}$  its area element.

The following corollary is used in the proof of theorem 6:

**Corollary 2.** *The Gâteaux derivative of the functional  $F(\Omega) = \int_{\Omega} f(\mathbf{r}) d\mathbf{r}$  in the direction of  $\mathbf{X}$  is given by*

$$\langle F'(\Omega), \mathbf{X} \rangle = - \int_{\partial\Omega} f(\mathbf{r}) \langle \mathbf{X}(\mathbf{r}), \mathbf{N}(\mathbf{r}) \rangle d\mathbf{a}(\mathbf{r}),$$

where  $\mathbf{N}$  is the unit inward normal to  $\partial\Omega$  and  $d\mathbf{a}$  its area element.

## References

---

- Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2), 77–87.
- Appell, J., & Chen, C.-J. (2006). How to solve Hammerstein equations. *Journal of integral equations and applications*, 18(3), 287–296.
- Blomquist, P., Wyller, J., & Einevoll, G. (2005). Localized activity patterns in two-population neuronal networks. *Physica D*, 206, 180–212.
- Bressloff, P. (2005). Spontaneous symmetry breaking in self-organizing neural fields. *Biological Cybernetics*, 93(4), 256–274.
- Buxhoeveden, D., & Casanova, M. (2002). The minicolumn hypothesis in neuroscience. *Brain*, 125, 935–951.
- Chalupa, L. M., & Werner, J. (Eds.). (2004). *The visual neurosciences*. Cambridge, MA: MIT Press.
- Colby, C., Duhamel, J., & Goldberg, M. (1995). Oculocentric spatial representation in parietal cortex. *Cereb. Cortex*, 5, 470–481.
- Coombes, S. (2005). Waves, bumps, and patterns in neural fields theories. *Biological Cybernetics*, 93(2), 91–108.
- David, O., Cosmelli, D., & Friston, K. J. (2004). Evaluation of different measures of functional connectivity using a neural mass model. *NeuroImage*, 21, 659–673.
- David, O., & Friston, K. J. (2003). A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage*, 20, 1743–1755.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.
- Delfour, M., & Zolésio, J.-P. (2001). *Shapes and geometries: Advances in design and control*. Philadelphia: SIAM.
- Dieudonné, J. (1960). *Foundations of modern analysis*. Orlando, FL: Academic Press.

- Dobrovinski, K. (2005). *Dynamics, stability and bifurcation phenomena in the nonlocal model of cortical activity* (UUDM Project Report 2005:8). Uppsala: Uppsala University, Department of Mathematics.
- Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics*, *61*, 353–430.
- Evans, L. (1998). *Partial differential equations*. Providence, RI: American Mathematical Society.
- Faugeras, O., Grimbert, F., & Slotine, J.-J. (2007). *Stability and synchronization in neural fields* (Tech. Rep. RR-6212). Sophia-Antipolis, France: INRIA.
- Freeman, W. (1975). *Mass action in the nervous system*. Orlando, FL: Academic Press.
- Funahashi, S., Bruce, C., & Goldman-Rakic, P. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.*, *61*, 331–349.
- Gerstner, W., & Kistler, W. M. (2002). Mathematical formulations of Hebbian learning. *Biological Cybernetics*, *87*, 404–415.
- Grimbert, F., & Faugeras, O. (2006). Bifurcation analysis of Jansen's neural mass model. *Neural Computation*, *18*(12), 3052–3068.
- Guo, Y., & Chow, C. (2005a). Existence and stability of standing pulses in neural networks: II: Stability. *SIAM Journal on Applied Dynamical Systems*, *4*, 249–281.
- Guo, Y., & Chow, C. C. (2005b). Existence and stability of standing pulses in neural networks: I. Existence. *SIAM Journal on Applied Dynamical Systems*, *4*(2), 217–248.
- Gutkin, B., Ermentrout, G., & O'Sullivan, J. (2000). Layer 3 patchy recurrent excitatory connections may determine the spatial organization of sustained activity in the primate prefrontal cortex. *Neurocomputing*, *32–33*, 391–400.
- Hadamard, J. (1968). Mémoire sur un problème d'analyse relatif à l'équilibre des plaques élastiques encastrées. In *Mémoire des savants étrangers*. Paris: CNRS.
- Haeusler, S., & Maass, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuits models. *Cerebral Cortex*, *17*, 149–162.
- Hammerstein, A. (1930). Nichtlineare Integralgleichungen nebst anwendungen. *Acta Math.*, *54*, 117–176.
- Hazewinkel, M. (Ed.). (2001). *Encyclopaedia of Mathematics*. Berlin: Springer.
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, USA*, *81*(10), 3088–3092.
- Hoppenstaedt, F., & Izhikevich, E. (1997). *Weakly connected neural networks*. New York: Springer-Verlag.
- Jansen, B. H., & Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics*, *73*, 357–366.
- Jansen, B. H., Zouridakis, G., & Brandt, M. E. (1993). A neurophysiologically-based mathematical model of flash visual evoked potentials. *Biological Cybernetics*, *68*, 275–283.
- Kandel, E., Schwartz, J., & Jessel, T. (2000). *Principles of neural Science* (4th ed.). New York: McGraw-Hill.
- Kishimoto, K., & Amari, S. (1979). Existence and stability of local excitations in homogeneous neural fields. *Journal of Mathematical Biology*, *7*(4), 303–318.

- Krasnosel'skii, M., Vainikko, G., Zabreiko, P., & Stetsenko, V. (1972). *Approximate solutions of operator equations* (trans. D. Louvish). Gröningen: Wolters-Noordhoff.
- Kubota, S., & Aihara, K. (2005). Analyzing global dynamics of a neural field model. *Neural Processing Letters*, 21, 133–141.
- Laing, C., & Troy, W. (2003). Two-bump solutions of Amari-type models of neuronal pattern formation. *Physica D*, 178(3), 190–218.
- Laing, C., Troy, W., Gutkin, B., & Ermentrout, G. (2002). Multiple bumps in a neuronal model of working memory. *SIAM J. Appl. Math.*, 63(1), 62–97.
- Lopes da Silva, F., Blanes, W., Kalitzin, S., Parra, J., Suffczynski, P., & Velis, D. (2003). Dynamical diseases of brain systems: Different routes to epileptic seizures. *IEEE Transactions in Biomedical Engineering*, 50(5), 540–548.
- Lopes da Silva, F., Hoeks, A., & Zetterberg, L. (1974). Model of brain rhythmic activity. *Kybernetik*, 15, 27–37.
- Lopes da Silva, F., van Rotterdam, A., Barts, P., van Heusden, E., & Burr, W. (1976). Model of neuronal populations: The basic mechanism of rhythmicity. In M. A. Corner & D. F. Swaab (Eds.), *Progress in brain research* (pp. 281–308). Amsterdam: Elsevier.
- Matsuoka, K. (1992). Stability conditions for nonlinear continuous neural networks with asymmetric connection weights. *Neural Networks*, 5, 495–500.
- Miller, E., Erickson, C., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.*, 16, 5154–5167.
- Mountcastle, V. (1957). Modality and topographic properties of single neurons of cat's somatosensory cortex. *Journal of Neurophysiology*, 20, 408–434.
- Mountcastle, V. (1997). The columnar organization of the neocortex. *Brain*, 120, 701–722.
- Nunez, P. (2005). *Electric fields of the brain: The neurophysics of EEG*. New York: Oxford University Press.
- Owen, M., Laing, C., & Coombes, S. (2007). Bumps and rings in a two-dimensional neural field: Splitting and rotational instabilities. *New Journal of Physics*, 9, 378–401.
- Pinto, D., & Ermentrout, G. (2001a). Spatially structured activity in synaptically coupled neuronal networks: 1. Traveling fronts and pulses. *SIAM J. of Appl. Math.*, 62, 206–225.
- Pinto, D., & Ermentrout, G. (2001b). Spatially structured activity in synaptically coupled neuronal networks: 2. Standing pulses. *SIAM J. of Appl. Math.*, 62, 226–243.
- Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes in C*. Cambridge: Cambridge University Press.
- Rubin, J., & Troy, W. (2004). Sustained spatial patterns of activity in neuronal populations without recurrent excitation. *SIAM Journal on Applied Mathematics*, 64(5), 1609–1635.
- Sokolowski, J., & Zolésio, J.-P. (1992). *Introduction to shape optimization: Shape sensitivity analysis*. Berlin: Springer-Verlag.
- Stoer, J., & Bulirsch, R. (1972). *Introduction to numerical analysis*. Berlin: Springer-Verlag.
- Thomson, A. M., & Bannister, A. P. (2003). Interlaminar connections in the neocortex. *Cerebral Cortex*, 13, 5–14.
- Tricomi, F. (1985). *Integral equations*. New York: Dover.

- van Rotterdam, A., Lopes da Silva, F., van den Ende, J., Viergever, M., & Hermans, A. (1982). A model of the spatial-temporal characteristics of the alpha rhythm. *Bulletin of Mathematical Biology*, *44*(2), 283–305.
- Wendling, F., Bartolomei, F., Bellanger, J., & Chauvel, P. (2001). Interpretation of interdependencies in epileptic signals using a macroscopic physiological model of the EEG. *Clinical Neurophysiology*, *112*(7), 1201–1218.
- Wendling, F., Bellanger, J., Bartolomei, F., & Chauvel, P. (2000). Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals. *Biological Cybernetics*, *83*, 367–378.
- Werner, H., & Richter, T. (2001). Circular stationary solutions in two-dimensional neural fields. *Biological Cybernetics*, *85*(3), 211–217.
- Wilson, H., & Cowan, J. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics*, *13*(2), 55–80.

---

Received December 7, 2007; accepted April 26, 2008.

## A Master Equation Formalism for Macroscopic Modeling of Asynchronous Irregular Activity States

**Sami El Boustani**

*elboust@unic.cnrs-gif.fr*

**Alain Destexhe**

*destexhe@iaf.cnrs-gif.fr*

*Unité de Neurosciences Intégratives et Computationnelles, CNRS,  
91198 Gif-sur-Yvette, France*

Many efforts have been devoted to modeling asynchronous irregular (AI) activity states, which resemble the complex activity states seen in the cerebral cortex of awake animals. Most of models have considered balanced networks of excitatory and inhibitory spiking neurons in which AI states are sustained through recurrent sparse connectivity, with or without external input. In this letter we propose a mesoscopic description of such AI states. Using master equation formalism, we derive a second-order mean-field set of ordinary differential equations describing the temporal evolution of randomly connected balanced networks. This formalism takes into account finite size effects and is applicable to any neuron model as long as its transfer function can be characterized. We compare the predictions of this approach with numerical simulations for different network configurations and parameter spaces. Considering the randomly connected network as a unit, this approach could be used to build large-scale networks of such connected units, with an aim to model activity states constrained by macroscopic measurements, such as voltage-sensitive dye imaging.

### 1 Introduction ---

Cortical activity in awake animals manifests highly complex behavior, often characterized by seemingly noisy activity. At the level of single neurons, the activity in awake animals is associated with considerable subthreshold fluctuations of the membrane potential and irregular firing (Matsumara, Cope, & Fetz, 1988; Steriade, Timofeev, & Grenier, 2001; Destexhe, Rudolph, & Paré, 2003). It is during this regime that the main computational tasks are performed, and understanding those network dynamics is a crucial step toward an analytical study of information processing in neural networks (Destexhe & Contreras, 2006).

Much effort has been devoted to the study of how such activity emerges. Balanced networks have been introduced as a possible model to generate

dynamical states similar to the biological ones (see Figure 1). Two antagonistic states have been highlighted: synchronous regular states (SR) and asynchronous irregular states (AI) (Brunel, 2000). AI states are of particular interest because their dynamical characteristics are very similar to those observed in awake animals. For conductance-based integrate-and-fire neuron networks, they have even been observed without external stimulation (Vogels & Abbott, 2005; El Boustani, Pospischil, Rudolph-Lilith, & Destexhe, 2007; Kumar, Schrader, Aertsen, & Rotter, 2008). Large networks (over 10,000 neurons) are required to yield states consistent with experimental data (El Boustani et al., 2007; Kumar et al., 2008).

In parallel to these studies, population measures of neural activity have also been of great interest, in particular through the emergence of new imaging techniques such as voltage-sensitive dye imaging or two-photon imaging. Although the relation between these signals and single-cell properties is still not completely clear, these measurements reveal structures and correlations over large distances (millimeters or centimeters). No model currently is able to describe neuronal dynamics in large-scale networks at such distance scales, and there is a need for theoretical models specifically designed to handle the temporal and spatial scales of optical imaging.

The type of model that seems most appropriate for such scales are mean-field approaches. Self-consistent mean-field approaches have been proposed and gave predictions about the network stability in a stationary regime (Amit & Brunel, 1997; Brunel, 2000; Latham, Richmond, Nelson, & Nirenberg, 2000; Hertz, Lerchner, & Ahmadi, 2004). However, first-order mean-field approximation fails to fully describe these networks because of their inherent dynamics, which can rely dramatically on activity fluctuations. Moreover, the large network limit is usually performed for randomly connected networks despite the lack of biological relevance.

In this letter, our aim is to obtain a macroscopic description of distributed neuronal activity during AI states, where the unit is not the neuron but a small network of neurons. The difficulty, however, is to obtain a description that captures the statistics of network activity while being consistent with single-cell behavior. For this reason, we introduce a mesoscopic description of neuronal activity, in which finite size effects are explicitly taken into account. We consider networks of typical sizes of a few thousand neurons, far away from the large network limit.

To obtain such a mesoscopic model, we use a master equation formalism appropriate for a second-order mean-field description of network activity. The AI states, characterized by low firing rates and exponential decrease of the activity autocorrelation, can be incorporated in such a framework. A complete description of the correlations and covariances can be extracted for timescales governed by the network time constants. Correlations and covariances in the neural dynamics convey crucial information and can be responsible for radical changes in network state according to the parameter regime. This question has already been addressed with a similar formalism



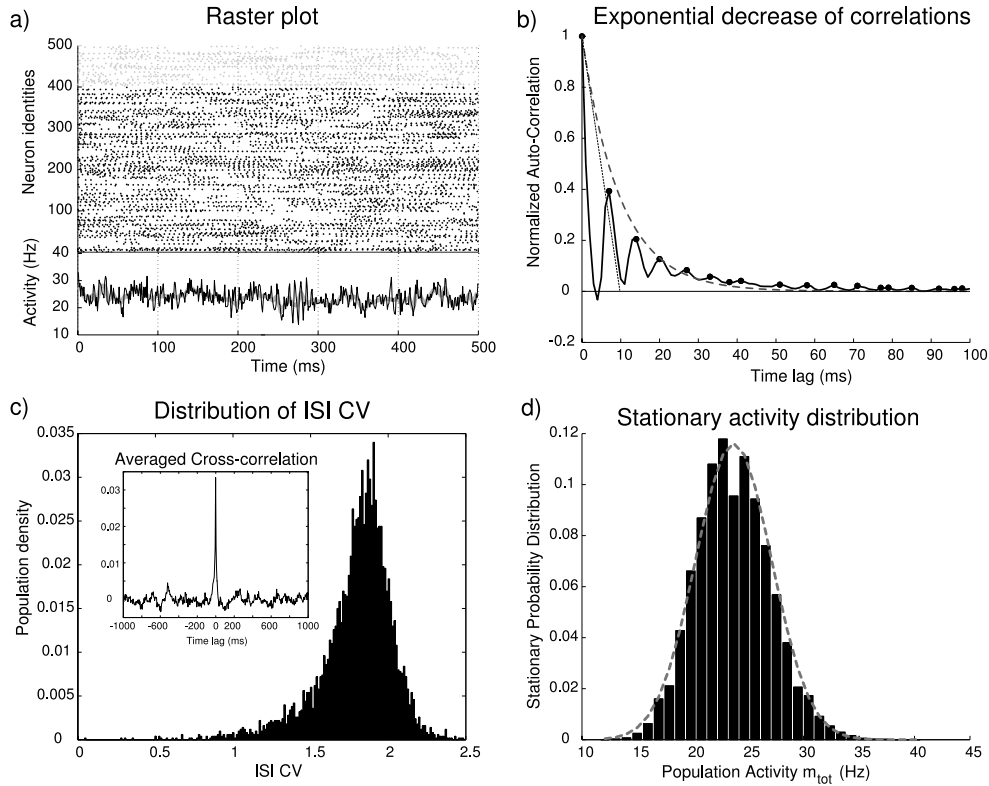


Figure 1: Example of a self-sustained asynchronous irregular (AI) state in a sparsely connected network of conductance-based neurons. The network contains 5000 neurons with a ratio of 4:1 between excitatory and inhibitory neurons and a connection probability  $p^{conn} = 0.01$ . Otherwise the neuron model is identical to Vogels and Abbott (2005) with  $\Delta g_{exc} = 7$  nS and  $\Delta g_{inh} = 100$  nS. (a) Top: Raster plot for a subset of excitatory (black dots) and inhibitory (gray dots) neurons. Bottom: Population activity with a bin size of 1 ms (black) and 5 ms (gray). The network has a mean activity of 23.68 Hz. (b) Autocorrelation of the total network activity. The dashed gray curve indicates the exponential envelope fit to the autocorrelation positive peaks. The dashed black curve indicates the slope at the origin. The decay time is equal to 10 ms, and the activity has been computed with a bin size of 1 ms. (c) Characterizing the asynchronous irregular state using statistical measures. The population distribution of the interspike interval coefficient of variation (CV) with a mean value of 1.7933, which is higher than 1 for the Poisson process. In the inset, the averaged pairwise cross-correlation computed with a bin size of 5 ms. The value at origin (0.034) estimates the network synchrony. This very low value indicates that there is no substantial synchrony among the neurons. (d) Stationary activity distribution of the network activity. The dashed gray curve indicates the gaussian fit made with the same mean value (23.68 Hz) and standard deviation (3.43 Hz). The activity has been computed with a time bin of 5 ms.

for binary neural networks (Ginzburg & Sompolinsky, 1994; Ohira & Cowan, 1993). Here we intend to develop a more general formalism, applicable to spiking neural networks (El Boustani & Destexhe, 2007).

A similar theory has been studied in parallel (Soula & Chow, 2007) with a discrete description of the network activity. Although a discrete description seems more natural for finite-size networks, several weaknesses appear when considering many populations, which is necessary for networks displaying spontaneously AI states. Indeed, the core of the theory is the transfer function, which maps the output firing rate of the neuron as a function of its synaptic input rates. Theoretical work has been done to obtain analytical transfer functions for a range of neuron models (Tuckwell, 1988; Brunel & Sergi, 1998; Brunel, 2000; Fourcaud & Brunel, 2002; Plesser & Gerstner, 2000). However, to take advantage of such results, we have to rely on a continuous description of network activity to link the neuron statistics to the network ones.

Those computations are not always possible, and a semianalytical approach has been used to tackle this problem (Kumar et al., 2008; Soula & Chow, 2007). In these studies, the neuron transfer function was determined numerically for every network state and used directly for a mean-field approach. This method becomes excessively time-consuming as soon as the network is slightly heterogeneous, which is the case when excitatory and inhibitory neurons have different intrinsic properties, for instance. Indeed, it is necessary to characterize numerically the transfer function for each population and conditionally to every population state. A continuous description is thus necessary to describe heterogeneous networks while benefiting from the theoretical work that has been made at the single-neuron level.

We consider different neuron models ranging from those for which a transfer function has been derived, to those for which an approximative model is required. For the latter, we suggest empirical models that can account for the network dynamics for a broad range of parameters. In particular, for conductance-based neurons, a transfer function can be found that gives a good qualitative description of different network regimes. Those transfer functions are detailed in the next sections and compared with numerical simulations for different model configurations. Finally, to gain some insight into the stability issues of self-sustained AI states and their lifetime, we use second-order statistics to discuss stability in different dynamical regimes, as well as discuss the limits of our approach. A preliminary version of this work has been published previously as a master's thesis (El Boustani, 2006) and a conference abstract (El Boustani & Destexhe, 2007).

## 2 Methods

---

All simulations were performed using the NEST simulator (<http://www.nest-initiative.org>) through the PyNN interface (<http://neuralensemble.org/PyNN>). We used different models of AI states in networks of spiking

neurons, based on previous work (Brunel, 2000; Mehring, Hehl, Kubo, Diesmann, & Aertsen, 2003; Vogels & Abbott, 2005). Network size will be 5000 neurons with a ratio of 1:4 between excitatory and inhibitory ( $\gamma = \frac{N_{inh}}{N_{tot}} = 0.2$ ), unless stated otherwise. For current-based models, an event-based strategy (Brette et al., 2007) was used to solve the network equations, and in the conductance-based model, a clock-based strategy was used with a time step of 0.1 ms, which is well beyond every time constant present in the system. In numerical and analytical models, we consider networks without any time delay in synaptic interactions. Based on previous work (Brunel & Hakim, 1999; Brunel, 2000; Vogels & Abbott, 2005), we expect this model to provide a broad range of AI states. In particular, the absence of interaction delay should decrease synchronous regions, as shown in Brunel (2000). Those states will be characterized using the usual statistical measures. To estimate the firing regularity, we will compute the mean interspike interval coefficient of variation. Neuron synchronization will be estimated by computing the mean pairwise cross-correlation among a set of 500 disjoint pairs and at time lag 0. This correlation coefficient is computed with a time bin of 5 ms, which gives a good estimation of synchrony among the neurons (Kumar et al., 2008). The measure is normalized so that it takes values between 0 (no synchrony) and 1 (complete synchrony). Those statistics quantities are illustrated in Figure 1c.

### 3 Results

---

We start by describing the general formalism in section 3.1, then consider different neuron models in sections 3.2 and 3.3, and end by illustrating the predictions of this formalism with numerical simulations in section 3.4. The symbol definitions can be found in Table 1 and the simulation parameters in Table 2.

**3.1 The Master Equation Formalism.** In this section, we develop the mathematical framework in which network dynamics will be described. The mean-field approach has been proven to be a powerful approach to describe networks of spiking neurons if pertinent approximations are done. In previous work, spontaneous activity and networks under stimulation have been studied with self-consistent equations for the neuron stationary mean firing rate (Amit & Brunel, 1997; Brunel, 2000). In these contexts, the individual neuron spike trains are considered as Poisson point processes, which allows one to take into account current fluctuations due to irregular firing. In particular, when a Fokker-Planck approach is used, the membrane potential distribution can be obtained and state diagrams drawn for the network activity. However, once the mean neuron firing rate has been determined, the self-consistent equations are studied without consideration to the population activity fluctuations and temporal dynamics. We propose a framework where a second-order description can be done at the

Table 1: Table of Symbols.

Symbol	Definition	Unit
<i>Network</i>		
$N_\mu$	Number of neuron in population $\mu$	-
$\gamma$	Inhibitory/excitatory neuron number ratio	-
$C_{\alpha\mu}$	Number of synaptic input from population $\alpha$ to $\mu$	-
$p_{\alpha\mu}^{conn}$	Proportion of synaptic connection from population $\alpha$ to $\mu$	-
$m_\mu$	Network activity	Hz
$m_\mu^{ext}$	External network activity	Hz
$T$	Network time constant	ms
<i>Neuron</i>		
$v_\mu$	Transfer function of a neuron in population $\mu$	Hz
$V_\mu$	Membrane potential	mV
$V_\mu^{rest}$	Resting membrane potential	mV
$V_\mu^{reset}$	Firing reset membrane potential	mV
$V_\mu^{threshold}$	Firing threshold potential	mV
$R_\mu$	Input resistance at rest	M $\Omega$
$G_\mu^L$	Leak conductance	nS
$C_\mu^{mem}$	Membrane capacitance	pF
$\tau_\mu^{mem}$	Resting membrane time constant	ms
$\tau_\mu^{ref}$	Firing refractory period	ms
<i>Synapse</i>		
$A_{\alpha\mu}$	Current synaptic strength from population $\alpha$ to $\mu$	pA
$J_{\alpha\mu}$	Voltage synaptic strength from population $\alpha$ to $\mu$	mV
$\Delta g_{\alpha\mu}$	Conductance synaptic strength from population $\alpha$ to $\mu$	nS
$E_\alpha$	Reversal potential for the $\alpha$ -type synapse	mV
$\tau_\alpha$	Synaptic time constant for the $\alpha$ -type synapse	ms

population level. In addition to using a stochastic process for the membrane potential, we invoke the master equation to obtain a stochastic process for the network activity using the corresponding neuron transfer function. In AI states, the activity autocorrelation decreases exponentially with the time lag (see Figure 1b), and our main hypothesis is that the network can be modeled as a Markov process for time steps in the order of this decay time.

*3.1.1 Main Hypothesis of the Phenomenological Model.* Cortical tissue is made of a great variety of neurons, which can be classified according to their biophysical properties. More precisely, electrophysiological measurements can be used to categorize neurons in their contribution to network dynamics. Although those properties exhibit a great diversity, for modeling, we can adopt a stereotypic description where only a few homogeneous populations are considered. In this letter, we consider only two populations, excitatory and inhibitory neurons, but the formalism can accommodate more classes if needed. To keep the model general, we assume that

Table 2: Network Configurations in Numerical Simulations.

Parameter	Value	Unit	
$\gamma$	0.2	-	
$T$	5 (except Figure 4)	ms	
$V_{\mu}^{rest}$	-60	mV	
$V_{\mu}^{reset}$	-60 (except Figure 5)	mV	
$V_{\mu}^{threshold}$	-50	mV	
$R_{\mu}$	100	M $\Omega$	
$\tau_{\mu}^{mem}$	20	ms	
$\tau_{\mu}^{ref}$	5	ms	
$E_{exc}$	0	mV	
$E_{inh}$	-80	mV	
Figure	$N$	$p^{conn}$	$(\tau_{exc}, \tau_{inh})$
2 and 4	5000	0.01	(1,3) ms
3a	15,000	0.01	(1,3) ms
3b and 3c	5000–25,000	0.01	(1,3) ms
1 and 5	5000	0.01	(5,10) ms
6a–6c	10,000	0.02	(5,10) ms
6d–6g	5000–10,000	0.01–0.02	(5,10) ms
7 and 10	10,000	0.01	(5,10) ms
11	10,000	0.02	(5,10) ms
12a and 12b	112,500	0.1	(0.3,0.3) ms
12d–12e	10,000	Variable	(5,10) ms

the network contains  $K$  homogeneous populations of neurons, denoted by  $1, 2, \dots, K$ . We define  $m_{\gamma}(t)$  as the network activity at time  $t$ ,

$$m_{\gamma}(t) = \lim_{\Delta t \rightarrow 0} \frac{n_{\gamma}(t - \Delta t, t)}{\Delta t N_{\gamma}}, \quad (3.1)$$

where  $n_{\gamma}(t - \Delta t, t)$  is the number of spikes emitted by population  $\gamma$  during the time interval  $[t - \Delta t, t]$  and  $N_{\gamma}$  is the population size for  $\gamma = 1, \dots, K$ . In order to build the master equation and obtain the desired Markovian description of network dynamics, we are interested in the conditional probability distribution for a short time interval  $T$ ,

$$P(\{m_{\gamma}(t)\} | \{m'_{\gamma}(t - T)\}), \quad (3.2)$$

with  $\gamma = 1, \dots, K$ . The system is assumed to be time invariant, so this probability depends only on the time constant  $T$ , and we write  $P_T(\{m_{\gamma}\} | \{m'_{\gamma}\})$ . This model is eventually intended to describe balanced network dynamics, which are made with sparse connectivity. For large enough networks, this property guarantees the existence of AI states in which pairwise correlations are negligible (van Vreeswijk & Sompolinsky, 1996; Brunel, 2000). We can

thus assume that the population-conditional probabilities are independent from each other beyond the timescale of  $T$ :

$$P_T(\{m_\gamma\} | \{m'_\gamma\}) = P_T(m_1 | \{m'_\gamma\}) \dots P_T(m_K | \{m'_\gamma\}). \quad (3.3)$$

Indeed, for a probabilistic system defined through a joint probability distribution, if the random variables are assumed to be independent, this distribution can be factorized in a product of marginal distributions describing each variable exclusively. As the network dynamics is assumed to be memoryless beyond the time interval  $T$ , we can then define a Markovian transition operator  $W(\{m_\gamma\} | \{m'_\gamma\})$  through the continuous master equation for population activities,

$$\begin{aligned} \partial_t P_t(\{m_\gamma\}) = & \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm'_\alpha (P_t(\{m'_\gamma\}) W(\{m_\gamma\} | \{m'_\gamma\}) \\ & - P_t(\{m_\gamma\}) W(\{m'_\gamma\} | \{m_\gamma\})), \end{aligned} \quad (3.4)$$

where  $m_\gamma \in [0, T^{-1}]$  and  $\partial_t P_t(\{m_\gamma\})$  is the time derivative of the probability distribution density. In this context, the population activity must be understood as the proportion of neurons that have fired at least once during the last period  $T$ , divided by the duration  $T$  and not as the instantaneous activity (see equation 3.1). The parameter  $T$  controls the temporal resolution of the model, and therefore the activity variables are bound by  $T^{-1}$ . However, if  $T^{-1}$  is larger than or equal to the neuron maximal firing rate, then we can avoid underestimation of the activity. Indeed, the fastest behavior in the network is defined by the neurons' maximal firing rate. In the SR regime, where excitation prevails over inhibition, the network firing rate is close to the neuron maximal firing rate. However, the time constant  $T$  cannot be taken as 0 because significant correlations at this scale would not be considered in this Markovian approach. We aim to describe network states with activity below 50 Hz, which can allow a quite large value for  $T$  without the risk of underestimation. An equivalent discrete formalism has been studied by Soula and Chow (2007) with a Markovian approach. Here we insist on keeping a continuous description characterized by the timescale  $T$  in order to obtain a framework entirely consistent with the Fokker-Planck approach at the neuron level. The crucial quantity, which must be small to allow this continuous description, is the activity resolution  $\Delta m = \frac{1}{NT}$ . It describes the minimal change in network activity due to a supplementary spike. It is thus necessary to find a good compromise between network size and time resolution to keep  $\Delta m$  small. The transition operator  $W(\{m_\gamma\} | \{m'_\gamma\})$  provides the rate of transition from state  $\{m'_\gamma\}$  to state  $\{m_\gamma\}$  giving the master equation its intuitive interpretation. It can be



defined using the conditional probability density, equation 3.3, by

$$\begin{aligned} W(\{m_\gamma\} | \{m'_\gamma\}) &= \lim_{T \rightarrow 0} \frac{P_T(\{m_\gamma\} | \{m'_\gamma\})}{T} \\ &= \lim_{T \rightarrow 0} \frac{\prod_{\alpha=1, \dots, K} P_T(m_\alpha | \{m'_\gamma\})}{T}. \end{aligned} \quad (3.5)$$

Therefore, we have to compute  $P_T(m_\alpha | \{m'_\gamma\})$  to fully specify the model. We will adopt a different definition that is more appropriate to our phenomenological model. We assume that the network follows a quasi-stationary evolution, which means that during time  $T$ , the system reaches a stationary state, determined by the previous state a time  $T$  earlier. This corresponds to the adiabatic hypothesis used in physics. Of course, this approximation no longer holds if the system is stimulated by a signal that possesses frequencies larger than  $T^{-1}$ , which can bring the system far from equilibrium. Therefore, and to avoid divergencies due to irrelevant high-order fluctuations, we define the transition function for finite  $T$ :

$$W(\{m_\gamma\} | \{m'_\gamma\}) = \frac{\prod_{\alpha=1, \dots, K} P_T(m_\alpha | \{m'_\gamma\})}{T}. \quad (3.6)$$

Similarly as mentioned before, equation 3.1 is reconsidered for finite  $T$  and is bounded by the maximal activity value. Hence, if  $T$  is small enough, the master equation formalism can be used. If we consider the limit  $T \rightarrow 0$ , the activity fluctuations become too important, and the transition function diverges. This parameter is equivalent to the bin size sampling in experiments, and it is well known that bin sizes that are too small give rise to irrelevant fluctuations in the population activity. Furthermore, for an infinitesimal bin size, those fluctuations become punctual, and the transition operator should be reduced to a two-dimensional matrix between the states where the network spikes or does not. To avoid this problem, we require that  $T$  has a finite value in the range of the network time constants. The Markovian approach is intended to describe the network dynamics responsible for the activity autocorrelation envelope (see Figure 1b). Correlations' fine structures in the AI regime are caused by residual global oscillations due to finite size effects. However, as the network size increases, the time constant  $T$  can take smaller values because the temporal finite-size effect in the autocorrelation vanishes because of sparse connectivity (Brunel, 2000). Thus, we can define a mesoscopic scale for sparsely connected networks in which the network is large enough to avoid temporal correlation finite-size effects but small enough to require second-order statistics to describe its population dynamics. However, the activity autocorrelation decreases exponentially on a timescale of the same order as the system time constants. A Markovian approach should not be acceptable below this timescale, and

we have to choose  $T$  carefully according to the regime under consideration. A  $T$  that is too large could underestimate the firing rate in high-activity regimes, and  $T$  values that are too small will overestimate the second-order statistics. This point has also been discussed in Soula and Chow (2007), but no mention has been made of the fact that the typical autocorrelation fine structure timescales of the network could be smaller than any time constant in the system (Gerstner, 2000), as shown in Figure 1b. Eventually, for a large enough network, we can consider  $T = v_{\max}^{-1}$ , where  $v_{\max}$  is the neuron maximum firing rate, for almost the whole range of the AI regime.

*3.1.2 Differential Equations for the Statistical Moments.* Using equation 3.4, we can obtain a hierarchy of first-order differential equations for the statistical moments. Indeed, the master equation solution provides the time evolution of the activity probability density. It generally cannot be solved exactly, but one can extract a hierarchy of equations for the statistical moments directly from the differential equation. For balanced networks, this set of equations can be stopped at the second order, and we avoid the closure problem. Indeed, the stationary activity distribution during AI states is well described by a gaussian function when the bin size is not too small (see Figure 1d). This can be understood thanks to the central limit theorem, because we are averaging out higher-order fluctuations during time steps  $T$  in the asynchronous activity. Of course, this can hold only if  $N$  or  $T$  is large enough.

If we close the master equation statistical moments hierarchy to the second order, we get (see appendix A)

$$\begin{aligned} \partial_t \langle m_\mu \rangle &= a_\mu(\{\langle m_\gamma \rangle\}) + \frac{1}{2} \partial_\lambda \partial_\eta a_\mu(\{\langle m_\gamma \rangle\}) c_{\lambda\eta} \\ \partial_t c_{\mu\nu} &= a_{\mu\nu}(\{\langle m_\gamma \rangle\}) + \partial_\lambda a_\mu(\{\langle m_\gamma \rangle\}) c_{\nu\lambda} + \partial_\lambda a_\nu(\{\langle m_\gamma \rangle\}) c_{\mu\lambda}, \end{aligned} \quad (3.7)$$

where  $\langle m_\mu \rangle$  is the mean population activity and  $c_{\mu\nu} = \langle (m_\mu - \langle m_\mu \rangle)(m_\nu - \langle m_\nu \rangle) \rangle$  is the activity covariance matrix. Here and in the following, we use Einstein index summation convention to avoid excessively awkward expressions. If an index is present in only one side of an equality, implicit summation over the whole range of value is understood.

The step moment functions  $a_\mu(\{\langle m_\gamma \rangle\})$  and  $a_{\mu\nu}(\{\langle m_\gamma \rangle\})$  are defined as follows:

$$\begin{aligned} a_\mu(\{\langle m_\gamma \rangle\}) &= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm'_\alpha (m'_\mu - \langle m_\mu \rangle) W(\{m'_\gamma\} | \{\langle m_\gamma \rangle\}) \\ &= \int_0^{1/T} dm'_\mu (m'_\mu - \langle m_\mu \rangle) \frac{P(m'_\mu | \{\langle m_\gamma \rangle\})}{T} \end{aligned}$$

$$\begin{aligned}
a_{\mu\nu}(\{\langle m_\gamma \rangle\}) &= \prod_{\alpha=1,\dots,K} \int_0^{1/T} dm'_\alpha(m'_\mu - \langle m_\mu \rangle) \\
&\quad \times (m'_\nu - \langle m_\nu \rangle) W(\{m'_\gamma\} | \{\langle m_\gamma \rangle\}) \\
&= \int_0^{1/T} dm'_\mu \int_0^{1/T} dm'_\nu (m'_\mu - \langle m_\mu \rangle) (m'_\nu - \langle m_\nu \rangle) \\
&\quad \times \frac{P(m'_\mu | \{\langle m_\gamma \rangle\}) P(m'_\nu | \{\langle m_\gamma \rangle\})}{T}. \tag{3.8}
\end{aligned}$$

The second and last identity has been obtained by using equation 3.5 given by the independence hypothesis. To complete the second-order development, we need to describe the correlation matrix of the network  $Corr_{\mu\nu}(t, t + \tau) = \langle (m_\mu(t) - \langle m_\mu(t) \rangle)(m_\nu(t + \tau) - \langle m_\nu(t + \tau) \rangle) \rangle$ . This is done in appendix B, and the resulting differential equation for a stationary state is given by

$$\partial_\tau Corr_{\mu\nu}(\tau) = \partial_\lambda a_\nu(\{\langle m_\gamma^{stat} \rangle\}) Corr_{\mu\lambda}(\tau), \tag{3.9}$$

where  $a_\nu(\{\langle m_\gamma \rangle\})$  is defined in equation 3.8 and  $\{\langle m_\gamma^{stat} \rangle\}$  is the stationary solution of the set in equation 3.7.

*3.1.3 A Model for the Transition Function.* Finally, to complete this framework, we need to specify the transition functions  $W(m_\alpha | \{m'_\gamma\})$  for  $\alpha = 1, \dots, K$ . These functions depend directly on the neuron properties. When a neuron is part of network dynamics and is exposed to intense synaptic inputs, it can be described using a transfer function that maps the neuron output firing rate to the synaptic inputs, regardless of their origin. In a recurrent neuron network, the synaptic input is provided by other neurons in the network and thus depends on the network activity. If we assume that we know the stationary transfer function of neurons in population  $\alpha$ ,  $v_\alpha$ , then the probability  $p_\alpha$  that a neuron in this population fires during time  $T$  given the previous state of the network  $\{m'_\gamma\}$  is

$$p_\alpha(\{m'_\gamma\}) \simeq v_\alpha(\{m'_\gamma\})T \leq 1.$$

This is a direct consequence of the adiabatic hypothesis we made in section 3.1.1. Indeed, the network in its previous state is characterized by a stationary transfer function so that the network activity is assumed to evolve near stationary states. During time  $T$ , each neuron can fire only once, so that we can express for the population  $\alpha$  the desired conditional probability  $P_T(m_\alpha | \{m'_\gamma\})$  with a binomial distribution using the independence

hypothesis,

$$P_T(m_\alpha | \{m'_\gamma\}) = \binom{N_\alpha}{m_\alpha N_\alpha T} p_\alpha(\{m'_\gamma\})^{m_\alpha N_\alpha T} (1 - p_\alpha(\{m'_\gamma\}))^{N_\alpha(1 - m_\alpha T)}, \quad (3.10)$$

where it is implicitly understood that we take the integer part of  $m_\alpha N_\alpha T$ . However, as population size  $N_\alpha$  increases, the correction due to the noninteger part becomes negligible, and we can apply the gaussian approximation by using the Stirling formula  $n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n}$ , leading to

$$P_T(m_\alpha | \{m'_\gamma\}) \simeq \sqrt{\frac{N_\alpha}{2\pi v_\alpha(\{m'_\gamma\})(1/T - v_\alpha(\{m'_\gamma\}))}} \times \exp \left[ -N_\alpha \frac{(m_\alpha - v_\alpha(\{m'_\gamma\}))^2}{2v_\alpha(\{m'_\gamma\})(1/T - v_\alpha(\{m'_\gamma\}))} \right]. \quad (3.11)$$

We see that this population conditional probability density follows a normal law with a variance that decreases when the neuron firing rate is near saturation  $v_\alpha = 1/T$  or in a quiescent state  $v_\alpha = 0$ . If symmetries are broken in the population  $\alpha$  and the neurons are not identical, this law is no longer valid. However, this law is a good first approximation because if the neuron firing rate in the population  $\alpha$  follows a particular distribution due to slight heterogeneities and  $N_\alpha$  are big enough, then the central limit theorem implies that the activity distribution is well described by a normal law with the same mean value. In this case, this mean value is the neuron transfer function in population  $\alpha$ , formally  $v_\alpha$ . Furthermore, the more  $N_\alpha$  increases, the narrower the normal distribution becomes. Thus, the theory describes large networks but also accounts for finite-size effects and is therefore close to a “mesoscopic” description. Indeed, if  $\{N_\alpha\}$  is taken to infinity, the normal law tends to a Dirac function in the sense of distribution theory. If  $T$  is kept constant while the network size is taken to infinity, the first-order mean field is recovered because fluctuations are completely averaged out and the second-order development is irrelevant. This phenomenological model stays in accordance with numerical models as long as  $T$  is kept finite. We finally get the transition function by dividing the conditional probability by  $T$  according to definition 3.6,

$$W(\{m_\gamma\} | \{m'_\gamma\}) = \frac{1}{T} \sqrt{\frac{\det(A)}{(2\pi)^K}} \exp \left[ -\frac{1}{2} (m_\mu - v_\mu(\{m'_\gamma\})) A_{\mu\nu} (m_\nu - v_\nu(\{m'_\gamma\})) \right], \quad (3.12)$$

where  $A_{\mu\nu} = \delta_{\mu\nu} \frac{N_\mu}{v_\mu(\langle m'_\gamma \rangle)(1/T - v_\mu(\langle m'_\gamma \rangle))}$ . The integral of this transition function over the entire state space gives trivially  $\frac{1}{T}$ , which can now be roughly interpreted as the degree of fluctuation taken into account. Indeed, if  $T \rightarrow \infty$ , the total integral goes to 0, which means that we are in the first-order mean-field approximation and the variance has no meaning, whereas if  $T \rightarrow 0$ , the total integral goes to infinity, which means that infinitesimal fluctuations are taken into account. These fluctuations are not relevant in our framework, and we keep  $T \sim v_{\max}^{-1}$ . If the interneuron correlations become substantial, the independence hypothesis no longer holds, and we have to consider another model to replace the binomial transition function. For large, sparsely connected networks, however, the independence hypothesis gives accurate results.

The whole formalism is completely described by the transition function, equation 3.12. The set of equations for the first- and second-order statistical moments can be further specified by injecting this transition function in the functions 3.8. We integrate over the complete real line because equation 3.12 is centered on  $0 < v_\alpha < T^{-1}$  with a variance that vanishes at the boundaries, so that the corrective terms are of order  $\mathcal{O}(e^{-N})$  and can be ignored for large enough populations. This is a natural consequence of the gaussian approximation. We then obtain

$$\begin{aligned} a_\mu(\{\langle m_\gamma \rangle\}) &= \frac{1}{T} \int_{-\infty}^{\infty} dm'_\mu (m'_\mu - \langle m_\mu \rangle) P(m'_\mu | \{\langle m_\gamma \rangle\}) \\ &= \frac{1}{T} (v_\mu - \langle m_\mu \rangle) \end{aligned} \quad (3.13)$$

$$\begin{aligned} a_{\mu\nu}(\{\langle m_\gamma \rangle\}) &= \int_{-\infty}^{\infty} dm'_\mu \int_{-\infty}^{\infty} dm'_\nu (m'_\mu - \langle m_\mu \rangle) \\ &\quad \times (m'_\nu - \langle m_\nu \rangle) W(\{\langle m'_\gamma \rangle\} | \{\langle m_\gamma \rangle\}) \\ &= \frac{\delta_{\mu\nu}}{T} \int_{-\infty}^{\infty} dm'_\mu (m_\mu'^2 - v_\mu^2 + v_\mu^2 - 2m'_\mu \langle m_\mu \rangle \\ &\quad + \langle m_\mu \rangle^2) P(m'_\mu | \{\langle m_\gamma \rangle\}) \\ &\quad + \frac{(1 - \delta_{\mu\nu})}{T} (v_\mu - \langle m_\mu \rangle)(v_\nu - \langle m_\nu \rangle) \\ &= \frac{1}{T} \left( \delta_{\mu\nu} \frac{v_\mu(1/T - v_\mu)}{N_\mu} + (v_\mu - \langle m_\mu \rangle)(v_\nu - \langle m_\nu \rangle) \right) \\ &= \frac{1}{T} (\delta_{\mu\nu} A_{\mu\mu}^{-1} + T^2 a_\mu a_\nu), \end{aligned} \quad (3.14)$$

where  $v_\mu = v_\mu(\{\langle m_\gamma \rangle\})$  is the transfer function of a neuron in the population  $\mu$ , which depends on the mean activity of every population. To get

every term in equation 3.7, we need to differentiate equation 3.13. The corresponding functions and derivatives give

$$\begin{aligned}
a_\mu(\{\langle m_\gamma \rangle\}) &= \frac{1}{T} (v_\mu - \langle m_\mu \rangle) \\
\partial_\lambda a_\mu(\{\langle m_\gamma \rangle\}) &= \frac{1}{T} (\partial_\lambda v_\mu - \delta_{\mu\lambda}) \\
\partial_\lambda \partial_\eta a_\mu(\{\langle m_\gamma \rangle\}) &= \frac{1}{T} \partial_\lambda \partial_\eta v_\mu.
\end{aligned} \tag{3.15}$$

Equations 3.7, 3.15, 3.9, and 3.14 provide a complete description of this master equation formalism, and we can write the set of differential equations according to those functions:

$$T \partial_t \langle m_\mu \rangle = (v_\mu - \langle m_\mu \rangle) + \frac{1}{2} \partial_\lambda \partial_\eta v_\mu c_{\lambda\eta} \tag{3.16}$$

$$\begin{aligned}
T \partial_t c_{\mu\nu} &= \delta_{\mu\nu} A_{\mu\mu}^{-1} + (v_\mu - \langle m_\mu \rangle)(v_\nu - \langle m_\nu \rangle) \\
&\quad + \partial_\lambda v_\mu c_{\nu\lambda} + \partial_\lambda v_\nu c_{\mu\lambda} - 2c_{\mu\nu}
\end{aligned}$$

$$T \partial_\tau \text{Corr}_{\mu\nu}(\tau) = (\partial_\lambda v_\nu (\{\langle m_\gamma^{stat} \rangle\}) - \delta_{\lambda\nu}) \text{Corr}_{\mu\lambda}(\tau) \tag{3.17}$$

We see that the determinant function for the first-order differential equation is the transfer function itself, which can be expected from the usual first-order analysis. The second-order statistics (covariances and correlations) are led mainly by the first derivative of the transfer function. We can expect higher-order statistical moments to depend on higher-order derivatives of this function in this framework. Thus, the set of transfer function  $\{v_\alpha\}$  for  $\alpha = 1, \dots, K$  plays a crucial role and should be determined to complete the model. In the next section, we consider different models.

