PHYSICS, COMPUTATION, AND THE MIND — ADVANCES AND CHALLENGES AT INTERFACES

Proceedings of the 12th Granada Seminar on Computational and Statistical Physics

La Herradura, Spain 17 – 21 September 2012

EDITORS

Pedro L. Garrido
Joaquín Marro
Joaquín J. Torres
J. M. Cortés

AIP
American Institute of Physics

AIP | CONFERENCE PROCEEDINGS ■ 1510

Garrido
Marro
Torres
Cortés

PHYSICS, COMPUTATION, AND THE MIND — ADVANCES AND CHALLENGES AT INTERFACES

1510

AIP

ISBN 978-0-7354-1128-9
ISSN 0094-243X

PHYSICS, COMPUTATION, AND THE MIND — ADVANCES AND CHALLENGES AT INTERFACES

To learn more about the AIP Conference Proceedings Series,
please visit **http://proceedings.aip.org**

PHYSICS, COMPUTATION, AND THE MIND — ADVANCES AND CHALLENGES AT INTERFACES

Proceedings of the 12th Granada Seminar on Computational and Statistical Physics

La Herradura, Spain 17 – 21 September 2012

EDITORS

Pedro L. Garrido
Joaquín Marro
Joaquín J. Torres
Universidad de Granada, Granada, Spain

J. M. Cortés
Biocruces Health Research Institute, Barakaldo, Spain

All papers have been peer reviewed.

SPONSORING ORGANIZATIONS
Institute Carlos I for Theoretical and Computational Physics
European Network for Scientific Computation CECAM
The European Physical Society
University of Granada

AIP
American Institute of Physics

Melville, New York, 2013
AIP | CONFERENCE PROCEEDINGS ■ 1510

Editors

Pedro L. Garrido
Joaquín Marro
Joaquín J. Torres

Instituto Carlos I de Física Teórica y Computacional
Facultad de Ciencias
Universidad de Granada
Granada 18071
Spain

Email: garrido@ugr.es
jmarro@ugr.es
jtorres@onsager.ugr.es

J. M. Cortés
Computational Neuroimaging Group
Biocruces Health Research Institute
Hospital Universitario de Cruces
Plaza de Cruces, s/n
E-48903 Barakaldo
Spain

E-mail: Jesus.m.cortes@gmail.com

ISBN 978-0-7354-1128-9
ISSN 0094-243X
Printed in the United States of America

AIP Conference Proceedings, **Volume 1510**
Physics, Computation, and the Mind—
Advances and Challenges at Interfaces

Table of Contents

Preface: Physics, Computation, and the Mind — Advances and Challenges at Interfaces

This volume originated at the 12th Granada Seminar, and contains the main lectures and a selection of contributed papers in that conference. This is the twelfth of a series of Granada Lectures previously published by:

- World Scientific (Singapore 1993),
- Springer Verlag (*Lecture Notes in Physics* volumes 448 and 493),
- Elsevier (*Computer Physics Communications* volumes 121 and 122), and
- American Institute of Physics (*Conference Proceedings Series*, volumes 574, 661, 779, 887, 1091 and 1332).

These books and the successive editions of the Seminar since 1990 are described in detail at http://ergodic.ugr.es/cp/. This *web* also contains updated information on the next edition.

The Granada Seminar is defined as a small topical, interdisciplinary conference whose pedagogical effort is especially aimed at young researchers. In fact, one interesting aspect of this meeting is the opportunity given to the youngest to present their results and to discuss their problems with leading specialists. There were in this edition a total of 57 lectures and 39 poster contributions. One hundred twenty participants came from 23 countries: Spain contributed with 35%, the rest of Europe including Iran, Israel, Russia and Turkey with 39%, and America with 21% (the rest came from Japan and Australia); half the participants received some, either total or partial support from the organization.

The 12th Granada Seminar was organized and mostly financed by the Institute *Carlos I* for Theoretical and Computational Physics, sponsored by the European Network for Scientific Computation CECAM, the European Physical Society and the University of Granada, and endorsed by the Spanish Physics Society and the American Physical Society.

We also wish to express gratitude to all those who have collaborated in making this event a success. In particular, we mention the remarkably high quality and friendly cooperation of invited speakers and other participants, whose personal effort enabled us to accomplish the goals of the Seminar, the Steering Committee's help in designing format and contents, and further *in situ* priceless collaboration from Joaquín J. Torres, J. Cortés and other colleagues, and from the Ph.D. students and postdocs in our group. This edition of the Seminar was held from 17 to 21 of September 2012 in the charming village of La Herradura, a remarkable spot of the *Tropical Coast* of Granada, Spain, where the participants enjoyed a paradisiacal setting.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 1-2 (2013); doi: 10.1063/1.4776494

Following the welcome speech, it is to be highlighted that this edition of the Granada Seminar is a straight consequence of the evolution and popularization of technologies such as encephalography, magnetic resonances, and positron emission tomography, which produce detailed static and dynamic pictures of the brain and its processes, and tiny electrodes and probes which detect and accurately measure the electric pulses generated in areas that are well localized and so small that include ideally only one or at most a few neurons and synapses today. While developing to so highly sophisticated standards, these technologies are becoming very widely used, so much so that are now familiar tools not only in specialized research laboratories but also in many hospitals. Therefore, myriads of valuable data concerning the structure and function of the nervous system are constantly becoming easily available to many scientists.

The old challenge of trying to understand what consciousness is, how intelligence can emerge from relatively unintelligent neurons, and why there are different levels of intelligence can therefore be accepted now. The data from laboratories and hospitals are certainly feeding scientists in neurobiology but also in the apparently distant mathematics and physics; particularly statistical physics which nowadays masters the modeling of complex systems is eager to handle good data on the brain, which is a paradigm of complexity.

But this is not the only justification for looking at the interfaces between disciplines. This meeting was also motivated by the fact that there is a slow though observable convergence between biology systems and digital ones which, in addition to intriguing, is a field with infinite applications that should prove be most relevant to humanity.

Finally, let me notice that an effort has been made by authors and editors to offer pedagogical notes here; in particular, each topic is comprehensively described within its scientific context. We try to mold the *Granada Lectures* into a series of books that help introduce the beginner to novel advances in statistical physics and to the creative use of computers in scientific research, as well as to serve as a work of reference for teachers, students and researchers.

Pedro L. Garrido and Joaquín Marro Granada, 21 November 2012

Granada Seminar Steering Committee

- Kurt Binder, Institute of Physics of Condensed Matter, University of Mainz
- Rafael de la Llave, Department of Mathematics, University of Texas, Austin
- Julio F. Fernández, Aragon Institute for Materials Science, C.S.I.C.
- Pedro L. Garrido, Institute "Carlos I" for Theoretical and Computational Physics, University of Granada
- Hans J. Herrmann, ETH, Zürich
- Hilbert J. Kappen, Department of Biophysics, University of Nijmegen
- Joel L. Lebowitz, Department of Mathematics, Rutgers University, New Jersey
- Michel Mareschal, ZCAM, Saragossa, and Université Libre de Bruxelles
- Joaquín Marro, Departamento de Electromagnetismo y Física de la Materia, University of Granada

Brain complexity born out of criticality

E. Tagliazucchi* and D. R. Chialvo†

*Neurology Department and Brain Imaging Center, Goethe University, Frankfurt am Main, Germany.
†Consejo Nacional de Investigaciones Científicas y Tecnológicas, Buenos Aires, Argentina.

Abstract. In this essay we elaborate on recent evidence demonstrating the presence of a second order phase transition in human brain dynamics and discuss its consequences for theoretical approaches to brain function. We review early evidence of criticality in brain dynamics at different spatial and temporal scales, and we stress how it was necessary to unify concepts and analysis techniques across scales to introduce the adequate order and control parameters which define the transition. A discussion on the relation between structural vs. dynamical complexity exposes future steps to understand the dynamics of the connectome (structure) from which emerges the cognitome (function).

Keywords: brain; phase transition; connectome.

WHY CRITICALITY?

Complexity, in simple terms, is all about how diversity and non-uniformity [1] arises from the uniform interaction of similar units. In all cases, the dynamics of the emergent complex behavior of the whole cannot be directly anticipated from the knowledge of the laws of motion of the isolated parts. Early forerunners of complexity science, namely statistical mechanics and condensed matter physics, have identified a peculiar scenario at which, under certain general conditions, such complexity can emerge: near the critical point of a second order phase transition. At this point, complexity appears as a product of the competition between ordering and disordering collective tendencies, such that the final result is a state with a wide variety of dynamic patterns exhibiting a mixture of order and disorder.

As argued elsewhere [2, 3] the dynamics of the human brain exhibit a large degree of concordance with those expected for a system near criticality. From the cognitive side, brain's complexity is an almost obvious statement: the ultimate products of such complexity are, for instance, the nearly unpredictable human behavior and the underlying subjective experience of consciousness, with its bewildering repertoire of possible contents. However, the proposal that the same mechanisms underlying physical complexity also underlie the biological complexity of the brain is surprisingly recent. The description of the dynamical rules governing neurons at the microscopic level [4] and the first mathematical demonstration of a second order phase transition in a many body system [5] were almost contemporaries. However, these developments were separated by a large temporal gap from the first proposals of criticality in brain dynamics [2, 6]. Being a relatively recent proposal, the consequences of such hypothesis are still far from clear. In the present essay we first discuss recent empirical evidence favoring this hypothesis, focusing on the presence of a second order phase transition in large scale brain dynamics

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 4-13 (2013); doi: 10.1063/1.4776495

[7], and then explore some possible consequences.

SCALE INVARIANCE IN BRAIN DYNAMICS: EARLY FINDINGS

The early evidences supporting criticality as a plausible dynamical regime for brain activity can be roughly classified according to the spatial and temporal resolutions at which they were obtained from the experimental techniques used for the recording of brain activity. The results exposed in this section are important to introduce the recent finding of a second order transition in large scale brain dynamics, since the strategy used for its uncovering can be regarded as a transfer of analysis techniques used at microscopic scales to the macroscopic large scale domain.

A landmark of the critical regime which occurs during a second order phase transition is the divergence of correlation length. As order emerges, the constituents of the system must organize themselves instantaneously. Also at this point, any external perturbation will also have the highest impact on the system, as measured by the susceptibility. For that to occur, the dynamics of individual units must be mutually influenced even for those which are macroscopically separated and not directly connected. The divergence of correlation length implies scale invariance (i.e., fractality) in the system, as the presence of a characteristic scale would violate the divergence required by the instantaneous onset of the ordered phase. In fMRI experiments, it has been demonstrated that functional connectivity networks are scale invariant [8] and, most remarkably, are virtually indistinguishable from correlation networks obtained in the Ising model at the onset of the second order phase transition [9]. Also, this scale invariance has been directly demonstrated for fMRI data [10], as well as the divergence in the correlation length [11].

We emphasize that this evidence of scale free brain dynamics at the large scale domain has insofar treated brain activity as a continuously variable. Whether this was the case or its continuous nature emerged as an artifact due to experimental and physiological blurring remained unknown. On the other hand, experiments at smaller (microscopic scales) have concentrated in the description of brain activity as discrete avalanching events, spreading throughout the cortex in a scale free fashion [12]. These scale free avalanches have been exposed using electrophysiological techniques in different settings [7, 13, 14]. Self organized critical systems are known to dissipate energy in form of power law distributed avalanches [15, 2] , hence, this is direct evidence favoring the hypothesis that brain achieves critical properties through self-regulation and does not require fine tuning of parameters. We re-emphasize that experiments at this scale, either single unit or Local Field Potential (LFP) recordings, demonstrate scale-free intermittence and bursting, with the intensity of the discrete burst obeying a power-law distribution.

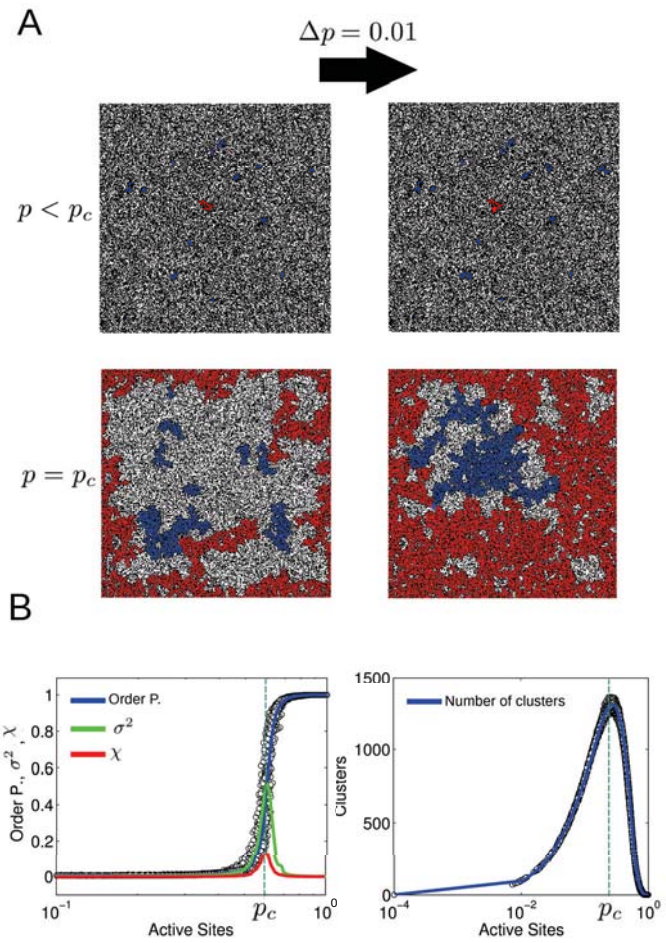


FIGURE 1. Order, disorder and susceptibilty in a second order percolation phase transition. A. The state of the system (black dots represent unnocuppied sites, white represents occupied sites) at sub-critical and critical concentrations. The effect of a small concentration change on the largest cluster (order parameter, red cluster) and other smaller clusters (blue) is shown, demonstrating how a small perturbation can induce large changes near the critical point (but not during the sub-critical phase) B. Order parameter, its variability (σ^2) and susceptibility (χ) as a function of the control parameter (left) and percolation diagram of active sites vs. number of clusters (right). In this case, the control parameter is defined as the ratio of occupied to the total number of sites (concentration of active sites).

A PHASE TRANSITION IN CORTICAL DYNAMIC AND ITS CONSEQUENCES

As compelling as the above experimental findings of scale invariance may be, they fail to reveal clues about its origin. Eventually, what is needed is to identify from the data the fundamental elements of a second order phase transition, namely the dynamical changes in an order parameter as a function of some control parameter. The derivation

of those is not required solely for theoretical reasons but is also of important practical relevance: a control parameter allows to quantify the "degree of criticality" present in the system. As in any finite non-equilibrium system, brain activity fluctuations impose spontaneous changes in the dynamical regime. An order and control parameter would allow, for instance, to dissipate the idea that the dynamical regime of the brain is fixed and therefore to study the impact of criticality in the spontaneously fluctuating aspects of behavior and cognition.

A leap forward in this direction provided for the first time a way to estimate these parameters [7]. The analysis was carried out at the level of whole-brain human fMRI data, an imaging modality known for its excellent spatial resolution. The control parameter defined here is roughly equivalent to the instantaneous global activity level. The order parameter is equivalent to the size of the largest cluster of activated cortical activity (normalized by the global activity level). With these definitions, a scatter plot of order vs. control parameters showed a sharp transition around a critical level of global activity, yielding a diagram which resembles those derived for other systems undergoing a second order phase transition.

Before exposing the rationale behind this choice of order and control parameters, we will briefly review the key insights which allowed their identification. As anticipated in the previous section, these are related to the fusion of analysis techniques used at different scales, allowing a unified interpretation of the dynamical properties of the brain.

The most important insight of this work was its departure from techniques which estimate time averages of activity (during extended periods of time) and to focus on charactering spatiotemporal instantaneous activity. Since the early days of fMRI, the Blood Oxygen Level Dependent (BOLD) signal is treated as a continuous smooth signal, even in spite of a large body of evidence showing that neural activity (at all scales) happens in bursts. Following this natural line of thought, we discarded low excursions of the hemodynamic signals and focused on large amplitude events [16]. This procedure is analogous to the LFP thresholding used at smaller scales to uncover avalanching activity [12, 17]. In formal terms, this is achieved by the reduction of the signal to a point process, which in turn is constructed by the introduction of a Poincaré section of the BOLD signal, as usually done in dynamical systems [18, 19, 20, 21, 22, 23]. Notably, the resulting point-process remarkably resembled the results of resting-state BOLD signal deconvolution, giving formal support for its introduction.

The point process allows, for the first time, the visualization *at each time step* of the brain spatial pattern of activation. Thus, this approach allows us to perform a true spatiotemporal analysis of fMRI data. This information is then used to identify both parameters: the control parameter, which is defined as the sum of all voxels above a threshold (i.e. those active in the point process) and the order parameter as the size of the largest cortical cluster. The computation of these parameters from the fMRI data is straightforward applying a cluster labeling algorithm to the spatiotemporal point process.

These definitions for order and control parameters are of clear interpretability in terms of degrees of order/disorder. Consider a moment in time in which the brain has a low value of the order parameter, i.e. the largest cluster of cortical activation is small compared to the total activated gray matter. It is possible to accommodate a vast number

of such clusters on the cortical surface, hence, the entropy is high, as corresponds to a disordered state. Following the increase of the control parameter a point is reached in which a giant cluster of cortical activity emerges, comprising and integrating many different functional systems. Since all activations have coalesced into this single large cluster, there is little room for variation: entropy is lower, as corresponds to an ordered state. In the extreme, a single cluster spans all active voxel, giving only one possible state (a globally activated state with zero entropy).

It is useful to view these parameters in analogy to those used to quantify the dynamical regime of the percolation model, in which occupied (active) sites are placed on a lattice with different concentrations (defined as the ratio of occupied to total sites) and the size of the largest cluster is taken as a measure of order. Figure 1 shows an illustration of the model and the evolution of different quantities due to changes in the control parameter, which in this case is the concentration of sites. Despite the fact that the percolation model has no dynamics, it serves an an useful example to introduce the ideas of order parameter, control parameter and the maximum variability and susceptibility of the critical state, as discussed below. In Fig. 2 we summarize the evidence of a second order phase transition

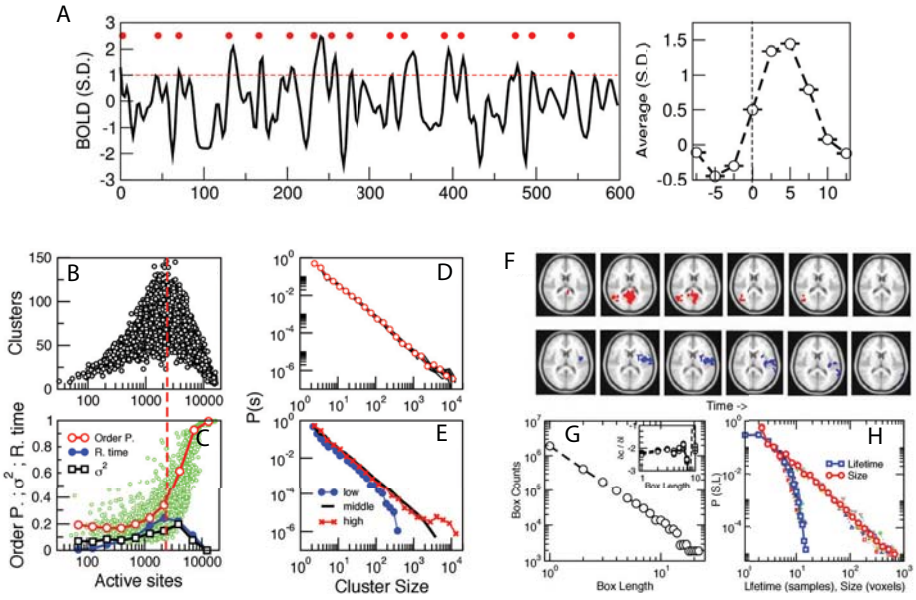


FIGURE 2. Identification of a second order phase transition in fMRI dynamics. A. Decomposition of an hemodynamic signal into a point process (left) which generates an stereotypical waveform similar to the hemodynamic response function, the fMRI signal response to a single and discrete electrophysiological activation (right). B. Percolation diagram of active sites vs. number of clusters. C. Order parameter, variability and residence times as a function of the control parameter (active sites). D. Distribution of cluster sizes. E. Distribution of cluster sizes at the three different dynamical regimes. F. Examples of avalanches triggered from the visual (up) and insular (bottom) cortex. G. Fractal dimension of the active sites obtained using a box counting algorithm. H. Distribution of avalanche sizes and avalanche lifetimes. Figure redrawn from [7].

found recently [7]. We introduce the point process decomposition of the hemodynamic signal, we show the evolution of the order parameter vs. the control parameter, as well as the variability and the residence times (the definition is introduced below), as was

similarly presented in Fig. 1 for the percolation model. This figure also presents results on the dynamical evolution of clusters, which spread as avalanches, a result consistent with previous findings at smaller scales [12].

Clusters (or "blobs", as they are usually refereed to in the neuroimaging community) represent co-activations (as evidenced by the point-process) of contiguous brain regions which are usually associated with sensory, motor, or higher order cognitive systems. For instance, statistics over time averages in a visual stimulation experiment would reveal a blob in the visual cortex (and similarly in other areas for other experiments). Seen this way, clusters represent a section or (level set) of the Statistical Parametric Map (SPM) which encodes the spatial distribution of the model fit statistical significance. Such model fit is performed during extended periods of time. However, the point process not only reveals that this activations appear spontaneously in the resting state but also shows their relationship with the dynamical regime: order means a shared co-activation (integration) of the processes underlying these clusters, disorder means an increased segregation. In between, the transition point arguably represents an optimum balance of segregation/integration. It also corresponds to the point in which the brain displays the higher variability in its repertoire of states, as evidenced by a peak in the variability of the order parameter vs. the control parameter (Fig. 2B, Fig. 2C).

As noted above, the brain fluctuates in and out of this regime. However, we demonstrated that most of the time it stays at the transition point, as evidenced by the computation of residence times (Fig. 2C. Residence times quantify how much time the system spends at each possible state of the space of dynamical variables. In this case, we have tracked the state of the system using the order and control parameters and we demonstrated that the system stays for more prolonged periods at the state corresponding to the critical control parameter. Incidentally, this may explain why direct evidence for the presence of a second order transition had to wait so long: since on average the brain spends most of the time in the transition point, any approach which is based on time averaged quantities is bound to highlight only the critical state, but not the super- (disordered) and sub- (ordered) critical states, as can be done with the point process approach.

Quantification of the cluster spatio-temporal evolution demonstrates scale-free avalanches spreading in fractal-like structures (with dimension slightly greater than two, see Fig. 2F, Fig. 2G, Fig. 2H). This result is a confirmation of what was previously known for brain dynamics at smaller scalers by means of electrophysiological experiments [12]. However, it highlights the fact that predictions based on properties of the critical state are manifest in all spatial and temporal scales of brain dynamics. Scale invariance dictates that avalanches of activity must follow the same distribution regardless of their size and that events as large as the size of the system must be found. Both predictions are clearly confirmed by analysis of fMRI dynamics in the computation of power law exponents. Importantly, the theory of phase transitions provides a set of critical exponents which may be computed for this data and shed light on a possible universality class for brain dynamics.

Finally, a peak in variability is related to a peak in the susceptibility of the system near the critical point of a second order phase transition. To introduce the concept of susceptibility, imagine a physical system upon which we exert forces. We expect such forces to elicit changes in the system. Susceptibility can be roughly defined as the ratio of the elicited response and the exerted influence. In finite critical systems,

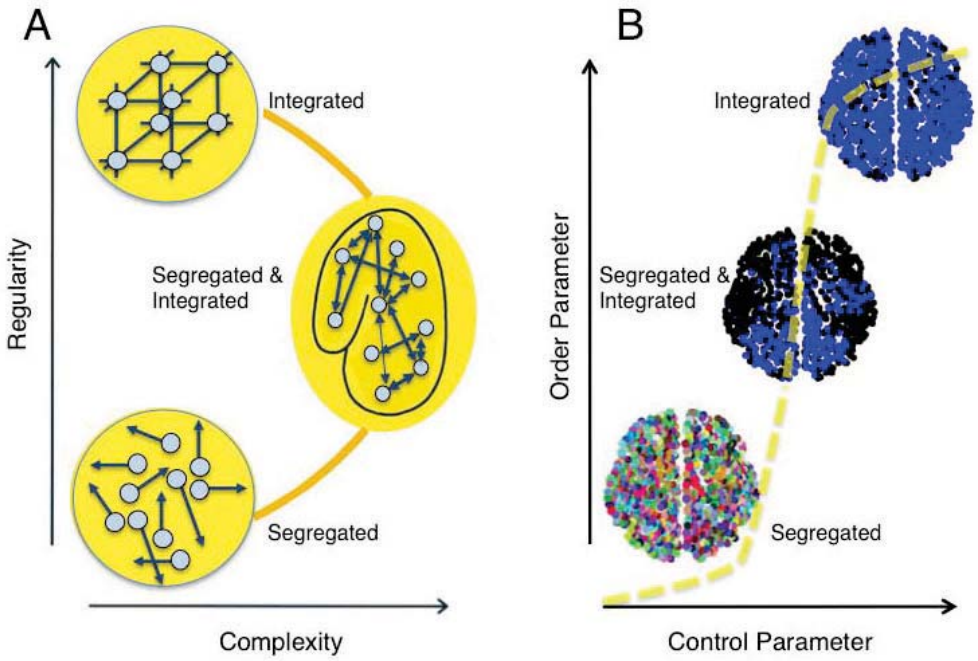


FIGURE 3. Two views of the integration-segregation dilemma. Panel A shows a *structural* point of view, in which neural structures exhibit different regularity and complexity. According to this, the structural connectivity of the human brain has a high complexity and intermediate regularity, with a crucial balance of integration and segregation acquired trough evolutionary selection [27]. Panel B depicts a *dynamical* alternative. According with the proposal discussed here, the three type of regimes in A (with higher or lower complexity) can be dynamically (and transiently) generated by any system undergoing a second order phase transition (even with trivial and homogeneous structural connectivity), because they represent the three generic dynamical regimes of the system. Colors in the three graphs label the clusters with coherent activity, thus in the most ordered regime (top right) the entire brain is active and coalesces into a single activated cortical cluster, whereas in the most disordered one (bottom left), each brain region acts independently. It is only at criticality (middle) that coherent clusters of all sizes are possible, thus optimizing the integration/segregation balance. Panel A redrawn from the figure in Box 2 of [27] rotated 90 degrees counterclockwise. Panel B redrawn from [28].

susceptibility is maximized at the critical point, therefore, the capacity of the system to react to external changes is maximized. Given the need of flexibility and reactiveness of the brain, the endowment of maximum susceptibility due to the critical state is a very attractive possibility from an evolutionary point of view. One must contrast this non-equilibrium view with that of many models (such as the attractor networks of the Hopfield model [29]) in which, after the system reaches its final state, it has a vanishing susceptibility (since small changes which could elicit large responses at the critical state are not capable of shifting the system from the equilibrium point or attractor). The relationship between equilibrium and non-equilibrium models and the inclusion of noise in dynamical equations will be discussed in the next session.

ORDER/DISORDER VIS A VIS INTEGRATION/SEGREGATION

As discussed already elsewhere [3] the idea of a continuous phase transition in brain dynamics is related to the ability of the brain to simultaneously integrate and segregate information, a point championed by Edelman, Tononi and Sporns [24, 25, 26, 27]. In their own words: *"Nervous systems facing complex environments have to balance two seemingly opposing requirements. First, there is a need quickly and reliably to extract important features from sensory inputs. This is accomplished by functionally segregated (specialized) sets of neurons, e.g. those found in different cortical areas. Second, there is a need to generate coherent perceptual and cognitive states allowing an organism to respond to objects and events, which represent conjunctions of numerous individual features. This need is accomplished by functional integration of the activity of specialized neurons through their dynamic interactions"* [24].

The cartoons in Fig. 3 show schematically that low complexity is expected for interactions occurring at both extreme degrees of order and disorder. It is only at the intermediate level that complexity peaks, when diverse mixtures of order and disorder are present. Panel A (taken from [27]) illustrates the point for the structural case, in which the brain structure molded trough evolution is able to optimize both integration and segregation. The case in panel B represents the dynamical scenario (taken from [28]) of a second order phase transition. In this case, as the control parameter is increased, the order parameter also increases, slowly first and then suddenly at the transition point. The overimposed pictures represent the top view of the brain from simulations of its activity at different regimes (unpublished data from [28]). The three examples are the type of clusters (as defined already for the results in Fig. 2) found at disordered, intermediate and ordered regimes. It is only at the critical regime that clusters of all sizes can be observed, which in dynamical terms represent the balance between high integration (a few big clusters) and high segregation (many small clusters). Indeed this mixture is characterized by a scale invariant distribution of cluster sizes (see Fig. 2D for the distribution found experimentally).

Before moving away from Fig. 3, it need to be noted that, although unlikely to be a realistic possibility, here and only in terms of considering mechanisms, the complexity exhibited by the networks depicted in Panel A could also arise from a phase transition. An example can be bond percolation, for which is known that the most complex network structures arise at the percolation threshold.

It is important to note two additional points about dynamics and structural complexity. First of all, we do not stress the segregation/integration balance by itself: it is considered simply a consequence of the physics of phase transitions, which are, arguably, of a more fundamental nature. Still, the relation between these concepts clearly deserves further theoretical and experimental investigations.

The second point to remark is a bit subtler: criticality endows an arbitrary graph with similar dynamic complexity (i.e., similar integration/segregation balance) than the structural complexity of the brain. However, this is not meant to imply that any arbitrary structure, provided with critical dynamics, can "think" like a brain. An understanding of the relation between the connectome (an exhaustive description of all *structural* connections) and what we call cognitome (an equally exhaustive description of the *functional* repertoire of the brain) and the role of dynamics is still ahead and much work

still needs to be done.

A lesson taught by the findings on critical phenomena in brain dynamics is that the explanative power of physical theories should not be disregarded in biology just because it provides all encompassing, holistic explanations as opposed to detailed, reductionistic descriptions of myriads of experimental facts. A glaring example is the question of what role noise plays in brain dynamics and where it is generated. Often theories postulate a "key role" for noise to adequately explain response variability in healthy brain function and proceed to explain the observations in the light of very detailed noisy processes. In the same direction, the information content of the brain BOLD signal's variability *per se* received increasing interest. For instance, it was shown recently [31] in a group of subjects of different age, that the BOLD signal variability (standard deviation) is a better predictor of the subject age than the average. Furthermore, additional work focused on the relation between the fMRI signal variability and task performance, and concluded that faster and more consistent performers exhibit significantly higher brain variability across tasks than the poorer performing subjects [32]. From the current perspective, all these observations have the same underlying explanation: maximum variability is a straightforward consequence of the critical regime as shown recently [11]. Noise is endogenous to non-equilibrium systems at the transition point and it needs not to be introduced as an ad-hoc equilibrium explanation every time a neurobiological fact displays a large degree of variability.

SUMMARY AND OUTLOOK

In this essay we have exposed recent experimental evidence demonstrating the presence of a second order phase transition in human brain dynamics, we have explored some of its consequences. In doing so, we have stressed that the application of these concepts should be aimed as a unifying physical explanation of the brain. We have also stressed that biological theories should take advantage of the all-encompassing framework provided by physical theories (in particular, that of critical phenomena) instead of relying on reductionistic ad-hoc explanations tailored for each experimental fact.

After establishing that resting state brain activity displays many signatures of sub, super and critical states, a question clearly needs to be addressed: what is the neurobiological role of these dynamical regimes? Since the brain continuously enters and leaves the critical regime (but remains most of the time at this point), one could be tempted to speculate about the possibility of a more permanent displacement. In other words, if criticality is important for healthy brain function, what happens if this property is lost? To gain insights on this question we propose to study brain states which radically differ from wakeful rest (by far the most studied condition in fMRI resting state analyses). Examples could be deep sleep, anesthesia, coma, as well as different states of consciousness induced by drug intake. We believe that new exciting venues of research will be open to clarify the role of these fluctuations around criticality in relation to these different neurobiological states.

ACKNOWLEDGMENTS

Work supported by NIH (USA), CONICET (Argentina) and the MCyT (Spain). ET was partially funded by the Bundesministerium fur Bildung und Forschung (grant 01 EV 0703) and LOEWE Neuronale Koordination Forschungsschwerpunkt Frankfurt (NeFF).

REFERENCES

1. P. Bak, and M. Paczuski, *Proc. Natl. Acad. Sci. USA* **92**, 6689–6696 (1995).
2. P. Bak, *How nature works: The science of self-organized criticality*, Copernicus Books, New York, 1996.
3. D. R. Chialvo, *Nature Phys.* **6**, 744–750 (2010).
4. A. L. Hodgkin, and A. F. Huxley, *J. Physiol.* **117**, 500–544 (1952).
5. L. Onsager, *Phys. Rev.* **65**, 117–149 (1944).
6. D. R. Chialvo, and P. Bak, *Neuroscience* **90**, 1137–1148 (1999).
7. E. Tagliazucchi, P. Balenzuela, D. Fraiman, and D. R. Chialvo, *Front. Physiol.* **3**, 15 (2012).
8. V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, *Phys. Rev. Lett.* **94**, 018102 (2005).
9. D. Fraiman, P. Balenzuela, J. Foss, and D. R. Chialvo, *Phys. Rev. E*, **79**, 061922 (2009).
10. P. Expert, R. Lambiotte, D. R. Chialvo, K. Christensen, H. J. Jensen, D. J. Sharp, and F. Turkheimer, *J. R. Soc. Interface* **8**, 472–479 (2011).
11. D. Fraiman, and D. R. Chialvo, *Front. Physiol.* **3**, 307 (2012).
12. J. M. Beggs, and D. Plenz, *J. Neurosci.* **23**, 11167–11177 (2003).
13. T. Petermann, T. C. Thiagarajan, M. A. Lebedev, M. A. L. Nicolelis, D. R. Chialvo DR, and D. Plenz, *Proc. Natl. Acad. Sci. USA* **106**, 15921–15926 (2009).
14. T. L. Ribeiro, M. Copelli, F. Caixeta, H. Belchior, D. R. Chialvo, M. A. L. Nicolelis, and S. Ribeiro, *PLoS ONE* **5** e14129 (2010).
15. P. Bak, C. Tang, and K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381–384 (1987).
16. E. Tagliazucchi, P. Balenzuela, D. Fraiman, P. Montoya, and D. R. Chialvo, *Neurosci. Lett.* **488**, 158–163 (2010).
17. T. C. Thiagarajan, M. A. Lebedev, M. A. Nicolelis, D. Plenz, *PLoS Biol.* **8**, e1000278 (2010).
18. P. Grassberger, and I. Procaccia, *Physica* **9D**, 189–208 (1983).
19. N. Packard, J. Crutchfield, J. D. Farmer, and R. Shaw, *Phys. Rev. Lett.* **45**, 712–716 (1980).
20. J. C. Roux, J. Rossi, S. Bachelart, and C. Vidal, *Phys. Lett. A* **77**, 391–393 (1980).
21. J. C. Roux, and H. Swinney, "Topology of chaos in a chemical reaction" in *Nonlinear phenomena in chemical dynamics,* C. Vidal, and A. Pacault Eds. Springer-Verlag, Berlin, 1981.
22. F. Takens, "Dynamical systems and turbulence, Warwick, 1980", D. A. Rand and L.-S. Young Eds., *Lecture Notes in Math* **898**, Springer-Verlag, Berlin, 1981.
23. R. Castro, and T. Sauer, *Phys. Rev. E* **55**, 287–290 (1997).
24. O. Sporns, G. Tononi, and G. M. Edelman, *Neural Netw.* **13**, 909–922 (2000).
25. G. Tononi, O. Sporns, and G. M. Edelman, *Proc Natl Acad Sci USA* **91**, 5033–5037 (1994).
26. G. Tononi, *BMC Neurosci.*. **5**, 42 (2004).
27. G. Tononi, G. M. Edelman, and O. Sporns, *Trends Cogn. Neurosci.*. **2**, 474–484 (1998).
28. A. Haimovici, E. Tagliazucchi, P. Balenzuela, D. R. Chialvo, "Brain organization into resting state networks emerges from the connectome at criticality". `arxiv.org/abs/1209.5353`.
29. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558 (1982).
30. E. T. Rolls, and G. Deco, *The noisy brain*. Oxford University Press, London, 2010.
31. D. Garret, N. Kovacevic, A. McIntosh, and C. Grady, *J. Neurosci.* **30**, 4914–4921 (2010).
32. D. Garrett, N. Kovacevic, A. R. McIntosh, and C. L. Grady, *J. Neurosci.*, **31**, 4496–4503 (2011).

Spontaneous neuronal activity as a self-organized critical phenomenon

L. de Arcangelis* and H. J. Herrmann†

*Department of Industrial and Information Engineering,
Second University of Naples, Aversa (CE), Italy.
†Institute Computational Physics for Engineering Materials, ETH, Zürich, CH.

Abstract. Neuronal avalanches are a novel mode of activity in neuronal networks, experimentally found in vitro and in vivo, and exhibit a robust critical behaviour. Avalanche activity can be modelled within the self-organized criticality framework, including threshold firing, refractory period and activity-dependent synaptic plasticity. The size and duration distributions confirm that the system acts in a critical state, whose scaling behaviour is very robust. Next, we discuss the temporal organization of neuronal avalanches. This is given by the alternation between states of high and low activity, named up and down states, leading to a balance between excitation and inhibition controlled by a single parameter. During these periods both the single neuron state and the network excitability level, keeping memory of past activity, are tuned by homeostatic mechanisms. Finally, we verify if a system with no characteristic response can ever learn in a controlled and reproducible way. Learning in the model occurs via plastic adaptation of synaptic strengths by a non-uniform negative feedback mechanism. Learning is a truly collective process and the learning dynamics exhibits universal features. Even complex rules can be learned provided that the plastic adaptation is sufficiently slow.

Keywords: spontaneous activity; neural networks; learning.
PACS: 05.65.+b, 05.45.Tp, 89.75.-k, 87.19.L-

INTRODUCTION

Spontaneous neuronal activity generally exhibits slow oscillations between high activity periods, or bursts, followed by substantially quiet periods. Bursts can last from a few to several hundreds of milliseconds and, if analysed at a finer temporal scale, show a complex structure in terms of neuronal avalanches. *In vitro* experiments allow to record avalanches [1, 2] from mature organotypic cultures of rat somatosensory cortex where they spontaneously emerge in superficial layers. The size and duration of neuronal avalanches follow power law distributions with very stable exponents, which is a typical feature of a system acting in a critical state, where large fluctuations are present and the response does not have a characteristic size. The same critical behaviour has been measured also *in vivo* from rat cortical layers during early post-natal development [3], from the cortex of awake adult rhesus monkeys [4], using microelectrode array recordings, as well as for dissociated neurons from rat hippocampus [5, 6] or leech ganglia [6]. Recent results have shown that the critical dynamics in the brain resting state is a necessary condition for many brain functions [7].

The quiet periods measured between bursts, also called down-states, can last up to several seconds. The emergence of these down-states can be attributed to a variety of mechanisms: a decrease in the neurotransmitter released by each synapse, either due to

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 14-24 (2013); doi: 10.1063/1.4776496

the exhaustion of available synaptic vesicles or to the increase of a factor inhibiting the release [8] such as the nucleoside adenosine [9], the blockage of receptor channels by the presence, for instance, of external magnesium [10], or else spike adaptation [11]. A down-state is therefore characterized by a *disfacilitation*, i.e. absence of synaptic activity, of a large number of neurons causing long-lasting returns to resting potentials [12] and its onset presents a high level of synchrony. Recently, it was shown analytically and numerically that self-organized critical behaviour characterizes up-states, whereas down-states are subcritical [13].

Whereas action potentials are rare during down-states, small amplitude depolarizing potentials, reminiscent of miniature potentials from spontaneous synaptic release, and some synaptic input occur at higher frequencies. The non-linear amplification of small amplitude signals contributes to the generation of larger depolarizing events bringing the system back into the up-state, as observed in cortical slabs [14], dissociated cultures [15] and slice cultures [16]. The analysis of the amount of time striatal spiny neurons [17, 18] and cortical pyramidal neurons [19] spend at each value of the membrane potential shows that both cell types toggle between two preferred membrane potentials [20]: A very negative one in the down state, and a more positive, depolarized one, in the up-state. The neuron up-state being just a few millivolts from the action potential threshold, suggests that during the up-state neurons respond faster and more selectively to synaptic inputs. For cortical neurons the up-state would be a metastable state, i.e. the membrane potential would soon decay down to the resting potential value, if network mechanisms would not sustain the activity. The up-state has therefore network, rather than cellular, properties.

The temporal organization of neuronal avalanches can be characterized by means of the waiting time distribution. Each avalanche is characterized by its size s_i, its starting and ending times, t_i^i and t_i^f. The properties of temporal occurrence are analysed by evaluating the distribution of waiting times $\Delta t_i = t_{i+1}^i - t_i^f$. This is a fundamental property of stochastic processes, widely investigated for natural phenomena [21] and able to discriminate between a simple Poisson and a correlated process. Indeed, in the first case the distribution is an exponential, whereas it exhibits a more complex behaviour with power law regime if correlations are present. For a wide variety of phenomena, earthquakes, solar flares, rock fracture, etc., this distribution always shows a monotonic behaviour. In a recent paper [22] this distribution has been analysed for freely behaving and anaesthetized rats. The distributions show consistently a decreasing behaviour. Universal scaling features are observed when waiting times are rescaled by the average occurrence rate for freely behaving rats, whereas curves for anaesthetized rats do not collapse onto a unique function.