**3.2 The Linear Model.** In this section, we illustrate some aspect of this master equation formalism, which could be harder to analyze once it is used for more realistic transfer functions. Following Soula and Chow (2007), we treat the simple case of a linear transfer function for a network of excitatory and inhibitory cells. Every neuron has the same transfer function (homogeneous intrinsic properties):

$$v(\{m_\gamma\}) = v_0 + k_{exc} m_{exc} + k_{inh} m_{inh}. \tag{3.18}$$

This problem is similar to a balanced network where all neurons receive the same synaptic input and have the same transfer function. In this case, the differential equations for the mean activities do not depend on the second-order moments. The set of second-order differential equations 3.16, can



thus be written with equation 3.15:

$$T \partial_t \langle m_\mu \rangle = v_0 + k_{exc} \langle m_{exc} \rangle + k_{inh} \langle m_{inh} \rangle - \langle m_\mu \rangle \quad (3.19)$$

$$\begin{aligned} T \partial_t c_{\mu\nu} = & \frac{\delta_{\mu\nu}}{N_\mu} v (T^{-1} - v) + (v - \langle m_\mu \rangle)(v - \langle m_\nu \rangle) \\ & + k_\lambda c_{\mu\lambda} + k_\lambda c_{\nu\lambda} - 2c_{\mu\nu}. \end{aligned} \quad (3.20)$$

This set of equations is linear, and we can look for the fixed point of the mean activity, equation 3.19, first. The solution of the linear system gives

$$m_0 = \langle m_{exc}^{FP} \rangle = \langle m_{inh}^{FP} \rangle = \frac{v_0}{1 - \Delta}, \quad (3.21)$$

where  $\Delta = k_{exc} + k_{inh}$  is the transfer function slope, which is modulated by excitatory and inhibitory strengths. The stability of this fixed point is given by the eigenvalues of the linear system, and we get the following values:

$$\lambda_1 = -1$$

$$\lambda_2 = \Delta - 1,$$

and we see as expected that the two populations converge to a common fixed point, whereas this fixed-point stability depends on the total slope  $\Delta$ , which must be inferior to unity. We can now study the remaining equations for the second-order moments 3.20. With the fixed-point 3.21, the system reduced to a simple set of three linear equations, which gives the fixed-point value for the covariance matrix. The solution is given by the following form:

$$\begin{aligned} \sigma^2(m_{exc})^{FP} &= \frac{m_0(1/T - m_0)}{2(\Delta - 2)(\Delta - 1)} \\ &\quad \times \left( \frac{(\Delta - 2)(1 - \Delta) + k_{exc}k_{inh}(k_{inh} - 1)}{(k_{exc} - 1)N_{exc}} + \frac{k_{inh}^2}{N_{inh}} \right) \\ \sigma^2(m_{inh})^{FP} &= \frac{m_0(1/T - m_0)}{2(\Delta - 2)(\Delta - 1)} \\ &\quad \times \left( \frac{k_{exc}^2}{N_{exc}} + \frac{(\Delta - 2)(1 - \Delta) + k_{inh}k_{exc}(k_{exc} - 1)}{(k_{inh} - 1)N_{inh}} \right) \\ c_{exc/inh}^{FP} &= -\frac{m_0(1/T - m_0)}{2(\Delta - 2)(\Delta - 1)} \left( \frac{k_{exc}(k_{inh} - 1)}{N_{exc}} + \frac{k_{inh}(k_{exc} - 1)}{N_{inh}} \right). \end{aligned} \quad (3.22)$$

Again we have to compute the eigenvalues of this linear system to study this fixed-point stability. We get

$$\lambda_3 = -2$$

$$\lambda_4 = 2(\Delta - 1)$$

$$\lambda_5 = \Delta - 2.$$

Those values guarantee the coherence in the covariance matrix. Nothing is learned about the network stability that was not already deduced from the mean activity set of equations. There is only a need for a balanced activity such that  $\Delta < 1$  to avoid the system's exploding. However, one has to remember that the fixed point is described with its fluctuations, so transitions between states are not excluded in a more complex system. We see from the set 3.22 that the covariance matrix entries decrease with the network size, and eventually we recover the first-order mean field when the networks are large enough, which is not the case with a description in terms of the number of spiking neurons. This relation is not exactly observed in small networks (Soula & Chow, 2007) due to a strong pairwise correlation, but if the network size is increased to biophysical scales, the relation is quite correct (Kumar et al., 2008). For finite-size networks, excitatory and inhibitory variances are different when sizes are different, even though the mean activity has exactly the same value in the stationary regime. Similar to Soula and Chow (2007), the network displays large fluctuations when it operates near the critical point  $\Delta = 1$ . Here again, we see that a description in terms of activity gives a unified description between neurons and networks where the limit  $N \rightarrow \infty$  is well defined. We give the correlation matrix (see equation B.5) to finish the example:

$$T \partial_\tau \text{Corr}_{\mu\nu}(\tau) = (\partial_\lambda v(\{m_\gamma^{FP}\}) - \delta_{\nu\lambda}) \text{Corr}_{\mu\lambda}(\tau) \quad (3.23)$$

$$= k_{exc} \text{Corr}_{\mu/exc}(\tau) + k_{inh} \text{Corr}_{\mu/inh}(\tau) - \text{Corr}_{\mu\nu}(\tau), \quad (3.24)$$

for which the eigenvalues are easily determined, and we find the same values as for the mean activity  $\lambda_1$  and  $\lambda_2$  with an algebraic multiplicity of 2. They describe the decrease of autocorrelation and cross-correlation. We see here again that  $\Delta = 1$  corresponds to a critical point where correlations are infinite spatially (between population activities) and temporally (within population activities).

### 3.3 Spiking Network Models.

*3.3.1 Network Structure.* Chaotic spontaneous activity as well as an asynchronous irregular regime have been observed in sparsely connected networks (Brunel, 2000; Brunel & Hakim, 1999; Kumar et al., 2008; van

Vreeswijk & Sompolinsky, 1996, 1998; Mehring et al., 2003; Vogels & Abbott, 2005). Sparse connections have been shown to be crucial to provide the network's irregular behavior. However, the connections do not need to be purely random (Mehring et al., 2003). A degree of locality in the connections can be tolerated as long as the correlations between neurons do not become critically strong and destroy the chaotic activity. We will consider sparsely connected networks with random connectivity. We will show in the numerical simulations that the model can still give good predictions when the neurons are locally sparsely connected. For the moment, every neuron from population  $\mu$  receives randomly  $C_{\alpha\mu}$  synaptic input from population  $\alpha$  where  $\frac{C_{\alpha\mu}}{N_\alpha} = p_{\alpha\mu}^{conn} < 1$ . Usually  $p_{\alpha\mu}^{conn}$  is taken between 1% and 10%.

**3.3.2 Current-Based Integrate-and-Fire Neurons.** We will consider current-based integrate-and-fire (IAF) neurons with the corresponding membrane potential equation,

$$\tau_\mu^{mem} \frac{d}{dt} V_{i_\mu}(t) = - (V_{i_\mu}(t) - V_\mu^{rest}) + R_\mu I_{i_\mu}(t) \mu \in [0, K] \quad \text{and} \quad i_\mu \in [0, N_\mu], \quad (3.25)$$

where  $V_\mu^{rest}$  is the resting potential and  $R_\mu$  and  $\tau_\mu^{mem}$  are, respectively, the membrane resistance and time constant of neurons in the population  $\mu$ . If the threshold  $V_\mu^{threshold}$  is crossed, the neuron emits a spike, and the membrane potential is clamped to the reset potential  $V_\mu^{reset}$  during a refractory period  $\tau_\mu^{ref}$ .  $I_{i_\mu}(t)$  is the external current coming from other neurons in the network  $I_{i_\mu}^{int}(t)$  or from an external source  $I_{i_\mu}^{ext}(t)$ . Because each population is homogeneous, we write  $\mu$  instead of  $i_\mu$  to simplify the notation.

To build the transition function, we have used a binomial law based on the independence approximation on timescale  $T$ . However, nothing was specified on the temporal structure of the spike trains emitted by each population during this timescale. If we assume the classical Poisson model for each neuron, the entire population can be modeled as a Poisson process too. The internal contribution is then represented as the convolution between a Poisson spike train and a postsynaptic potential function  $PSP_{\alpha\mu}(t)$  from population  $\alpha$  to population  $\mu$ ,

$$R_\mu I_\mu^{int}(t) = \sum_{\alpha=1, \dots, K} \int_{\mathbb{R}} PSP_{\alpha\mu}(t-s) N_{\alpha\mu}(ds), \quad (3.26)$$

where  $N_{\alpha\mu}(ds)$  are Poisson point processes describing the incoming spike trains. We discuss different synapse functions in the next section. We consider external stimulation currents also as Poisson spike trains with rate  $m_\alpha^{ext}$ .

One of the advantages of a description in term of continuous activity is that we can benefit from the Fokker-Planck approach to compute different transfer functions for our neurons. Although neurons can have different firing rates during the dynamics, we consider that, independently for every population, the law of large numbers prevails after time  $T$ , and only the mean firing rate of the whole process should be considered. Therefore, if we have  $C_{\alpha\mu}$  incoming spike trains from population  $\alpha$  to population  $\mu$  with a mean firing rate  $m_\alpha$ , then the total spike train can be considered a Poisson process of total rate  $C_{\alpha\mu}m_\alpha$ . Using the diffusion approximation, we can deduce the free (without spike mechanism) membrane potential distribution. This can be used as a first approximation to estimate the neuron transfer function (Amit & Brunel, 1997). Once the Fokker-Planck solution  $P_\mu(V)$  is found, we consider that the output firing rate is given by the distribution tail that lies above the threshold divided by the membrane time constant:

$$v_\mu = \frac{1}{\tau_\mu^{mem}} \int_{V_\mu^{threshold}}^{\infty} dVP(V) = \frac{1}{2\tau_\mu^{mem}} \left( 1 + \operatorname{erf} \left( \frac{\langle V_\mu \rangle - V_\mu^{threshold}}{\sqrt{2}\sigma(V_\mu)} \right) \right). \quad (3.27)$$

This approximation is valid as long as the membrane time constant leads the dynamics, more specifically in the AI regime. If the spike mechanism is included in the Fokker-Planck approach, an exact solution can be obtained for some synapse type for the IAF neuron transfer function. Brunel (2000) showed that for instantaneous synapses (Dirac functions), the solution is given by the inverse mean interspike interval of the first passage problem with white noise (Tuckwell, 1988). More recently (Fourcaud & Brunel, 2002), a correction has been added to take into account the colored noise produced by exponential synapses as long as the ratio  $\sqrt{\frac{\tau_{syn}}{\tau_{mem}}}$  is small compared to unity. The transfer function for this first-order correction is similar to the one used in Brunel (2000) but with a corrective term  $\Delta h_\mu \sim 1.03 \sqrt{\frac{\tau_{syn}}{\tau_{mem}}}$ ,

$$v_\mu = \left( \tau_\mu^{ref} + \tau_\mu^{mem} \sqrt{\pi} \int_{\frac{V_\mu^{reset} - \langle V_\mu \rangle}{\sqrt{2}\sigma(V_\mu)} + \Delta h_\mu}^{\frac{V_\mu^{threshold} - \langle V_\mu \rangle}{\sqrt{2}\sigma(V_\mu)} + \Delta h_\mu} du e^{u^2} (1 + \operatorname{erf}(u)) \right)^{-1}, \quad (3.28)$$

where the membrane potential statistics are given by the white noise model and the new synaptic weights must be normalized to ensure the correspondence between exponential synapses and Dirac synapses when the synaptic time constants are taken to 0. We have the following relation,  $J_\alpha^{Dirac} = J_\alpha^{Exp} \frac{\tau_\alpha}{\tau_\mu^{mem}}$ , where  $J_\alpha$  and  $\tau_\alpha$  are the synaptic strength and time constant corresponding to synapse  $\alpha$ . If we consider a balanced network made of excitatory and inhibitory neurons, the largest synaptic time constant will mainly be responsible for the corrective term. In all simulations presented

here, the inhibitory synaptic time constant will be at least twice as large as the excitatory time constant, so that the corrective term must be of the order  $\Delta h_\mu \sim 1.03 \sqrt{\frac{\tau_{inh}}{\tau_{mem}}}$ . However, in the following, we will consider network models where the synaptic time constant can be large compared to the membrane time constant (Vogels & Abbott, 2005), so we decided to adopt a different model for the transfer function. Instead of using the membrane potential mean value and variance computed with white noise, we will use shot noise processes to deduce those values for different synapses and replace  $\langle V_\mu \rangle$  and  $\sigma^2(V_\mu)$  in the transfer function, equation 3.28, without the corrective term:

$$v_\mu = \left( \tau_\mu^{ref} + \tau_\mu^{mem} \sqrt{\pi} \int_{\frac{V_\mu^{reset} - \langle V_\mu \rangle}{\sqrt{2\sigma(V_\mu)}}}^{\frac{V_\mu^{threshold} - \langle V_\mu \rangle}{\sqrt{2\sigma(V_\mu)}}} du e^{u^2} (1 + erf(u)) \right)^{-1}. \quad (3.29)$$

The mean membrane potential  $\langle V_\mu \rangle$  and the variance  $\sigma^2(V_\mu)$  depend on the incoming activity and the chosen synapse. This point is discussed in the next section, where different types of synapses are considered. As long as the model allows, we can also estimate the coefficient of variation of the interspike interval by using the recurrence relation developed for the first-passage problem (Tuckwell, 1988). For the model introduced in Brunel (2000), this quantity is given in a stationary point by

$$C V_\mu^2 = 2\pi \left( \tau_\mu^{mem} \langle m_\mu \rangle^{stat} \right)^2 \int_{\frac{V_\mu^{reset} - \langle V_\mu \rangle^{stat}}{\sigma(V_\mu)^{stat}}}^{\frac{V_\mu^{threshold} - \langle V_\mu \rangle^{stat}}{\sigma(V_\mu)^{stat}}} dx e^{x^2} \times \int_{-\infty}^x dy e^{y^2} (1 + erf(y))^2. \quad (3.30)$$

It is thus possible to access second-order statistics at the single-neuron level as well as the network level as long as the neuron model is specified. This is a powerful aspect of this theory, and it will play a crucial role in applying this framework to the study of voltage-sensitive dyes optical imaging data. Indeed, this type of signal is proportional to the subthreshold membrane potential, and we need a simple dynamic description that can give us access to the membrane potential distribution. This is done on timescales  $T$  with the network activity differential equation. Therefore, it is necessary to have a description at both levels. The formal equation for the interspike interval coefficient of variation (ISI CV) is not always determined, however, Kumar et al. (2008) noticed that the activity variance is very informative on the spiking regularity for self-sustained balanced networks. In some cases, the activity variance could describe the firing irregularity without the need for an exact equation for ISI CV. The limitation in the analytical derivation of the transfer function and the ISI CV depends on the chosen synapse or, equivalently, the nature of correlation in the current input. In the next section, we present different kind of synapses, some of which allow exact analytical results.

*Synapse models.* In the preceding section, we mentioned the Fokker-Planck approach. We need to determine the relation between the incoming spike train statistics and the membrane potential probability distribution. In particular, it is necessary to compute the mean and the variance according to the firing rate. Under the Poisson approximation, shot noise theory provides the required relations through Campbell's theorem,

$$\langle V_\mu \rangle = V_\mu^{rest} + \sum_{\alpha=1, \dots, K} C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \int_{\mathbb{R}} dt \text{PSP}_{\alpha\mu}(t) \quad (3.31)$$

$$\sigma^2(V_\mu) = \sum_{\alpha=1, \dots, K} C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \int_{\mathbb{R}} dt \text{PSP}_{\alpha\mu}^2(t), \quad (3.32)$$

where  $\text{PSP}_{\alpha\mu}(t)$  with  $\alpha, \mu \in \{1, \dots, K\}$  are the membrane potential time courses elicited by population  $\alpha$  synapses on population  $\mu$ . We will consider different synapse functions and compute these functions using equation 3.25, which is exactly solvable.

- *Dirac synapses (instantaneous).* Our first model is the Dirac synapse current,

$$\text{syn}_{\alpha\mu}(t) = A_{\alpha\mu} \tau_\mu^{mem} \delta(t),$$

with  $A_{\alpha\mu}$  the synaptic strength. Once integrated through equation 3.25, the membrane potential equation, this synapse gives the following postsynaptic potential,

$$\text{PSP}_{\alpha\mu}(t) = R_\mu A_{\alpha\mu} e^{-t/\tau_\mu^{mem}} H(t),$$

where  $H(t)$  is the Heaviside function. Using equations 3.31 and 3.32, we finally get the desired statistics characteristics,

$$\begin{aligned} \langle V_\mu \rangle &= V_\mu^{rest} + \tau_\mu^{mem} \sum_{\alpha=1, \dots, K} J_{\alpha\mu} C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \\ \sigma^2(V_\mu) &= \frac{1}{2} \sum_{\alpha=1, \dots, K} \tau_\mu^{mem} J_{\alpha\mu}^2 C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}), \end{aligned} \quad (3.33)$$

with  $J_{\alpha\mu} = R_\mu A_{\alpha\mu}$  the potential increment.

- *Exponential synapses.* A more realistic model would include a decaying tail to the synaptic current with a time constant specific to each synapse type. The next step in modeling realistic synapses is the exponential synapse model,

$$\begin{aligned} \text{syn}_{\alpha\mu}(t) &= A_{\alpha\mu} e^{-t/\tau_\alpha} H(t) \\ \text{PSP}_{\alpha\mu}(t) &= \frac{J_{\alpha\mu} \tau_\alpha}{\tau_\mu^{mem} - \tau_\alpha} (e^{-t/\tau_\mu^{mem}} - e^{-t/\tau_\alpha}) H(t), \end{aligned}$$



where  $\tau_\alpha$  is the decay time constant for synapses coming from the population  $\alpha$ , so that we finally get

$$\begin{aligned}\langle V_\mu \rangle &= V_\mu^{rest} + \sum_{\alpha=1, \dots, K} J_{\alpha\mu} \tau_\alpha C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \\ \sigma^2(V_\mu) &= \frac{1}{2} \sum_{\alpha=1, \dots, K} \frac{J_{\alpha\mu}^2 \tau_\alpha^2}{\tau_\alpha + \tau_\mu^{mem}} C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}).\end{aligned}\quad (3.34)$$

- $\alpha$ -synapses. To include a finite rising time of the synaptic current, we use alpha functions, which correspond to the following forms:

$$\begin{aligned}syn_{\alpha\mu}(t) &= A_{\alpha\mu} \frac{t}{\tau_\alpha} e^{1-t/\tau_\alpha} H(t) \\ PSP_{\alpha\mu}(t) &= J_{\alpha\mu} e \left( \frac{-te^{-t/\tau_\alpha}}{\tau_\mu^{mem} - \tau_\alpha} + \frac{\tau_\mu^{mem} \tau_\alpha}{(\tau_\mu^{mem} - \tau_\alpha)^2} (e^{-t/\tau_\mu^{mem}} - e^{-t/\tau_\alpha}) \right) H(t)\end{aligned}$$

and

$$\begin{aligned}\langle V_\mu \rangle &= V_\mu^{rest} + e \sum_{\alpha=1, \dots, K} J_{\alpha\mu} \tau_\alpha C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \\ \sigma^2(V_\mu) &= \sum_{\alpha=1, \dots, K} (2\tau_\mu^{mem} + \tau_\alpha) \left( \frac{e J_{\alpha\mu} \tau_\alpha}{2(\tau_\mu^{mem} + \tau_\alpha)} \right)^2 C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}).\end{aligned}\quad (3.35)$$

*Neuron transfer function in the master equation.* Now that we have specified the membrane potential mean value and variance, we have to incorporate these functions in the neuron transfer function to derive the network dynamics equations. We compute the necessary functions using the free membrane potential transfer function, as well as the phenomenological transfer function defined with the Campbell's theorem in the previous section. We thus use the definitions—equation 3.27 or 3.29 into 3.15—to get a complete description of the set of differential equations. We make some definitions to simplify the computation for equation 3.27,

$$\begin{aligned}Q_\mu &= \langle V_\mu \rangle - V_\mu^{threshold} \\ &= \Phi_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) + (V_\mu^{rest} - V_\mu^{threshold}),\end{aligned}\quad (3.36)$$

where  $\Phi_{\alpha\mu}$  depends on the chosen synapses,

$$\Phi_{\alpha\mu}^{Dirac} = \tau_\mu^{mem} J_{\alpha\mu} C_{\alpha\mu} \quad \Phi_{\alpha\mu}^{Exp} = \tau_\alpha J_{\alpha\mu} C_{\alpha\mu} \quad \Phi_{\alpha\mu}^{\alpha syn} = e \tau_\alpha J_{\alpha\mu} C_{\alpha\mu}, \quad (3.37)$$

and similarly

$$\begin{aligned}K_\mu &= 2\sigma^2(V_\mu) \\ &= \Psi_{\alpha\mu} (m_\alpha + m_\alpha^{ext}),\end{aligned}\quad (3.38)$$

with the corresponding synapse functions,

$$\begin{aligned}\Psi_{\alpha\mu}^{Dirac} &= \tau_{\mu}^{mem} J_{\alpha\mu}^2 C_{\alpha\mu} & \Psi_{\alpha\mu}^{Exp} &= \frac{J_{\alpha\mu}^2 \tau_{\alpha}^2}{\tau_{\alpha} + \tau_{\mu}^{mem}} C_{\alpha\mu} \\ \Psi_{\alpha\mu}^{asym} &= \frac{(2\tau_{\mu}^{mem} + \tau_{\alpha})}{2} \left( \frac{e J_{\alpha\mu} \tau_{\alpha}}{(\tau_{\mu}^{mem} + \tau_{\alpha})} \right)^2 C_{\alpha\mu},\end{aligned}\quad (3.39)$$

such that the transfer function, equation 3.27, can be written as

$$v_{\mu} = \frac{1}{2\tau_{\mu}^{mem}} \left( 1 + erf \left( \frac{Q_{\mu}}{\sqrt{K_{\mu}}} \right) \right). \quad (3.40)$$

To obtain the full set of differential equations, equation 3.16, we need to compute the step function derivatives, equation 3.15,

$$\partial_{\lambda} a_{\mu}(\{m_{\gamma}\}) = \frac{1}{T} \left( \frac{e^{-\frac{Q_{\mu}^2}{K_{\mu}}}}{\sqrt{\pi} \tau_{\mu}^{mem}} \left( \frac{2K_{\mu} \Phi_{\lambda\mu} - Q_{\mu} \Psi_{\lambda\mu}}{2K_{\mu}^{3/2}} \right) - \delta_{\lambda\mu} \right)$$

and for the second derivative,

$$\begin{aligned}\partial_{\lambda} \partial_{\eta} a_{\mu}(\{m_{\gamma}\}) &= \frac{e^{-\frac{Q_{\mu}^2}{K_{\mu}}}}{\sqrt{\pi} T \tau_{\mu}^{mem}} \\ &\cdot \left( \frac{2K_{\mu} (2Q_{\mu}^2 - K_{\mu}) (\Phi_{\lambda} \Psi_{\eta} + \Phi_{\eta} \Psi_{\lambda}) + Q_{\mu} (3K_{\mu} - 2Q_{\mu}^2) \Psi_{\lambda} \Psi_{\eta} - 8K_{\mu}^2 Q_{\mu} \Phi_{\lambda} \Phi_{\eta}}{4K_{\mu}^{7/2}} \right).\end{aligned}$$

If we consider the transfer function, equation 3.29, we have to define slightly different functions,

$$\begin{aligned}Q_{\mu}^{th} &= V_{\mu}^{threshold} - (V_{\mu}^{rest} + \Phi_{\alpha\mu}(m_{\alpha} + m_{\alpha}^{ext})) \\ Q_{\mu}^{re} &= V_{\mu}^{reset} - (V_{\mu}^{rest} + \Phi_{\alpha\mu}(m_{\alpha} + m_{\alpha}^{ext})) \\ K_{\mu} &= \Psi_{\alpha\mu}(m_{\alpha} + m_{\alpha}^{ext}),\end{aligned}\quad (3.41)$$

where  $\Phi_{\alpha\mu}$  and  $\Psi_{\gamma\mu}$  depend on the chosen synapses and are defined as previously. We also define

$$\begin{aligned}x_{\mu}^{th} &= \frac{Q_{\mu}^{th}}{\sqrt{K_{\mu}}} \\ x_{\mu}^{re} &= \frac{Q_{\mu}^{re}}{\sqrt{K_{\mu}}}\end{aligned}$$

such that equation 3.29 can be written in a shorter form:

$$v_\mu = \left( \tau_\mu^{ref} + \tau_\mu^{mem} \sqrt{\pi} \int_{x_\mu^{re}}^{x_\mu^{th}} du e^{u^2} (1 + \operatorname{erf}(u)) \right)^{-1}.$$

As before, we compute the step function derivatives, equation 3.15,

$$\begin{aligned} \partial_\lambda a_\mu(\{m_\gamma\}) &= -\frac{1}{T} (v_\mu^2 \tau_\mu^{mem} \sqrt{\pi} (e^{(x_\mu^{th})^2} (1 + \operatorname{erf}(x_\mu^{th})) \partial_\lambda x_\mu^{th} \\ &\quad - e^{(x_\mu^{re})^2} (1 + \operatorname{erf}(x_\mu^{re})) \partial_\lambda x_\mu^{re}) + \delta_{\lambda\mu}) \\ \partial_\lambda \partial_\eta a_\mu(\{m_\gamma\}) &= \frac{2\partial_\lambda v_\mu \partial_\eta v_\mu}{T v_\mu} - \frac{v_\mu^2 \tau_\mu^{mem} \sqrt{\pi}}{T} \\ &\quad \times \left( \left( 2x_\mu^{th} e^{(x_\mu^{th})^2} (1 + \operatorname{erf}(x_\mu^{th})) + \frac{2}{\sqrt{\pi}} \right) \partial_\lambda x_\mu^{th} \partial_\eta x_\mu^{th} \right. \\ &\quad + e^{(x_\mu^{th})^2} (1 + \operatorname{erf}(x_\mu^{th})) \partial_\lambda \partial_\eta x_\mu^{th} \\ &\quad - \left. \left( 2x_\mu^{re} e^{(x_\mu^{re})^2} (1 + \operatorname{erf}(x_\mu^{re})) + \frac{2}{\sqrt{\pi}} \right) \partial_\lambda x_\mu^{re} \partial_\eta x_\mu^{re} \right. \\ &\quad \left. - e^{(x_\mu^{re})^2} (1 + \operatorname{erf}(x_\mu^{re})) \partial_\lambda \partial_\eta x_\mu^{re} \right) \end{aligned}$$

with

$$\begin{aligned} \partial_\lambda x_\mu^{th/re} &= \frac{-Q_\mu^{th/re} \Psi_{\lambda\mu} - 4K_\mu \Phi_{\lambda\mu}}{4K_\mu^{3/2}} \\ \partial_\lambda \partial_\eta x_\mu^{th/re} &= \frac{4K_\mu (\Phi_{\lambda\mu} \Psi_{\eta\mu} + \Phi_{\eta\mu} \Psi_{\lambda\mu}) + 3Q_\mu^{th/re} \Psi_{\lambda\mu} \Psi_{\eta\mu}}{16K_\mu^{5/2}}. \end{aligned}$$

From now on, the current-based model can be used to perform a parameter space exploration. The stationary solutions can be numerically computed for any set of parameters and compared with network simulations. This is done in section 3.4.

**3.3.3 Conductance-Based Models.** For the conductance-based model, every synaptic event results in an increase in the corresponding conductance, and the integrate-and-fire equation can be written as

$$C_\mu^{mem} \frac{d}{dt} V_\mu(t) = G_\mu^L (V_\mu^{rest} - V_\mu(t)) + G_{\alpha\mu}(t) (E_\alpha - V_\mu(t)), \quad (3.42)$$

where  $C_\mu^{mem}$  and  $G_\mu^L$  are the membrane capacitance and leak conductance, respectively, such that  $\frac{C_\mu^{mem}}{G_\mu^L} = \tau_\mu^{mem}$  is the resting membrane time constant.  $G_{\alpha\mu}$  is the total conductance of the synaptic set  $\alpha$  and  $E_\alpha$  the corresponding reversal potential. Similar to the current-based model, the synaptic input can be modeled by Poisson processes, and the total conductance of the synapses  $\alpha$  is

$$G_{\alpha\mu}(t) = \int_{\mathbb{R}} g_{\alpha\mu}(t-s) N_{\alpha\mu}(ds), \quad (3.43)$$

where  $g_{\alpha\mu}(t)$  is the conductance time course elicited by an incoming spike from population  $\alpha$ . We can write equation 3.42 in an analogous form to the current-based model with an effective membrane time constant  $\tau_\mu^{eff}(t)$ ,

$$\tau_\mu^{eff}(t) \frac{d}{dt} V_\mu(t) = -V_\mu(t) + \frac{G_\mu^L V_\mu^{rest} + G_{\alpha\mu}(t) E_\alpha}{G_\mu^{tot}(t)}, \quad (3.44)$$

where the total conductance and the effective time constant are defined as follows:

$$G_\mu^{tot}(t) = G_\mu^L + \sum_{\alpha=1, \dots, K} G_{\alpha\mu}$$

$$\tau_\mu^{eff}(t) = \frac{C_\mu^{mem}}{G_\mu^{tot}(t)}.$$

Even when the synaptic input is considered as white noise, equation 3.42 cannot be solved, and there is no exact solution for the transfer function. We can use equation 3.27 as a first approximation, but we need to compute the mean membrane potential  $\langle V_\mu \rangle$  and the variance  $\sigma^2(V_\mu)$  in the conductance-based model. Equation 3.44 can be approximated by the following effective current-based equation (Kuhn, Aertsen, & Rotter, 2004),

$$\langle \tau_\mu^{eff} \rangle \frac{d}{dt} V_\mu(t) = -V_\mu(t) + \frac{G_\mu^L V_\mu^{rest} + \langle G_{\alpha\mu} \rangle E_\alpha}{\langle G_\mu^{tot} \rangle}, \quad (3.45)$$

with

$$\langle G_{\alpha\mu} \rangle = C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \int_{\mathbb{R}} ds g_{\alpha\mu}(ds)$$

$$\langle G_\mu^{tot} \rangle = G_\mu^L + \sum_{\alpha=1, \dots, K} \langle G_{\alpha\mu} \rangle$$

$$\langle \tau_\mu^{eff} \rangle = \frac{C_\mu^{mem}}{\langle G_\mu^{tot} \rangle}$$

*Synapse model.* Following Kuhn et al. (2004), we can deduce from equation 3.45 a good approximation for the mean membrane potential and the variance. The mean membrane potential is given by the following form,

$$\langle V_\mu \rangle = \langle \tau_\mu^{eff} \rangle \left( \frac{V_\mu^{rest}}{\tau_\mu^{mem}} + \frac{E_\alpha}{C_\mu^{mem}} \langle G_{\alpha\mu} \rangle \right) \quad (3.46)$$

$$= \langle \tau_\mu^{eff} \rangle \left( \frac{V_\mu^{rest}}{\tau_\mu^{mem}} + \Phi_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \right), \quad (3.47)$$

where

$$\Phi_{\alpha\mu} = C_{\alpha\mu} \frac{E_\alpha}{C_\mu^{mem}} \int_{\mathbb{R}} ds g_{\alpha\mu}(s)$$

depends on the chosen synapse. Using this mean membrane potential, equation 3.45 can be integrated for a synaptic event while the neuron is clamped around the mean value  $\langle V_\mu \rangle$ . The corresponding  $PSP_{\alpha\mu}(t)$  allows us to compute the variance:

$$\sigma^2(V_\mu) = C_{\alpha\mu} (m_\alpha + m_\alpha^{ext}) \int_{\mathbb{R}} ds PSP_{\alpha\mu}^2(s) \quad (3.48)$$

$$= \frac{\Psi_{\alpha\mu}}{2} (m_\alpha + m_\alpha^{ext}). \quad (3.49)$$

We have to specify the form of  $\Phi_{\alpha\mu}$  and  $\Psi_{\alpha\mu}$  for different synapses. For the conductance-based model, we consider only exponential synapses and  $\alpha$ -synapses. Following the same computation as in the current-based model, we can determine the desired functions,

$$\Phi_{\alpha\mu}^{Exp} = \frac{\tau_\alpha \Delta g_{\alpha\mu} C_{\alpha\mu} E_\alpha}{C_\mu^{mem}} \quad \Phi_{\alpha\mu}^{\alpha syn} = \frac{e \tau_\alpha \Delta g_{\alpha\mu} C_{\alpha\mu} E_\alpha}{C_\mu^{mem}} \quad (3.50)$$

and

$$\Psi_{\alpha\mu}^{Exp} = \frac{C_{\alpha\mu}}{\tau_\alpha + \langle \tau_\mu^{eff} \rangle} \left( \frac{\tau_\alpha \Delta g_{\alpha\mu} \langle \tau_\mu^{eff} \rangle (E_\alpha - \langle V_\mu \rangle)}{C_\mu^{mem}} \right)^2 \quad (3.51)$$

$$\Psi_{\alpha\mu}^{\alpha syn} = \frac{1}{2} C_{\alpha\mu} (2 \langle \tau_\mu^{eff} \rangle + \tau_\alpha) \left( \frac{e \tau_\alpha \Delta g_{\alpha\mu} \langle \tau_\mu^{eff} \rangle (E_\alpha - \langle V_\mu \rangle)}{C_\mu^{mem} (\tau_\alpha + \langle \tau_\mu^{eff} \rangle)} \right)^2, \quad (3.52)$$

where  $\Delta g_{\alpha\mu}$  is the synaptic strength.

*Neuron transfer function in the master equation.* The main difference with the current-based model is that the effective time constant is now activity dependent, and  $\Psi_{\alpha\mu}$  depends on the mean membrane potential. We consider

that the leading time constant in the neuron dynamic is the effective time constant whatever the regime. The corresponding approximated transfer function can be written as

$$v_\mu = \frac{1}{2\langle\tau_\mu^{eff}\rangle} \left( 1 + \operatorname{erf} \left( \frac{\langle V_\mu \rangle - V_\mu^{threshold}}{\sqrt{2}\sigma(V_\mu)} \right) \right). \quad (3.53)$$

The computation is thus more complicated but still straightforward. We first define as previously the functions

$$Q_\mu = \langle V_\mu \rangle - V_\mu^{threshold} \quad (3.54)$$

$$K_\mu = 2\sigma^2(V_\mu) \quad (3.55)$$

$$x_\mu = \frac{Q_\mu}{\sqrt{K_\mu}}. \quad (3.56)$$

Because those functions now depend on  $m_\alpha$  in a more intricate way, the step function derivatives, equation 3.15, must be written as follows:

$$\begin{aligned} \partial_\lambda a_\mu(\{m_\gamma\}) = & \frac{1}{T} \left( \frac{e^{-x_\mu^2}}{\sqrt{\pi}\langle\tau_\mu^{eff}\rangle} \left( \frac{2K_\mu\partial_\lambda Q_\mu - Q_\mu\partial_\lambda K_\mu}{2K_\mu^{3/2}} \right) \right. \\ & \left. - v_\mu \frac{\partial_\lambda \langle\tau_\mu^{eff}\rangle}{\langle\tau_\mu^{eff}\rangle} - \delta_{\lambda\mu} \right), \end{aligned} \quad (3.57)$$

and for the second derivative,

$$\begin{aligned} & \partial_\lambda \partial_\eta a_\mu(\{m_\gamma\}) \\ &= \frac{e^{-x_\mu^2}}{\sqrt{\pi}T\langle\tau_\mu^{eff}\rangle} \left( \frac{1}{4K_\mu^{7/2}} \left( 2K_\mu(2Q_\mu^2 - K_\mu)(\partial_\lambda Q_\mu \partial_\eta K_\mu + \partial_\eta Q_\mu \partial_\lambda K_\mu) \right. \right. \\ & \quad + 4K_\mu^3 \partial_\eta \partial_\lambda Q_\mu - 8K_\mu^2 Q_\mu \partial_\lambda Q_\mu \partial_\eta Q_\mu \\ & \quad \left. \left. + Q_\mu(3K_\mu - 2Q_\mu^2) \partial_\lambda K_\mu \partial_\eta K_\mu - 2K_\mu^2 Q_\mu \partial_\eta \partial_\lambda K_\mu \right) \right. \\ & \quad - \left( \frac{2K_\mu \partial_\lambda Q_\mu - Q_\mu \partial_\lambda K_\mu}{2K_\mu^{3/2}} \right) \frac{\partial_\eta \langle\tau_\mu^{eff}\rangle}{\langle\tau_\mu^{eff}\rangle} \\ & \quad - \left( \frac{2K_\mu \partial_\eta Q_\mu - Q_\mu \partial_\eta K_\mu}{2K_\mu^{3/2}} \right) \frac{\partial_\lambda \langle\tau_\mu^{eff}\rangle}{\langle\tau_\mu^{eff}\rangle} \\ & \quad \left. + 2 \frac{\partial_\lambda \langle\tau_\mu^{eff}\rangle \partial_\eta \langle\tau_\mu^{eff}\rangle}{T\langle\tau_\mu^{eff}\rangle^2} - \frac{\partial_\lambda \partial_\eta \langle\tau_\mu^{eff}\rangle}{T\langle\tau_\mu^{eff}\rangle} \right), \end{aligned}$$



with

$$\begin{aligned}
Q_\mu &= \langle \tau_\mu^{eff} \rangle \left( \frac{V_\mu^{rest}}{\tau_\mu^{mem}} + \Phi_{\alpha\mu}(m_\alpha + m_\alpha^{ext}) \right) - V_\mu^{threshold} \\
\partial_\lambda Q_\mu &= \partial_\lambda \langle \tau_\mu^{eff} \rangle \left( \frac{V_\mu^{rest}}{\tau_\mu^{mem}} + \Phi_{\alpha\mu}(m_\alpha + m_\alpha^{ext}) \right) + \langle \tau_\mu^{eff} \rangle \Phi_{\lambda\mu} \\
\partial_\eta \partial_\lambda Q_\mu &= \partial_\eta \partial_\lambda \langle \tau_\mu^{eff} \rangle \left( \frac{V_\mu^{rest}}{\tau_\mu^{mem}} + \Phi_{\alpha\mu}(m_\alpha + m_\alpha^{ext}) \right) + \partial_\eta \langle \tau_\mu^{eff} \rangle \Phi_{\lambda\mu} \\
&\quad + \partial_\lambda \langle \tau_\mu^{eff} \rangle \Phi_{\eta\mu}
\end{aligned}$$

and

$$\begin{aligned}
K_\mu &= \Psi_{\alpha\mu}(m_\alpha + m_\alpha^{ext}) \\
\partial_\lambda K_\mu &= \partial_\lambda \Psi_{\alpha\mu}(m_\alpha + m_\alpha^{ext}) + \Psi_{\lambda\mu} \\
\partial_\eta \partial_\lambda K_\mu &= \partial_\lambda \partial_\eta \Psi_{\alpha\mu}(m_\alpha + m_\alpha^{ext}) + \partial_\eta \Psi_{\lambda\mu} + \partial_\lambda \Psi_{\eta\mu}.
\end{aligned}$$

The derivatives of functions  $\langle \tau_\mu^{eff} \rangle$  and  $\Psi_{\alpha\mu}$  are given in appendix C. The conductance-based model is now completely specified and can be used for a parameter space exploration as well.

**3.4 Numerical Results.** Simulations were done to compare the predicted stationary states given by the master equation and the corresponding neuron networks. In section 3.4.1, we explore different parameter spaces for the current-based neuron model, and the conductance-based neuron network is studied in section 3.4.2. Section 3.4.3 is devoted to more general models where a more realistic connectivity scheme is used.

*3.4.1 Current-Based Models.* We consider here two types of state diagrams based on the literature. The three concerned parameters are the external excitatory firing rate stimulation  $m_{exc}^{ext}$  and the excitatory/inhibitory synaptic strength  $A_\mu$  with  $\mu \in \{exc, inh\}$ . With those parameters, we were interested in the  $(m_{exc}^{ext}, g)$  space where  $g = \frac{A_{inh}}{A_{exc}}$ , as in Brunel (2000) and Mehring et al. (2003). The second type of state diagram is generated by varying independently the excitatory and inhibitory synaptic strength ( $A_{exc}, A_{inh}$ ) by feeding the network with a constant current in every neuron to sustain the activity (Vogels & Abbott, 2005). For each simulation, we want to investigate whether the model can provide a good prediction for the first and second statistical moments. Therefore, we systematically compare the mean activity and the standard deviation. First-order predictions have already been studied for similar network models (Brunel, 2000), and the corrective term

due to the second-order development does not contribute much. However, because we suggest using another neuron transfer function in the following, it is still interesting to perform the first-order analysis. For those simulations,  $N = 5000$  neurons with a probability connection  $p^{conn} = 0.01$ . Neurons' intrinsic properties are homogeneous, with a membrane time constant  $\tau_{\mu}^{mem} = 20$  ms, a refractory period  $\tau_{\mu}^{ref} = 5$  ms, the resting potential and the reset potential  $V_{\mu}^{rest} = V_{\mu}^{reset} = -60$  mV, the threshold potential  $V_{\mu}^{threshold} = -50$  mV, and the membrane resistance  $R_{\mu}^{mem} = 100$  M $\Omega$  for  $\mu \in \{exc, inh\}$ . Synaptic properties and external stimulations depend on the parameter space under consideration.

*The  $(m_{exc}^{ext}, g)$  parameter space.* For this state diagram we took exponential synapses with  $\tau_{exc} = 1$  ms and  $\tau_{inh} = 3$  ms for the excitatory and inhibitory synaptic time constants. The external stimulation will be considered in the  $\nu_{th}$  unit which is the frequency needed to bring the mean membrane potential to threshold with an excitatory Poisson input. As we are using exponential synapses we have, based on equation 3.34,

$$\nu_{th} = \frac{V^{threshold} - V^{rest}}{J_{exc} C_{exc} \tau_{exc}},$$

where  $J_{exc} = R^{mem} A_{exc}$ , we have omitted the second index because of network homogeneity. We chose  $A_{exc} = 0.02$  pA such that the resulting EPSP peak is  $J_{exc} = 2$  mV. The inhibitory synaptic strength is defined by  $A_{inh} = g A_{exc}$  where  $g$  is the second parameter. We represent in Figure 2a the network mean cross-correlation and interspike interval to characterize the asynchronous irregular states.

Three regions can be outlined: a broad AI domain, an intermediate asynchronous regular (AR) region where neurons begin to fire periodically, and saturated synchronous regular (SR) states. We were interested in the AI regime in which the Markovian approximation is assumed to apply. Therefore, we limited our analysis of the parameter regime to firing rates below  $1/\tau^{mem} = 50$  Hz. Above this frequency, long-range correlations in each neuron firing appear, and the analytical framework is not well suited to predict the macroscopic quantities. This is manifest by the emergence of a new peak near 0 in the ISI CV distribution (not shown), whereas interneuron correlations reach very small values. The absence of synchronous irregular states (SI) for high firing rates is essentially due to the absence of interaction delays between the neurons (Brunel, 2000). When a finite homogeneous delay is present, SI states could occur in the transitory region (Mehring et al., 2003); however, if the delays are drawn randomly from a broad distribution, this SI transition region is replaced by AR states. In the low-activity region, there is an increase of synchrony concomitant with a decrease in the mean ISI CV typical of slow SI states. For the whole region, as long as the network

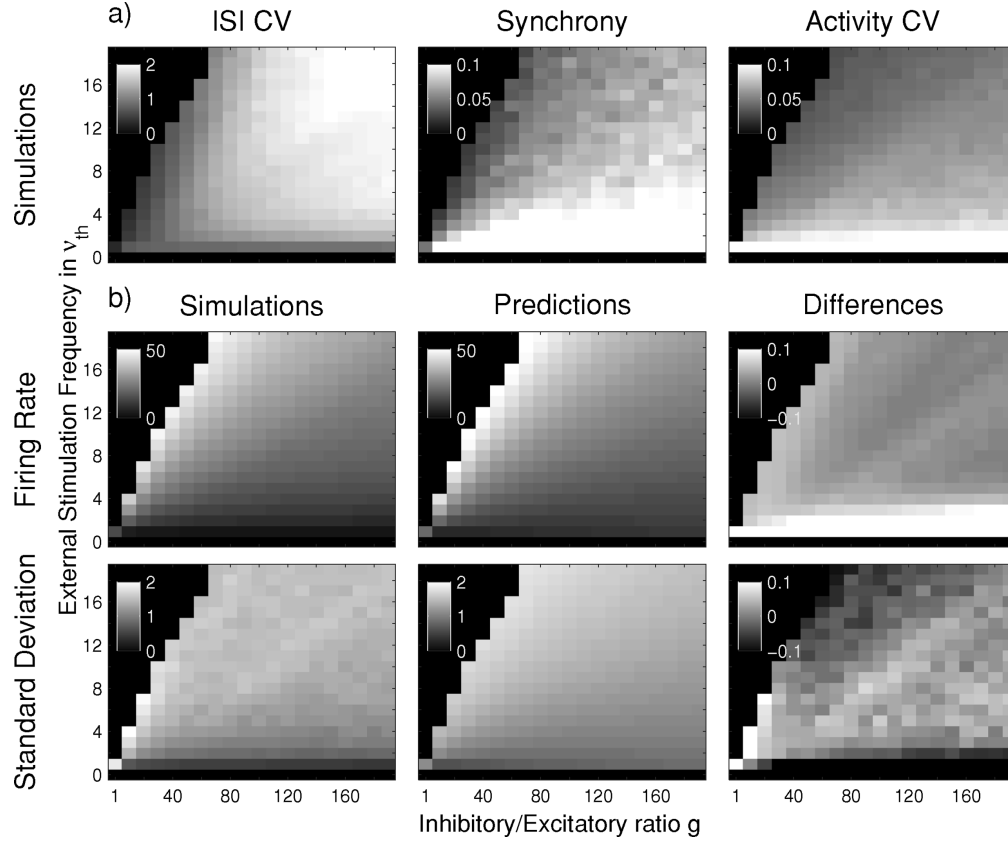


Figure 2: Characterization of the AI states and the first- and second-order statistics of the excitatory population activity in the  $(m_{exc}^{ext}, g)$  parameter space, similar to Brunel (2000). The network contains  $N = 5000$  neurons randomly connected with probability  $p^{conn} = 0.01$  and current-based synaptic interactions. Every statistical quantity has been computed with a time bin of  $T = 5$  ms, and the analytical model was solved with the same parameters. (a) The AI region is delimited using the mean ISI CV and the mean pairwise cross-correlation. In the right panel, we compute the activity CV to evaluate the validity of the independence hypothesis. (b) Top: Mean activity estimated from numerical simulations (left) and computed using the master equation formalism (middle). In the right panel, the relative difference between measured and predicted values. Bottom: The activity standard deviation is estimated from numerical simulations and compared as well with the mean-field predictions.

is homogeneous and the neurons are independent, the synchrony phase diagram should be well predicted by the activity second-order statistics. More precisely, it should be directly related to the activity coefficient of variation (CV). Therefore, in the last panel in (see Figure 2a), we computed the activity CV, which can be compared with the synchrony panel. Both diagrams exhibit the same tendency according to the parameter regime. Thus, this quantity could be estimated in the framework of the independence

hypothesis as long as the first- and second-order statistics are well predicted.