Recently, the waiting time distribution has been measured for different cultures of rat cortex slices [23]. The curves exhibit a complex non-monotonic behaviour with common features: an initial power law regime and a local minimum followed by a more or less pronounced maximum (Fig. 1). More precisely, all curves show an initial power law regime between 10 and about 200ms, with an average exponent 2.15 ± 0.32. For $\Delta t > 200ms$ curves can become quite different with the common characteristics of a local minimum located at $200ms < \Delta t_{\min} < 1s$, followed by a more or less pronounced maximum at $\Delta t \simeq 1 - 2s$. This behaviour is not usually observed in natural phenom-

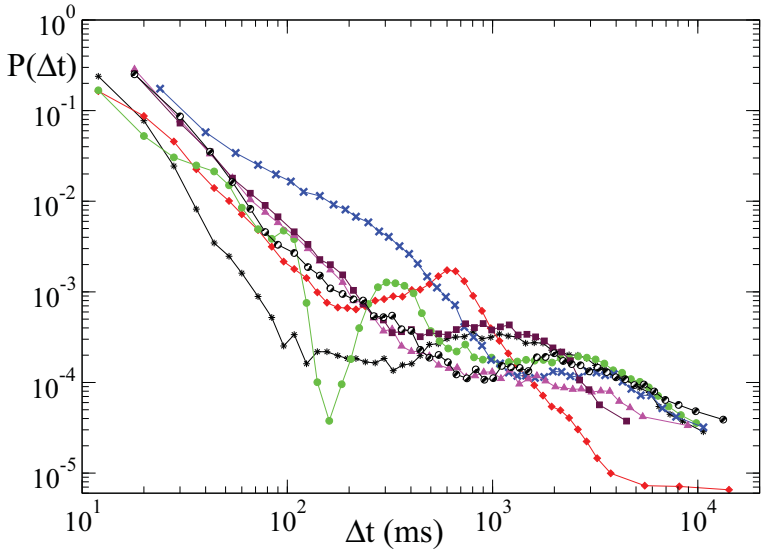


FIGURE 1. (Color online) The distribution of waiting times for seven different slices of rat cortex. Experiments were performed on coronal slices from rat dorsolateral cortex attached to a poly-D-lysine coated 60-microelectrode array and grown at 35.5 ^{o}C in normal atmosphere and standard culture medium without antibiotics for 4-6 weeks before recording. Avalanche activity was measured from cortex-striatum-substantia nigra triple cultures or single cortex cultures as reported previously[1].

ena and suggests that avalanche occurrence is not a pure Poisson process. In order to investigate the origin of this behaviour, we simulate avalanche activity by a neuronal network model [24, 25, 26], which is able to reproduce the scaling properties of neuronal avalanches.

NEURONAL MODEL

We here discuss a neuronal network model inspired in self-organized criticality ideas [27]. The model implements several physiological properties of real neurons: a continuous membrane potential, firing at threshold, synaptic plasticity and pruning [24, 25, 26]. We consider N neurons at the nodes of the chosen network, characterized by their potential v_i. The neuron positions will then be ordered in space for regular lattices and small world networks, organized in a hierarchical manner for the Apollonian network and randomly chosen in two dimensions for the scale-free and fully connected networks. To each neuron we assign an out-going connectivity degree, k_{out}, on the chosen lattice. We implement scale-free networks according to the distribution measured by fMRI measurements of ongoing activity in humans [28]. Each neuron has a degree equal to a random number between $k_{out}^{min} = 2$ and $k_{out}^{max} = 100$ according to the probability distribution $n(k_{out}) \propto k_{out}^{-2}$. The two neurons are chosen according to a distance dependent probability, $p(r) \propto e^{-r/5<r>}$, where r is their spatial distance [29]. Once the network of output connections is established, we identify the resulting degree of in-connections, k_{in_j}, for each neuron j. To each synaptic connection we assign an initial random strength g_{ij}, where $g_{ij} \neq g_{ji}$, and to each neuron randomly either an excitatory or an inhibitory

character, with a fraction p_{in} of inhibitory synapses. Whenever at time t the value of the potential at a site i is above a certain threshold $v_i \geq v_{max}$, the neuron sends action potentials which arrive to each of the k_{out_i} pre-synaptic buttons and lead to a total production of neurotransmitter proportional to v_i. As a consequence, the total charge that could enter into connected neurons is proportional to $v_i k_{out_i}$. Each of them receives charge in proportion to the strength of the synapses g_{ij}

$$v_j(t+1) = v_j(t) \pm \frac{v_i(t) k_{out_i}}{k_{in_j}} \frac{g_{ij}(t)}{\sum_k g_{ik}(t)} \tag{1}$$

where the sum is extended to all out-going connections of i. The normalization by k_{in_j} in Eq.(2) insures a controlled firing behaviour for neurons with a high number of in-going terminals, whereas the plus or minus sign is for excitatory or inhibitory synapses, respectively. In regular networks neurons have the same number of ingoing and outgoing connections, therefore Eq.(1) reduces to the simpler expression $v_j(t+1) = v_j(t) \pm v_i(t) \frac{g_{ij}(t)}{\sum_k g_{ik}(t)}$. The same consideration holds for small world networks.

The firing rate of real neurons is limited by the refractory period, i.e. the brief period after the generation of an action potential during which a second action potential is difficult or impossible to elicit. The practical implication of refractory periods is that the action potential does not propagate back toward the initiation point and therefore is not allowed to reverberate between the cell body and the synapse. In our model, once a neuron fires, it remains quiescent for one time step and it is therefore unable to accept charge from firing neighbours. This ingredient indeed turns out to be crucial for a controlled functioning of our numerical model. In this way an avalanche of charges can propagate far from the input through the system. The initial values of the neuron potentials are uniformly distributed random numbers and the value of v_{max} is fixed equal to 6 in all simulations. Moreover, a small fraction (10%) of neurons is chosen to be output sites, i.e. an open boundary, with a zero fixed potential, playing the role of sinks for the charge. They model neurons connected to neurons not belonging to the slice and avoid that an excess to charge influx would lead to supercritical behaviour. Each time neuronal activity stops in the network, an external stimulus is necessary to trigger further activity, which therefore mimics the nutrients from the bath needed to keep a real neuronal network alive. This stimulus consists in increasing the potential of a random neuron by a random quantity uniformly distributed between 0 and v_{max}.

During the propagation of an avalanche according to Eq. (1), we identify the bonds connecting two successively active neurons, namely neurons whose activity is correlated. The strength of their connections is increased proportionally to the activity of the synapse, namely the membrane potential variation of the post-synaptic neuron induced by the presynaptic neuron

$$g_{ij}(t+1) = g_{ij}(t) + \alpha(v_j(t+1) - v_j(t))/v_{max} \tag{2}$$

where α a dimensionless parameter. Once an avalanche of firings comes to an end, the strength of all inactive synapses is reduced by the average strength increase per bond

$$\Delta g = \sum_{ij,t} \delta g_{ij}(t)/N_a \tag{3}$$

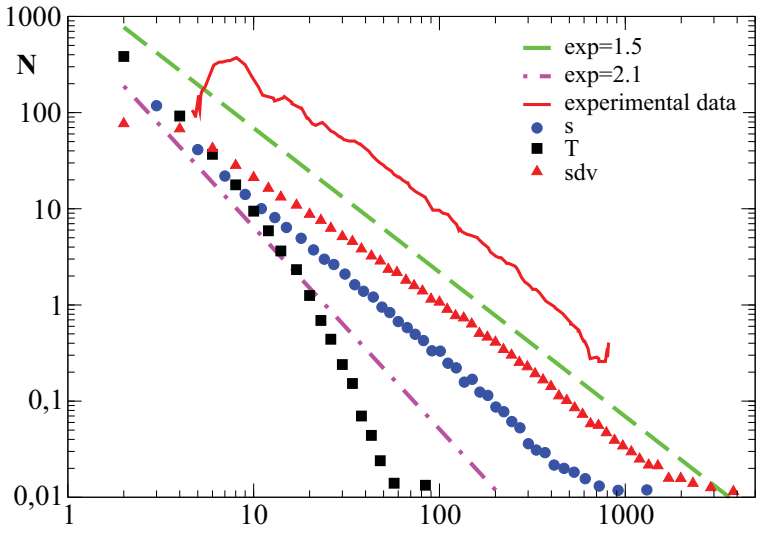


FIGURE 2. The distributions of avalanche size (circles), duration (square) and the total potential variation during one avalanche (triangle) for 100 configurations of scale-free network with $N = 16000$ neurons ($\alpha = 0.6$, $N_p = 10000$, $p_{in} = 0.05$). The dashed line has a slope -1.5, whereas the dot-dashed line has a slope -2.1. The continuous line represents the experimental distribution of avalanche sizes in rat cortex slices. Experimental data are shifted for better comparison.

where N_a is the number of bonds active in the previous avalanche. Here α is the only parameter controlling both the strengthening and the weakening rule in the Hebbian plasticity and represents the ensemble of all possible physiological factors influencing synaptic plasticity. By implementing these rules, our neuronal network "memorizes" the most used paths of discharge by increasing their strength, whereas the less solicited synapses slowly atrophy. Indeed, once the strength of a bond is below an assigned small value $g_t = 10^{-4}$, we remove it, i.e. set its strength equal to zero, which corresponds to the so-called pruning.

We implement synaptic plasticity rules during a series of N_p stimuli in order to let the activity tune the synaptic strengths, initially set at random. The extension of the plastic adaptation procedure then represents the level of experience, or *age*, of the system, whose response we monitor over a time-scale much shorter than the one needed for structural adaptation.

Avalanche activity

After "aging" the system applying plasticity rules during N_p external stimuli, we submit the system to a new sequence of stimuli with no modification of synaptic strengths. The response of the system to this second sequence models the spontaneous activity of a trained neuronal network with a given level of experience. We analyse this activity by measuring the avalanche size distribution $n(s)$ and the time duration distribution $n(T)$.

We measure the distribution of neuronal avalanche sizes, defined either as the total number of firing neurons, or as the sum of their voltage variations during an avalanche.

This distribution exhibits a power law behaviour, with an exponent equal to 1.5 ± 0.1, quite stable with respect to parameters. This scaling behaviour is also robust for densities of inhibitory synapses up to 10%, whereas the scaling behaviour is lost for higher densities. Moreover, the distribution of avalanche temporal durations is also a power law with an exponent close to -2.0. Both these values show an excellent agreement with experimental data. Extensive studies have verified that the critical behaviour of avalanche distributions does not depend on parameter values or network properties (regular, small-world, Apollonian networks). Results imilar to the ones for random initial conductances are found for equal initial conductances. The dependence of the critical behaviour on synaptic strengths has been recently investigated in networks of integrate-and-fire neurons [30]. Only for fully connected networks, which undergo plastic adaptation routines of different length, all networks exhibit supercritical behaviour, namely an excess of very large avalanches, due to the high level of connectivity in the system [31]. Moreover, these scaling properties do not depend on system size, indicating that the network is in a critical state and self-regulates, by adjusting synaptic strengths, producing the observed scale-invariant behaviour.

Implementation of up and down states

In order to model the waiting time distribution measured experimentally,we implement the alternation between up and down-states, both at the level of a single neuron potential and a network state [23]. At the end of each avalanche we measure its size in terms of the sum of depolarizations δv_i of all active neurons, $s_{\Delta v} = \sum \delta v_i$. If the last avalanche is larger than a threshold, $s_{\Delta v} > s_{\Delta v}^{min}$, the system transitions into a down-state and neurons active in the last avalanche become hyperpolarized proportionally to their previous activity, namely we reset

$$v_i = v_i - h\delta v_i \tag{4}$$

where $h > 0$. This equation implies that each neuron is hyperpolarized proportionally to its previous activity, i.e. its potential is the lower, the higher its potential variation in the previous avalanche δv_i. This rule introduces a short range memory at the level of a single neuron and models a number of possible mechanisms: the local inhibition experienced by a neuron, due to spike adaptation, adenosine accumulation, synaptic vesicle depletion, etc.

Conversely, if the avalanche just ended has a size $s_{\Delta v} \leq s_{\Delta v}^{min}$, the system either will remain, or will transition into an up-state. All neurons firing in the previous avalanche are not set equal to zero resting potential but to the depolarized value

$$v_i = v_{\max}(1 - s_{\Delta v}/s_{\Delta v}^{min}) \tag{5}$$

The neuron potential then depends on the response of the whole network via $s_{\Delta v}$, in agreement with measurements of the neuronal membrane potential which remains close to the firing threshold in the up-state. $s_{\Delta v}^{min}$ controls the extension of the up-state and therefore the level of excitability of the system. The high activity in the up-state must be

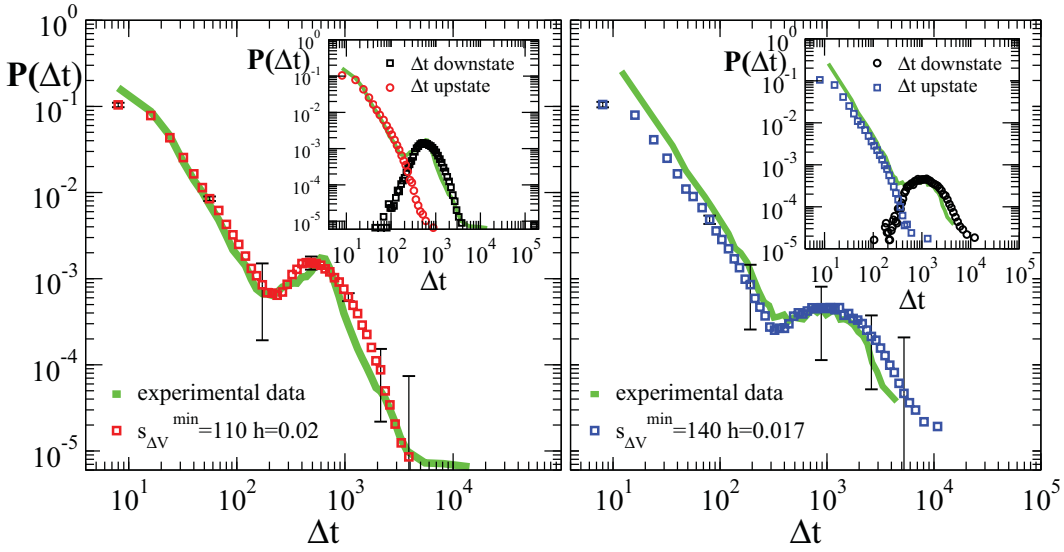


FIGURE 3. (Color online) Waiting time distributions measured experimentally are compared with the average numerical distributions for 100 networks with $N = 64000$ neurons. Left: numerical curve ($s_{\Delta v}^{min} = 140$ and $h = 0.017$) fitting the experimental curve with blue squares in Fig. 1; Right: numerical curve ($s_{\Delta v}^{min} = 110$ and $h = 0.02$) fitting the experimental curve with red diamonds in Fig. 1. For both experimental curves the best agreement is reached by numerical curves with $R \simeq 10^{-4}$. In the inset the waiting time distribution evaluated separately in the up and downstate. Statistical error bars not shown are comparable to the symbol size.

sustained by collective effects in the network, otherwise the depolarized potentials would soon decay to zero, and therefore the random stimulation in the up-state has an amplitude that depends on past activity. Eqs. (4) and (5) each depend on a single parameter, h and $s_{\Delta v}^{min}$, which introduce a memory effect at the level of single neuron activity and the entire system, respectively. In order to reproduce the behavior observed experimentally, the parameters $s_{\Delta v}^{min}$ and h are controlled separately. However, simulations show that the ratio $R = h/s_{\Delta v}^{min}$ is the only relevant quantity controlling the temporal organization of avalanches.

Numerical simulations show that the system indeed switches between up and down states, with different temporal durations. The numerical waiting time distributions (Fig. 3) exhibit the non-monotonic behaviour of the experimental curves, where the position of the minimum is controlled by the value of $s_{\Delta v}^{min}$ and the power law regime scales with the same exponent ~ -2 as experimental data. The different contribution from the two states is reflected in the activity temporal scale. The up-state generates strongly clustered avalanches, originating the power law regime of the waiting time distribution, whose extension depends on $s_{\Delta v}^{min}$. Large Δt between avalanches generated in the upstate are observed with a very small probability, which increases with decreasing h. Conversely, the waiting time distribution evaluated in the down-state has a bell-shaped behaviour centered at large intertimes which depends on h, i.e. for a larger disfacilitation of the network the probability to observe intermediate waiting times decreases in favour of long Δt. The presence of the minimum and the height of the relative maximum are sample dependent (Fig. 1) and for each sample the agreement between numerical and experimental data

depends on the subtle balance between excitation and inhibition. For different samples, optimal agreement is realized when the ratio $R = h/s_{\Delta v}^{min} \simeq 10^{-4}$. Enhancing excitation, by increasing the threshold value $s_{\Delta v}^{min}$, produces a major shift in the data. Increasing inhibition, by increasing the parameter h, generates the opposite effect, recovering the good agreement with experimental data. Interestingly, the avalanche size and duration distributions also reproduce the experimental scaling behaviour for the parameter values expressing the balance between excitatory and inhibitory components. The abrupt transition between the up and down-state, controlled by a threshold mechanism, generates the minimum observed experimentally.

LEARNING

Here we discuss the learning performance of this neuronal network acting in a critical state [26]. We apply Boolean rules at input neurons and we monitor the response at one output neuron. These nodes are randomly placed in the network under the condition that they are not boundary sites and they are mutually separated on the network by k_d nodes. k_d represents the chemical distance on the network and plays the role of the number of hidden layers in a perceptron. We test the ability of the network to learn different rules: AND, OR, XOR and a random rule RAN with three inputs. A single learning step requires the application of the entire sequence of states at the input neurons, monitoring the state of the output neuron. For each rule the binary value 1 is identified with the output neuron firing, namely the neuron membrane potential at a value greater or equal to v_{max} at some time during the activity. Conversely, the binary state 0 at the output neuron corresponds to the physiological state of a real neuron which has been depolarized but fails to reach the firing threshold membrane potential during the entire avalanche propagation. Once the input sites are stimulated, we let the avalanche evolve to its end according to Eq. 1. If at the end of the avalanche the propagation of charge did not reach the output neuron, we consider that the state of the system was unable to respond to the given stimulus, and as a consequence to learn. We therefore increase uniformly the potential of all neurons by units of a small quantity, $\beta = 0.01$, until the configuration reaches a state where the output neuron is first perturbed. We then compare the state of the output neuron with the desired output.

Plastic adaptation is applied to the system according to a non-uniform negative feedback algorithm. Namely, if the output neuron is in the correct state according to the rule, we keep the value of synaptic strengths. Conversely, if the response is wrong we modify the strengths of those synapses involved in the information propagation by $\pm\alpha/d_k$, where d_k is the chemical distance of the presynaptic neuron from the output neuron. The sign of the adjustment depends on the mistake made by the system: If the output neuron fails to be in a firing state we increase the used synapses by a small additive quantity proportional to α. Synaptic strengths are instead decreased by if the expected output 0 is not fulfilled. This adaptation rule intends to mimic the feedback to the wrong answer triggered locally at the output site, for instance by some hormons, and propagating backward towards the input sites.

We first analyse the dependence of the learning performance on the number of neurons. Indeed, as the system size increases the number of highly connected neurons be-

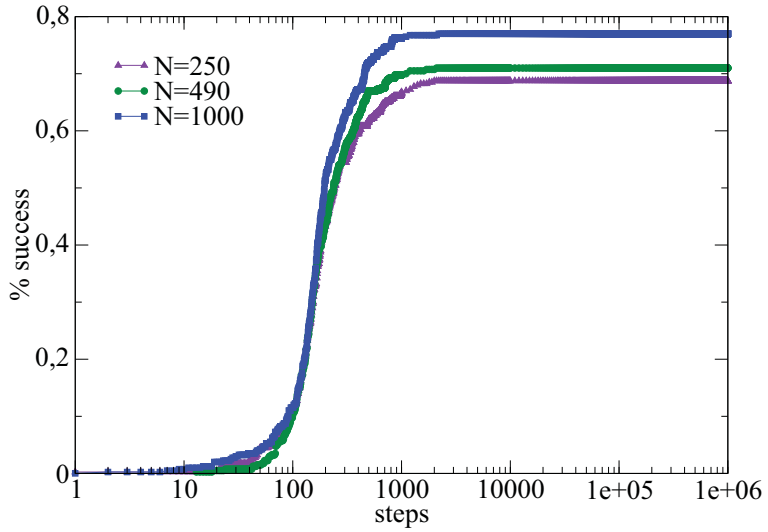


FIGURE 4. Percentage of configurations learning the XOR rule as function of the number of learning steps, for $\alpha = 0.005$, $k_d = 3$, the initial minimum connectivity degree equal to 3 and different numbers of neurons N (from 250 to 1000 bottom to top).

comes larger. A well connected system provides better performances, therefore we could expect that the size dependence reflects the same effect. In Fig. 4 we show data for a set of parameters and different system sizes. The learning performance indeed improves with the system size since the overall level of connectivity improves for larger systems. Next we check the ability of the system to learn the different rules. The fraction of configurations learning the AND rule versus the number of learning steps for different values of the plastic adaptation strength α. We notice that the larger the value of α the sooner the system starts to learn the rule, however the final percentage of learning configurations is lower. The final rate of success increases as the strength of plastic adaptation decreases. This result is due to the highly non-linear dynamics of the model, where firing activity is an all or none event controlled by the threshold. The result that all rules give a higher percentage of success for weaker plastic adaptation, is in agreement with recent experimental findings on visual perceptual learning, where better performances are measured when minimal changes in the functional network occur as a result of learning [32].

CONCLUSIONS

Extensive simulations have shown that a novel brain model with activity dependent plasticity, implemented on different lattices is able to capture the main statistical features of spontaneous brain activity. The ingredients of the model are close to most functional and topological properties of real neuronal networks. The avalanche size and duration distributions show a power law behaviour with exponents 1.5 ± 0.1 and 2.0 ± 0.1, respectively, compatible the values found experimentally for neuronal avalanches. The complex non-monotonic temporal organization of neuronal avalanches is controlled by

the system balance level between excitation and inhibition expressed by the parameter *R*. Avalanches are temporally correlated in the up-state, whereas down-states are long term recovery periods where memory of past activity is erased. The good agreement with experimental data indicates that the transition from an up-state to a down-state has a high degree of synchronization. Moreover it confirms that alternation between up and down-states is the expression of a homeostatic regulation which, during periods of high activity, is activated to control the excitability of the system and avoid pathological behaviour. These collective effects must be supported by the single neuron behaviour, which toggles between two preferential states, a depolarized one in the up-state and a hyperpolarized one in the down-state. The model confirms that the depolarized neuron state is a network effect: the avalanche activity itself determines how close to the firing threshold a neuron stays in the up-state. Conversely, the hyperpolarized state is a form of temporal auto-correlation in the neuron activity. Finally, the investigation of the learning ability of this model has evidenced that the learning dynamics is a cooperative mechanism where all neurons contribute to select the right answer and negative feedback is provided in a non-uniform way. Despite the complexity of the problem and the high number of degrees of freedom involved at each step of the iteration, the system can learn successfully even complex rules. The neuronal network has a “universal" learning dynamics, even complex rules can be learned provided that the plastic adaptation is sufficiently slow.

REFERENCES

1. J. M. Beggs, and D. Plenz, *J. Neurosci.* **23**, 11167–11177 (2003).
2. J. M. Beggs, and D. Plenz, *J. Neurosci.* **24**, 5216–5229 (2004).
3. E. D. Gireesh, and D. Plenz, *Proc. Nat. Acad. Sci. USA* **105**, 7576–7581 (2008).
4. T. Petermann, T. C. Thiagarajan, M. A. Lebedev, M. A. L. Nicolelis, D. R. Chialvo, and D. Plenz, *Proc. Nat. Acad. Sci. USA* **106**, 15921–15926 (2009).
5. A. Mazzoni, F. D. Broccard, E. Garcia-Perez, P. Bonifazi, M. E. Ruaro, and V. Torre, *PLoS ONE* 2(5), e439 1–12 (2007).
6. V. Pasquale, P. Massobrio, L. L. Bologna, M. Chiappalone, and S. Martinoia, *Neuroscience* **153**, 1354–1369 (2008).
7. E. Tagliazucchi, P. Balenzuela, D. Fraiman, and D. Chialvo, *Front. Physiol.* **3**, article 15 (2012).
8. K. J. Staley, M. Longacher, J. S. Bains, and A. Yee, *Nature Neurosci.* **1**, 201–209 (1998).
9. S. M. Thompson, H. L. Haas, and B. H. Gähwiler, *J. Physiol.* **451**, 347–363 (1992).
10. E. Maeda, H. P. Robinson, and A. Kawana, *J. Neurosci.* **15**, 6835–6845 (1995).
11. M. V. Sanchez-Vives, L. G. Nowak, and D. A. McCormick, *J. Neurosci.* **20**, 4286–4299 (2000).
12. I. Timofeev, F. Grenier, and M. Steriade, *Proc. Nat. Acad. Sci. USA* **98**, 1924–1929 (2001).
13. D. Millman, S. Mihalas, A. Kirkwood, and E. Niebur, *Nature Phys.* **6**, 801–805 (2010).
14. I. Timofeev, F. Grenier, M. Bazhenov, T. J. Sejnowski, and M. Steriade, *Cereb. Cortex* **10**, 1185–1199 (2000).
15. D. Eytan, and S. Marom, *J. Neurosci.* **26**, 8465–8476 (2006).
16. D. Plenz, and A. Aertsen, *Neuroscience* **70**, 893–924 (1996).
17. D. Plenz, and S. T. Kitai, *J. Neurosci.* **18**, 266–283 (1998).
18. E. A. Stern, D. Jaeger, and C. J. Wilson, *Nature* **394**, 475–478 (1998).
19. M. O. Cunningham et al, *Proc. Nat. Acad. Sci. USA* **103**, 5597–5601 (2006).
20. A. Hasenstaub, R. N. Sachdev, and D. A. McCormick, *J. Neurosci.* **27**, 9607–9622 (2007).
21. L. de Arcangelis, C. Godano, E. Lippiello, and M. Nicodemi, *Phys. Rev. Lett.* **96**, 0511021–4 (2006).
22. T. L. Ribeiro, M. Copelli, F. Caixeta, H. Belchior, D. R. Chialvo, M. A. L. Nicolelis, and S. Ribeiro, *PLoS ONE* 5(11), e14129 1–14 (2010).

23. F. Lombardi, H. J. Herrmann, C. Perrone-Capano, D. Plenz, and L. de Arcangelis, *Phys. Rev. Lett.* **108**, 2287031–5 (2012).
24. L. de Arcangelis, C. Perrone-Capano, and H. J. Herrmann, *Phys. Rev. Lett.*, **96**, 0281071–4 (2006).
25. G. L. Pellegrini, L. de Arcangelis, H. J. Herrmann, and C. Perrone-Capano, *Phys. Rev. E* **76**, 0161071–9 (2007).
26. L. de Arcangelis, and H. J. Herrmann, *Proc. Natl. Acad. Sci. USA* **107**, 3977–3981 (2010).
27. P. Bak, *How nature works. The science of self-organized criticality*, Springer, New York, 1996.
28. V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, *Phys. Rev. Lett.* **94**, 0181021-4 (2005).
29. B. Roerig, and B. Chen , *Cereb. Cortex* **12**, 187–198 (2002).
30. A. Levina, J.M. Herrmann, and T. Geisel, *Nature Phys.* **3**, 857 (2007); *Phys. Rev. Lett.* **102**, 1181101-4 (2009).
31. L. de Arcangelis, *Eur. Phys. J. Special Topics* **205**, 243-257 (2012).
32. C. M. Lewis, A. Baldassarre, G. Committeri, G. L. Romani, and M. Corbetta, *Proc. Natl. Acad. Sci. USA* **106**, 17558–17563 (2009).

The emergence of spontaneous activity in neuronal cultures

J. G. Orlandi*, E. Alvarez-Lacalle†, S. Teller*, J. Soriano* and J. Casademunt*

*Departament d'ECM, Facultat de Física, Universitat de Barcelona. Martí i Franqués 1, 08028 Barcelona, Spain.
†Departament Física Aplicada, EETAC. Universitat Politècnica de Catalunya BarcelonaTech. Esteve Terrades 5, 08860 Castelldefels, Spain.

Abstract. *In vitro* neuronal networks of dissociated hippocampal or cortical tissues are one of the most attractive model systems for the physics and neuroscience communities. Cultured neurons grow and mature, develop axons and dendrites, and quickly connect to their neighbors to establish a spontaneously active network within a week. The resulting neuronal network is characterized by a combination of excitatory and inhibitory neurons coupled through synaptic connections that interact in a highly nonlinear manner. The nonlinear behavior emerges from the dynamics of both the neurons' spiking activity and synaptic transmission, together with biological noise. These ingredients give rise to a rich repertoire of phenomena that are still poorly understood, including the emergence and maintenance of periodic spontaneous activity, avalanches, propagation of fronts and synchronization. In this work we present an overview on the rich activity of cultured neuronal networks, and detail the minimal theoretical considerations needed to describe experimental observations.

Keywords: neuronal cultures; spontaneous activity.
PACS: 87.19.L-, 87.19.lj, 87.18.Tt

INTRODUCTION

Understanding the activity and network structure of neurons in the mammalian brain, and relating it to a particular brain process or function, is one of the major challenges of modern neuroscience [1]. One of the most fundamental question to address is the emergence of spontaneous activity in neuronal assemblies, a crucial mechanism in living neuronal networks involved in the correct formation, survival and refinement of neuronal circuits. Indeed, rhythmic spontaneous episodes of activity are widespread in neuronal tissues in the form of brain rhythms [2]. However, spontaneous activity is not limited to naturally–formed *in vivo* neuronal tissues. Brain slices and *in vitro* neuronal cultures are also spontaneously active. Specifically, *in vitro* neuronal circuits lead to networks that self–organize, grow and mature to constitute a spontaneously active network [3]. The robust presence of spontaneous activity in such different structures hints at the presence of general mechanisms —in both neuronal dynamics and connectivity— that initiate and control it.

Neuronal cultures are typically prepared from specific regions of embryonic rat brains. After dissociation and plating, neurons connect within a day, and show rich spontaneous activity as early as day *in vitro* 5-6 [4]. Activity is typically monitored through Calcium Fluorescence Imaging, which can simultaneously record thousands of neurons in large

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 25-27 (2013); doi: 10.1063/1.4776497

areas at $\sim$ 20 ms resolution, or through Multi-Electrode Arrays (MEAs) [3], which deliver high, $\sim \mu$s resolution but with a limited number of recording sites, on the order of 100. The emergence of spontaneous activity in cultures is characterized by bursting episodes of collective neuronal activation combined with quieter inter–bursts intervals.

A good starting point to study activity–initiation mechanisms in cultures is to consider networks that are almost unidimensional, i.e. where the width of the culture is much smaller than the characteristic diameter of the neuron's dendritic tree (around 300 μm). The predominant activity pattern in these lines is the appearance of pulses (bursts) of neuronal activity that initiate in localized zones along the line, termed burst initiation zones. This observation was first investigated by Feinerman *et al.* [5, 6], who showed that these pulses propagate at constant velocity, sequentially activating the neurons in their path. Indeed, the line behaves like a chain–like network that extends with equal probability towards both ends of the system. The velocity of the activity front predominantly depends on the connectivity properties of the network, with velocities that double when inhibitory synapses are blocked. The same authors also reported that the nucleation of the front predominantly occurs in the regions with higher connectivity [6].

The natural step after investigation 1D neuronal networks is to test whether these basic behaviors, namely activity–initiation features and burst propagation, also appear in bi–dimensional neuronal cultures. However, this endeavor is difficult to assess since one has to access large areas, and with both high temporal and spatial resolution. In this context, and as we treat below, modeling and in silico simulations may provide a first hint on spontaneous activity driving mechanisms in much more complex networks.

MODELING NEURONAL CULTURES DYNAMICS

The theoretical and numerical analysis [7] of one–dimensional cultures showed that, to reliably reproduce the experimental observations, simple Integrate–and–fire models are not sufficient. The neuron models must include a slow variable that mimics the presence of slow K^+ channels in the soma as well as spike frequency adaptation. The set of equations that include these ingredients while keeping the minimum number of components are described in the two–dimensional Izhikevich model [8]. Furthermore, to fully describe the observed activity, the chemical synapses connecting any two neurons must have short–term depression [9], which takes into account the limited number of available neurotransmitters. This limitation results in a decrease of synaptic efficiency with use, while synapses recover slowly at a constant rate.

In the absence of stimuli, neurons fire spontaneously with a characteristic frequency that follows Poisson statistics. Although the source of spontaneous activity is still unclear, miniature post–synaptic currents generated by the spontaneous release of neurotransmitter's vesicles at the synapses seems to be the leading candidate [10]. This *minis* can be considered as a shot noise, i.e. a Poisson process acting at each synapse and that generates post–synaptic currents, but of smaller amplitude than the evoked ones.

DISCUSSION

The picture of burst initiation becomes less clear when more detailed topological properties are included in systems beyond unidimensional cultures. Studies using MEAs in two dimensional cultures have not revealed any clear pattern of activity, and several studies have theorized about the driving mechanisms of bursting behavior. Those studies range from the concept of leader neurons [11] to the hypothesis of different network structures and neuronal subtypes [12]. It was also suggested, based on both experiments and theoretical models, that activity may be described as neuronal avalanches with distributions that resemble those observed in Self–Organized Criticality [13].

On the other hand, the emergence of collective activity, just as the network reaches maturity, can be explained statically as a problem of quorum percolation [4, 14]. When a critical fraction of neurons become active at the same time, they can excite all their neighbors and in turn, through an iterative process, the entire network. However, a dynamical description of the process that includes all the relevant time scales as well as the network properties [15] is still pending.

In conclusion, the initiation and propagation of spontaneous activity is a challenging problem that has not been fully addressed yet. There are still several open questions concerning the identification of fundamental ingredients at the dynamical and topological levels, in particular the incorporation of realistic connectivity maps.

ACKNOWLEDGMENTS

We thank O. Stetter and D. Battaglia for fruitful discussions. We acknowledge financial support from MICINN (Spain) under projects FIS2010-21924-C02-02, FIS2011-28820-C02-01 and the Generalitat de Catalunya under project 2009-SGR-00014.

REFERENCES

1. E. Bullmore, and O. Sporns, *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
2. G. Buzsáki, and A. Draguhn, *Science* **304**, 1926–1929 (2004).
3. J.-P. Eckmann, O. Feinerman, L. Gruendlinger, E. Moses, J. Soriano, and T. Tlusty, *Phys. Rep.* **449**, 54–76 (2007).
4. J. Soriano, M. Rodríguez Martínez, T. Tlusty, and E. Moses, *Proc. Natl. Acad. Sci. USA* **105**, 13758–13763 (2008).
5. O. Feinerman, M. Segal, and E. Moses, *J. Neurophysiol.* **94**, 3406–3416 (2005).
6. O. Feinerman, M. Segal, and E. Moses, *J. Neurophysiol.* **97**, 2937–2948 (2007).
7. E. Alvarez-Lacalle, and E. Moses, *J. Comput. Neurosci.* **26**, 475–493 (2009).
8. E. M. Izhikevich, *IEEE Trans. Neural Netw. (1990-2011)* **14**, 1569–1572 (2003).
9. M. V. Tsodyks, and H. Markram, *Proc. Natl. Acad. Sci. USA* **94**, 719–723 (1997).
10. D. Cohen, and M. Segal, *J. Neurophysiol.* **106**, 2314–2321 (2011).
11. J.-P. Eckmann, E. Moses, O. Stetter, T. Tlusty, and C. Zbinden, *Front. Comput. Neurosci.* **4** (2010).
12. M. Shein, V. Volman, N. Raichman, Y. Hanein, and E. Ben-Jacob, *Phys. Biol.* **5**, 036008 (2008).
13. A. Mazzoni, F. D. Broccard, E. Garcia-Perez, P. Bonifazi, M. E. Ruaro, and V. Torre, *PLoS ONE* **2**, e439 (2007).
14. O. Cohen, A. Keselman, E. Moses, M. Rodríguez Martínez, J. Soriano, and T. Tlusty, *Europhys. Lett.* **89**, 18008 (2010).
15. T. Tlusty, and J.-P. Eckmann, *J. Phys. A: Math. Theor.* **42**, 5004 (2009).

Critical and resonance phenomena in neural networks

A. V. Goltsev[*,†], M. A. Lopes[*], K.-E. Lee[*] and J. F. F. Mendes[*]

[*]*Department of Physics & I3N, University of Aveiro, 3810-193 Aveiro, Portugal.*
[†]*Ioffe Physico-Technical Institute, 194021, St. Petersburg, Russia.*

Abstract. Brain rhythms contribute to every aspect of brain function. Here, we study critical and resonance phenomena that precede the emergence of brain rhythms. Using an analytical approach and simulations of a cortical circuit model of neural networks with stochastic neurons in the presence of noise, we show that spontaneous appearance of network oscillations occurs as a dynamical (non-equilibrium) phase transition at a critical point determined by the noise level, network structure, the balance between excitatory and inhibitory neurons, and other parameters. We find that the relaxation time of neural activity to a steady state, response to periodic stimuli at the frequency of the oscillations, amplitude of damped oscillations, and stochastic fluctuations of neural activity are dramatically increased when approaching the critical point of the transition.

Keywords: neuronal networks; brain rhythms; phase transitions; band pass filter; resonance; noise.
PACS: 87.19.lc, 87.19.lj, 87.19.ll, 87.19.lm, 87.19.ln

INTRODUCTION

Brain rhythms contribute in every aspect of brain function from sensory and cognitive processing, and memory to motor control [1]. Origin and physiological functions of brain rhythms are a topic problem in neuroscience. Brain rhythms are also related to many unusual phenomena observed in the brain. Interactions between billions of neurons give rise to phase transitions, self-organization, and critical phenomena [2, 3]. Phase transitions were observed, for example, in human bimanual coordination [4, 5, 6, 7, 8] and in living neural networks stimulated by electric fields [9]. There are evidences that epileptic seizures, alpha and gamma oscillations, and the ultraslow oscillations of BOLD fMRI patterns emerge as a result of non-equilibrium phase transitions. Neural avalanches are one more example of critical collective phenomena observed in the brain [10, 3].

Various resonance phenomena were also observed in the brain. Experimental investigations of CA1 neuronal networks from mammalian brain demonstrated that stochastic resonance can enhance effects of intrinsic 4-10 Hz hippocampal theta and 40 Hz gamma oscillations [11]. Recently, using a functional imaging technique, Sasaki *et al.* [12] revealed that the majority of rat CA1 neurons act collectively like a band-pass filter. Damped oscillations and the Berger effect are also related to brain rhythms. The Berger effect manifests itself in activation of alpha waves on the electroencephalogram when the eyes are closed and diminution of alpha waves when they are opened [13].

In the present paper, we study collective dynamics of neural networks composed by excitatory and inhibitory neurons in the presence of noise. Based on exact analytical calculations and numerical simulations, we show that spontaneous emergence of network

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 28-35 (2013); doi: 10.1063/1.4776498

oscillations occurs as a dynamical (non-equilibrium) phase transition at a critical level of noise. The transition manifests itself in slowing down of the relaxation of a perturbed neural activity to a steady state, a strong enhancement of stochastic fluctuations of activities of neural populations and an increase of the linear response function to afferent periodic stimuli at the frequency of neural oscillations. We show that near to the critical boundary, neural networks act as damped harmonic oscillators or band-pass filters that pass frequencies within a certain range and attenuate frequencies outside that range.

CORTICAL CIRCUIT MODEL

We use a cortical circuit model [14] composed of N_e pyramidal cells (excitatory neurons) and N_i interneurons (inhibitory neurons) that form a sparsely connected network. The probability that there is a synaptic connection between two neurons is c/N where $N = N_e + N_i$ is the total number of neurons and c is the mean degree. This network has the structure of a directed classical random graph (or Erdős-Rényi graph) with the Poisson degree distribution $P_n(c) = c^n e^{-c}/n!$ where n is the number of presynaptic neurons. Neurons receive sporadic inputs from a remote part of the cortex and synaptic noise. Neurons fire with a constant firing frequency ν that is the same for both excitatory and inhibitory neurons. The total input V_m to a neuron with index m, $m = 1,2,\ldots,N$, is the sum of random spikes from noise, excitatory and inhibitory neurons,

$$V_m(t) = \sum_{n=1}^{N} k_n(t) a_{nm} J_{nm} + \xi(t), \tag{1}$$

where $k_n(t)$ is the number of spikes that arrive from presynaptic neuron n during the time interval $[t-\tau, t]$, τ is the integration time. Below we will consider the case $\tau\nu \leq 1$ when the number of spikes $k_n(t)$ is 1 or 0. If we assume that the emissions times of spikes of different neurons are uncorrelated, then the parameter $\tau\nu$ has a meaning of the probability that a postsynaptic neuron receives a spike from an active presynaptic neuron during time τ. Furthermore, a_{nm} is the adjacency matrix, i.e., $a_{nm} = 1$ if there is a direct edge from neuron n to neuron m, otherwise $a_{nm} = 0$. J_{nm} is the efficacy of the synapse connecting neuron n with neuron m. J_{nm} is positive if presynaptic neuron n is excitatory and it is negative if the neuron is inhibitory. $\xi(t)$ is the number of random spikes from noise that neuron m receives during the time interval $[t-\tau, t]$. We use the Gaussian distribution for $\xi(t)$,

$$G(\xi) = A \exp\Big[-\frac{(\xi - \langle n \rangle)^2}{2\sigma^2}\Big], \tag{2}$$

where A is the normalization constant, σ^2 is the variance, $\langle n \rangle$ is the mean number of random spikes determined by the mean rate ω_{rs}, $\langle n \rangle = \omega_{rs}\tau$. Note that noise in our model is actually shot noise. According to Schottky's theorem, the intensity of this noise is proportional to $\langle n \rangle$.

We consider stochastic neurons. Their response on input is a stochastic process that occurs with a certain rate. Two rules determine dynamics of stochastic neurons [14]:

1. If the total input $V_m(t)$ at an inactive excitatory or inhibitory neuron m at time t is at least a certain threshold Ω (i.e., $V_m(t) \geq \Omega$), then this neuron is activated at a rate μ_e or μ_i, respectively, and fires with a cyclic frequency ν.
2. Active excitatory (inhibitory) neuron m is inactivated at a rate μ_e (μ_i) if $V_m(t) < \Omega$.

We assume that $1/\mu_e$ and $1/\mu_i$ are of the order of the first spike latencies of excitatory and inhibitory neurons, respectively. We introduce the ratio

$$\alpha \equiv \mu_i/\mu_e \tag{3}$$

that plays an important role in our model, as it will be shown below. The advantage of this model with stochastic neurons is that it can be solved analytically.

In numerical simulations, we studied sparsely connected networks of size $N = 10^3 - 10^5$ and applied the following algorithm. We divided time t into intervals of width $\Delta t = \tau$. At each time step, for each neuron, we calculated the input Eq. (1), taking into account that each active presynaptic neuron contributes with a spike with probability $\tau\nu$. The number of random spikes from noise in this input is generated according to the Gaussian distribution, Eq. (2). Then, with the probability $\tau\mu_a$, $a = e,i$, we updated the states of all neurons using the stochastic rules formulated above. We used the following parameters: the fraction of excitatory neurons is $g_e = N_e/N = 75\%$, the fraction of inhibitory neurons is $g_i = N_i/N = 25\%$, the mean number of connections $c = 1000$ (750 excitatory and 250 inhibitory connections), the threshold $\Omega = 30$, and the variance of noise $\sigma^2 = 10$. Following [15], we chose $J_{ie} = J_{ii} \equiv J_i$, $J_{ee} = J_{ei} \equiv J_e$, and $J_i = -3J_e$. These parameters agree with anatomical estimates for cortex. In cortex, the fraction g_i of inhibitory neurons is between 0.15 and 0.3, the mean number of synaptic connections c is about 7000. The threshold Ω is between 15 and 30 in neural networks *in vivo* [9] and about $30-400$ in the brain. The level of noise $\langle n \rangle$ was varied in the interval $0-150$ spikes per integration time τ. We also assumed that, for simplicity, $\tau\nu = 1$ and $\tau\mu_e = 0.1$.