In Figure 2b, we computed and compared the mean excitatory activity and the standard deviation of excitatory activity for the parameter space. For each point, a simulation was run for 10 s in order to have an acceptable evaluation of the second-order statistics. The corresponding stationary quantities were numerically computed using the master equation formalism, equation 3.16, with the transfer function, equation 3.29. To compare our prediction with simulations, we computed the relative difference,

$$\Delta(m_{exc}^{ext}, g) = \frac{O_{Simulation}(m_{exc}^{ext}, g) - O_{Prediction}(m_{exc}^{ext}, g)}{|O_{Simulation}(m_{exc}^{ext}, g)| + |O_{Prediction}(m_{exc}^{ext}, g)|},$$

where  $O$  can be either the mean activity or the standard deviation. First-order and second-order statistics are in good agreement over almost the entire AI region with a relative error smaller than 0.1. However, the relative error is larger in the low-activity regime, as can be seen in Figure 2b. The reasons are twofold. On the one hand, for small networks exhibiting low activity, the synaptic input impinging on each neuron is equally low, and the diffusion approximation at the membrane potential level is not a good approximation anymore. On the other hand, the increase of pairwise correlations among neurons could partially invalidate the model, as can be seen from the discrepancies between the synchrony and the activity CV. For very large networks, a mean-field approach can be recovered even in the low-activity regime, and it provides accurate predictions (Kumar et al., 2008).

As the *in vivo* activity in awake animals usually displays low firing rates, we further investigated whether larger networks could generate more realistic cortical activity regimes. In Figure 3a, we show the results of a simulation of a larger network ( $N = 15,000$  neurons) while keeping other parameters exactly the same as used in the simulation of a smaller network described in Figure 2. It is apparent that the region between 6 Hz and 10 Hz, which was not well predicted previously, can now be better described by the mean-field approach (with relative errors smaller than 0.1). This region is delimited by black curves in Figure 3a.

To pursue this analysis, we chose a row of parameters in the phase diagram and computed the mean firing rate numerically and with the master equation for various network sizes. We found that the relative error for each network size increased very slightly for smaller firing rates, as expected (see Figure 3b). However, the relative error still stays below 0.1, even for firing rates around 5 Hz. The corresponding predicted firing rates are given in Figure 3c for the different network sizes. The error bars are computed with the relative error using  $|O_{Prediction}(m_{exc}^{ext}, g) - \Delta(m_{exc}^{ext}, g)|$ . In the right panel, we show for a particular point of the phase diagram  $(m_{exc}^{ext}, g) = (6 \cdot v_{th}, 141)$  a decrease in firing rate as well as an increase in the relative error. The

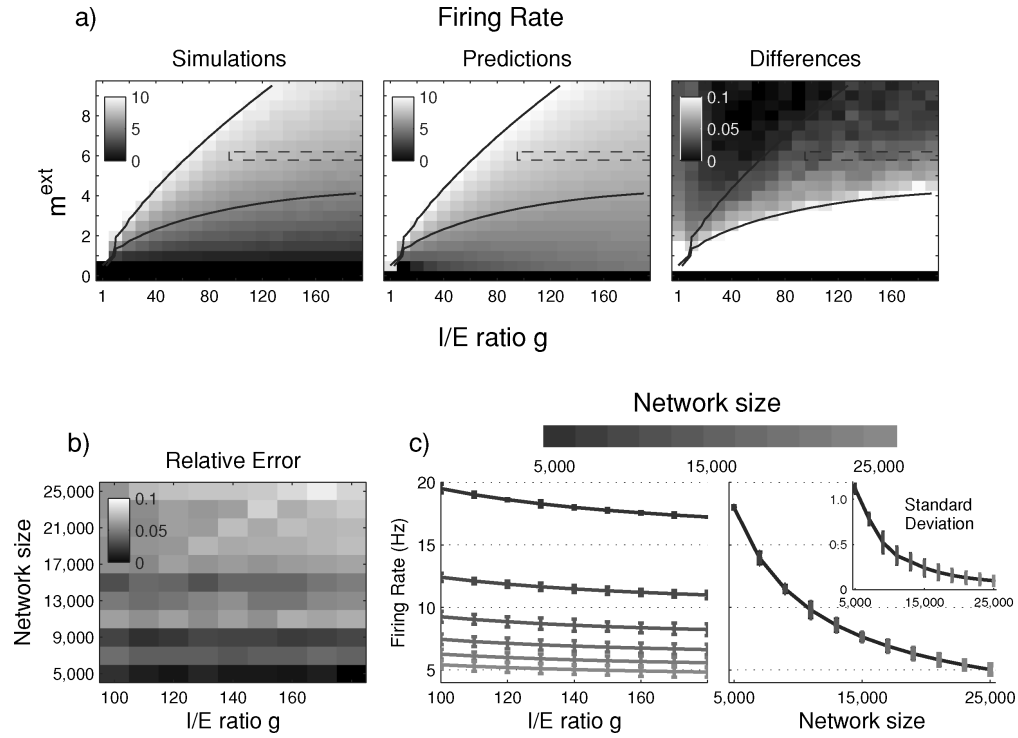


Figure 3: Master equation predictions for low firing rates depend on network size. (a) A large network of  $N = 15,000$  neurons has been used to draw a similar phase diagram as in Figure 2. The region between the two black curves exhibits firing rates smaller than 10 Hz (higher curve) and relative errors smaller than 0.1 (lower curve). The predictions are improved for a broad range of frequencies compared to a smaller network. (b) The relative error between numerical simulations and predictions for a row of parameters in the phase diagram (indicated by a dashed rectangle in *a* for different network sizes. (c) The relative error is slightly increasing for larger networks, but the firing rate decreases at the same time. Left: Predicted firing rate in the row for each network size. The error bars indicate the relative error in units of the corresponding predicted value. The firing rates decrease quickly compared to the error. Right: The firing rate decreases as a function of network size, for a particular point in the phase diagram  $(m_{exc}^{ext}, g) = (6 \cdot v_{th}, 141)$ . Inset: Decay of the standard deviation with the network size. All networks had identical parameters as in Figure 2.

predicted firing rate seems to decrease exponentially compared to the relative error, which does not change much. A similar tendency has been found for the standard deviation. The inset of the right panel of Figure 3c shows that the error remains almost constant while the activity standard deviation decreases for larger networks. The activity coefficient of variation indicates that the averaged pairwise correlations also decrease. Therefore, based on this result, it should be possible to find large enough networks to cover a realistic range of firing frequencies.

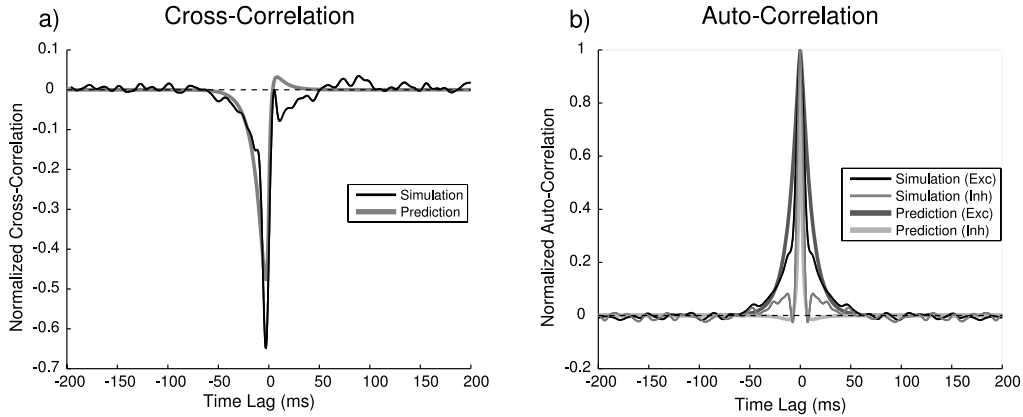


Figure 4: Correlation functions for the specific point  $(m_{exc}^{ext}, g) = (8 \cdot v_{th}, 41)$  of Figure 2. (a) Cross-correlation between inhibitory and excitatory population activity for the master equation (thick line) and the numerical simulations (thin line). (b) Autocorrelation functions computed from the master equation for the inhibitory and excitatory activity (thick lines) and with  $T = 10$  ms. The same functions have been estimated from the numerical simulations (thin lines). The activity traces have been computed with a bin size of 1 ms and then filtered with a gaussian function to get a smoother curve. The gaussian standard deviation was chosen to be  $1.5T$  to match the corresponding function without filtering.

Covariance and inhibitory population statistics are equally well predicted. To compare the activity correlations, we chose a point in the parameter space for which we estimated the inhibitory and excitatory autocorrelation and the cross-correlation. The corresponding functions were computed from the master equation using the last equation of equation 3.16 and the stationary mean activity computed for that particular state (see Figure 4). To keep a good resolution of the numerical functions, we took a time bin  $T = 10$  ms and filter the signal with a gaussian function with a standard deviation equal to  $1.5T$ . This gives a good prediction except for some part of the function, which can be due to a temporal finite-size effect. Indeed, the network is homogeneous, so the only difference between both populations is the size. If those populations are not large enough, residual oscillations that could not be predicted in our framework occur in the correlation functions. These oscillations also appear when the bin size  $T$  is too small (see Figure 1b). According to equation 3.16, we know that the interplay between network size and the considered timescale must be carefully taken into account to validate the model. For a very large network size or time constant  $T$ , the fast oscillations completely vanish. Moreover, we know from equation 3.16 that the correlation matrix depends directly on the neuron transfer function first derivative. More precisely, the predicted correlation exponential decrease as well as the activity standard deviation are led by this function. Therefore, regions where the autocorrelation and

the cross-correlation functions can be correctly described are regions where the second-order statistics match the numerical simulations.

*The  $(A_{exc}, A_{inh})$  parameter space.* To reproduce results obtained in Vogels and Abbott (2005), we chose a similar model to probe the  $(A_{exc}, A_{inh})$  parameter space. In the AI regime, current-based neurons cannot sustain activity without external stimulation, and one needs to inject a constant input in each neuron by bringing the resting potential to  $V^{rest} = -49$  mV, as in the original letter. We also took exponential synapses with excitatory and inhibitory time constant  $\tau_{exc} = 5$  ms and  $\tau_{inh} = 10$  ms, respectively. The only free parameters remaining are the current quantal increments  $(A_{exc}, A_{inh})$ . We first computed the mean ISI CV, the mean pairwise cross-correlation, and the activity CV to find the boundaries of the AI region. As in the previous section, the state diagram was drawn in the AI regime for the excitatory population with the master equation prediction using equation 3.29 and the neuron network simulations (see Figure 5). The mean activity prediction is in good agreement with the numerical simulations, with a relative error smaller than 0.1, as shown in the first row of Figure 5b. The relative error is larger for the low-activity regime ( $< 8$  Hz), where it can reach 0.2. This can be understood as previously; in this region, the synaptic input is not strong enough to fulfill the required conditions for the diffusion approximation. This error can be reduced for larger networks (see Figure 3). The prediction for the standard deviation matches quite well except for the region of high and low activity, where the difference is more substantial. The latter can be expected from the first-order comparison. Concerning the former, the region is at the boundary of the AR regime, and the Markov hypothesis begins to fail because of the emergence of regularities in the neuron individual firing.

Similar simulations were done using the transfer function, equation 3.28, and the resulting predictions were worse compared to previous results (not shown). This transfer function has been obtained as a first-order approximation for exponential synaptic noise when the ratio  $\sqrt{\frac{\tau_{syn}}{\tau_{mem}}}$  is small. In the numerical simulations here, the inhibitory synaptic time constant is only half the membrane time constant, and the assumption underlying equation 3.28 begins to fail. The phenomenological transfer function, equation 3.29, therefore seems more appropriate for these neuron properties.

**3.4.2 Conductance-Based Models.** Conductance-based neuron networks have been shown to display self-sustained activity for specific parameter space (Vogels & Abbott, 2005). For the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space, two regions have been identified exhibiting, respectively, AI and SR states (see Figures 6a–6c).

In this section, we study a similar model with the resting membrane time constant  $\tau^{mem} = 20$  ms, the membrane resistance  $R^{mem} = 100$  M $\Omega$ , the



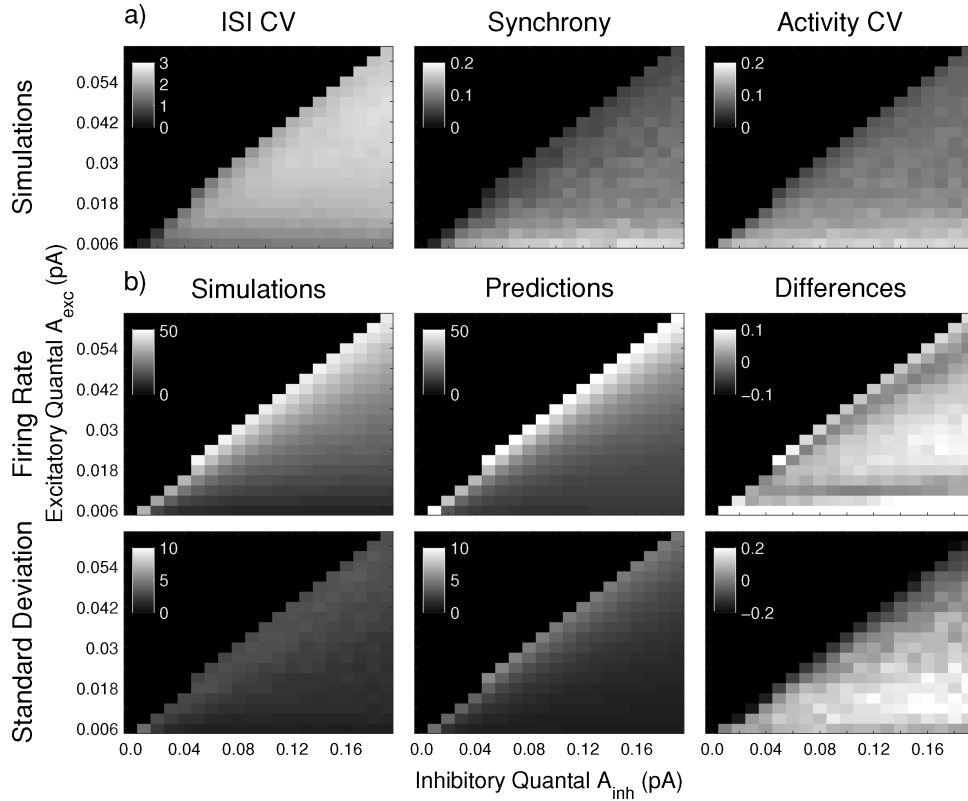


Figure 5: Characterization of AI states and the first- and second-order statistics of the excitatory population activity in the  $(A_{exc}, A_{inh})$  parameter space, in a Vogels-Abbott-type current-based network. The network contains  $N = 5000$  neurons randomly connected with probability  $p^{conn} = 0.01$ . Every statistical quantity has been computed with a time bin of  $T = 5$  ms, and the analytical model was solved with the same parameters. (a) The AI region is delimited using the mean ISI CV and the mean pairwise cross-correlation. In the right panel, we compute the activity CV to evaluate the validity of the independence hypothesis. (b) Top: Mean activity estimated from numerical simulations (left) and computed using the master equation formalism (middle). In the right panel, the relative difference between measured and predicted values. Bottom: The activity standard deviation is estimated from numerical simulations and compared as well with the mean-field predictions.

resting and reset membrane potential  $V^{rest} = V^{reset} = -60$  mV, the threshold  $V^{threshold} = -50$  mV, and the refractory period  $\tau^{ref} = 5$  ms. Synaptic time constant and reversal potential are taken to be  $\tau_{exc} = 5$  ms,  $E_{exc} = 0$  mV,  $\tau_{inh} = 10$  ms, and  $E_{inh} = -80$  mV for excitatory and inhibitory synapses, respectively. Given the state diagram (see Figures 6a–6c), we see that the AI region is surrounded by unstable states, which is not present in the current-based correspondent (compare with Figure 5). This is a specificity of the conductance-based model and an interesting issue that should be discussed

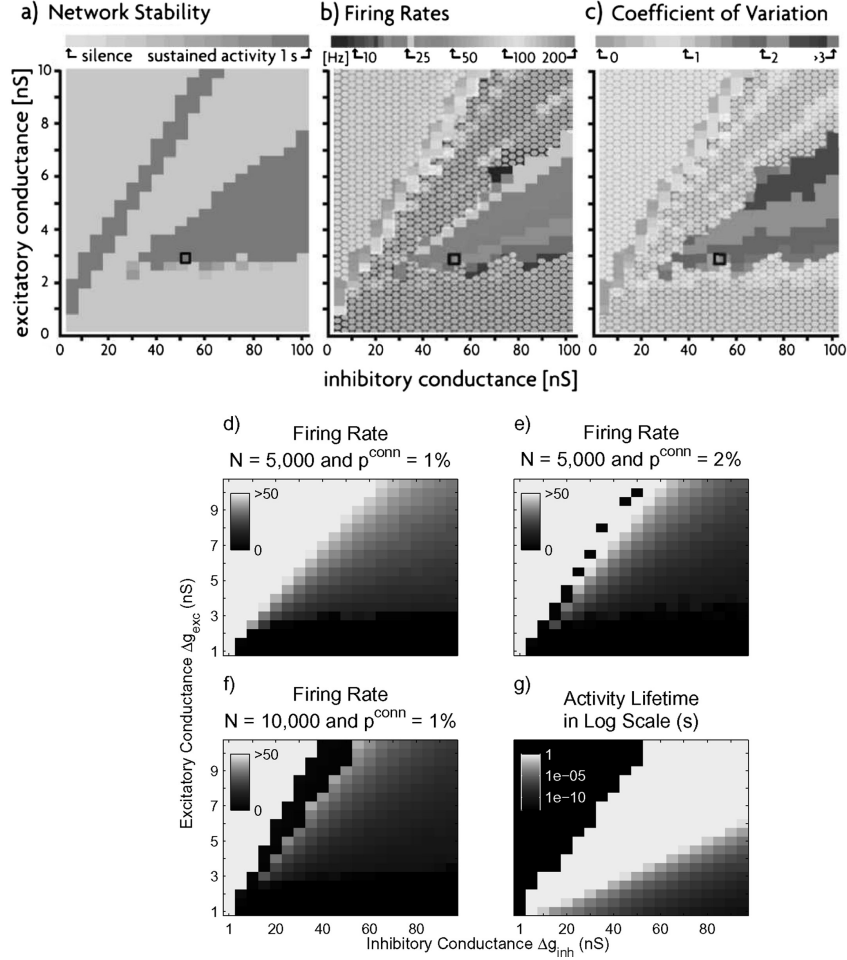


Figure 6: Sources of instabilities in Vogels-Abbott-type networks. Study of the stability (a), the mean activity (b), and the mean ISI CV (c) in the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space for a self-sustained conductance-based neuron network. The network in panels a to c contains  $N = 10,000$  neurons, each randomly connected with  $p^{conn} = 0.02$  of the population. The network is considered stable if its activity lasts longer than 1 s. (a–c modified, with authorization, from Vogels and Abbott, 2005). (d–f) Excitatory population mean activity for different network structures. For these numerical simulations, the network is considered stable if its activity lasts longer than 3 s. The transitory region between the AI and the SR domain in the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space is sensitive to synaptic input fluctuation. (d) Network of  $N = 5000$  neurons with  $p^{conn} = 0.01$ . The AR domain is fully stable. (e) Network of  $N = 5000$  neurons with  $p^{conn} = 0.02$ . Some disparate points in the AR domain lose their stability. (f) Network of  $N = 10,000$  neurons with  $p^{conn} = 0.01$ . The AR domain is almost completely unstable. (g) Network lifetime in log scale computed in the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space for a network of  $N = 10,000$  neurons and  $p^{conn} = 0.01$ . We took  $m_{crit} = 1$  Hz for the critical activity because no self-sustained activity of 1 Hz could be produced with this network model. Furthermore, this prediction has been computed by using the effective transfer function introduced below. Stability criteria match qualitatively with  $f$  for the low-activity AI region.

about the network stability in numerical simulation and the corresponding master equation prediction.

*Stability in self-sustained networks.* According to the stationary solution of the master equation, the active state could be unstable or stable in the usual sense. However, for stable points, there is an important consideration to take into account to match the numerical simulations. The framework presented here is a mean-field solution that does not take into account every finite-size effect that can be encountered in numerical models. For instance, stable states in the model with very weak activity are not sustained in network simulation. In some cases, the network activity fluctuations are too strong and can bring the dynamics to a quiescent state after a transient period. It has been shown that for multi-unit systems such as neural networks, a quasi-stationary state could survive for a period that is exponentially proportional to the system size (Crutchfield & Kaneko, 1988). More particularly, in self-sustained conductance-based networks, the state loses its stability when the network activity exhibits fluctuations that are too strong (Kumar et al., 2008). Therefore, we can adopt a criterion to keep those states with large fluctuations according to the timescale we are interested in. For instance, in Vogels and Abbott (2005), a network is said to be stable if it can sustain its activity longer than a second. Those networks will eventually fall into a quiescent state after a long transient. In our framework, once the equations are solved, we can compute the total mean activity and the total variance. For a balanced network, we have

$$\begin{aligned}\langle m_{tot} \rangle &= \langle (1 - \gamma)m_{exc} + \gamma m_{inh} \rangle \\ &= (1 - \gamma)\langle m_{exc} \rangle + \gamma \langle m_{inh} \rangle \\ \sigma^2(m_{tot}) &= \langle ((1 - \gamma)m_{exc} + \gamma m_{inh})^2 \rangle - \langle (1 - \gamma)m_{exc} + \gamma m_{inh} \rangle^2 \\ &= (1 - \gamma)^2 \sigma^2(m_{exc}) + \gamma^2 \sigma^2(m_{inh}) + 2(1 - \gamma)\gamma c_{exc/inh}.\end{aligned}$$

As  $m_{exc}$  and  $m_{inh}$  are described as a normal law defined by only the first two statistical moments, the total activity also follows a normal law with the following characteristics:

$$\begin{aligned}m_{tot} &\sim \mathcal{N}((1 - \gamma)\langle m_{exc} \rangle + \gamma \langle m_{inh} \rangle, (1 - \gamma)^2 \sigma^2(m_{exc}) \\ &\quad + \gamma^2 \sigma^2(m_{inh}) + 2(1 - \gamma)\gamma c_{exc/inh}).\end{aligned}\tag{3.58}$$

Therefore, we can use a simple criterion for stability by saying that the survival time is inversely proportional to the probability for the activity to be below a critical value  $m_{crit}$  near 0. Indeed, the spontaneous AI state is said to be quasi-stationary such that its activity is described by a stationary distribution. Eventually the dynamics will fall into this quiescent state, which is represented in mean-field theory by the probability below  $m_{crit}$

normalized by the minimal time step,

$$T_{survival} = \frac{T}{P(m_{tot} < m_{crit})} \quad (3.59)$$

$$= \frac{T}{F(m_{tot} = m_{crit})}, \quad (3.60)$$

where  $F(m_{tot} = m)$  is the repartition function. A similar criterion has been proposed in Kumar et al. (2008) but based on a first-order semianalytical model. In their argument, they mention a critical time window that corresponds to our network effective time constant  $T$ . In numerical simulations, they could estimate its value around 1 ms, which is very short compared to other time constants in the network model. This is encouraging for considering the master equation formalism well beyond large time constants. We can easily see from equations 3.16 and 3.22 that the covariance matrix is inversely proportional to the number of neurons in each population, so that  $T_{survival}$  grows exponentially with the number of neurons. Discrepancy from this result can be caused by residual correlations between neurons that could invalidate the binomial model. We can further estimate the network lifetime dependency on other parameters. We show in Figure 6g in a log scale the network lifetime evaluated with equation 3.59 for the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space for a network of  $N = 10,000$  neurons with  $p^{conn} = 0.01$ . We see that the AI region considered unstable in Figure 6f matches the region where the lifetime is beyond the order of a second. Therefore, the state diagram for self-sustained networks is sensitive to the time-scale under consideration.

Despite the stability issue, which is due to the limited lifetime of the network activity, there is finally an important source of instability that can also cause the network to reach the quiescent state. We can see numerically in Figures 6d to 6f that if the subthreshold membrane potential fluctuations are too strong, this induces supplementary fluctuations at the network level, which can destroy the activity spontaneously. These strong fluctuations can be caused by high levels of connectivity. Therefore, the upper AI region, which is predicted to be stable, can be numerically unstable because of those strong fluctuations. In Kumar et al. (2008), this region is not systematically unstable, and the lack of evidence with a second-order theory seems to reveal an artifact of the simulation or the breakdown of the model prediction. The sensitive region lay between the AI and the SR regime in which correlations can become critical. Indeed, the AR region displays correlated spike trains, which is not taken into account within the Poisson hypothesis, and this could go beyond the Markovian prediction. Kumar et al. (2008), have also suggested that the initial stimulation could contribute to this instability. Indeed, the initial stimulation must be adequate to allow the network to reach a self-sustained state. Increasing the network size will diminish the

firing rate, as predicted by the numerical and analytical transfer function, making the network more likely to shut down given the fluctuation state. This instability is reduced for sparser connectivities.

*The  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space.* We focus on the last configuration shown in Figure 6 to check the validity of the theoretical prediction. For each stable point of the state diagram, the set of differential equation 3.16, is solved with the transfer function, equation 3.53. The resulting analysis (see Figure 7) shows that the predictions are qualitatively correct, but the relative difference increases rapidly for low activity. There is a narrow region in the AI regime where the prediction is good for the first and second statistical moments. This important discrepancy is due to the approximation made for the transfer function, which is not derived for a complete conductance-based neuron with threshold. Similar results have already been reported in Kuhn et al. (2004) using the same transfer function.

To circumvent this problem, a semianalytical method has been proposed (Soula & Chow, 2007; Kumar et al., 2008), where the conductance-based neuron transfer function is determined numerically and then used directly in the model for a first-order or second-order prediction. Although this approach can give good predictions in the AI regime, it requires computing numerically the transfer function for each point of the state diagram, which can be time-consuming for heterogeneous networks. Indeed, the transfer function is determined by computing the neuron firing rate for a given population and for every state of each population in the network. Therefore, it is necessary to find an effective analytical function that can provide a better approximation. Although we cannot have an exact expression, it is still possible to fit a phenomenological model to the numerical simulations. Based on previous observations, two parameters seem critical in the transfer function: the time constant  $\tau$  in the denominator and a corrective term  $\Delta h$ , which takes into account colored noise in the synaptic input (Brunel & Sergi, 1998; Fourcaud & Brunel, 2002). The phenomenological function based on equation 3.27 can thus be written

$$v = \frac{1}{2\tau} \left( 1 + \operatorname{erf} \left( \frac{\langle V \rangle - V^{threshold}}{\sqrt{2}\sigma(V)} + \Delta h \right) \right). \quad (3.61)$$

Considering the limited region of the whole  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space in Figure 7, we can estimate the total error made for the set of transfer function, equation 3.61, according to the free parameters  $(\tau, \Delta h)$ . The total error is estimated by taking the sum of the relative difference between the simulation mean activity and the mean activity given by equation 3.61 in absolute value normalized by the number of stable points. For this optimization problem, there is a global minimum for which the error is small (see Figure 8). This computation depends on the network configuration and

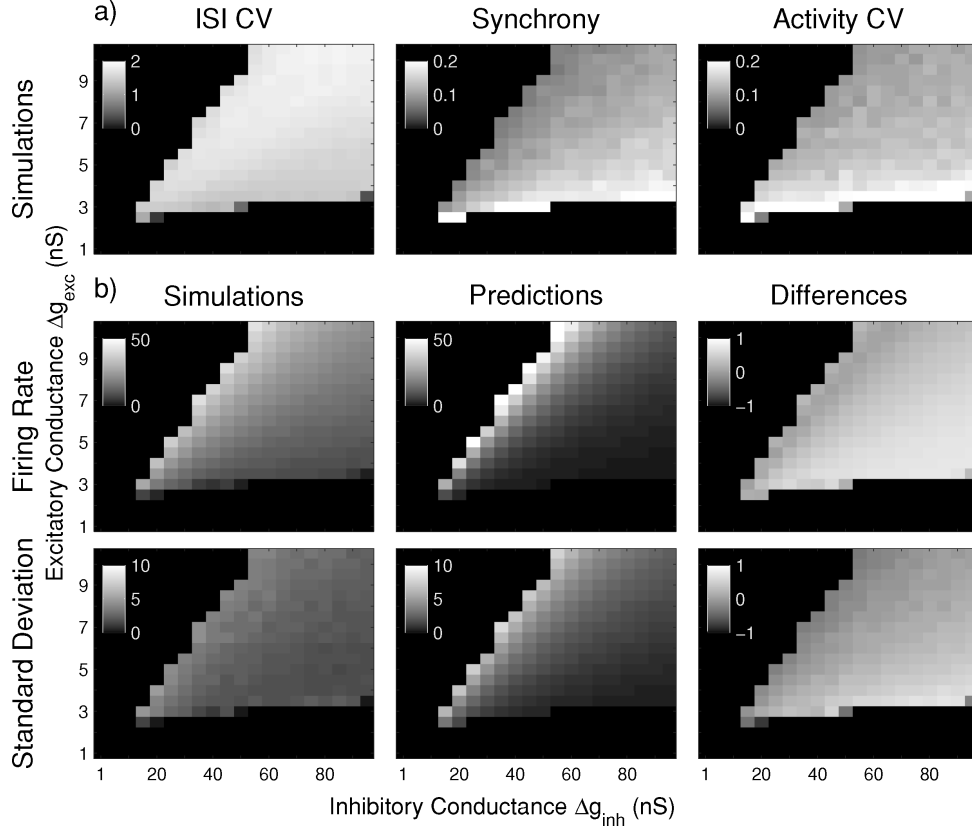


Figure 7: AI states characterization and the first- and second-order statistics of the excitatory population activity in the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space in a Vogels-Abbott-type conductance-based network. The network contains  $N = 10,000$  neurons randomly connected with probability  $p^{conn} = 0.01$ . Every statistical quantity has been computed with a time bin of  $T = 5$  ms, and the analytical model was solved with the same parameters. (a) The AI region is delimited using the mean ISI CV and the mean pairwise cross-correlation. In the right panel, we compute the activity CV to evaluate the validity of the independence hypothesis. (b) Top: Mean activity estimated from numerical simulations (left) and computed using the master equation formalism with the neuron transfer function, equation 3.53 (middle). In the right panel, the relative difference between measured and predicted values. Bottom: The activity standard deviation is estimated from numerical simulations and compared as well with the mean-field predictions.

the portion of the state diagram used to estimate the error. Therefore, this procedure gives a local good approximation that is acceptable as long as the network parameters (here the synaptic strengths) are kept in the fitted region.

The second-order statistics depends on the transfer function behavior around the stationary point through the first and second derivative of this

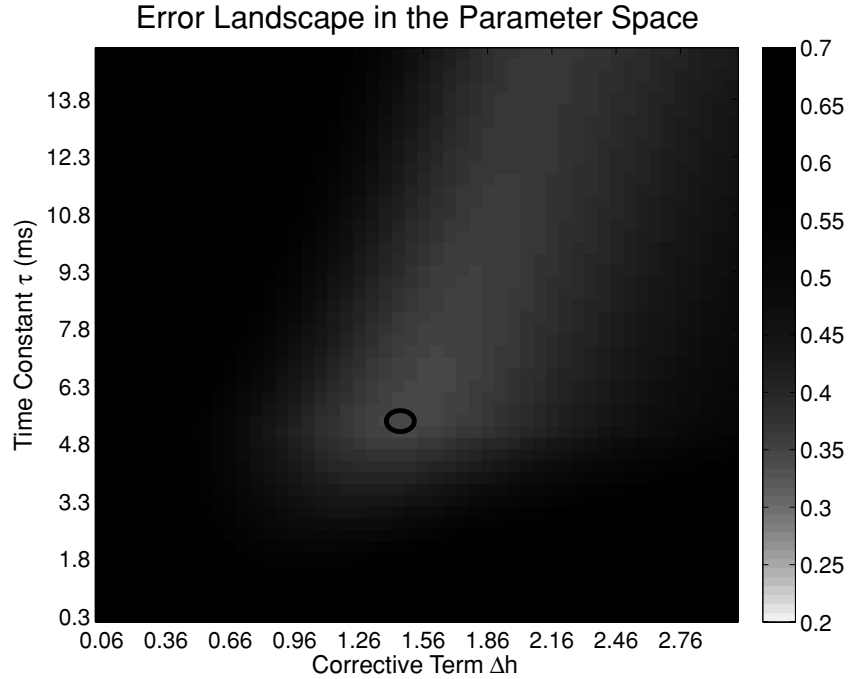


Figure 8: Error landscape for the optimization problem that determines the best parameter set for the effective transfer function, equation 3.61. It has been computed for a network of  $N = 10,000$  neurons and  $p^{comm} = 0.01$ . The minimal solution (circle) gives  $\tau = 5.3$  ms and  $\Delta h = 1.5$ .

function. In particular, the covariance and correlation matrix are strongly affected by the slope of the transfer function in the stationary point. Therefore, we compare for some point of the parameter space the effective transfer function with the semianalytical approach (see Figure 9). For each network configuration, the transfer function of a neuron is computed numerically and compared with the optimized transfer function. The local behavior around the stationary point is in good agreement with the latter.

Using the optimized transfer function, we can compare the prediction to the numerical simulation (see Figure 10). The error is reasonable compared to Figure 7, especially for the standard deviation for which prediction is much improved. Therefore, for a given network configuration, there is an effective transfer function that can provide a good description of the network dynamics in a large part of the parameter space. This allows us to avoid the semianalytical approach but requires solving an optimization problem based on numerical simulations.

This method could be useful when considering several network units described by the master equation. Indeed, eventually we would like to apply the master equation formalism to large-scale cortical recordings. In order to do so, it will be necessary to acquire a more realistic transfer function than



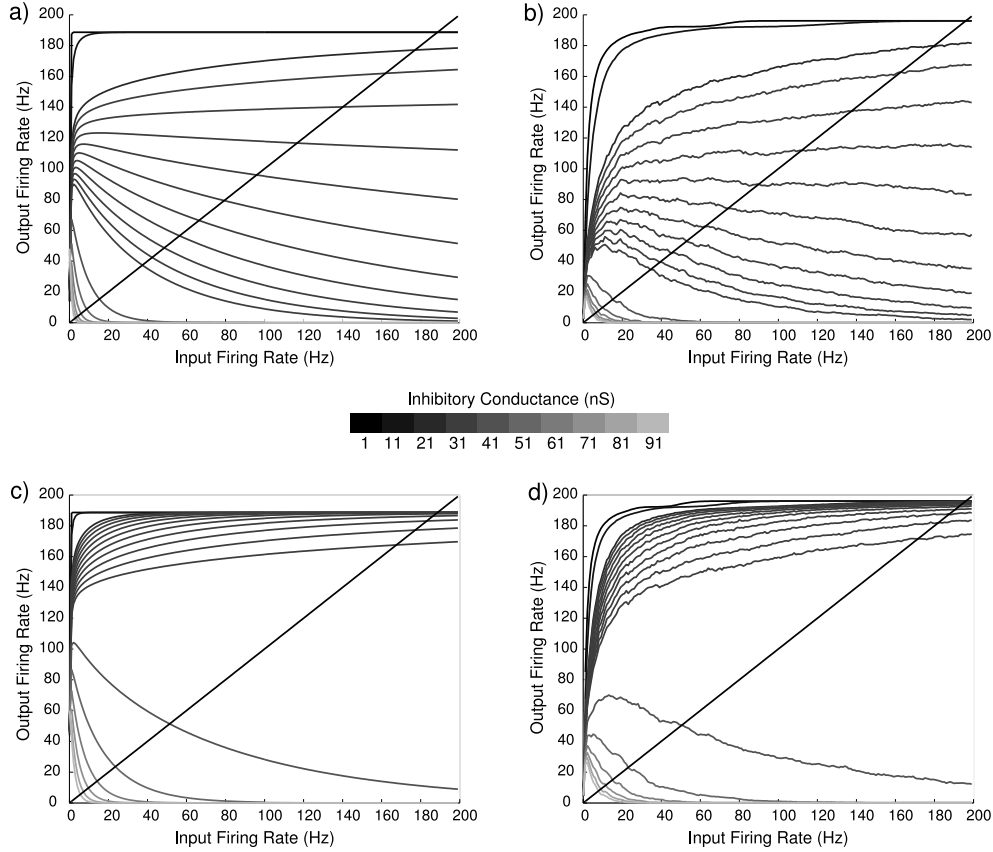


Figure 9: Transfer function estimated with the semianalytical approach compared with the optimized effective transfer function for several points of the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space. For each figure, the excitatory synaptic strength is fixed, and the inhibitory synaptic strength take the values  $\Delta g_{inh} = 1, 11, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 41, 51, 61, 71, 81,$  and  $91$  nS from the top-most curve to the lowest monotonically. In the first-order mean-field approximation, the network stationary activity is found by imposing that the input and output rate must be equal. The intersection with the diagonal line marks this first-order solution. (a) Effective transfer function and (b) numerical transfer function for  $\Delta g_{exc} = 7$  nS. (c) Effective transfer function and (d) numerical transfer function for  $\Delta g_{exc} = 10$  nS.

the one obtained from integrate-and-fire neurons. Using the optimization strategy, we could consider a large family of functions based on the usual theoretical results and obtain a representative transfer function adapted to high-conductance states. This could be done with dynamic-clamp in vitro experiments by finding autoconsistent solutions for a broad range of stimulation regimes. The resulting transfer function would then be used as a kernel in the master equation formalism to represent the dynamical property of the corresponding neuron population in the network unit. In

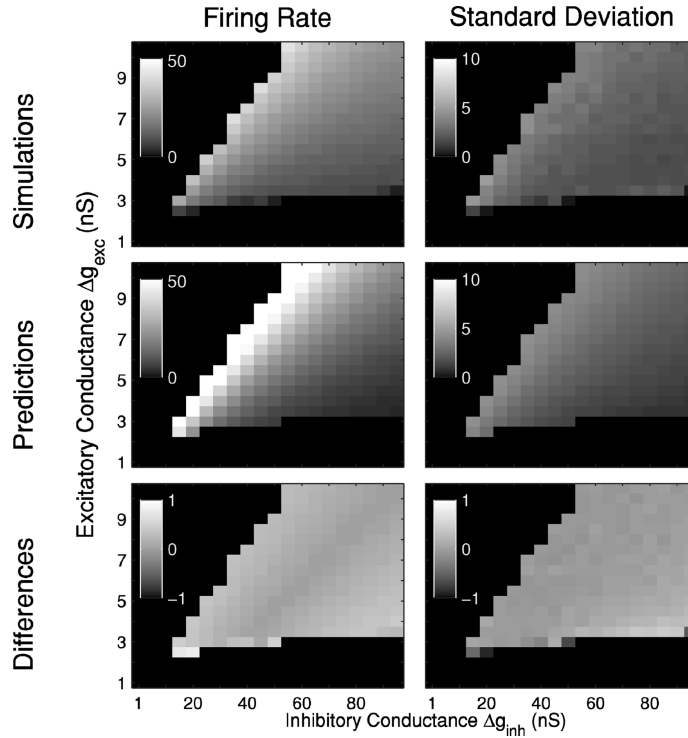


Figure 10: First- and second-order statistics of the excitatory population activity in the  $(\Delta g_{exc}, \Delta g_{inh})$  parameter space. A Vogels-Abbott network was simulated with  $N = 10,000$  conductance-based neurons randomly connected with probability  $p^{conn} = 0.01$ . The moments are computed with a time bin of  $T = 5$  ms, and the analytical model was solved with the same parameter and an effective transfer function (see equation 3.61). (Top) Mean activity and standard deviation estimated from numerical simulations. (Middle) Mean activity and standard deviation computed from the master equation formalism. (Bottom) Relative difference between measured and predicted values.

the context of voltage-sensitive dyes optical imaging, this unit would be associated with a small set of neighbor pixels. Once effective couplings between these units are extracted from data, it should be possible to obtain a theoretical comparative model in order to study activity propagation within large-scale cortical areas, in particular, dynamical phase transitions according to different network-controlled conditions.

*3.4.3 Effect of Topology and Heterogeneity.* In this section, we discuss the generality of the master equation formalism and possible sources of deviations from the predictions. Heterogeneity in the synaptic input across the network can create a bias in the mean activity of the network that is not taken into account in a mean-field model. Indeed, if neurons in the network do not receive exactly the same number of synapses, the firing rate distribution in the network will tend to be skewed compared to the sharp gaussian

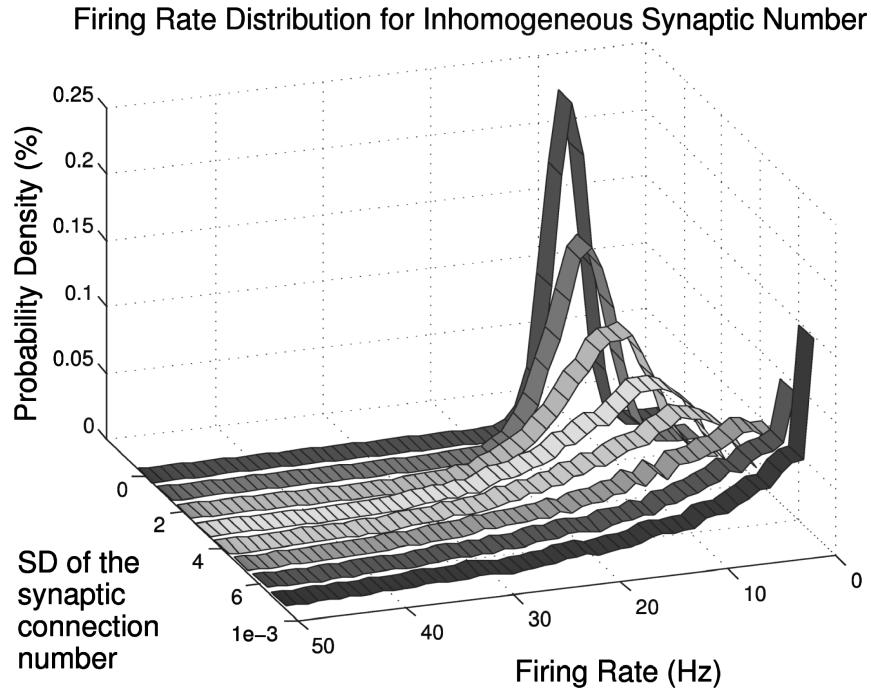


Figure 11: The neuron firing rate distribution in the AI state according to the degree of heterogeneity in the number of incoming synapses per neuron. The network contains  $N = 10,000$  neurons with a mean connectivity proportion  $p^{conn} = 0.02$ . The synaptic strengths are given by  $(\Delta g_{exc}, \Delta g_{inh}) = (6, 67)$  nS. The number of incoming synapses is taken from a gaussian distribution. From the back curve to the front curve, the standard deviation is increasing, resulting in skewer curves.

distribution for a homogeneous system. If the incoming connection number is gaussian distributed, the skewness of the firing rate distribution will increase with the standard deviation of the gaussian (see Figure 11), resulting in a shift of the mean activity. Similar results were reported by van Vreeswijk and Sompolinsky (1998) for heterogeneous thresholds. Therefore, predictions given by the master equation could lose some accuracy if the network is not perfectly homogeneous. However, this kind of heterogeneity does not create dramatic changes in the mean firing rate, so the model still provides good predictions.

Another interesting aspect of the theory concerns the connectivity schemes of the network. When the theoretical framework was built, two important hypotheses were made that directly concern network circuitry. We assumed that the connectivity is sparse and that the spiking probability is independent from one neuron to another at each time step. Any connectivity scheme that can account for those two hypotheses should be describable by this master equation formalism. Although connections in the cortex are highly specific, they are known to exhibit the sparseness property. As

we are describing macroscopic quantities of the network dynamics, high-order structure in the connectivity should not affect the analytical validity. However, the independence hypothesis can be broken if substantial correlations appear between neurons. In our simulations, we used a random connectivity scheme to avoid this situation. This choice is reasonable for very small networks, but for larger networks, it is necessary to consider a more realistic connectivity scheme. In previous work (Mehring et al., 2003), a numerical study of the  $(m_{exc}^{ext}, g)$  parameter space has been made for networks of locally random connected neurons with periodic boundary conditions. In this model, each neuron was connected to a fixed proportion  $p^{conn} = 0.1$  of its neighbors according to a gaussian probability law (see Figure 12c). The standard deviation is taken to be 0.3 mm for a square network with boundary length of 2 mm. This network contains  $N = 112,500$  current-based neurons with a ratio of 4:1 between excitatory and inhibitory neurons. Neurons interact with a fixed delay of 1.5 ms. The network is homogeneous with membrane time constant  $\tau^{mem} = 10$  ms, refractory period  $\tau^{ref} = 2$  ms, resting and reset potential  $V^{reset} = V^{rest} = -70$  mV, and threshold  $V^{threshold} = -50$  mV.  $\alpha$ -synapses were used with  $\tau_{exc} = \tau_{inh} = 0.3$  ms, and the excitatory synaptic strength was chosen such that the EPSP peak equals 0.14 mV. We implemented the same network in the formalism with the transfer function (see equation 3.29) and computed the excitatory mean activity to compare with their results (see Figures 12a and 12b). For this macroscopic quantity, it seems that correlations due to local connections do not dramatically alter the first-order mean-field predictions. This is very encouraging, and we hope to get qualitatively good descriptions of large-scale cortical networks based on this generic behavior of balanced networks. Of course, those who are interested in higher-order statistics of the dynamics have to rely on a more specific model of the network connectivity.

We investigated this question for a network that integrates some realistic features. Based on the anatomical data for the rat (DeFelipe, Alonso-Nanclares, & Arellano, 2002), we modeled a portion of the cortex as a layer with periodic boundary conditions respecting the superficial neuron density. This is defined as  $\rho_{ve} = \rho_s \sim 28,183$  neurons/mm<sup>2</sup> with  $\rho_v \sim 61,670$  neurons/mm<sup>3</sup> for layer 2–3 volume density, and  $e \sim 0.457$  mm the depth of these layers. We also took distance-dependent delays with a homogeneous propagation speed of  $v_{prop} = 5$  mm/ms. Intrinsic neuron properties are identical to the model used in Vogels and Abbott (2005), and the network contains  $N = 10,000$  neurons. Connections for a neuron follow a uniform law on a disc centered on the neuron. We considered as free parameters the radius of the connectivity disc and the proportion of connected neurons  $p^{conn}$  inside the disc (see Figure 12f). For a given synaptic strength set  $(\Delta g_{exc}, \Delta g_{inh}) = (6, 67)$  nS, we computed the mean excitatory activity as well as the mean interspike interval coefficient of variation to probe the first- and second-order properties of the dynamics (see Figures 12d and 12e). We notice the existence of isostatistics lines where

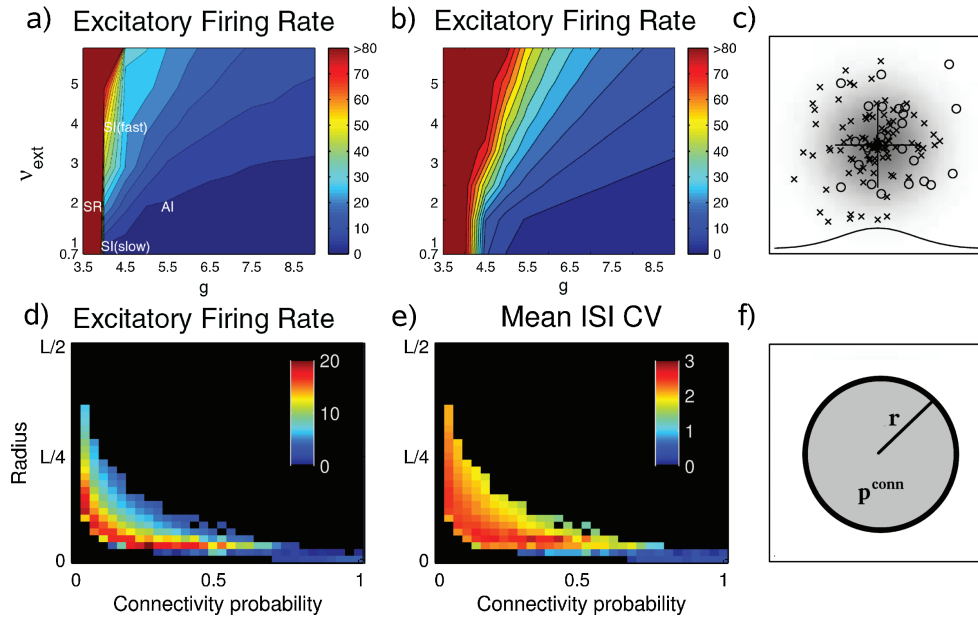


Figure 12: Effect of local network connectivity on macroscopic statistics. (a–c) Comparison of mean excitatory activity for a locally randomly connected current-based network (Mehring et al., 2003) in the  $(m_{exc}^{ext}, g)$  parameter space. (a) Numerical simulations with gaussian distribution (panel reproduced, with permission, from Mehring et al., 2003). (b) First-order mean-field predictions. (c) The gaussian distributed connectivity scheme. (d–f) Effect of local uniformly random connectivity. The parameter space is described by the connectivity disc radius and the connection probability inside the disc. (d) Mean excitatory activity. (e) Mean excitatory interspike interval coefficient of variation. (f) The uniform local random connectivity scheme.

the network displays the same macroscopic behavior. Those lines lie on network configurations where the number of inputs per neuron is constant. Indeed, if the disc radius is increased, the number of incoming synapses will also increase, and it is necessary to decrease the connection probability to recover the same statistical properties. This confirms the previous results shown in Figures 12a and 12b. Interneuron correlations due to more local connections do not seem to invalidate drastically the mean-field predictions as long as the number of connections per neuron is kept fixed and homogeneous. Note that in Figure 12, the case of the randomly connected network does not appear as a stable state because of the distant-dependent delays. Indeed, self-sustained activity cannot occur if the interactions among the network are too slow and we see a large domain of the state diagram that is unstable.

Following the discussion started at the end of the previous section, this result provides an encouraging foundation for large-scale modeling. Indeed, at first sight, it seems intractable to obtain a realistic model of

mesoscopic cortical dynamics when considering the high specificity of the wiring. However, this should not be relevant if population quantities are considered. Extrinsic optical imaging data provide such a quantity that could be related to the master equation variables. Therefore, all high-order patterns in the connectivity should be translated by statistical principles in effective macroscopic coupling available from large-scale data. Further investigations should be devoted to this question to ensure that specific high-order characteristics would not have a dramatic impact on global network dynamics.

#### 4 Discussion

---

In this letter, we have proposed a mean-field description of AI cortical activity states in balanced networks. We have considered a master equation formalism to describe the activity of networks of size  $N$  for timescales larger than a characteristic time  $T$ . These numbers were kept finite in order to account for finite-size effects and thus obtain a “mesoscopic” level of description. The resulting phenomenological theory provides a dynamical description of spiking neuron networks that sought to predict state diagrams, but could also be used beyond stability analysis. Furthermore, this framework can be used for any type of neuron as long as its transfer function is known. We obtained a closed set of equations for the mean and variance of the activity and showed that the state diagrams predicted by the mean-field model well matched the diagrams obtained numerically for different types of networks proposed previously (Brunel, 2000; Mehring et al., 2003; Vogels & Abbott, 2005).

Studies of network dynamics have shown that higher-order statistics are crucial in balanced networks. Indeed, chaotic behavior in those systems is produced by the balanced dynamics at the membrane potential level (van Vreeswijk & Sompolinsky, 1996, 1998; Brunel, 2000). Once the mean membrane potential is “clamped” at its subthreshold value, only fluctuations can bring the neuron to fire, thereby providing an irregular firing rate. However, although firing irregularity can be estimated from the stationary interspike interval coefficient of variation (Tuckwell, 1988), no model is able to describe the dynamics of second-order statistics at the network level. The Markovian approach is directly constrained by the population activity correlation’s fine structure during AI states and the minimal bin size to capture network dynamics. According to the numerical model, those values seem to be of the same order, which considerably simplifies the choice of the parameter  $T$ .

The master equation model directly relies on the neuron transfer function, which can be chosen according to the desired network model. This transfer function can be exactly determined in current-based networks provided that the input spike trains follow Poisson processes with Dirac synapses (Tuckwell, 1988; Brunel, 2000). However, to cover a broad range of synapse models, we adopt a phenomenological function that can account

for first- and second-order network statistics. For conductance-based models, no exact solution of the equation can be found. We also have to seek approximations to obtain the transfer function. Two main solutions have been proposed: estimating the transfer function numerically (Kumar et al., 2008; Soula & Chow, 2007) or using an approximation with the mean membrane potential of the conductance-based model and the variance taken from an effective current-based model Kuhn et al. (2004). To stay in an analytical framework, we decided to use the second approach, which can easily be implemented in our model.

In current-based models, the predicted state diagrams are in good agreement with the numerical simulations. In this case, the discrepancies are likely to be due to residual correlations caused by finite size effects or by the phenomenological transfer function. Indeed, for activity in small networks that is too low, the diffusion approximation is no longer legitimate, and the corresponding transfer function leads to incorrect predictions. We showed that those frequencies can be better described for larger networks, providing a good model for cortical dynamical regimes. In conductance-based models, the predicted diagrams match the numerical simulations qualitatively but give poor quantitative predictions. It is, however, possible to find an optimized transfer function from a set of functions described by two free parameters. This significantly improves the predictions. In section 3.4.2, we discussed the possibility of adapting this method to dynamics-clamp recordings in order to obtain biophysically more realistic transfer functions directly estimated from real neurons. At a large-scale level, this model could faithfully represent the behavior of conductance-based balanced networks and is therefore a good candidate for building macroscopic models of local field potentials or optical imaging data.

Preliminary results have shown that considering more local connectivities—instead of random schemes—does not alter significantly the master equation predictions as long as the sparseness is strong enough. Therefore, first- and second-order activity statistics do not require an exact description of the network structure. However, heterogeneity among the neurons can be responsible for slight discrepancies between simulations and predictions. Although our model does not take into account specific delays between neurons, for random delays of the order of  $T$ , the numerical simulations are even closer to predictions (data not shown). Indeed, global oscillations are destroyed by heterogeneous delays, and the AI region is larger. We are currently working on a systematic study of phase diagrams and their dependence on various parameters.

In conclusion, we have proposed here a mean-field approach to describe the activity of large networks but still staying in the finite-size regime. Such a “mesoscopic” description constitutes a first step toward obtaining a large-scale model of cerebral cortex tissue. The typical size of the networks considered here ( $N \sim 5000$  neurons) can be thought of representing the population of cortical neurons seen under one or several pixels of optical



imaging data. Typical values are  $100 \times 100$  pixels, covering from about  $3 \times 3$  mm to  $3 \times 3$  cm of cortical tissue, which gives about 5 to 5000 neurons per pixel in superficial layers according to neuronal densities published previously (Braitenberg & Schüz, 1998). Thus, constructing a  $100 \times 100$  network of such populations, each described by a master equation analogous to the model presented here, would be possible if the connectivity between adjacent and distant populations could be incorporated in the formalism. This important addition will require studying interconnected networks of neurons in AI states, which constitutes a natural extension of the modeling effort examined in this letter.

### Appendix A: Mean Activity and Covariance Matrix Differential Equations

---

In this appendix, we compute the set of differential equations for the mean activity and the covariance matrix from the master equation. For bin-sized activity defined on the time interval  $T$ , the central limit theorem allows one to stop the statistical moment hierarchy at the second order. In other words, we consider a gaussian approximation of the stochastic process. We note the mean activity  $\langle m_\mu \rangle$  so that

$$\begin{aligned}
& \partial_t \langle m_\mu \rangle \\
&= \partial_t \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha m_\mu P_t(\{m_\gamma\}) \\
&= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha m_\mu \partial_t P_t(\{m_\gamma\}) \\
&= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha \prod_{\beta=1, \dots, K} \int_0^{1/T} dm'_\beta (m_\mu P_t(\{m'_\gamma\}) W(\{m_\gamma\} | \{m'_\gamma\}) \\
&\quad - m_\mu P_t(\{m_\gamma\}) W(\{m'_\gamma\} | \{m_\gamma\})) \\
&= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha \prod_{\beta=1, \dots, K} \int_0^{1/T} dm'_\beta (m'_\mu - m_\mu) W(\{m'_\gamma\} | \{m_\gamma\}) P_t(\{m_\gamma\}) \\
&= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha a_\mu(\{m_\gamma\}) P_t(\{m_\gamma\}) \\
&= \langle a_\mu(\{m_\gamma\}) \rangle, \tag{A.1}
\end{aligned}$$

with

$$a_\mu(\{m_\gamma\}) = \prod_{\beta=1, \dots, K} \int_0^{1/T} dm'_\beta (m'_\mu - m_\mu) W(\{m'_\gamma\} | \{m_\gamma\}).$$

To obtain the first-order equation, we develop this function around each populations mean activity values to the second order,

$$a_\mu(\{m_\gamma\}) = a_\mu(\{\langle m_\gamma \rangle\}) + \partial_\lambda a_\mu(\{\langle m_\gamma \rangle\}) \cdot (m_\lambda - \langle m_\lambda \rangle) + \frac{1}{2} \partial_\lambda \partial_\eta a_\mu(\{\langle m_\gamma \rangle\}) \cdot (m_\lambda - \langle m_\lambda \rangle)(m_\eta - \langle m_\eta \rangle) + \mathcal{O}(\delta m^3),$$

which becomes, after averaging,

$$\langle a_\mu(\{m_\gamma\}) \rangle = a_\mu(\{\langle m_\gamma \rangle\}) + \frac{1}{2} \partial_\lambda \partial_\eta a_\mu(\{\langle m_\gamma \rangle\}) c_{\lambda\eta},$$

where we have introduced the covariance matrix  $c_{\lambda\eta} = \langle (m_\lambda - \langle m_\lambda \rangle)(m_\eta - \langle m_\eta \rangle) \rangle$  containing the second-order moments. This gives the first-order set of equations:

$$\begin{aligned} \partial_t \langle m_\mu \rangle &= a_\mu(\{\langle m_\gamma \rangle\}) + \frac{1}{2} \partial_\lambda \partial_\eta a_\mu(\{\langle m_\gamma \rangle\}) c_{\lambda\eta} \\ a_\mu(\{m_\gamma\}) &= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm'_\alpha (m'_\alpha - m_\mu) W(\{m'_\gamma\} | \{m_\gamma\}). \end{aligned} \quad (\text{A.2})$$

To compute the second-order equations, we proceed as before:

$$\begin{aligned} \partial_t c_{\mu\nu} &= \partial_t \langle (m_\mu - \langle m_\mu \rangle)(m_\nu - \langle m_\nu \rangle) \rangle \\ &= \partial_t \langle m_\mu m_\nu \rangle - \partial_t (\langle m_\mu \rangle \langle m_\nu \rangle) \\ &= \partial_t \langle m_\mu m_\nu \rangle - \langle m_\nu \rangle \partial_t \langle m_\mu \rangle - \langle m_\mu \rangle \partial_t \langle m_\nu \rangle. \end{aligned} \quad (\text{A.3})$$

Following the previous equations, we can write,

$$\begin{aligned} \partial_t \langle m_\mu m_\nu \rangle &= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha \\ &\quad \times \prod_{\beta=1, \dots, K} \int_0^{1/T} dm'_\beta (m'_\beta m'_\nu - m_\mu m_\nu) W(\{m'_\gamma\} | \{m_\gamma\}) P_t(\{m_\gamma\}). \end{aligned}$$

Using the relation

$$m'_\mu m'_\nu - m_\mu m_\nu = (m'_\mu - m_\mu)(m'_\nu - m_\nu) + m_\nu (m'_\mu - m_\mu) + m_\mu (m'_\nu - m_\nu),$$

we have

$$\partial_t \langle m_\mu m_\nu \rangle = \langle a_{\mu\nu}(\{m_\gamma\}) \rangle + \langle m_\nu a_\mu(\{m_\gamma\}) \rangle + \langle m_\mu a_\nu(\{m_\gamma\}) \rangle \quad (\text{A.4})$$

with

$$a_{\mu\nu}(\{m_\gamma\}) = \prod_{\beta=1,\dots,K} \int_0^{1/T} dm'_\beta (m'_\mu - m_\mu)(m'_\nu - m_\nu) W(\{m'_\gamma\} | \{m_\gamma\}).$$

Inserting equation A.4 into A.3, and using A.1, we get

$$\begin{aligned} \partial_t c_{\mu\nu} &= \langle a_{\mu\nu}(\{m_\gamma\}) \rangle + \langle m_\nu a_\mu(\{m_\gamma\}) \rangle + \langle m_\mu a_\nu(\{m_\gamma\}) \rangle \\ &\quad - \langle m_\nu \rangle \partial_t \langle m_\mu \rangle - \langle m_\mu \rangle \partial_t \langle m_\nu \rangle \\ &= \langle a_{\mu\nu}(\{m_\gamma\}) \rangle + \langle a_\mu(\{m_\gamma\}) \cdot (m_\nu - \langle m_\nu \rangle) \rangle \\ &\quad + \langle a_\nu(\{m_\gamma\}) \cdot (m_\mu - \langle m_\mu \rangle) \rangle. \end{aligned}$$

The first term can be expanded to the second order as we did for the first-order set of equations. Concerning the two other terms, we have,

$$\begin{aligned} &\langle a_\mu(\{m_\gamma\}) \cdot (m_\nu - \langle m_\nu \rangle) \rangle \\ &= \langle (m_\nu - \langle m_\nu \rangle) \cdot (a_\mu(\{m_\gamma\}) \\ &\quad + \partial_\lambda a_\mu(\{m_\gamma\}) \cdot (m_\lambda - \langle m_\lambda \rangle)) \rangle + \mathcal{O}(\delta m^3) \\ &= \partial_\lambda a_\mu(\{m_\gamma\}) \langle (m_\nu - \langle m_\nu \rangle)(m_\lambda - \langle m_\lambda \rangle) \rangle + \mathcal{O}(\delta m^3) \\ &= \partial_\lambda a_\mu(\{m_\gamma\}) c_{\nu\lambda} + \mathcal{O}(\delta m^3), \end{aligned}$$

and the same for the third term but with  $\mu$  and  $\nu$  inverted. The second-order set of equations can finally be written (van Kampen, 2003)

$$\begin{aligned} \partial_t c_{\mu\nu} &= a_{\mu\nu}(\langle \{m_\gamma\} \rangle) + \partial_\lambda a_\mu(\langle \{m_\gamma\} \rangle) c_{\nu\lambda} + \partial_\lambda a_\nu(\langle \{m_\gamma\} \rangle) c_{\mu\lambda} \\ a_{\mu\nu}(\{m_\gamma\}) &= \prod_{\alpha=1,\dots,K} \int_0^{1/T} dm'_\alpha (m'_\mu - m_\mu)(m'_\nu - m_\nu) W(\{m'_\gamma\} | \{m_\gamma\}). \end{aligned} \tag{A.5}$$

## Appendix B: Correlation Matrix Differential Equation

In this appendix, we compute the set of differential equations that describes the correlation matrix in a stationary state. The derivation is slightly different from the computation in appendix A. A similar computation has been done in Ginzburg and Sompolinsky (1994). The derivative can be written as:

$$\begin{aligned} \partial_\tau \text{Corr}_{\mu\nu}(\tau) &= \partial_\tau \langle (m_\mu(t) - \langle m_\mu(t) \rangle) (m_\nu(t + \tau) - \langle m_\nu(t + \tau) \rangle) \rangle \\ &= \partial_\tau \langle (m_\mu(t) m_\nu(t + \tau)) - \langle m_\mu(t) \rangle \langle m_\nu(t + \tau) \rangle \rangle. \end{aligned} \tag{B.1}$$

If we consider the first term of equation B.1,

$$\begin{aligned}
\partial_\tau \langle m_\mu(t) m_\nu(t + \tau) \rangle &= \\
&= \partial_\tau \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha \prod_{\beta=1, \dots, K} \int_0^{1/T} dm'_\beta m'_\mu m'_\nu P(\{m'_\gamma\}, t \\
&\quad + \tau \mid \{m_\gamma\}, t) P_t(\{m_\gamma\}) \\
&= \prod_{\alpha=1, \dots, K} \int_0^{1/T} dm_\alpha m_\mu \left( \prod_{\beta=1, \dots, K} \int_0^{1/T} dm'_\beta m'_\nu \partial_\tau P(\{m'_\gamma\}, t \right. \\
&\quad \left. + \tau \mid \{m_\gamma\}, t) \right) P_t(\{m_\gamma\}).
\end{aligned}$$

The conditional probability  $P(\{m'_\gamma\}, t + \tau \mid \{m_\gamma\}, t)$  is also a solution of the master equation, so that the term in the bracket is similar to equation A.1, and we can write

$$\partial_\tau \langle m_\mu(t) m_\nu(t + \tau) \rangle = \langle m_\mu(t) a_\nu(\{m_\gamma(t + \tau)\}) \rangle. \quad (\text{B.2})$$

The second term in equation A.1 is simply

$$\begin{aligned}
\partial_\tau \langle m_\mu(t) \rangle \langle m_\nu(t + \tau) \rangle &= \partial_\tau \langle m_\mu(t) \rangle \partial_\tau \langle m_\nu(t + \tau) \rangle \\
&= \langle m_\mu(t) \rangle \langle a_\nu(\{m_\gamma(t + \tau)\}) \rangle.
\end{aligned} \quad (\text{B.3})$$

Combining equations B.2 and B.3 into B.1, we get

$$\partial_\tau \text{Corr}_{\mu\nu}(\tau) = \langle (m_\mu(t) - \langle m_\mu(t) \rangle) a_\nu(\{m_\gamma(t + \tau)\}) \rangle. \quad (\text{B.4})$$

As we are considering time-scales that are beyond the decreasing time of the network activity correlations, we can develop  $a_\nu(\{m_\gamma(t + \tau)\})$  around the mean values to the second order,

$$\begin{aligned}
a_\nu(\{m_\gamma(t + \tau)\}) &= a_\nu(\{\langle m_\gamma(t + \tau) \rangle\}) + \partial_\lambda a_\nu(\{\langle m_\gamma(t + \tau) \rangle\}) \\
&\quad \times (m_\lambda(t + \tau) - \langle m_\lambda(t + \tau) \rangle) + \mathcal{O}(\delta m^2),
\end{aligned}$$

so that equation B.4 becomes, to second order,

$$\partial_\tau \text{Corr}_{\mu\nu}(\tau) = \partial_\lambda a_\nu(\{\langle m_\gamma(t + \tau) \rangle\}) \text{Corr}_{\mu\lambda}(\tau). \quad (\text{B.5})$$

We are interested in the activity correlations in the stationary state; therefore, the mean activities  $\{\langle m_\gamma \rangle\}$  no longer depend on  $\tau$ . Moreover, the initial

conditions for those differential equations are given by the stationary values of the covariance matrix.

### Appendix C: Conductance-Based Synaptic Functions for the Master Equation

---

In this appendix, we compute the remaining functions necessary to use the conductance-based model. The effective time constant is activity dependent in the conductance-based model, and it is given by

$$\langle \tau_\mu^{eff} \rangle = \frac{1}{\frac{1}{\tau_\mu^{mem}} + \sum_{\alpha=1, \dots, K} \frac{\langle G_{\alpha\mu} \rangle}{C_\mu}}.$$

We need to compute the first and second derivatives:

$$\begin{aligned} \partial_\lambda \langle \tau_\mu^{eff} \rangle &= -\langle \tau_\mu^{eff} \rangle^2 \frac{\Phi_{\lambda\mu}}{E_\lambda} \\ \partial_\eta \partial_\lambda \langle \tau_\mu^{eff} \rangle &= 2\langle \tau_\mu^{eff} \rangle^3 \frac{\Phi_{\eta\mu}}{E_\eta} \frac{\Phi_{\lambda\mu}}{E_\lambda}. \end{aligned}$$

The function  $\Psi_{\alpha\mu}$  depends on the chosen synapse, and we compute the corresponding derivatives for exponential and  $\alpha$ -synapses:

- Exponential synapses. The function is given by

$$\Psi_{\alpha\mu}^{Exp} = \frac{C_{\alpha\mu}}{\tau_\alpha + \langle \tau_\mu^{eff} \rangle} \left( \frac{\tau_\alpha \Delta g_{\alpha\mu} \langle \tau_\mu^{eff} \rangle (E_\alpha - \langle V_\mu \rangle)}{C_\mu} \right)^2,$$

so that

$$\begin{aligned} \partial_\lambda \Psi_{\alpha\mu}^{Exp} &= \Psi_{\alpha\mu}^{Exp} \left( \partial_\lambda \langle \tau_\mu^{eff} \rangle \left( \frac{2}{\langle \tau_\mu^{eff} \rangle} - \frac{1}{\langle \tau_\mu^{eff} \rangle + \tau_\alpha} \right) \right. \\ &\quad \left. - \frac{2\partial_\lambda Q_\mu}{(E_\alpha - \langle V_\mu \rangle)} \right) \\ \partial_\eta \partial_\lambda \Psi_{\alpha\mu}^{Exp} &= \frac{\partial_\eta \Psi_{\alpha\mu} \partial_\lambda \Psi_{\alpha\mu}}{\Psi_{\alpha\mu}} + \Psi_{\alpha\mu} (\partial_\eta \partial_\lambda \langle \tau_\mu^{eff} \rangle \\ &\quad \times \left( \frac{2}{\langle \tau_\mu^{eff} \rangle} - \frac{1}{\langle \tau_\mu^{eff} \rangle + \tau_\alpha} \right) + \partial_\eta \langle \tau_\mu^{eff} \rangle \partial_\lambda \langle \tau_\mu^{eff} \rangle \\ &\quad \times \left( \frac{1}{(\langle \tau_\mu^{eff} \rangle + \tau_\alpha)^2} - \frac{2}{\langle \tau_\mu^{eff} \rangle^2} \right) \\ &\quad \left. - \frac{2\partial_\lambda Q_\mu \partial_\eta Q_\mu}{(E_\alpha - \langle V_\mu \rangle)} - \frac{2\partial_\lambda \partial_\eta Q_\mu}{(E_\alpha - \langle V_\mu \rangle)^2} \right). \end{aligned}$$

- $\alpha$ -synapses. Similarly

$$\Psi_{\alpha\mu}^{\alpha sym} = \frac{1}{2} C_{\alpha\mu} (2\langle\tau_{\mu}^{eff}\rangle + \tau_{\alpha}) \times \left( \frac{e\tau_{\alpha} \Delta g_{\alpha\mu} \langle\tau_{\mu}^{eff}\rangle (E_{\alpha} - \langle V_{\mu} \rangle)}{C_{\mu} (\tau_{\alpha} + \langle\tau_{\mu}^{eff}\rangle)} \right)^2,$$

so that

$$\begin{aligned} \partial_{\lambda} \Psi_{\gamma\mu} &= 2\Psi_{\gamma\mu} \left( \partial_{\lambda} \langle\tau_{\mu}^{eff}\rangle \left( \frac{1}{(2\langle\tau_{\mu}^{eff}\rangle + \tau_{\gamma})} + \frac{1}{\langle\tau_{\mu}^{eff}\rangle} \right. \right. \\ &\quad \left. \left. - \frac{1}{(\langle\tau_{\mu}^{eff}\rangle + \tau_{\gamma})} \right) - \frac{\partial_{\lambda} Q_{\mu}}{(E_{\gamma} - \langle V_{\mu} \rangle)} \right) \\ \partial_{\eta} \partial_{\lambda} \Psi_{\gamma\mu} &= \frac{\partial_{\lambda} \Psi_{\gamma\mu} \partial_{\eta} \Psi_{\gamma\mu}}{\Psi_{\gamma\mu}} \\ &\quad + \left( \partial_{\eta} \partial_{\lambda} \langle\tau_{\mu}^{eff}\rangle \left( \frac{1}{(2\langle\tau_{\mu}^{eff}\rangle + \tau_{\gamma})} + \frac{1}{\langle\tau_{\mu}^{eff}\rangle} \right. \right. \\ &\quad \left. \left. - \frac{1}{(\langle\tau_{\mu}^{eff}\rangle + \tau_{\gamma})} \right) \right. \\ &\quad \left. + \partial_{\eta} \langle\tau_{\mu}^{eff}\rangle \partial_{\lambda} \langle\tau_{\mu}^{eff}\rangle \left( \frac{1}{(\langle\tau_{\mu}^{eff}\rangle + \tau_{\gamma})^2} \right. \right. \\ &\quad \left. \left. - \frac{2}{(2\langle\tau_{\mu}^{eff}\rangle + \tau_{\gamma})^2} - \frac{1}{\langle\tau_{\mu}^{eff}\rangle^2} \right) \right. \\ &\quad \left. - \frac{\partial_{\eta} \partial_{\lambda} Q_{\mu}}{(E_{\gamma} - \langle V_{\mu} \rangle)} - \frac{\partial_{\eta} Q_{\mu} \partial_{\lambda} Q_{\mu}}{(E_{\gamma} - \langle V_{\mu} \rangle)^2} \right). \end{aligned}$$

## Acknowledgments

---

The research was supported by the CNRS, ANR, and the European Community (FACETS grant FP6 15879). We thank Olivier Marre, Pierre Yger, Nicolas Brunel, and Romain Brette for fruitful discussions and comments on the manuscript.

## References

---

- Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex*, *7*, 237–252.
- Braitenberg, V., & Schüz, A. (1998). *Cortex: Statistics and geometry of neuronal connectivity* (2nd ed.). Berlin: Springer-Verlag.
- Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., et al. (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *J. Comput. Neurosci.*, *23*, 349–398.
- Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.*, *8*, 183–208.
- Brunel, N., & Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.*, *11*, 1621–1671.
- Brunel, N., & Sergi, S. (1998). Firing frequency of leaky integrate-and-fire neurons with synaptic currents dynamics. *J. Theor. Biol.*, *195*, 87–95.
- Crutchfield, J. P., & Kaneko, K. (1988). Are attractors relevant to turbulence? *Phys. Rev. Lett.*, *60*, 2715–2718.
- DeFelipe, J., Alonso-Nanclares, L., & Arellano, J.I. (2002). Microstructure of the neocortex: Comparative aspects. *J. Neurocytology*, *39*, 299–316.
- Destexhe, A., & Contreras, D. (2006). Neuronal computations with stochastic network states. *Science*, *314*, 85–90.
- Destexhe, A., Rudolph, M., & Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nature Reviews Neurosci.*, *4*, 739–751.
- El Boustani, S. (2006). *Information transport in networks during irregular activity states* (in French). Master's thesis, Ecole Normale Supérieure, France.
- El Boustani, S., & Destexhe, A. (2007). Mesoscopic model of balanced neuron networks using a master equation formalism. (abstract) In *Computation and Neural Systems 2007 Conference*. Available online at <http://www.cnsorg.org>.
- El Boustani, S., Pospischil, M., Rudolph-Lilith, M., & Destexhe, A. (2007). Activated cortical states: Experiments, analyses and models. *J. Physiol. Paris* *101*, 99–109.
- Fourcaud, N., & Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput.*, *14*, 2057–2110.
- Gerstner, W. (2000). Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Comput.*, *12*, 43–89.
- Ginzburg, I., & Sompolinsky, H. (1994). Theory of correlations in stochastic neural networks. *Phys. Rev.*, *50*, 3171–3191.
- Hertz, J., Lerchner, A., & Ahmadi, M. (2004). Mean field methods for cortical network dynamics. *Comput. Neurosci.*, *3146*, 71–89.
- Kuhn, A., Aertsen, A., & Rotter, S. (2004). Neuronal integration of synaptic input in the fluctuation-driven regime. *J. Neurosci.*, *24*, 2345–2356.
- Kumar, A., Schrader, S., Aertsen, A., & Rotter, S. (2008). The high-conductance state of cortical networks. *Comput. Neurosci.*, *20*, 1–43.
- Latham, P. E., Richmond, B. J., Nelson, P. G., & Nirenberg, S. (2000). Intrinsic dynamics in neuronal networks. I. Theory. *J. Neurophysiol.*, *83*, 808–827.



- Matsumura, M., Cope, T., & Fetz, E. E. (1988). Sustained excitatory synaptic input to motor cortex neurons in awake animals revealed by intracellular recording of membrane potentials. *Exp. Brain Res.*, *70*, 463–469.
- Mehring, C., Hehl, U., Kubo, M., Diesmann, M., & Aertsen, A. (2003). Activity dynamics and propagation of synchronous spiking in locally connected random networks. *Biol. Cybern.*, *88*, 395–408.
- Ohira, T., & Cowan, J. D. (1993). Master-equation approach to stochastic neurodynamics. *Phys. Rev. E*, *48*, 2259–2266.
- Plesser, H. E., & Gerstner, W. (2000). Noise in integrate-and-fire neurons: From stochastic input to escape rates. *Neural Comput.*, *12*, 367–384.
- Soula, H., & Chow, C. C. (2007). Stochastic dynamics of a finite-size spiking neural network. *Neural Comput.*, *19*, 3262–3292.
- Steriade, M., Timofeev, I., & Grenier, F. (2001). Natural waking and sleep states: A view from inside neocortical neurons. *J. Neurophysiol.*, *85*, 1969–1985.
- Tuckwell, H. C. (1988). *Introduction to theoretical neurobiology*. Cambridge: Cambridge University Press.
- van Kampen, N. G. (2003). *Stochastic processes in physics and chemistry*. Amsterdam: North-Holland Personal Library.
- van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, *274*, 1724–1726.
- van Vreeswijk, C., & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.*, *10*, 1321–1371.
- Vogels, T. P., & Abbott, L. F. (2005). Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.*, *25*, 10786–10795.

---

Received February 12, 2008; accepted May 16, 2008.

## Spike-Frequency Adapting Neural Ensembles: Beyond Mean Adaptation and Renewal Theories

**Eilif Muller**

*emueller@kip.uni-heidelberg.de*

**Lars Buesing**

*lbuesing@kip.uni-heidelberg.de*

**Johannes Schemmel**

*schemmel@kip.uni-heidelberg.de*

**Karlheinz Meier**

*meierk@kip.uni-heidelberg.de*

*Kirchhoff Institute for Physics, University of Heidelberg, 69120 Heidelberg, Germany*

We propose a Markov process model for spike-frequency adapting neural ensembles that synthesizes existing mean-adaptation approaches, population density methods, and inhomogeneous renewal theory, resulting in a unified and tractable framework that goes beyond renewal and mean-adaptation theories by accounting for correlations between subsequent interspike intervals. A method for efficiently generating inhomogeneous realizations of the proposed Markov process is given, numerical methods for solving the population equation are presented, and an expression for the first-order interspike interval correlation is derived. Further, we show that the full five-dimensional master equation for a conductance-based integrate-and-fire neuron with spike-frequency adaptation and a relative refractory mechanism driven by Poisson spike trains can be reduced to a two-dimensional generalization of the proposed Markov process by an adiabatic elimination of fast variables. For static and dynamic stimulation, negative serial interspike interval correlations and transient population responses, respectively, of Monte Carlo simulations of the full five-dimensional system can be accurately described by the proposed two-dimensional Markov process.

### 1 Introduction ---

Spike-frequency adaptation (SFA) refers to the intrinsic property of certain neurons to fire with gradually increasing interspike intervals (ISIs) in response to a steady injection of suprathreshold current. SFA is ubiquitous: It has been observed in many neural systems of diverse species (Fuhrmann, Markram, & Tsodyks, 2002). In the mammalian visual system, for example, the majority of retinal ganglion cells (RGCs) (O'Brien, Isayama, Richardson, & Berson, 2002), geniculate relay neurons (Smith, Cox, Sherman, & Rinzel,

2001), and neocortical and hippocampal regular spiking pyramidal neurons (McCormick, Connors, Lighthall, & Prince, 1985) exhibit SFA.

The in-vitro conditions used to experimentally verify the presence of SFA are far from the operational mode of a typical neuron in a network. Given cortical neuron firing rates and interconnectivity, each neuron there is under intense bombardment by both excitatory and inhibitory synapses. These mutually opposing showers of excitation and inhibition induce highly irregular fluctuations of the membrane potential reminiscent of a random walk. The resulting dominance of the mean synaptic conductances over the leak results in a markedly shortened effective membrane time constant, a dynamical regime known as the high-conductance state (Destexhe, Rudolph, & Paré, 2003; Shelley, McLaughlin, Shapley, & Wielaard, 2002). In this regime, action potentials are emitted when the membrane potential chances across the firing threshold and the resulting ISIs appear stochastic and are, for adapting neurons, roughly gamma distributed (Softky & Koch, 1993; Destexhe, Rudolph, Fellous, & Sejnowski, 2001; Dayan & Abbott, 2001).

Conductance-based phenomenological models for SFA and related relative refractory mechanisms are standard and given in Dayan and Abbott (2001) and Koch (1999) and recently generalized in Brette and Gerstner (2005). Benda and Herz (2003) show that a large class of biophysical mechanisms that induce SFA can be reduced to these conductance-based phenomenological models. Similar but current-based adaptation mechanisms have been studied in van Vreeswijk and Hansel (2001) and the related threshold fatigue model for adaptation, also known as dynamic threshold, in Chacron, Pakdaman, and Longtin (2003) and Lindner and Longtin (2003). See Ermentrout, Pascal, and Gutkin (2001) for a bifurcation analysis of  $I_{ahp}$ , the afterhyperpolarization current, a calcium-dependent potassium current, and  $I_m$ , the muscarinic slow voltage-dependent potassium current, two biophysical mechanisms behind SFA.

Mean-adaptation approximations for the firing rate of populations of spike-frequency adapting neurons augmenting the standard Wilson and Cowan equations (Wilson & Cowan, 1972) were devised in Latham, Richmond, Nelson, and Nirenberg (2000) and Fuhrmann et al. (2002) and used to study the synchronizing effects of SFA. Universal mean-adaptation methods for modeling the firing rate of adapting neurons subject to suprathreshold noise-free current input are given in Benda and Herz (2003). In La Camera, Rauch, Lüscher, Senn, and Fusi (2004), mean-adaptation methods are investigated to describe the static and dynamic firing rates of a large class of integrate-and-fire neuron models with current-based and dynamic threshold adaptation mechanisms driven by noisy input currents. The phenomenological firing rate relaxation dynamics of previous Wilson and Cowan studies is replaced in La Camera et al. (2004) with a firing rate that depends instantaneously on filtered synaptic currents, as suggested in Fourcaud and Brunel (2002) and Renart, Brunel, and Wang (2004). While

for the Wilson and Cowan approaches, the relaxation time constant is a free parameter, the approach due to La Camera et al. (2004) has no free parameters, and excellent agreement is reported in the static and dynamic case for several neuron models.

Originally introduced in Knight (1972) and recently the subject of intense study, population density formalisms provide powerful tools to understand neural ensemble and network behavior in a quantitative way (Brunel, 2000; Omurtag, Knight, & Sirovich, 2000; Nykamp & Tranchina, 2000, 2001; Fourcaud & Brunel, 2002; Meffin, Burkitt, & Grayden, 2004; Renart et al., 2004). Such studies are mostly restricted to exactly solvable white noise input cases, with notable exceptions (Nykamp & Tranchina, 2001; Fourcaud & Brunel, 2002). In Fourcaud and Brunel (2002), the key observation is made that colored input noise due to synaptic filtering results in a nonzero probability density near threshold and allows neurons to respond instantaneously to injected currents. This provides the theoretical basis for studies such as La Camera et al. (2004) and will also play an important role in the work here. Conductance-based neurons with finite synaptic time constants are treated in Rudolph and Destexhe (2003a, 2005), Richardson (2004), Richardson and Gerstner (2005), though only in the subthreshold regime, limiting their applicability for understanding firing rate, and networks dynamics. The problem with threshold has yet to be solved exactly, however, it is treated in Moreno-Bote and Parga (2004, 2005).

For neurons without SFA driven by noisy input, an alternate and fruitful approach is to apply renewal theory as presented in detail in Gerstner and Kistler (2002). With the defining characteristic of renewal theory being that successive ISIs are statistically independent, these models neglect by definition the observation in Chacron et al. (2003) and Lindner and Longtin (2003) that SFA induces negative serial ISI correlations.

While the great majority of excitatory neurons exhibit SFA, there has yet to be a population density treatment accounting for it, given the difficulty in treating the added dimension analytically and numerically. We present here a study whereby the ensemble behavior of adapting neurons in the high-conductance state can be understood in a quantitative way.

We start by considering in section 2 how to go beyond the renewal theory formalism of Gerstner and Kistler (2002) by introducing a dependence between ISIs, resulting in a Markov model described by a master equation. A connection to renewal theory is found by a suitable variable transformation, and expressions for the ISI distribution and conditional ISI distribution are derived. We then consider in section 3 the full five-dimensional master equation of the canonical conductance-based integrate-and-fire neuron model driven by Poisson spike trains augmented by SFA and a relative refractory mechanism of the form given in Dayan and Abbott (2001). By applying an adiabatic elimination of fast relaxing variables (Haken, 1983; Gardiner, 1984), we argue that this five-dimensional master equation can be approximated by a two-dimensional master equation of the same form as the

“beyond renewal theory” Markov model proposed in section 2. In section 4, we determine the generalized hazard function required for the Markov model by fitting to Monte Carlo simulations of the full system, given that the firing rate of the neuron model we employ has yet to be solved exactly. By reasoning as in Fourcaud and Brunel (2002), Renart et al. (2004), and La Camera et al. (2004), we show how the generalized hazard function applies in the dynamic case by accounting for synaptic filtering. In section 5, we provide numerical methods for solving the master equations and generating realizations of the proposed Markov processes. In section 6, predictions for ISI correlations and conditional ISI distributions in the static case, and firing rates in the dynamic case due to the proposed Markov model are compared to Monte Carlo simulations of the full system. Finally in section 7, the master equation is employed to analyze the domain of validity of mean-adaptation approaches.

## 2 Beyond Renewal Theory

---

Gerstner and Kistler (2002) demonstrate that for spike response models (a generalization of integrate-and-fire neuron models), the statistical ensemble of a single neuron with noise can be described using methods of inhomogeneous renewal theory, as reviewed in appendix C.

The basic assumption of inhomogeneous renewal theory is that the state of the modeled system can be described by a single state variable,  $\tau$ , the time since last renewal, or age of the system, and time  $t$ . The limiting probability density for the neuron to spike, or more generally, for the system to renew after surviving a time interval  $\tau$ ,

$$\rho(\tau, t) = \lim_{\Delta t \rightarrow 0^+} \frac{\text{prob}\{> 0 \text{ renewals in } [t, t + \Delta t) \mid \tau\}}{\Delta t}, \quad (2.1)$$

also known as the hazard function (Cox, 1962), is a function of time,  $t$ , and age,  $\tau$ .<sup>1</sup> Thus, subsequent interspike intervals (ISIs) are by definition independent and uncorrelated.

As Gerstner and Kistler (2002, pp. 245), stated, “A generalization of the [renewal] population equation to neuron models with [spike-frequency] adaptation is not straightforward since the [renewal] formalism assumes that only the last spike suffices. . . . A full treatment of adaptation would involve a density description in the high-dimensional space of the microscopic neuronal variables [as in] (Knight, 2000).”

In section 3 we provide a full treatment of the density description mentioned above. However, before we proceed, it is instructive to consider what

---

<sup>1</sup>For our discussion of renewal processes, we follow the notation of Cox (1962) but use  $\tau$  to denote age,  $t$  to denote time, and  $\rho$  instead of  $h$  to denote the hazard function, as in appendix C.

a model might look like that allows for a dependence between subsequent ISIs.

Consider the standard phenomenological model for spike-frequency adaptation (SFA) proposed in Dayan and Abbott (2001) where a given neuron model is augmented with a conductance  $g_s(t)$  that makes the jump  $g_s(t + dt) = g_s(t) + q_s$  when the neuron spikes at time  $t$  and is otherwise governed by

$$\frac{dg_s(t)}{dt} = -\frac{1}{\tau_s} g_s(t). \quad (2.2)$$

Now consider a neuron that has  $g_s$  as a state variable and a probability density to fire of the form

$$h_g(g_s, t) = \lim_{\Delta t \rightarrow 0^+} \frac{\text{prob}\{> 0 \text{ spikes in } [t, t + \Delta t] \mid g_s\}}{\Delta t}, \quad (2.3)$$

where  $g_s$  evolves in time by equation 2.2. This process is analogous to a renewal process, but now with a single state variable,  $g_s$ , which is not reset at each occurrence of a spike but slowly forgets with a timescale of  $\tau_s$  due to equation 2.2. For a model of this form, it is possible for correlations to arise between subsequent ISIs. We refer to both the renewal hazard function,  $\rho(\tau, t)$ , and the  $h_g(g_s, t)$  defined here as hazard functions, as they both represent a probability density of the system to spike.

It is straightforward to show that the ensemble of such neurons is governed by a master equation of the form

$$\begin{aligned} \frac{\partial}{\partial t} P(g_s, t) = & \frac{\partial}{\partial g_s} \left[ \frac{g_s}{\tau_s} P(g_s, t) \right] \\ & + h_g(g_s - q_s, t) P(g_s - q_s, t) \\ & - h_g(g_s, t) P(g_s, t), \end{aligned} \quad (2.4)$$

where  $P(g_s, t)$  is the distribution of state variables  $g_s$  with  $P(g_s < 0, t) \equiv 0$ . The distribution  $P(g_s, t)$  is analogous to the distribution of ages,  $f^-(\tau, t)$ , of renewal theory, and equation 2.4 is analogous to the renewal theory equation C.7, both given in appendix C. The model defined by equation 2.4 is referred to as the 1D Markov (1DM) model throughout the text. (See Table 1 for an overview of the models considered in the text.)