Dynamical behavior of the model is described by the fractions $\rho_e(t)$ and $\rho_i(t)$ of active excitatory and inhibitory neurons, respectively, at time t. We will call them 'activities' of the neural populations. Using the rules of the stochastic dynamics formulated above and assuming that activities are changed slightly during the integration time τ, in the infinite size limit $N \to \infty$, we find a rate equation [14],

$$\frac{d\rho_a(t)}{\mu_a dt} = f_a(t)(1-\rho_a(t)) - \rho_a(t) + \Psi_a(\rho_e(t),\rho_i(t)). \tag{4}$$

for $a = e,i$. The function $\Psi_a(\rho_e,\rho_i)$ is the probability that at time t the input to a randomly chosen excitatory or inhibitory neuron is at least the threshold Ω. For the model under consideration $\Psi_i(\rho_e,\rho_i) = \Psi_e(\rho_e,\rho_i) \equiv \Psi(\rho_e,\rho_i)$, where

$$\Psi(\rho_e,\rho_i) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{\xi=-\infty}^{\infty} \Theta(J_e k + J_i l + \xi - \Omega) G(\xi) P_k(g_e\rho_e\widetilde{c}) P_l(g_i\rho_i\widetilde{c}). \tag{5}$$

Here $\Theta(x)$ is the Heaviside step function, the parameter $\widetilde{c}$ is defined as $\widetilde{c} \equiv c\nu\tau$, and $P_k(g_e\rho_e\widetilde{c})$ and $P_l(g_i\rho_i\widetilde{c})$ are the probabilities that a randomly chosen neuron receives k

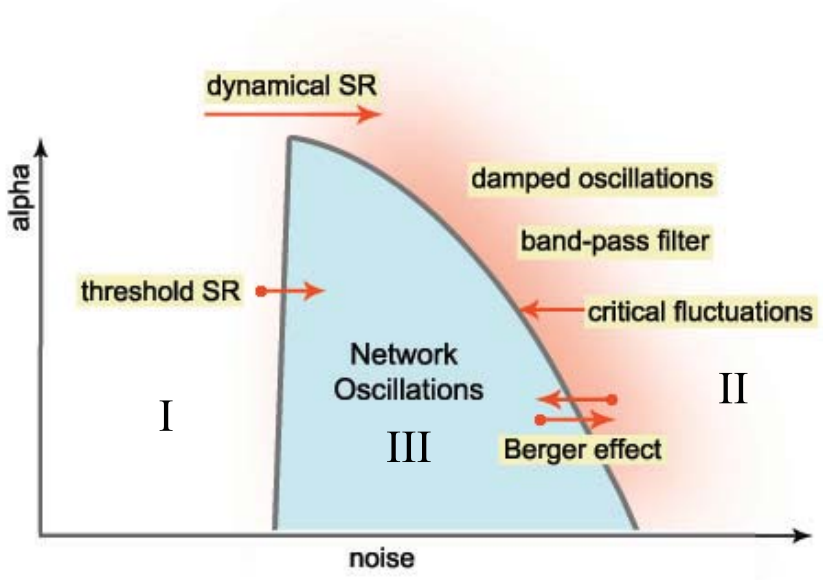


FIGURE 1. Schematic phase diagram of the cortical model and critical and resonance phenomena near the critical boundary of the non-equilibrium phase transition to sustained network oscillations.

spikes from active presynaptic excitatory and l spikes from inhibitory neurons, respectively, during the time window τ at given activities ρ_e and ρ_i. The functions $f_e(t)$ and $f_i(t)$ represent a rate of spontaneous activation of excitatory and inhibitory neurons, respectively, by stimulus, for example, an electric field. The rate equation (4) is similar to the Wilson-Cowan equations [16, 17], see also [14]. Equation (4) is asymptotically exact in the limit $N \to \infty$.

Steady states of the neural populations can be found from Eq. (4), supposing $d\rho_a/dt = 0$ in the limit $t \to \infty$. If $\rho_e(t)$ and $\rho_i(t)$ at time t are close to steady state activities $\rho_e(\infty)$ and $\rho_i(\infty)$, then Eq. (4) enables us to describe relaxation of $\rho_a(t)$ to the steady state. We introduce

$$\delta\rho_a(t) \equiv \rho_a(t) - \rho_a(\infty) = Re(A_a e^{-\gamma t}) \tag{6}$$

where A_a is a complex amplitude. Using the standard perturbation theory, we solve Eq. (4) in the first order in $\delta\rho_a(t)$. We find

$$\gamma_\pm = \frac{1}{2}(B_1 + B_2) \pm \frac{1}{2}\Big[(B_1 - B_2)^2 + 4\alpha D_{ei} D_{ie}\Big]^{1/2}, \tag{7}$$

where we introduced parameters $B_1 = 1 - D_{ee}$, $B_2 = \alpha(1 - D_{ii})$, $D_{ab} = d\Psi_a(\rho_e, \rho_i)/d\rho_b$ for $a, b = e, i.$ respectively. Derivatives D_{ab} are determined by the activities $\rho_e(\infty)$ and $\rho_i(\infty)$ from the non-linear equation Eq. (4) when $d\rho_a/dt = 0$ [14]. The real and imaginary parts of the complex rate γ ($\gamma_r \equiv Re(\gamma_-)$ and $\gamma_i \equiv Im(\gamma_-)$) determine the relaxation rate and the angular frequency of damped oscillations, respectively. Notice that the period of the oscillations equals $2\pi/\gamma_i$.

Analyzing behavior of γ_r and γ_i in dependence on α and $\langle n \rangle$, we obtain the phase diagram in Fig. 1. One can see that there are three regions. There is a region I (small noise level and/or large α) where the relaxation of the neural activity to a steady state is exponential ($\gamma_r > 0$ and $\gamma_i = 0$). In region II, the neural activity relaxes in a form of damped oscillations ($\gamma_r > 0$ and $\gamma_i \neq 0$). In region III, network oscillations are sustained. A similar phase diagram was found in [14] for a simpler model. If α is above a critical value α_t, that corresponds to the α-coordinate of the top point of the region III in Fig. 1, then with increasing the noise level $\langle n \rangle$, the activities ρ_e and ρ_i in the steady state undergo a first-order phase transition at a critical noise level n_c. A similar discontinuous transition was observed in living neural networks *in vitro* when living neural networks were stimulated by an electric field [9]. Neuronal avalanches are precursors of this phase transition. Activation (or inactivation) of one neuron can trigger avalanche process of activation (or inactivation) of a cluster of neurons. In cortex, neuronal avalanches have been observed experimentally [10], see the review [3].

If the parameter $\alpha < \alpha_t$, sustained networks oscillations appear in a certain 'optimal' range of the noise level $\langle n \rangle$ between two critical points. Weak noise can not stimulate network oscillations. Too strong noise over-activates neural networks and only damped oscillations can occur. The critical boundary of region with the sustained oscillations is determined by the condition that the relaxation rate γ_r is zero,

$$\gamma_r = Re(\gamma_-) = 0, \tag{8}$$

where the complex frequency γ_- is given by Eq. (7). For the parameters given above and $\tau = 10$ ms, frequencies of the oscillations lie in the range of brain waves (1– 100 Hz).

LINEAR RESPONSE FUNCTION AND BAND-PASS FILTER BEHAVIOR

Now we study critical phenomena that precede the non-equilibrium phase transition from asynchronious dynamics to sustained oscillations. For this purpose we calculate the linear response of the neural network to a time-dependent stimulus $f_e(t)$ and $f_i(t)$ in Eq. (4) for region I and II. Here we are not studying a response in region III that needs a special consideration. A response of the neural population $a = e, i$ to a weak stimulus $f_a(t)$ is determined by the linear response function $\chi_{ab}(t-t')$,

$$\Delta\rho_a(t) \equiv \rho_a(t) - \rho_a(\infty) = \sum_{b=e,i} \int_{-\infty}^{t} \chi_{ab}(t-t') f_b(t') dt'. \tag{9}$$

Solving Eq. (4) in the linear-response regime, we find that in the regions I and II the neural network behaves as a damped oscillator driven by a force $F_e(t)$,

$$\frac{d^2\Delta\rho_e(t)}{dt^2} + 2\zeta\omega_0 \frac{d\Delta\rho_e(t)}{dt} + \omega_0^2 \Delta\rho_e(t) = F_e(t), \tag{10}$$

(see, for example, in [18]). Here we introduced the damping ratio $\zeta = \gamma_r/\omega_0$ and a frequency $\omega_0 = (\gamma_r^2 + \gamma_i^2)^{1/2}$. In region I, the network is critically damped because $\zeta = 1$

and it is underdamped in region II, where $\zeta < 1$. In the case $f_e(t) \neq 0$ and $f_i(t) = 0$, the force $F_e(t)$ equals $F_e(t) = (1-\rho_e(\infty))(B_2 f_e(t) + d f_e(t)/dt)$. The parameter B_2 was defined above. Solving Eq. (10) leads to a response function,

$$\chi_{ee}(t-t') = X_e e^{-\gamma_r(t-t')} \sin\Big[\gamma_i(t-t') + \Phi_e\Big]. \tag{11}$$

where $X_e = (1-\rho_e(\infty))[1+(B_2-\gamma_r)^2/\gamma_i^2]^{1/2}$ and $\Phi_e = \tan^{-1}[\gamma_i/(B_2-\gamma_r)]$ (one finds a similar result for X_i and Φ_i of inhibitory neurons). If $\gamma_r > 0$, then Eq. (11) shows loss of memory in the neural network with increasing time interval $t-t'$. If γ_r tends to zero, the memory becomes long-range. The Fourier transform $\widetilde{\chi}_{ee}(\omega)$ of the linear response function is

$$\widetilde{\chi}_{ee}(\omega) = \frac{(1-\rho_e)(i\omega + B_2)}{\omega_0^2 - \omega^2 + 2i\zeta\omega_0\omega}. \tag{12}$$

Equation (12) shows that at $\zeta < 1$, the neural network acts as a band-pass filter. The spectral intensity as a function of ω has a maximum at a resonance frequency $\omega_r \approx \omega_0\sqrt{1-2\zeta^2}$ at $\zeta < 1/\sqrt{2}$. The maximum value $\|\widetilde{\chi}_{ee}(\omega_r)\|^2$ depends on the noise level $\langle n \rangle$. When approaching the critical point, $\gamma_r \to 0$, the value $\|\widetilde{\chi}_{ee}(\omega_r)\|^2$ diverges as $\widetilde{\chi}_{ee}(\omega_r) \propto 1/\gamma_r^2 \to \infty$, while the angular frequency of damped oscillations γ_i tends to the frequency of stable network oscillations. This behavior signals that, in this regime, in the presence of noise, a neuronal network can amplify periodic signals. This amplification may be a mechanism of stochastic resonance observed in brain [19].

The band pass filter behavior described by Eq. (12) seems to be supported by measurements of response of rat CA1 neurons to afferent stimulation *in vitro* [12]. These measurements revealed that the majority of rat CA1 neurons act collectively like a band-pass filter and fire synchronously in response to a limited range of presynaptic firing rates ($20-40$ Hz) that are in the range of gamma oscillations in the rat hippocampus [20]. One can also note that, a long time ago, a number of characteristics of a band-pass filter behavior and a resonance response on sin wave trains already have been observed in EEG recordings of alpha activity [21]. Based on Eq. (12), we suggest that band-pass filter behavior observed in Sasaki et al. [12] and Tweel [21] is a manifestation of the critical phenomena near to the transition to neural network oscillations.

STOCHASTIC FLUCTUATIONS OF NEURONAL ACTIVITY

EEG measurements demonstrate that brain activity always contains a stochastic component. In this section we will show that stochastic fluctuations are enhanced when a neural network is close to the critical point of the non-equilibrium phase transition. For characterizing stochastic fluctuations, we introduce the autocorrelation function

$$C_{ab}(t) = \frac{1}{T}\int_0^T \delta\rho_a(t_1)\delta\rho_b(t_1+t)dt_1, \tag{13}$$

where $\delta\rho_a(t) = \rho_a(t) - \overline{\rho}_a$ describes fluctuations of activity $\rho_a(t)$ of population a, $a = e, i$, around the mean value $\overline{\rho}_a$ (see, for example, Ref. [22]). $C_{ab}(t)$ is a measure

of correlations between values of $\delta\rho_a(t_1)$ and $\delta\rho_b(t_1+t)$ at two different instants separated by a lag t and averaged over an arbitrary large time window T. The Wiener-Khintchine theorem states that the power density spectrum of the fluctuations is the Fourier transform of the autocorrelation function.

For calculating the autocorrelation function, one uses the standard method [22, 23]. In the deterministic equation (4), we assume that $f_a(t)$ is a stochastic force that satisfies conditions $\langle f_a(t)\rangle = 0$ and $\langle f_a(t) f_b(t')\rangle = f_0^2 \delta(t-t')\delta_{a,b}$. If fluctuations are small, the autocorrelation function may be found in the linear response theory [22, 23]. Assuming, for simplicity, $f_i(t) = 0$, we obtain Eq. (9) that leads to

$$C_{ee}(t) = 2\pi f_0^2 \int_{-\infty}^{\infty} e^{i\omega t} \|\widetilde{\chi}_{ee}(\omega)\|^2 d\omega, \tag{14}$$

where the linear response function $\widetilde{\chi}_{ee}(\omega)$ is given by Eq. (12). In the region of damped oscillations, the autocorrelation function $C_{ee}(t)$ has a form

$$C_{ee}(t) = A_e e^{-\gamma_r |t|} \cos\Big(\gamma_i |t| + \Phi_e\Big). \tag{15}$$

The parameter A_e and the phase Ψ_e behave as $A_e \propto 1/\gamma_r$ and $\Phi_e \propto \gamma_r/\gamma_i$ at small γ_r. For inhibitory neurons we obtain a similar behavior. Thus, stochastic fluctuations of activities of excitatory and inhibitory neural populations are enhanced when approaching the critical point $\gamma_r = 0$, Eq. (8), of the emergence of network oscillations (see Fig. 1). However, the linear-response approximation is not valid when fluctuations become sufficiently large. This occurs near to the non-equilibrium phase transition and non-perturbative methods are required for calculating $C_{ab}(t)$.

CONCLUSION

In the present paper, using a cortical model with stochastic neurons, we have showed that, in neuronal networks, spontaneous appearance of sustained network oscillations occurs as a non-equilibrium phase transition. The critical point is determined by the level of noise, structure of the neural network, the balance between excitatory and inhibitory neurons, and other parameters. We have found critical and resonance phenomena that precede the transition. The important property of this transition is that, at the critical point, the relaxation time of the neuronal activity to a steady state becomes infinite in the infinite size limit. An increase of the response of neural networks to periodic afferent stimulations and a strong enhancement of stochastic fluctuations of activities of neural populations are also the critical phenomena that precede the transition. Note, that these phenomena are general properties of second-order phase transitions observed in physical, chemical and biological systems (see, for example, Stanley [24], Haken [25], Kelso [8]). These critical phenomena have been observed near the non-equilibrium phase transition in human hand movements [4, 5, 6, 7, 8]. The noise-induced nonequilibrium phase transition found in [26] is one more example of a phase transition with similar critical phenomena. Furthermore, we have demonstrated that near to the critical point, neuronal networks behave as damped harmonic oscillators or band-pass filters in agreement with

band-pass filter behavior observed *in vitro* in networks of CA1 neurons in mammalian brain [12]. We suggest that band-pass filter behavior is a manifestation of critical phenomena near to the transition to network oscillations.

We have also demonstrated that, in the cortical model, stochastic neural activity generated by a stochastic force is similar to spontaneous alpha activity observed in EEG recordings of both a normal man and human epileptic seizures of petit mal activity [27].

ACKNOWLEDGMENTS

This work was partially supported by the following projects PTDC: SAU-NEU/ 103904/2008, FIS/108476/2008, MAT/114515/2009, and PEst-C/CTM/LA0025/2011. K. E. Lee was supported by FCT under grant SFRH/BPD/ 71883/2010, M. A. L. was supported by FCT under Grant No. SFRH/BD/ 68743/2010.

REFERENCES

1. G. Buzsaki, *Rhythms of the Brain*, Oxford University Press, USA, 2006.
2. D. R. Chialvo, *Nature Phys.* **2**, 301–302 (2006).
3. D. R. Chialvo, *Nature Phys.* **6**, 744–750 (2010).
4. J. A. Kelso, *Am. J. Physiol.: Reg. Int. & Comp. Physiol.* **246**, R1000–R1004 (1984).
5. H. Haken, J. A. S. Kelso, and H. Bunz, *Biol. Cybern.* **51**, 347–356 (1985).
6. J. Kelso, J. Scholz, and G. Schoner, *Phys. Lett. A* **118**, 279 – 284 (1986).
7. J. Scholz, J. Kelso, and G. Schoner, *Phys. Lett. A* **123**, 390 – 394 (1987).
8. J. A. S. Kelso, *Front. Hum. Neurosci.* **4**, 23 (2010).
9. J.-P. Eckmann, O. Feinerman, L. Gruendlinger, E. Moses, J. Soriano, and T. Tlusty, *Phys. Rep.* **449**, 54–76 (2007).
10. J. M. Beggs, and D. Plenz, *J. Neurosci.* **23**, 11167–11177 (2003).
11. B. J. Gluckman, T. I. Netoff, E. J. Neel, W. L. Ditto, M. L. Spano, and S. J. Schiff, *Phys. Rev. Lett.* **77**, 4098–4101 (1996).
12. T. Sasaki, R. Kimura, M. Tsukamoto, N. Matsuki, and Y. Ikegaya, *J. Physiol.* **574**, 195–208 (2006).
13. R. Hari, and R. Salmelin, *Trends Neurosci.* **20**, 44 – 49 (1997).
14. A. V. Goltsev, F. V. de Abreu, S. N. Dorogovtsev, and J. F. F. Mendes, *Phys. Rev. E* **81**, 061921 (2010).
15. D. J. Amit, and N. Brunel, *Cereb. Cortex* **7**, 237–252 (1997).
16. H. R. Wilson, and J. D. Cowan, *Biophys. J.* **12**, 1 – 24 (1972).
17. H. R. Wilson, and J. D. Cowan, *Biol. Cybern.* **13**, 55–80 (1973).
18. R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II, Nonequilibrium Statistical Mechanics*, Springer-Verlag, New York, 1985.
19. F. Moss, L. M. Ward, and W. G. Sannita, *Clin. Neurophysiol.* **115**, 267–281 (2004).
20. J. Csicsvari, B. Jamieson, K. D. Wise, and G. Buzsáki, *Neuron* **37**, 311 – 322 (2003).
21. L. H. Tweel, *Docum. Ophthal.* **18**, 287–304 (1964).
22. C. W. Gardiner, *Handbook of stochastic methods : for physics, chemistry and the natural sciences*, Springer, Berlin, Heidelberg, New York, 2002.
23. P. Hänggi, and H. Thomas, *Phys. Rep.* **88**, 207–319 (1982).
24. H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, Oxford, 1971.
25. H. Haken, *Synergetics. An introduction. Nonequilibrium phase transitions and self organization in physics, chemistry and biology.*, Springer, Berlin, New York, 1983, 3rd edn.
26. C. Van den Broeck, J. M. R. Parrondo, and R. Toral, *Phys. Rev. Lett.* **73**, 3395–3398 (1994).
27. A. Babloyantz, and A. Destexhe, *Proc. Natl. Acad. Sci. USA.* **83**, 3513–3517 (1986).

Critical behavior near a phase transition between retrieval and non-retrieval regimes in a LIF network with spatiotemporal patterns

S. Scarpetta* and A. de Candia†

*Dipartimento di Fisica "E.R.Caianiello" Università di Salerno, INFN gruppo collegato di Salerno
Via Ponte don Melillo, Stecca 9, 84084 Fisciano (SA), Italy.
†Dipartimento di Scienze Fisiche, Università di Napoli "Federico II", CNR-SPIN and INFN,
Sezione di Napoli
Complesso Universitario di Monte S. Angelo, Via Cintia, Edificio 6, 80126 Naples, Italy.

Abstract. We study the associative memory dynamics of a network of spiking integrate and fire neurons with Poisson noise. We introduce an order-parameter, and we study the critical regime at the transition between the region of persistent replay of stored patterns and the region of no-reply. At critical spiking threshold the order parameter fluctuations are maximized, as expected for a phase transition. Moreover we also found that, at the critical point, the avalanche size and duration distributions follow power laws. In conclusion our simple model suggests that avalanche power laws in cortical spontaneous activity may be the effect of a network at the critical point between successful associative memory regime and no-retrieval regime.

Keywords: associative memory; phase transitions; criticality.
PACS: 87.18.Sn, 87.19.lv, 87.19.lj

INTRODUCTION

It has been conjectured that the cortex operates near a critical point [1, 2, 3], as reflected also by the power laws of avalanches size distribution, and maximization of fluctuations. Notably, also large scale fMRI analysis [4] demonstrates that the resting brain spends most of the time near the critical point of a second order transition, and exhibits avalanches of activity ruled by the same dynamical and statistical properties described for neuronal events at smaller scales.

Here we study the different regimes of a network of integrate and fire units, whose learning mechanism is based on the Spike-Time-Dependent Plasticity, subject to external noise. The temporal patterns we consider are periodic spatiotemporal patterns of spikes. The storage capacity of this model, in absence of noise, has been studied in [5, 6]. Here we study the spontaneous dynamics, in absence of any stimulation, when Poissonian noise is added to the post-synaptic potential of the units. An important result of this paper is the study of the critical point and of the different regimes observed by changing the excitability parameters of the network. To characterize the transition between the regime of permanent replay and the regime of no-replay, we evaluate the order parameter and its fluctuations as a function of the spiking threshold of the coupled neurons. Finally, avalanches size and duration distributions are studied at this critical point.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 36-43 (2013); doi: 10.1063/1.4776499

THE MODEL

The spiking model has been introduced in [5, 7]. Here we briefly review the model and add the noise term. We have N coupled units. The single unit model is a Leaky Integrate-and-Fire (LIF) in the Spike Response Model (SRM) formulation [9, 8]. In this formulation, the post-synaptic membrane potential of neuron i is given by:

$$h_i(t) = \sum_j J_{ij} \sum_{\hat{t}_j > \hat{t}_i} \varepsilon(t - \hat{t}_j) + \eta_i(t), \tag{1}$$

where $\eta_i(t)$ is a Poissonian noise, J_{ij} are the synaptic connections, $\varepsilon(t)$ describes the response kernel to incoming spikes, and the sum over $\hat{t}_j$ runs over all pre-synaptic firing times following the last spike of neuron i. Namely, each pre-synaptic spike j, with arrival time $\hat{t}_j$, is supposed to add to the membrane potential a post-synaptic potential of the form $J_{ij}\varepsilon(t - \hat{t}_j)$, where

$$\varepsilon(t - \hat{t}_j) = K \left[\exp\left(-\frac{t - \hat{t}_j}{\tau_m} \right) - \exp\left(-\frac{t - \hat{t}_j}{\tau_s} \right) \right] \Theta(t - \hat{t}_j) \tag{2}$$

where τ_m is the membrane time constant (here 10 ms), τ_s is the synapse time constant (here 5 ms), $\Theta(t)$ is the Heaviside step function, and K is a multiplicative constant chosen so that the maximum value of $\varepsilon(t)$ is one.

The Poissonian noise $\eta_i(t)$ is modelled as

$$\eta_i(t) = \sum_{\hat{t}_{\text{noise}} > \hat{t}_i} J_{\text{noise}} \varepsilon(t - \hat{t}_{\text{noise}}). \tag{3}$$

The times $\hat{t}_{\text{noise}}$ and the strengths J_{noise} are extracted randomly and independently for each neuron i, The intervals between times $\hat{t}_{\text{noise}}$ on the single neuron are extracted from a Poissonian distribution $P(\delta t) \propto e^{-\delta t / \tau_{\text{noise}}}$, while the strength J_{noise} is extracted for each time $\hat{t}_{\text{noise}}$ from a Gaussian distribution with mean $\bar{J}_{\text{noise}}$ and standard deviation $\sigma(J_{\text{noise}})$.

When the membrane potential $h_i(t)$ exceeds the spiking threshold θ^i_{th}, a spike is scheduled, and the membrane potential is reset to the resting value zero. While in previous work [5] we used a unique value of spiking threshold θ^i_{th} for all units, here we use two values of θ^i_{th}, a low threshold θ_{th1} for $N_1 < N$ to model units more sensible to noise, and a higher threshold θ_{th2} for the others $N_2 = N - N_1$ units.

Numerical simulations of the dynamics are performed for a network with P stored patterns, where connections J_{ij} are determined via the learning rule inspired to the STDP, previously introduced in [10, 11, 12, 7, 13]. The synaptic strength of the connection J_{ij}, due to a periodic spike train of period T^μ , when the learning time is longer than the period T^μ of the learned pattern, is formulated as follows:

$$\delta J^\mu_{ij} = \sum_{n=-\infty}^{\infty} A(t^\mu_j - t^\mu_i + nT^\mu) \tag{4}$$

where t^μ_j are the spike times of the neuron j in the pattern μ, and the learning window $A(\tau)$, used to model STDP, is the one introduced and motivated by [14], $A(\tau) =$

$a_p e^{-\tau/T_p} - a_D e^{-\eta\tau/T_p}$ if $\tau > 0$, $A(\tau) = a_p e^{\eta\tau/T_D} - a_D e^{\tau/T_D}$ if $\tau < 0$, with the same parameters used in [14] to fit the experimental data of [15], $a_p = \gamma[1/T_p + \eta/T_D]^{-1}$, $a_D = \gamma[\eta/T_p + 1/T_D]^{-1}$, with $T_p = 10.2$ ms, $T_D = 28.6$ ms, $\eta = 4$, $\gamma = 42$. Notably, this function $A(\tau)$ satisfies the balance condition $\int_{-\infty}^{\infty} A(\tau)d\tau = 0$. When $A(\tau)$ is used in Eq. (4) to learn phase-coded patterns with uniformly distributed phases, then the balance condition assures that the averaged connections $(1/N)\sum_j J_{ij}$ are of order $1/\sqrt{N}$, and therefore it assures a balance between excitation and inhibition.

The spike patterns used in this work are periodic spatio-temporal sequences, made up of one spike per cycle and each of which has a phase ϕ_j^μ randomly chosen from a uniform distribution in $[0, 2\pi)$. The set of timing of spikes of unit j can be defined as

$$t_j^\mu + nT^\mu = (\frac{\phi_j^\mu}{2\pi} + n)T^\mu.$$

Thus, each pattern μ is characterized by the period T^μ (or frequency $\nu^\mu = 1/T^\mu$) and the specific phases of spike ϕ_j^μ of the neurons $j = 1, .., N$. When multiple phase coded patterns are stored, the learned connections are simply the sum of the contributions from individual patterns, namely

$$J_{ij} = \sum_{\mu=1}^{P} \delta J_{ij}^\mu. \tag{5}$$

THE ORDER PARAMETER

To measure quantitatively the success of the retrieval, in analogy with the Hopfield model, we introduce a dynamical order parameter, which estimates the overlap between the network collective activity during the spontaneous dynamics and the stored spatiotemporal pattern. This quantity is maximal (equal to one) when collective activity is periodic and the ordering of spiking times coincide with that of the stored pattern, and is order $\simeq 1/\sqrt{N}$ when the spike timings are uncorrelated with the stored ones.

The order parameter is defined by

$$m^\mu = \max_{T^w} \frac{1}{\langle N_s \rangle} \langle |M(t, T^w)| \rangle, \tag{6}$$

with

$$M(t, T^w) = \sum_{\substack{j=1,\dots,N \\ t<t_j^*<t+T^w}} e^{-i2\pi t_j^*/T^w} e^{i2\pi t_j^\mu/T^\mu} \tag{7}$$

where t_j^* is the spike timing of neuron j during the spontaneous dynamics, T^w is a "probe" period of the collective spontaneous periodic dynamics, the average $\langle \cdots \rangle$ is done on the starting time t of the window, and $\langle N_s \rangle$ is the average number of spikes on a window of time T^w.

We then define the fluctuations of the order parameter by

$$\sigma(m^\mu)^2 = \max_{T^\mu} \frac{1}{\langle N_s \rangle^2} \left[\left\langle |M(t, T^w)|^2 \right\rangle - \langle |M(t, T^w)| \rangle^2 \right]. \tag{8}$$

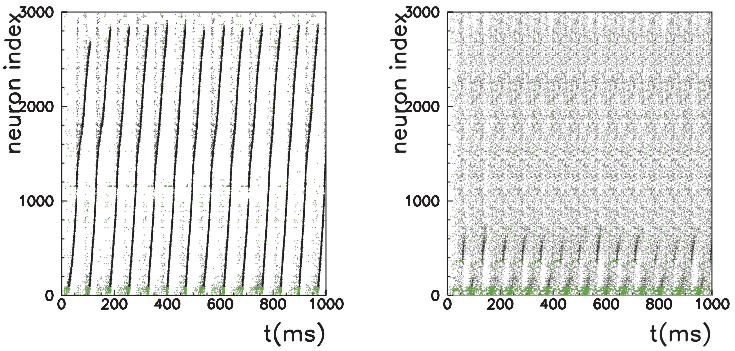


FIGURE 1. Regime of correct replay. Spontaneous dynamics without any stimulation in a noisy environment with $\theta_{th2} = 80$, $\theta_{th1} = 26$, $N = 3000$, $\nu^{\mu} = 3Hz$. Spikes are shown with units sorted on the vertical axes according to order of units in the first (a) or second (b) stored pattern. Units with threshold $\theta_{th1} = 26$ are shown in green, the other in black. Permanent replay of first pattern is observed, initiated by the noisy units (low threshold, more sensible to noise).

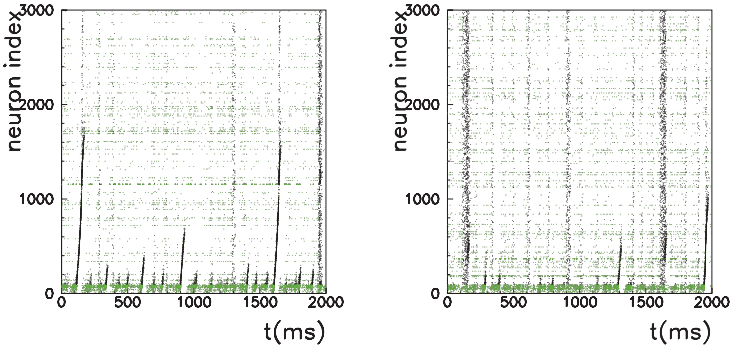


FIGURE 2. Critical regime. Spontaneous dynamics without any stimulation in a noisy environment with $\theta_{th2} = 90$, $\theta_{th1} = 26$, $N = 3000$, $\nu^{\mu} = 3Hz$. Spikes are shown with units sorted on the vertical axes according to order of units in the first (a) or second (b) stored pattern. Short transient replays of the two patterns are initiated, from time to time.

Note that, when the dynamics is such that the pattern is not replayed continuously, but there are short and incoherent segments of different patterns (as in Fig. 2), it is not possible to evaluate consistently the time scale at which the pattern is replayed. For this reason, we define the order parameter looking at the time window T^w which maximizes it. Of course, when the pattern is periodic, this coincides with the period of the pattern, but our definition works also when short replays are hidden in a non-periodic spike train, such as here and in many experimental situations.

The phase transition and the critical point

An important result of this paper is the study of the different regimes observed by changing the spiking thresholds of the units.

We explore the parameter space, finding that there's a transition between persistent replay regime and no-replay regime and at the transition point there's a critical behavior, in which short transient replays are initiated by noise and then fade away.

We perform numerical simulations of our network embedded in the noisy environment described previously, in absence of any external stimulation.

Figures 2 and 1 show the spontaneous dynamics of a network of $N = 3000$ units, $P = 2$ stored patterns with $\nu^\mu = \underline{3}$ Hz, and randomly chosen phases ϕ_i^μ, with noise given by Eq. (3) with $\tau_{noise} = 1$ ms, $J_{noise} = 0$, $\sigma(J_{noise}) = 5$, $N_1 = 200$ units with $\theta_{th1} = 26$ and the other $N - N_1 = 2800$ units with $\theta_{th2} = 80, 90$, respectively in Figs. 2 and 1.

As shown in many raster plot of in-vitro spontaneous dynamics with neuronal avalanches, there is often a small subset of units which has a higher spiking rate then the other units. These units with spontaneous higher spiking rate are modelled here as a subset of units with lower spiking threshold, therefore more sensible to the Poissonian noise which acts on the membrane potentials of all the units of the network. If some of these low-threshold units belongs to one of the stored pattern and have consecutive phases in this pattern, then it may happen that these units are able to initiate a collective replay of the pattern. To check this hypothesis, therefore, we consider that for each stored pattern there is a small subset of N_1/P units, with consecutive phases in the pattern, that have low spiking threshold θ_{th1}. While the others $N - N_1$ units will have threshold θ_{th2}.

To put in evidence the replay of different patterns, we show the raster plot of the network dynamics with different sorting on the vertical axes. In Fig. 2 the raster plot of the network dynamics is shown with a different sorting of units on the vertical axes. In Fig. 2a, the units are sorted according to increasing values of stored phases in pattern $\mu = 1$, while in Fig. 2b units are sorted according to $\mu = 2$. It can be seed that, from time to time, there is a short transient replay of one of the two the patterns. When pattern $\mu = 1$ is recalled, a short sorted sequence of spikes appears in Fig. 2a, while when pattern $\mu = 2$ is retrieved, a short sorted sequence of spikes appears in Fig. 2b.

At a lower value of the threshold θ_{th2}, namely $\theta_{th2} = 80$, the first pattern that is replayed (randomly chosen by the noise) lasts for very long times, seemingly permanently, as shown in Fig. 1a and 1b, where only pattern $\mu = 1$ is replayed.

In Fig. 3a we show the order parameter (circles) and its fluctuations (stars) for three values of θ_{th2}, namely $\theta_{th2} = 80$, 90, 100, with $\theta_{th1} = 26$ and $N = 3000$. The behavior for the other values of θ_{th1} (not shown), namely $\theta_{th1} = 22$,24, 28, and 30, is similar, but the value $\theta_{th1} = 26$ gives the highest values of the order parameter.

Then we fix T^w to the value that maximize the order parameter, and we study the order parameter m^μ and its fluctuations $\sigma(m^\mu)$ as a function of the spiking threshold θ_{th2}, with $\theta_{th1} = 26$ fixed, as shown in Fig. 3b. At low spiking threshold ($\theta_{th2}/N = 0.027$) the order parameter is high and fluctuations are low, indicating that, as shown in Fig. 1a, the noise is able to initiate a successful long-lasting replay of the stored pattern. At high threshold ($\theta_{th2}/N = 0.033$) both order parameter and its fluctuations are low. At the dynamic critical point ($\theta_{th2}/N = 0.03$) between the two regimes, the fluctuations of

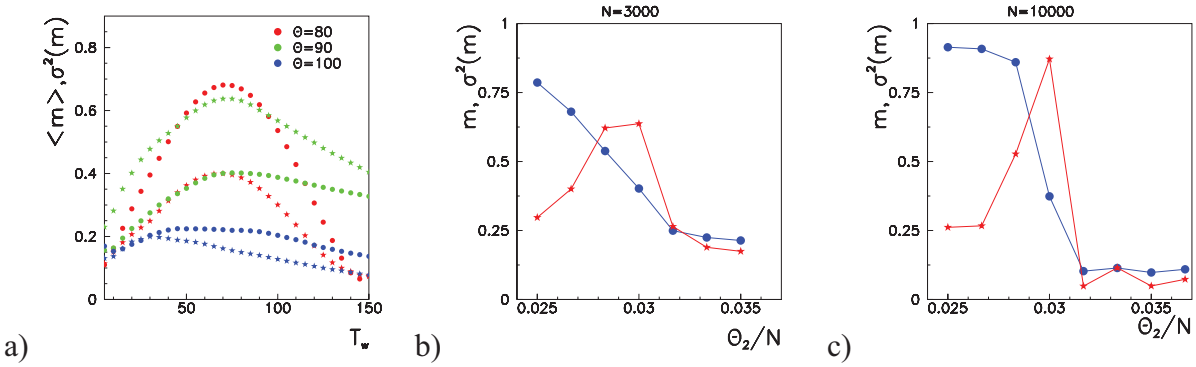


FIGURE 3. a) Order parameter (circles) and its fluctuations (stars) as a function of the chosen window T^w, with $\theta_{th1} = 26$ and for different θ_{th2}, for $N = 3000$. b) and c) Order parameter and its fluctuations as a function of θ_{th2}/N, for $N = 3000$ and $N = 10000$.

the order parameter are maximized, as expected in a phase transition.

We further characterize the critical behavior studying the behavior of order parameter and its fluctuations in a larger network with $N = 10000$. We performed numerical simulations of a network with $N = 10000$, and J_{ij} given by Eq. (4) and (5), with $P = 2$ stored patterns at $\nu^{\mu} = 3$ Hz, noise given by Eq. (3) with $tau_{noise} = 1$ ms, $\bar{J}_{noise} = 0$, $\sigma(J_{noise}) = 5$. Analogously to the case with $N = 3000$, where N_1/N was 0.0667, here we choose $N_1 = 667$ units with θ_{th1} and the other $N - N_1$ units with θ_{th2}. The behavior as a function of θ_{th2} has been studied, with θ_{th1}/N and θ_{th2}/N in the same range as in the case $N = 3000$. Indeed previous investigations (see Fig. 6d of [5]) has shown that the relevant parameter is the threshold scaled by the network size. The numerical simulations of the $N = 10000$ network confirm that at the transition between high order parameter regime and low order parameter regime, there's a critical point in which fluctuations of order parameter are maximized. The fluctuations are larger when the size of the network increases.

NEURAL AVALANCHES AT THE CRITICAL POINT

In order to characterize the noise-induced collective dynamics near the critical point, we study the inter-spikes-intervals statistics and the sizes and durations of avalanches of spikes.

As usual, an avalanche is defined by the activity that occurs over some contiguous set of bins that all have non-zero activity (i.e., at least one spike in each bin), but are preceded and succeeded by at least one bin of zero activity.

The distribution of inter-spike intervals among consecutive spikes of the network (ISI) is shown in Fig. 4 for spiking threshold $\theta_{th2} = 80, 90, 105$ and $N = 3000$.

For the three threshold values $\theta_{th2} = 80, 90, 105$ the average ISI is respectively $0.03, 0.1$ and $0.15ms$. A unique bin width, equal to $0.1ms$ i.e. the order of magnitude of the network ISI, is used to bin the spike trains and evaluates the avalanches. For each avalanche, we measure its duration T in ms, its size s defined as the total number of

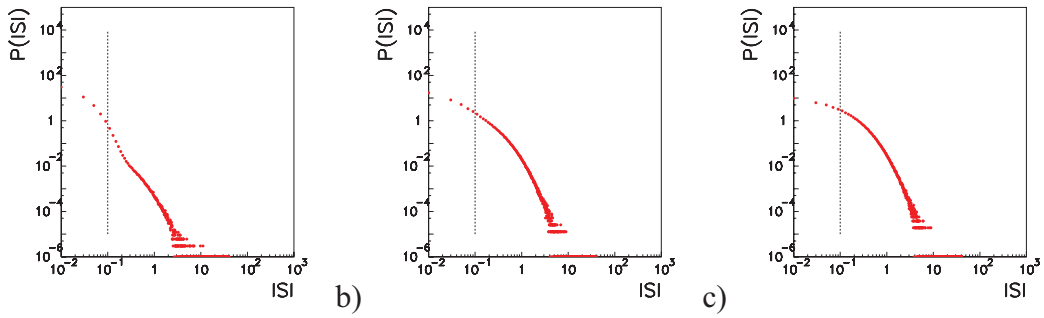


FIGURE 4. Network inter spike intervals at $\theta_{th2} = 80, 90, 105$ respectively in a,b,c. N=3000, all parameters are the same used in Figs. 1,2.

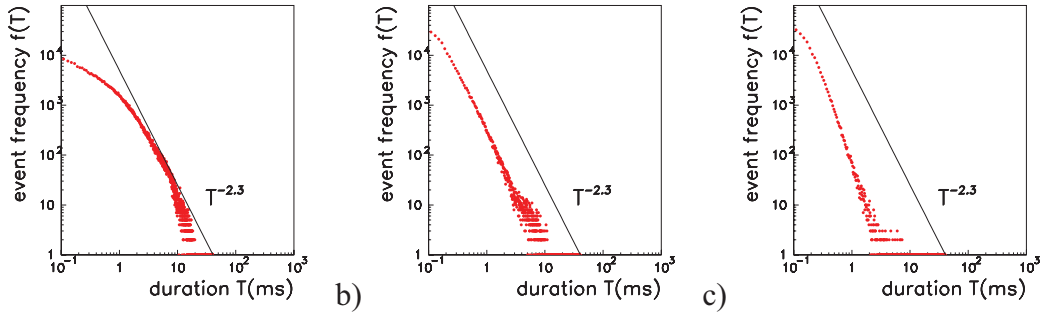


FIGURE 5. Distribution of avalanches duration at $\theta_{th2} = 80, 90, 105$ respectively in a,b,c. N=3000, all parameters are the same used in Figs. 1,2. For comparison the line $f(T) = T^{-2.3}$ is shown. The distribution approaches a power law at the critical point $\theta_{th2} = 90$.

spikes within the avalanche, and its shape defined as the temporal profile of the spikes of the avalanche.

Figure 5 shows the duration distribution for the three regimes, showing that at the critical point $\theta_{th2} = 90$ the duration distribution approaches a power law, with exponent close to $\alpha = 2.3$. Figure 6 shows the sizes distribution at the three different spiking thresholds. At the critical point (6.b) the size distribution is a power law, with critical exponent $\beta = 2$. Finally the Fig. 7 shows the sizes $s(T)$ of the avalanche of duration T, as a function of duration T. Again at the critical point the function approach a power law, with exponent $k = 1.3$, in agreement with results of [2].