Understanding the connection of the 1DM model to its renewal theory cousin is facilitated by transforming  $g_s$  to a pseudo-age variable  $t_s$  with  $\frac{d}{dt}t_s = 1$  by<sup>2</sup>

$$t_s = \eta(g_s) := -\tau_s \log(g_s/q_s). \quad (2.5)$$

The hazard function  $h_g(g_s, t)$  becomes  $h(t_s, t) = h_g(\eta^{-1}(t_s), t)$ , a hazard function as in equation 2.1 of the pseudovariable  $t_s$  but defined also for  $t_s < 0$ . The distribution of states  $P(g_s, t)$  becomes  $P(t_s, t)$ , where they are related by

$$P(t_s, t) = P(g_s = \eta^{-1}(t_s), t) \frac{d}{dt_s} \eta^{-1}(t_s). \quad (2.6)$$

The reset condition is not  $t_s \mapsto 0$  as for a renewal process, but  $t_s \mapsto \eta(g_s + q_s)$ , where the right-hand side can be expressed in terms of  $t_s$  using the relation  $g_s = \eta^{-1}(t_s)$ . Defining the reset mapping,  $\psi(t_s)$ , such that the reset condition becomes  $t_s \mapsto \psi(t_s)$ , it follows that

$$\begin{aligned} \psi(t_s) &= \eta(\eta^{-1}(t_s) + q_s) \\ &= -\tau_s \log\left(\exp\left(\frac{-t_s}{\tau_s}\right) + 1\right), \end{aligned} \quad (2.7)$$

with its inverse given by

$$\psi^{-1}(t_s) = -\tau_s \log\left(\exp\left(\frac{-t_s}{\tau_s}\right) - 1\right), \quad (2.8)$$

whereby  $\psi(\psi^{-1}(t)) = t$  and  $\psi^{-1}(\psi(t)) = t$  as required by the definition of the inverse.

The variable  $t_s$  is then a general state variable that no longer represents the time since the last spike, as in renewal theory. Since  $\psi : \mathbb{R} \rightarrow \mathbb{R}^-$ , it follows that all trajectories are reinserted at negative pseudo-ages, and it can be seen from the form of  $\psi$  that “younger” spiking trajectories are reinserted at more negative pseudo-ages. This dependence of the reinserted state on the state just prior to spiking yields a Markov process (Risken, 1996), which cannot be described by renewal theory.

---

<sup>2</sup>We follow the convention throughout the text of using positional arguments for functions and labeled arguments for derivatives. Probability distributions are excepted from this rule, as they are not functions but densities. The notation “:=” denotes definition of a function and its positional arguments.



The master equation in terms of  $t_s$  takes the form

$$\begin{aligned} \frac{\partial}{\partial t} P(t_s, t) = & -\frac{\partial}{\partial t_s} P(t_s, t) \\ & + \begin{cases} -h(t_s, t)P(t_s, t), & t_s \geq 0 \\ h(\psi^{-1}(t_s), t)P(\psi^{-1}(t_s), t) - h(t_s, t)P(t_s, t) & t_s < 0, \end{cases} \end{aligned} \quad (2.9)$$

revealing the advantage of the variable transformation  $g_s \rightarrow t_s$ : The deterministic drift term in equation 2.4 for the exponential decay of  $g_s$  is transformed to a constant drift term in  $t_s$  analogous to age in renewal theory. As a result, much can be calculated by analogy to renewal theory, and we are freed from the difficulty of treating the nonconstant drift toward zero in equation 2.4 numerically. We will see in later sections that  $h(t_s, t)$  is in practice approximately of the form

$$h(t_s, t) = a(t) \exp(-b(t)q_s \exp(-t_s/\tau_s)) \quad (2.10)$$

when modeling spike-frequency adapting neurons in the high-conductance state, where  $a(t)$  and  $b(t)$  are determined by the stimulus.

For the static case where  $h(t_s, t) \equiv h(t_s)$ ,  $P(t_s)$  can be found from equation 2.9 by setting  $\partial/\partial t P(t_s, t) = 0$ . The resulting equation for  $t_s \geq 0$ ,

$$\frac{\partial}{\partial t_s} P(t_s) = -h(t_s)P(t_s), \quad (2.11)$$

is exactly as for a renewal process. The solution is the homogeneous survival function,

$$P(t_s) = k\mathcal{W}(t_s, 0), \quad (2.12)$$

where

$$k^{-1} = \int_{-\infty}^{\infty} \mathcal{W}(t_s, 0) dt_s \quad (2.13)$$

is a constant of normalization, and the survival function,

$$\mathcal{W}(\Delta t, t_s^0) = \exp\left(-\int_0^{\Delta t} h(t_s^0 + s) ds\right), \quad (2.14)$$

and analogously the inhomogeneous survival function,

$$\mathcal{W}(\Delta t, t_s^0, t) = \exp\left(-\int_0^{\Delta t} h(t_s^0 + s, t + s) ds\right), \quad (2.15)$$

represent the probability that a system with initial state  $t_s^0 \in \mathbb{R}$  will survive for a time  $\Delta t$ , and a time  $\Delta t$  after  $t$ , respectively and are analogous to the survival function of renewal theory as discussed in appendix C, except for the explicit dependence on the initial state  $t_s^0$ . For  $t_s < 0$ , we solve  $P(t_s)$  numerically by discretizing and integrating back from  $t_s = 0$ .

The distribution of pseudo-ages just prior to spiking at  $t$ ,  $P^*(t_s, t)$ , is related to  $P(t_s, t)$  by

$$P^*(t_s, t) = \frac{h(t_s, t)P(t_s, t)}{\alpha(t)}, \quad (2.16)$$

where

$$\alpha(t) = \int_{-\infty}^{\infty} h(t_s, t)P(t_s, t)dt_s \quad (2.17)$$

is a normalizing constant and also the firing rate of the ensemble.

The distribution of pseudo-ages just after spiking at  $t$ ,  $P^\dagger(t_s, t)$ , is related to  $P^*(t_s, t)$  by transforming variables by the reset mapping (see equation 2.7) for a probability distribution:

$$P^\dagger(t_s, t) = P^*(\psi^{-1}(t_s), t) \frac{d}{dt_s} \psi^{-1}(t_s). \quad (2.18)$$

**2.1 Computing Renewal Quantities.** The various quantities of renewal theory such as the ISI distribution, hazard function, and survival function are of interest and are straightforward to calculate for the proposed Markov process.

First, the renewal survival function,  $\mathcal{F}(\tau, t)$ , the probability that a system that spiked at  $t$  will survive the time interval  $\tau$ , is given by

$$\mathcal{F}(\tau, t) = \int_{-\infty}^{\infty} \mathcal{W}(\tau, t_s, t)P^\dagger(t_s, t)dt_s. \quad (2.19)$$

The ISI distribution,  $f(\tau, t)$ , the probability that a neuron that spiked at  $t$  will survive for an interval  $\tau$  and subsequently spike at  $t + \tau$ , is

$$f(\tau, t) = \int_{-\infty}^{\infty} h(t_s + \tau, t + \tau)\mathcal{W}(\tau, t_s, t)P^\dagger(t_s, t)dt_s. \quad (2.20)$$

Equivalently in terms of  $P^*(t_s, t)$ ,

$$f(\tau, t) = \int_{-\infty}^{\infty} h(\psi(t_s) + \tau, t + \tau) \mathcal{W}(\tau, \psi(t_s), t) P^*(t_s, t) dt_s. \quad (2.21)$$

The hazard function of the system in a renewal sense,  $\rho(\tau, t)$ , where  $\tau$  is a true age, is by definition the firing rate of the subpopulation that previously spiked at time  $t - \tau$ . Thus,

$$\rho(\tau, t) = \int_{-\infty}^{\infty} h(t_s, t) P(t_s, t | \text{spike at } t - \tau) dt_s, \quad (2.22)$$

where the state distribution of the system given a spike at  $t - \tau$ ,  $P(t_s, t | \text{spike at } t - \tau)$ , can be determined by reasoning that it is the distribution of states just after spiking with arguments  $t_s - \tau$  and  $t - \tau$ ,  $P^\dagger(t_s - \tau, t - \tau)$ , which subsequently survive the interval  $\tau$ ,

$$P(t_s, t | \text{spike at } t - \tau) = k_1 \mathcal{W}(\tau, t_s - \tau, t - \tau) P^\dagger(t_s - \tau, t - \tau), \quad (2.23)$$

where  $k_1$  is the normalization factor,

$$k_1^{-1} = \int_{-\infty}^{\infty} \mathcal{W}(\tau, t_s - \tau, t - \tau) P^\dagger(t_s - \tau, t - \tau) dt_s, \quad (2.24)$$

and by inspection of equation 2.19,

$$k_1^{-1} = \mathcal{F}(\tau, t - \tau), \quad (2.25)$$

such that

$$\rho(\tau, t) = \frac{1}{\mathcal{F}(\tau, t - \tau)} \int_{-\infty}^{\infty} h(t_s, t) \mathcal{W}(\tau, t_s - \tau, t - \tau) P^\dagger(t_s - \tau, t - \tau) dt_s. \quad (2.26)$$

Clearly, the numerator is just  $f(\tau, t - \tau)$ , resulting in

$$\rho(\tau, t) = \frac{f(\tau, t - \tau)}{\mathcal{F}(\tau, t - \tau)}. \quad (2.27)$$

This verifies that the standard renewal theory relation that  $f(\tau) = \rho(\tau) \mathcal{F}(\tau)$ , generalized for the inhomogeneous case, still holds even though the underlying stochastic process is not a renewal process. It is interesting to note that in the inhomogeneous case, there is an alternate definition for the ISI distribution that is equally sensible: define  $\hat{f}(\tau, t)$  as the probability that

a neuron that spiked at  $t - \tau$  will survive the interval  $\tau$  and subsequently spike at  $t$ . This is the ISI distribution that treats the spike at  $t$  as the final spike of the ISI rather than the initial spike as in equation 2.21. If one prefers this alternate definition of the ISI distribution, as in Gerstner and Kistler (2002), then one has

$$\hat{f}(\tau, t) = \int_{-\infty}^{\infty} h(t_s + \tau, t) \mathcal{W}(\tau, t_s, t - \tau) P^\dagger(t_s, t - \tau) dt_s, \quad (2.28)$$

implying that  $\hat{f}(\tau, t) = f(\tau, t - \tau)$ , and equation 2.27 becomes

$$\rho(\tau, t) = \frac{\hat{f}(\tau, t)}{\mathcal{F}(\tau, t - \tau)}. \quad (2.29)$$

**2.2 Correlations.** In this section, an expression for the joint serial ISI distribution,  $f(\tau_{i+1}, \tau_i, t)$ , will be derived for the proposed Markov process and shown to exhibit ISI correlations.

Recall the definition of the absence of correlations between two random variables:  $\tau_i$  and  $\tau_{i+1}$  are uncorrelated (independent) if and only if

$$f(\tau_{i+1}, \tau_i) = f(\tau_{i+1})f(\tau_i), \quad (2.30)$$

where  $f(\tau_{i+1}, \tau_i)$  is the joint probability distribution of two back-to-back ISIs in the homogeneous case.

For the inhomogeneous case, a separation of this joint distribution  $f(\tau_{i+1}, \tau_i, t)$  by Bayes' theorem,

$$f(\tau_{i+1}, \tau_i, t) = f(\tau_{i+1}, t | \tau_i) f(\tau_i, t - \tau_i), \quad (2.31)$$

reveals a subtlety: The time argument of  $f(\tau_i, t)$ , the marginal distribution of  $\tau_i$ , must be retarded by  $\tau_i$ . This is due to the fact that for  $\tau_i$  to precede  $\tau_{i+1}$  at  $t$ , it must occur at  $t - \tau_i$ . Given that  $f(\tau, t)$  is known, it is left to determine an expression for  $f(\tau_{i+1}, t | \tau_i)$ . This can be achieved using equation 2.21 by replacing  $P^*(t_s, t)$  with the conditional distribution of states just prior to spiking given a spike at  $t - \tau_i$ , which is denoted by  $P^*(t_s, t | \tau_i)$ .

The distribution  $P^*(t_s, t | \tau_i)$ , the conditional distribution of states just prior to spiking, given a spike at  $t - \tau_i$ , takes the form

$$P^*(t_s, t | \tau_i) = k_2 h(t_s, t) P(t_s, t | \text{spike at } t - \tau_i), \quad (2.32)$$

where  $k_2$  is a normalization factor, and an expression for  $P(t_s, t | \text{spike at } t - \tau_i)$  was given in equation 2.23. By inspection of equation 2.22, it can be seen that  $k_2^{-1} = \rho(\tau_i, t)$ . This results in

$$P^*(t_s, t | \tau_i) = \frac{h(t_s, t) \mathcal{W}(\tau_i, t_s - \tau_i, t - \tau_i) P^\dagger(t_s - \tau_i, t - \tau_i)}{f(\tau_i, t - \tau_i)}, \quad (2.33)$$

where the denominator,  $\rho(\tau_i, t) \mathcal{F}(\tau_i, t - \tau_i)$ , was replaced by  $f(\tau_i, t - \tau_i)$  using equation 2.27.

Plugging this expression for  $P^*(t_s, t | \tau_i)$  into equation 2.21 yields

$$\begin{aligned} f(\tau_{i+1}, \tau_i, t) = \\ f(\tau_{i+1}, t | \tau_i) f(\tau_i, t - \tau_i) = \int_{-\infty}^{\infty} h(\psi(t_s) + \tau_{i+1}, t + \tau_{i+1}) \mathcal{W}(\tau_{i+1}, \psi(t_s), t) \\ \times h(t_s, t) \mathcal{W}(\tau_i, t_s - \tau_i, t - \tau_i) P^\dagger(t_s - \tau_i, t - \tau_i) dt_s, \end{aligned} \quad (2.34)$$

an inhomogeneous expression for the joint ISI distribution of two successive ISIs.

It is instructive to verify that for the case of a renewal process, equation 2.34 predicts no correlations. For a renewal process,  $\psi(t_s) = 0$  and  $P^\dagger(t_s, t) = \delta(t_s)$ , such that equation 2.34 becomes

$$f(\tau_{i+1}, \tau_i, t) = h(\tau_{i+1}, t + \tau_{i+1}) \mathcal{W}(\tau_{i+1}, 0, t) \cdot h(\tau_i, t) \mathcal{W}(\tau_i, 0, t - \tau_i). \quad (2.35)$$

In addition, the ISI distribution given by equation 2.20 reduces to

$$f(\tau, t) = h(\tau, t + \tau) \mathcal{W}(\tau, 0, t). \quad (2.36)$$

Thus, it can be seen by inspection that equation 2.35 is of the form

$$f(\tau_{i+1}, \tau_i, t) = f(\tau_{i+1}, t) f(\tau_i, t - \tau_i), \quad (2.37)$$

implying as expected that successive ISIs are uncorrelated for a renewal process.

### 3 Connection to a Detailed Neuron Model

---

In this section we show that the full five-dimensional master equation for the canonical conductance-based integrate-and-fire neuron model driven by Poisson spike trains, augmented by mechanisms for SFA and a relative refractory period, can be reduced to a two-dimensional generalization of the 1DM model by an adiabatic elimination of fast variables.

**3.1 Neuron Model, Adaptation, Input.** Following Rudolph and Destexhe (2003a, 2005), Richardson (2004), Richardson and Gerstner (2005), we consider the equations for the membrane potential,  $v(t)$ , and excitatory and inhibitory synaptic conductances,  $g_e(t)$  and  $g_i(t)$ , of the conductance-based integrate-and-fire neuron driven by Poisson spike trains:

$$c_m \frac{dv(t)}{dt} = g_l(E_l - v(t)) + g_e(t)(E_e - v(t)) + g_i(t)(E_i - v(t)) \quad (3.1)$$

$$\frac{dg_e(t)}{dt} = -\frac{1}{\tau_e} g_e(t) + q_e \mathcal{S}_e(t) \quad (3.2)$$

$$\frac{dg_i(t)}{dt} = -\frac{1}{\tau_i} g_i(t) + q_i \mathcal{S}_i(t), \quad (3.3)$$

where  $c_m$  represents the membrane capacitance,  $g_l$  the leak conductance,  $E_x$  the various reversal potentials,  $q_x$  the quantal conductance increases, and  $\tau_x$  the synaptic time constants. The exact parameters used are given in appendix A. The excitatory and inhibitory input spike trains,  $\mathcal{S}_x(t)$  with  $x \in \{e, i\}$ , respectively, are given by

$$\mathcal{S}_x(t) = \sum_k \delta(t - s_{x,k}), \quad (3.4)$$

where  $s_{x,k}$  are the spike times of a realization of an inhomogeneous Poisson process (Papoulis & Pillai, 1991). Thus,  $\mathcal{S}_x(t)$  satisfies the constraints

$$\langle \mathcal{S}_x(t) \rangle = v_x(t) \quad (3.5)$$

$$\langle \mathcal{S}_x(t) \mathcal{S}_x(t') \rangle = v_x(t) v_x(t') + v_x(t') \delta(t - t'). \quad (3.6)$$

Here  $v_x(t)$  represents the time-varying rate of the inhomogeneous Poisson process, and  $\langle \rangle$  represents the expectation value over the ensemble of realizations. In what follows, all Poisson processes are assumed inhomogeneous unless otherwise stated.

To put the neuron in a state of high conductance, it is bombarded by  $N_e = 1000$  and  $N_i = 250$  excitatory and inhibitory Poisson processes, all with rate functions  $\lambda_e(t)$  and  $\lambda_i(t)$ , respectively, so that

$$v_x(t) = N_x \lambda_x(t). \quad (3.7)$$

A simple thresholding mechanism approximates the action potential dynamics of real neurons: If  $v(t)$  exceeds the threshold,  $v_{th}$ ,  $v(t)$  is reset to  $v_{reset}$ . Analogous to the input spike train, we can thus define the output

spike train,

$$\mathcal{A}(t) = \sum_k \delta(t - s_k), \quad (3.8)$$

where  $s_k$  are the times of membrane potential threshold crossings enumerated by  $k$ .

SFA and a relative refractory period can both be modeled with the addition of a current to equation 3.1 of the form proposed in Dayan and Abbott (2001),

$$g_y(t)(E_y - v(t)), \quad (3.9)$$

where  $E_y$  is a reversal potential. The conductance  $g_y(t)$  is governed by

$$\frac{dg_y(t)}{dt} = -\frac{1}{\tau_y}g_y(t) + q_y\mathcal{A}(t), \quad (3.10)$$

where  $\tau_y$  and  $q_y$  are the time constant and quantal conductance increase of the mechanism. We label SFA and the relative refractory mechanism by the subscripts  $y = s$  and  $y = r$ , respectively. Defining

$$\beta_v(v, g_e, g_i, g_s, g_r) := g_l(E_l - v) + \sum_{\mu=e,i,s,r} g_\mu(E_\mu - v) \quad (3.11)$$

and for  $\mu = e, i, s, r$ ,

$$\beta_\mu(g_\mu) := -\frac{1}{\tau_\mu}g_\mu, \quad (3.12)$$

the five-dimensional system of coupled differential equations describing the conductance-based spike-frequency adapting relative refractory integrate-and-fire neuron driven by Poisson spike trains is:

$$c_m \frac{dv(t)}{dt} = \beta_v(v(t), \dots, g_r(t)) - (V_{\text{th}} - V_{\text{reset}})\mathcal{A}(t) \quad (3.13)$$

$$\frac{dg_x(t)}{dt} = \beta_x(g_x(t), t) + q_x\mathcal{S}_x(t) \quad (3.14)$$

$$\frac{dg_y(t)}{dt} = \beta_y(g_y(t), t) + q_y\mathcal{A}(t), \quad (3.15)$$

where  $x \in \{e, i\}$  and  $y \in \{s, r\}$ . We refer to equations 3.13 to 3.15 as the full five-dimensional (5DF) model throughout the text (see the model overview in Table 1). The parameters used are given in Table 3.

**3.2 Ensemble Behavior.** It is natural to look for an ensemble description of equations 3.13 to 3.15, given that the input is described in terms of an ensemble.

Equations 3.13 to 3.15 are a set of concurrent first-order differential equations, that is, the right-hand sides at time  $t$  are functions of the instantaneous values of the state variables,  $(v(t), g_e(t), g_i(t), g_s(t), g_r(t))$ , implying no delays or memory effects are to be modeled. The system is therefore a Markov process, and given an initial distribution  $P(v, g_e, g_i, g_s, g_r, t_0)$  for some  $t_0$ , the evolution of  $P(v, g_e, g_i, g_s, g_r, t)$  can be described by a suitable master equation (Risken, 1996). For the system in question here, the master equation takes the form

$$\begin{aligned} \frac{\partial}{\partial t} P(v, g_e, g_i, g_s, g_r, t) = & -\text{div} J(v, g_e, g_i, g_s, g_r, t) \\ & + \delta(v - v_{\text{reset}}) J_v(v_{\text{th}}, g_e, g_i, g_s - q_s, g_r - q_r, t), \end{aligned} \quad (3.16)$$

where the probability current density,  $J$ , is a vector with components

$$J_v(v, g_e, g_i, g_s, g_r, t) = \beta_v(v, g_e, g_i, g_s, g_r, t) P(v, g_e, g_i, g_s, g_r, t) \quad (3.17)$$

$$J_\mu := \beta_\mu(g_\mu, t) P(v, g_e, g_i, g_s, g_r, t) \quad (3.18)$$

with  $\mu \in \{s, r\}$ . (For  $J_e$  and  $J_i$ , see appendix B.) The  $\delta$  term in equation 3.16 implements the reinsertion of probability flux that crosses the threshold. Furthermore, we define  $P(v, g_e, g_i, g_s, g_r, t) = 0$  if one or more of the conductances  $g_e, \dots, g_r$  is negative.

There exists a wealth of literature treating master equations of conductance and current-based integrate-and-fire neuron models in the absence of adaptation and relative refractory mechanisms (Knight, 1972; Gerstner, 1995; Brunel, 2000; Omurtag et al., 2000; Nykamp & Tranchina, 2000; Knight, Omurtag, & Sirovich, 2000; Gerstner, 2000; Fourcaud & Brunel, 2002; Rudolph & Destexhe, 2003a; Richardson, 2004; Richardson & Gerstner, 2005). The usual approach is to make the so-called diffusion approximation yielding generally a Fokker-Planck equation for the membrane potential, and perhaps one or two other dimensions treating synaptic conductances.

We present here a novel approach applicable for neurons in the high-conductance state whereby the variables  $v, g_e, g_i$  are eliminated by a technique known as an adiabatic elimination of fast variables (Haken, 1983; Gardiner, 1984), and the system is reduced to a master equation for the two-dimensional marginal probability distribution,  $P(g_s, g_r, t)$ , of the slow variables,  $g_s$  and  $g_r$ . As we will see, the membrane potential,  $v$ , and the synaptic conductances,  $g_e$  and  $g_i$ , are thus encapsulated in the hazard function,  $h_g(g_s, g_r, t)$ . We treat here the static input case,  $\lambda_e, \lambda_i$ . The case for dynamic external input  $\lambda_e(t), \lambda_i(t)$  is treated in section 4.



We follow here the intuitive treatment of adiabatic elimination given in Haken (1983). We begin by integrating  $P(v, \dots, g_r)$  over the fast variables  $v, g_e, g_i$ , yielding the marginal distribution for the slow variables  $g_s, g_r$ ,

$$P(g_s, g_r, t) = \int_0^\infty \int_0^\infty \int_{-\infty}^{v_{\text{th}}} P(v, g_e, g_i, g_s, g_r, t) dv dg_e dg_i. \quad (3.19)$$

Integrating equation 3.16 over  $v, g_e, g_i$  yields

$$\begin{aligned} \frac{\partial}{\partial t} P(g_s, g_r, t) = & - \sum_{\mu=s,r} \frac{\partial}{\partial g_\mu} (\beta_\mu(g_\mu) P(g_s, g_r, t)) \\ & - \int_0^\infty \int_0^\infty \beta_v(v_{\text{th}}, g_e, g_i, g_s, g_r) P(v_{\text{th}}, g_e, g_i, g_s, g_r, t) dg_e dg_i \\ & + \int_0^\infty \int_0^\infty \beta_v(v_{\text{th}}, g_e, g_i, g_s - q_s, g_r - q_r) \\ & \times P(v_{\text{th}}, g_e, g_i, g_s - q_s, g_r - q_r, t) dg_e dg_i. \end{aligned} \quad (3.20)$$

For details of the calculation, see appendix B. Now we separate the marginal distribution for the slow variables from the full distribution by Bayes' theorem, resulting in

$$P(v, g_e, g_i, g_s, g_r, t) = P(v, g_e, g_i, t | g_s, g_r, t) P(g_s, g_r, t), \quad (3.21)$$

and make the adiabatic approximation as in Haken (1983) that

$$P(v, g_e, g_i, t | g_s, g_r, t) \approx P^{(g_s, g_r)}(v, g_e, g_i, t), \quad (3.22)$$

where  $P^{(g_s, g_r)}(v, g_e, g_i, t)$  is the solution to the three-dimensional master equation for the canonical conductance-based integrate-and-fire neuron with a constant bias current,  $I(g_s, g_r) = g_s(E_s - v) + g_r(E_r - v)$ , with neither SFA nor the relative refractory mechanism. This implies we assume that  $v, g_e, g_i$  are immediately at equilibrium given the slow variables, or in other words, the system responds adiabatically to the dynamics of the slow variables  $g_s, g_r$ . The adiabatic assumption ensures the two-dimensional process  $(g_s(t), g_r(t))$  is a Markov process.

Now defining the hazard function,

$$h_g(g_s, g_r, t) := \int_0^\infty \int_0^\infty \beta_v(v_{\text{th}}, g_e, g_i, g_s, g_r) P^{(g_s, g_r)}(v_{\text{th}}, g_e, g_i, t) dg_e dg_i, \quad (3.23)$$

the master equation, 3.20, becomes

$$\begin{aligned} \frac{\partial}{\partial t} P(g_s, g_r, t) = & - \sum_{\mu=s,r} \frac{\partial}{\partial g_\mu} (\beta_\mu(g_\mu) P(g_s, g_r, t)) \\ & - h_g(g_s, g_r, t) P(g_s, g_r, t) \\ & + h_g(g_s - q_s, g_r - q_r, t) P(g_s - q_s, g_r - q_r, t). \end{aligned} \quad (3.24)$$

We refer to the model defined by equation 3.24 as the 2D Markov (2DM) model throughout the text (see the model overview in Table 1). Since no analytical solution is yet known for  $P^{(g_s, g_r)}(v_{\text{th}}, g_e, g_i, t)$  in equation 3.23,  $h_g(g_s, g_r)$  was extracted from Monte Carlo simulations of equations 3.13 to 3.15, as will be discussed in section 4.1. Then given a solution to the master equation,  $P(g_s, g_r, t)$ , the firing rate of the ensemble, denoted by  $\alpha(t)$ , is determined by the expectation value of the hazard function  $h_g(g_s, g_r, t)$  over  $P(g_s, g_r, t)$ :

$$\alpha(t) = \int_0^\infty \int_0^\infty h_g(g_s, g_r, t) P(g_s, g_r, t) dg_s dg_r. \quad (3.25)$$

For the adiabatic approximation, the assumption that  $g_s$  is slow compared to  $v, g_e, g_i$  is easily justified as the timescale of  $g_s$  is on the order of 100 ms, while the timescale of  $v$  is on the order of 2 ms in the high-conductance state. The timescale of the mean and standard deviation of  $g_e$  and  $g_i$  are on the order of  $\tau_e = 1.5$  ms and  $\tau_i = 10$  ms, respectively, while the fluctuations of  $g_e$  and  $g_i$  are the source of stochasticity of the system and are on a still shorter timescale.

The timescale of  $g_r$  is significantly faster than  $g_s$ , though its treatment as a slow variable is also justifiable, but in a somewhat indirect manner. As has been argued in Fourcaud and Brunel (2002) and Renart et al. (2004), for neurons with synaptic time constants comparable to or larger than the effective membrane time constant and driven by sufficient input noise, as is the case here, the firing rate follows the input current almost instantaneously. It is this property that allows the dynamic firing rate to be treated as a function of the time-dependent means and variances of the synaptic conductances in La Camera et al. (2004), a method we follow in section 4. This suggests that such modulations do not push the system far from equilibrium and that the system returns to equilibrium on a timescale faster than that of the synaptic means ( $\tau_e, \tau_i$ ). Since over the domain of the  $g_r$  trajectory for which the integrals on the right-hand side of equation 3.20 are nonzero,  $g_r$  has a timescale comparable to the mean of the synapses, the argument applies equally to  $g_r$ . However, since  $g_r$  is spike triggered, we leave  $g_r$  in the master equation, while the synaptic variables,  $g_e$  and  $g_i$ , determine  $h_g(g_s, g_r, t)$  and can be treated outside the master equation formalism.

Methods to undertake a rigorous analysis of the error in the adiabatic approximation are beyond the scope of this letter. What follows are a variety of numerical comparisons to demonstrate the accuracy and domain of applicability of the proposed approximation.

## 4 Methods

---

In this section we provide methods for determining appropriate homogeneous and inhomogeneous hazard functions for the 1DM, 2DM, and renewal models. Since no analytical expression for equation 3.23, or the renewal hazard function of the 5DF model is yet known, we approach the problem by fitting the homogeneous hazard functions determined by 5DF Monte Carlo simulations in the static case. The inhomogeneous functions are then constructed from the homogeneous ones by discretizing time and taking one homogeneous hazard function for the duration of a single time bin.

### 4.1 Determining the Static Hazard Function for Markov Models.

Given a finite subset of the possible realizations of the Poisson input spike trains, the 5DF model equations, 3.13 to 3.15, can be integrated for each input realization. Any statistical quantity of interest can then be approximated by averaging or histogramming over this finite set of trajectories. This approach is known as the Monte Carlo method. By increasing the number of trials in this finite set of realizations, the statistical quantities determined by the Monte Carlo method converge to the true quantities. Therefore, Monte Carlo simulations are used for determining the unknown hazard functions as well as later benchmarking the reduced master equations.

By Monte Carlo simulations of the 5DF model under static stimulation, the quantities  $P^*(g_s + g_r)$ ,  $P(g_s + g_r)$ , and  $\alpha(t)$  can be obtained. Then analogous to equation 2.16, we can determine  $h_g(g_s, g_r)$  by

$$h_g(g_s, g_r) = h_g(g_s + g_r) = \frac{\alpha P^*(g_s + g_r)}{P(g_s + g_r)}, \quad (4.1)$$

where we treat the sum of the conductances,  $g_s + g_r$ , rather than each independently because we have chosen their reversal potentials to be equal (see appendix A). It was found that  $h_g(g_s, g_r)$  can be fit well by a function of the form

$$h_g(g_s, g_r) = a \exp(-b \cdot (g_s + g_r)), \quad (4.2)$$

where  $a$  and  $b$  are fit parameters. Some typical fits for various excitatory Poisson input rates are shown in Figure 1. For the 1DM model, the same fit parameters were used, but with  $g_r = 0$ . Transforming to  $(t_s, t_r)$  by the

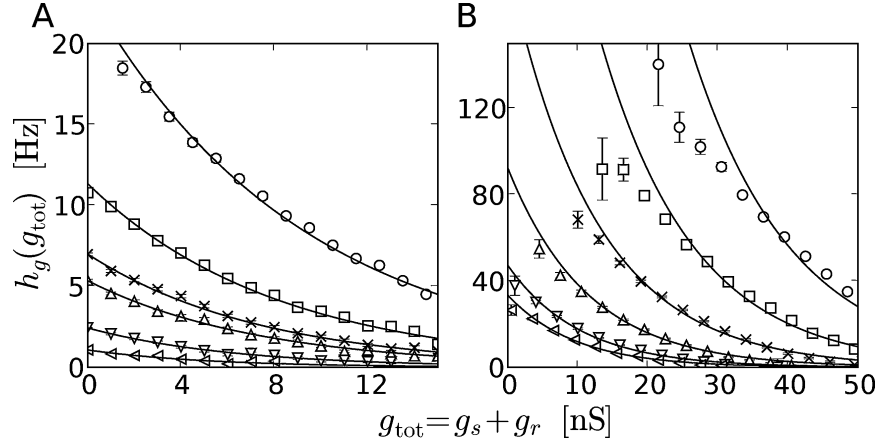


Figure 1:  $h_g(g_s, g_r) = h_g(g_s + g_r)$  as a function of  $g_{\text{tot}} = g_s + g_r$ , as determined from 5DF Monte Carlo (data points, 1000 trials per  $\lambda_e$ , 10 s per trial,  $dt = 0.01$  ms) by equation 4.1, was found to be approximately exponential for a range of excitatory stimulation rates,  $\lambda_e$ , with the inhibitory stimulation rate fixed at  $\lambda_i = 11.4$  Hz. For the definition of the 5DF model, see Table 1. The exponential fits (lines) are good for low rates ( $\leq \lambda_e = 5.26$  Hz,  $\nabla$ :  $\lambda_e = 5.56$  Hz,  $\triangle$ :  $\lambda_e = 5.88$  Hz,  $\times$ :  $\lambda_e = 6.01$  Hz,  $\square$ :  $\lambda_e = 6.25$  Hz,  $\circ$ :  $\lambda_e = 6.67$  Hz) in A, but poorer for  $g_s$  near zero for high rates ( $\leq \lambda_e = 6.90$  Hz,  $\nabla$ :  $\lambda_e = 7.14$  Hz,  $\triangle$ :  $\lambda_e = 7.69$  Hz,  $\times$ :  $\lambda_e = 8.33$  Hz,  $\square$ :  $\lambda_e = 9.09$  Hz,  $\circ$ :  $\lambda_e = 10.0$  Hz) in B.

inverse of equation 2.5, we have

$$h(t_s, t_r) = a \exp \left( -b \cdot \left( q_s \exp \left( \frac{-t_s}{\tau_s} \right) + q_r \exp \left( \frac{-t_r}{\tau_r} \right) \right) \right). \quad (4.3)$$

**4.2 Constructing Inhomogeneous Hazard Functions.** Now given the hazard functions determined under static input statistics, the inhomogeneous hazard function given time-varying Poisson input rates  $\lambda_e(t)$ ,  $\lambda_i(t)$  can be constructed by accounting for synaptic filtering.

The homogeneous hazard functions given static stimulation rates  $\lambda_e$ ,  $\lambda_i$  determined by the recipes in section 4.1 are the hazard functions given synaptic conductance distributions parameterized by  $\langle g_{e,i} \rangle$ , neglecting higher-order moments. It can be shown that

$$\frac{d}{dt} \langle g_x(t) \rangle = -\frac{1}{\tau_x} (\langle g_x(t) \rangle - q_x \tau_x N_x \lambda_x(t)), \quad (4.4)$$

with  $x \in \{e, i\}$ , a low-pass filter equation of the quantity  $q_x \tau_x N_x \lambda_x(t)$  with a cutoff frequency of  $2\pi/\tau_x$  (Gardiner, 1985; La Camera et al., 2004).

As argued in Fourcaud and Brunel (2002) and Renart et al. (2004), the firing rate of neurons with nonzero synaptic time constants driven by

sufficient noise follows their input currents instantaneously. Then the hazard function  $h_g(g_s, g_r, t)$  here is also determined instantaneously by the mean synaptic conductances. Therefore, inhomogeneous parameters  $a(t)$  and  $b(t)$  in equation 4.3 can be determined by interpolating the parameters determined from static  $\langle g_e \rangle$  and  $\langle g_i \rangle$  with the instantaneous dynamic  $\langle g_e(t) \rangle$  and  $\langle g_i(t) \rangle$  determined by integrating equation 4.4 for some given arbitrary time-varying input spike trains parameterized by  $\lambda_e(t)$ ,  $\lambda_i(t)$ . Thus, we have the hazard function

$$h_g(g_s, g_r, t) = a(t) \exp(-b(t) \cdot (g_s + g_r)). \quad (4.5)$$

A similar approach was taken in La Camera et al. (2004), except that we do not correctly account for the dynamics of the standard deviation of the synaptic conductance by the fitting approach used here. This could be remedied given an analytically solvable neuron model as was used in La Camera et al.

In this study, we investigate only time-varying excitation. Treating inhibition in addition would require additional fits and two-dimensional interpolation of the resulting parameters but would yield no new results for this study.

**4.3 Comparing to Renewal Theory Models.** In inhomogeneous renewal theory, only the time since the last spike (age) enters into the hazard function (Gerstner & Kistler, 2002). While such theories cannot account for ISI correlations due to SFA, they can account for much of the gradual increase in excitability that follows a spike due to SFA by an appropriately chosen hazard function. Perhaps surprisingly, such models are sufficient to exhibit “adapting” transients to step stimuli. Like the 2DM model, we seek to calibrate such renewal models to the 5DF system and assess their suitability for modeling the ensemble firing rate under dynamic stimuli. Sufficient for such a comparison is a recipe for specifying the hazard function as a function of the static stimulus. The dynamic case can then be constructed by using the effective synaptic filtered stimulus to determine the inhomogeneous hazard function at each instant in time, as for the Markov models in the previous section.

For the static case, one can determine the hazard function as a function of the stimulus by interpolating the ISI distribution due to 5DF Monte Carlo and applying the standard renewal theory result that

$$\rho(\tau) = \frac{f(\tau)}{\mathcal{F}(\tau)}, \quad (4.6)$$

where the renewal theory survival function is given by

$$\mathcal{F}(\tau) = \int_{s=\tau}^{\infty} f(s) ds. \quad (4.7)$$

The renewal model will thus reproduce the ISI distribution of 5DF Monte Carlo under static stimulation. This process is numerically unstable for large  $\tau$ , and for the dynamic case too costly. Another approach is to determine the renewal hazard function by equation 2.26, with one caveat: since the resulting renewal hazard function must be uniquely determined by the stimulus,  $P(t_s, t_r, t)$  in equation 2.26 must be replaced by  $P_{\infty}(t_s, t_r, t)$ , the instantaneous equilibrium distribution, or asymptotic state distribution for large time resulting from a  $h(t_s, t_r, t)$  fixed at the instantaneous value at time  $t$ . The renewal hazard function determined by this recipe, combined with the renewal master equation C.7, defines what we subsequently refer to as the effective renewal (ER) model (see the model overview in Table 1). Typical hazard functions are shown in Figure 2. Indeed, the renewal hazard functions determined by equation 2.26 are consistent with those of 5DF Monte Carlo determined by equation 4.6.

Simulation of the ER model implies that the master equation for  $P(t_s, t_r, t)$  must be allowed to converge to  $P_{\infty}(t_s, t_r, t)$  for each time step where the stimulation changes. This is costly and makes the ER models much less efficient to simulate than the 1DM and 2DM models, but allows a direct comparison of renewal models with the 1DM and 2DM models and 5DF Monte Carlo. When the renewal hazard function is known a priori, as would be the case for a gamma renewal process, or when the hazard functions can be fit by simple functions, the renewal theory ensemble equations given in appendix C are comparatively more efficient to simulate than the 1DM and 2DM models.

## 5 Numerics

---

In this section we describe the numerical techniques applied to solve the 1DM and 2DM master equations, generate realizations of the 1DM and 2DM processes, and solve the 5DF neuron model equations.

**5.1 Numerical Solution of Master Equations.** We solved the 1DM and 2DM master equations numerically by discretizing  $P(t_s, t)$  and  $P(t_s, t_r, t)$ , respectively, applying the exponential Euler method for the death term, and reinserting the lost probability by walking the negative time domain and fetching the probability sources of each bin determined by equation 2.8. We present the one-dimensional case here, which can be generalized to two dimensions.

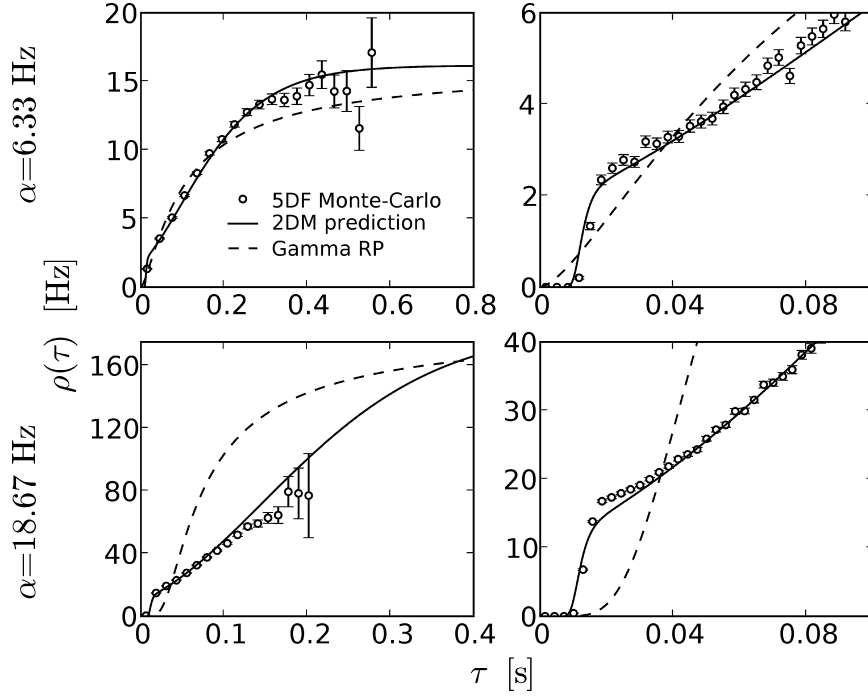


Figure 2: The renewal hazard function,  $\rho(\tau)$ , for synaptic input rates ( $\lambda_e = 6.5$  Hz,  $\lambda_i = 11.4$  Hz) and ( $\lambda_e = 8.3$  Hz,  $\lambda_i = 11.4$  Hz) resulting in an ensemble firing rate of  $\alpha = 6.33$  Hz (top row), and  $\alpha = 18.67$  Hz (bottom row), respectively. The renewal hazard function for 5DF Monte Carlo (circles) was computed by equation 4.6 with a spike train of  $10^4$  s. The renewal hazard function due to the 2DM model (solid line) was determined by equation 2.26. The renewal hazard function for a gamma renewal process (dashed line) equal to the 2DM renewal hazard function at large  $\tau$  and with the same average firing rate was computed as discussed in section C.2. The small  $\tau$  region is shown blown up in the right column. For the definition of the 2DM and 5DF models, see Table 1.

We discretize  $P(t_s, t)$  on equally spaced grids  $t_s^i$  and  $t^j$  with grid spacings  $\Delta t_s := t_s^{i+1} - t_s^i$  and  $\Delta t := t^{j+1} - t^j$ , respectively, with  $\Delta t_s = \Delta t$ , such that  $P(t_s, t) \rightarrow P^{i,j}$ . The discretized form of the master equation 2.9 then takes the form

$$P^{i+1,j+1} = P^{i,j} \exp(-\Delta t \cdot h(t_s^i, t^j)) + P_r^{i,j}, \quad (5.1)$$

where the first term is the exponential Euler computation of loss of probability due to the death term. On the left-hand side, the first superscript of  $P$ ,  $i + 1$ , leads  $i$  by one to implement the constant drift of  $t_s$ ,  $dt_s/dt = 1$ . The

Table 1: Overview of the Models and Quick Reference for the Equations and Sections for the Models.

Model	Description	Keywords	Defined in (section)	Equations	Calibration to 5DF Recipe in (section)	Numerics in (section)
1DM	One-dimensional Markov process	Beyond renewal theory, spike-frequency adaptation, master equation, ensemble	2	2.4, 2.9	4.1, 4.2	5.1 (master equation), 5.3 (realizations)
5DF	Conductance-based spike-frequency adapting relative refractory integrate-and-fire neuron driven by Poisson spike trains	Full five-dimensional system, Monte Carlo, reference, benchmark	3.1, A	3.13–3.15	NA	5.2
2DM	Two-dimensional Markov process	Adiabatic elimination (of 5DF), spike-frequency adaptation, relative refractory period, master equation, ensemble	3.2	3.24	4.1, 4.2	5.1 (master equation), 5.3 (realizations)
ER	Effective renewal theory model	Inhomogeneous, master equation	C	C.7	4.3, 4.2	C.1 (master equation), C.2 (realizations)



reinserted probability,  $P_r^{i,j}$ , is computed for  $t_s^i + \Delta t_s < 0$  by

$$\begin{aligned}
 P_r^{i,j} := & \sum_{m=i_{\text{rif}}(t_s^i)}^{i_{\text{rif}}(t_s^{i+1})-1} P_d^{m,j} \\
 & + \frac{\psi^{-1}(t_s^{i+1}) - i_{\text{rif}}(t_s^{i+1})}{\Delta t_s} P_d^{i+1,j} \\
 & - \frac{\psi^{-1}(t_s^i) - i_{\text{rif}}(t_s^i)}{\Delta t_s} P_d^{i,j},
 \end{aligned} \tag{5.2}$$

where  $P_d^{i,j}$  is the loss of probability computed by

$$P_d^{i,j} := \Delta t \cdot P^{i,j} \cdot h(t_s^i, t^j), \tag{5.3}$$

and  $i_{\text{rif}}$  refers to the “reinserted-from” index, which satisfies

$$t_s^{i_{\text{rif}}(t_s^i)} \leq \psi^{-1}(t_s^i) < t_s^{i_{\text{rif}}(t_s^i)} + \Delta t_s. \tag{5.4}$$

The first term in equation 5.2 is just a sum of all  $P_d^{i,j}$  except the fractional last bin, which sends probability to the interval  $t \in [t_s^i, t_s^i + \Delta t_s)$ . The second two terms subtract the fractional first and add the fractional last bins of  $P_d^{i,j}$ , which are reinserted to the interval, and thus implement a sort of anti-aliasing of the reinsertion mapping.

**5.2 Neuron Simulations.** Monte Carlo simulations of the full-system (5DF Monte Carlo) were performed by solving equations 3.13 to 3.15 using the NEST simulation environment (Diesmann & Gewaltig, 2002) with a time step of 0.01 ms.

**5.3 Generating Realizations of the Proposed Markov Processes.** Generating realizations of the proposed 1DM or 2DM processes is straightforward: The thinning method for a general hazard function described in Devroye (1986) can be applied. The quantity  $h_{\text{max}} = \max_{t_s, t} (h(t_s, t))$  for the 1DM case or  $h_{\text{max}} = \max_{t_s, t_r, t} (h(t_s, t_r, t))$  for the 2DM case must be known. The variables  $t$  and  $t_s$  (1DM), or  $t, t_s$ , and  $t_r$  (2DM) are required and can have initial values of zero. Sequential intervals are generated using a homogeneous Poisson process with hazard rate  $\rho = h_{\text{max}}$ . Given one such interval,  $\Delta t_i$ , it is added to  $t$  and  $t_s$  (1DM), or  $t, t_s$ , and  $t_r$  (2DM). Next, a spike is generated at time  $t$  with probability  $h(t_s, t)/h_{\text{max}}$  (1DM) or  $h(t_s, t_r, t)/h_{\text{max}}$  (2DM), and if a spike is generated,  $t_s \mapsto \psi_s(t_s)$ , and  $t_r \mapsto \psi_r(t_r)$ , where  $\psi_s$  and  $\psi_r$  refer to the reinsertion mappings as in equation 2.7 with the respective parameters for the two mechanisms.

## 6 Results

---

In this section we compare ISI distributions, static ISI correlations, and firing rate dynamics of the 1DM, 2DM, and ER models to 5DF Monte Carlo.

**6.1 Interspike Interval Distributions.** The predictions due to the ER and 2DM model are in excellent agreement with the static ISI distribution of 5DF Monte Carlo. The prediction due to the 1DM model neglects refractory effects and is therefore poor for low ISIs, as can be seen in Figure 3.

**6.2 Interspike Interval Correlations.** In this section we investigate correlations between subsequent ISIs, a feature of the proposed 1DM and 2DM models that is absent by definition in renewal theory models of spike statistics.

The correlation coefficient,  $r$ , for a finite number of observations is defined by

$$r^2 = \frac{(\sum(x_i - \bar{x})(y_i - \bar{y}))^2}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}, \quad (6.1)$$

and is the standard measure by which to quantify correlations between two random variables  $x, y$ , where  $x_i, y_i$  denote the individual observations and  $\bar{x}, \bar{y}$  denote the means.

The correlation coefficients of subsequent ISIs under static stimulation were calculated for 100 runs of 100 s, and the mean and deviation in the mean are given in Table 2. Indeed, the renewal process shows no ISI correlations. For low and high firing rates, the difference between the correlation coefficients for 5DF Monte Carlo and realizations of the 2DM model is consistent with zero. Both exhibit negative ISI correlations, implying short ISIs are generally followed by long ISIs and vice versa, as has been observed in previous experimental and theoretical studies (Longtin & Racicot, 1997; Chacron, Longtin, & Maler, 2000; Chacron et al., 2005; Nawrot et al., 2007).

The conditional ISI distribution,  $f(\tau_{i+1}|\tau_i)$  can be computed for the 1DM and 2DM models by equation 2.34. Predictions due to the 2DM model are in agreement with 5DF Monte Carlo for low and high rates, and both long and short  $\tau_i$ , as shown in Figure 3. Applying equation 2.34, we can compute the quantity

$$\langle \tau_{i+1} | \tau_i \rangle_{\tau_{i+1}} = \int_0^\infty \tau_{i+1} f(\tau_{i+1} | \tau_i) d\tau_{i+1}. \quad (6.2)$$

As discussed in Whittaker and Robinson (1967), this is a linear function of  $\tau_i$  for normal distributions, the slope of which is the correlation coefficient. As the ISI distributions here are not normal, there are deviations from linearity,

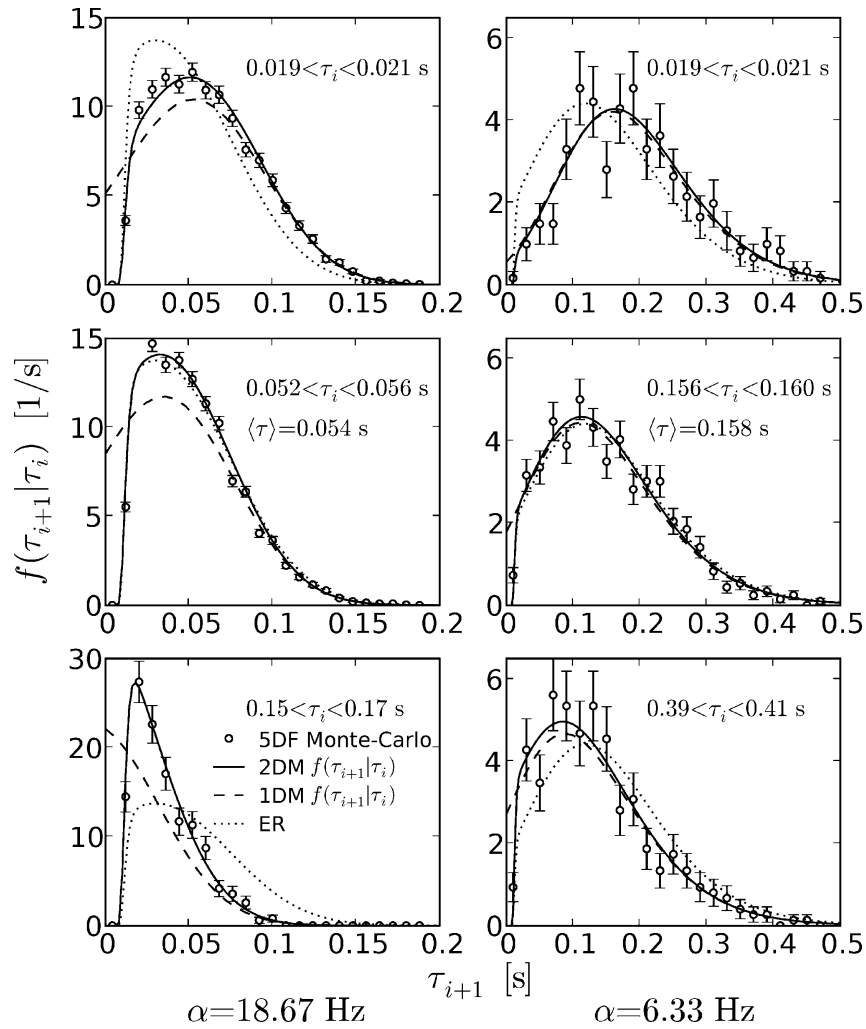


Figure 3: Comparison of the conditional ISI distributions due to 5DF Monte Carlo with predictions due to effective renewal theory (ER, dotted line), the 1DM model (dashed line, determined by equation 2.34), and the 2DM model (solid line, determined by the 2D generalization of equation 2.34). For the definition of the 1DM, 2DM, 5DF, and ER models, see Table 1. The left column shows three representative conditional ISI distributions for an ensemble firing rate of  $\alpha = 18.67$  Hz ( $\lambda_e = 8.3$  Hz,  $\lambda_i = 11.4$  Hz), and the right column shows the same for  $\alpha = 6.33$  Hz ( $\lambda_e = 6.5$  Hz,  $\lambda_i = 11.4$  Hz). The upper two plots show the conditional ISI distribution for  $\tau_i$  much shorter than the mean. The middle two plots show the conditional ISI distribution for  $\tau_i$  equal to the mean. The lower two plots show the conditional ISI distribution for  $\tau_i$  much longer than the mean. The preceding ISI,  $\tau_i$ , is given on each plot, and a small interval around  $\tau_i$  is used to compute the distributions from 5DF Monte Carlo to yield sufficient statistics. The theoretical predictions of the conditional ISI distributions using the 2DM model are in good agreement with 5DF Monte Carlo for all situations considered. The ISI distribution due to 5DF Monte Carlo is consistent with the renewal ISI distribution only when the preceding ISI is equal to the mean ISI (middle row). Spike trains of duration  $10^4$  s were used. Error bars represent the relative error in the histogram bin counts,  $1/\sqrt{\text{count}}$ .

Table 2: Serial ISI Correlation Coefficients for Monte Carlo Simulations of the Full Five-Dimensional System (5DF), Realizations of the One- and Two-Dimensional Markov Models (1DM, 2DM), and Realizations of the Effective Renewal Theory Model (ER).

Model	Correlation coefficient
$\alpha = 6.33$ Hz	
5DF	$-0.148 \pm 0.004$
2DM	$-0.147 \pm 0.003$
1DM	$-0.160 \pm 0.003$
ER	$0.0042 \pm 0.0043$
$\alpha = 18.67$ Hz	
5DF	$-0.235 \pm 0.002$
2DM	$-0.236 \pm 0.002$
1DM	$-0.283 \pm 0.002$
ER	$0.001 \pm 0.002$

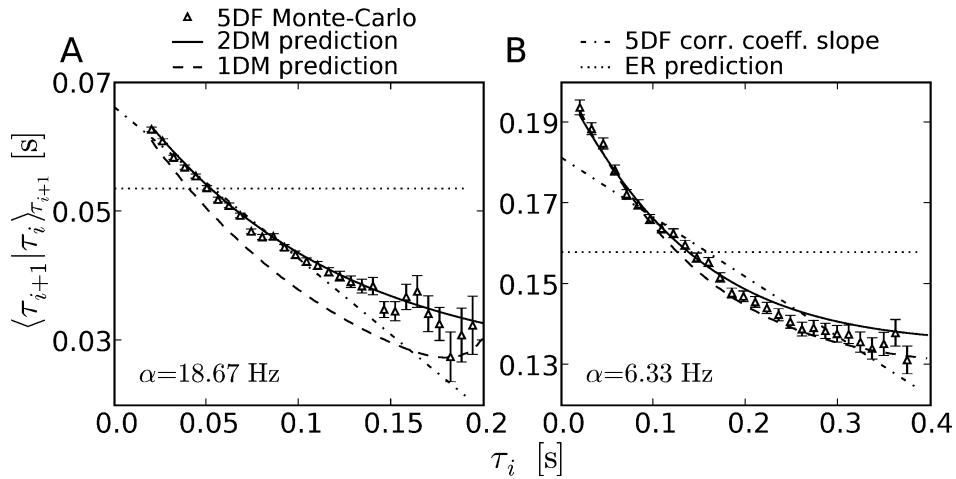


Figure 4: Mean of the conditional ISI distribution as a function of the preceding ISI,  $\tau_i$ , for high-ensemble firing rates (A, ( $\lambda_e = 8.3$  Hz,  $\lambda_i = 11.4$  Hz),  $\alpha = 18.67$  Hz) and low-ensemble firing rates (B, ( $\lambda_e = 6.5$  Hz,  $\lambda_i = 11.4$  Hz),  $\alpha = 6.33$  Hz). The data points (triangles) shown are for 5DF Monte Carlo. Theoretical predictions due to the 1DM (dashed line), 2DM (solid line), and ER (dotted line) models are shown. For the definition of the 1DM, 2DM, 5DF, and ER models, see Table 1. A linear function with slope equal to the serial ISI correlation coefficient would be the functional form if the ISI distributions were normal. Thus, these linear functions are plotted for comparison.

as shown in Figure 4. Predictions due to equation 6.2 for the 2DM model are consistent with 5DF Monte Carlo for both low and high rates, as seen in Figure 4. Thus, the 2DM model is indistinguishable from the full system when considering static correlations.

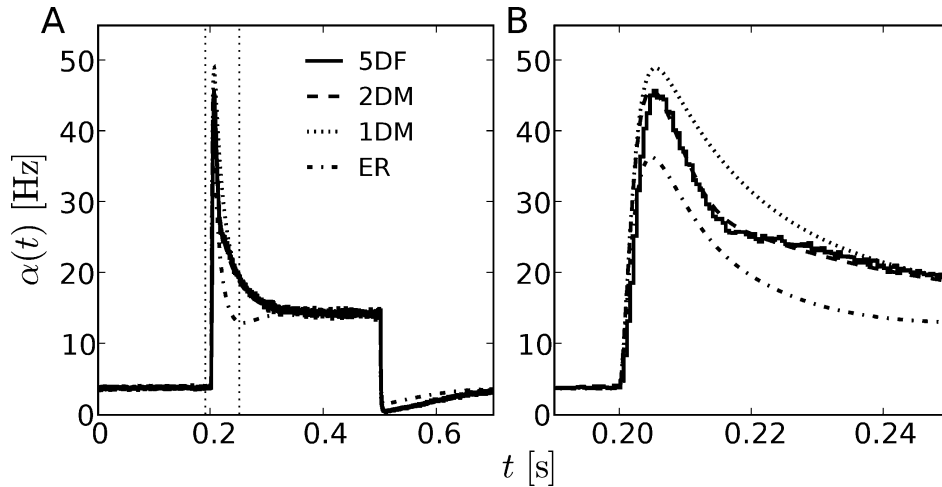


Figure 5: (A) The ensemble firing rate,  $\alpha(t)$ , in response to a moderate step stimulus, determined by 5DF Monte Carlo ( $5 \cdot 10^5$  trials, solid line), and numerical solution of the 1DM (dotted line), 2DM (dashed line), and ER (dash-dotted line) master equations. For the definition of the 5DF, 1DM, 2DM, and ER models, see Table 1. (B) The region outlined by the dashed rectangle is enlarged, showing consistency between the two-dimensional Markov (2DM) model and the full system (5DF Monte Carlo) apart from a 0.5 ms lead of the 2DM solution. This discrepancy is likely due to the neglected membrane potential dynamics.

**6.3 Firing Rate Dynamics.** In this section, we compare the ensemble firing rates of the 1DM, 2DM, and ER models to 5DF Monte Carlo for small to large step stimuli, and for random fluctuating stimuli generated by an Ornstein-Uhlenbeck process.

We subject the neural ensemble to a dynamic stimulus by specifying a time-varying excitatory Poisson input rate,  $\lambda_e(t)$ . Given the time-dependent hazard function determined by the Poisson input rates as described in section 4.2, the ensemble firing rate,  $\alpha(t)$ , of the 1DM and 2DM models can be calculated by solving equations 2.9 and 3.24, respectively, for the time-dependent state distribution, and subsequently applying equation 2.17 or the 2D generalization of it. For the ER model, the hazard function was calculated by the methods discussed in section 4.3, and the ensemble firing rate was determined by solving equation C.7.

For weak step stimuli that do not bring the system too far from equilibrium, all models (ER, 1DM, 2DM) faithfully reproduce the step stimulus response of 5DF Monte Carlo (not shown). For moderate step stimuli, only the 2DM model faithfully reproduces the step stimulus response of 5DF Monte Carlo, shown in Figure 5. For large step stimuli, the 2DM model starts to deviate from 5DF Monte Carlo, as seen in Figure 6.

The effective renewal theory (ER) model does a reasonable job of predicting the ensemble firing rate of the system to low-amplitude step stimuli.

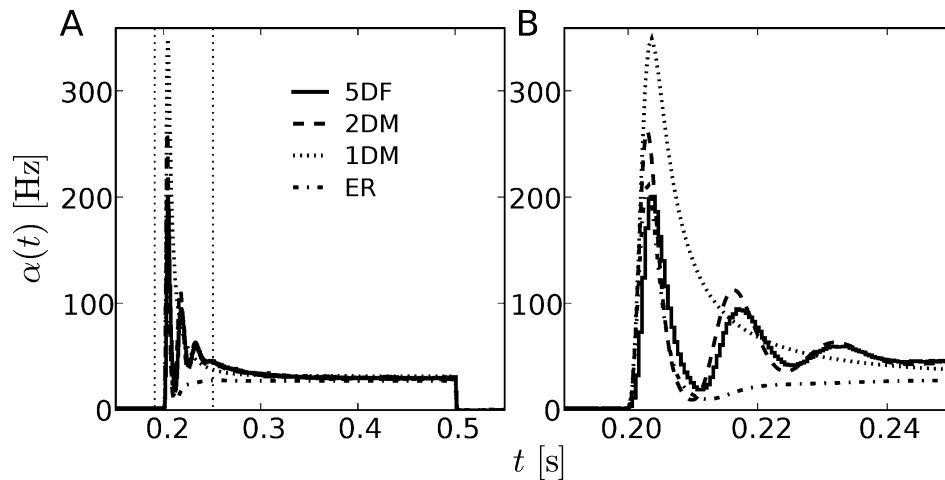


Figure 6: (A) The ensemble firing rate,  $\alpha(t)$ , in response to a large step stimulus, determined by 5DF Monte Carlo ( $5 \cdot 10^5$  trials, solid line), and numerical solution of the 1DM (dotted line), 2DM (dashed line), and ER (dashed dotted line) master equations. For the definition of the 5DF, 1DM, 2DM, and ER models, see Table 1. (B) The region outlined by the dashed rectangle is enlarged, showing disagreement between the two-dimensional Markov (2DM) model and the full system (5DF Monte Carlo).

This is perhaps surprising, since we do not expect renewal models to faithfully reproduce the dynamical responses of spike-frequency adapting neurons, as renewal models do not account for the dependencies of the firing probability density (hazard function) on spikes prior to the most recent. However, this shows that if the ensemble is not pushed far from equilibrium, knowledge of just the last spike is sufficient to predict the firing rate.

We further compared 5DF Monte Carlo and predictions of the 2DM model for a stimulus,  $v_e(t)$ , generated by an Ornstein-Uhlenbeck (OU) process. Let  $\zeta(t)$  be an OU process with mean of 10 Hz, standard deviation of 0.6 Hz, and time constant of 0.2 s. Then the excitatory synaptic inputs were supplied with  $v_e(t) = N_e \zeta(t)$ , with  $N_e = 1000$  being the number of excitatory synaptic inputs.

The ensemble firing rates for the 2DM model, its adiabatic solution, and 5DF Monte Carlo are shown in Figure 7. The adiabatic solution of the 2DM model is defined as the system that at each instant in time has a distribution of states equal to the instantaneous equilibrium distribution,  $P_\infty(t_s, t_r, t)$ , the asymptotic state distribution for large time resulting from a  $h(t_s, t_r, t)$  fixed at the instantaneous value at time  $t$ . The firing rate of the adiabatic 2DM model is then calculated by

$$\alpha(t) = \int_{-\infty}^{\infty} h(t_s, t_r, t) P_\infty(t_s, t_r, t) dt_s dt_r. \quad (6.3)$$

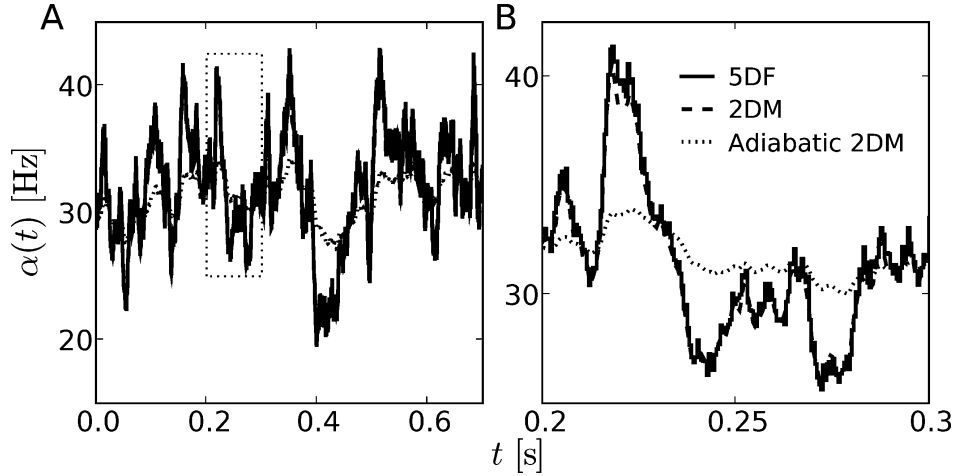


Figure 7: (A) The ensemble firing rate,  $\alpha(t)$ , in response to an Ornstein-Uhlenbeck process stimulus (as described in the text), determined by 5DF Monte Carlo (solid line), numerical solution of the 2DM master equation (dashed line), and the adiabatic solution (adiabatic 2DM, dotted line) computed by equation 6.3. For the definition of the 5DF, and 2DM models, see Table 1. (B) The region outlined by the dashed rectangle is enlarged, showing consistency between the two-dimensional Markov (2DM) model and the full system (5DF Monte Carlo).

By comparison of the ensemble firing rates of the 2DM model with its adiabatic solution in Figure 7, we can see that the system is being driven from equilibrium by the stimulus. Furthermore, the behavior of the 2DM model far from equilibrium captures the ensemble firing rate dynamics of 5DF Monte Carlo faithfully. This is a robust result under variation of neuron parameters and stimuli, so long as the ensemble is not pushed too far from equilibrium, as for the large step stimulus in Figure 6.

## 7 Beyond Mean-Adaptation Approximations

In this section we show that statistical moment theory approximations such as the mean-adaptation theories due to La Camera et al. (2004) can be derived from the 1DM master equation. The approach generalizes, and we derive the next order moment theory approximation, a mean+variance-adaptation theory and use it to clarify the domain of validity of mean-adaptation theories.

Recall the 1DM master equation for a spike-frequency adapting neuron,

$$\begin{aligned} \frac{\partial}{\partial t} P(g_s, t) = & \frac{\partial}{\partial g_s} \left[ \frac{g_s}{\tau_s} P(g_s, t) \right] \\ & + h_g(g_s - q_s, t) P(g_s - q_s, t) \\ & - h_g(g_s, t) P(g_s, t), \end{aligned} \quad (7.1)$$

where  $P(g_s, t)$  is the probability density of the adaptation state variable,  $g_s$ , and  $P(g_s < 0, t) = 0$ . The ensemble firing rate,  $\alpha(t)$ , is given by

$$\alpha(t) = \int_{-\infty}^{\infty} h_g(g_s, t)P(g_s, t)dg_s. \quad (7.2)$$

The mean adaptation variable is

$$\langle g_s(t) \rangle = \int_{-\infty}^{\infty} g_s P(g_s, t)dg_s. \quad (7.3)$$

Multiplying equation 7.1 by  $g_s$  and integrating over  $g_s$  yields the time evolution of the mean,  $\langle g_s(t) \rangle$ ,

$$\frac{d\langle g_s(t) \rangle}{dt} = -\frac{1}{\tau_s} \langle g_s(t) \rangle + q_s \alpha(t). \quad (7.4)$$

By Taylor expanding  $h_g(g_s)$  in equation 7.2 around  $\langle g_s(t) \rangle$ , and keeping up to linear terms, a mean-adaptation theory as in La Camera et al. (2004) results. Keeping up to quadratic terms, we have

$$\alpha(t) \approx \alpha(\langle g_s(t) \rangle, \langle \Delta g_s^2(t) \rangle) = h_g(\langle g_s(t) \rangle) - \frac{1}{2} h_g''(\langle g_s(t) \rangle) \cdot \langle \Delta g_s^2(t) \rangle, \quad (7.5)$$

where the  $h_g'(\langle g_s(t) \rangle)$  term vanishes by a cancellation of means. A mean+variance-adaptation theory results, but we require the time evolution of the variance. Multiplying equation 7.1 by  $(g_s - \langle g_s(t) \rangle)^2$  and integrating over  $g_s$  yields the time evolution of the variance,  $\langle \Delta g_s^2(t) \rangle$ ,

$$\begin{aligned} \frac{d\langle \Delta g_s^2(t) \rangle}{dt} &= -\frac{2}{\tau_s} \langle \Delta g_s^2(t) \rangle + q_s^2 \alpha(t) \\ &\quad + 2q_s \int_0^{\infty} (g_s - \langle g_s(t) \rangle) h_g(g_s, t) P(g_s, t) dg_s. \end{aligned} \quad (7.6)$$

Approximating  $h_g(g_s) \approx h_g(\langle g_s(t) \rangle) + h_g'(\langle g_s(t) \rangle)(g_s - \langle g_s(t) \rangle)$ , equation 7.6 becomes

$$\begin{aligned} \frac{d\langle \Delta g_s^2(t) \rangle}{dt} &\approx -\frac{2}{\tau_s} \langle \Delta g_s^2(t) \rangle + q_s^2 \alpha(\langle g_s(t) \rangle, \langle \Delta g_s^2(t) \rangle) \\ &\quad + 2q_s h_g'(\langle g_s(t) \rangle) \cdot \langle \Delta g_s^2(t) \rangle, \end{aligned} \quad (7.7)$$



which has a steady state

$$\langle \Delta g_s^2 \rangle = \frac{1}{2} \frac{q_s^2 \alpha \langle g_s \rangle, \langle \Delta g_s^2 \rangle}{\frac{1}{\tau} - q_s h'_g \langle g_s \rangle}. \quad (7.8)$$

Thus, the mean+variance-adaptation theory consistency relation for the adapted equilibrium firing rate,  $\alpha^*$ , reads

$$\alpha^* = h_g(q_s \tau_s \alpha^*) + \frac{1}{4} h''_g(q_s \tau_s \alpha^*) \left[ \frac{q_s^2 \alpha^*}{\frac{1}{\tau} - q_s h'_g(q_s \tau_s \alpha^*)} \right]. \quad (7.9)$$

Higher-order moment theories can be derived by keeping higher terms in the expansions in equations 7.5 and 7.7, and computing the necessary equations for the time evolution of higher statistical moments from the master equation 7.1.

**7.1 Validity of Mean-Adaptation Theories.** In this section we give a heuristic criterion for the validity of mean-adaptation theories in the static case, and the improved accuracy of the mean+variance-adaptation theory is demonstrated by a numerical example. It is illustrative to first investigate the exactly solvable leaky integrate-and-fire neuron driven by white noise for the parameters considered in La Camera et al. (2004), and subsequently contrast the findings to the 5DF model defined by equations 3.13 to 3.15.

It can be seen by inspection of equation 7.9 that if  $h''_g(g_s) \approx 0$  over the regime where  $P(g_s)$  is appreciably nonzero, then the mean-adaptation consistency relation,

$$\alpha^* = h_g(q_s \tau_s \alpha^*), \quad (7.10)$$

as in La Camera et al. (2004), results.

First, we use the 1DM master equation to verify the mean-adaptation theory for the leaky integrate-and-fire neuron driven by white noise considered in La Camera et al. (2004). The hazard function,  $h_g(g_s, t)$ , is referred to there as the response function in the presence of noise and has the exact solution,

$$h_g(g_s, t) = \left[ \tau \int_{\frac{C V_r - (m - g_s) \tau}{\sigma \sqrt{\tau}}}{\frac{C \theta - (m - g_s) \tau}{\sigma \sqrt{\tau}}} \sqrt{\pi} e^{x^2} (1 + \operatorname{erf}(x)) dx \right]^{-1}, \quad (7.11)$$

due to Siegert (1951), Ricciardi (1977), and Amit and Tsodyks (1991), where  $V_r$  is the reset potential,  $\theta$  is the threshold,  $\tau$  is the membrane potential,  $C$  is the membrane capacitance, and  $\operatorname{erf}(x) = (2/\sqrt{\pi}) \int_0^x e^{-t^2} dt$  is the error

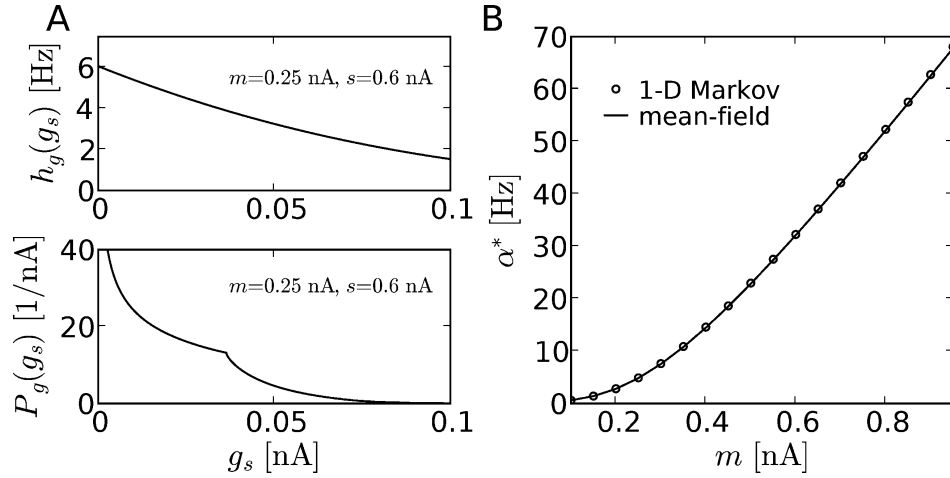


Figure 8: (A, top) The hazard function,  $h_g(g_s)$ , and (A, bottom) the equilibrium distribution of adaptation states,  $P(g_s)$ , in the low-firing rate regime ( $\alpha^* = 4.83$  Hz, mean current input  $m = 0.25$  nA and noise  $\sigma = s \cdot \sqrt{2}$  ms with  $s = 0.6$  nA) of the leaky integrate-and-fire neuron (LIF) used in La Camera et al. (2004).  $P(g_s)$  was determined by numerical solution of the 1DM master equation using the hazard function given in equation 7.11. Neuron parameters:  $C = 0.5$  nF,  $\tau_m = 20$  ms,  $\theta = 20$  mV,  $V_r = 10$  mV. Adaptation parameters:  $\tau_s = 110$  ms,  $q_s \cdot \tau_s = 4$  pA  $\cdot$  s. For comparison, the neuron and adaptation parameters are as for Figure 1a in La Camera et al. (2004), except  $\tau_r = 0$  ms and  $\tau_s = 110$  ms. For the definition of the 1DM model, see Table 1. The hazard function is nearly linear over the distribution of states; thus, terms depending on the variance of  $P(g_s)$  in equation 7.9 can be neglected, and mean-adaptation theories will yield good approximations to the adapted ensemble firing rate. (B) The adapted ensemble firing rate,  $\alpha^*$ , for a range of mean current inputs,  $m$ , determined by numerical solution of the 1DM master equation (circles), and the mean-adaptation theory consistency relation (solid line).

function. As in La Camera et al. (2004),  $m$  and  $\sigma$  are the mean and standard deviation of the input current. Upon firing, the adaptation current,  $g_s$ , makes a jump of  $q_s$  and relaxes with a time constant  $\tau_s$ . As can be seen in Figure 8A,  $h_g(g_s)$  is quite near linear over the regime where  $P(g_s)$  is appreciably nonzero, and predictions of the adapted firing rate due to a mean-adaptation theory are in excellent agreement with the 1DM master equation as shown in Figure 8B. The mean+variance-adaptation theory helps us to understand this: agreement is good because  $h_g''(g_s) \approx 0$  over the regime where  $P(g_s)$  is appreciably nonzero for all firing rates considered.

For the 5DF models defined by equations 3.13 to 3.15, we have  $h_g(g_s) \approx a \cdot \exp(-bg_s)$ . As can be seen in Figure 9A,  $h_g(g_s)$  has an appreciable second derivative over  $P(g_s)$ , and thus we expect mean-adaptation equilibrium ensemble firing rate predictions to deviate from the ensemble firing rate of the 1DM master equation. Indeed, such deviations are observed and are

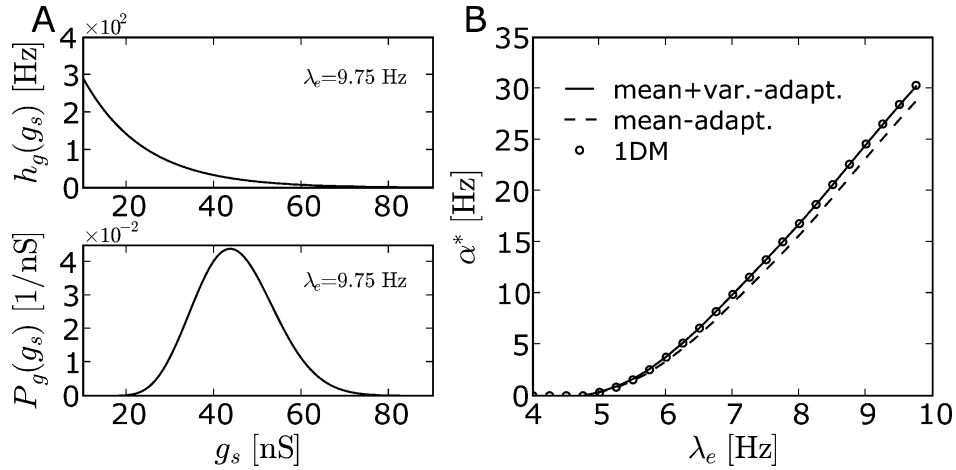


Figure 9: (A, top) The hazard function,  $h_g(g_s)$ , and (A, bottom) the equilibrium distribution of adaptation states,  $P(g_s)$ , determined by numerical solution of the 1DM master equation. The hazard function,  $h_g(g_s)$ , was determined by fitting to 5DF Monte Carlo as in Figure 1 with  $\lambda_e = 9.75$  Hz,  $\lambda_i = 11.4$  Hz. For the definition of the 5DF and 1DM model, see Table 1. The hazard function has nonzero curvature ( $h_g''(g_s) > 0$ ) over the distribution of states; thus, terms depending on the variance of  $P(g_s)$  in equation 7.9 cannot be neglected, and mean-adaptation theory predictions for the adapted ensemble firing rate are expected to be in error. (B) The adapted ensemble firing rate,  $\alpha^*$ , for a range of Poisson input rates,  $\lambda_e$ , determined by solution of the 1DM master equation (circles), the mean-adaptation theory consistency relation (dashed line), and the mean+variance-adaptation consistency relation (solid line). As expected, mean-adaptation theory predictions for the adapted firing rate are corrected by the mean+variance-adaptation theory consistency relation in equation 7.9.

corrected by the mean+variance-adaptation consistency relation, as seen in Figure 9B. Thus, a heuristic condition for the validity of mean-adaptation theories is that we must have  $h_g''(g_s) \approx 0$  over the regime where  $P(g_s)$  is appreciably nonzero. Less heuristically, the contributions due to the second term (and all neglected higher-order terms) on the right-hand side of equation 7.9 must vanish compared to the first. When this condition is violated, higher-order moment theories such as the mean+variance-adaptation theory given here, or the 1DM master equation, should be applied to determine the ensemble firing rate.

For the neuron models considered here, the accuracy of the mean+variance-adaptation theory was also verified in the dynamic case for an OU stimulus as in Figure 7, as shown in Figure 10.

## 8 Discussion

In this letter, we propose a one-dimensional Markov process (the 1DM model) for modeling neural ensemble activity and spike train statistics

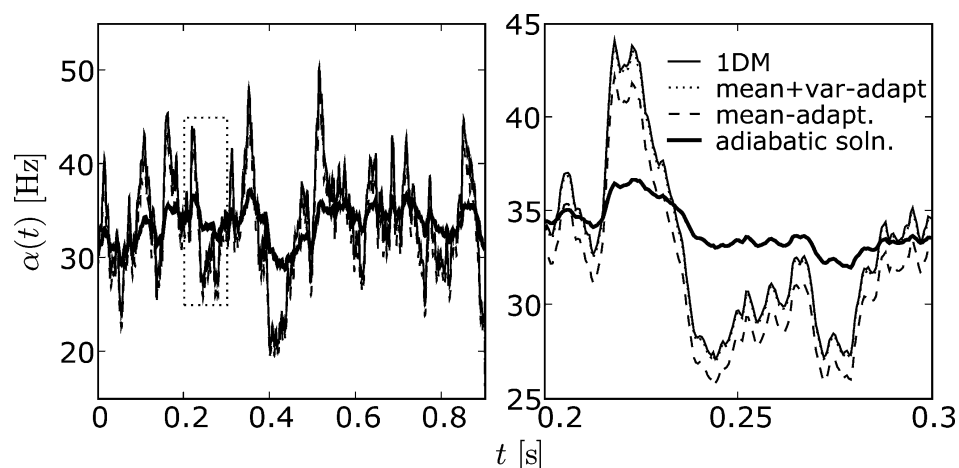


Figure 10: (A) The ensemble firing rate,  $\alpha(t)$ , in response to an Ornstein-Uhlenbeck process stimulus (as for Figure 7), determined by the 1DM model (solid line), the adiabatic solution (thick solid line) computed by the mean+variance-adaptation consistency relation equation 7.9, the dynamic mean+variance-adaptation theory equations 7.4 to 7.6 (dotted line), and the dynamic mean-adaptation theory equations (dashed line). (B) The region outlined by the dashed rectangle is enlarged, showing consistency between the 1DM model and the mean+variance-adaptation theory, while predictions due to the mean-adaptation theory are poor. For the definition of the 1DM model, see Table 1.

that goes beyond renewal theory by accounting for interspike interval (ISI) correlations due to spike-frequency adaptation (SFA) mechanisms without the need to model the high-dimensional space of the microscopic neuronal state variables.

We demonstrated that the full five-dimensional master equation of a conductance-based integrate-and-fire neuron with SFA and a refractory mechanism driven by Poisson spike trains (the 5DF model) can be reduced to a two-dimensional master equation plus filtering differential equations accounting for synaptic dynamics (the 2DM model), under an adiabatic elimination of the fast variables  $v$ ,  $g_e$ ,  $g_i$ , assuming the neuron has nonzero synaptic time constants and is in the high-conductance state. The resulting 2DM master equation is a two-dimensional generalization of the Markov process proposed at the outset as an extension of renewal theory to account for ISI correlations.

We presented methods for generating inhomogeneous realizations of the proposed 1DM and 2DM models and a method for solving their master equations numerically. The 2DM model was shown to accurately predict firing rate profiles of the full system under dynamic stimulation and conditional ISI distributions and serial ISI correlations under static stimulation.

It was shown that mean-adaptation theories for spike-frequency adapting neurons with noisy inputs such as in La Camera et al. (2004) and higher-order statistical moment theories can be derived from the 1DM master equation as long as one neglects the refractory period. A heuristic condition for the validity of mean-adaptation theories was derived and found to be violated for the neuron model (5DF) and parameters considered here. Furthermore, a mean+variance-adaptation theory was derived that corrected the ensemble firing rate predictions of mean-adaptation theories in this case.

**8.1 Comparison with Other Studies of Adaptation.** Studies of the firing rates of networks and ensembles of spike-frequency adapting neurons due to Latham et al. (2000) and Fuhrmann et al. (2002) augment a Wilson and Cowan equation (Wilson & Cowan, 1972) for the firing rate with a mean adaptation variable.

As is typical of the Wilson and Cowan approach, the ensemble firing rate,  $\alpha$ , enters a differential equation of the form

$$\tau_e \frac{d\alpha}{dt} = -\alpha + h_g(\langle g_s(t) \rangle, \dots), \quad (8.1)$$

where  $h_g(\langle g_s(t) \rangle, \dots)$  is the static firing rate given the input and the mean adaptation, and  $\tau_e$  is the timescale for relaxation to a firing rate equilibrium. As is suggested in Fuhrmann et al. (2002),  $\tau_e$  is determined mainly by the membrane time constant of the neuron, but is also affected by the mean amplitude of the input, and is treated there as a free parameter.

It has been argued in Gerstner (2000), Brunel, Chance, Fourcaud, and Abbott (2001), Fourcaud and Brunel (2002), Renart et al. (2004), and La Camera et al. (2004) that for current and conductance-based synapses with nonzero time constants and biological input statistics, the ensemble firing rate responds instantaneously to input currents, and synaptic filtering dominates. In this case, the Wilson and Cowan equation for  $\alpha$  can be replaced by an instantaneous f-I function, and the synaptic currents or conductances modeled by relaxation equations for their means and variances. This is the approach taken in La Camera et al. (2004). Thus, one sidesteps the concerns mentioned in Fuhrmann et al. (2002) that the Wilson and Cowan equation “cannot be rigorously derived from the detailed integrate-and-fire model” and has been “shown not to accurately describe the firing rate dynamics [by] (Gerstner, 2000).”

The models due to Latham et al. (2000), Fuhrmann et al. (2002), and La Camera et al. (2004) all approximate the evolution of the ensemble of adaptation variables by its mean value and are therefore mean-adaptation theories. La Camera et al. (2004) state that such mean-adaptation theories are a good approximation under the assumption that “adaptation is slow compared to the timescale of the neural dynamics. In such a case, the

feedback [adaptation] current . . . is a slowly fluctuating variable and does not affect the value of  $s$  [the standard deviation of the input current].”

La Camera et al. (2004) explore adaptation time constants on the order of 100 ms under the assumption that the adaptation dynamics are “typically slower than the average ISI.” They report that “for irregular spike trains the agreement is remarkable also at very low frequencies, where the condition that the average ISI be smaller than  $\tau_N$  [the time constant of adaptation] is violated. This may be explained by the fact that although  $\langle ISI \rangle > \tau_N$ , the ISI distribution is skewed towards smaller values, and [the mean adaptation current proportional to the firing rate] . . . is still a good approximation.”

In section 7 we used the 1DM master equation to derive a mean+variance-adaptation theory, the next correction to the mean-adaptation theories in La Camera et al. (2004), yielding another explanation for the success reported there. We found that the error in the firing rate in La Camera et al. remained small because the hazard function used there is a nearly linear function of the adaptation variable in the interesting regime where  $P(g_s)$  is appreciably nonzero. Thus, perturbing contributions to the average firing rate from deviations of the adaptation variable above and below the mean over the course of one ISI roughly cancel on average. For the neuron model (5DF) and parameters considered here, the hazard function has an appreciable nonlinearity resulting in erroneous predictions of the firing rate when using a mean-adaptation theory. The mean+variance-adaptation theory derived here corrected the predictions.

It is appropriate to reiterate that both the 1DM master equation and the resulting mean+variance-adaptation theory approximation considered here neglect refractory dynamics. It was demonstrated by the adiabatic reduction of the 5DF model to the 2DM model that the inclusion of a relative refractory period requires a two-dimensional master equation. Indeed, as shown in Figure 6, oscillations emerge for large and fast stimulation changes that are qualitatively captured by the 2DM model but not by the 1DM model. It remains to be seen if a two-dimensional mean- or mean+variance-adaptation theory can be constructed that accounts for this behavior, and under what conditions it can be reduced to a one-dimensional model by simply rescaling the firing rate by  $f' = 1/(1/f + \tau_{\text{eff}})$ , as in La Camera et al. (2004) for the absolute refractory period case, where  $\tau_{\text{eff}}$  is some effective absolute refractory period of the relative refractory mechanism.

In Benda and Herz (2003), a thorough mathematical analysis of several well-known mechanisms for SFA based on biophysical kinetics is performed for the case of a suprathreshold current. A universal phenomenological mean-adaptation model for such biophysical mechanisms for SFA is introduced with much the same form as later used in La Camera et al. (2004) for the case of noisy drive. Methods are given to completely parameterize the model using quantities that can be easily measured by standard recording techniques. Implications for signal processing are considered there and in subsequent publications (Benda, Longtin, & Maler, 2005).

In Chacron et al. (2003), a novel approach is taken compared to Latham et al. (2000), Fuhrmann et al. (2002), Benda and Herz (2003), and La Camera et al. (2004). There an expression is derived for the serial correlation coefficient of ISIs in the static case by employing a Markov chain. In their analysis, they define a quantity analogous to the static distribution  $P^\dagger(t_s)$  here. In their framework, they prove that adaptation of the threshold fatigue form used there results in ISI correlations as have been observed experimentally (Longtin & Racicot, 1997; Chacron et al., 2000; Nawrot et al., 2007). Their expression, however, contains integrals that require “the computation of the FPT [first-passage time] PDF of the Ornstein-Uhlenbeck process through an exponential boundary. Given that no general analytical expression is available for this quantity, derivation of the correlation from the integrals can be computationally more demanding than estimating the same quantities from simulations” (Chacron et al., 2003). Subsequently, only simulations are performed, and the derived expression is never compared to the simulated result. Thus, they miss an important benchmark to ensure the calculations are correct. It is possible that the numerical techniques used here could be applied to compute a prediction for the correlation coefficient by the expression they derive and subsequently compared to the simulated result.

Mean-adaptation theories cannot be used to model the correlation between subsequent ISIs, as they do not preserve the ensemble statistics. Our approach is that one simply not replace the trajectory of the adaptation variable,  $g_s$ , by its mean. This resolves the problem in the development in La Camera et al. (2004) that the mean input current and instantaneous  $g_s$  have an equal role in determining the instantaneous firing rate, and  $g_s$  cannot be consistently replaced by its mean. What results is the 1DM master equation presented here. We subsequently calculate an expression for the inhomogeneous conditional ISI distribution and compare it to 5DF Monte Carlo in the static case. Furthermore, we calculate the variation of the mean of the conditional ISI distribution as a function of the preceding ISI, a generalization of the serial correlation coefficient of ISIs, and compare it to 5DF Monte Carlo. By our Markov process master equation, we avoid the difficulty encountered in Chacron et al. (2003) of treating the first passage times of an Ornstein-Uhlenbeck process through an exponential boundary, while capturing the full inhomogeneous ensemble dynamics in a framework that is tractable.

**8.2 On the Adiabatic Reduction of the Master Equation.** Under the assumption that the neuron is in the high-conductance state due to biologically realistic noisy inputs, we show that the 5DF master equation for the conductance-based spike frequency adapting relative refractory integrate-and-fire neuron model used here can be reduced to the 2DM master equation by an adiabatic elimination of fast variables. The variables that remain are those of SFA and the relative refractory mechanism, and the form is

analogous to the 1DM master equation proposed to extend renewal theory to a class of Markov processes that account for SFA.

The adiabatic reduction does not solve explicitly the firing rate of the given neuron model (without adaptation or the refractory mechanism) or rely on such a solution. We leave the firing rate dynamics of the given neuron model (without adaptation or the refractory mechanism) encapsulated in the hazard function (see equation 3.23). The approach applies to other models of adaptation such as the adapting threshold models in Chacron et al. (2003) and the current-based adaptation models in La Camera et al. (2004).

Concerning the generality of the adiabatic elimination for the adaptation variable, we expect it could be applied to a larger class of formally spiking neuron models with fast, intrinsic dynamics compared to adaptation. For those interested in modeling a class of neurons where a solution to equation 3.23 already exists, the framework can be easily and immediately applied. The fitting methods presented allow the connection to be made between models for which an explicit solution for the hazard function is unknown and the 1DM and 2DM models presented here. What results is a reduced state space to explore for functional implications.