Therefore the same critical value of the threshold which gives the maximization of the fluctuation of the order parameter gives also a power law avalanches distribution. This is in agreement with the picture discussed previously showing that at critical threshold there are transient reactivations of different stored patterns which last for different durations, and the reactivation may be as large as the full network or involve only a short number of units. These results suggest that the critical avalanches observed experimentally may be the manifestation of a system at the dynamical critical point of a phase transition between a permanent retrieval regime with stable dynamical attractors, and a non-retrieval regime.

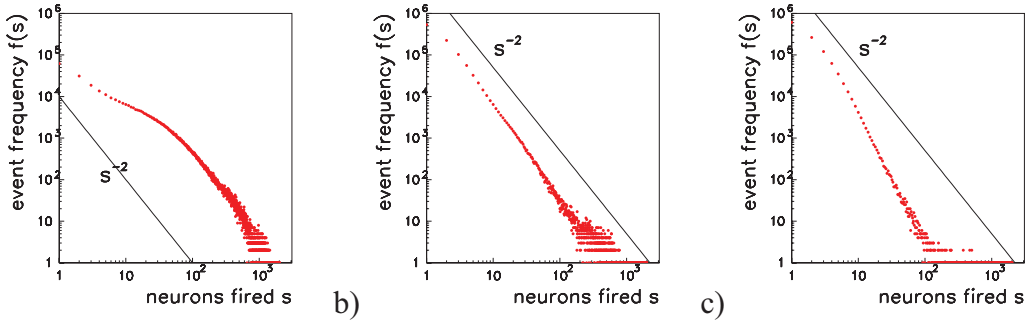


FIGURE 6. Distribution of avalanches sizes at $\theta_{th2} = 80, 90, 105$ respectively in a,b,c. N=3000, all parameters are the same used in Figs. 1,2. For comparison the line $f(s) = s^{-2}$ is shown. The distribution approaches a power law at the critical point $\theta_{th2} = 90$.

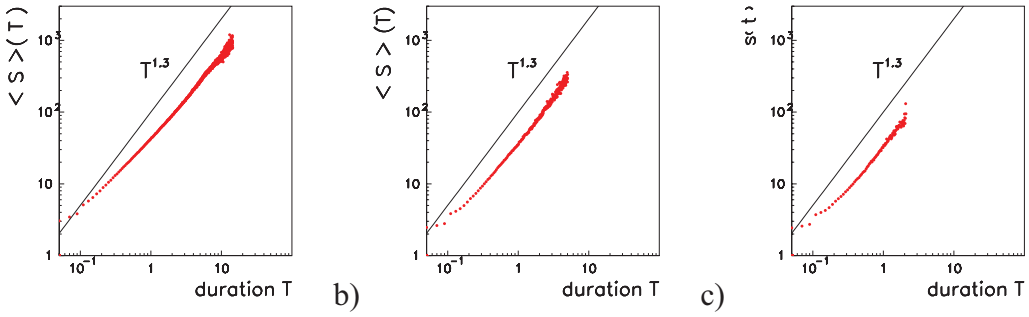


FIGURE 7. Avalanche size as a function of duration,at $\theta_{th2} = 80, 90, 105$ respectively in a,b,c. N=3000, all parameters are the same used in Figs. 1,2. For comparison the line $< s > (T) = T^{1.3}$ is shown.

REFERENCES

1. D. R. Chialvo, *Nature Phys.* **6**, 744–750 (2010).
2. N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. Lee DeVille, K. A. Dahmen, J. M. Beggs, and T. C. Butler, *Phys. Rev. Lett.* **108**, 208102 (2012)
3. D. Plenz, and T. C. Thiagarajan, *Trends Neurosci.* **30**(3), 101–110 (2007).
4. E. Tagliazucchi, P. Balenzuela, D. Fraiman, and D. R. Chialvo., *Front. Physiol.* **3**, 15 (2012).
5. S. Scarpetta, and F. Giacco, *J. Comput. Neurosci.*, 1–18 (2012).
6. S. Scarpetta, F. Giacco, and A. De Candia, *Europhys. Lett.* **95**(2), 28006 (2011).
7. S. Scarpetta, A. de Candia, and F. Giacco, *Front. Synap. Neurosci.* **2**, 1–12 (2010).
8. W. Gerstner, R. Ritz, and J.L. van Hemmen, *Biol. Cybern.* **69** 503–515 (1993)
9. W. Gerstner and W. Kistler, *Spiking neuron models: single neurons, populations, plasticity.* Cambridge University Press, NY, 2002.
10. S. Scarpetta, L. Zhaoping, and J. Hertz, *Spike-Timing-Dependent Learning for Oscillatory Networks.* Advances in Neural Information Processing Systems 13, MIT Press, 2001.
11. S. Scarpetta, L. Zhaoping, and J. Hertz, *Neural Comput.* **14**, 2371–96 (2002).
12. M. Yoshioka, S. Scarpetta, and M. Marinaro, *Phys. Rev. E* **75** 051917 (2007)
13. F. Giacco, and S. Scarpetta, *Front. Artif. Intel. & Appl.* **234**, 265-274 (2011).
14. H. Abarbanel, R. Huerta, and M. I. Rabinovich, *Proc. Natl. Acad. Sci. USA* **99**, 10132–10137 (2002).
15. G. Q. Bi, and M. M. Poo, *J. Neurosci.* **18**, 10464–10472 (1998).

Observing scale-invariance in non-critical dynamical systems

C. Gros and D. Marković

Institute for Theoretical Physics, Goethe University Frankfurt, 60438 Frankfurt a.M., Germany.

Abstract. Recent observation for scale invariant neural avalanches in the brain have been discussed in details in the scientific literature. We point out, that these results do not necessarily imply that the properties of the underlying neural dynamics are also scale invariant. The reason for this discrepancy lies in the fact that the sampling statistics of observations and experiments is generically biased by the size of the basins of attraction of the processes to be studied. One has hence to precisely define what one means with statements like 'the brain is critical'. We recapitulate the notion of criticality, as originally introduced in statistical physics for second order phase transitions, turning then to the discussion of critical dynamical systems. We elucidate in detail the difference between a 'critical system', viz a system on the verge of a phase transition, and a 'critical state', viz state with scale-invariant correlations, stressing the fact that the notion of universality is linked to critical states. We then discuss rigorous results for two classes of critical dynamical systems, the Kauffman net and a vertex routing model, which both have non-critical states. However, an external observer that samples randomly the phase space of these two critical models, would find scale invariance. We denote this phenomenon as 'observational criticality' and discuss its relevance for the response properties of critical dynamical systems.

Keywords: criticality; critical states; scale invariance; observational criticality; dynamical systems.
PACS: 64.60.av, 64.60.aq, 64.60.De, 05.45.-a, 05.65.+b, 05.70.Jk

INTRODUCTION

The notion of criticality stems from statistical mechanics and is fundamentally related to the deeply routed concept of universality [1, 2]. As critical equilibrium systems show scale invariance it is natural to assume that the same would hold for critical non-equilibrium systems [3, 4]. The situation is however substantially more complex for classical dynamical systems far from equilibrium and the subject of our deliberations. The discussion will revolve around three central concepts.

CRITICAL SYSTEM A system is denoted critical when being located right on the transition point of a second order phase transition [5, 6].

CRITICAL STATE The state of a thermodynamic or dynamical system is denoted critical when exhibiting scale invariance [5, 7]. Critical thermodynamic systems dispose always of a critical state, critical dynamical systems not necessarily.

OBSERVATIONAL CRITICALITY The experimental observation of a dynamical system generically involves a stochastic sampling of its phase space. Scale invariance may be observed for a critical dynamical system which does not dispose of a critical state [8, 9, 10].

This dichotomy is caused by the difference between mean and typical properties. It turns out that for critical dynamical systems the scaling behavior of the typical

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 44-53 (2013); doi: 10.1063/1.4776500

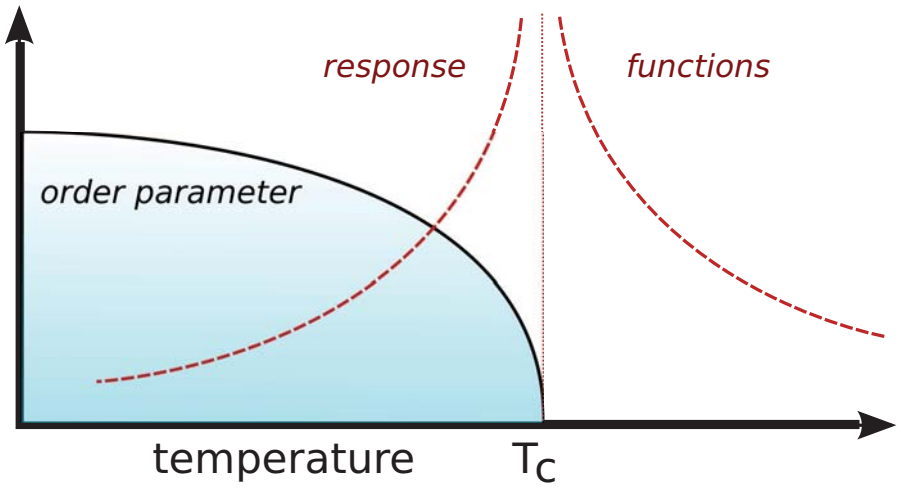


FIGURE 1. Illustration of a second order phase transition. The low-temperature phase is characterized by an order parameter which drops continuously to zero at the critical temperature T_c. The system becomes increasingly susceptible to perturbations coupling to the order parameter close to the transition point, the respective response functions diverge algebraically.

attractor may differ qualitatively from the scaling of the mean attractor, as defined by randomly sampling a phase space.

We will start by recapitulating the central notions of the theory of critical thermodynamic systems, stressing the fact that the scale invariance, which is observed in this case, is deeply intertwined with the concept of universality. We will then discuss two examples of critical dynamical systems for which the scaling behavior at criticality is, at least in parts, exactly known.

CRITICALITY IN STATISTICAL PHYSICS

In statistical physics a phase transition is termed a second order phase transition when the ordering process starts continuously at the critical temperature T_c, when lowering the temperature T of the system, compare Fig. 1. Otherwise, when the low-temperature state discontinuously appears, one speaks of a transition of first order. The theory of critical phenomena deals with second order phase transitions [11].

Scaling towards criticality. For a second order phase transition there are precursors of the impending transitions, which can be measured experimentally using appropriate probes. For example, applying an external magnetic field to a ferromagnetic system will lead to a strong response, in terms of the induced magnetization, close to the transition. In general this response will diverge as

$$\sim \frac{1}{|T - T_c|^{\gamma}} , \tag{1}$$

where $\gamma > 0$ is the critical exponent [1]. Power-laws like Eq. (1) are denoted scale invariant, as they do not change their functional form when rescaling the argument via $|T - T_c| \to c|T - T_c|$, where $c > 0$ is an arbitrary scaling factor.

Critical state. At criticality, $T = T_c$, the thermodynamic state is very special, its correlation function being scale invariant both in the spatial and the temporal domain. For a magnetic system, with moments $S(\mathbf{x})$ at $\mathbf{x}$, the equal time correlation function

$$D(\mathbf{r}) \equiv D(\mathbf{x}-\mathbf{y}) = \left\langle S(\mathbf{x})S(\mathbf{y})\right\rangle - \langle S\rangle^2$$

obeys the scaling relations

$$D(\mathbf{r}) \propto \begin{cases} \mathrm{e}^{-r/\xi} & T \neq T_c \\ r^{-\alpha} & T = T_c \end{cases}, \qquad \xi \propto \frac{1}{|T-T_c|^z}, \tag{2}$$

with ξ being termed the correlation length and z the critical dynamical exponent [12, 13].

Absence of microscopic length scales. The scaling of the correlation function (2) is very intriguing, since it implies that all microscopic scales (length, time, energy, *etc.*) become irrelevant at criticality. As an example consider the Schrödinger equation

$$i\hbar\frac{\partial\Psi(t,\mathbf{r})}{\partial t} = -E_R\left(a_0^2\Delta + \frac{2a_0}{|\mathbf{r}|}\right)\Psi(t,\mathbf{r}) \qquad E_R = \frac{me^4}{2\hbar^2}, \qquad a_0 = \frac{\hbar^2}{me^2}$$

which determines the properties of most matter we know. The Schrödinger equation contains two scales, the Rydberg energy $E_R = 13.6\,\mathrm{eV}$, which determines the energy level spacing, and the Bohr radius $a_0 = 0.53\text{Å}$, which determines the extension of the atoms. Any Hamiltonian known is characterized by corresponding scales, but these become irrelevant at criticality and do not determine the magnitude of the critical exponents.

Universality. The symmetry of the high-temperature phase is broken at a second order phase transition. For example, in a magnetic systems with classical moments, these magnetic moments point in any direction for $T > T_c$, the symmetry of the high temperature phase is O(3), the symmetry group of the sphere. In the low-temperature phase the magnetic moments point however predominantly into a specific direction, breaking spontaneously the O(3) symmetry of the order parameter.

A central result of the modern theory of phase transitions is now that the critical exponents are determined solely by two factors: the dimensionality of the system and the symmetry of the order parameter. This relation is termed 'universality' as it allows to classify second order phase transitions into a relatively small number of distinct classes [5, 1, 2]. Results obtained using a given microscopic model are valid for all models within the same universality class. Universality is the core to our understanding of second-order phase transition, the scale invariance of the critical state being a manifestation of it.

[1] Critical exponents may differ for $T < T_c$ and $T > T_c$

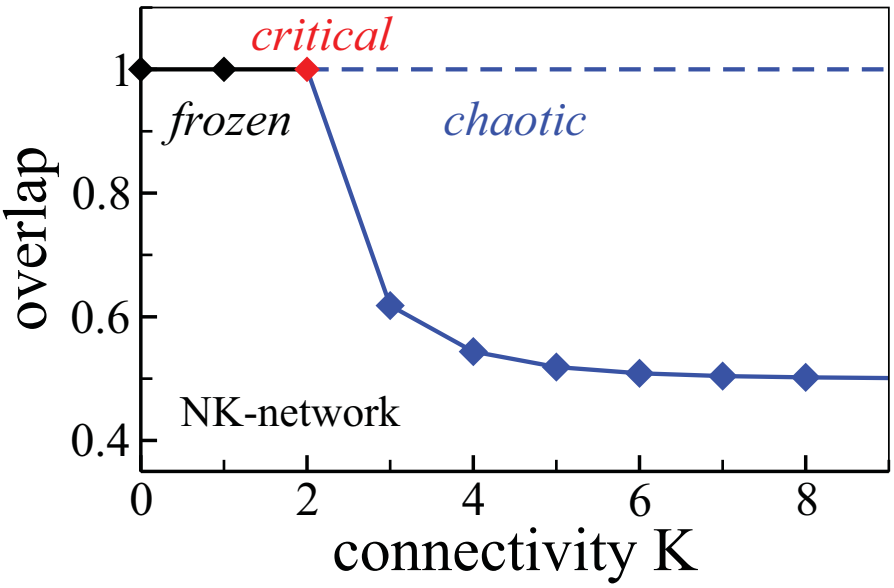


FIGURE 2. The evolution of the order parameter of the NK-network. Shown is the overlap, as given by Eq. (4), in the long-time limit, of two initially close trajectories. In the frozen state the overlap becomes maximal, since the two trajectories flow into the same attractor. In the chaotic state two initially close states diverge, the Lyapunov exponent is positive.

BOOLEAN NETWORKS

In equilibrium thermodynamics one studies systems in the thermodynamic limit where the number of components N becomes infinitely large, $N \to \infty$. Phase transitions hence take place, in statistical physics, in systems made-up of many similar units. We consider here an equivalent setting for non-equilibrium phase transitions. A dynamical system can be described as a set of N differential equations,

$$\frac{d}{dt}x_i(t) \;=\; f_i(x_1,..,x_N;\eta), \qquad\qquad i=1,..,N\,, \tag{3}$$

where f_i determines the time evolution of the dynamical variables $x_i(t)$ which are related to each of the system's elements. Here η denotes a generic control parameter. Random Boolean networks are defined by three specifications [14].

BOOLEAN VARIABLES The variables $x_i \in \{0,1\}$ are Boolean and the time $t = 0,1,2,\ldots$ discrete.

RANDOM COUPLING FUNCTIONS The coupling functions are Boolean, $f_i \in \{0,1\}$, and selected randomly.

CONNECTIVITY The coupling functions are determined by only a subset of K randomly selected controlling elements and not by all N Boolean variables. Hence the term 'Boolean network'. The control parameter K is denoted connectivity.

Random Boolean networks are also termed NK- or Kauffman nets [15]. They show a phase transition for connectivity $K = 2$, being regular for $Z < 2$ and chaotic for $Z > 2$

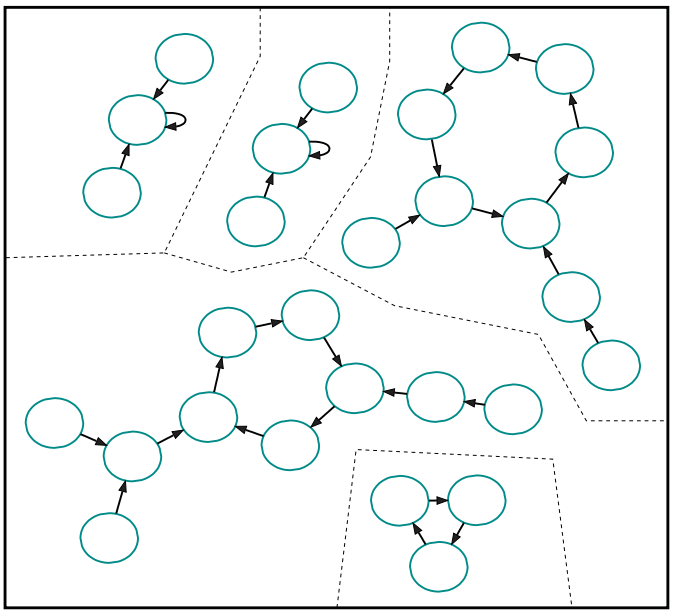


FIGURE 3. There are many cyclic attractors in the phase space of boolean networks and routing models. Each attractor comes with its distinct basin of attraction, which is made up of the cycle itself together with all points of phase space flowing into the attractor.

[16].

$K<2$	$K=2$	$K>2$
frozen	critical	chaotic

The order parameter is given by the overlap

$$\lim_{t\to\infty}\Big(1-||\mathbf{y}-\mathbf{x}||\Big) \tag{4}$$

of two initially close trajectories $\mathbf{x}(t)$ and $\mathbf{y}(t)$, where $||..||$ denotes the Manhattan distance, that is, the sum of the absolute differences of coordinates of $\mathbf{x}$ and $\mathbf{y}$. In the frozen phase the overlap is maximal, since close-by trajectories will end up in the same attractor, see Fig. 2. The dynamics becomes chaotic however for $Z>2$, and two trajectories diverge, with their mutual overlap decreasing.

Attractors and cycles. The time evolution of any dynamical network with finite phase space, which is $\Omega=2^N$ for the NK net, is determined by the number and the size of its cyclic attractors. The Kauffman net is critical for $Z=2$ and one may ask the question to which extend this criticality is reflected in the statistics of its attractors.

Any attractor comes with a respective basin of attraction, as illustrated in Fig. 3, defined as the set of all points in phase space flowing into the attractor. In the ordered phase a small number of attractors with large basins of attraction dominates phase space and the dynamics is hence very stable, nearby trajectories converge. In the chaotic phase, for $Z>2$, the number of attractors is however very large and the size of their respective

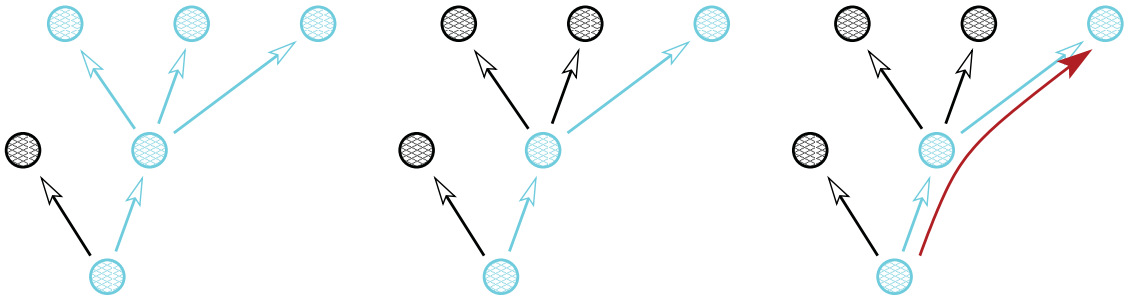


FIGURE 4. Illustration of information spreading on networks. When information spreads diffusively (left), it may be passed on to any number of subsequent vertices. When information is conserved (center), the information can be considered as a package which can be passed on only to a single downstream site. Alternatively one can consider information routing (right), where an incoming package is routed to an outgoing link.

basins of attraction correspondingly smaller. Nearby trajectories tend to diverge, being attracted by different cycles.

Finite-size scaling. To calculate the properties of a dynamical or thermodynamic system directly in the thermodynamic limit is most of the time difficult or impossible. Alternatively one can evaluate the quantity of interest for finite systems size N and then extrapolate to large system size, a procedure denoted finite-size scaling. For scale invariant states, like the critical thermodynamic state, finite size scaling involves power-laws. The reason is that there are no length scales at criticality in statistical physics and power-laws are the only scale invariant relations. Conversely we expect finite-size scaling to be algebraic whenever the underlying state is critical, viz scale invariant.

Initial numerical calculation for the $Z = 2$ Kauffman net did indeed find that the number of attractors, scales polynomial, like $\sqrt{N}$ [15]. The same scaling relation was also found for the mean cycle length. However it has recently been show rigorously, that the number of attractors actually increases faster than any power of N, viz super-polynomial [16, 17]. The intrinsic state of the critical $Z = 2$ Kauffman net is hence not scale invariant.

Observational scale invariance. The phase space $\Omega = 2^N$ of the NK network increases exponentially with system size N. Numerical studies have hence to resort to appropriate statistical sampling of phase space. Actually, this is also what an experimental observer would do when examining a dynamical system at a random starting time. It may now be the case that a relatively small number of attractors dominate phase space and the results of a statistical sampling procedure, see Fig. 3.

In order to illustrate this scenario we discuss now a fictional example. Let's assume that there are big attractors of the order of $\sqrt{N}$, each having on the average a basin of attraction of the size

$$\sim \frac{\Omega}{\sqrt{N}} = \frac{2^N}{\sqrt{N}} .$$

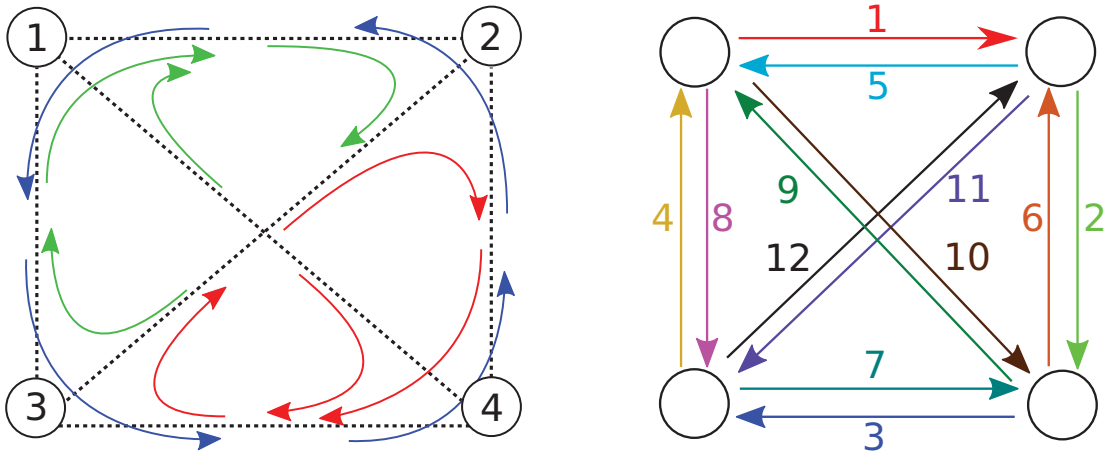


FIGURE 5. Illustration of of a $N = 4$ sites vertex routing model which has (left) three cyclic attractors. Note that more than one cycle can pass through any given vertex, as the phase space (right) is made up by the collection of the $N(N-1) = 12$ directed links.

There could be in addition a very large number of point attractors, each having a basin of attraction of size one. For example the number of point attractors could scale super-polynomial like

$$\sim 2^{\sqrt{N}} .$$

In this case their combined relative contribution

$$\sim \frac{2^{\sqrt{N}}}{\Omega} = \frac{2^{\sqrt{N}}}{2^N} = \frac{1}{2^{\sqrt{N}}}$$

to phase space would still vanish in the thermodynamic limit $N \to \infty$. This is what happens for the $Z = 2$ Kauffman net. The typical attractor is very small and not seen by a stochastic sampling procedure. A relatively small number of big attractors with large basins of attraction dominate phase space and determine the statistics as sampled by an external observer.

VERTEX ROUTING MODELS

Criticality and conservation laws are intrinsically related. A branching process is critical, to give an example, when the average number of offspring is equal to the number of parents, that is, when average activity remains constant. It is hence possible to construct critical dynamical systems when incorporating a conservation of activity levels. An example for this procedure are vertex routing models [18].

Information can spread diffusively or via routing processes, see Fig. 4. For the later case one considers information packages transmitted at every vertex via randomly selected routing tables. The phase space is hence given by the collection of directed links, the phase space volume $\Omega = N(N-1)$ scales algebraically. More than one cycle can hence pass through a given vertex. The number of cycles passing through a given model

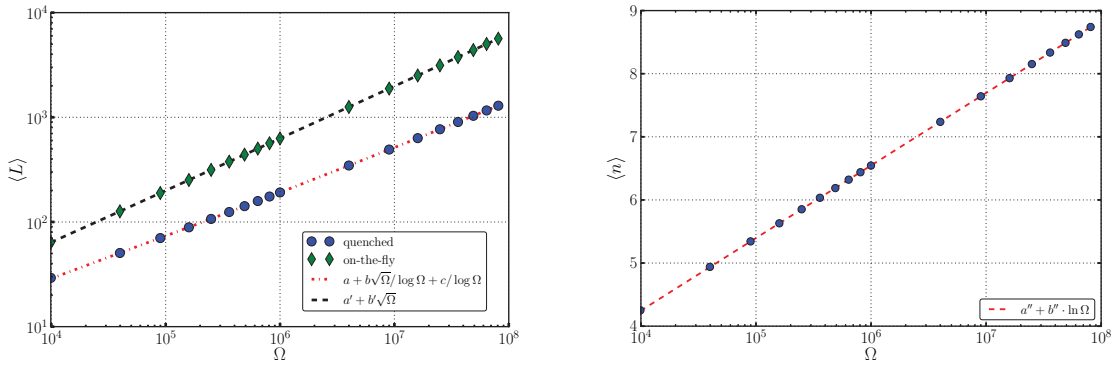


FIGURE 6. Exact results for the vertex routing model. The mean cycle length (left) for both quenched and on-the-fly dynamics and the the mean cycle number (right), which can be evaluated only for quenched dynamics.

can be viewed as a measure for information centrality which has a non-trivial distribution in the thermodynamic limit [18].

Exact solution. The routing dynamics can be mapped to a random walk in configuration space, the collection of directed links, and solved exactly [19, 14]. The number $\langle C_L \rangle(N)$ of cycles of length L is given by

$$\langle C_L \rangle(N) \;=\; \frac{N((N-1)^2)!}{L(N-1)^{2L-1}((N-1)^2+1-L)!}\,, \tag{5}$$

for fully connected graphs with N vertices. In addition to the exact expression (5) for the intrinsic cycle length distribution of the routing model, one can also derive the distribution of cycle length an observer would find when randomly sampling phase space. In this case the probability to find a given cycle of length L is weighted by the size of its basin of attraction. The resulting cycle length distribution is

$$\langle C_L \rangle(N) \;\propto\; \sum_{t=L}^{L_{max}} \frac{((N-1)^2)!}{(N-1)^{2t}((N-1)^2+1-t)!}\,. \tag{6}$$

Algorithmically the difference between the expressions (5) and (6) is equivalent to quenched deterministic and on-the-fly stochastic dynamics. Quenched dynamics is present when the routing tables are selected once at the start and then kept fixed, whereas for on-the-fly dynamics one randomly generates an entrance to a routing table 'on the fly', viz only when needed.

Scaling of the vertex routing model. One can evaluate the exact expressions (5) and (6) for very large system size N, the results are shown in Fig. 6, respectively for the average cycle length $\langle L \rangle$ and the overall number of cycles. Only relative quantities can be evaluated with on-the-fly dynamics and hence $\langle L \rangle$ but not the total number of cycles present. The results are given in Table 1, where we have included also results for a modified vertex routing model, a Markovian variant. On-the-fly routing results in

TABLE 1. The scaling behavior of the vertex routing model (first row) and of a modified routing model with nor routing memory (second row). Corrections $\sim \log(N)$ are present for quenched dynamics, viz for the intrinsic model behavior. An observer would however obey power-law scaling, as given by the on-the-fly dynamics, which can evaluate only relative quantities (and not the overall number of cycles).

		quenched	on the fly
vertex routing	number of cycles	$\log(N)$	–
	mean cycle length	$N/\log(N)$	N
markovian model	number of cycles	$\log(N)$	–
	mean cycle length	$\sqrt{N}/\log(N)$	$\sqrt{N}$

power-law scaling for the average cycle length, in contrast to the exact properties of the respective model, which contains logarithmic corrections.

DISCUSSION

When probing a dynamical or thermodynamical system, like the brain or a magnet, one needs to perturb the system and measure the resulting response. The probing protocol may be considered unbiased when the phase space is probed homogeneously. If the dynamical system being probed contains attractors, or attractor relics [20, 21], these will dominate the statistics of the response. It may now happen that properties of the attractors, like the cycle length for the case of cyclic attractors, have a highly non-trivial statistics in the sense, that the characterizing properties of the typical attractor differ qualitatively from the average behavior probed by random sampling phase space. In this the intrinsic or typical properties of the system differ from the one an observer would find when sampling phase space randomly.

We have argued in this study, that this situation does indeed occur for critical dynamical systems, at least for the classes of critical systems for which exact results are known, Boolean networks and vertex routing models. We believe that further investigation into this question is warranted for additional classes of critical dynamical systems, in order to examine the question whether power-law scaling is independent, or conditional, on universality in critical dynamical systems. This is an open issue. Here we found that the intrinsic state of two critical dynamical systems is not scale invariant, a property typically associated with universality in thermodynamics, but experimentally probing the system stochastically would result in power-law scaling.

ACKNOWLEDGMENTS

We thank the German Science Foundation (DFG) for financial support.

REFERENCES

1. L. Kadanoff, *Physica A* **163**, 1–14 (1990).
2. H. Stanley, *Rev. Mod. Phys.* **71**, 358–366 (1999).
3. H. Hinrichsen, *Adv. in Phys.* **49**, 815–958 (2000).
4. S. Lübeck, *Int. J. Mod. Phys. B* **18**, 3977–4118 (2004).
5. D. Stauffer, M. Ferer, and M. Wortis, *Phys. Rev. Lett.* **29**, 345–349 (1972).
6. R. Balescu, *NASA STI/Recon Technical Report A* **76**, 32809 (1975).
7. L. Kadanoff, S. Nagel, L. Wu, and S. Zhou, *Phys. Rev. A* **39**, 6524 (1989).
8. D. Sornette, *Int. J. Mod. Phys. C* **13**, 133–136 (2002).
9. M. Newman, *Contemp. Phys.* **46**, 323–351 (2005).
10. D. Sornette, *arXiv preprint arXiv:0707.2194* (2007).
11. J. Yeomans, *Statistical mechanics of phase transitions*, Oxford University Press, USA, 1992.
12. F. Wang, and C. Hu, *Phys. Rev. E* **56**, 2310 (1997).
13. P. Hohenberg, and B. Halperin, *Rev. Mod. Phys.* **49**, 435 (1977).
14. C. Gros, *Complex and adaptive dynamical systems: A primer*, Springer Verlag, 2008.
15. S. Kauffman, *J. Theor. Biol.* **22**, 437–467 (1969).
16. J. Socolar, and S. Kauffman, *Phys. Rev. Lett.* **90**, 68702 (2003).
17. T. Mihaljev, and B. Drossel, *Phys. Rev. E* **74**, 046101 (2006).
18. D. Markovic, and C. Gros, *New J. Phys.* **11**, 073002 (2009).
19. M. Kruskal, *The American Mathematical Monthly* **61**, 392–397 (1954).
20. C. Gros, *New J. Phys.* **9**, 109 (2007).
21. C. Gros, *Cogn. Comput.* **1**, 77–90 (2009).

Interplay activity–connectivity: dynamics in patterned neuronal cultures

E. Tibau, Ch. Bendiksen, S. Teller, N. Amigó and J. Soriano

Departament d'ECM, Facultat de Física, Universitat de Barcelona, Martí i Franqués 1, E-08028 Barcelona, Spain.

Abstract. The ability of a neuronal tissue to efficiently process and transmit information depends on both the intrinsic dynamical properties of the neurons and the connectivity between them. One of the few experimental systems where one can vary the connectivity of a neuronal network in a control manner are neuronal cultures. Here we show that, by combining neuronal cultures with different pattering techniques, we can control and dictate the connectivity of neuronal networks. The emerging cultures are characterized by a rich spontaneous activity, but with some dynamical traits that can be ascribed to the underlying, engineered wiring architecture. Simple patterned cultures can be obtained by plating neurons onto predefined topographical molds, which guide neurons and connections through complex paths. In contrast to homogeneous cultures, characterized by an on/off behavior where all neurons fire in a short time window, patterned cultures show more complex spatio–temporal dynamics, and with varying propagation paths and velocities. Patterned cultures provide a valuable tool to understand not only the interplay activity–connectivity, but also aspects such as the emergence and maintenance of spontaneous activity, synchronization, or the presence of specific dynamic motifs.

Keywords: engineered neuronal cultures; spontaneous activity; wave propagation; models of neuronal activity; percolation.
PACS: 87.85.Wc, 87.18.Sn, 87.19.lh, 87.19.lj, 87.19.ll, 87.19.lq, 87.19.lv, 87.19.lw, 64.60.ah

INTRODUCTION

Any nervous system, from the smallest neuronal network of a worm to the sophisticated human brain, consists of an intricate mesh of connections that constitute the neuronal circuitry. This architecture, together with the processing units that are the neurons, define the computational capabilities of the neuronal network. However, the whole picture of the relationship between connectivity, neuronal activity and function remains elusive, and its characterization is one of the major challenges of modern neuroscience.

Uncovering the complexity of a neuronal network, and ultimately the human brain, is obviously a daunting task. The exploration of the brain at a macroscopic scale (top-down approach) by non–invasive activity–measuring techniques (such as fMRI, EEG or MEG) in combination with graph–theoretical analyses [1] is shedding light on diverse brain functions as well as the interplay between brain areas. However, despite the clinical relevance of macroscopic approaches and their prominent role in cognitive neuroscience [2], the sheer size of the brain, the limited spatial resolution of the imaging techniques, and the simultaneous occurrence of several processes, hinders the identification of relevant neuronal network processes. The emergence and maintenance of complex brain rhythms [3] as well as the neuron–level mechanisms behind electrical or optical stimulation [4] are some pivotal aspects that are not properly addressed by macroscopic

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 54-63 (2013); doi: 10.1063/1.4776501

approaches.

At the other extreme, microscopic (bottom–up) approaches are based on the detailed description of connections, brain cells, and biochemical processes. One of the most prominent examples is the Blue Brain Project [5]. However, microscopic approaches may become unaccessible for their cumbersomeness and over–detail [6], and often fail at providing a clear picture of the emergence of collective phenomena.

Given the above difficulties, mesoscopic approaches are gradually catching the attention of the neuroscience community [7] as an intermediate exploratory platform. The goal is to identify basic processes and dynamical aspects that emerge from networks constituted by hundreds to thousands of neurons. The mesoscopic approach is proving successful to understand the emergence of function from connectivity and activity data. In the worm C. *elegans*, for instance, although its precise wiring diagram was unraveled about 25 years ago, it was not until mesoscopic tools were introduced —with the prominent support of network theory tools [8, 9]— that a clear picture of animal's function started to take shape [10, 11].

C. *elegans* is one of the most important mesoscopic model systems in neuroscience and a valuable tool to understand the interplay connectivity–activity–function [10, 11]. However, its fixed wiring architecture can be viewed as a drawback that limits the exploration of other network configurations and scenarios. By engineering a living neuronal network one could shape the connectivity of the network and ascribe the emerging neuronal dynamics to the underlying architecture. Such a powerful concept forms the basis of patterned (or engineered) neuronal cultures.

Neuronal cultures

Neuronal cultures in general [12, 13] constitute one of the most attractive experimental tools for the neuroscience and physics communities alike. Cultures are typically prepared by isolating neurons from specific neuronal tissues (such as the retina, spinal cord, or brain cortex) and plating them in biocompatible substrates (Fig. 1a). With adequate culturing conditions, neurons quickly connect to one another to form a new, spontaneously active neuronal network within a week [13, 14, 15]. Spontaneous activity in cultures is characterized by bursts of neuronal activity combined with quiescent inter–burst intervals. Several factors determine the structure and frequency of occurrence of these network bursts, including the level of maturation of the network circuitry, neuronal excitability, and the balance between excitation and inhibition [14, 16].

The activity of hundreds to thousands of cells *in vitro* can be simultaneously monitored using calcium fluorescence imaging techniques (Fig. 1b) [17, 18]. A major drawback of this technique, however, is that the typical frame rate during acquisition is slower than the cell's firing dynamics. Furthermore, the poor signal–to–noise ratio is such to make hard the detection of elementary firing events. On the contrary, other techniques such as multielectrode arrays (MEAs, Fig. 1c) [13, 19, 16] allow to discriminate single spikes, but are restricted to smaller populations of neurons. Additionally, the source of the spiking events can be often difficult to resolve since neurons are, in general, not located at the recording sites. Calcium imaging, despite its limitations, is a fast evolving

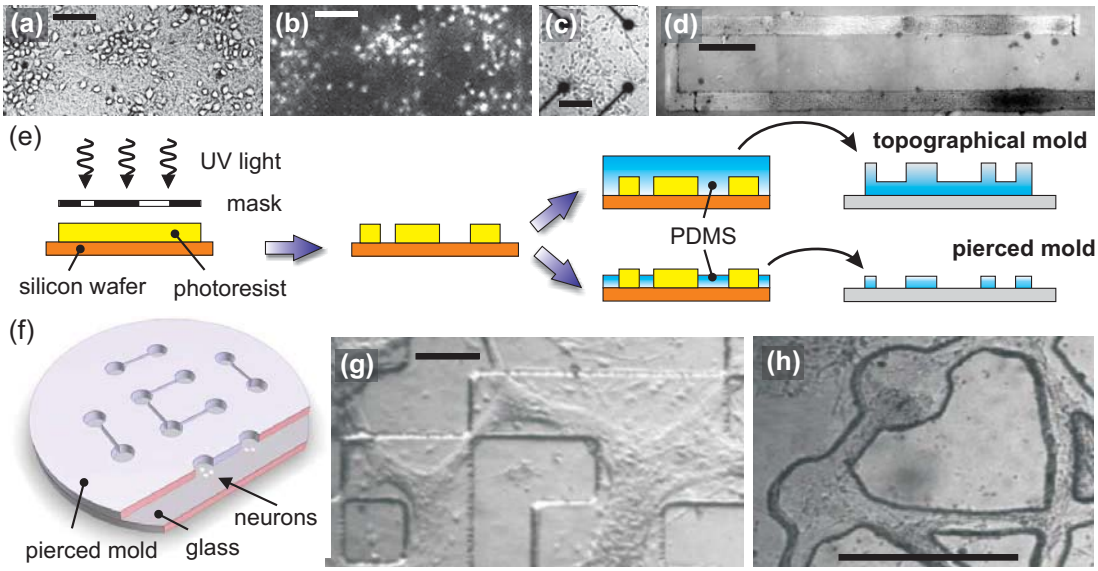


FIGURE 1. Neuronal cultures. (a)–(b) Neurons on a glass cover slip, showing a bright field (left) and calcium fluorescence (right) images. Circular objects in (a) and bright spots in (b) are neurons. (c) Culture on a multielectrode arrays system. Black circles are electrodes. (d) One–dimensional neuronal culture on a chemically patterned line 170 μm wide and 12 mm long. (e) Sketch of the preparation of PDMS patterned molds through photo–lithography. (f) Confinement of neurons using a pierced PDMS mold. (g)–(h) Patterned networks growing at the top (left) and bottom (right) parts of a PDMS topographical mold. For the latter, the valleys of the mold correspond to circular cavities and channels. Scale bars are 100 μm in all images except (d), that is 500 μm.

technique that can monitor a large population of neurons, both *in vitro* and *in vivo*.

Neuronal cultures, for their accessibility and versatility, provide a platform for studying both fundamental features and complexity in neuronal networks. These studies include, among others, self–organization and criticality [20, 21], development [15, 19, 16], connectivity [15, 22], and patterns of activity [14, 16]. They also constitute a pivotal tool to understand the influence of noise in neuronal circuits [23] and the emergence of synchronization [24, 25]. Neuronal cultures are also perfectly suited platforms to perturb neuronal networks in a control manner. Several strategies have been devised in the last years to stimulate neuronal networks *in vitro*, either locally (at a single neuron level) or globally (the entire network). Electric [26] and magnetic [27] stimulation have been for several years the standard protocols to act on neuronal networks, until the relatively recent advent of optical tools through optogenetics [28].

In cultures, the process of isolating and plating the neurons erases any information on the wiring architecture of the native tissue. Therefore, neurons in culture typically connect to their closest neighbors, giving rise to networks whose connectome and dynamics may dramatically differ from the native ones. This aspect is one of the major criticisms to neuronal cultures. In studies with brain slices [29], for instance, the connectivity is preserved along the plane of the slice, making them more adequate for investigations in which the circuitry is crucial, such as in the understanding of rhythmic activity in the visual cortex [29].

The relative simplicity of the connectivity diagram in a neuronal culture can be viewed, however, as a positive aspect. First, the connectivity of the culture can be modified through patterning, providing a very valuable tool to study the interplay between activity and connectivity. Second, cultures serve as a unique laboratory in the quest for identifying general mechanisms governing the dynamics in neuronal tissues, such as synchronization or the emergence of robust activity patterns. And third, the process of growth, maturation, and stabilization of a neuronal culture is related to concepts as important as self–organization or criticality, of pivotal importance in Physics.

In the present work we illustrate the potential of neuronal cultures to address diverse problems. Specifically, we show how engineered neuronal cultures become excellent tools to explore the subtle interplay between activity and connectivity, a paradigm of neuroscience whose comprehension is just starting to be uncovered.