The generality of treating the relative refractory mechanism as a slow variable in the adiabatic elimination is less clear. There are some issues that need to be clarified before one could specify the class of neurons to which it applies. Specifically, the relationship between the requirement that the neuron be in the high-conductance state (small effective  $\tau_m$ ) and the requirement that the synapses have nonvanishing time constants ( $\tau_e > 0$ ) resulting in a nonvanishing probability at threshold ( $P(v_{th}, \dots) > 0$ ) remains to be thoroughly investigated. The delta-conductance-based approach in Meffin et al. (2004), for example, does not satisfy the second requirement. The nonvanishing probability at threshold seems to be a criterion for the neuron to respond quickly to the synaptic signal (Fourcaud & Brunel, 2002; Renart et al., 2004).

An important step in the reduction is the treatment of the synaptic conductances. As their statistics are assumed to instantaneously determine the equilibrium statistics of the membrane potential, they were removed from the master equation. Then differential equations were found for their first statistical moments (means) in terms of the rate of the Poisson process input, as in La Camera et al. (2004). One weakness of the fitting approach used here is that we cannot account for the dynamics of the second central moment (variance), as was done in La Camera et al., and modeling both dynamic excitation and inhibition simultaneously requires a laborious fitting of a two-dimensional space of synaptic inputs. Further work will apply methods such as those due to Moreno-Bote and Parga (2004) to obtain a solution to equation 3.23 without fitting, thus allowing ensemble studies of adapting network models and analysis as in Latham et al. (2000) with the rigor of, for example, (Brunel, 2000), and the possibility for quantitative agreement with traditional large-scale network simulations.



**8.3 Beyond Renewal Theory.** We reviewed standard results of inhomogeneous renewal theory in appendix C and uncovered a conceptual error often made in the literature when using the intensity-rescaling transformation to make the transition from homogeneous (static) to inhomogeneous (dynamic) renewal theory. We clarified and remedied the problem by presenting a correct renewal process generation scheme as discussed in Devroye (1986).

By means of a variable transformation, we provided a link between the 1DM model and inhomogeneous renewal theory methods, allowing direct comparison and contrast. The 1DM master equation was found to have an analogous structure to the renewal master equation, but with a state space spanning the whole real line. Furthermore, the 1DM state is not reborn to a universal initial value upon spiking, as in renewal theory (zero age), but reinserted to a state that is a function of the state just prior to spiking. This fact introduces a memory into the system and results in negative ISI correlations as reported in Chacron et al. (2003).

Due to the detailed adiabatic reduction and fitting, we proposed the nested exponential form of the hazard function as given by equation 2.10 and the state-dependent reinsertion function as given by equation 2.7 for the conductance-based SFA mechanism considered here. The hazard function (perhaps time-dependent) and the reinsertion function together are a complete specification of the proposed Markov model given an initial distribution of states. We provided a numerical recipe to efficiently generate inhomogeneous realizations of the proposed Markov process.

With an additional dimension for a relative refractory mechanism, the Markov process faithfully reproduces the transient dynamics and ISI correlations of 5DF Monte Carlo, as expected by the adiabatic reduction. The same comparison between a one-dimensional Markov process and a neuron model without the relative refractory mechanism was not done, as we found that without refractory mechanisms, the neuron model used exhibited a high probability to spike just after spiking due to correlations in the synaptic conductance on the timescale of the refractory mechanism. We feel this is a bug rather than a feature of neuron models without a refractory mechanism. Thus we chose not to build a Markov process to account for it. Furthermore, the proposed relative refractory mechanism requires only slightly more effort than treating an absolute refractory period, as done in Nykamp and Tranchina (2001). When the hazard function calibrated for the 2DM model is used directly for the 1DM model, reasonable agreement to refractory neuron models was still observed for the moderate firing rates considered.

**8.4 Suprathreshold Stimuli.** For large and rapid changes in stimulus that bring the ensemble predominantly into the suprathreshold regime, the predictions due to numerical solutions of the 2DM model deviated somewhat from 5DF Monte Carlo simulations, as seen in Figure 6. The reasons

for this are twofold. First, the stimulus brings us into a regime where the exponential fitting procedure for  $h(t_s, t_r)$  begins to fail and was poorly populated with data points. This fact likely accounts for the larger amplitude of the oscillations of the 2DM model compared to 5DF Monte Carlo. It is likely that a choice of function that improves the fit in this regime or a proper analytical solution for  $h(t_s, t_r)$  would improve the agreement here. Second, following the large stimulus change, a large portion of the population is in the suprathreshold regime where neurons make large migrations from the reset potential directly to the threshold following a spike. The 2DM model completely neglects the dynamics of the membrane potential and thus this migration period, resulting in the phase lead over the full system.

A closer inspection of Figure 6 reveals a transition from predominantly supra- to predominantly subthreshold firing. Shortly after stimulus onset, a large portion of the population fires almost immediately and is reinserted with the adaptation conductance increased by  $q_s$ , that is, a mass exodus in phase space. For the 2D case, the neurons also start a refractory period upon reinsertion; in the 1D case, they do not. The stimulus is sufficiently strong that in the 2D case, it is still suprathreshold following the refractory period. In the 1D case, there is no refractory period, and the neurons can fire immediately following a spike cycle, and no lull is seen in the firing rate following the initial mass exodus. For the 2D case, and even the renewal case, the system is refractory following the mass exodus, and a lull in the firing rate results, to peak again as the neurons are released from the refractory state. With the accumulation of adaptation, the fraction of the ensemble participating in subsequent exodus events is ever diminished as more and more neurons enter the subthreshold regime where neurons survive for highly variable durations following the refractory period. Thus, for large stimuli that keep the neuron suprathreshold over several spikes, the population is initially synchronized, firing at a rate determined by the refractory mechanism. As adaptation accumulates, spiking becomes more irregular, and the neurons desynchronize.

The desynchronization of neurons driven by suprathreshold current has been observed experimentally in Mainen and Sejnowski (1995). It is shown there that this transition to the subthreshold regime due to adaptation is not strictly necessary for the neurons to desynchronize due to the constant presence of noise. However, adaptation is also a mechanism by which an otherwise irregularly firing neural ensemble is synchronized at a stimulus onset. Following such a synchronization, the transition from the predominantly suprathreshold regime to the predominantly subthreshold regime induced by the accumulation of adaptation is akin to a transition from a noisy oscillator firing mode to a point process firing mode. While the ensemble would gradually desynchronize in the noisy oscillator firing mode, the transition to the point process firing mode over only a few spikes ensures this occurs much more rapidly. Thus, adaptation works to both synchronize and then desynchronize at changes in stimulus. This is an implementation of novelty

detection and is related to the high-pass filtering properties of SFA reported in Benda et al. (2005). Evidence for a synchronization-desynchronization coding scheme for natural communication signals was recently reported for the spike-frequency adapting P-unit electroreceptors of weakly electric fish in Benda, Longtin, and Maler (2006).

## 9 Conclusion

---

In this letter, we have focused on establishing a framework for rigorously treating the dynamic effects of spike-frequency adaptation and refractory mechanisms on neural ensemble spiking. The resulting master equation formalism unifies renewal theory models and previous studies on adaptation (e.g., Latham et al., 2000; Fuhrmann et al., 2002; Chacron et al., 2003; Benda & Herz, 2003; La Camera et al., 2004) into an ensemble, or population density framework such as those due to Knight (1972, 2000), Brunel (2000), Omurtag et al. (2000), Nykamp and Tranchina (2000), Fourcaud and Brunel (2002), Richardson (2004), Rudolph and Destexhe (2005), Meffin et al. (2004), and Moreno-Bote and Parga (2004). The resulting methods are new and powerful tools for accurately modeling spike-frequency adaptation, an aspect of neuron dynamics ubiquitous in excitatory neurons that has been largely ignored in neural ensemble studies thus far due to the added difficulties of treating the extra state variable.

By distilling the detailed neuron model down to two essential dimensions, spike-frequency adaptation and a relative refractory period, using an adiabatic elimination, their central role in perturbing neural firing is emphasized. Future work will employ the framework to examine the functional implications of spike-frequency adaptation. Given the variety of intriguing and prominent consequences such as interspike interval correlations and transient synchronization following stimulus changes, one is compelled to question if spike-frequency adaptation can be neglected when considering, for example, the dynamic nature of the neural code (Shadlen & Newsome, 1998; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997), the propagation of synchrony (Abeles, 1991; Diesmann, Gewaltig, & Aertsen 1999), or the function of spike-timing-based learning rules (Gerstner, Kempter, van Hemmen, & Wagner, 1996; Song, Miller, & Abbott, 2000).

## Appendix A: Neuron and Adaptation Model Parameters

---

The parameters of the 5DF neuron model given in equations 3.13 to 3.15 were determined by fitting to a single-compartment Hodgkin-Huxley (HH) model of a pyramidal neuron under various conditions using NEURON (Hines & Carnevale, 1997) as described in Muller (2003). The HH model and parameters are taken from Destexhe, Contreras, and Steriade (1998).

The phenomenological mechanism for spike-frequency adaptation (SFA) used here, the counterpart to the M-current and AHP-current mechanisms

in the HH model, was inspired by (Dayan & Abbott, 2001), and similar models have been proposed in (Koch, 1999) and Fuhrmann et al. (2002), and more recently generalized in Brette and Gerstner (2005).

Additionally, a relative refractory period (RELREF) mechanism identical to the SFA mechanism was added, but with a much shorter time constant and a much larger conductance increase. Both the SFA and RELREF mechanisms consist of an action potential (AP) activated and exponentially decaying conductance coupled to an inhibiting reversal potential so that the standard membrane equation takes the form:

$$c_m \frac{dv(t)}{dt} = g_l(E_l - v(t)) + g_e(t)(E_e - v(t)) + g_i(t)(E_i - v(t)) \\ + g_s(t)(E_s - v(t)) + g_r(t)(E_r - v(t)).$$

If  $v$  exceeds the threshold  $v_{\text{th}}$ :

- $v$  is reset to  $v_{\text{reset}}$ .
- $g_s \mapsto g_s + q_s$ .
- $g_r \mapsto g_r + q_r$ .
- The time of threshold crossing is added to the list of spike times.

All conductances,  $g_x(t)$ , where  $x \in \{s, r, e, i\}$  are governed by an equation of the form

$$\frac{dg_x(t)}{dt} = -\frac{1}{\tau_x} g_x(t).$$

The arrival of a spike at a synapse triggers  $g_x \mapsto g_x + q_x$  for  $x \in \{e, i\}$ . Poisson processes were used to supply spike trains to the 1000 excitatory and 250 inhibitory synapses, where Poisson rates in the range 3 to 20 Hz were used as described in the text for each specific simulation. The synaptic model and parameters were directly transferred from the HH models, while the remaining parameters, as determined by fits to the HH model, are given in Table 3.

## Appendix B: Further Details on the Adiabatic Reduction

In this appendix, we give the mathematical steps that lead from equation 3.16 to 3.20 in a more detailed way.

For notational simplicity, we introduce the five-dimensional state variable  $x = (v, g_e, g_i, g_s, g_r)$ . The indices 1, 2, 3, 4, 5 shall correspond to  $v, e, i, s, r$ , as used in the definition of the neuron model in equations 3.13 to 3.15 (e.g.,  $\tau_2 := \tau_e$ ). The partial derivatives with respect to  $x_\mu$  are denoted by  $\partial_\mu$  and with respect to time by  $\partial_t$ . Furthermore, we define  $P(x_1, x_2, x_3, x_4, x_5) = 0$  if one or more of the conductances  $x_2, \dots, x_5$  is negative.

Table 3: Neuron and Synapse Model Parameters Used for Simulations of the Full System (5DF) Given by Equations 3.13 to 3.15.

Parameter	Description	Value
$v_{\text{th}}$	Threshold voltage	$-57$ mV
$v_{\text{reset}}$	Reset voltage	$-70$ mV
$c_m$	Membrane capacitance	$289.5$ pF
$g_l$	Membrane leak conductance	$28.95$ nS
$E_l$	Membrane reversal potential	$-70$ mV
$q_r$	RELREF quantal conductance increase	$3214$ nS
$\tau_r$	RELREF conductance decay time	$1.97$ ms
$E_r$	RELREF reversal potential	$-70$ mV
$q_s$	SFA quantal conductance increase	$14.48$ nS
$\tau_s$	SFA conductance decay time	$110$ ms
$E_s$	SFA reversal potential	$-70$ mV
$E_{e,i}$	Reversal potential of excitatory and inhibitory synapses, respectively	$0$ mV, $-75$ mV
$q_{e,i}$	Excitatory and inhibitory synaptic quantal conductance increase	$2$ nS
$\tau_{e,i}$	Excitatory and inhibitory synaptic decay time	$1.5$ ms, $10.0$ ms

The master equation governing the evolution of the probability density  $P(x, t)$  may be formulated as a conservation equation:

$$\partial_t P(x, t) = -\text{div} J(x, t) + \delta(x_1 - v_{\text{reset}}) J_1(v_{\text{th}}, x_2, x_3, x_4 - q_4, x_5 - q_5, t). \quad (\text{B.1})$$

The second term on the right-hand side of equation B.1 accounts for neurons that cross the threshold surface  $x_1 = v_{\text{th}}$  at time  $t$  with the state variables  $(v_{\text{th}}, x_2, x_3, x_4 - q_4, x_5 - q_5)$  and are reinserted to  $(v_r, x_2, x_3, x_4, x_5)$ .

The probability current  $J(x, t)$  is determined by the underlying stochastic differential equations 3.13 to 3.15. The components  $J_\mu(x, t)$  for  $\mu = 1, \dots, 5$  consist of the current due to the drift terms,  $\beta_\mu(x)$ , and for  $\mu = 2, 3$  of additional currents due to the excitatory and inhibitory input Poisson spike trains, respectively. The drift term for the membrane potential reads

$$\beta_1(x) := \frac{1}{c_m} \left( \sum_{\mu=2}^5 x_\mu (E_\mu - x_1) + g_l (E_l - x_1) \right). \quad (\text{B.2})$$

For the conductances  $x_\mu$  with  $\mu = 2, \dots, 5$ , the drift terms are:

$$\beta_\mu(x) = \beta_\mu(x_\mu) := -\frac{1}{\tau_\mu} x_\mu. \quad (\text{B.3})$$

The components of the probability current for  $\mu = 1, 4, 5$  obey the equation

$$J_\mu(x, t) = \beta_\mu(x)P(x, t). \quad (\text{B.4})$$

For the excitatory synaptic conductance  $x_2$ , the component  $J_2(x, t)$  is

$$\begin{aligned} J_2(x, t) = & \beta_2(x)P(x, t) \\ & + \int_0^{x_2} \left[ \int_0^\infty W_2(y_2, y_1, t)P(x_1, y_1, x_3, \dots, x_5)dy_2 \right] dy_1 \\ & - \int_0^{x_2} \left[ \int_0^\infty W_2(y_1, y_2, t)P(x_1, y_2, x_3, \dots, x_5)dy_2 \right] dy_1. \end{aligned} \quad (\text{B.5})$$

The component  $J_3(x, t)$  has a similar form with obvious modifications. Since the synaptic input is modeled as a Poisson process, the transition rates  $W_\mu(y_1, y_2, t)$  for  $\mu = 2, 3$  may be written as

$$W_\mu(y_1, y_2, t) = \nu_\mu(t)\delta(y_1 - (y_2 + q_\mu)), \quad (\text{B.6})$$

given the presynaptic firing rates  $\nu_\mu(t)$ . The diffusion approximation can be obtained by a Kramers-Moyal expansion of the components  $J_2$  and  $J_3$  (Gardiner, 1985).

**B.1 Integration.** To obtain an equation for the marginal probability distribution,  $P(x_4, x_5, t)$ , one integrates equation B.1 over  $x_1, x_2, x_3$ . The integral of the terms  $\partial_\mu J_\mu(x, t)$  on the right-hand side in B.1 for  $\mu = 2, 3$  vanishes due to the boundary condition that the probability current vanishes in the limit  $x_\mu \rightarrow 0$  and  $x_\mu \rightarrow \infty$  for  $\mu = 2, 3$ :

$$\int_0^\infty \partial_\mu J_\mu(x, t)dx_\mu = \lim_{x_\mu \rightarrow \infty} J_\mu(x, t) - J_\mu(x, t)\Big|_{x_\mu=0} = 0. \quad (\text{B.7})$$

The component  $J_1(x, t)$  yields a nonvanishing contribution:

$$\begin{aligned} & \int_0^\infty \int_0^\infty \left( \int_{-\infty}^{v_{\text{th}}} \partial_1 J_1(x, t)dx_1 \right) dx_2 dx_3 \\ & = \int_0^\infty \int_0^\infty J_1(v_{\text{th}}, x_2, \dots, x_5, t)dx_2 dx_3. \end{aligned} \quad (\text{B.8})$$

The reinsertion term involves an integration over a delta distribution:

$$\begin{aligned} & \int_0^\infty \int_0^\infty \left( \int_{-\infty}^{v_{\text{th}}} \delta(x_1 - v_{\text{reset}})J_1(v_{\text{th}}, x_2, x_3, x_4 - q_4, x_5 - q_5, t)dx_1 \right) dx_2 dx_3 \\ & = \int_0^\infty \int_0^\infty J_1(v_{\text{th}}, x_2, x_3, x_4 - q_4, x_5 - q_5, t)dx_2 dx_3. \end{aligned} \quad (\text{B.9})$$

Integration of the left-hand side in equation B.1 results in:

$$\int_0^\infty \int_0^\infty \int_{-\infty}^{v_{\text{th}}} (\partial_t P(x, t)) dx_1 dx_2 dx_3 = \partial_t P(x_4, x_5, t). \quad (\text{B.10})$$

Plugging these results into equation 3.16 yields:

$$\begin{aligned} \partial_t P(x_4, x_5, t) = & - \sum_{\mu=4,5} \partial_\mu (\beta_\mu(x_\mu) P(x_4, x_5, t)) \\ & + \int_0^\infty \int_0^\infty J_1(v_{\text{th}}, x_2, x_3, x_4 - q_4, x_5 - q_5, t) dx_2 dx_3 \\ & - \int_0^\infty \int_0^\infty J_1(v_{\text{th}}, x_2, \dots, x_5) dx_2 dx_3. \end{aligned} \quad (\text{B.11})$$

Returning to the initial notation and using the definition for  $J_1(x, t) = \beta_1(x, t)P(x, t)$  yields equation 3.20.

### Appendix C: Inhomogeneous Renewal Processes ---

The proposed Markov models can be easily understood by analogy to inhomogeneous renewal theory. We review some standard results thereof, which define the renewal models used in the text.

Poisson and gamma renewal processes, both special cases of a renewal process, are used extensively to model the spike train statistics of cortical neurons (van Rossum, Bi, & Turrigiano, 2000; Song & Abbott, 2001; Rudolph & Destexhe, 2003b; Shelley et al., 2002), and are treated in detail in (Gerstner & Kistler, 2002) in sections 5.2, 5.3, 6.2.2, and 6.3.2. While the treatment in section 6.2.2 is developed for spike response model neurons with escape noise and in section 6.3.2 for populations of neurons satisfying a few basic assumptions, it is not explicitly stated there that the analysis is that of an arbitrary inhomogeneous renewal process, though it is mentioned in Gerstner (2001). We first reiterate this fact by producing the main results of section 6.2.2 and 6.3.2 of Gerstner and Kistler (2002) using an inhomogeneous generalization of the notation of Cox (1962), a classic reference work on homogeneous renewal theory. Second, we present a recipe for efficiently generating spike trains of a general inhomogeneous renewal process.

In what follows, by working exclusively with the inhomogeneous hazard function, we avoid the pitfall of studies that erroneously assume an intensity-rescaling transformation of a stationary gamma renewal process

with parameter  $\gamma$ , or  $a$  yields an inhomogeneous gamma renewal process with parameter  $\gamma$ , or  $a$  (Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Gazères, Borg-Graham, & Frégnac, 1998).

A renewal process is defined here to be *inhomogeneous* when its hazard function takes the form  $\rho(\tau, t)$ , where  $\tau$  denotes the age, and  $t$  denotes the explicit dependence of the hazard function on time. The ensemble firing rate<sup>3</sup> at  $t$ , denoted by  $\alpha(t)$ , is the expectation value,

$$\alpha(t) = \langle \rho(t) \rangle = \int_0^\infty \rho(s, t) f^-(s, t) ds, \quad (\text{C.1})$$

where  $f^-(\tau, t)$  denotes the probability density function (PDF) of times since last renewal, also called the backward recurrence time in Cox (1962). The PDF,  $f^-(\tau, t)$ , can be determined by reasoning that the probability that the system has an age in the interval  $(\tau, \tau + \Delta\tau)$  is equal to the probability that there is a renewal in the time interval  $(t - \tau, t - \tau + \Delta\tau)$  and that the system subsequently survives until  $t$ . This yields

$$f^-(\tau, t) = \alpha(t - \tau) \mathcal{F}(\tau, t - \tau), \quad (\text{C.2})$$

where  $\mathcal{F}(\Delta t, t)$  is the inhomogeneous survival function, representing the probability that the system will survive for a time  $\Delta t$  after  $t$ . Generalizing equation 1.2.10 in Cox (1962) for the inhomogeneous case, we have

$$\mathcal{F}(\Delta t, t) = \exp\left(-\int_0^{\Delta t} \rho(s, t + s) ds\right). \quad (\text{C.3})$$

Plugging equation C.2 into C.1 results in the equivalent of equations 6.44 and 6.45 of Gerstner and Kistler (2002).

A differential formulation of equations C.1 to C.3 is possible. First, note that age increases with increasing  $t$  and thus

$$\frac{d\tau}{dt} = 1.$$

This suggests a transform of the age variable  $\tau \rightarrow \tau' = t - \tau$ , as in equation 6.46 of Gerstner and Kistler (2002). This new age variable,  $\tau'$ , is stationary with the evolution of  $t$ . We define the stationary backward recurrence time PDF as

$$f_s^-(\tau', t) := f^-(t - \tau', t).$$

---

<sup>3</sup>The ensemble firing rate is referred to as the population activity,  $A(t)$ , in Gerstner and Kistler, 2002.



Thus,

$$\begin{aligned}\frac{d}{d\tau} f_s^-(\tau', t) &= \frac{\partial}{\partial t} f_s^-(\tau', t) \\ &= \frac{\partial}{\partial t} (\alpha(\tau') \mathcal{F}(t - \tau', \tau')),\end{aligned}$$

and differentiation of equation C.3 yields  $\frac{d}{dt} \mathcal{F}(t - \tau', \tau') = \mathcal{F}(t - \tau', \tau') \rho(t - \tau', t)$  whereby we have

$$\begin{aligned}\frac{d}{dt} f_s^-(\tau', t) &= -\alpha(\tau') \mathcal{F}(t - \tau', \tau') \rho(t - \tau', t) \\ \frac{d}{dt} f_s^-(\tau', t) &= -f_s^-(\tau', t) \rho(t - \tau', t).\end{aligned}\tag{C.4}$$

This relation determines  $\frac{d}{dt} f_s^-(\tau', t)$  for  $\tau' \in (-\infty, t)$ . Additionally, we need to ensure that the normalization of  $f_s^-(\tau, t)$  is preserved, namely, that

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial t} f_s^-(\tau, t) d\tau = 0.\tag{C.5}$$

Splitting the integral into three regions of interest, we have

$$\begin{aligned}\lim_{\Delta t \rightarrow 0^+} \left[ \int_{-\infty}^{t-\Delta t} \frac{\partial}{\partial t} f_s^-(s, t) ds \right. \\ \left. + \int_{t-\Delta t}^{t+\Delta t} \frac{\partial}{\partial t} f_s^-(s, t) ds \right. \\ \left. + \int_{t+\Delta t}^{\infty} \frac{\partial}{\partial t} f_s^-(s, t) ds \right] = 0.\end{aligned}$$

Since  $f_s^-(\tau' > t, t) = 0$ , the third integral is zero. We then have

$$\lim_{\Delta t \rightarrow 0^+} \int_{t-\Delta t}^{t+\Delta t} \frac{\partial}{\partial t} f_s^-(s, t) ds = - \int_{-\infty}^t \frac{\partial}{\partial t} f_s^-(s, t) ds.$$

Since the contribution from equation C.4 in the integral on the left-hand side is vanishing, we need to add an additional term to  $\frac{d}{dt} f_s^-(\tau, t)$ , which is zero for all  $\tau \neq t$  but which has a finite integral when integrating around  $t$ . This can be achieved by addition of a  $\delta$ -function term:

$$\frac{d}{dt} f_s^-(\tau, t) \rightarrow \frac{d}{dt} f_s^-(\tau, t) - \delta(\tau - t) \int_{-\infty}^t \frac{\partial}{\partial t} f_s^-(s, t) ds.$$

Clearly the factor behind the  $\delta$ -function is  $\alpha(t)$ . Thus, we have the final form,

$$\frac{d}{dt} f_s^-(\tau', t) = \begin{cases} -f_s^-(\tau', t)\rho(t - \tau', t), & \tau' < t \\ 0, & \tau' \geq t \end{cases} + \alpha(t)\delta(\tau' - t), \quad (\text{C.6})$$

defined for  $\tau' \in (-\infty, \infty)$ . Equation C.6 expressed in terms of  $f^-(\tau, t)$  results in the master equation,

$$\frac{\partial}{\partial t} f^-(\tau, t) = -\frac{\partial}{\partial \tau} f^-(\tau, t) - f^-(\tau, t)\rho(\tau, t) + \alpha(t)\delta(\tau), \quad (\text{C.7})$$

defined for  $\tau \in [0, \infty)$ . This is exactly equation 6.43 in Gerstner and Kistler (2002). It is analogous to equation 2.9, but with reinsertion to  $\tau = 0$  after a spike.

**C.1 Numerical Solution of the Renewal Master Equation.** As the renewal master equation C.7 is just a special case of the 1DM master equation, it can be solved with the same numerical techniques as described in section 5.1. The content of the  $\delta$  term in equation C.7 is that all probability lost due to the death term (the second term on the rhs) is accumulated and reinserted to the  $\tau = 0$  bin. Thus, we are spared the complication of treating state-dependent reinsertion of probability, as was necessary for the 1DM and 2DM master equations.

**C.2 Generating Random Numbers of a General Hazard Function.** Random numbers with a general hazard function can be generated by the thinning method as discussed in Devroye (1986) and summarized here. The maximum of the hazard function,  $\rho_{\max} = \max_{\tau, t}(\rho(\tau, t))$ , must be known. Sequential event intervals are generated using a homogeneous Poisson process with a rate of  $\rho_{\max}$ . The resulting spike train is then sequentially thinned, given the event time  $t$  and time since last event  $\tau$ , by the rule:

1. Generate a uniform random number,  $T$ , on  $[0, 1)$ .
2. if  $\rho(\tau, t)/\rho_{\max} > T$ , keep the event in the spike train; otherwise remove it.

The remaining event times are consistent with the prescribed hazard function.

For the case of random number generation for a GRP, evaluation of  $\rho(\tau, t)$  using equation 4.6 is numerically unstable for large  $\tau$  and costly. An implementation of the algorithm (Shea, 1988) for calculating the cumulative hazard function of a gamma renewal process is available in `pgamma.c` of the Mathlib of the R statistics environment (Ihaka & Gentleman, 1996)

under the GNU Public License. Alternatively, the logarithm of the function `gsl_sf_gamma_inc_Q` provided by the GNU Scientific Library can be used. The hazard function can then be calculated by a simple discrete difference calculation of the derivative. Time dependence can be introduced by giving a time dependence to the parameters of the gamma distribution.

## Acknowledgments

---

This work is supported in part by the European Union under the grant IST-2005-15879 (FACETS). Thanks to M. Rudolph, A. Destexhe, M. Diesmann, N. Brunel, M. Pospischil and N. Nurpeissov for helpful discussion. Thanks to the reviewers for their many detailed and insightful comments which have had a significant and positive impact on this letter.

## References

---

- Abeles, M. (1991). *Corticonics*. Cambridge: Cambridge University Press.
- Amit, D. J., & Tsodyks, M. V. (1991). Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate-spikes, rates and neuronal gain. *Network*, 2, 259–273.
- Barbieri, R., Quirk, M., Frank, L., Wilson, M., & Brown, E. (2001). Construction and analysis of non-Poissonian stimulus-response models of neural spiking activity. *Journal of Neuroscience Methods*, 105, 25–37.
- Benda, J., & Herz, A. (2003). A universal model for spike-frequency adaptation. *Neural Computation*, 15, 2523–2564.
- Benda, J., Longtin, A., & Maler, L. (2005). Spike-frequency adaptation separates transient communication signals from background oscillations. *Journal of Neuroscience*, 25, 2312–2321.
- Benda, J., Longtin, A., & Maler, L. (2006). A synchronization-desynchronization code for natural communication signals. *Neuron*, 52, 347–358.
- Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94, 3637–3642.
- Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8, 183–208.
- Brunel, N., Chance, F. S., Fourcaud, N., & Abbott, L. F. (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Physical Review Letters*, 86, 2186–2189.
- Chacron, M., Longtin, A., & Maler, L. (2000). Suprathreshold stochastic firing dynamics with memory in p-type electroreceptors. *Physical Review Letter*, 85, 1576–1579.
- Chacron, M., Pakdaman, K., & Longtin, A. (2003). Interspike interval correlations, memory, adaptation, and refractoriness in a leaky integrate-and-fire model with threshold fatigue. *Neural Computation*, 15, 253–278.
- Cox, D. R. (1962). *Renewal theory*. London: Methuen.

- Dayan, P., & Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.
- Destexhe, A., Contreras, D., & Steriade, M. (1998). Mechanisms underlying the synchronizing action of corticothalamic feedback through inhibition of thalamic relay cells. *Journal of Neurophysiology*, *79*, 999–1016.
- Destexhe, A., Rudolph, M., Fellous, J., & Sejnowski, T. J. (2001). Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience*, *107*, 13–24.
- Destexhe, A., Rudolph, M., & Paré, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nature Reviews Neuroscience*, *4*, 739–751.
- Devroye, L. (1986). *Non-uniform random variate generation*. New York: Springer-Verlag.
- Diesmann, M., & Gewaltig, M.-O. (2002). NEST: An environment for neural systems simulations. In T. Plesser & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen, Beiträge zum Heinz-Billing-Preis 2001* (Vol. 58, pp. 43–70). Göttingen: Ges. für Wiss. Datenverarbeitung.
- Diesmann, M., Gewaltig, M.-O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, *402*, 529–533.
- Ermentrout, B., Pascal, M., & Gutkin, B. (2001). The effects of spike frequency adaptation and negative feedback on the synchronization of neural oscillators. *Neural Computation*, *13*, 1285–1310.
- Fourcaud, N., & Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Computation*, *14*, 2057–2110.
- Fuhrmann, G., Markram, H., & Tsodyks, M. (2002). Spike frequency adaptation and neocortical rhythms. *Journal of Neurophysiology*, *88*, 761–770.
- Gardiner, C. W. (1984). Adiabatic elimination in stochastic systems. I. formulation of methods and application to few-variable systems. *Physical Review A*, *29*, 2814–2823.
- Gardiner, C. W. (1985). *Handbook of stochastic methods*. Berlin: Springer-Verlag.
- Gazères, N., Borg-Graham, L. J., & Frégnac, Y. (1998). A phenomenological model of visually evoked spike trains in cat geniculate nonlagged x-cells. *Visual Neuroscience*, *15*, 1157–1174.
- Gerstner, W. (1995). Time structure of the activity in neural network models. *Physical Review E*, *51*, 738–758.
- Gerstner, W. (2000). Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Computation*, *12*, 43–89.
- Gerstner, W. (2001). Coding properties of spiking neurons: Reverse and cross-correlations. *Neural Networks*, *14*, 599–610.
- Gerstner, W., Kempter, R., van Hemmen, J. L., & Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, *383*, 76–78.
- Gerstner, W., & Kistler, W. (2002). *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge: Cambridge University Press.
- Haken, H. (1983). *Synergetics* (3rd ed.). Berlin: Springer-Verlag.
- Hines, M. L., & Carnevale, N. T. (1997). The NEURON simulation environment. *Neural Computation*, *9*, 1179–1209.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

- Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *Journal of General Physiology*, *59*, 734–766.
- Knight, B. W. (2000). Dynamics of encoding in neuron populations: Some general mathematical features. *Neural Computation*, *12*, 473–518.
- Knight, B. W., Omurtag, A., & Sirovich, L. (2000). The approach of a neuron population firing rate to a new equilibrium: An exact theoretical result. *Neural Computation*, *12*, 1045–1055.
- Koch, C. (1999). *Biophysics of computation: Information processing in single neurons*. New York: Oxford University Press.
- La Camera, G., Rauch, A., Lüscher, H.-R., Senn, W., & Fusi, S. (2004). Minimal models of adapted neuronal response to in vivo-like input currents. *Neural Computation*, *16*, 2101–2124.
- Latham, P. E., Richmond, B. J., Nelson, P. G., & Nirenberg, S. (2000). Intrinsic dynamics in neuronal networks. I. theory. *Journal of Neurophysiology*, *83*, 808–827.
- Lindner, B., & Longtin, A. (2003). Nonrenewal spike trains generated by stochastic neuron models. In *SPIE Proceedings 5114 (Fluctuations and Noise)* (pp. 209–218). Bellingham, WA: SPIE.
- Longtin, A., & Racicot, D. (1997). Spike train patterning and forecastability. *Biosystems*, *40*, 111–118.
- Mainen, Z., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, *268*, 1503–1505.
- McCormick, D., Connors, B., Lighthall, J., & Prince, D. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, *54*, 782–806.
- Meffin, H., Burkitt, A. N., & Grayden, D. B. (2004). An analytical model for the “large, fluctuating synaptic conductance state” typical of neocortical neurons in-vivo. *Journal of Computational Neuroscience*, *16*, 159–175.
- Moreno-Bote, R., & Parga, N. (2004). Role of synaptic filtering on the firing response of simple model neurons. *Physical Review Letters*, *92*, 92.028102.
- Moreno-Bote, R., & Parga, N. (2005). Membrane potential and response properties of populations of cortical neurons in the high conductance state. *Physical Review Letters*, *94*, 088103.
- Muller, E. (2003). *Simulation of high-conductance states in cortical neural networks*. M.Sc. thesis, University of Heidelberg.
- Nawrot, M. P., Boucsein, C., Rodriguez-Molina, V., Aertsen, A., Grün, S., & Rotter, S. (2007). Serial interval statistics of spontaneous activity in cortical neurons in vivo and in vitro. *Neurocomputing*, *70*, 1717–1722.
- Nykamp, D. Q., & Tranchina, D. (2000). A population density approach that facilitates large-scale modeling of neural networks: Analysis and an application to orientation tuning. *Journal of Computational Neuroscience*, *8*, 51–63.
- Nykamp, D. Q., & Tranchina, D. (2001). A population density approach that facilitates large-scale modeling of neural networks: Extension to slow inhibitory synapses. *Neural Computation*, *13*, 511–546.
- O’Brien, B. J., Isayama, T., Richardson, R., & Berson, D. M. (2002). Intrinsic physiological properties of cat retinal ganglion cells. *Journal of Physiology*, *538*, 787–802.

- Omurtag, A., Knight, B. W., & Sirovich, L. (2000). On the simulation of large populations of neurons. *Journal of Computational Neuroscience*, *8*, 51–63.
- Papoulis, A., & Pillai, S. U. (1991). *Probability, random variables and stochastic processes* (4th ed.). New York: McGraw-Hill.
- Renart, A., Brunel, N., & Wang, X.-J. (2004). Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. In J. Feng (Ed.), *Computational neuroscience: A comprehensive approach* (pp. 431–490). New York: Chapman&Hall/CRC.
- Ricciardi, L. (1977). *Diffusion processes and related topics in biology*. Berlin: Springer-Verlag.
- Richardson, M. J. E. (2004). Effects of synaptic conductance on the voltage distribution and firing rate of spiking neurons. *Physical Review E*, *69*(5), 051918.
- Richardson, M. J. E., & Gerstner, W. (2005). Synaptic shot noise and conductance fluctuations affect the membrane voltage with equal significance. *Neural Computation*, *17*, 923–947.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.
- Risken, H. (1996). *The Fokker-Planck equation: Methods of solution and applications* (2nd ed.). Berlin: Springer-Verlag.
- Rudolph, M., & Destexhe, A. (2003a). Characterization of subthreshold voltage fluctuations in neuronal membranes. *Neural Computation*, *15*, 2577–2618.
- Rudolph, M., & Destexhe, A. (2003b). The discharge variability of neocortical neurons during high-conductance states. *Neuroscience*, *119*, 885–873.
- Rudolph, M., & Destexhe, A. (2005). An extended analytic expression for the membrane potential distribution of conductance-based synaptic noise. *Neural Computation*, *17*, 2301–2315.
- Shadlen, M., & Newsome, W. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, *18*, 3870–3896.
- Shea, B. L. (1988). Algorithm AS 239, incomplete gamma function. *Applied Statistics*, *37*, 466–473.
- Shelley, M., McLaughlin, D., Shapley, R., & Wielaard, D. (2002). States of high conductance in a large-scale model of the visual cortex. *Journal of Computational Neuroscience*, *13*, 93–109.
- Siegert, A. J. F. (1951). On the first passage time probability function. *Physical Review*, *81*, 617–623.
- Smith, G. D., Cox, C. L., Sherman, S. M., & Rinzel, J. (2001). A firing-rate model of spike-frequency adaptation in sinusoidally-driven thalamocortical relay neurons. *Thalamus and Related Systems*, *1*, 135–156.
- Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, *13*, 334–350.
- Song, S., & Abbott, L. F. (2001). Cortical remapping through spike timing-dependent synaptic plasticity. *Neuron*, *32*, 1–20.
- Song, S., Miller, K., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, *3*, 919–926.

- van Rossum, M. C. W., Bi, G. Q., & Turrigiano, G. G. (2000). Stable Hebbian learning from spike timing-dependent synaptic plasticity. *Journal of Neuroscience*, *20*(23), 8812–8821.
- van Vreeswijk, C. A., & Hansel, D. (2001). Patterns of synchrony in neural networks with spike adaptation. *Neural Computation*, *13*, 959–992.
- Whittaker, E. T., & Robinson, G. (1967). *The calculus of observations: A treatise on numerical mathematics* (4th ed.). New York: Dover.
- Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.

---

Received November 7, 2005; accepted February 15, 2007.

Sami El Boustani and Alain Destexhe\*

# Brain dynamics at multiple scales: can one reconcile the apparent low-dimensional chaos of macroscopic variables with the seemingly stochastic behavior of single neurons?

(in press, 2009)

**Abstract** Nonlinear time series analyses have suggested that the human electroencephalogram (EEG) may share statistical and dynamical properties with chaotic systems. During slow-wave sleep or pathological states like epilepsy, correlation dimension measurements display low values, while in awake and attentive subjects, there is not such low dimensionality, and the EEG is more similar to a stochastic variable. We briefly review these results and contrast them with recordings in cat cerebral cortex, as well as with theoretical models. In awake or sleeping cats, recordings with microelectrodes inserted in cortex show that global variables such as local field potentials (local EEG) are similar to the human EEG. However, in both cases, neuronal discharges are highly irregular and exponentially distributed, similar to Poisson stochastic processes. To attempt reconcile these results, we investigate models of randomly-connected networks of integrate-and-fire neurons, and also contrast global (averaged) variables, with neuronal activity. The network displays different states, such as “synchronous regular” (SR) or “asynchronous irregular” (AI) states. In SR states, the global variables display coherent behavior with low dimensionality, while in AI states, the global activity is high-dimensionally chaotic with exponentially distributed neuronal discharges, similar to awake cats. Scale-dependent Lyapunov exponents and  $\varepsilon$ -entropies show that the seemingly stochastic nature at small scales (neurons) can coexist with more coherent behavior at larger scales (averages). Thus, we suggest that brain activity obeys similar scheme, with seemingly stochastic dynamics at small scales (neurons), while large scales (EEG) display more coherent behavior or high-dimensional chaos.

## 1 Introduction

A number of methods from nonlinear dynamical systems have been applied to the different states of the human EEG. Early studies have calculated correlation dimensions from EEG and reported evidence for low-dimensional chaos for slow-wave sleep [Babloyantz et al., 1985; Mayer-Kress et al., 1988], as well as for pathological states such as epilepsy [Babloyantz & Destexhe, 1986; Frank et al., 1990] or the terminal state of Creutzfeldt-Jakob disease [Destexhe et al., 1988]. These findings have been confirmed by numerous studies [reviewed in Destexhe, 1992; Elbert et al., 1994; Korn & Faure, 2003]. Interestingly, in these studies, the EEG dynamics during wakefulness or REM sleep did not show evidence for low-dimensional dynamics. These results were also corroborated with other measurements such as Lyapunov exponents, entropies, power spectral densities, autocorrelations and symbolic dynamics [Destexhe, 1992].

The existence of low-dimensional chaotic dynamics in systems with such large degrees of freedom is surprising. Chaotic dynamics in extended coupled systems have been studied extensively for the last decades and is still a matter of intense investigation. In particular, for network of simplified neurons, as the spin-glass models, mean-field theories have proven the existence of stable chaotic attractors [Sompolinsky et al., 1988; van Vreeswijk & Sompolinsky, 1996; van Vreeswijk & Sompolinsky, 1998]. These systems have been shown to exhibit chaotic persistence regarding parameter changes [Albers et al., 2006], a property that makes chaotic dynamics more common than exceptional [Sprott, 2008]. From a computational point of view, chaotic behaviour or nearly chaotic regime (edge of chaos) can be optimal for information processing [Bertschinger & Natschläger, 2004; Legenstein & Maass, 2007]. In network models of spiking neurons, chaotic regimes have also been studied [Cessac, 2008; Cessac & Viéville, 2008] and can be present only as transients during which the system is numerically indistinguishable from a usual chaotic attractor. However the lifetime of these transient periods is known to increase exponentially with network size [Crutchfield & Kaneko, 1988; Tél &

---

Unité de Neurosciences Intégratives et Computationnelles (UNIC),  
CNRS, Gif-sur-Yvette, France

\*: Corresponding author at the following address:

UNIC, Bat 33, CNRS, 1 Avenue de la Terrasse, 91198 Gif sur Yvette,  
France.

E-mail: Destexhe@unic.cnrs-gif.fr



Lai, 2008] making them more relevant in practice than the real stable attractor for large enough networks.

In this paper, we intend to study a specific spiking neuron model which displays these properties and yield biophysical behaviour relevant to understand EEG data. Classical nonlinear tools as the correlation dimension or the Lyapunov exponents have given some insight on macroscopic quantities as the EEG or the overall activity of numerical models but are often criticized because of their limitation for low-dimensional dynamics [Kantz & Schreiber, 2004]. In order to distinguish between microscopic dynamics and collective behaviour, we borrow recent tools developed in the context of high-dimensional systems and which offer analysis at different scales. In particular, finite-size Lyapunov exponent [Aurell et al., 1997] as well as  $\varepsilon$ -entropy [Gaspard & Wang, 1993] provide a scale-dependent description of spiking neuron networks and could detect whether or not low-dimensional dynamics prevail at a macroscopic scales. These tools will be applied to EEG data as well as numerical models.

## 2 Methods

In this section we briefly describe the analytical and numerical tools which will be used to probe the nonlinear nature of the EEG recordings or the numerical simulations. Most of these analysis are extensively used in the literature. Furthermore, we used the TISEAN toolbox [Hegger et al., 1999] to perform most of our nonlinear analysis. We also give information about the numerical model.

### 2.1 Phase space reconstruction

We used the method of time-delayed vectors of the time series, which yields reconstructed attractors topologically equivalent to the original attractor of the system [Takens, 1981; Sauer et al., 1991]. We chose a fixed delay parameter determined by the first minimum of the mutual information [Fraser & Swinney, 1986]. The embedding dimension was chosen such that any self-similar asymptotic behavior saturates beyond this dimension, indicating a successful attractor reconstruction.

### 2.2 Correlation dimension

The correlation dimension was measured from the correlation integral [Grassberger & Procaccia, 1983] which is estimated by using

$$C(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N \Theta(\varepsilon - \|\mathbf{x}(i) - \mathbf{x}(j)\|) \quad (1)$$

where  $\mathbf{x}(i) \in \mathbb{R}^m$  is the  $m$ -dimensional delay vector and  $\varepsilon$  is a threshold distance which reflects the scale under consideration. Moreover we used a Theiler window [Theiler, 1990]

according to the temporal correlation of each time series to avoid spurious estimation. If the correlation integral manifests a power-law behavior over a range of scale which saturates with increasing embedding dimension, the correlation dimension is simply the corresponding power-law coefficient.

### 2.3 Recurrence plots

This tool provides an intuitive visualization of recurrences in trajectories on the attractor [Eckmann et al., 1987]. The recurrence matrix is defined by

$$RP(i, j) = \Theta(\varepsilon - \|\mathbf{x}(i) - \mathbf{x}(j)\|) \quad (2)$$

where the matrix entry is equal to 1 if the trajectory  $\mathbf{x}$  at  $i$  is in the  $\varepsilon$ -neighbor of itself at  $j$  and 0 otherwise. Stereotypic patterns in the recurrence plot can indicate the existence of periodic behavior, low-dimensional chaos or stochastic-like dynamics. In particular, diagonal lines are typical of periodic dynamics whereas clouds of dots are produced by a stochastic component. For an extensive review on recurrence plot, the reader should refer to [Marwan et al., 2006].