EXPERIMENTS IN PATTERNED NEURONAL CULTURES

Engineered neuronal networks are not a novelty. They were introduced over two decades ago and are under continuous development [30]. Often combined with microfluidics [31, 32], the use of engineered cultures lays in the interest for immobilizing neurons in predefined locations [33], or to restrict their development along pre–patterned areas [36]. They also bring the possibility to 'tailor', design small neuronal circuits [34, 35]. In other words, engineered cultures bring the opportunity to dictate the connectivity of the network and engender very different architectures. These cultures proved successful in addressing questions as diverse as synchronized oscillations [35], front propagation [36], information coding [37], activity initiation mechanisms [38], neuronal logic [39], mutual synchronization [40], and the role of connectivity in activity propagation [41].

Although there are several strategies to control the connectivity of a neuronal culture through patterning, two of the most successful methods are chemical confinement and physical trapping. The first one consists in the imprint of adhesive proteins in predefined locations of a substrate, giving rise to a pattern where neurons adhere. An illustrative and powerful design consists in a long line several mm long and just few tens of microns wide. As shown in Fig. 1d, neurons adhere solely to the coated areas, leaving the rest of the substrate unpopulated. This method forms the basis for the preparation of *one–dimensional* neuronal cultures [36, 37, 38]. It can also be extended to form more complex, two–dimensional structures, for instance to create small islands of neurons connected to one another, a design known as *clustered neuronal cultures* [34, 35]. The second method, physical trapping, is based in the preparation of PDMS topographical (Fig. 1e) or pierced (Fig. 1e-f) molds using photolithographic techniques.

Photolithography is a process that first creates a negative relief of a photo–mask (typically printed on transparency film) out of a resin. The procedure involves a photosensitive resist that is irradiated with UV light through the photo–mask to engrave the desired pattern on the substrate (black areas of the pattern are protected, while the transparent ones are exposed). This process induces cross–linking in the irradiated parts, and after development one obtains a resin mold with the negative relief of the photo–mask. PDMS is next poured over the relief and cured to get the final mould, where the black areas of the original design correspond to the valleys of the PDMS topography. One typically

obtains topographical structures that are 50 μm deep. By chemically treating either the top or bottom parts the PDMS mold ((Figs. 1g and 1h, respectively) one can prepare cultures that grow solely in the treated areas and following complex paths.

In addition to topographical molds, one can also prepare resins $200 - 300$ μm high and pour the PDMS just below the resin level. In this way one can design pierced PDMS molds with complex shapes. Neurons are then plated with the mold in contact with the substrate, thus confining the neurons solely in the mold openings.

Patterned cultures offer immense possibilities for the freedom and versatility of the designs that one can create. To illustrate this potential, we provide below experimental results on two typical systems, namely an unidimensional culture, and a bidimensional one formed by interconnected islands of neurons.

One–dimensional cultures

We use calcium fluorescence imaging in all our experiments. Spontaneous neuronal activity is monitored through a high–speed CMOS camera attached to a microscope, providing images with a size of 800×400 pixels that cover a maximum area of 6×3 mm^2. By processing the acquired images we can study each burst in detail by identifying single neurons in the images and associating them to regions of interest (ROIs). We monitor neuronal activity at acquisition speeds in the range $50 - 200$ frames per second, allowing the detection of the onset times of neuronal activation with good precision.

Unidimensional neuronal cultures are obtained by imposing a dimension of the culture much smaller than the characteristic diameter of the neuron's dendritic tree (typically around 300 μm). Unidimensional cultures were initially introduced by Feinerman *et al.* [36] by chemically confining neurons in a narrow band of about 200 μm wide, as shown in Fig. 1d. An alternative approach that we use consists in a pierced mold (conceptually similar to the one shown in Fig. 1f) with cavities of similar shape and dimensions of the ones reported by Feinerman and coworkers. In our experiments we leave both excitation and inhibition active.

Spontaneous activity in these lines (Fig. 2a) is characterized by pulses of activity (bursts) that initiate in specific areas of the line and propagate through it. An example of a passing front in our experiments is shown in Fig. 2b. The velocity of a propagating pulse is determined by measuring the difference in the firing times of consecutive regions along the line (Figs. 2b–d). For this particular realization we considered 35 ROIs, each including approximately 50 neurons. With an acquisition speed of 100 frames/s we can precisely identify the differences in activation times of the different regions (Fig. 2c), and by plotting the position of each ROI along the line as a function of the onset times of firing we can determine with good precision the velocity of the front, which is typically around $\simeq 70$ mm/s (Fig. 2d).

Our experiments and, especially, the ones by Feinerman *et al.* [36, 37] also show that activity initiates in localized zones along the line, termed *burst initiation zones* [38], whose origin and properties are still source of much debate. On the other hand, the front itself propagates at a constant velocity, sequentially activating the neurons in their path. This is somehow expected since axons grow parallel to the line and therefore neurons

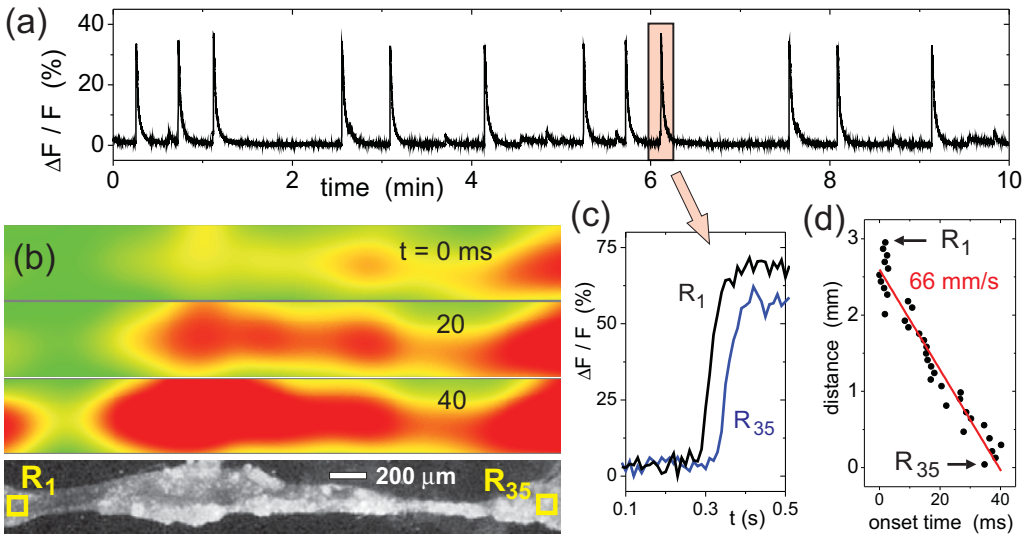


FIGURE 2. Experiments in 1D networks. (a) Spontaneous activity recorded through fluorescence imaging. Fluorescence peaks are bursting episodes, each of them corresponding to the generation and propagation of an activity front. (b) Advance of a propagating front that crosses the field of view from right to left at $\simeq 70$ mm/s. Color coding is proportional to fluorescence amplitude. The bottom panel shows an image of the field of view depicting the first and last regions of interest (ROIs). In total, 35 ROIs are positioned along the line. (c) Detail of the fluorescence signal as the front crosses the first and last ROIs. The sharp increase in fluorescence signal marks the onset time of neuronal activation. (d) Horizontal position of the ROIs as a function of the their onset times, with the linear fit providing the velocity of the front. Fluorescence amplitude is shown normalized respect to the baseline F_0, with $\Delta F = F - F_0$.

easily connect to their neighbors, forming a chain–like network that extends with equal probability towards both ends of the line.

Detailed investigations on initiation and front propagation mechanisms in these lines show that the velocity of the activity front predominantly depends on the connectivity properties of the network [37, 41]. This includes the balance between excitatory and inhibitory neurons [37, 38] and the topological features of the network itself [41]. For instance, networks with both excitation and inhibition provide velocities around 60 mm/s, which almost doubles when inhibition is blocked. The additional treatment of the neuronal culture with pharmacological agents that stimulate the formation of connections, such as BDNF or NT3, increase the velocity of the front an additional $20-30\%$ [41].

The existence of localized burst initiation zones is ascribed to both a stronger connectivity and recurrent network activity at the vicinity of the initiation zone [38]. These and other investigations also identified two characteristic velocities, a slow one associated to the recruitment process (pulse generation), and a fast one that corresponds to the actual propagation of the front. These results beautifully evidence that the initiation and propagation of spontaneous activity depends on a subtle interplay between neuronal dynamics and connectivity, a problem of intense study in 1D networks [38, 42, 43] that is also being explored in 2D systems.

Two–dimensional cultures

The investigation of the initiation and propagation of spontaneous activity in 2D networks is more difficult since one has to access large areas, and with both high temporal and spatial resolution. However, a general feature of standard 2D cultures is that their general dynamic behavior is also characterized by bursts of activity, where neurons in the network fire collectively, almost at unison, in a short time window (Fig. 3a). By *standard* cultures we refer to those characterized by a homogeneous distribution of neurons in a substrate, as in Fig. 1a. For these cultures, we found difficult the identification of burst initiation areas and the properties of the propagating front, although preliminary results provide propagation speeds in the range $100 - 150$ mm/s.

To start investigating the influence of connectivity in the dynamic behavior of 2D cultures, we consider 2D patterned cultures that slightly differ from the homogeneous case by allowing a higher aggregations between neurons. The culture is prepared by depositing patches of the adhesive protein poly–l–lysine over a glass substrate. The preparation is combined with a PDMS pierced mold containing circular cavities 3 mm wide, so that the entire culture exactly fits in the field of view of the camera. This is advantageous since we monitor the activity of the entire network population. The patterned culture that finally results is formed by small assemblies of neurons connected to one another, leading to a network architecture that combines short range connections within assemblies with long range connections across the culture.

In the experiments we leave both excitation and inhibition active. A detail of the patterned network, together with the raster plot of network activity for a single burst, is shown in Fig. 3b. The raster plot is ordered by time of neuronal activation for clarity. A fluorescence image of the entire culture is provided in Fig. 3c, where the bright patches are small aggregates of neurons containing $\simeq 20 - 40$ cells.

A first analysis of the activity profiles reveals diverse phenomenologies that are not observed in the standard, homogeneous neuronal culture. First, from the raster plot it is clear that the velocity of the front is slower and fluctuates strongly during its evolution. Second, as shown in the sequence of Fig. 3d, the front advances in a non–uniform fashion, suggesting that local inhomogeneities in connectivity may determine the properties of the front. And third, we observe that activity preferentially initiates in specific spots at the edge of the culture, similarly to the burst initiation zones observed in the 1D system.

For the front shown in Fig. 3d we calculated its velocity by measuring the average advance of the front in its normal direction, and obtained 12 mm/s with a variability of more than 50% depending on the region of the culture. Interestingly, the shape of the front and its velocity also varied from burst to burst.

These experiments highlight the rich repertoire of dynamic patterns that neuronal cultures can display. The experiments also open interesting questions to be addressed in the future, such as the importance of local fluctuations in front propagation, the role of inhibition, or at what extend the distribution and size of the cell assemblies is crucial.

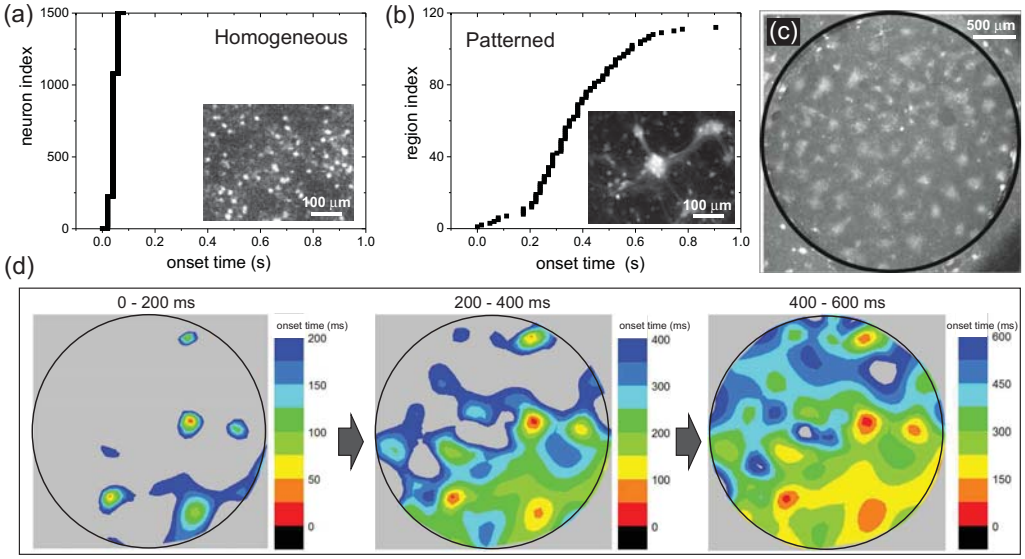


FIGURE 3. Experiments in 2D patterned networks. (a)-(b) Raster plots comparing the activity in a standard, homogeneous neuronal culture and a patterned one. The former shows a fast propagation, while the latter displays a much slower propagation with varying velocity. The images illustrate the structure of the network. Bright spots are firing neurons. (c) Fluorescence image of the entire patterned culture, formed by small assemblies of neurons (bright areas) connected to one another. The network is confined inside a PDMS circular cavity 3 mm in diameter (black outline). (d) Evolution of a spontaneously generated front in the patterned culture. Each contour shows the order of neuronal activation within a given time frame. The front is rich in structure, and with a propagation velocity in the range 5 – 12 mm/s.

DISCUSSION

Our studies in both 1D and 2D patterned cultures exemplify how a relatively simple experimental concept can help understanding the interplay activity–connectivity in living neuronal networks.

One of the limitations of the experiments is the use of calcium imaging for monitoring activity. Its relatively poor temporal resolution hinders single neuronal spiking events and makes difficult the characterization of fast network events prior burst initiation. The continuous development of new calcium indicators, particularly those genetically encoded, brings better capabilities for precise detection of neuronal activity. However, superior time resolution is still a major drawback in calcium imaging studies. Hence, several investigations are addressing the problem of spontaneous activity through MEAs for their high temporal resolution. Those studies proved highly valuable. For the problem of burst initiation, for instance, some works identified a subset of neurons termed *leaders* that were always the first to ignite, and that induced the firing of the entire neuronal population. It was also observed that activity in this subset increased exponentially in a short time window prior to the burst [44]. The authors hypothesized that such a fast recruitment required the leader neurons to be present all over the network, possibly forming a subnetwork of highly connected neurons [45, 46].

The existence of leader neurons is still an open issue. Given the poor density of recording sites in some experiments, it is difficult to precisely establish sequences of neuronal

activation. Recent experiments in burst initiation mechanisms could not identify leader neurons [14], and there is a strong debate on the possibility that they may be a 'side effect' of avalanching mechanisms or other complex dynamical processes during activity recruitment and burst build–up. Another possibility is that, since connectivity in cultures increases with neuronal density [15], the strong fluctuations in the spatial distribution of neurons may give rise to highly connected regions that are foci of activity. Hence, given the poor spatial resolution of MEAs, it may occur that leader neurons actually correspond to regions with distinct topological properties that favor initiation, rather than a special kind of cells.

To advance in the comprehension of initiation mechanisms, and to fully use the potential of patterned cultures, an important aspect that needs attention is the characterization of the structure of the network, i.e. its layout of connections. For the 1D experiments reported in the literature [38, 39] the authors used green fluorescence protein (GFP) labeling to identify connections and extract statistical information on neuronal processes. In combination with the rich theoretical studies of pulse propagation in 1D networks [43, 47] the authors built realistic simulations and models that were later compared and fit to experimental data. Such a successful interplay between experiments and modeling is still lacking in 2D patterned cultures. However, new technological developments oriented towards connectivity visualization and identification, or the improvement in connectivity reconstruction algorithms [48], are gradually providing more data and resources for modeling. In this sense, statistical physics is providing indirect yet powerful methods to extract information on neuronal connectivity, for instance with the combination of experiments with percolation and network theory tools [15, 49, 50, 51].

Network theory has boosted in the last years both our understanding and description of living neuronal networks. The development and characterization of patterned cultures may serve as a unique platform to validate network theory concepts and develop new ideas. Indeed, the possibility that these cultures offer to dictate network architecture and monitor the subsequent dynamics is enormously attractive for the complex networks community, and at a well controlled, accessible mesoscopic scale.

ACKNOWLEDGMENTS

We thank J. Orlandi, E. Álvarez-Lacalle and J. Casademunt for fruitful discussions and insight. We acknowledge financial support from Ministerio de Ciencia e Innovación (Spain) under projects FIS2008-06335-C02-01, FIS2009-07523, and FIS2010-21924-C02-02, and the Generalitat de Catalunya under project 2009-SGR-00014.

REFERENCES

1. E. Bullmore, and O. Sporns, *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
2. D. Zhang, and M. E. Raichle, *Nat. Rev. Neurology* **6**, 15–28 (2010).
3. G. Buzsáki, and A. Draguhn, *Science* **304**, 1926–1929 (2004).
4. K. Smith, *Nature* **466** (suppl.), S15–S18 (2010).
5. H. Markram, *Nat. Rev. Neurosci.* **7**, 153–160 (2006).
6. M. M. Waldrop, *Nature* **482**, 456–458 (2012).

7. J. W. Bohland, *PLoS Comput. Biol.* **5**, e1000334 (2009).
8. R. Albert, and A. L. Barabasi, *Rev. Mod. Phys.* **74**, 47–97 (2002).
9. S. Boccaletti, *et al.*, *Phys. Rep.* **424**, 175–308 (2006).
10. A. Arenas, *et al.*, *New J. Phys.* **10**, 053039 (2008).
11. A. Arenas, *et al.*, *Lect. Not. Comput. Sci.* **5151**, 9–18, Springer (2008).
12. M. A. Colicosand, and N. I. Syed, *J. Exp. Biol.* **209**, 2312–2319 (2006).
13. J.-P. Eckman, O. Feinerman, L. Gruendlinger, E. Moses, J. Soriano, and T. Tlusty, *Phys. Rep.* **449**, 54–76 (2007).
14. E. Cohen, M. Ivenshitz, V. Amor–Baroukh, V. Greenberger, and M. Segal, *Brain Res.* **1235**, 21–30 (2008).
15. J. Soriano, M. Rodríguez Martínez, T. Tlusty, and E. Moses. *Proc. Natl. Acad. Sci. USA* **105**, 13758–13763 (2008).
16. D. A. Wagenaar, J. Pine, and S. M. Potter, *BMC Neurosci.* **7**, 11 (2006).
17. C. Stosiek, O. Garaschuk, K. Holthoff, and A. Konnerth, *Proc. Natl. Acad. Sci. USA* **100**, 7319–7324 (2003).
18. C. Grienberger, and A. Konnerth,, *Neuron* **73**, 862–885 (2012).
19. S. Marom, and G. Shahaf, *Q. Rev. Biophys.* **35**, 63–87 (2002).
20. V. Pasquale, P. Massobrio, L. L. Bologna, M. Chiappalone, and S. Martinoia, *Neuroscience* **153**, 1354–1369 (2008).
21. C. Tetzlaff, S. Okujeni, U. Egert, F. Wörgötter, and M. Butz, *PLoS Comput. Biol.* **6**, e1001013 (2010).
22. J. Erickson, A. Tooker, Y.-C. Tai, and J. Pine, *J. Neurosci. Methods* **175**, 1–16 (2008).
23. M. D. McDonnell, and L. M. Ward, *Nat. Rev. Neurosci.* **12**, 415–426 (2011).
24. A. Arenas, *et al.*, *Phys. Rep.* **469**, 93–153 (2008).
25. D. Eytan, and S. Marom, *J. Neurosci.* **26**, 8465–8476 (2006).
26. D. A. Wagenaar, R. Madhavan, J. Pine, and S. M. Potter, *J. Neurosci.* **25**, 680–688 (2005)
27. A. Rotem, and E. Moses, *Biophys. J* **94**, 5065–5078 (2008).
28. O. Yizhar, L. E. Fenno, T. J. Davidson, M. Mogri, and K. Deisseroth, *Neuron* **71**, 9–34 (2011).
29. M. V. Sanchez–Vives, and D. A. McCormick, *Nat. Neurosci.* **3**, 1027–1034 (2000).
30. B. C. Wheeler, and G. J. Brewer, *Proc. IEEE* **98**, 398–406 (2010).
31. J. W. Park, B. Vahidi, A. M. Taylor, S. W. Rhee, and N. L. Jeon, *Nat. Protocols* **1**, 2128–2136 (2006).
32. J. Wang, *et al.*, *Lab Chip* **7**, 644–652 (2009).
33. M. Shein, *et al.*, *Biomed. Microdevices* **11**, 495–501 (2009).
34. R Sorkin, *et al. J. Neural. Eng.* **3**, 95–101 (2006).
35. M. Shein–Idelson, E. Ben–Jacob, and Y. Hanein, *PLoS ONE* **5**, e14443 (2010).
36. O. Feinerman, M. Segal, and E. Moses, *J. Neurophysiol.* **94**, 3406-3416 (2005).
37. O. Feinerman, and E. Moses, *J. Neurosci.* **26**, 4526–4534 (2006).
38. O. Feinerman, M. Segal, and E. Moses, *J. Neurophysiol.* **97**, 2937–2948 (2007).
39. O. Feinerman, A. Rotem, and E. Moses, *Nat. Phys.* **4**, 967–973 (2008).
40. I. Baruchi, V. Volman, N. Raichman, M, Shein, and E. Ben–Jacob, *Eur. J. Neurosci.* **28**, 1825–1835 (2008).
41. S. Jacobi, J. Soriano, and E. Moses, *J. Neurophysiol.* **104**, 2932–2939 (2010).
42. S. Jacobi, and E. Moses, *J. Neurophysiol.* **97**, 3597–3606 (2007).
43. E. Álvarez–Lacalle, and E. Moses, *J. Comput. Neurosci.* **26** 475–493 (2009).
44. D. Eytan, and S. Marom, *J. Neurosci.* **26**, 8465–8476 (2006).
45. J-P. Eckmann, S. Jacobi, S. Marom, E. Moses, and C. Zbinden *New J. Phys.* **10**, 015011 (2008).
46. J-P. Eckmann, E. Moses, O. Stetter, T. Tlusty, and C. Zbinden *Front. Comput. Neurosci.* **4**, 132 (2010).
47. R. Osan, and G. B.Ermentrout, *Physica D* **163**, 217–235 (2002).
48. O. Stetter, D. Battaglia, J. Soriano, and T. Geisel *PLoS Comput. Biol.* **8**, e1002653 (2012).
49. I. Breskin, J. Soriano, E. Moses, and T. Tlusty. *Phys. Rev. Lett.* **97**, 188102 (2006).
50. J. Soriano, I. Breskin, E. Moses, and T. Tlusty, "Percolation approach to study connectivity in living neural networks", in *9th Granada Seminar 2006 – Cooperative Behavior in Neural Systems*, edited by P. L. Garrido, J. Marro, & J. J. Torres, *AIP Conference Proceedings* **887**, pp. 96–106, American Institute of Physics, New York, 2006.
51. O. Cohen, *et al.*, *Europhys. Lett.* **89**, 18008 (2010)

From structure to function, via dynamics

O. Stetter*, J. Soriano†, T. Geisel* and D. Battaglia*

*Max Planck Institute for Dynamics and Self-organization and Bernstein Center for Computational Neuroscience, Am Faßberg 17, D-37077 Göttingen, Germany.
†Departament d'ECM, Facultat de Física, Universitat de Barcelona, Martí i Franqués 1, E-08028, Barcelona, Spain.

Abstract. Neurons in the brain are wired into a synaptic network that spans multiple scales, from local circuits within cortical columns to fiber tracts interconnecting distant areas. However, brain function require the dynamic control of inter-circuit interactions on time-scales faster than synaptic changes. In particular, strength and direction of causal influences between neural populations (described by the so-called *directed functional connectivity*) must be reconfigurable even when the underlying structural connectivity is fixed. Such directed functional influences can be quantified resorting to causal analysis of time-series based on tools like Granger Causality or Transfer Entropy. The ability to quickly reorganize inter-areal interactions is a chief requirement for performance in a changing natural environment. But how can manifold functional networks stem "on demand" from an essentially fixed structure? We explore the hypothesis that the self-organization of neuronal synchronous activity underlies the control of brain functional connectivity. Based on simulated and real recordings of critical neuronal cultures in vitro, as well as on mean-field and spiking network models of interacting brain areas, we have found that "function follows dynamics", rather than structure. Different dynamic states of a same structural network, characterized by different synchronization properties, are indeed associated to different functional digraphs (functional multiplicity). We also highlight the crucial role of dynamics in establishing a structure-to-function link, by showing that whenever different structural topologies lead to similar dynamical states, than the associated functional connectivities are also very similar (structural degeneracy).

Keywords: connectivity; functional connectivity; causality; oscillations and synchrony; neural dynamics; neuronal cultures; neuroimaging.
PACS: 87.10.-e, 87.18.Sn, 87.19.L, 87.19.le, 87.19.lf, 87.19.lj, 87.19.lh, 87.19.lm, 87.19.lo

INTRODUCTION

Flexible transmission of information is a core feature of biological systems. For instance, the firing activity of neurons conveys information about the external world or internal brain states. Arguably, the correct timing of the exchanged signals is crucial for a correct relay of information through complex networks. A natural device to achieve such temporal coordination might be self-organized synchronization. Oscillatory synchronization, in particular, has been observed in interaction networks arising in very diverse domains. In particular, consistent experimental evidence as been cumulated for the role played in perception and cognition by oscillatory coherence in neural circuits at multiple scales [1, 2]. Notably, according to the "communication-through-coherence" hypothesis [3], information exchange between two neuronal populations is enhanced when the oscillations of their coherent activity is suitably phase-locked with a suitable phase-relation. Therefore the efficiency and the directionality of information transmission between neuronal populations is affected by changes in their synchronization pattern, as also advocated

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 64-73 (2013); doi: 10.1063/1.4776502

by modeling studies [4, 5]. Furthermore synchronization in networks of spiking neurons can arise in many forms, ranging from sparsely synchronized periodic oscillations [6], to low-dimensional chaotic rhythms [7, 8] to temporally-irregular avalanche-like bursting [9].

The circuits of the brain must enact a sweeping amount of functions. How can a flexible control of local computations or global "brain states" be achieved despite the fact that anatomic interconnections are essentially fixed on fast timescales relevant for perception or behavior? In systems neuroscience, a distinction is made between structural and directed functional connectivities [10]. Structural connectivity describes actual synaptic connections. On the other hand, directed functional connectivity is estimated from time-series of simultaneous neural recordings using causal analysis [11], to quantify, beyond correlation, directed influences between brain areas.

Here, we revisit recent theoretical and modeling work [5, 12] showing that the relation between structural and functional connectivity in neural circuits is far from being trivial. Indeed, if structural connectivity constrains at least partially the kinds of dynamics that a given neural circuit can display, nevertheless it does not constrain them fully. As a result, multiple functional connectivities, associated to different dynamical states, can stem out of a system with a given structural connectivity, a phenomenon that we call *functional multiplicity*. In parallel, a same specific functional connectivity can be generated by very different structural circuits displaying strongly similar dynamics, a phenomenon that we refer to as *structural degeneracy*.

To highlight these two phenomena, we simulate large networks of spiking neurons representing systems at different scales, i.e. synchronously bursting cultures of dissociated neurons [12] and mesoscopic motifs involving brain areas undergoing a coherent oscillatory activity [5]. For both these systems, directed functional connectivity is estimated from synthetic activity time-series, allowing to establish a tight correspondence between ground-truth dynamics (usually unknown in the case of causal analyses of experimental recordings) and emergent functional connectivity for a wide spectrum of considered network topologies.

STATE-DEPENDENT TRANSFER ENTROPY

Throughout this contribution, we will characterize directed functional interactions in terms of Transfer Entropy [13], an information-theoretic implementation of the Granger Causality concept [14, 15]. Following the original idea of Clive Granger, causal analysis intends to go beyond correlation analysis, by quantifying under a strict operational definition how forecasting of the evolution of a given time-series is improved by he consideration of a time-series of the activity of a potential cause.

Let us consider a pair of continuous time-series, describing the dynamics of two different neural circuit elements, like e.g. LFPs or EEGS from different brain areas in a meso- or macroscale neural circuit, or optical recordings of single neuron activity in a micro-scale circuit.

Let denote these two time-series as $\Lambda_X(t)$ and $\Lambda_Y(t)$ and let quantize them into B discrete levels $\ell_1, \ldots, \ell_B$ (equally sized for simplicity). The continuous-valued time-series are thus converted into strings of symbols $\tilde{\Lambda}_X(t)$ and $\tilde{\Lambda}_Y(t)$ from a small alphabet

[16].

Two transition probability matrices, $(P_{XY,Y}(\tau))_{ijk} = P[\tilde{\Lambda}_Y(t) = \ell_i|\tilde{\Lambda}_Y(t-\tau) = \ell_j, \tilde{\Lambda}_X(t-\tau) = \ell_k]$ and $(P_{Y,Y}(\tau))_{ij} = P[\tilde{\Lambda}_Y(t) = \ell_i|\tilde{\Lambda}_Y(t-\tau) = \ell_j]$, where the lag τ is an arbitrary temporal scale on which causal interactions are probed, are then sampled as normalized multi-dimensional histograms over very long symbolic sequences.

Then, following [13], the causal influence $\mathrm{TE}_{XY}(\tau)$ of circuit element X on circuit element Y is defined as the Transfer Entropy:

$$\mathrm{TE}_{XY}(\tau) = \sum P_{XY,Y}(\tau) \log_2 \frac{P_{XY,Y}(\tau)}{P_{Y,Y}(\tau)} \tag{1}$$

where the sum runs over all the three indices i, j and k of the transition matrices.

This quantity represents the Kullback-Leibler divergence [17] between the transition matrices $P_{XY,Y}(\tau)$ and $P_{Y,Y}(\tau)$, analogous to a distance between probability distributions. Therefore, $\mathrm{TE}_{XY}(\tau)$ will vanish if and only if $P_{XY,Y}(\tau)$ and $P_{Y,Y}(\tau)$ coincide, i.e. if the transition probabilities between different activity values of circuit element Y do not depend on past activity values of circuit element X. Conversely, this quantity will be strictly positive if these two transition matrices differ, i.e. if the past activity values of circuit element X affect the evolution of the activity in circuit element Y.

Such expression for Transfer Entropy can straightforwardly be generalized to higher Markov order descriptions by conditioning transition probabilities over longer vectors of past activity values, e.g. by considering, for order $p = 2$, $(P_{XY,Y}(\tau, p=2))_{ij_1j_2k_1k_2} = P[\tilde{\Lambda}_Y(t) = \ell_i|\tilde{\Lambda}_Y(t-\tau) = \ell_{j_1}, \tilde{\Lambda}_Y(t-2\tau) = \ell_{j_2}, \tilde{\Lambda}_X(t-\tau) = \ell_{k_1}, \tilde{\Lambda}_X(t-2\tau) = \ell_{k_2},]$ and $(P_{Y,Y}(\tau, p=2))_{ij_1j_2} = P[\tilde{\Lambda}_Y(t) = \ell_i|\tilde{\Lambda}_Y(t-\tau) = \ell_{j_1}, \tilde{\Lambda}_Y(t-2\tau) = \ell_{j_2}]$.

More importantly, to analyze directed functional interactions in different dynamical states separately, a further *state conditioning* can be introduced. Let $\vec{\mathbf{S}}_T$ a vector describing the history of the entire system (i.e. not only the two considered circuit elements X and Y but the whole neural circuit to which they belong) over a specified time-window T. We define then an arbitrary set of constraints $\mathscr{K}$ that $\vec{\mathbf{S}}_T$ must satisfy as a *state selection filter*. Then we will define:

$$\mathrm{TE}_{XY,\mathscr{K}}(\tau) = \sum_{\vec{\mathbf{S}}_T \text{ satisfies } \mathscr{K}} P_{XY,Y}(\tau) \log_2 \frac{P_{XY,Y}(\tau)}{P_{Y,Y}(\tau)} \tag{2}$$

where the sampling is restricted only to (possibly disconnected) sections of the time-series $\tilde{\Lambda}_X(t)$ and $\tilde{\Lambda}_Y(t)$ corresponding to time-intervals during which the collective system state $\vec{\mathbf{S}}_T$ satisfies the set of filtering constraints $\mathscr{K}$.

Although the definition is very general, we will consider in the following two simple type of filters, one based on the average activity of the system within a considered time-window and another one based on phase relations to be fulfilled between different system components.

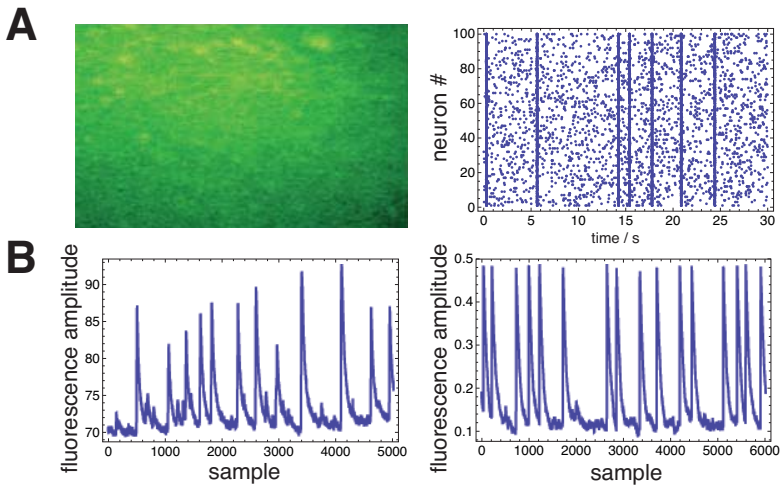


FIGURE 1. *Bursting neuronal cultures* in vivo *and* in silico. A: a frame of a calcium imaging movie of the dynamics of a real culture is compared with the raster plot of a simulated network of integrate-and fire neurons. B: examples of real (left) and synthetic (right) time-series of average calcium fluorescence. Highly synchronous network bursts are manifested by fluorescence peaks, in both experiment and simulations. For details of simulations see [12].

FUNCTIONAL INTERACTIONS IN BURSTING CULTURES

Analysis of neuronal cultures *in vitro* is emerging as a versatile paradigm [18] in the quest for uncovering neuronal connectivity [19] and its interplay with dynamics. Using calcium imaging techniques, the activity of order 10^2–10^3 cells in *in vitro* can be simultaneously monitored, even if with a time-resolution of few tenths of a ms, i.e. slower than the cell's firing by an order of magnitude. We have introduced in [12] an algorithm for the reconstruction of the connectivity of cultured networks based on calcium fluorescence time-series. Our method is based on Transfer Entropy (TE)[13], an information-theory based generalization of Granger Causality [11]. An important advantage of our approach is its model-free nature, not assuming specific models of neuronal activity or network connectivity and not being constrained to linear interactions between nodes.

To benchmark our reconstruction algorithm we have introduced an *in silico* model of *in vitro* cultures, designed to reproduce the occurrence of temporally irregular switching between states of weak-rate asynchronous activity and states of highly synchronous activity, commonly denoted as "network bursts" [19]. All the details, including simulation parameters, can be found in [12]. Here we briefly mention that we simulated the spontaneous spiking dynamics of networks of excitatory integrate-and-fire neurons, matching typical experimental conditions. Network bursts occur in our model thanks to the introduction of limited neurotransmitter resources [20]. Realistic bursting rates and distributions could be obtained for very diverse structural topologies, notably with arbitrary clustering levels [12]. Synthetic calcium fluorescence time series were then produced based on this spiking dynamics. Figure 1 shows a comparison between real

and simulated calcium fluorescence signals. Network bursts are evident in both real and synthetic traces.

We extract then directed functional connectivity based on time-series x_n and y_n of (high-passed) simulated calcium fluorescence, evaluating a state-dependent TE for every pair of nodes X and Y:

$$\mathrm{TE}^{*}_{Y\to X}(\tilde{g}) = \sum P(x_{n+1}, x_n, y_{n+1}, y_n \,|\, \tilde{g}_* < g_{n+1} < \tilde{g}^*) \cdot \log \frac{P(x_{n+1}|x_n, y_{n+1}, y_n \,|\, \tilde{g}_* < g_{n+1} < \tilde{g}^*)}{P(x_{n+1} \,|\, x_n, \tilde{g}_* < g_{n+1} < \tilde{g}^*)} \tag{3}$$

Comparing Eq. (3) with the general formulation of state-dependent TE given by equation (1), we observe that state selection filtering is performing through inspection and constraining of a variable g_n given by fluorescence averaged over the entire network. Datapoints are sampled for the evaluation of TE between two network nodes only when the average network fluorescence falls within a specific range with lower and upper bounds given respectively by $\tilde{g}_*$ and $\tilde{g}^*$. Such conditioning on mean fluorescence corresponds to a crude way of restricting the analysis to a specific dynamic regime. For instance, asynchronous inter-burst periods or synchronous bursting epochs are associated to different mean fluorescence ranges (Fig. 2A).

Another modification with respect to the basic definition, of technical rather than conceptual importance, is *instantaneous feedback*, i.e. the appearance of y_{n+1} in Eq. (3), accounting for possible causal interactions faster than the poor sampling resolution of the used calcium imaging technique.

Functional multiplicity in bursting cultures

Functional networks associated to different dynamical regimes are obtained by including into the network all the edges whose generalized TE score is above a certain threshold specified a priori. For synthetic data the overlap between the reconstructed functional network and the known ground-truth structural connectivity can be evaluated for different choices of the threshold, and results of this comparison can be summarized by receiver-operating-characteristic (ROC) curves, for different dynamical regimes. A shown in Fig. 2B, functional topologies in inter-burst and in bursting regimes are very different and can be quite different from structural topology.

When considering very low fluorescence level, what we see is essentially noise (Figure 2B, regime I). Correspondingly, links are entered into the reconstructed functional network practically at random, as indicated by a diagonal ROC curve. We consider then intermediate fluorescence levels, associated to an activity significantly above baseline, but not yet elevated as in fully developed bursts. In this inter-bursts regime (Fig. 2B, regime II), the retrieved functional network is strongly correlated with the underlying structure, since detected causality reflect primarily the direct influence of pre-synaptic neurons on post-synaptic targets. When bursts are fully developed (Fig. 2B, regime III), the network is a critically excitable state, where the firing of a single neuron can trigger an avalanche of firing extending even to neurons not structurally connected to it. In this regime, therefore the retrieved functional network reflects communities of tightly

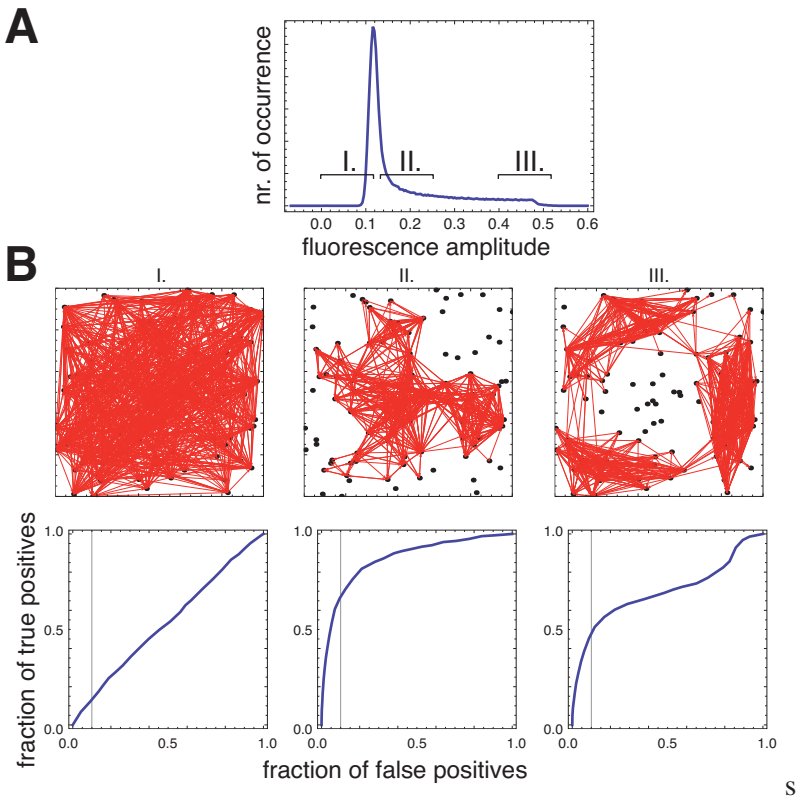


FIGURE 2. *Functional multiplicity in simulated cultures.* A: the distribution of mean fluorescence levels g_n presents an initial gaussian rise, followed by a transition region and then by a (initially power-law) tail. B: directed functional connectivity networks retrieved in different dynamical regimes by Transfer Entropy. Also shown are ROC curves for network reconstruction from different dynamical regimes, describing quality of overlap between functional networks and the underlying structural connectivity. A vertical line denotes the position on the ROC curve corresponding to the depicted networks. The considered regimes are: I. Noise-dominated weak rate regime. II. Inter-bursts regime with intermediate firing rate. III. Fully developed bursts regime. See [12] for details.

synchronous firing, rather than structural topology. The ROC curve indicates thus a poorer quality structural reconstruction, even if the localization and the extension of synchronous communities continue to be shaped, roughly, by structure (as denoted by a better-than-random ROC curve).

The existence of such different topologies of functional interactions stemming out of different dynamical ranges of a same structural network provides a perfect example of the notion of functional multiplicity, previously defined in the introduction.