### 2.4 Finite-size Lyapunov and $\varepsilon$ -entropy

Scale-dependent nonlinear analysis have been greatly studied the last decade and have come in handy to distinguish deterministic (possibly chaotic) dynamics from stochastic noise. In this paper, we will mainly consider two quantities : the Finite-Size Lyapunov Exponent (FSLE) and the  $\varepsilon$ -entropy. The former has been introduced in the context of developed turbulence [Aurell et al., 1997] and has proven to be more suited for a broad range of system which dynamics can exhibit low-dimensional chaotic behavior only on large-scale [Shibata & Kaneko, 1998; Cencini et al., 1999; Gao et al., 2006]. Roughly speaking, if we want to quantify the sensitivity to initial conditions on large scales, it is necessary to consider perturbations which are not infinitesimal. Therefore, the FSLE can be defined as follows

$$\lambda(\delta) = \frac{1}{\langle T_r(\delta) \rangle} \left\langle \ln \left( \frac{\Delta_r}{\delta} \right) \right\rangle \quad (3)$$

where for a perturbation between two initial trajectories  $\delta$ ,  $T_r$  is the minimal time required for those trajectories to be separated by a distance  $\Delta_r$  greater than or equal to  $\delta r$  where  $r = 2$  is usually taken. The brackets in Eq. 3 denotes averages on the attractor for many realizations. The  $\varepsilon$ -entropy is a generalization of the Kolmogorov-Sinai entropy rate [Gaspard & Wang, 1993] which is defined for a finite scale  $\varepsilon$  and time delay  $\tau$  by

$$h(\varepsilon, \tau) = \lim_{m \rightarrow \infty} h_m(\varepsilon, \tau) \quad (4)$$

$$= \frac{1}{\tau} \lim_{m \rightarrow \infty} \frac{1}{m} H_m(\varepsilon, \tau) \quad (5)$$

with

$$h_m(\varepsilon, \tau) = \frac{1}{\tau} (H_{m+1}(\varepsilon, \tau) - H_m(\varepsilon, \tau)) \quad (6)$$

where  $H_m(\varepsilon, \tau)$  is the entropy estimated with box partition of the phase space for box size given by  $\varepsilon$  on the attractor reconstructed with a time delay  $\tau$  and an embedding dimension  $m$ . This quantity exhibits a plateau on particular scales if a deterministic low-dimensional dynamics occurs at these scales. It can thus be used to describe large scale dynamics independently of the small scale noisy behavior or can be used to distinguish noise from chaos in some cases [Cencini, 2000].

### 2.5 Avalanche analysis

To identify the presence of scale invariance, typical of self-organized critical states [Jensen, 1998], we used an ‘‘avalanche analysis’’ methods [Beggs and Plenz, 2003]. This method consists of detecting ‘‘avalanches’’ as clusters of contiguous events separated by silences, by binning the system’s activity in time windows (1 ms to 16 ms were used). Cluster of events were defined from the spike times among the ensemble of simultaneously recorded neurons. The scale invariance was determined from the distribution of avalanche size, calculated as the total number of events [Beggs and Plenz, 2005].

### 2.6 Network of Integrate-and-Fire neurons

Networks of ‘‘integrate-and-fire’’ neurons were simulated according to models and parameters published previously [Vogels & Abbott, 2005; El Boustani et al., 2007]. The network consisted of 5,000 neurons, which were separated into two populations of excitatory and inhibitory neurons, forming 80% and 20% of the neurons, respectively. All neurons were connected randomly using a connection probability of 2%.

The membrane equation of cell  $i$  was given by:

$$C_m \frac{dV_i}{dt} = -g_L(V_i - E_L) + S_i(t) + G_i(t), \quad (7)$$

where  $C_m = 1 \mu\text{F}/\text{cm}^2$  is the specific capacitance,  $V_i$  is the membrane potential,  $g_L = 5 \times 10^{-5} \text{ S}/\text{cm}^2$  is the leak conductance density and  $E_L = -60 \text{ mV}$  is the leak reversal potential. Together with a cell area of  $20,000 \mu\text{m}^2$ , these parameters give a resting membrane time constant of 20 ms and an input resistance at rest of  $100 \text{ M}\Omega$ . The function  $S_i(t)$  represents the spiking mechanism intrinsic to cell  $i$  and  $G_i(t)$  stands for the total synaptic current of cell  $i$ . Note that in this model, excitatory and inhibitory neurons have the same properties.

In addition to passive membrane properties, IF neurons had a firing threshold of  $-50 \text{ mV}$ . Once the Vm reaches threshold, a spike is emitted and the membrane potential is reset to  $-60 \text{ mV}$  and remains at that value for a refractory period of 5 ms. This model was inspired from a previous publication

reporting self-sustained irregular states [Vogels and Abbott, 2005].

Synaptic interactions were conductance-based, according to the following equation for neuron  $i$ :

$$G_i(t) = - \sum_j g_{ji}(t)(V_i - E_j), \quad (8)$$

where  $V_i$  is the membrane potential of neuron  $i$ ,  $g_{ji}(t)$  is the synaptic conductance of the synapse connecting neuron  $j$  to neuron  $i$ , and  $E_j$  is the reversal potential of that synapse.  $E_j$  was 0 mV for excitatory synapses, or  $-80 \text{ mV}$  for inhibitory synapses.

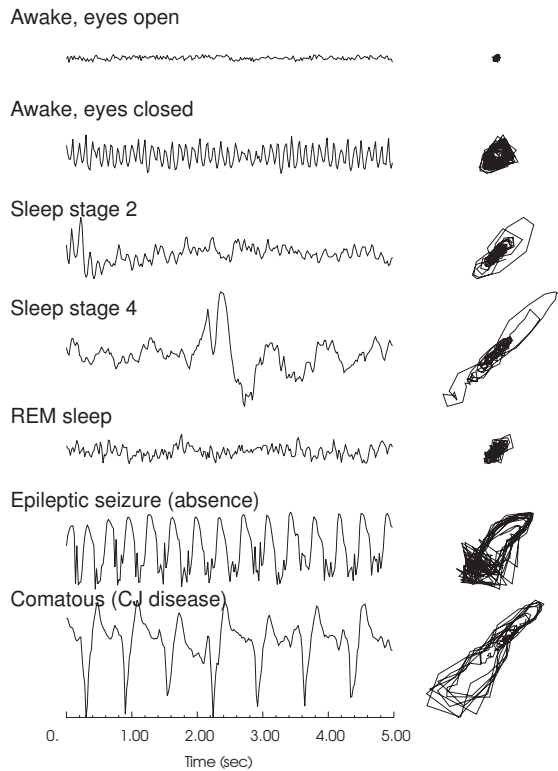
Synaptic interactions were implemented as follows: when a spike occurred in neuron  $j$ , the synaptic conductance  $g_{ji}$  was instantaneously incremented by a quantum value ( $q_e = 6 \text{ nS}$  and  $q_i = 67 \text{ nS}$  for excitatory and inhibitory synapses, respectively) and decayed exponentially with a time constant of  $\tau_e = 5 \text{ ms}$  and  $\tau_i = 10 \text{ ms}$  for excitation and inhibition, respectively.

## 3 Evidence for chaotic dynamics in EEG activity

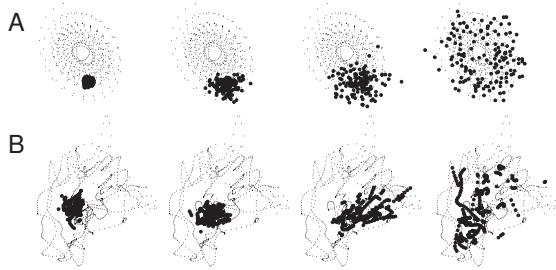
Human EEG recordings during different brain states are illustrated in Fig. 1, along with a 2-dimensional representation of the phase portrait obtained from each signal. During wakefulness (eyes open) and REM sleep, the dynamics is characterized by low-amplitude and very irregular EEG activity, while during deep sleep, the EEG displays slow waves (‘‘delta waves’’) of large amplitude. Oscillatory dynamics with a frequency around 10 Hz is seen in the occipital region when the eyes are closed (‘‘alpha rhythm’’). Pathological states, such as epilepsy or comatous states, display large amplitude oscillations, which are strikingly regular.

A first evidence for chaotic dynamics is that EEG dynamics display a prominent sensitivity to initial conditions. This sensitivity is illustrated in Fig. 2 for the alpha rhythm (awake eyes closed) and slow-wave sleep (stage IV). A set of close initial conditions is defined by choosing a neighboring points in phase space, and this set of points is followed in time. The divergence of the trajectories emanating from each initial condition is evident from the illustration of Fig. 2, and is actually exponential. This exponential divergence betrays the presence of at least one positive Lyapunov exponent. A more quantitative investigation using numerical methods [Wolf et al., 1985] reveals the presence of positive Lyapunov exponents for all EEG states investigated [Destexhe, 1992].

EEG dynamics also displays other characteristics of chaotic dynamical systems, such as broad-band power spectra (not shown) and fractal attractor dimensions. This latter point was investigated by a number of laboratories, and is summarized in Fig. 3. Correlation integrals  $C(r)$  are calculated from the reconstructed phase portraits for different embedding dimensions (Fig. 3A-B). The correlation dimension  $d$  is obtained by estimating the scaling of  $C(r)$  by using logarithmic representations, in which the slope directly gives an



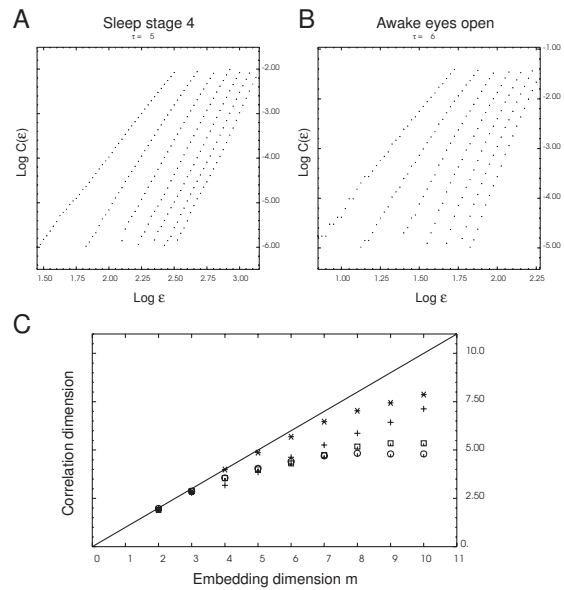
**Fig. 1** Electroencephalogram signals and phase portraits during different brain states in humans. Left: 5 seconds of EEG activity in different brain states (same amplitude scale). Right: 2-dim phase portrait of each signal. Modified from [Destexhe, 1992].



**Fig. 2** Illustration of the sensitivity to initial conditions in EEG dynamics. A cluster of neighboring points in phase space (leftmost panels) is followed in time and is shown on the same phase portrait after 200 ms, 400 ms and 3 seconds (from left to right; same data as in Fig. 1). A. Awake eyes closed; B. Sleep stage IV. Modified from [Destexhe, 1992].

estimate of  $d$ . The calculation of  $d$  as a function of the embedding dimension (Fig. 3C) saturates to a constant value for some states (such as deep sleep or pathologies), but not for others. This is the case for the EEG during wakefulness, which dimension  $d$  does not saturate, which is a sign of high dimensionality.

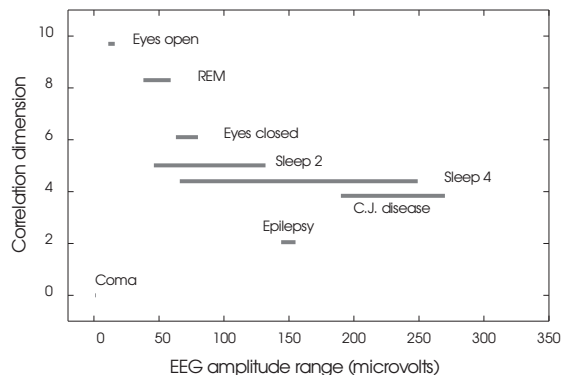
The correlation dimensions obtained for different brain states are represented in Fig. 4 as a function of the mean



**Fig. 3** Correlation dimension calculated from different brain states in humans.

A. Correlation integrals  $C(r)$  calculated from sleep stage IV. B. Correlation integrals calculated from Awake eyes open. C. Correlation dimension as a function of embedding dimension for different brain states. Symbols: \* = Awake eyes open, + = REM sleep, squares = sleep stage 2, circles = sleep stage 4. Modified from [Destexhe, 1992].

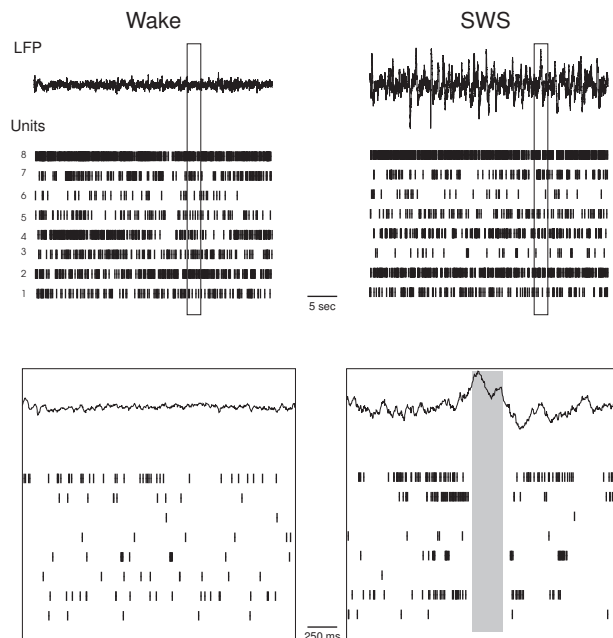
amplitude of the EEG signal. This representation reveals a “hierarchy” of dimensionalities for the EEG. The aroused states, such as wakefulness and REM sleep, are characterized by high dimension and no sign of slow-wave activity. As the brain drifts towards sleep, the dimensionality decreases and attains its lowest level during the deepest phase of sleep, in which the EEG is dominated by large-amplitude slow waves. A further decrease is seen during pathologies, which are also dominated by slow-wave activity.



**Fig. 4** Dimension – amplitude representation for different brain states. The correlation dimension of the EEG is shown as a function of the amplitude range (maximal amplitude deflection calculated over 1 second periods). Modified from [Destexhe, 1992].

#### 4 Evidence for stochastic dynamics in brain activity

The above results are consistent with the idea that awake brain activity may be associated with high-dimensional dynamics, perhaps analogous to a stochastic system. To further investigate this aspect, we have examined data from animal experiments in which both microscopic (cells) and macroscopic (EEG) activities can be recorded. The correspondence between these variables is shown in Fig. 5 for cat cerebral cortex during wakefulness and slow-wave sleep.



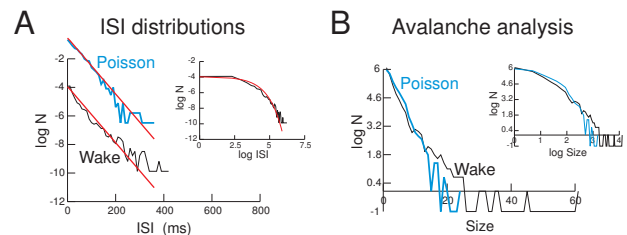
**Fig. 5** Distributed firing activity and local field potentials in cat cortex during wake and sleep states.

Recordings were done using eight bipolar tungsten electrodes in cat parietal cortex (data from [Destexhe et al., 1999]). The irregular firing activity of 8 multi-units is shown at the same time as the LFP recorded in electrode 1. During wakefulness, the activity is sustained and irregular (see magnification below). During slow-wave sleep (SWS), the activity is similar as wakefulness, except that “pauses” of firing occur in all cells, and in relation to the slow waves (one example is shown in gray in the bottom graph). The boxes in the top graphs are shown in bottom graphs at 20 times higher temporal resolution.

Those electrical measurements were made using tungsten microelectrodes directly inserted in cortical gray matter [Destexhe et al., 1999]. This recording system enables the extraction of two signals: a global signal, similar to the EEG, which is called “local field potential” (LFP) and reflects the averaged electrical activity of a large population of neurons. In addition, single neurons can be distinguished and can be extracted. These signals are shown and compared in Fig. 5. During wakefulness, the LFP shows low-amplitude irregular activity, while neuronal discharges seem random. Slow-wave sleep is characterized by dominant delta waves in the LFP, as in human sleep stage IV. The occurrence of slow

waves is correlated with a concerted pause in the firing of the neurons [Destexhe et al., 1999].

Analyzing the spike discharge of single neurons revealed that the interspike-intervals (ISI) are exponentially distributed, in a manner indistinguishable from a Poisson stochastic process (Fig. 6A). Performing an avalanche analysis revealed that the distribution of avalanche size from the neuronal discharges was also exponential (Fig. 6B). The same scaling could be explained by uncorrelated Poisson processes, as if the neurons discharged randomly and independently. This analysis was reproduced from a previous study (Bedard et al., 2006).



**Fig. 6** Analysis of neuronal activity in awake cat cerebral cortex.

A. Interspike interval (ISI) distributions computed from extracellularly recorded neurons in wakefulness (natural logarithms). A Poisson process with the same average rate is shown for comparison (blue; curve displaced upwards for clarity). The inset shows a log-log representation. The red curves indicate the theoretical value for Poisson processes. B. Avalanche analysis of extracellular recordings in the awake cat (natural logarithms). The same analysis was performed on surrogate data (blue; Poisson processes). The inset shows the same data in log-log representation. Modified from Bedard et al., 2006.

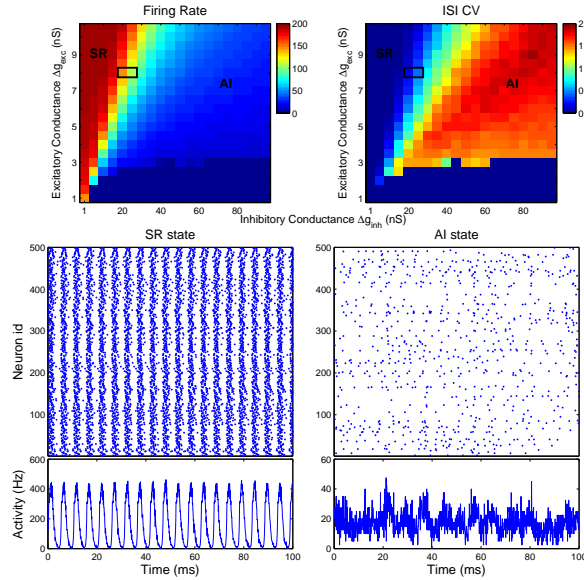
Thus, these data and analysis show that the dynamics of neuronal activity in the awake cerebral cortex is similar to stochastic processes. The statistics of the ISI distributions, as well as the collective dynamics (“avalanches”) cannot be distinguished from a Poisson stochastic processes.

#### 5 Models of irregular dynamics in neuronal networks

We next consider if this type of dynamics can be found in theoretical models of neuronal networks. We consider randomly-connected networks of excitatory and inhibitory point neurons, in which firing activity is described by the “integrate-and-fire” model, while synaptic interactions are conductance-based. As shown in previous publications (Vogels and Abbott, 2005; El Boustani et al., 2007), this model can display states of activity consistent with recordings in awake neocortex, as shown in Fig. 7. The network can display different types of states, such as “synchronous regular” (SR) states, or “asynchronous irregular” (AI), both of which are illustrated in Fig. 7. In SR states, the activity is oscillatory and synchronized between neurons, while in AI states, neurons are desynchronized and fire irregularly, similar to recordings in awake cats (compare with cells in Fig. 5). The averaged activity of the network is coherent and of high amplitude



in SR states, but is very noisy and of low amplitude in AI states, similar to the EEG or LFP activity seen in wakefulness (compare with EEGs in Fig. 1 and LFPs in Fig. 5).

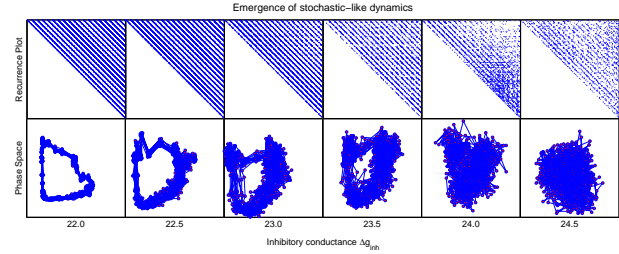


**Fig. 7** Different network states in randomly-connected networks of spiking neurons.

Top: state diagrams representing the average firing rate (left) and the firing irregularity (ISI CV, right) of a randomly-connected network of 5,000 integrate-and-fire neurons with a sparse connectivity of 2%. The phase diagram is drawn according to the excitatory and inhibitory quantal conductance. For increasing inhibition the dynamics undergo a transition from synchronous firing among the neuron and regular firing to a state where the synchrony has substantially decreased and neurons fire irregularly. The black rectangle indicates a transition from the SR to the AI regime. We selected this region for the rest of the study (see next figures). Bottom: the two different states, “Synchronous Regular” (SR, left) and “Asynchronous Irregular” (AI, right). The corresponding regions in the state diagram are indicated on top. For both panels, the upper part show a raster plot of a 500 neurons taken randomly from the network. Each dot is a spike emitted by the corresponding neuron across time. In the lower part, the mean firing rate computed among the whole network with a time bin of 0.1 ms. The asynchronous activity is reflected through the less fluctuating mean activity.

The time series of these averaged network activities (bottom graphs in Fig. 7) were analyzed similarly to the EEG in Section 3. We were particularly interested in the transition region between AI and SR states where the firing loses its coherence. This trajectory in the phase diagram is shown as a black rectangle in the upper panels of Fig. 7 and is characterized by an abrupt drop of activity concomitant with an increasing irregular firing. Similar results can be obtained by reducing the excitatory synaptic strength instead. However, to avoid a loss of stability, we preferred to study the transition by increasing the inhibitory synaptic strength. Phase portraits of SR and AI states, as well as the corresponding recurrence plots, are illustrated in Fig. 8. As expected, SR states (leftmost panel in Fig. 8) display a limit cycle phase portrait, while AI states (rightmost panel) appear as a dense

unstructured attractor. When the dynamics is dominated by inhibition the limit cycle is rapidly blurred by small-scale fluctuations which become ubiquitous for AI states. However, even though the phase portrait does not display any structure, the recurrence plot still exhibit a strong deterministic component (diagonal line) which is completely lost for a state well beyond the boundary region (noisy recurrence plot).

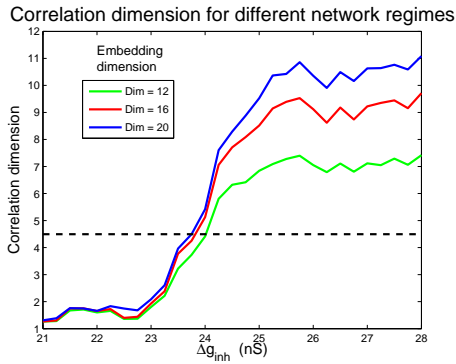


**Fig. 8** Recurrence plot and phase portraits of different network states. From left to right, different step of the transition from the SR state (leftmost) to AI state (rightmost). All states differed by the value of the inhibitory conductance  $\Delta g_{inh}$ , as indicated. The top graphs indicate the recurrence plot. The diagonals indicate the existence of a dominant periodic component. As the slaved degrees of freedom are unleashed with increasing inhibition, the recurrence plot is blurred by the stochastic-like component emerging at larger scales. This is the manifestation of the transition to a high-dimensional dynamics. For network states in the middle of the AI region, there is almost no visible recurrence and the recurrence plot looks like that of a stochastic process. The bottom graphs show bi-dimensional phase space reconstruction of the corresponding state. We can clearly see the “noisy” limit cycle which progressively degenerates into a dense high-dimensional attractor confirming what is shown by the recurrence plot.

To quantify this progressive release of degree of freedom, we estimated the correlation dimension on the overall activity for those different states. Fig. 9 shows the computed dimension for three different embedding dimensions. The network activity displays a low-dimensional dynamics as expected for SR state whereas the dimension suddenly increases with increasing inhibition conductance. Moreover, the corresponding values don’t saturate with increasing embedding dimension suggesting a high-dimensional attractor. Therefore, beyond the dotted black line, the correlation dimension is not a relevant measure anymore.

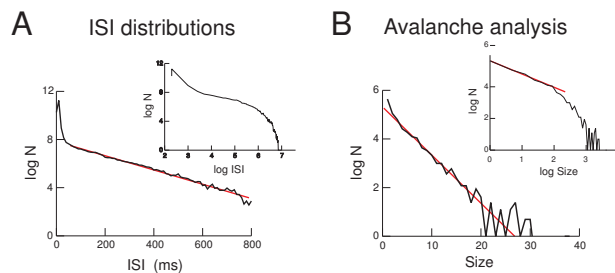
We also analyzed the spiking activity of the model using the same statistical tools as for experimental data. The distribution of ISIs during AI states was largely dominated by an exponential component (Fig. 10A, very similar to experimental data (compare with Fig. 6A)). Analyzing population activity, through “avalanche analysis”, displayed exponential distributions (Fig. 10B), also similar to experimental data (compare with Fig. 6B). Power-law scaling was also present for a limited range (see Fig. 10B, inset). Similar results were obtained for different network states (not shown).

One can ask what would be the effect of a more specific connectivity architecture on these results. In previous work, we have shown that macroscopic properties in these



**Fig. 9** Transition to a high-dimensional attractor in the AI regime. The correlation dimension is plotted according to the network state from the SR to the AI regime. Saturation of the scaling region in the correlation integral is shown by illustrating the estimated correlation dimension for three embedding dimensions. Below the dotted line, the measure is considered to saturate correctly whereas above this line, the measure can not be trusted anymore. In particular for low inhibition conductances, there are severe shifts with increasing embedding dimension. The dotted line is given by the disappearance of the plateau in the  $\epsilon$ -entropy of Fig. 12.

networks are conserved for more local connectivity, as long as the the connexions remain sparse [El Boustani & Destexhe, 2009]. Therefore, the phase diagram for topological networks owns similar structures as those displayed in Fig. 7. However, in the limit of “first-neighbor” connectivity, correlations between neurons become significantly stronger and the resulting neuronal dynamics are more regular and far from biological observations. Thus, even though more local architectures can result in more correlated activity, these correlations have to remain small in order to preserve the network stability and the rich repertoire of regimes. Our results are thus general for locally-connected networks as long as the sparseness is respected.



**Fig. 10** Analysis of the network dynamics in a model of asynchronous irregular states.

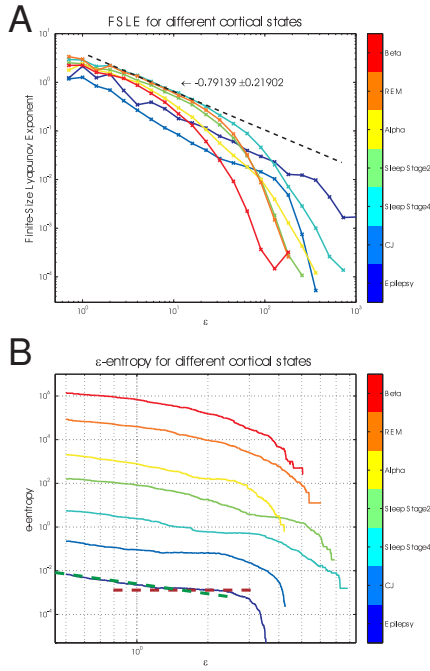
A. ISI distributions during the AI state in a randomly-connected network of 16,000 integrate-and-fire neurons. ISI distributions are exponential (noisy trace), as predicted by a Poisson process (red line; inset: log-log representation using natural logarithms). B. Absence of avalanche dynamics in this model (same description as Fig. 6B). The red lines indicates regions of exponential scaling; the red line in inset indicates a region with power-law scaling. Modified from El Boustani et al., 2007.

## 6 How to reconcile these data ?

The above data show that in both human, cat and models, the dynamics can show clear signs of coherence and low dimensionality at the level of macroscopic measurements (EEG, LFPs, averages), while microscopically, neuronal dynamics are highly irregular and resemble stochastic processes. In an attempt to reconcile these observations, we rely on recently introduced generalization of classical nonlinear tools. In particular, Finite-Size Lyapunov Exponent [Aurell et al., 1997; Shibata & Kaneko, 1998; Cencini et al., 1999; Gao et al., 2006] and the  $\epsilon$ -entropy [Cencini et al., 2000] have proven to be valuable measures to probe system which display different behaviors at different scales. In the present context, where low-dimensional dynamics can take place in top of a highly irregular neuronal behavior, those measures appear as the most natural.

They were applied to the human EEG, as shown in Fig. 11. The FSLE in the upper panel neither possess a plateau nor behave according to a power-law except in the small-scale limit where the stochastic-like component is dominant. For Creutzfeldt-Jacob disease as well as for epileptic seizure, however, it seems that a small region displays a almost scale-free behavior which would indicate a low-dimensional chaotic dynamics. Following [Shibata & Kaneko, 1998], it seems that the cortical activity lives in a high-dimensional attractor which could be chaotic. Therefore, the FSLE does not really help to untangle the different scales dynamics here. However, when resorting to the  $\epsilon$ -entropy, we get a different picture. Fig. 11B clearly shows that most of these cortical states manifest a large plateau on large-scale. These plateau is a signature of low-dimensional dynamics [Cencini et al., 2000] whereas the small-scale power-law behavior is the signature of stochastic-like dynamics produced by high-dimensional attractor. In accordance with Fig. 3, the dynamics during awake eyes open and REM sleep do not own a plateau, and their attractor dimension is too high to be distinguished from noise. However, it should be noted that those different dynamics are produced by the same system hence being mainly modulated by endogenic factors. In light of this result, we can conclude that there is no contradiction between the experimental data acquired at the macroscopic level (EEG) and the at the microscopic level (Spiking Activity). For large network of coupled units, averaged quantities can display low-dimensional structured dynamics while being seemingly stochastic at a smaller scale.

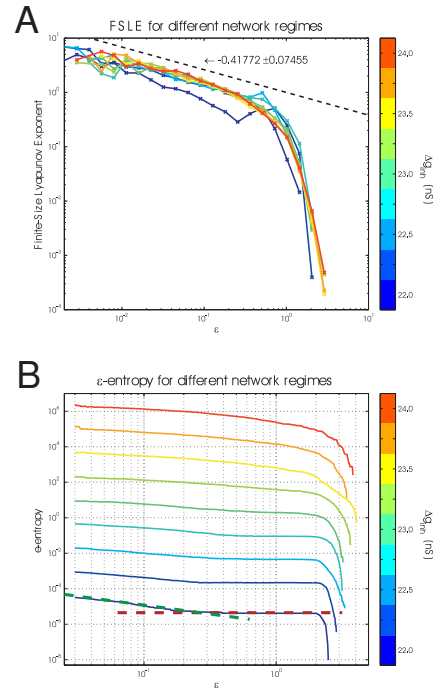
We also evaluated the same quantities from the averaged activities of the numerical model. In Fig. 12A, we see as before that the FSLE does not exhibit any scale-free region except for the small-scale limit. Thus we can not rely on this measure to distinguish small- and large- scale behavior. In contrast, the  $\epsilon$ -entropy in Fig. 12B yields a large and distinct plateau in the SR regime. The disappearance of this plateau have been used as a criteria to draw the dotted black line in Fig. 9. For inhibitory conductances larger than  $\Delta g_{inh} \simeq 23.5$  (nS), the  $\epsilon$ -entropy slowly converges to its small-scale stochastic-like behavior. We thus recover a very



**Fig. 11** Scale-dependent Lyapunov exponent and Epsilon-entropy for different brain states.

The scale-dependent Lyapunov exponent (A) and  $\epsilon$ -entropy (B) were calculated from the same EEG states as shown in Fig. 1. The FSLE (A) does not show any plateau and if the inverse FSLE is plotted in a semilog scale on the x-axis, no low-dimensional chaotic behavior can be detected for most states (data not shown). Following [Shibata & Kaneko, 1998] example for general coupled maps, the macroscopic activity seems to exhibit high-dimensional chaotic dynamics. For small scales  $\epsilon \rightarrow 0$ , a scale-free region is found with a power-law exponent around 0.791 (black dotted line) as expected from the microscopic stochastic-like dynamics [Cencini et al., 2000; Gao et al., 2006]. Because of the absence of a large scale-free region in the semilog region, we can not define a macroscopic Lyapunov exponent from this measure. However, for the Creutzfeldt-Jacob disease and the epileptic seizure, the FSLE have a free-scale region on a broad range which is different from the other cortical states. This could indicate a low-dimensional chaotic dynamic even though it is not reflected in the semilog scale. Indeed, the  $\epsilon$ -entropy (B) manifests a clear plateau (red dotted line) for the epileptic seizure, the Creutzfeldt-Jacob disease, sleep stage 2 and 4 and the alpha waves. These plateau indicates the existence of a low-dimensional attractor on the corresponding scales [Cencini et al., 2000] which can sometimes be easily visualized with an embedding procedure (see Fig. 1). For REM and awake states, this plateau disappears leaving the dynamics in a stochastic-like (high-dimension) state at all scales. The small-scale behavior is identical to the FSLE (green dotted line).

comparable behaviour to the EEG data where pathological states or deep sleep produce structured activity at the EEG level and awake state or REM are comparable to the dynamics of AI regime, as suggested previously [van Vreeswijk & Sompolinsky, 2006; van Vreeswijk & Sompolinsky, 2008; Vogels & Abbott, 2005; El Boustani et al., 2007; Kumar et al., 2008].



**Fig. 12** Scale-dependent Lyapunov exponent and Epsilon-entropy for different network states.

The Finite-Size Lyapunov Exponent (A) and  $\epsilon$ -entropy (B) were calculated from the same network states as shown in Fig. 7. The FSLE behaves almost identically for every network state. There is no evidence for a low-dimensional chaotic dynamics which would be indicated by a scale-free region in intermediate scales. The power-law at small-scale owns a coefficient around 0.418 (black dotted line) characterizing the stochastic-like dynamic (high-dimensional deterministic) at those scales. Even though no evidence for low-dimension chaos can be found from the FSLE, the  $\epsilon$ -entropy (B) has a large plateau (red dotted line) for state lying in the SR region (low inhibitory conductance). This plateau is shortened when the network dynamics is driven toward the AI regime where eventually the dynamics is indistinguishable from stochastic process. The small-scale behavior is identical to the FSLE (green dotted line).

## 7 Discussion

In this paper, we have briefly reviewed correlation dimension analyses of human EEG, which revealed a hierarchy of brain states, where the dimensionality varies approximately inversely to the level of arousal (Fig. 4). In particular, awake and attentive subjects display high dimensionalities, while deep sleep of pathological states show evidence for low dimensionalities. These low dimensions are difficult to reconcile with the fact that these signals emanate from the activity of millions (if not billions) of neurons. In cat cerebral cortex, the cellular activity during wake and sleep are highly irregular, and exponentially distributed like stochastic (Poisson) processes, a feature which is also difficult to reconcile with low dimensionalities at the EEG level, even though LFP data do not manifest low-dimensional dynamics (data not shown). We performed similar analyses on computational

models, which also display these apparently coherent activities at the level of large-scale averages, while the microscopic activity is highly irregular. In particular SR states can display coherent behavior at large scales, while AI states do not show evidence for coherence, similar to recordings in awake cats and humans.

The nature of the dynamics exhibited by models is assimilable to high-dimensional chaos. It has been shown that AI states in such models shut down after some time, and are thus transient in nature [Vogels & Abbott, 2005; El Boustani et al., 2007; Kumar et al., 2008; El Boustani & Destexhe, 2009]. This lifetime has been estimated to increase exponentially with the network size [Kumar et al., 2008; El Boustani & Destexhe, 2009], and can reach considerable times (beyond any reasonable simulation time; [Kumar et al., 2008]). Moreover, other recent studies [Cessac, 2008; Cessac & Viéville, 2008] have obtained analytical results on similar models where “transient chaotic-like regimes” were found. More precisely, these regimes are periodic, but with a period which also grows exponentially with network size. These results are reminiscent of the non-attractive chaotic manifolds extensively discussed in the literature [Crutchfield & Kaneko, 1988; Dhamala et al., 2001; Dhamala & Lai, 2002; Tél & Lai, 2008; Zillmer et al., 2006]. Hence, even though the chaotic nature of the dynamics is not inherent of the underlying system, the network spends a long enough period trapped in this transient dynamics indistinguishable from chaos from a numerical point of view.

To characterize the dynamics at different scales, we estimated quantities such as FSLE or  $\varepsilon$ -entropy, which clearly show that microscopic scales (neurons) tend to be very high-dimensional and complex, in many ways similar to “noise”, while more coherent behavior can be present at large scales. This analysis is consistent with the recently proposed concept of “macroscopic chaos”, where a very high-dimensional microscopic dynamics coexists with low dimensionality at the macroscopic level. In this paper, the chosen numerical model is known to be purely deterministic and can display highly irregular spiking patterns close to stochastic processes or noise. From a nonlinear analysis point of view, most of this dynamics is indistinguishable from a random process. However, as soon as a large scale behaviour emerges, the  $\varepsilon$ -entropy can keep track of it and still manifests the small-scale irregular fluctuations.

We conclude that numerical models of recurrent neuronal networks, with conductance-based integrate-and-fire neurons, can be assimilated to high-dimensional chaotic systems, and are in many ways similar to the EEG. Moreover, even though they exhibit limit-cycle regimes for several parameter sets, this collective dynamics is built on top of irregular and high-dimensional neuronal activity which is only apparent at small-scales. Interestingly, this scheme reminds fluid dynamics, where a seemingly random microscopic dynamics may also coexist with more coherent behavior at large scales. This is the case for example close to the transition to turbulence, where fluids can show evidence for low-dimensional chaos [Brandstater et al., 1983]. More devel-

oped turbulence, however, does not show such evidence, presumably because a large number of degrees of freedom have been excited and the high-dimensional dynamics is present at all scales. It is possible that similar considerations apply to brain dynamics, especially when considering the recent debate about the nature of the ongoing activity of visual primary cortex (V1). Recordings made using voltage-sensitive dyes imaging (VSDI) in this cortical area have suggested that spontaneous activity consists of a seemingly-random replay of sensory-evoked orientation maps of activity in anesthetized [Kenet et al., 2003], but not in awake animals [Omer et al., 2008]. These observations have raised the question of whether the cortical activity of V1 could be described by a single-state high-dimensional attractor or a low-dimensional multistable attractor [Goldberg et al., 2004]. It is likely that at the VSDI scale, both scenarios are possible at the same time but at different scales, which would be consistent with the present results.

---

## Acknowledgments

Research supported by the CNRS, ANR and the European Community (FACETS grant FP6 15879). More details are available at <http://cns.iaf.cnrs-gif.fr>

---

## References

- Albers, D.J., Sprott, J.C. & Crutchfield, J.P. [2006] “Persistent chaos in high dimensions,” *Phys. Rev. E* **74**:057201
- Aurell, E., Boffetta, G., Crisanti, A., Paladin, G. & Vulpiani, A. [1997] “Predictability in the large : an extension of the concept of Lyapunov exponent,” *J. Phys. A* **30**(1), 1-26
- Babloyantz, A. & Destexhe, A. [1986] “Low dimensional chaos in an instance of epileptic seizure”, *Proc. Natl. Acad. Sc. USA* **83**, 3513-3517.
- Babloyantz, A., Nicolis, C. & Salazar, M. [1985] “Evidence for chaotic dynamics of brain activity during the sleep cycle”, *Phys. Lett. A* **111**, 152-156.
- Bédard, C., Kröger, H. & Destexhe, A. [2006] “Does the 1/f frequency scaling of brain signals reflect self-organized critical states?”, *Phys. Rev. Lett.* **97**(11) : 118102
- Beggs, J. & Plenz, D. [2003] Neuronal avalanches in neocortical circuits. *J. Neurosci.* **23**, 11167-11177.
- Bertschinger, N. & Natschläger, T. [2004] “Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks,” *Neural Comp.* **16**, 1413-1436
- Brandstater, A., Swift, J., Swinney, H.L., Wolf, A., Farmer, J.D., Jen, E. & Crutchfield, J.P. [1983] “Low-dimensional chaos in a hydrodynamic system”. *Phys. Rev. Lett.* **51**: 1442-1445.
- Cencini, M., Falcioni, M., Vergni, D. & Vulpiani A. [1999] “Macroscopic chaos in globally coupled maps,” *Physica D* **130**, 58-72
- Cencini, M., Falcioni, M., Olbrich, E., Kantz, H. & Vulpiani A. [2000] “Chaos or noise: Difficulties of a distinction,” *Physica D* **130**, 58-72
- Cessac, B. [2008] “A discrete time neural network model with spiking neurons; Rigorous results on the spontaneous dynamics,” *J. Math. Biol.* **56**, 311-345
- Cessac, B. & Viéville, T. [2008] “On Dynamics of Integrate-and-Fire Neural Networks with Conductance Based Synapses,” *Front. Comput. Neurosci.* **2**(2)
- Crutchfield, J.P. & Kaneko, K. [1988] “Are attractors relevant to turbulence?,” *Phys. Rev. Lett.* **60**(26), 2715-2718



- Destexhe, A. [1992] *Nonlinear Dynamics of the Rhythmical Activity of the Brain* (in French), Doctoral Dissertation (Université Libre de Bruxelles, Brussels, 1992). <http://cns.iaf.cnrs-gif.fr/alain.thesis.html>
- Destexhe, A., Contreras, D. & Steriade, M. [1999] "Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states", *J. Neurosci.* **19**, 4595-4608.
- Destexhe, A., Sepulchre, J.A. & Babloyantz, A. [1988] "A comparative study of the experimental quantification of deterministic Chaos". *Phys. Lett. A* **132**: 101-106.
- Dhamala, M., Lai, Y.-C. & Holt, R.D. [2001] "How often are chaotic transients in spatially extended ecological systems?," *Phys. Lett. A* **280**, 297-302
- Dhamala, M. & Lai, Y.-C. [2002] "The natural measure of nonattracting chaotic sets and its representation by unstable periodic orbits," *Int. J. Bif. Chaos* **12**(12), 2991-3005
- Eckmann, J.P., Kamphorst, S.O. & Ruelle, D. [1987] "Recurrence Plots of Dynamical Systems," *Europhys. Lett* **5**, 973977.
- Elbert, T., Ray, W.J., Kowalik, Z.J., Skinner, J.E., Graf K.E. & Birbaumer, N. [1994] "Chaos and physiology: deterministic chaos in excitable cell assemblies", *Physiol. Rev.* **74**: 1-47.
- El Boustani, S., Pospischil, M., Rudolph-Lilith, M. & Destexhe, A. [2007] "Activated cortical states: experiments, analyses and models", *J. Physiol. Paris* **101**: 99-109.
- El Boustani, S. & Destexhe, A. [2009] "A master equation formalism for macroscopic modeling of asynchronous irregular activity states," *Neural Comp.* **21**: 46-100
- Frank, G.W., Lookman, T., Nerenberg, M.A.H., Essex, C. & Lemieux, J. [1990] "Chaotic time series analysis of epileptic seizures," *Physica* **46D**: 427-438.
- Fraser, A.M. & Swinney H.L. [1986] "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A* **33**, 1134-1140
- Gaspard, P. & Wang, X.J. [1993] "Noise, chaos, and  $(\tau, \epsilon)$ -entropy per unit time," *Phys. Rep.* **235**(6), 321-373.
- Goldberg, J.A., Rokni, U., & Sompolinsky, H. [2004] "Patterns of ongoing activity and the functional architecture of the primary visual cortex," *Neuron* **42**, 489-500
- Grassberger, P. & Procaccia, I. [1983] "Measuring the strangeness of strange attractors," *Physica D* **9**, 189208
- Gao, J.B., Hu, J., Tung, W.W. & Cao Y.H. [2006] "Distinguishing chaos from noise by scale-dependent Lyapunov exponent," *Phys. Rev. E* **74**:066204
- Hegger, R., Kantz, H. & Schreiber, T. [1999] "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos* **9**: 413-435
- Jensen, H.J. [1998] *Self-Organized Criticality. Emergent Complex Behavior in Physical and Biological Systems*. (Cambridge University Press, Cambridge UK).
- Kantz, H. & Shreiber, T. [2004] "Nonlinear time series analysis," Cambridge University Press
- Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A. & Arieli, A. "Spontaneously emerging cortical representations of visual attributes," *Nature* **425**, 954-956
- Korn, H. & Faure, P. [2003] "Is there chaos in the brain? II. Experimental evidence and related models", *C.R. Biol.* **326**, 787-840.
- Kumar, A., Schrader, S., Aertsen, A. & Rotter, S. [2008] "The High-Conductance State of Cortical Networks", *Neural Comput.* **20**, 1-43
- Legenstein, R. & Maass, W. [2007] "Edge of chaos and prediction of computational performance for neural circuit models," *Neural Net.* **20**, 323-334
- Marwan, N., Romano, M.C., Thiel, M. & Kurths, J. [2006] "Recurrence plots for the analysis of complex systems" *Physics Rep.* **438**, 237:329
- Mayer-Kress, G., Yates, F.E., Benton, L., Keidel, M., Tirsh, W., Poppl, S.J. & Geist, K. [1988] "Dimension analysis of nonlinear oscillations in brain, heart and muscle". *Math. Biosci.* **90**: 155-182.
- Sauer, T., Yorke, J.A. & Casdagli, M. [1991] "Embedology," *J. Stat. Phys.* **65**, 579-616
- Shibata, T. & Kaneko, K. [1998] "Collective chaos," *Phys. Rev. Lett.* **81**(9):4116
- van Vreeswijk, C.A. & Sompolinsky, H. [1996] "Chaos in neuronal networks with balanced excitatory and inhibitory activity," *Science* **274**, 1724-1726
- van Vreeswijk, C.A. & Sompolinsky, H. [1998] "Chaotic balanced state in a model of cortical circuits," *Neural Comp.* **10**,1321-1372
- Sompolinsky, H., Crisanti, A. & Sommers, H.J [1988] "Chaos in random neural networks," *Phys. Rev. Lett.* **61**, 259-262
- Sprott, J.C. [2008] "Chaotic dynamics on large networks," *Chaos* **18**:023135
- Takens, F. [1981] "Detecting strange attractors in turbulence," *Lecture notes in mathematics*, Vol.898. Dynamical systems and turbulence, Springer, Berlin
- Theiler, J. [1990] "Estimating fractal dimension," *J. Opt. Soc. Amer. A* **7**, 1055
- Tél, T. & Lai, Y.-C. [2008] "Chaotic transients in spatially extended systems," *Physics Rep.* **460**, 245-275
- Vogels, T.P. & Abbott, L.F. [2005] "Signal propagation and logic gating in networks of integrate-and-fire neurons", *J. Neurosci.* **25**, 10786-10795.
- Wolf, A., Swift, J.B., Swinney, H.L. & Vastano, J.A. [1985] "Determining Lyapunov exponents from a time series", *Physica D* **16**, 285-317.
- Zillmer R, Livi R, Politi A & Torcini A. [2006] "Desynchronization in diluted neural networks", *Phys. Rev. E.* **74**: 036203.