Structural degeneracy in bursting cultures

To address the problem of the structural-to-functional connectivity inter-relations under a different angle, we construct bursting neuronal cultures model with very different

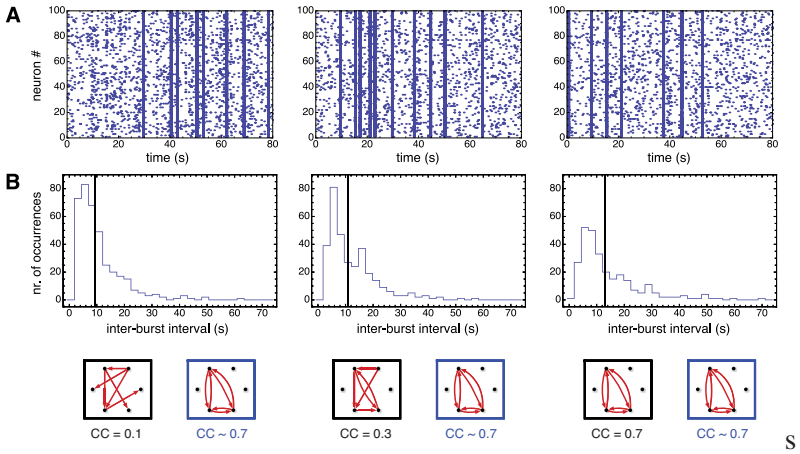


FIGURE 3. *Structural degeneracy in simulated cultures.* A: Examples of spike raster plots for three structural networks with different clustering coefficients (respectively 0.1, 0.3 and 0.5, from left to right), showing that their underlying spiking dynamics are similar. B: Histograms of the inter-burst intervals (IBIs), with the vertical lines indicating the mean of each distribution. C: panels below the IBI distributions illustrate graphically the amount of clustering in the actual structural network (black frame and text) and in the functional network reconstructed from fluorescence range 3 (bursting regime) as given by Fig. 2 (blue frame and text). To very different degrees of structural clustering correspond equivalent elevated levels of functional clustering, due to the common bursting statistics. Figure adapted from [12].

structural topologies. Figure 3 illustrates the dynamic behavior of three networks, designed to have different clustering coefficients but the same total number of links. The synaptic coupling between neurons was adjusted in each network using an automated procedure to obtain bursting activities with comparable bursting rates (see [12] for details on the procedure and on the models). The simulated spiking dynamics is shown in the raster plots of Fig. 3A. These three networks display indeed very similar bursting dynamics, not only in terms of the mean bursting rate, but also in terms of the entire inter-burst interval (IBIs) distribution, shown in Fig. 3B.

Based on these bursting dynamics, we extracted the functional connectivity of the three differently-clustered structural networks, associated to the dynamics range III defined by Fig. 2, i.e. the fully-developed burst regime. The extracted bursting-regime functional networks had always an elevated clustering level, close to 0.7. This contrasted with the actual structural clusterings, varying in a broad range between 0.1 and 0.5 (see Fig. 3C).

We stress that our procedure for the automatic generation of networks with similar bursting dynamics was not guaranteed to converge for such a wide range of clustering coefficients. Thus, the illustrative simulations of Fig. 3 provide genuine evidence that the relation between network dynamics and network structural clustering is not trivially "one-to-one". In this sense, Fig. 3 provides a typical example of the phenomenon denoted as structural degeneracy in the introduction, i.e. many structures can map to a same function.

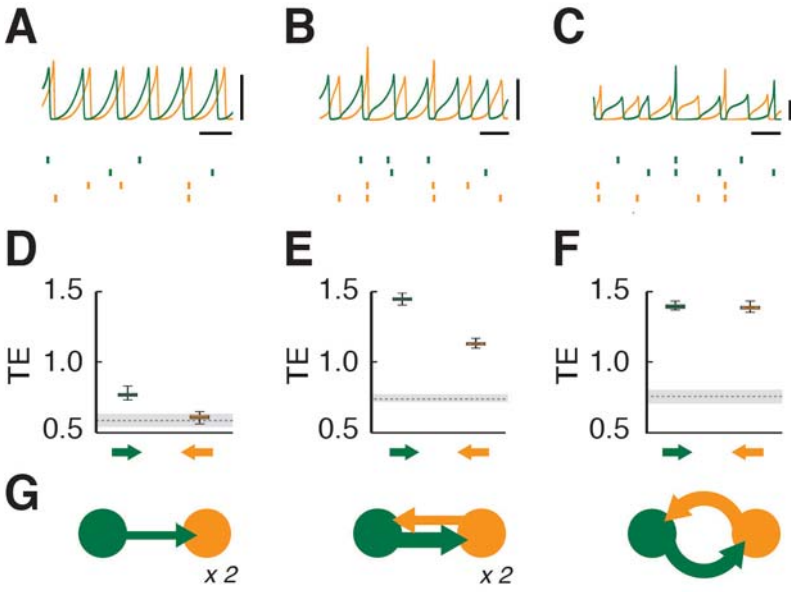


FIGURE 4. *State-dependency of functional connectivity in simple motifs of interacting populations.* Shown here are dynamical states and resulting directed functional connectivites of a motif of $N = 2$ brain areas structurally connected in a symmetric way. A–C: simulated "LFPs" and spike trains of the two populations for three different strengths of the symmetric inter-areal coupling, leading to more or less regular phase-locked states. D–E: Transfer entropies for the two possible directions of functional interaction, associated to the dynamic states in panels A–C. A grey band indicates threshold for statistical significancy of a causal interaction. G: graphic depiction of the functional interactions between the two areas, as captured by Transfer Entropy, in the states that can then be described of effective entrainment (A), leaky effective entrainment (B) and mutual entrainment (C). A multiplier factor indicate multistability between motifs with same topology but different direction. Figure adapted from [5].

FUNCTIONAL INTERACTIONS IN OSCILLATING MOTIFS

Moving then to a larger scale, we simulate structural motifs involving a small number of coupled brain areas. A local area is modeled as a random network of thousands of excitatory and inhibitory spiking neurons. In addition to diluted inhibition and excitation within each area, long-range excitation between areas is also introduced. Details of the model are given in [5], but parameters are selected in such a way that isolated areas undergo a collective oscillation at a frequency of $\sim 40 - 60$ Hz. When connected into a motif, with identical probability of long-range connections in all directions, these locally-generated oscillations engage into phase-locked states. For increasing coupling strengths, these synchronous oscillations can become chaotic leading to perturbation of precise phase-locking. Note that despite the regularity of collective activity, as tracked for instance by average membrane potential (as a proxy for "Local Field Potential", or "LFP"), individual neurons continue to fire very irregularly (see Fig. 3A–C). In such sparsely synchronized states, individual spike trains can be very entropic, i.e. convey potentially large amount of information, even when the ongoing oscillation is periodic.

In a broad range of conditions (notably, when local inhibition is strong [7]), populations lock in out-of-phase configurations, in which some areas lead in phase over the others. The symmetry of such phase-locked states is weaker than the structural motif full symmetry. Due to this spontaneous symmetry breaking, anisotropy of functional interactions can then emerge, as revealed by ordinary Transfer Entropy [13] between "LFP" time-series (cfr. Fig. 3D–F) or mutual information between spike trains (cfr. [5]).

The existence of such phase-locked stable states provide a natural way to define the state selection filters $\mathscr{K}$ to be used in evaluating Eq. (1). TE within a specific state can be computed by sampling only over epochs in which the inter-areal phase shift fluctuates in close vicinity to one of the possible stable phase-lockings.

TE analyses can be summarized in graphical form by drawing the possible "functional motifs" that a given structural motif can generate. In Fig. 3G arrows of increasing thickness indicate statistically significant causal interactions of increasing strength. Thus, the net information transfer over a $N = 2$ fully symmetric structural topology can be unidirectional (Fig. 3A and D) or bidirectional anisotropic (Fig. 3B and E) or balanced (Fig. 3C and F), depending on the coupling strength. Furthermore, multistability between motifs with different dominant directionality exists whenever the symmetry of the functional motif is broken, in such a way that rewiring of directed functional connectivity can be achieved just through suitable transient perturbations to the ongoing oscillations [5].

FUNCTION FROM STRUCTURE?

The architect Louis Sullivan first popularized a tag line stating that "form follows function". The two examples just reviewed certainly disclose that "function doesn't follow structure (trivially)": functional connectivity can for instance display a clustered community structure (Fig. 2) or be strongly anisotropic (Fig. 4) even when structural connectivity is homogeneous or simmetric. Furthermore, very different structural connectivities can give rise t very similar functional connectivities (Fig. 3). Thus, these examples rather showed that "function follows dynamics", since the properties of the dynamical states supported by a given structure determine the resulting functional connectivities.

Still and all, functional connectivity patterns of whole-brain activity are known to be strongly determined by structure [21]. Note that, in our examples, structure was fixed a priori, but, in nature (or in the dish) networks are shaped by spontaneous growth and eventually, on longer time-scales, evolution. Which is then the optimization goal that self-organized design tries to achieve? We don't know the answer, but if, as sometimes speculated [22], brain structural topology had developed such to lead to rich repertoires of possible dynamics, it might well be that Louis Sullivan's motto applies as well to the description of living neural circuits at multiple scales, even if only through an indirect detour involving nonlinear dynamics. As a matter of fact, for evolution or development, the problem of engineering a circuit with a given function, could be nothing else than the design of a structural connectivity acting as an emergent "functional collectivity" [23] with suitable properties.

ACKNOWLEDGMENTS

The authors would like to thank the Granada Seminar organizing committee members. They are also very grateful to Dante Chialvo for illuminating discussions.

REFERENCES

1. R. Eckhorn et al, *Biol. Cybern.* **60**, 121–130 (1988).
2. F. Varela, J. P. Lachaux, E. Rodriguez, and J- Martinerie, *Nat. Rev. Neurosci.* **2**, 229–239 (2001).
3. P. Fries, *Trends Cogn. Sci.* **9**, 474–480 (2005).
4. A. Buehlmann, and G. Deco, *PLoS Comp. Biol.* **6,** e100934 (2010).
5. D. Battaglia, A. Witt, F. Wolf, and T. Geisel, *PLoS Comp. Biol.* **8**, e1002438 (2012).
6. N. Brunel, and X. J. Wang, *J. Neurophysiol.* **90**, 415–430 (2003).
7. D. Battaglia, N. Brunel, and D. Hansel, *Phys. Rev. Lett.* **99**, 238106 (2007).
8. D. Battaglia, and D. Hansel, *PLoS Comput. Biol.* **7**, e1002176 (2011).
9. A. Levina, J. M. Herrmann, and T. Geisel, *Nat. Phys.* **3**, 857–860 (2007).
10. K. J. Friston, *Brain Connectivity* vol. 1, pp. 13-36, 2011.
11. B. Gourévitch, R. L. Bouquin-Jeannès, and G. Faucon, *Biol. Cybern.* **95**, 349–369 (2006).
12. O. Stetter, D. Battaglia, J. Soriano, and T. Geisel, *PLoS Comput. Biol.* **8**, e1002652 (2012).
13. T. Schreiber, *Phys. Rev. Lett.* **85**, 461–464 (2000).
14. C. W. J. Granger, *Econometrica* **37**, 424–438 (1969).
15. M. Ding M, Y. Chen, and S.L. Bressler SL, "Granger causality: Basic theory and application to neuroscience", in *Handbook of Time Series Analysis*, editors S. Schelter, N. Winterhalder, and J. Timmer, Wiley, Wienheim, 2006.
16. M. Staniek, and K. Lehnertz, *Phys. Rev. Lett.* **100**, 158101-5 (2008).
17. D. MacKay, *Information theory, inference, and learning algorithms*, Cambridge University Press, Cambridge (UK), 2003.
18. J. P. Eckmann et al., *Phys. Rep.* **449**, 54–76 (2007).
19. J. Soriano, M. R. Martinez, T. Tlusty, and E. Moses, *Proc. Natl. Acad. Sci. USA* **105**, 13758–13763 (2008).
20. M. Tsodyks, A. Uziel, and H. Markram, *J. Neurosci.* **20**, 1–5 (2000).
21. C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns, *Proc. Natl. Acad. Sci. USA* **104**, 10240–10245 (2007).
22. G. Deco, V. K. Jirsa, and R. McIntosh, *Nat. Rev. Neurosci.* **12**, 43–56 (2011).
23. D. Fraiman, P. Balenzuela, J. Foss, and D. R. Chialvo, *Phys. Rev. E.* **79**, 061922 (2009).

Identification of informative subgraphs in brain networks

D. Marinazzo[*], G. Wu[*,†], M. Pellicoro[**] and S. Stramaglia[**]

[*]*Department of Data Analysis, Faculty of Psychology and Pedagogical Sciences, University of Ghent, Belgium.*
[†]*Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China.*
[**]*Dipartimento di Fisica, Università degli Studi di Bari and INFN Bari, Italy.*

Abstract. Measuring directed interactions in the brain in terms of information flow is a promising approach, mathematically treatable and amenable to encompass several methods. Here we present a formal expansion of the transfer entropy to put in evidence irreducible sets of variables which provide information for the future state of each assigned target. Multiplets characterized by a large contribution to the expansion are associated to informational circuits present in the system, with an informational character (synergetic or redundant) which can be inferred from the sign of the contribution.

Keywords: mutual Information; transfer entropy; Granger causality.
PACS: 05.45.Tp, 87.19.L-

INTRODUCTION

The inference of couplings between dynamical subsystems, from data, is a topic of general interest. Transfer entropy [1], which is related to the concept of Granger causality [2], has been proposed to distinguish effectively driving and responding elements and to detect asymmetry in the interaction of subsystems. By appropriate conditioning of transition probabilities this quantity has been shown to be superior to the standard time delayed mutual information, which fails to distinguish information that is actually exchanged from shared information due to common history and input signals [3, 4]. On the other hand, Granger formalized the notion that, if the prediction of one time series could be improved by incorporating the knowledge of past values of a second one, then the latter is said to have a *causal* influence on the former. Initially developed for econometric applications, Granger causality has gained popularity also in neuroscience (see, e.g., [5, 6, 7, 8, 9]). A discussion about the practical estimation of information theoretic indexes for signals of limited length can be found in [10]. Transfer entropy and Granger causality are equivalent in the case of Gaussian stochastic variables [11]: they measure the information flow between variables [12]. Recently it has been shown that the presence of redundant variables influences the estimate of the information flow from data, and that maximization of the total causality is connected to the detection of groups of redundant variables [13].

In recent years, information theoretic treatment of groups of correlated degrees of freedom have been used to reveal their functional roles as memory structures or those capable of processing information [14]. Information theory suggests quantities that re-

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 74-84 (2013); doi: 10.1063/1.4776503

veal if a group of variables is mutually redundant or synergetic [15, 16]. Most approaches for the identification of functional relations among nodes of a complex networks rely on the statistics of motifs, subgraphs of k nodes that appear more abundantly than expected in randomized networks with the same number of nodes and degree of connectivity [17, 18].

An interesting approach to identify functional subgraphs in complex networks, relying on an exact expansion of the mutual information with a group of variables, has been presented in [19]. In this work we generalize these results to show a formal expansion of the transfer entropy which puts in evidence irreducible sets of variables which provide information for the future state of the target. Multiplets of variables characterized by an high value, unjustifiable by chance, will be associated to informational circuits present in the system. Additionally, in applications where linear models are sufficient to explain the phenomenology, we propose to use the exact formula for the conditioned mutual information among Gaussian variables so as to get a computationally efficient approach. An approximate procedure is also developed, to find informational circuits of variables starting from few variables of the multiplet by means of a greedy search. We illustrate the application of the proposed expansion to a toy model and two real EEG data sets.

The paper is organized as follows. In the next section we describe the expansion and motivate our approach. In section III we report the applications of the approach and describe our greedy search algorithm. In section IV we draw our conclusions.

EXPANSION

We start describing the work in [19]. Given a stochastic variable X and a family of stochastic variables $\{Y_k\}_{k=1}^n$, the following expansion for the mutual information, analogous to a Taylor series, has been derived there:

$$\begin{aligned} &S(X|\{Y\}) - S(X) = -I(X;\{Y\}) = \\ &\textstyle\sum_i \frac{\Delta S(X)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} + \cdots + \frac{\Delta^n S(X)}{\Delta Y_i \cdots \Delta Y_n}, \end{aligned} \tag{1}$$

where the variational operators are defined as

$$\frac{\Delta S(X)}{\Delta Y_i} = S(X|Y_i) - S(X) = -I(X;Y_i), \tag{2}$$

$$\frac{\Delta^2 S(X)}{\Delta Y_i \Delta Y_j} = -\frac{\Delta I(X;Y_i)}{\Delta Y_j} = I(X;Y_i) - I(X;Y_i|Y_j), \tag{3}$$

$$\frac{\Delta^3 S(X)}{\Delta Y_i \Delta Y_j \Delta Y_k} = I(X;Y_i|Y_k) - I(X;Y_i|Y_j,Y_k) - I(X;Y_i) + I(X;Y_i|Y_j), \tag{4}$$

and so on.

Now, let us consider $n+1$ time series $\{x_\alpha(t)\}_{\alpha=0,\ldots,n}$. The lagged state vectors are denoted

$$Y_\alpha(t) = (x_\alpha(t-m),\ldots,x_\alpha(t-1)),$$

m being the window length.

Firstly we may use the expansion (1) to model the statistical dependencies among the x variables at equal times. We take x_0 as the target time series, and the first terms of the expansion are

$$W_i^0 = -I(x_0;x_i) \tag{5}$$

for the first order;

$$Z_{ij}^0 = I(x_0;x_i) - I\left(x_0;x_i|x_j\right) \tag{6}$$

for the second order; and so on. We note that

$$Z_{ij}^0 = -\mathscr{I}\left(x_0;x_i;x_j\right),$$

where $\mathscr{I}\left(x_0;x_i;x_j\right)$ is the *interaction information*, a well known information measure for sets of three variables [20]; it expresses the amount of information (redundancy or synergy) bound up in a set of variables, beyond that which is present in any subset of those variables. Unlike the mutual information, the interaction information can be either positive or negative. Common-cause structures lead to negative interaction information . As a typical example of positive interaction information one may consider the three variables of the following system: the output of an XOR gate with two independent random inputs (however some difficulties may arise in the interpretation of the interaction information, see [21]). It follows that positive (negative) Z_{ij}^0 corresponds to redundancy (synergy) among the three variables x_0, x_i and x_j.

In order to go beyond equal time correlations, here we propose to consider the flow of information from multiplets of variables to a given target. Accordingly, we consider

$$S(x_0|\{Y_k\}_{k=1}^n) - S(x_0) = -I(x_0;\{Y_k\}_{k=1}^n), \tag{7}$$

which measures to what extent all the remaining variables contribute to specifying the future state of x_0. This quantity can be expanded according to (1):

$$\begin{array}{l} S\left(x_0|\{Y_k\}_{k=1}^n\right) - S(x_0) = \\ \sum_i \frac{\Delta S(x_0)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(x_0)}{\Delta Y_i \Delta Y_j} + \cdots + \frac{\Delta^n S(x_0)}{\Delta Y_i \cdots \Delta Y_n}. \end{array} \tag{8}$$

A drawback of the expansion (7) is that it does not remove shared information due to common history and input signals; therefore we choose to condition it on the past of x_0, i.e. Y_0. To this aim we introduce the conditioning operator $\mathscr{C}_{Y_0}$:

$$\mathscr{C}_{Y_0} S(X) = S(X|Y_0),$$

and observe that $\mathscr{C}_{Y_0}$ and the variational operators (2) commute. It follows that we can condition the expansion (8) term by term, thus obtaining

$$\begin{array}{l} S\left(x_0|\{Y_k\}_{k=1}^n, Y_0\right) - S(x_0|Y_0) = -I\left(x_0;\{Y\}_{k=1}^n|Y_0\right) = \\ \sum_i \frac{\Delta S(x_0|Y_0)}{\Delta Y_i} + \sum_{i>j} \frac{\Delta^2 S(x_0|Y_0)}{\Delta Y_i \Delta Y_j} + \cdots + \frac{\Delta^n S(x_0|Y_0)}{\Delta Y_i \cdots \Delta Y_n}. \end{array} \tag{9}$$

The first order terms in the expansion are given by:

$$A_i^0 = \frac{\Delta S(x_0|Y_0)}{\Delta Y_i} = -I(x_0;Y_i|Y_0), \tag{10}$$

and coincide with the bivariate transfer entropies $i \to 0$ (times -1). The second order terms are

$$B^0_{ij} = I(x_0; Y_i | Y_0) - I\left(x_0; Y_i | Y_j, Y_0\right), \tag{11}$$

and may be seen as a generalization of the interaction information $\mathscr{I}$; hence a positive (negative) B^0_{ij} corresponds to a redundant (synergetic) flow of information $\{i, j\} \to 0$. The typical examples of synergy and redundancy, in the present framework of network analysis, are the same as in the static case, plus a delay for the flow of information towards the target. The third order terms are

$$\begin{aligned} C^0_{ijk} = \;& I\left(x_0; Y_i | Y_j, Y_0\right) + I(x_0; Y_i | Y_k, Y_0) \\ & - I(x_0; Y_i | Y_0) - I\left(x_0; Y_i | Y_j, Y_k, Y_0\right), \end{aligned} \tag{12}$$

and so on.

The generic term in the expansion (9),

$$\Omega_k = \frac{\Delta^k S(x_0 | Y_0)}{\Delta Y_i \cdots \Delta Y_k}, \tag{13}$$

is symmetrical under permutations of the Y_i and, remarkably, statistical independence among any of the Y_i results in vanishing contribution to that order. Therefore each nonvanishing accounts for an irreducible set of variables providing information for the specification of the target: the search for for informational multiplets is thus equivalent to search for terms (13) which are significantly different from zero. Another property of (9) is that the sign of each term is connected to the informational character of the corresponding set of variables, see [19]).

For practical applications, a reliable estimate of conditional mutual information from data should be used. Non parametric methods are recommendable when nonlinear effects are relevant. However, a conspicuous amount of phenomenology in brain can be explained by linear models: therefore, for the sake of computational load, In this work we adopt the assumption of Gaussianity and use the exact expression that holds in this case [11], which reads as follows. Given multivariate Gaussian random variables X, W and Z, the conditioned mutual information is

$$I(X; W | Z) = \frac{1}{2} \ln \frac{|\Sigma(X|Z)|}{|\Sigma(X|W \oplus Z)|}, \tag{14}$$

where $|\cdot|$ denotes the determinant, and the partial covariance matrix is defined

$$\Sigma(X|Z) = \Sigma(X) - \Sigma(X,Z)\Sigma(Z)^{-1}\Sigma(X,Z)^{\top}, \tag{15}$$

in terms of the covariance matrix $\Sigma(X)$ and the cross covariance matrix $\Sigma(X,Z)$; the definition of $\Sigma(X|W \oplus Z)$ is analogous.

The statistical significance of (13) can be assessed by observing that it is the sum of terms like (14) which, under the null hypothesis $I(X; W|Z) = 0$, have a χ^2 distribution. Alternatively, statistical testing may be done using surrogate data obtained by random temporal shuffling of the target vector x_0; the latter strategy is the one we use in this work.

Greedy search of multiplets

Given a target variable, the time required for the exhaustive search of all the subsets of variables, with a statistically significant information flow (13), is exponential in the size of the system. It follows that the exact search for large multiplets is computationally unfeasible, hence we adopt the following approximate strategy. We start from a pair of variables with non-vanishing second order term B w.r.t. the given target. We consider these two variables as a *seed*, and aggregate other variables to them so as to construct a multiplet. The third variable of the subset is selected among the remaining ones as those that, jointly with the previously chosen variable, maximize the modulus $|C|$ of the corresponding third order term. Then, one keeps adding the rest of the variables by iterating this procedure. Calling Z_{k-1} the selected set of k - 1 variables, the set Z_k is obtained adding, to Z_{k-1}, the variable, among the remaining ones, with the greatest modulus of Ω_k. These iterations stop when Ω_k, corresponding to Z_k, is not significantly different from zero (the Bonferroni correction for multiple comparisons is to be applied at each iteration); Z_{k-1} is then recognized as the multiplet originated by the initial pair of variables chosen as the seed.

We apply this strategy to the following toy model

$$\begin{array}{lll} x_0(t) & = a\,\eta(t-1)+\sigma\xi_0(t), & \\ x_\alpha(t) & = b_\alpha\,\eta(t)+\sigma_1\xi_\alpha(t), & \alpha=1,\ldots,m \\ x_\beta(t) & = \sigma_2\xi_\beta(t), & \beta=m+1,\ldots,m+M \end{array} \tag{16}$$

where ξ and η are i.i.d. unit variance Gaussian variables. In this model the target x_0 is influenced by the process η; variables x_α, $\alpha=1,\ldots,m$, are a mixture of η and noise ξ, whilst the remaining M variables are pure noise. Estimates of Ω_k are based on time series, generated from (16) and 1000 samples long. The results are displayed in Fig 5. Firstly we consider the case $m=20$ and $M=0$, with all the twenty variables driving the target with equal couplings b_α; in Fig. 5-A we depict the term Ω_k corresponding to the k-th iteration of the greedy search. We note that Ω_k has alternating sign and its modulus decreases with k. In Fig. 5-B we consider another situation, with $m=10$ and $M=10$, the ten non-zero couplings b_α being non-uniform. Ω_k still shows alternating sign, and Ω_k vanishes for $k>9$; hence the multiplet of ten variables is correctly identified. The order of selection is related to the strength of the coupling: variables with stronger coupling are selected first.

APPLICATIONS

Electroencephalographic recordings

In this subsection we show the application of the proposed expansion, truncated at the second order. To this aim we turn to real electroencephalogram (EEG) data, the window length m being fixed by cross validation. Firstly we consider recordings obtained at rest from 10 healthy subjects. During the experiment, which lasted for 15 min, the subjects were instructed to relax and keep their eyes closed. To avoid drowsiness, every minute

the subjects were asked to open their eyes for 5 s. EEG was measured with a standard 10-20 system consisting of 19 channels [26]. Data were analyzed using the linked mastoids reference, and are available from [27].

For each subject we consider several epochs of 4 seconds in which the subjects kept their eyes closed. For each epoch we compute the second order terms at equal times Z^0_{ij} and the lagged ones B^0_{ij}; then we average the results over epochs. In order to visualize these results, for each target electrode we plot a on a topographic scalp map the pairs of electrodes which are redundant or synergetic with respect to it. Both quantities are distributed with a clear pattern across the scalp. Interactions at equal times are one order of magnitude higher than the lagged interactions, and are dominated by the effect of spatial proximity, see Fig. 1. On the other hand, B^0_{ij} show a richer dynamics, such as interhemispheric communications and predominance redundancy to and from the occipital channels, see Fig. 2, reflecting the prominence of the occipital rhythms when the subjects rest with their eyes closed.

As another example we consider intracranial EEG recordings from a patient with drug-resistant epilepsy and which has thus been implanted with an array of 8×8 cortical electrodes and two depth electrodes with six contacts. The data are available at [29] and described in [28]. For each seizure data are recorded from the preictal period, the 10 seconds preceding the clinical onset of the seizure, and the ictal period, 10 seconds from the clinical onset of the seizure. We analyze data corresponding to eight seizures and average the corresponding results.

For each electrode we compute the lagged influences B^0_{ij}, obtaining for each electrode the pair of other electrodes with redundant or synergetic contribution to its future. The patient has a putative epileptic focus in a deep hippocampal region, with the seizure that then spreads to the cortical areas. In Fig. 3 we report the values of coefficients B taking as the target a cortical electrode located on the putative cortical focus: we report the values of B^0_{ij} corresponding to all the couple of the electrodes, as well as their sum over electrode j. It is clear how the redundancy increases during the seizure. On the other hand, for sensors from 70 to 76, corresponding to a depth electrode, the redundancy is higher in the preictal period, reflecting the fact that the seizure is already active in its primary focus even if not yet clinically observable. The values of B corresponding to this electrode are reported in Fig. 4.

In Fig. 6 we consider again the EEG data from healthy subjects with closed eyes [27], and apply the greedy search taking C3 as the target and $\{C4, C6\}$ as the seed. We find a subset of 9 variables influencing the target. The fact that the sign of Ω_k is alternating, as in the previous model, suggests that the channels in this set correspond to a single source which is responsible for the inter-hemispheric communication towards the target electrode C3. In Fig. 7 we take O1 as the target and $\{F3, C5\}$ as the seed. A subset of 11 variables is found which describes the information flow from the frontal to the occipital cortex.

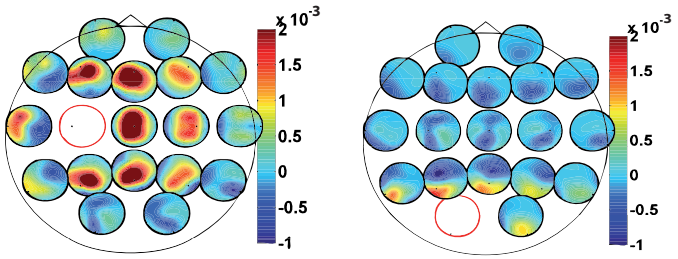


FIGURE 1. The instantaneous components Z_{ij}^0 for two target electrodes, C3 on the left and O1 on the right. The target electrode is in white, and for each of the other electrodes i on the map, the value of Z_{ij}^0 is displayed for the other electrodes.

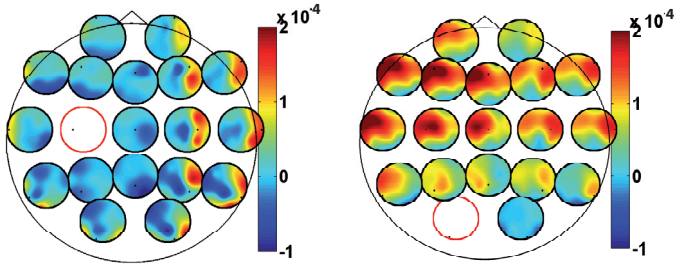


FIGURE 2. The lagged components B_{ij}^0 for two target electrodes, C3 on the left and O1 on the right. The target electrode is in white, and for each of the other electrodes i on the map, the value of B_{ij}^0 is displayed for the other electrodes.

Functional Magnetic Resonance Imaging data

We used two resting state datasets from a public repository [1]. Data were acquired by using of single-shot gradient echo planar imaging (EPI) sequence (repetition time [TR]: 2000ms and 1400ms; echo time: 30ms; slices: 33; thickness: 3mm; gap: 0.6mm; field of view: $200 \times 200mm^2$; in-plane resolution: 64×64; flip angle: 90°). Preprocessing of resting-state imagesl was performed using the Statistical Parametric Mapping software (SPM8, http://www.fil.ion.ucl.ac.uk/spm), including slice-timing corrected relative to middle axial slice for the temporal difference in acquisition among different slices, realigned with the corresponding 3-D structure image, head motion correction(for all subjects, the translational or rotational parameters of a data set did not exceed $\pm 1mm$ or $\pm 1°$), spatial normalization into a standard stereotaxic space, parameters from normalizing 3-D structure images to the Montreal Neurological Institute T1 template in SPM8 were written to fMRI images then resampled to 3-mm isotropic voxels. The functional images were segmented into 90 regions of interest (ROIs) using automated anatomical labeling (AAL) template [22]. For each subject, the representative time series of each ROI was obtained by averaging the fMRI time series across all voxels in the ROI. Sev-

[1] $http://www.nitrc.org/projects/fcon\_1000/$

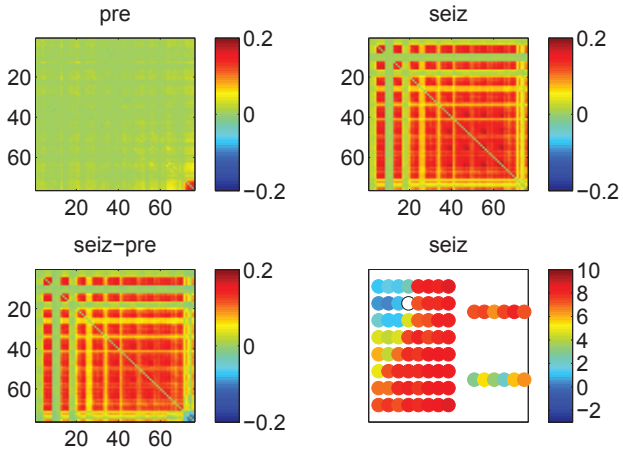


FIGURE 3. The lagged second order terms B^0_{ij} for a cortical electrode (in white) right before and during the clinical onset of a seizure, and the sum over the second electrode of the pair in the lower right panel.

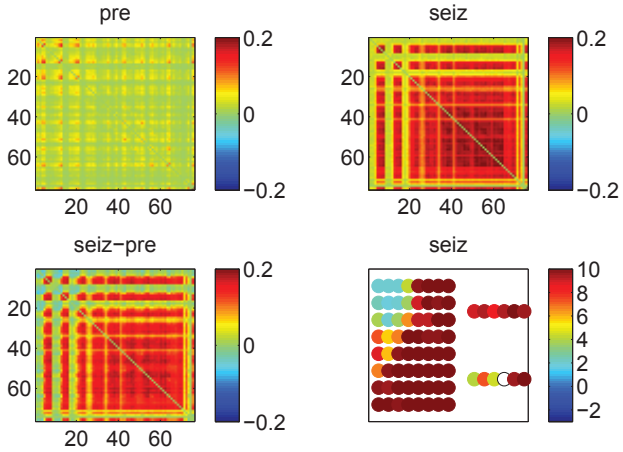


FIGURE 4. The lagged second order terms B^0_{ij} for a depth electrode (in white) right before and during the clinical onset of a seizure, and the sum over the second electrode of the pair in the lower right panel.

eral procedures were used to remove possible spurious variances from the data through linear regression. These were 1) six head motion parameters obtained in the realigning step, 2) signal from a region in cerebrospinal fluid, 3) signal from a region centered in the white matter. 4) global signal averaged over the whole brain. The hemodynamic response function was deconvolved from the BOLD time series.

We start reporting in Fig. 8 the histograms from the first three order terms, computed exactly, confronted with the results obtained by reshuffling the time series. We can observe, in the graph of the third order, that redundant and synergetic contributions seem to have different probability distributions.

Coming to the greedy decomposition, a convenient way to visualize the results is

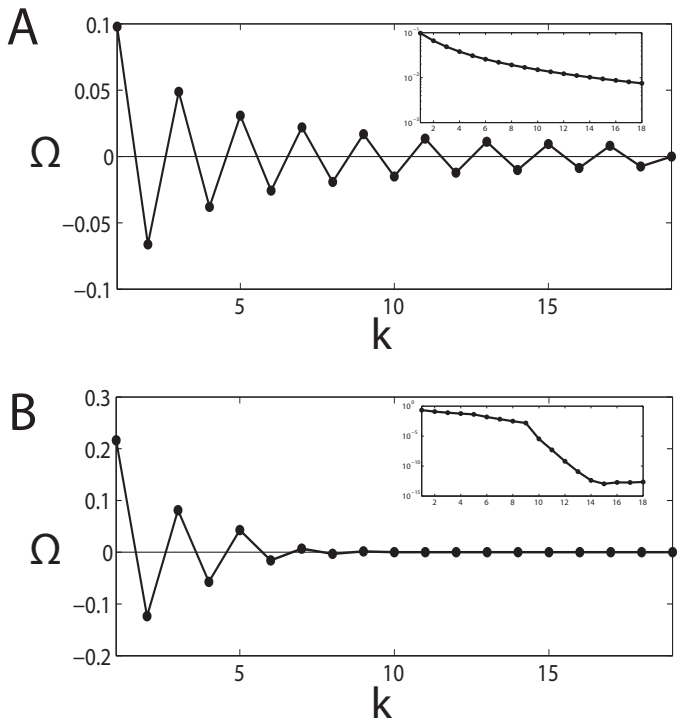


FIGURE 5. Ω_k as a function of the multiplet size k for a model in which one variable is influenced by all the other variables or by part of them. In (A) $m = 20$ and $M = 0$: all the 20 variables influence the target with unitary weight. In (B) $m = 10$ and $M = 10$; the weights b_α are [1.75 1.75 1 1 1 1 .5 .5 .5 .5]. The insets show the logarithm of the absolute value of Ω_k. The first point $k = 1$, in both plots, represents the initial pair of variables chosen as the seed, i.e. $\{1,2\}$. The other parameters are, in both cases, $a, \sigma, \sigma_1, \sigma_2 = 0.5$.

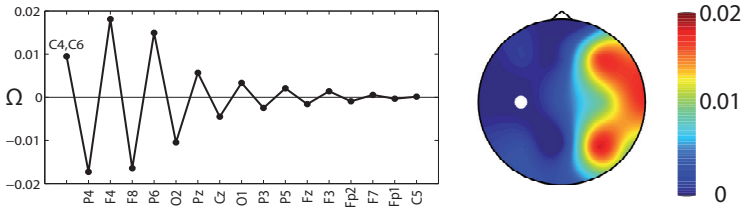


FIGURE 6. Informative contributions to the target electrode C3. Left: information contribute from the resulting multiplet when time series from a given electrode are added to the existing multiplet, starting from the pair (C4,C6) which is the one which shares the most of information on the future of the target time series. Channels P4, F4, F8, P6, O2, Pz and Cz are recognized to belong to the same multiplet as C4 and C6, whilst including $O2$ leads to a Ω_k which is not significantly different from zero. Right: the absolute value of this contributions plotted on a scalp map.

to count, for a given target, how many times another variable appears in redundant of synergetic multiplets. In Fig. 9 we report these findings for the left precuneus, which has been previously reported as a sink hub (mostly receiving information from other brain

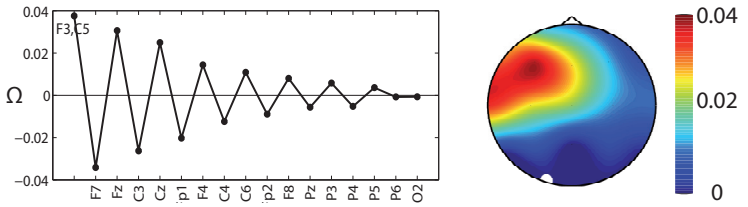


FIGURE 7. Informative contributions to the target electrode O1. Left: information contribute from the resulting multiplet when time series from a given electrode are added to the existing multiplet, starting from the pair (F3,C5) which is the one which shares the most of information on the future of the target time series. Nine channels (F7, Fz, C3, Cz, Fp1, F4, C4, C6, Fp2) are recognized to belong to the same multiplet, for the remaining variables Ω_k is not significantly different from zero. Right: the absolute value of this contributions plotted on a scalp map.

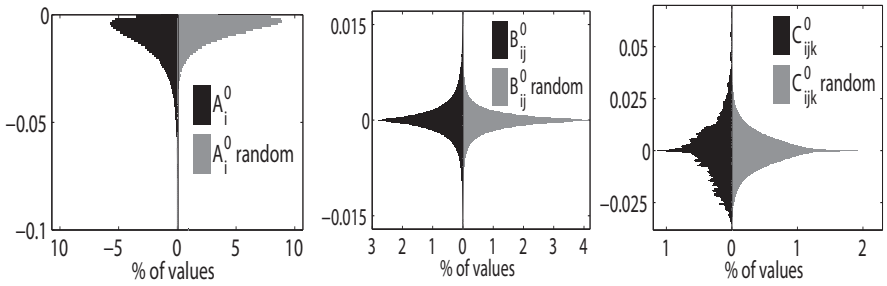


FIGURE 8. Distribution of the first three order terms in the decomposition of the transfer entropy the resting state fMRI. The results are confronted with those obtained reshuffling the time series. Left: first order; Center: second order; Right: third order.

regions)[24]. From the figure is evident the balance between functional segregation and integration in the brain.

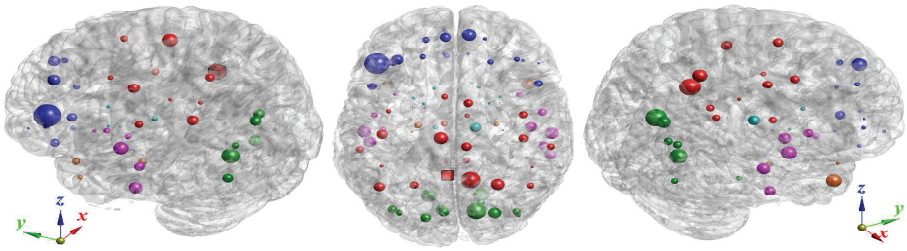


FIGURE 9. Most redundant regions for left precuneus. The size of the regions is proportional to the number of times that the region is present in a redundant multiplet. The color corresponds to each of the six subsystems of the resting brain (Brown: Medial Temporal, Light Blue: Subcortical, Green: Occipital, Dark Blue: Frontal, Purple: Temporal, Red: Parietal-(Pre)motor.

REFERENCES

1. T. Schreiber, *Phys. Rev. Lett.* **85**, 461–464 (2000).
2. C. W. J. Granger, *Econometrica* **37**, 424–438 (1969).
3. M. Staniek, and K. Lehnertz, *Phys. Rev. Lett.* **100**, 158101 (2008).
4. M. Lindner, R. Vicente, V. Priesemann, and M. Wibral, *BMC Neurosci.* **12**, 119 (2011).
5. K.J. Blinowska, R. Kus, and M. Kaminski, *Phys. Rev. E* **70**, 50902–50905(R) (2004).
6. D. A. Smirnov, and B. P. Bezruchko, *Phys. Rev. E* **79**, 46204–46212 (2009).
7. M. Dhamala, G. Rangarajan, and M. Ding, *Phys. Rev. Lett.* **100**, 18701–18704 (2008).
8. D. Marinazzo, M. Pellicoro, and S. Stramaglia, *Phys. Rev. Lett.* **100**, 144103–144107 (2008).
9. L. Faes, A. Porta, and G. Nollo, *Phys. Rev. E* **78**, 26201 (2008).
10. A. Porta, *et al.*, *Methods Inf. Medi.* **49**, 506–510 (2010).
11. L. Barnett, A. B. Barrett, and A. K. Seth, *Phys. Rev. Lett.* **103**, 238701–23704 (2009).
12. K. Hlavackova-Schindler, M. Palus, M. Vejmelka, and J. Bhattacharya, *Phys. Rep.* **441**, 1 (2007).
13. L. Angelini, M. de Tommaso, D. Marinazzo, L. Nitti, M. Pellicoro, and S. Stramaglia, *Phys. Rev. E* **81**, 37201 (2010).
14. A. Borst, and F. E. Theunissen, *Nat. Neurosci.* **2**, 947–957 (1999).
15. E. Schneidman, W. Bialek, and M. J. Berry II, *J. Neurosci.* **23**, 11539–11553 (2003).
16. L. M. A. Bettencourt, *et al.*, *Phys. Rev. E* **75**, 21915–21924 (2007).
17. R. Milo, *et al.*, *Science* 298, 824–827 (2002).
18. E. Yeger-Lotem, *et al.*, *Proc. Natl. cad. Sci. U.S.A.* 101, 5934–5939 (2004).
19. L. M. A. Bettencourt, V. Gintautas, and M. I. Ham, *Phys. Rev. Lett.* **100**, 238701–238704 (2008).
20. W. J. McGill, *Psychometrika* **19**, 97–116 (1954).
21. A. J. Bell, "The co-information lattice", In *Proceedings of ICA 2003,* Nara, Japan, April 2003.
22. N. Tzourio-Mazoyer, *et al.*, *Neuroimage* **15**, 273–289 (2002).
23. R. Salvador, *et al.*, *Cereb. Cortex* **15**, 1332–1342 (2005).
24. A. Cavanna, and M. R. Trimble, *Brain*, **129**(3), pp 564–583 (2006).
25. R. Leech, R. Braga, and D. J. Sharp, *J. Neurosci.* **32**, 215–222 (2012).
26. G. Nolte, A. Ziehe, V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, and K. R. Müller, *Phys. Rev. Lett.* **100**, 234101 (2008).
27. `http://clopinet.com/causality/data/nolte/`, accessed may 2012.
28. M. A. Kramer, E. D. Kolaczyk, and H. E. Kirsch, *Epilepsy Research* **79**, 173 (2008).
29. `http://math.bu.edu/people/kolaczyk/datasets.html`, accessed may 2012.
30. P. L. Williams, and R. D. Beer, "Generalized Measures of Information Transfer", preprint arXiv:1102.1507 (2011).
31. J. T. Lizier, and M. Rubinov, "Multivariate construction of effective computational networks from observational data", preprint MPI MIS Preprint 25/2012.

Signal transmission competing with noise in model excitable brains

J. Marro[*], J. F. Mejias[*,†], G. Pinamonti[*,**] and J. J. Torres[*]

[*] *Institute Carlos I for Theoretical and Computational Physics, and Departmento de Electromagnetismo y Física de la Materia, University of Granada, 18071 - Granada, Spain.*
[†] *Department of Physics and Center for Neural Dynamics, University of Ottawa, K1N-6N5 Ottawa, ON, Canada.*
[**] *Present address: Dipartimento di Fisica, Università degli Studi di Trieste, Italy.*

Abstract. This is a short review of recent studies in our group on how weak signals may efficiently propagate in a system with noise-induced excitation-inhibition competition which adapts to the activity at short-time scales and thus induces excitable conditions. Our numerical results on simple mathematical models should hold for many complex networks in nature, including some brain cortical areas. In particular, they serve us here to interpret available psycho–technical data.

Keywords: complex neural networks; adaptive synapses; stochastic multiresonance; excitable brains.
PACS: 05.40.-a, 87.10.Mn, 87.10.Rt, 87.18.Sn, 87.18.Tt, 87.85.D-, 89.75.-k

INTRODUCTION

Transmission in natural media is sometimes observed to be spontaneously optimized by the system itself, so that weak signals are able to go through without damping while overcoming an apparently predominant noisy environment. Analysis of the relevant situations have in the past associated this outstanding feature of nature with various interesting phenomenologies (see, for instance, [1]-[7] and references therein), including enhancement of dynamic range, stochastic resonance, coherence among noise and signal, nonequilibrium phase transitions, etc.

The actual perspective after three decades of effort suggests that such assorted phenomenology always seems originated in the same basic mechanism, namely, kind of excitability which is in practice due to competition between opposing, say, excitatory/inhibitory tendencies. Such a competition concerns, for example, the biophysical processes driving dynamics of actual synapses in the nervous systems. This significantly influences the transmission of information which, encoded in spike trains or in waves of action potentials generated in a given neuron reaches other neurons. As a matter of fact, the postsynaptic response to incoming stimuli has been demonstrated to be an activity–dependent process in such a way that, for instance, may be enhanced and/or depressed at short time scales depending on the stimuli frequency [8, 9, 10].

We here review some of our recent effort, and also present some new material on this topic as communicated in the 12th Granada Seminar. We first focus on the influence of different possible basic hypothesis on the transmission of weak signals by using simple, familiar partial–differential–equation models [11, 12]. The information thus collected

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 85-93 (2013); doi: 10.1063/1.4776504

is then applied to study transmission in complex networks [13], which is relevant to many fields of science [14, 15] including neuroscience [16, 17]. The result is a detailed description of the microscopic mechanisms that seem to originate, and the relevant phenomenology that has been reported to accompany lack of dissipation and resonances during the propagation of signals through diverse excitable media. We report qualitative agreement of our main model with available experimental data.

A FEED-FORWARD MODEL

A simple *perceptron* type of model will first serve us to correlate different basic hypothesis with emergent behavior. Let us assume a unit, which one may interpret as one of the *neurons* in a network, acted on by a set of other N units or *neurons,* as illustrated in Fig. 1A. The relation is via complex links, channels or *synapses* which induce excitability. Specifically, following [18], for instance, we assume that any link, $i = 1,\ldots,N$, has r_i random components —to be interpreted as neurotransmitter emitting vacuoles, for example— and that each is activated with probability u and stays silent afterwards for a time interval τ, this being a random variable with exponential distribution $p_t(\tau)$ of mean $\tau_{\rm rec}$ at time t. The activation of i at time t_i generates a current, $I_i(t)$, that evolves according to

$$\frac{dI_i(t)}{dt} = -\frac{I_i(t)}{\tau_{\rm in}} + \sum_{\ell=1}^{r_i} J_{i,\ell}\,\delta(t-t_i), \tag{1}$$

where $\tau_{\rm in}$ characterizes the transmission duration, and $J_{i,\ell}$ is the change in component ℓ if this is active which occurs with probability $u[1-p_t(\tau)]$ ($J_{i,\ell}$ is zero otherwise). The variations with i of this change as well as those of r_i may be assumed, for simplicity, to be Gaussian distributed with mean and standard deviation (J,Δ_J) and (r,Δ_r), respectively. The total current is then $I_N(t) = \sum_i I_i(t)$, and the voltage in response to an input signal $A(t)$ may be written as $dV/dt = F(V,I_N,A)$, i.e., a dynamics that depends on the nature of each unit which is determined by function F. A familiar choice is the *integrate–and–fire* (IF) model with refractory period $\tau_{\rm refrac}$ in which F is linear with V ([19]; see also, for instance, [20]). However, this means a fixed firing threshold θ, which is a poor description for most purposes [21, 22]. Therefore, one may alternatively assume that units are of the FitzHugh–Nagumo (FHN) type [23, 24, 25]. It follows, which may be more realistic [26], that the threshold unit adapts to the input current. The resulting dynamics is

$$\frac{dV(t)}{\varepsilon\,dt} = V(t)\,[V(t)-a]\,[1-V(t)] - W(t) + \frac{\rho}{\varepsilon\,\tau_m}\,[A(t)+I_N(t)], \tag{2}$$

where ρ is a resistance to transform current into voltage, and $W(t)$ is a slow variable, accounting for the refractory time of the unit, which satisfies $dW(t)/dt = bV(t) - cW(t)$. The (presumed weak) signal may be chosen as $A(t) \equiv A_0\cos(2\pi f_{\rm signal}t)$ but other choices, including Poissonian trends have been used in our studies [27]. A situation of interest is one in which all the, say, *presynaptic neurons* are in a steady state in which they fire independently, uncorrelated from each other with certain mean frequency $f_{\rm noise}$.

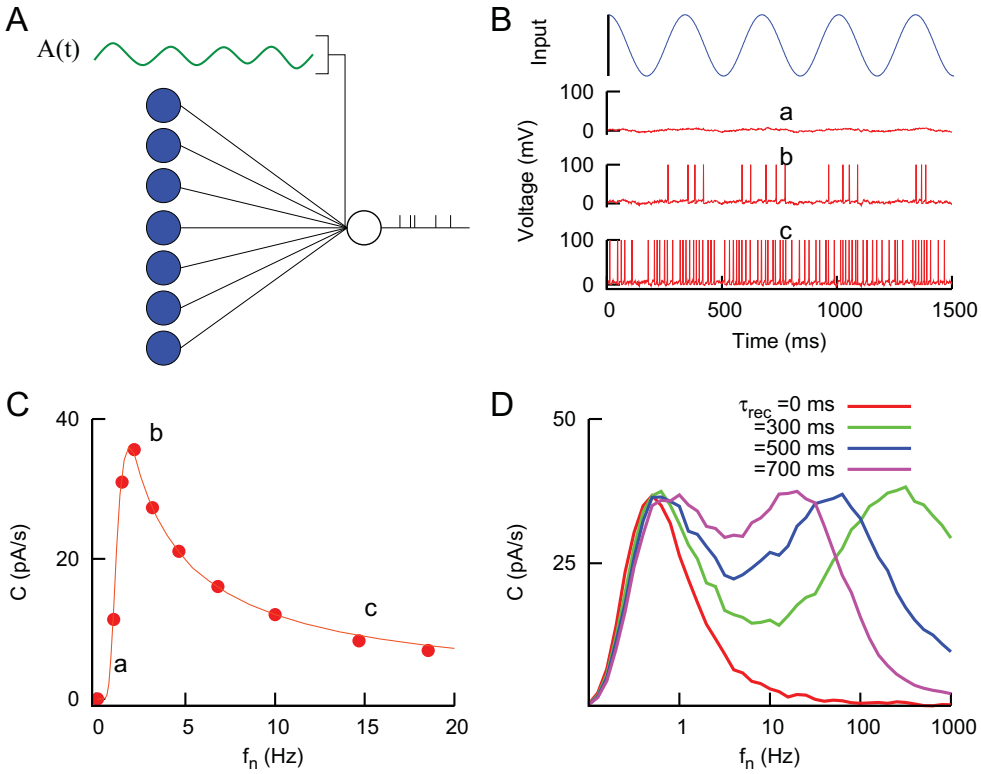


FIGURE 1. (A) Feed-forward, *perceptron* setting in which N units (full circles) transmit noisy, generally uncorrelated activity to a single unit (empty circle) which simultaneously receives a weak structured signal $A(t)$. (B) Time variation of the input signal (top) and response series at low (a), medium (b) and high (c) levels of noisy, for static links and IF neurons. (C) For the same case, the input-output correlation *versus* $f_{\rm noise}$ indicating the corresponding ranges of noise as in the previous graph. (D) The same for dynamic links and FHN neurons as one varies $\tau_{\rm rec}$. Other details are given in the main text and in [27].

The result is a noisy I_N with a difficult dependence on $f_{\rm noise}$. The question is then how $A(t)$ of frequency $f_{\rm signal}$ can go through as one modifies $f_{\rm noise}$.

A convenient measure of this is the input-output correlation function:

$$C(f_{\rm noise}) = \frac{1}{\Delta t}\int_{t_0}^{t_0+\Delta_t} m(t)A(t), \tag{3}$$

for small Δt, where $m(t)$ is the firing rate —mean value of activity; see next section— of the, say, *postsynaptic neuron*. Figures 1B and 1C illustrate the behavior of the system when links are static, namely, $u = 1$ and $\tau_{\rm rec} = 0$, in which case $p_t(\tau)$ transforms into a delta function and $J_{i,\ell}$ is constantly nonzero. The first of these figures shows typical outputs for the voltage at the postsynaptic neuron as one modifies the firing frequency of the presynaptic population, $f_{\rm noise}$. When this is very small, as in (a), the generated current I_N on the postsynaptic neuron can only induce sub-threshold behavior which is weakly correlated if at all with $A(t)$. Increasing $f_{\rm noise}$ increases both I_N and its fluctuations so that a condition is reached in which the postsynaptic neuron fires so frequently, as in (c), which obscures $A(t)$. However, there is an intermediate value of $f_{\rm noise}$, as in (b),

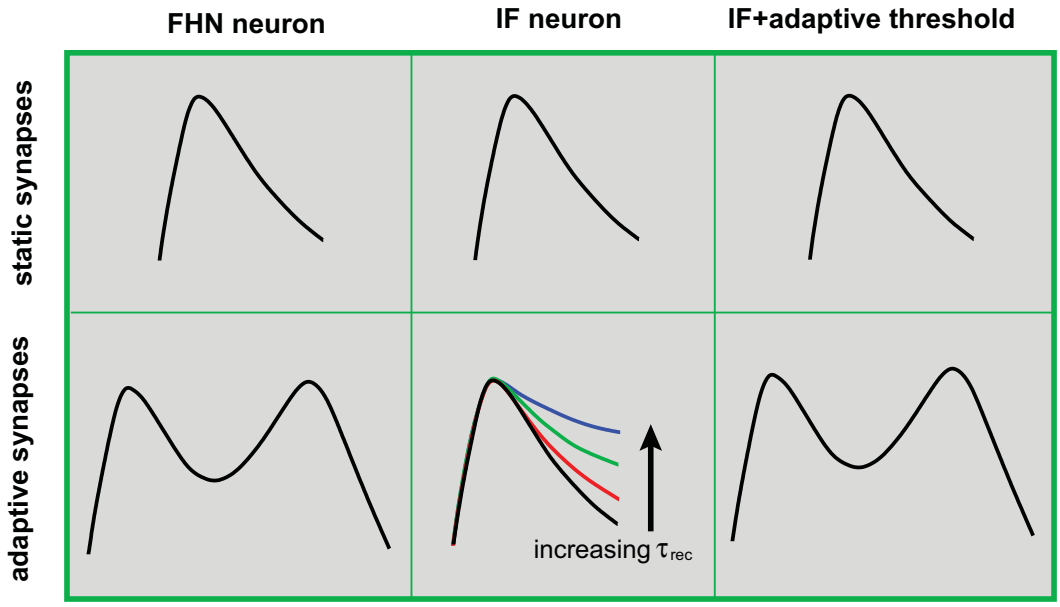


FIGURE 2. Schematic form of the input-output correlation function in transmission experiments *versus* the level of ambient noise. The first row is for static links while the second one concerns links with short-time adaptation to the system activity. The three columns correspond to different neuron models as indicated. See details in the main text.

at which the neuron fires strongly correlated with $A(t)$, that is, several action potentials are emitted each time $A(t)$ is at a maximum. This interesting behavior, often known as stochastic resonance, is clearly revealed by the peak of the function $C(f_{\text{noise}})$ in Fig. 1C. The situation is even more interesting when the links have an active dynamic behavior and neurons are of the FHN type as described above. This is illustrated in Fig. 1D showing, for a range of τ_{rec} values, that two maxima of the correlation then occur so that the transmission is optimized for two ranges of rather well separated noise frequencies. The same occurs if, in addition to synaptic depression as in these examples, we assume facilitation [12]. The precise location of the high and low frequency peaks depends on the level of depression and facilitation which is adopted at the links. Interesting enough, this indicates one how to get a better design of devices for the controlling of signals, for instance.

A systematic study of emergent behavior using variations of our model above reveals the situation illustrated in Fig. 2. A summary is that, under rather general conditions a weak signal may successfully compete with noise and show stochastic resonance —as in the cases in Fig. 2 corresponding to the three graphs in the upper row and to the center one in the bottom row— while transmission through several different levels of noise requires some special conditions. That is, it is not sufficient to have complex dynamic links but the relevant units need also have adaptive thresholds as, for example, in the FHN neuron model.

A main question is wether this behavior, namely, undamped transmission of weak signals in some cases, which in the above abstract model is due to kind of local excitability, may spontaneously ensue from cooperation in complex attractor networks. To explore this issue, let us consider next a networked set of binary units, say, $\mathbf{s} = \{s_i\}$, where s_i at node i is assigned either 0 or 1. The links or *synapses* $i \leftrightarrow j = 1, ..., N$ between units are of intensity or weight given by $\overline{\omega}_{ij}(t) = \omega_{ij} x_j(t)$. Here, $x_i(t) \in [0,1]$ is a dynamic variable to be determined and $\omega_{ij} = [Np(1-p)]^{-1} \sum_\mu (\xi_i^\mu - p)(\xi_j^\mu - p)$ is the maximum of conductance (ω_{ii} is defined to be 0). This choice somewhat modifies the familiar Hebbian prescription [32]; it still involves a set of patterns, e.g., $\xi^\mu = \left\{\xi_i^\mu = 0, 1\right\}$ with $\mu = 1, \ldots, P$, as if previously stored in the links, but we are specifically concerned here with random patterns having a given value for the symmetry parameter of the set, $p \equiv \left\langle \xi_i^\mu \right\rangle_{i,\mu}$.

The resulting network [13], which can sensibly metaphor a brain cortical area, evolves with time as a parallel, cellular automaton, namely, by stochastic changes of the whole set $\mathbf{s} = \{s_i\}$ at each time according to probabilities:

$$P_i\{s_i(t+1) = \sigma\} = \frac{1}{2} + \left(\sigma - \frac{1}{2}\right) \tanh\left[I_i(t) T^{-1}\right], \ \forall i, \tag{4}$$

where $\sigma = 1$ or 0, and T is the temperature of the underlaying bath which controls the stochasticity of dynamics. Here, $I_i(t) = 2\left[h_i(t) - \theta_i + A(t)\right]$ stands for the total input on each unit, $h_i(t) = \sum_j \overline{\omega}_{ij} s_j(t)$ is the net current from others on unit i, θ_i are thresholds for firing, and $A(t)$ is the (weak) external signal; for simplicity, these are taken as $\theta_i = \frac{1}{2}\sum_j \overline{\omega}_{ij}$ and $A(t) = A_0 \cos(2\pi f t)$. A simple choice to determine the weights $\overline{\omega}_{ij}(t)$ consists in assuming that $x_i(t)$ changes with time according to the map [33]:

$$x_i(t+1) = x_i(t) + \frac{1 - x_i(t)}{\alpha} - \beta x_i(t) s_i(t). \tag{5}$$

This depicts a sawtooth–shaped time change, with α and β measuring the teeth width and depth, respectively, which describes a competition of effects in the channels weights associated to their "fatigue". That is, the link of intensity $\omega_{ij} x_j$ is debilitated as β is increased, while decreasing α makes x to recover its maximum value more rapidly. The channel weight effectively remains constant in practice if such a recovery becomes very fast, so that one sometimes speaks of "$\alpha = 0$" as the limit of static synapses which characterizes the standard Ising and Hopfield cases [34, 32]. The motivation for the ansatz (5) are differential equations —such as the ones we developed in the previous section— trying to account for the fact that electrical stimulation due to local and even spatially extended activity may induce short–term plasticity leading to depression and sometimes also facilitation of the channel transmission [8, 35]. As a simpler minded alternative to (5), one may assume that $x_i(t)$ changes so rapidly with time that its action may be described in the relevant time scale by means of the stationary distribution

$$p(x_i) = \zeta(\mathbf{s})\, \delta(x_i - \Phi) + [1 - \zeta(\mathbf{s})]\, \delta(x_i - 1). \tag{6}$$

That is, with probability $\zeta(\mathbf{s})$, which will in general depend on the network activity —in practice, it happens not to be essential wether the model considers either local or global activity here—, the weights are changed by a factor Φ but remain unchanged otherwise. Depending on the value of Φ, this rule may simulate nodes excitability or potentiation or fatigue of the connections as a function of the degree of order in the system. The standard, Hopfield–Hebb model then corresponds to $\Phi = 1$. A convenient choice here happens to be $\zeta(\mathbf{s}) = (1+P/N)^{-1}\sum_{\mu}(m^{\mu})^2$ defined in terms of the overlap $m^{\mu} = [Np(1-p)]^{-1}\sum_i(\xi_i^{\mu} - p)\,s_i$ between the system and each of the stored patterns, μ.

The detailed study of the model using either (5) or (6) shows no essential differences for $A(t) = 0\ \forall t$ [36]. A main result in the specific case (6) is that, tuning properly parameter values, it exhibits familiar equilibrium phases, namely, a disordered high-T phase —corresponding to the *paramagnetic* phase in condensed matter— in which (the stationary values of) all the overlaps are practically zero, a low-T phase with conventional order —corresponding to *ferromagnetism*— in which the global activity converges with time towards one of the attractors $\left\{\xi_i^{\mu}\right\}$, so that it is often taken as a model example of associative memory, and a —say, *spin-glass*— phase in which convergence is towards a mixture of stored patterns. In addition, and most interesting, the system may be tuned to exhibit nonequilibrium phases [34]. Namely, (*i*) one in which there is a rapid and rather irregular roaming among the attractors —thus closely mimicking, for example, long-time structural changes and oscillations that have been associated with reaction–diffusion phenomena in physics and chemistry, as well as efficient, say, *states of attention* that are of interest in neuroscience—, (*ii*) one which is mainly characterized by oscillations between one of the stored patterns and its negative or corresponding *antipattern*, and (*iii*) one with quite irregular, apparently chaotic roaming randomly interrupted by pattern–antipattern oscillations [36]. The case (5) induces similar though relatively simpler behavior, e.g., the most involved behavior (*iii*) does not seem to fully develop in this case.

The relevant order in this system may be described by monitoring the firing rate, explicitly defined as $m(t) = N^{-1}\sum_i s_i(t)$, which is in fact sometimes recorded in laboratory experiments. Though hardly experimentally accessible, also interesting to illustrate in detail the system behavior is the overlap of the actual state with each pattern μ, $m^{\mu}(t)$ as defined above. Furthermore, we are also interested in the input-output correlation which is defined now as

$$C_f = \lim_{\Delta t\to\infty}\frac{1}{\Delta t}\int_{t_0}^{t_0+\Delta t} m(t)e^{ift}dt, \tag{7}$$

i.e., the Fourier coefficient at frequency f of the output firing rate. The relevant correlation, to be denoted $C(T)$ in the following, is signal dependent, e.g., we define it in the cosinus case as the value of $C(f,T) \equiv \left|C_f\right|^2/A_0^2$ computed at the frequency of the input signal. Some main results of our study are summarized in Fig. 3 [37].

This confirms occurrence of stochastic resonance in complex networks, and it allows one to associate main features of the phenomenon to details of the system, namely, variations of the network link weights which induce some essential excitability. The simplest situation of a single resonance peak, as in Fig. 3A, is for attractor networks with fixed connection weights. The peak in this case occurs at a relatively large level of noise

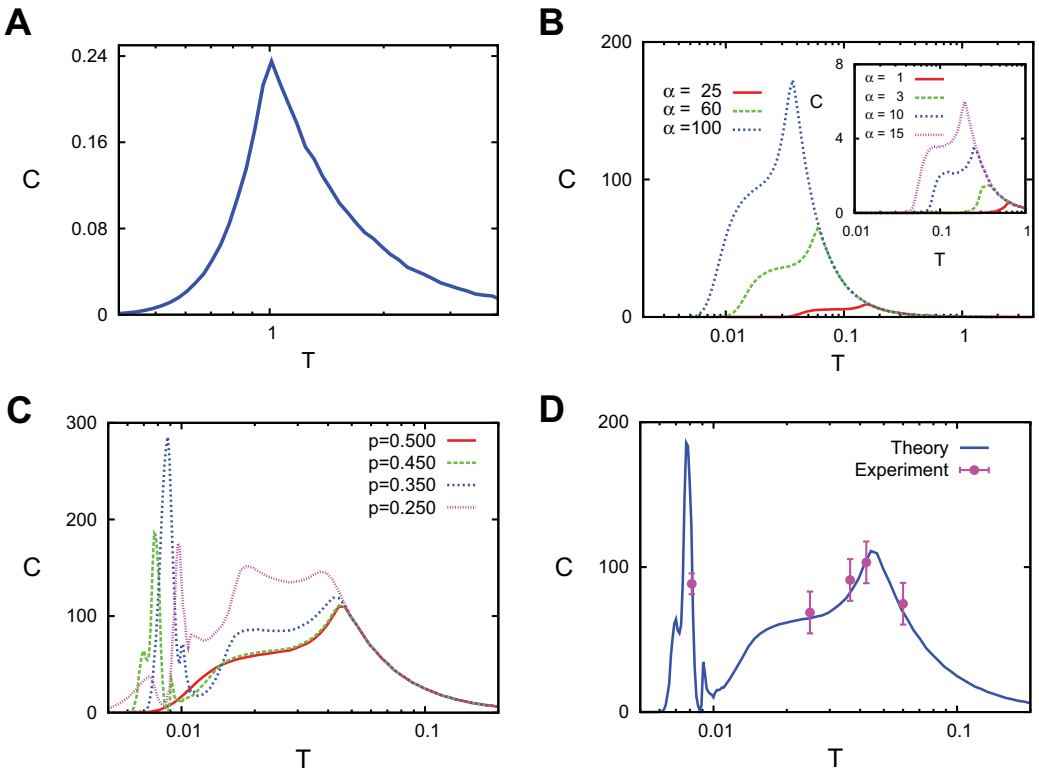


FIGURE 3. The input-output correlation $C(T)$ for the complex network model with a single stored pattern having assymetry parameter p. (A) The case with static links and $p = \frac{1}{2}$. (B) Same but with links affected by synaptic depression. Increasing parameter α (different curves) the resonance peak moves towards lower noise levels. A plateau suggests tendency to a new resonance at low T. (C) Links here are depressed with $\alpha = 80$. Varying p (different curves) allows for varios resonances. (D) Recent data from an experiment concerning eye blinks in the presence of auditory noise are compared here with our prediction for $\alpha = 80$ and $p = 0.45$. See main text and [37] for details.

(corresponding in this model to the underlying bath temperature), around the familiar second-order transition between memory and disordered phases. The potential barrier separating local minima is not too high in this case, and the noise helps the weak signal in overcoming it thus driving the system activity and producing a maximum of correlation. This is not very realistic, however; networks are generally complex in the sense that connection weights are not homogeneous nor constant with time, often adapting to the activity, in such a way that functional connections have in fact abnormal, often power-law distributions [14]-[17]. Allowing for such situation, as we did above with rather general, still expected realistic activity-dependent links, the interesting behavior illustrated in Figs. 3B and 3C, which includes the so-called *stochastic multi-resonance* phenomena emerges. For instance, Fig. 3B shows how synaptic depression induces a change of the resonance peak toward lover values of the noise as well as the emergence of a plateau of relatively high activity-signal correlation at an even lower level of noise. Interesting enough, this announces the possibility of having further resonances, which is confirmed when our model is implemented with another realistic feature, namely, when the set of patterns which determines the mean values of the connection weights

are assumed to be asymmetric, as is certainly always expected to be the case in nature (where the symmetry $p = \frac{1}{2}$ is quite difficult in practice). Such asymmetry together with short-term depression of connections induce oscillations of $m(t)$ at low noise levels as in a first-order phase transition resulting in correlation with the signal near the transition point. Even a third peak appears in the plateau for a high degree of asymmetry, e.g., $p = \frac{1}{4}$ in our numerical experiments, as may correspond to many actual situations. Note that simplicity dictated our illustrations to be concerned with just a single stored pattern, but we also checked that the relevant phenomenology remains when increasing the number of patterns and using other types of signals.

Unfortunately, even though the transmission of signals in difficult conditions which we are interested in here bears, both theoretically and from a practical point of view a great interest, and one may think of various related experiments which do not seem to be specially involved, the fact is that there is no much data to contrast with our theory. An exception is a psycho-physical simple experiment on eye blink reflex in the presence of auditory noise [38]. This shows resonance and, as Fig. 3D shows, we are able to fit that data to our theory by transforming the level of auditory noise in dB into our noise parameter T using a nonlinear relationship; we refer to [13] for details.

CONCLUSION

We have briefly reviewed some recent results concerning phenomena, including stochastic resonance associated with the transmission of signals competing with noise. Our study involves two types of models, namely, partial differential equations and complex networks, and focus on the effects of assuming time-varying connections which depend on the current activity and transform the system into an excitable one. Though the network model in "A Complex Network" section may appear to be rather different from the simpler one in "A Feed-Forward Model" section, so that they may seem to induce resonances by different basic mechanisms, this is not so; for instance, the concern on the pattern asymmetry in the former influences the firing threshold θ_i, and therefore the network excitability as it was also the case in the latter. The result in both cases is intriguing emergent behavior that shows varied multi-resonances that may easily be tuned by changing the models parameters. In addition to a well-defined theoretical reference, our study thus opens a way to many applications. We also present a first contact with experimental data. No doubt new related experiments would be very useful at this moment.

ACKNOWLEDGMENTS

This work was supported by the Junta de Andalucía, project FQM–01505, and the MICINN–FEDER, project FIS2009–08451.

REFERENCES

1. R. Benzi, A. Sutera, and A. Vulpiani, *J. Phys. A: Math. Gen*. **14**, L453–L457 (1981).
2. K. Wiesenfeld, and F. Moss, *Nature* **373**, 33–36 (1995).
3. P. Jung, and G. Mayer-Kress, *Phys. Rev. Lett.* **74**, 2130–2133 (1995).
4. B. Lindner, J. García-Ojalvo, A. Neiman, and L. Schimansky-Geier, *Phys. Rep.* **392**, 321–424 (2004).
5. F. Sagués, J. M. Sancho, and J. García-Ojalvo, *Rev. Mod. Phys.* **79**, 829–882 (2007).
6. Sen Pei *et al*, *Phys. Rev. E* **86**, 021909 (2012).
7. Ch. Lotze *et al*., *Science* **338**, 779–782 (2012).
8. M. V. Tsodyks, and H. Markram, *Proc. Natl. Acad. Sci. USA* **94**, 719–723 (1997).
9. L. F. Abbott, and W. G. Regehr, *Nature* **431**, 796–803 (2004).
10. J. M. Cortes, J. J. Torres, J. Marro, P. L. Garrido, and H. J. Kappen, *Neural Comput.* **18**, 614–633 (2006).
11. J. J. Torres, J. Marro, and J. F. Mejias, "*New J. of Phys.* **13**, 053014 (2011).
12. J. F. Mejias, and J. J. Torres, *PLoS One* **6**, e17255 (2011).
13. G. Pinamonti, J. Marro, and J. J. Torres, *PLoS One* in press (2012).
14. O. Kinouchi, and M. Copelli, *Nature Phys.* **2**, 348–351 (2008).
15. E. D. Bullmore, and O. Sporns, *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
16. C. Castellano, S. Fortunato, and V. Loreto, *Rev. Mod. Phys.* **81**, 591–646 (2009).
17. V. M. Yakovenko, and J. B. Rosser, Jr., *Rev. Mod. Phys.* **81**, 1703–1725 (2009).
18. J. de la Rocha, R. Moreno, and N. Parga, *Neurocomputing* **58-60**, 313–319 (2004).
19. L. Lapicque, *J. Physiol. Pathol. Gen.* **9**, 620–635 (1907).
20. W. Gerstner, and W. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity*, Cambridge Univ. Press, 2002.
21. S. Shinomoto, Y. Sakai, and S. Funahashi, *Neural Comput.* **11**, 935–951 (1999).
22. J. Feng, *Neural Netw.* **14**, 955–975 (2001).
23. R. FitzHugh, *Biophys. J.* **1**(6), 445–466 (1961).
24. J. Nagumo, S. Arimoto, and S. Yoshizawa, *Proc. IRE,* **50**, 2061–2071 (1962).
25. E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*, The MIT Press, Cambridge, MA, 2007.
26. G. G. Izus, R. R. Deza, and H. S. Wio, *Phys. Rev. E* **58**, 93–98 (1998).
27. The parameter values used for the perceptron-like model in the figures are $f_{\text{signal}} = 5$ Hz, $\tau_{\text{m}} = 10$ ms, $\tau_{in} = 3$ ms and $N = 200$, and we specifically used in Figs. 1B and 1C $A_0 = 10$ pA, static synapses corresponding to $J_{i\ell} = J = 48$ pA, $u = 1$, $\tau_{\text{rec}} = 0$ and $r_i = r = 0$, and IF neurons corresponding to $\rho = 0.1\,\text{G}\Omega$, $\tau_{\text{refrac}} = 5$ ms, $\theta = 10$ mV, and $\varepsilon = 1$. Figure 1D is for $u = 0.5, J = 3 \pm 1$ pA, $r = 50 \pm 0.1$, $\rho = 0.1\,\text{G}\Omega/\text{mV}$, $\varepsilon = 1000\,\text{ms}^{-1}$ which makes the (dimensionless) voltage $V(t) = 1$ correspond to 100 mV and the time variable to be within the ms range, $a = 0.001$, $b = 3.5\,(\text{ms})^{-1}$, $c = 1\,(\text{ms})^{-1}$, and $A_0 = 50$ pA. These values are perfectly consistent with reported physiological ranges; see, for instance, [28, 29, 30, 31].
28. H. Markram *et al.*, *J. Physiol.* **500**, 409–440 (1997).
29. L. Herrero *et al.*, *J. Compar. Neurol.* **401**, 411–428 (1998).
30. Z. Gil, B. W. Connors, and Y. Amitai, *Neuron* **23**, 385–397 (1999).
31. C. Koch, *Biophysics of computation: information processing in single neurons*, Oxford Univ. Press, 1999.
32. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558 (1982).
33. L. Pantic, J. J. Torres, H. J. Kappen, and S. C. A. M. Gielen, *Neural Comput.* **14**, 2903–2923 (2002).
34. J. Marro, and R. Dickman, *Nonequilibrium phase transitions in lattice systems*, Cambridge University Press, Cambridge, 2005.
35. Y. Jimbo, T. Tateno, and H. P. C. Robinson, *Biophys. J.* **76**, 670–678 (1999).
36. S. de Franciscis, J. J. Torres, and J. Marro, *Phys. Rev. E* **82**, 041105 (2010).
37. The parameter values used for the complex network model in Fig. 3 are $N = 1000$ and $f_{\text{signal}} = 0.04$ $(\text{MCS})^{-1}$ and, for each specific plot, as follows: $\alpha = 0$ (A), variable (B) and 80 (C and D), $\beta = 0.5$ (B, C and D) and it is not relevant for A, $p = 0.05$ (A and B), variable (C) and 0.45 (D), and $A_0 = 0.005$ (A) and 0.01 (B, C and D).
38. H. Yasuda, T. Miyaoka, J. Horiguchi, A. Yasuda, Peter Hanggi, and Y. Yamamoto, *Phys. Rev. Lett.*, **100**, 118103 (2008).

Does the phenomenon of stochastic amplification of fluctuations play a relevant role in cortical dynamics?

J. Hidalgo*, L. F. Seoane†, J. M. Cortés** and M. A. Muñoz*

*Departamento de Electromagnetismo y Física de la Materia e Instituto de Física Teórica y Computacional Carlos I. Universidad de Granada, E-18071 Granada, Spain.
†Bernstein Center for Computational Neuroscience, Technische Universität Berlin, Germany.
**IKERBASQUE, The Basque Foundation for Science, E-48011, Bilbao, Spain. Biocruces Health Research Institute. Hospital Universitario de Cruces. E-48903, Barakaldo, Spain.

Abstract. We review the mechanism of stochastic amplification of fluctuations in the context of fast cortical oscillations observed during up-states both *in vitro* and *in vivo*. For this purpose, we employ minimalistic models based on short-time synaptic depression with or without synaptic facilitation and compare results with empirical observations. The phenomenon of stochastic amplification of fluctuations is shown to be relevant and robust against different regulatory mechanisms and model specificities. In particular, by introducing synaptic facilitation as a possible manner to dynamically tune the synaptic efficacy, we show that, beyond resonancy details, the mechanism responsible for stochastic amplification is robust and persists along a wide range in the synaptic parameters space. In passing, we explain why a similar stochastic amplification cannot possibly be observed in cortical down-states.

Keywords: brain oscillations; Up and Down states; fluctuations; noise induced phenomena; collective behaviour.
PACS: 87.10.Mn, 87.18.Tt, 87.19.L-, 87.19.lc, 87.19.lm, 87.19.ln

INTRODUCTION

Deciphering the diverse patterns of global activity recorded in the brain and associating them with behavioral states are major challenges in Neuroscience [1, 2, 3]. High-frequency neural activity in the β and γ ranges ($10 : 100$ Hz) has been related to a plethora of cognitive tasks such as working memory [4], selective attention [5], or response to sensory cues [6]; abnormal fast oscillations have been implicated in seizures and pathologies [7]. On the other hand, slow *delta* waves ($0.5 : 2$ Hz) become preponderant during the deepest stages of sleep, under anesthesia or even during quiet wakefulness [8, 9, 10] and might play an important role in neural plasticity and in the consolidation of new memories [11].

Remarkably, neural activity can be spontaneously generated at the cortical-network level even in the absence of external stimulation. For instance, slow δ waves have been observed both *in vivo* and in slice preparations under different experimental protocols [12, 13, 14, 15] in the form of *up-and-down states* in which a large fraction of neurons alternate between two different stable membrane-potential states: the *down-state* –with a high degree of hyper-polarization and very low activity– and the depolarized *up-state*, with high synaptic and spiking activity[16, 9]. The coherent alternation between such

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 94-100 (2013); doi: 10.1063/1.4776506

two, up and down states gives rise to low frequency δ waves [17] with exciting possible implications [18].

Two interesting empirical observations that we would like to clarify are the following:

- high-frequency oscillations have been observed to occur *within* the active (up) intervals of slow oscillations but not in down states [19, 20, 21]. In particular, the associated up-state power spectra develop a non-trivial peak at some frequency in the β band, between 20 and $30Hz$, together with a substantial increase in the spectral power all along the β/γ range [22].
- while global network measurements reveal robust oscillations in the β/γ range in the up-state, individual membrane potentials or synaptic events detected at the intracellular level do not show any trace of oscillations in this range of frequencies.

This suggests, first that oscillations are a collective phenomenon emerging at the network level, and second, that there is no global synchronization locking the rhythms of individual neurons to the systemic one.

Trying to shed light on these issues –and following two recent publications by Wallace et al. [23] and ourselves [24]– we employ minimalistic models able to generate up and down states. As we shall illustrate by working with simple models with short-time synaptic plasticity, the mechanism of stochastic amplification of fluctuations explains in a parsimonious and elegant way all the above-mentioned phenomenology.

STOCHASTIC AMPLIFICATION OF FLUCTUATIONS IN A SIMPLISTIC MODEL FOR UP-AND-DOWN STATES

Minimalistic models for neural dynamics are those in the seminal works of Wilson and Cowan [25] and Amari [26]. These represent the activity of the system through a global (or "mean-field") variable –the population-averaged firing rate– and assume a deterministic evolution for it. Models for network bistability require of some regulatory mechanism such as synaptic depression [27, 28] or a balanced combination of excitatory and inhibitory neurons [29, 30, 31, 32, 33], providing for a negative feedback loop and thus allowing for network self-regulation. These can be easily implemented in simple deterministic models as the above-mentioned ones, allowing for a description of up and down states as corresponding to fixed points of high and low firing-rate, respectively.

Spontaneous transitions between these two deterministic states (i.e. fixed points of the dynamics) can be mimicked by switching-on some stochasticity, able to induce transitions between them. A simple instance of this is the work of Holcman and Tsodyks [27], who introduced a noise source into a simple dynamical model for neural with activity-dependent synaptic plasticity [34]: the noiseless version of the model presents bistability while the noisy version exhibits up-and-down states (see also [35]).

The model of **Markram and Tsodyks** [36, 34] is described by the mean voltage potential, v, and the variable x, which measures the level of available synaptic resources (e.g. neurotransmitters). A dynamical equation for x allows us to model short-time synaptic depression (STSD). The mean voltage grows owing to both external and internal inputs and decreases owing to leakage processes; synaptic resources are consumed in the process of generating internal outputs and spontaneously recover to a target max-

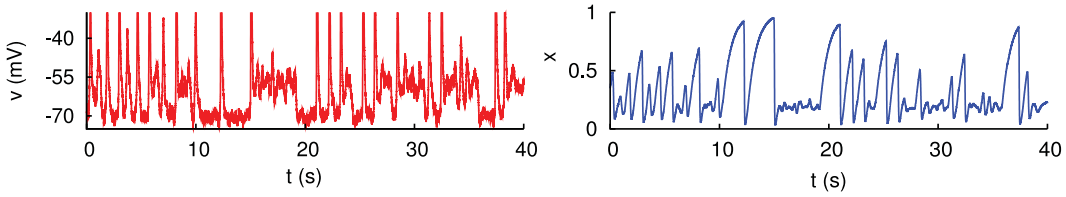


FIGURE 1. Average potential and synaptic utility in computer simulations of the model of Markram and Tsodyks; the system exhibits up and down transitions. In this simulation, noise variances are $\sigma_v = 2.2$ mV/$\sqrt{\tau}$, $\sigma_x = 0$, and time-step 10^{-4} s

imum value, fixed here to $x = 1$:

$$\dot{v} = g_v(v,x) = \frac{1}{\tau}(v_r - v + \omega u x f(v)) \quad (1)$$

$$\dot{x} = g_x(v,x) = \frac{1-x}{\tau_r} - u x f(v), \quad (2)$$

where τ and τ_r are the characteristic times of voltage-leakage and synaptic-recovery, respectively, v_r is the resting potential, ω the amplitude of internal inputs, u the release probability of the neurotransmitters, and –finally– the firing rate function, $f(v)$, is assumed to be of the form $f(v) = \alpha(v - T)$ if $v \geq T$, where T is a threshold value, and $f(v) = 0$ otherwise. Physiologically plausible parameters values are given by $\tau = 0.05$ s, $\tau_r = 0.8$ s, $v_r = -70$ mV, $\omega = 12.6$ mV/Hz, $u = 0.5$, $T = 2.0$ mV, and $\alpha = 1.0$ Hz/mV. For the chosen parameters, there are two stable fixed points (as well as a saddle-point between them). One of them corresponds to a sustained up-state $v^* = v_r + 12.7865$ and $x^* = 0.18817$, and the other to a down-state $v^* = v_r$ and $x^* = 1$ (see Fig. 1). The system experiments a Hopf-bifurcation when decreasing ω, appearing a stable limit cycle with sustained oscillations [27, 37]. This set of equations is deterministic; as in [27], we add noise to have into account some possible stochastic sources (such as irregular external inputs, finite size, or irregular and limited connectivity to name but a few). For simplicity, we add uncorrelated Gaussian white noises $\eta_v(t)$, $\eta_x(t)$ with respective variances σ_v, σ_x, but the forthcoming results do not depend crucially on this choice.

To analyze fluctuations around either of the fixed points, we define $\delta v = v - v^*$ and $\delta x = x - x^*$, and linearize around any of the fixed points:

$$\begin{aligned} \dot{\delta v} &= a_{vv}\delta v + a_{vx}\delta x + \eta_v(t) \\ \dot{\delta x} &= a_{xv}\delta v + a_{xx}\delta x + \eta_x(t), \end{aligned} \quad (3)$$

where $a_{zz'} = \frac{\partial g_z}{\partial z'}(v^*, x^*)$ (z and z' standing for either v or x) are the elements of the Jacobian matrix A evaluated either at the up or the down state. A standard lineal stability analysis of A for the used parameters reveals that the down-state is a node (two real negative eigenvalues) while the up-state is a focus (two complex eigenvalues with negative real part). Defining $\widetilde{z}(\omega) = \mathscr{F}[z(t)]$ as the Fourier transform of either $\delta v(t)$ and $\delta x(t)$, it is straightforward to compute their power spectrum as $P_z(\omega) = \langle |\widetilde{z}(\omega)|^2 \rangle$,

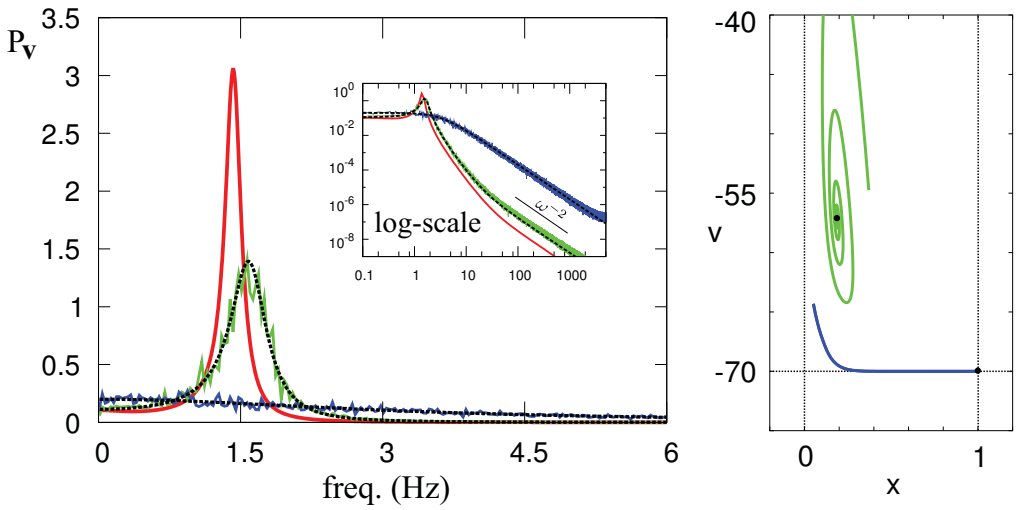


FIGURE 2. Holcman-Tsodyks' model, left: Power spectrum of fluctuations in up- and down-states for average membrane potential v. The main plots show the power-spectra in lineal scale: a marked peak appears for the up-state (green curves) near ≈ 1.5 Hz. Instead, there is no track of similar peaks for down states (blue curves). Observe the excellent agreement between simulation results (noisy curves) and analytical results, Eq. (4) (black dashed lines). Red curve represents the power spectrum for the up-state when synaptic facilitation is incorporated to the model. The peak moves slightly to a lower frequency, and spectrum becomes sharper. The inset represents double-logarithmic plots of the same quantities as in the main plots; in all cases there is a tail w^{-2} revealing the presence of fluctuations at many different scales. All spectra have been generated with $\sigma_z^2 = 0.01 z^*/\tau$, and normalized to unit area. Right: deterministic trajectories for the model without facilitation using different initial conditions. The up-state is a focus (complex eigenvalues), while the down-state is a node (real eigenvalues).

which takes the usual form:

$$P_z(\omega) = \frac{\alpha_z + \sigma_z^2 \omega^2}{\left[\Omega^2 - \omega^2\right]^2 + \Gamma^2 \omega^2} \tag{4}$$

where $\alpha_z = a_{zz'}^2 \sigma_{z'}^2 + a_{z'z'}^2 \sigma_z^2$ and $\sigma_z^2 = \langle \eta_z^2 \rangle$; while $\Omega^2 = \det(A)$ and $\Gamma = \mathrm{Tr}(A)$ do not depend on the noise amplitude. The resulting $P_v(\omega)$ is represented in Fig. 2. Observe that in the limit of small noise amplitude, both spectra exhibit a maximum nearby $\omega = \sqrt{\Omega^2 - \Gamma^2/2}$ (where the denominator of Eq. (4) has a minimum) assuming that the solution of the previous equation exists. Observe the presence of a non-trivial peak for the up-state spectrum, indicating the existence of noise-induced quasi-cycles. This effect, called *stochastic amplification of fluctuations*, has been recently applied in the context of population oscillations in Ecology[38] and other fields such as Epidemiology [39]. It requires the presence of a focus (more specifically, complex eigenvalues with $\mathrm{Im}[\lambda] > \mathrm{Re}[\lambda]$) in the deterministic dynamics plus some additional source of stochasticity. In a nutshell, the gist of the mechanism is as follows: the system tries to relax to the fixed point, but noise "kicks" it away, amplifying some frequencies which are closely related (but not identical) to that of the deterministic damped oscillations. On the other hand, if the system decays towards a deterministic node (i.e. with real eigenvalues), no frequencies are amplified whatsoever (see Fig. 2). This is what happens in

the down-state, where the crossed coupling terms vanish when $f(v) = 0$; accordingly, the derivative of the denominator in Eq. (4) $w^2 = \Omega^2 - \Gamma^2/2 = -(a_{vv}^2 + a_{xx}^2) < 0$ does not vanish for any real value, resulting in the absence of a non-trivial peak in the power-spectrum.

As a second step, we can check the robustness of the mechanism of Stochastic Amplification in the model with synaptic plasticity when we introduce short-term synaptic facilitation. Following Tsodyks and Markram[36], we write a new equation for the release probability of available neurotransmitters, $u = u(t)$ which was taken to be a constant above. Without activity, it recovers to its baseline U_0 with time constant τ_f, while in the presence of activity it increases proportionally to $(1-u)$.

$$\dot{u} = \frac{U_0 - u}{\tau_f} + U_0(1-u)f(v). \tag{5}$$

Fixing $U_0 = 0.05$ and $\tau_f = 1.5$ s, we find that the stable fixed point corresponding to the up-state shifts to $v^* = v_r + 12.5921, x^* = 0.2005, u^* = 0.4708$. On the other hand, the down-state remains at $v^* = v_r, x^* = 1, u^* = U_0$. Computing the power spectrum for each variable we can generalize Eq. 4 to the case with an arbitrary number of coupled equations, obtaining

$$P_z(\omega) = \frac{[\mathrm{Adj}(A - i\omega 1)\langle \vec{\eta}\,\vec{\eta}^{\,t}\rangle \mathrm{Adj}(A^t + i\omega 1)]_{zz}}{\det(A^2 + \omega^2 1)} \tag{6}$$

where Adj stands for the adjoint matrix (transpose of the cofactors). In the limit of small noise amplitude, we find a peak at the frequency that minimizes the denominator $\det(A^2 + \omega^2 1)$. Again, a non-trivial peak appears in the spectra only for the up state (at 1.4 Hz), while the distribution becomes sharper even if its structure remains essentially unchanged (see red curve in Fig. 2). Therefore, the mechanism of stochastic amplification of fluctuations described above is robust to the inclusion of synaptic facilitation.

DISCUSSION AND CONCLUSIONS

We have presented a very simple model, including synaptic plasticity as a key ingredient, able to reproduce up and down states. The dynamics is given by two coupled mean-field equations representing the mean activity of the population and the level of synaptic depression (or resources). While the deterministic system presents two attractors, i.e. bi-stability, up and down transitions appear when noise is added to the system.

First, we analyzed the fluctuations around each fixed point by analyzing the power spectrum of fluctuations for each variable. Its structure, or, more precisely, the existence of a peak, depends essentially on the trajectories near the deterministic attractor. If spiral trajectories are present (i.e. focus fixed point) noise may amplify some frequencies in a resonant way, resulting in a pronounced peak in the power spectrum. This can occur in the up-state, where the variables are strongly coupled because of the feedback-loop between excitation and depression. We have verified that the mechanism is robust against the incorporation of synaptic facilitation to the model. Similarly, the same framework

explains why (owing to the effective decoupling of equations) a similar resulting frequency cannot be observe in down states (in accordance with experiments).

These results are in perfect agreement with experimental results (e.g. [22]). However, in order to improve the result for the characteristic frequency a obtain a result closer to empirical values, more detailed models are required. This path has been followed in [24] where –by considering a simple network-version of the model above in which the role of individual neurons can be explicitly followed– we have shown that the peak frequency shifts towards empirically observed values and, more remarkably– that individual neurons follow a rhythm much faster than the emerging collective one, to which they do not lock [22, 24, 23], in excellent agreement with observations.

Summing up, a simple deterministic model able to reproduce up and down states, does also include non-trivial oscillations within the up state but not in the down state when some noise source is switched on. The mechanism of stochastic amplification of fluctuations can explain the structure of the power spectra and other highly non-trivial features of cortical oscillations in a simple, elegant and parsimonious way.

ACKNOWLEDGMENTS

We acknowledge financial support from the Spanish MICINN-FEDER under project FIS2009-08451, from Junta de Andalucía Proyecto de Excelencia P09FQM-4682. L.S. acknowledges the financial support of Fundación P. Barrié de la Maza and funding grant 01GQ1001A.

REFERENCES

1. G. Buzsáki, *Rhythms of the Brain*, Oxford University Press, USA, 2006.
2. P. Dayan, and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, Cambridge, MIT Press, USA, 2006.
3. W. J. Freeman, *Neurodynamics*, Springer, New York, USA, 2000.
4. C. Tallon-Baudry, O. Bertrand, F. Peronnet, and J. Pernier, *J. Neurosci.* **18**, 4244–4254 (1998).
5. P. N. Steinmetz, A. Roy, P. J. Fitzgerald, S. S. Hsiao, K. O. Johnson, and E. Niebur, *Nature* **404**, 187–190 (2000).
6. A. K. Roopun, S. J. Middleton, M. O. Cunningham, F. E. N. LeBeau, A. Bibbig, M. A. Whittington, and R. D. Traub, *Proc. Natl. Acad. Sci. USA* **103**, 15646–15650 (2006).
7. M. M. McCarthy, C. Moore-Kochlacs, X. Gu, E. S. Boyden, X. Han, and N. Kopell, *Proc. Natl. Acad. Sci. USA* **108**, 11620–11625 (2011).
8. M. Steriade, D. McCormick, and T. Sejnowski, *Science* **262**, 679–685 (1993).
9. A. Destexhe, S. Hughes, M. Rudolph, and V. Crunelli, *Trends Neurosci.* **30**, 334–342 (2007).
10. C. C. H. Petersen, T. T. G. Hahn, M. Mehta, A. Grinvald, and B. Sakmann, *Proc. Natl. Acad. Sci. USA* **100**, 13638–13643 (2003).
11. R. Stickgold, *Nature* **437**, 1272–1278 (2005).
12. M. V. Sanchez-Vives, and D. A. Mccormick, *Nat. Neurosci.* **3**, 1027–1034 (2000).
13. R. Cossart, D. Aronov, and R. Yuste, *Nature* **423**, 283–288 (2003).
14. Y. Shu, A. Hasenstaub, and D. A. McCormick, *Nature* **419**, 288 (2003).
15. C.-y. T. Li, M.-m. Poo, and Y. Dan, *Science* **324**, 643–646 (2009).
16. R. Yuste, J. N. MacLean, J. Smith, and A. Lansner, *Nat. Rev. Neurosci.* **6**, 477–483 (2005).
17. C. Wilson, *Scholarpedia Journal* **3**, 1410 (2008).
18. D. A. McCormick, *Curr. Biol.* **15**, R294–R296 (2005).
19. M. Steriade, F. Amzica, and D. Contreras, *J. Neurosci.* **16**, 392–417 (1996).

20. M. Mukovski, S. Chauvette, I. Timofeev, and M. Volgushev, *Cereb. Cortex* **17**, 400–414 (2007).
21. S. Fujisawa, N. Matsuki, and Y. Ikegaya, *Cereb. Cortex* **16**, 639–654 (2006).
22. A. Compte, R. Reig, V. F., Descalzo, M. A. Harvey, G. D. Puccini, and M. V. Sanchez-Vives, *J. Neurosci.* **17**, 13828–13844 (2008).
23. E. Wallace, M. Benayoun, W. van Drongelen, and J. Cowan, *PLoS ONE* **6**, e14804 (2011).
24. J. Hidalgo, L. Seoane, J. Cortes, and M. Munoz, *PLoS ONE* **7**, e40710 (2012).
25. H. R. Wilson, and J. D. Cowan, *Biophys. J.* **12**, 1–24 (1972).
26. S. Amari, *IEEE Trans. Syst. Man Cybern.* **2**, 643–657 (1972).
27. D. Holcman, and M. Tsodyks, *PLoS Comput. Biol.* **2**, e23 (2006).
28. D. Millman, S. Mihalas, A. Kirkwood, and E. Niebur, *Nature Phys.* **6**, 801–805 (2010).
29. N. Parga, and L. F. Abbott, *Front. Neurosci.* **1**, 57–66 (2007).
30. M. Bazhenov, I. Timofeev, M. Steriade, and T. Sejnowski, *J. Neurosci.* **22**, 8691–8704 (2002).
31. A. Compte, M. V. Sanchez-Vives, D. A. McCormick, and X.-J. Wang, *J. Neurophysiol.* **89**, 2707–2725 (2003).
32. O. Melamed, O. Barak, G. Silberberg, H. Markram, and M. Tsodyks, *J. Comput. Neurosci.* **25**, 308–316 (2008).
33. A. Destexhe, *J. Comput. Neurosci.* **27**, 493–506 (2009).
34. M. V. Tsodyks, K. Pawelzik, and H. Markram, *Neural Comput.* **10**, 821–835 (1998).
35. J. F. Mejias, H. J. Kappen, and J. J. Torres, *PLoS ONE* **5**, e13651 (2010).
36. M. V. Tsodyks, and H. Markram, *Proc. Natl. Acad. Sci. USA* **94**, 719–723 (1997).
37. L. Pantic, J. Torres, H. Kappen, and S. Gielen, *Neural Comp.* **14**, 2903–2923 (2002).
38. A. J. McKane, and T. J. Newman, *Phys. Rev. Lett.* **94**, 218102 (2005).
39. D. Alonso, A. J. McKane, and M. Pascual, *J. R. Soc. Interface* **4**, 575–58 (2007).

Learning pattern recognition and decision making in the insect brain

R. Huerta

BioCircuits Institute, University of California San Diego, La Jolla 92032, USA.

Abstract. We revise the current model of learning pattern recognition in the Mushroom Bodies of the insects using current experimental knowledge about the location of learning, olfactory coding and connectivity. We show that it is possible to have an efficient pattern recognition device based on the architecture of the Mushroom Bodies, sparse code, mutual inhibition and Hebbian leaning only in the connections from the Kenyon cells to the output neurons. We also show that despite the conventional wisdom that believes that artificial neural networks are the bioinspired model of the brain, the Mushroom Bodies actually resemble very closely Support Vector Machines (SVMs). The derived SVM learning rules are situated in the Mushroom Bodies, are nearly identical to standard Hebbian rules, and require inhibition in the output. A very particular prediction of the model is that random elimination of the Kenyon cells in the Mushroom Bodies do not impair the ability to recognize odorants previously learned.

Keywords: pattern recognition; decision making; learning; memory formation.
PACS: 87.85.dq, 87.85.Ng, 87.55.de, 87.55.kh, 87.57.nm, 87.64.Aa, 87.85.D-

INTRODUCTION

The process of deciding what action to take based on the current and future expected external/internal state is typically called decision making [1, 2, 3, 4, 5]. There are two key critical information processing components ubiquitous in the decision making process: i) the prediction of one's action on the environment, *i.e.*, regression, and ii) a pattern recognition problem to discriminate situations, *i.e.*, classification. Both tasks require models to substantiate the action of decision making, and the processes and mechanisms by which those models are learned reveal plausible mechanistic explanations of learning in the brain [6, 7, 8].

In this paper we want to elaborate on the decision making mechanisms that require learning using the most primitive form of all sensory modalities: chemical sensing. This is the sensory modality that coexisted with all forms of life on earth, from the living bacterias to the human brain and remains puzzling and enigmatic despite being so primordial. The insect brain is our choice to understand the underpinnings of learning because they rely on the olfactory modality and they are simpler than the mammalian counterparts. Moreover, the main brain areas dealing with olfactory processing are fairly well known due to the simplicity of the structural organization [9, 10, 11, 12, 13, 14, 15, 16], the nature of the neural coding [17, 18, 19, 20, 21, 22, 23, 24, 25, 26], the advent of the genetic manipulation techniques that isolate brain areas during the formation of memories [27, 28, 29, 30], and the extensive odor conditioning experiments that shed light into the dynamics of learning during discrimination tasks [31, 32, 33, 34, 35, 6].

The main areas where we will concentrate our efforts to understand learning are the

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 101-119 (2013); doi: 10.1063/1.4776507

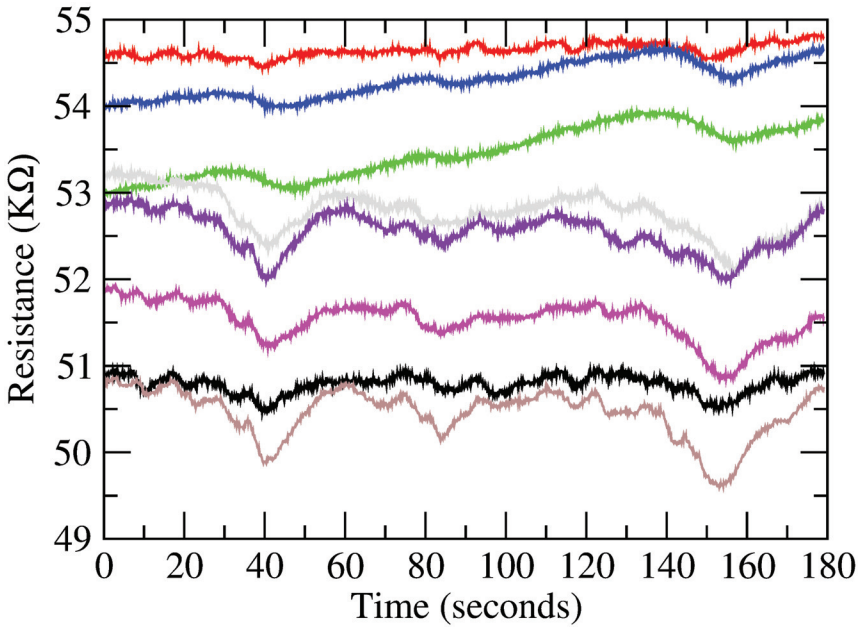


FIGURE 1. Recordings using an artificial sensor array [43, 48, 49, 50, 44] during carbon monoxide presence for 180 seconds in a wind tunnel under turbulent flow.

mushroom bodies [36, 37, 38, 5]. These are responsible for memory formation [36]. There are two additional layers of critical importance for odor processing just in front of the mushroom bodies which are the antennas and the antennal Lobes, but, although memory traces are present [39, 40, 41, 42], their primary function might be signal processing, feature extraction or information filtering [41]. Each of those processing layers are very different in their anatomical and physiological properties. Therefore, since our goal is to understand the mechanisms of learning, we first direct our efforts to the memory formation in the mushroom bodies (MBs) before understanding what specific aspects of the antennal lobe (AL) and the antenna are important.

THE STRUCTURAL ORGANIZATION OF THE INSECT BRAIN

The nature of the olfactory stimulus is stochastic due to the unreliable information carrier. The wind transports gases by turbulent flows that induces complex filaments of gas (see sensor responses in Fig. 1 in [43, 44, 45] and also recordings using a ionization detector in [46, 47]). The nature of the olfactory information differs very markedly from other sensory modalities like vision or audition. The information is intermittent and unreliable, yet evolution has provided to these primitive nervous systems the ability to extract all the necessary information for survival. The brain modules involved in pattern recognition in olfaction are the antennas, the antennal lobes (ALs) and the mushroom bodies (MBs).

Early code

The sensors in the antenna are called the olfactory receptor cells. They are also present in mammals [51] and we still do not have the sensor technology capable of reaching their reaction times, selectivity and stability [48, 49, 50]. Each type of olfactory receptor cell in the antenna connects to a specific glomerulus in the AL [52, 53, 54]. Thus, a chemosensory map of receptor activity in the antenna is represented in the AL. This genetically encoded architecture induces a stimulus-dependent spatial code in the glomeruli [55, 56, 57, 23, 58]. Moreover, the spatial code is maintained across individuals of the same species [59] as would be expected given the genetic structure. In principle this peripheral olfactory structure already seems to be able to discriminate among odors at this early stage. However, the ability to discriminate depends on the number of possible odors, their concentrations, and the complexity in the presence of mixtures [25].

Temporal dynamics in the Antennal Lobe

The antennal lobe receives the input from the olfactory receptor cells that deliver the information into particular sets of glomeruli. The neural network in the AL is made of projection neurons (PNs), which are excitatory, and lateral neurons (LNs), which are mostly inhibitory. The PNs and the LNs connect to each other via the glomeruli. The glomeruli structure induces a bipartite graph of connections that contrasts to the standard directed Bernoulli-induced graphs typically used in AL models [60, 61, 62, 63, 64, 65] with a few exceptions [66]. Moreover, the connections via the glomeruli may be complicated enough because they can be presynaptic [67, 68].

The odor stimuli processed by insects are not constant in time because insects move and the odor plumes flow through the air. The coding mechanism of the AL has to deal with this because the insect needs to detect the odor class, the source and the distance to the odor [16, 69] (see Fig. 2). Since the early works of [70, 71, 72] many experiments have demonstrated the presence of spatio-temporal patterns in the first relay station of the olfactory system of invertebrates and vertebrates [73, 74, 75, 76, 77, 22, 78, 79, 80]. This dynamics results from the interplay of excitatory and inhibitory neurons [74, 81, 82]. There is some debate about the function of temporal coding in behavior, because individuals react faster solving discrimination task than the structure of the temporal code indicates [83, 84, 85]. However there is evidence that by blocking inhibition in the AL insects lack the ability to discriminate between similar odors [33, 74]. And, second, the distance to the source of the odorant is encoded in the intermittentcy of the turbulent flow [43, 44, 45, 46, 47]. The further away from the source the sensors are, the slower the peak frequency of the recordings becomes. Thus, these two facts point out at the need for temporal code to solve pattern recognition (what gas) and regression estimation (how far and where). In fact, these two different functions may use separate pathways in the insect brain [69].

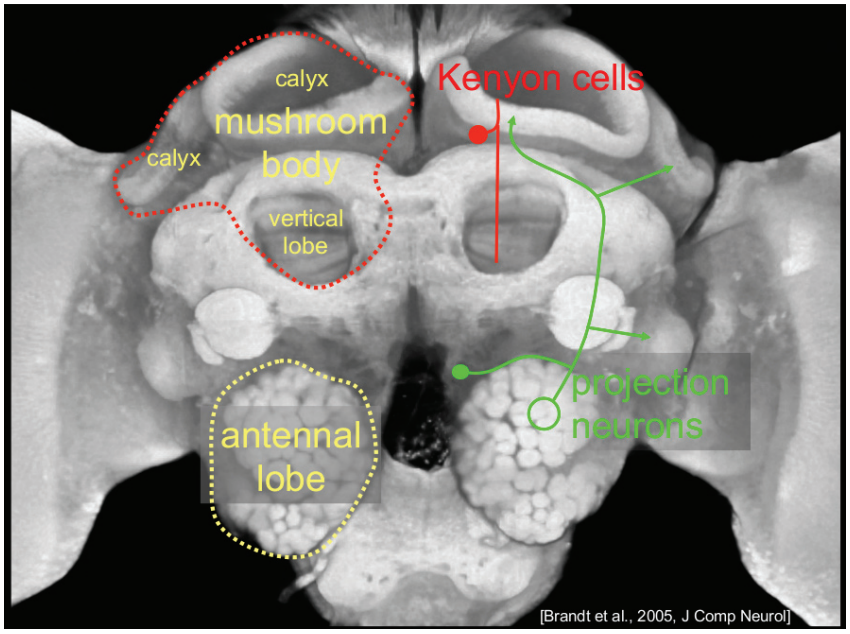


FIGURE 2. Anatomy of the honey-bee brain (courtesy of Robert Brandt, Paul Szyszka, and Giovanni Galizia). The antennal lobe is circled in dashed yellow and the MB is circled in red. The projection neurons (in green) send direct synapses to the Kenyon cells (KCs) in the calyx which is part of the MB.

The Formation of Memories in the Mushroom Bodies

Gain control. It is known that increasing odor concentrations recruit increasing numbers of glomeruli [86] and the activity of the projection neurons within the AL [87, 88, 89]. From the behavioral point of view, increasing the synapses of the local inhibitory neurons makes gases repulsive and to change the number of excitatory synapses makes odorants attractive [90]. This study shows how important the tight equilibrium of the excitatory and inhibitory network is in the AL. A simple transduction of the PN activity into the MB would increase the number of active KCs as well. However, mean firing rates of the PNs that send the output to the MBs have constant firing rates regardless of the gas concentration [91] and recordings of the PNs in the MBs show concentration independence[92][1]. Moreover, the *drosophila* shows that calcium activity is also independent of odor concentration in the KC neurons [30]. Therefore, the activity of the Kenyon cells KCs (see Fig. 2) appears to be heavily regulated and has been shown to generate sparse activity in Honeybees and Locust [94, 95, 96], which is consistent with the overwhelming predictions of associate memory and pattern recognition models [97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107].

[1] Despite there is evidence of gain control at least at the MB level, we still are missing good controls for the concentration of the gases delivered in the antenna [93]. For example, for 1-Hexanol dilution on mineral oil over 1% the concentration of the gas in the air saturates at 200-400 particle per millions.

Learning. Most importantly, the MBs undergo significant synaptic changes [108, 109] and it is known for quite a long time that play a key role in learning odor conditioning [110, 111, 112, 113, 27, 114, 115, 116]. This situates the MBs as the center of learning in the insects. A notion that has been promoted by Martin Heisenberg all along [36].

Inhibition. Another important factor to be able to organize learning and in particular pattern recognition in the Mushroom Bodies is inhibition or more specifically lateral inhibition in the output layer. The notion of lateral inhibition to improve information output has been around for a while [117]. Without lateral inhibition in the output it is not possible to organize competition to have neurons responding for a particular set of stimulus [118, 104]. In fact, it has been shown that there is strong lateral inhibition in the β-lobes in the MBs in [119] which is consistent with standing theoretical models [117, 120, 118, 104, 121].

Temporal code. We have argued that to estimate the distance to the source, temporal processing is required due to the turbulent transport of the gas (see Fig. 1). Further analysis of the data in [95] shows that at the early stages of the processing, right after the stimulus induces a reaction in the insect brain, the MBs can have better ability to discriminate. At later stages, the receptive fields or sensitivity of the MBs become broader [122]. Perhaps at this level slow lateral excitation between the KCs may better encode temporal information of the plume [123]. This implies that the discrimination and recognition of the gas may happen quickly, but the gas concentration estimation requires temporal integration over long time scales as shown in gas source localization using artificial sensor arrays [43, 48, 49, 50, 44].

THE COMPUTATIONAL ORGANIZATION OF THE INSECT BRAIN

In the previous section we tried to succinctly summarize some of the most relevant facts that are needed to build a pattern recognition device in olfaction. These are not by any means all of them and are not necessarily fully consistent with each other, but despite their differences there is more coherence than dissonance with the elements required to have an efficient pattern recognition device. In Fig. 3 we depict the basic model that we use to analyze the computational properties of the MBs.

The simplest model. If we want to understand first the role that the connectivity of the insect brain plays in pattern recognition problems, one has to chose the simplest possible model that complies with the integration properties of neurons. The basic concept is that whenever there is sufficient synaptic input arriving into a neuron, it is going to fire, respond, or transmit information to another group of neurons. A classic model of a neuron that is still successfully used today is the McCullough-Pitts neuron [124]. It is remarkable that it is still used despite being 70 years old and it is used to get estimates of the degree and strength of connections of network architectures to be implemented in

more realistic models[2]. The McCullough-Pitts (MP) neuron is expressed as

$$y_j = \Theta\Big(\sum_{i=1}^{N_{\rm AL}} c_{ji} x_i - \theta_{\rm KC}\Big) \quad j = 1,2,...,N_{\rm KC}. \tag{1}$$

$\mathbf{x}$ is the state vector of the AL neurons (see Fig. 3). It has dimension $N_{\rm AL}$, where $N_{\rm AL}$ is the number of AL neurons. The components of the vector $\mathbf{x} = [x_1, x_2, ..., x_{N_{\rm AL}}]$ are 0's and 1's. $\mathbf{y}$ is the state vector for the KC layer or Calyx; it is $N_{\rm KC}$ dimensional. The c_{ij} are the components of the connectivity matrix which is $N_{\rm AL} \times N_{\rm KC}$ in size; its elements are also 0's and 1's. $\theta_{\rm KC}$ is an integer number that gives the firing threshold in a KC. The Heaviside function $\Theta(\cdot)$ is unity when its argument is positive and zero when its argument is negative. This model can be generalized by replacing the Heaviside function by a nonlinear increasing function and can also be recast in the format of an ordinary differential equation to obtain the a Grossberg-type [126, 127] or Wilson-Cowan models [128].

Advantages and challenges. The MP model is adequate to answer limits in performance of pattern recognition devices for fast operation which is sufficient to account for the fast reliable code observed in the AL [129, 83, 84, 85]. It is also very useful to establish the equivalence with classical pattern recognition devices like the support vector machines (SVMs) [130, 131, 132, 133, 134]. However, it fails at comprehending the role of time in the brain [135, 136, 137, 138, 139] and thus by itself cannot easily solve the regression or distance-to-source estimation problem. Even if the system can recognize efficiently objects, it has to be controlled and regulated within the circuit itself and from other brain areas. This is a challenging problem that we do not address in this paper and requires models with a proper description of the time scales in the brain.

Information Conservation in the Mushroom Bodies

Hypothesis. The main hypothesis is that the Mushroom Bodies are a large screen where one can discriminate objects much more easily. The theoretical basis for discrimination on a large screen to discriminate more easily was already proposed by Thomas Cover [140] and later within the framework of support vector machines [141]. In addition, sparse code is a very useful component to achieving a powerful pattern recognition device as observed in the MBs and as already mentioned the theoretical support for sparse code is extensive [97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107].

Sparse code. The evidence of sparse code in the Calyx is found in the locust [95, 96] and the honeybee [94, 95]. The prevailing theoretical idea to make the code stable over time from the AL to the MB is using forward inhibition [142, 106, 96]. In what follows we will assume that neural circuits are placed in an stable sparse mode.

The AL-MB circuit as an injective function. As we are not addressing the temporal aspects of the system for now, the input for our classification system is an early sin-

[2] See for example the transition from a MP model in [104] to a realistic spiking model in [106] and the model of learning in [125] that resembles closely the model in [105].

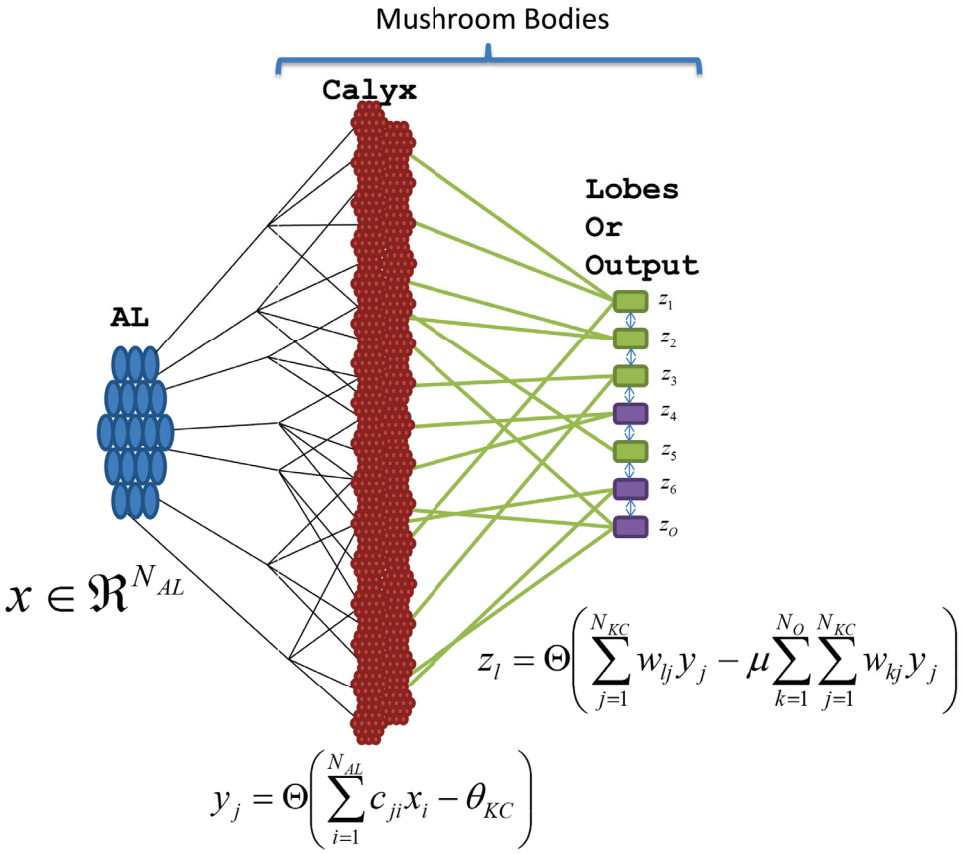


FIGURE 3. The equivalent model of the MBs. We denote by $\mathbf{x}$ as the AL code. $\mathbf{y}$ is the code on the Calyx or the KC neurons, that sometimes we will refer to $\mathbf{y} = \Phi(\mathbf{y})$ in the context of SVMs, and $\mathbf{z}$ is the output of the MBs. Note that all the output neurons inhibit each other with a factor μ.

gle snapshot of information when the antenna hits the plume. The hypothesis for the nonlinear transformation from the AL to the MB is then that every such snapshot or codeword in the AL has a unique corresponding codeword in the MB: The nonlinear transformation needs to be an injective function at least in a statistical sense. In [143] it was proposed to select the parameter values that allow constructing such an injective function from the AL to the KC layer with very high probability.

To determine the statistical degree of injectivity of the connectivity between the AL and KC, we first calculate the probability of having identical outputs given different inputs for a given connectivity matrix: $P(\mathbf{y} = \mathbf{y}' \,|\, \mathbf{x} \neq \mathbf{x}', C)$, where C is one of the possible connectivity matrices (see [143] for details) and the notation $\mathbf{x} \neq \mathbf{x}'$ is $\{(\mathbf{x}, \mathbf{x}') : \mathbf{x} \neq \mathbf{x}'\}$. We want this probability, which we call the probability of *confusion*, to be as small as possible, on the average over all inputs and over all connectivity matrices.

We write this average as $P(\text{confusion}) = \left\langle \langle P(\mathbf{y} = \mathbf{y}' \,|\, \mathbf{x} \neq \mathbf{x}', C) \rangle_{\mathbf{x} \neq \mathbf{x}'} \right\rangle_C$, where $\langle \cdot \rangle_{\mathbf{x} \neq \mathbf{x}'}$ is the average over all non-identical input pairs $(\mathbf{x}, \mathbf{x}')$, and $\langle \cdot \rangle_C$ is the average over all connectivity matrices C. This gives us a measure of *injectivity*, the opposite of confusion,

as

$$I = 1 - P(\text{confusion}), \tag{2}$$

The closer I is to 1, the better is our statistically injective transformation from the states $\mathbf{x}$ of the AL to the states $\mathbf{y}$ of the KCs.

There are two parameters of the model that can be adjusted using the measure of injectivity. One is the probability p_C of having a connection between a given neuron in the AL and a given KC. The second is the firing threshold θ_{KC} of the KCs. Fixed parameters in the model are the probability p_{AL} of having an active neuron in the AL layer, the number N_{AL} of input neurons, and the number N_{KC} of KCs. p_C and θ_{KC} can be estimated using the following inequality

$$I \leq 1 - \{{p_{\mathrm{KC}}}^2 + (1 - p_{\mathrm{KC}})^2 + 2\sigma^2\}^{N_{\mathrm{KC}}}, \tag{3}$$

where p_{KC} is the firing probability of a single neuron in the KC layer. It can be calculated for inputs and connection matrices generated by a Bernoulli process with probabilities p_{AL} and p_C as

$$p_{KC} = \sum_{i=\theta_{KC}}^{N_{AL}} \binom{N_{AL}}{i} (p_{AL} p_C)^i (1 - p_{AL} p_C)^{N_{AL} - i}. \tag{4}$$

where the summatory starts at the the threshold level at which the neurons can fire. This probability has variance (σ^2) for all the prior probabilities of the inputs $\mathbf{x}$ and connectivity matrices. This type of connectivity can be very unstable for perturbations of activity in the input[144]. As can be seen in Fig. 4 where small variations of the probability of activation of AL neurons can lead to a very sharp change in the MBs [145]. This unstability makes necessary to have gain control mechanisms to regulate the sparse activity as proposed in [142, 106, 96, 146] via forward inhibition or by synaptic plasticity [147]. The regulation of sparseness via plasticity from the AL to the MB is an unlikely mechanism to generate sparseness because it actually reduces the information content on the KCs [148].

The formula for the probability of confusion can be intuitively understood if we assume that the activity of every KC is statistically independent from the activity of the others. If so, the probability of confusion in one output neuron is the sum of the probability of having a one for two inputs plus the probability of having a zero for both: $p_{\mathrm{KC}}^2 + (1 - p_{\mathrm{KC}})^2$. Thus, the probability of confusions in all N_{KC} output neurons is $(p_{\mathrm{KC}}^2 + (1 - p_{\mathrm{KC}})^2)^{N_{\mathrm{KC}}}$ in the approximation of independent inputs. This bound on I should be close to unity for any set of parameter values we choose. The inequality for the measure of injectivity becomes an equality for sparse connectivity matrices.

Information preservation versus discrimination and stability. If one takes realistic physiological values [143] one can summarize the expression of confusion just in terms of $n_{KC} \ll N_{KC}$, which is the total number of simultaneously active neurons of the KC layer, as follows:

$$P(\mathbf{x} = \mathbf{x}') \propto e^{-2 \cdot n_{KC}}. \tag{5}$$

which means that ideally to improve injectivity, the system should be placed as far as it can from sparse code reaching the maximum at $n_{KC} = N_{KC}/2$. However, first,

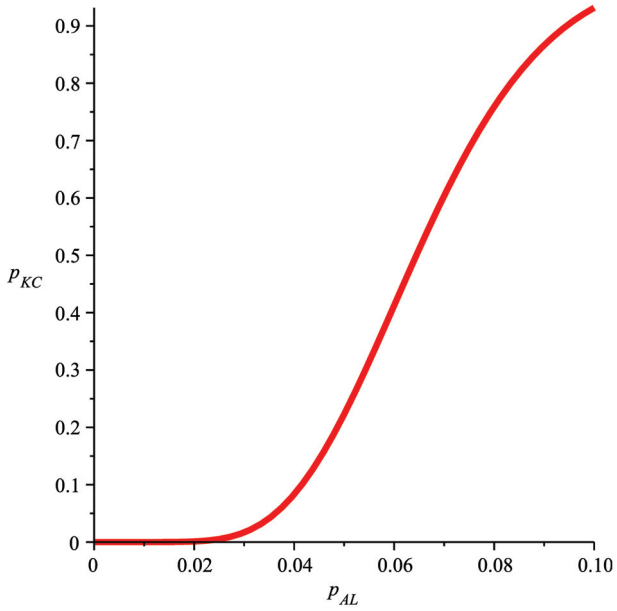


FIGURE 4. Probability of activation of the KC neurons as a function of the probability of the activation off the AL neurons. $N_A = 1000, p_C = 0.15$ and $\theta_{KC} = 10$.

the expression of injectivity in Eq. (3) saturates very rapidly, and, second, in terms of classification performance or memory storage one wants the opposite [97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107]. And in terms of stability as shown in Fig. 4 using realistic parameter values it is difficult to place the MB activity in moderate levels of activity between 10 and 50 percent for all possible inputs and for all the concentration levels (see [145] for details). In fact, a 5% variation in activity in the AL can switch the KCs from sparse activity to having almost all neurons responding to the input. This instability in the statistics of the input is not a desirable property of a pattern recognition device.

Learning Pattern Recognition in the Mushroom Bodies

Rationale. The evidence of learning odor conditioning in the MBs has mounted over the years. Although there has been plasticity shown in the AL network [41], the role played there is data tuning or preprocessing for the pattern recognition device, which is the MB. As we will show the MBs can be shown to be nearly equivalent to Support Vector Machines (SVM) [130] not only in terms of architecture but also in terms of learning. The data tuning or preprocessing in the dynamical system which is the AL can be shown to improve the performance of the pattern recognition device [44]. How this can be carried out in biologically plausible manner remains a mystery.

Beyond olfaction. Despite our main effort has been on olfaction, models of learning in the MBs have increased recently due to the multimodal nature of the MBs [149]. For

example, Wu and Gao's model of decision making of the visual information has the the center of the decision making in the MBs too [8]. The MBs not only have olfactory information but also contextual information, making the MB an integrative center that takes about 35% of the neurons in the insect brain [150].

The model. In [104, 105] we propose a basic model of learning in the MBs which is based on the MP neurons where the key component is to have the output neurons of the MB inhibiting each other (see Fig. 3) as

$$z_l = \Theta\Big(\sum_{j=1}^{N_{\rm KC}} w_{lj}\cdot y_j - \mu \sum_{k=1}^{N_O}\sum_{j=1}^{N_{\rm KC}} w_{kj}\cdot y_j\Big),\ \ l=1,\ldots,N_{\rm O}. \tag{6}$$

Here, the label O denotes the MB Lobes and μ denotes the level of inhibition between the output neurons. The output vector $\mathbf{z}$ of the MB lobes has dimension $N_{\rm O}$. The $N_{\rm KC}\times N_{\rm O}$ connectivity matrix are subjected to Hebbian learning [151] but implemented in a stochastic form. Synaptic changes do not occur in a deterministic manner [152, 153]. Axons are believed to make additional connections to dendrites of other neurons in a stochastic manner, suggesting that the formation or removal of synapses to strengthen or weaken a connection between two neurons is best described as a stochastic process [154, 153].

Output coding and classification decision. We do not know where the final decision or odor classification is taking place in the insect. It is even possible that we will never know because the neural layers from the periphery to the motor action are connected by feedback loops. This intricate connections make difficult to isolate areas of the brain during the realization of particular functions. What we can argue from the theoretical point of view is that the decision of what type of gas is presented outside in the antenna can take place in the output neurons of the MBs, $\mathbf{z}$, with a high odor recognition performance. This performance is much higher than any other location of the layers involved in olfactory processing.

Inhibition in the output. We hypothesized that mutual inhibition exists in the MB lobes and, in joint action with 'Hebbian' learning is able to organize a non-overlapping response of the decision neurons [104]. Recently this hypothesis was verified in [119] showing that the inhibition in the output neurons is fairly strong, and in [108] where plasticity was found from the KC neurons into the output ones.

Reinforcement Signal. One important aspect of learning in the insect brain is that it is not fully supervised. The reward signal is delivered to many areas of the brain as good (octopamine[3]), $r=+1$, or bad (dopamine), $r=-1$. The Mushroom Bodies are innervated by huge neurons that receive direct input from the gustatory areas [155]. They play a critical role in the formation of memories [5] and the can remain activated for long periods of time releasing octopamine into not only the MBs but also the ALs and other areas of the brain. In addition, the delay between the presence of the stimulus and the reward has an impact[156] in learning memories. The learning rules that one can use to have the system learn to discriminate are not unique [105]. For example one can

[3] Note that in the mammalian brain is just the opposite.

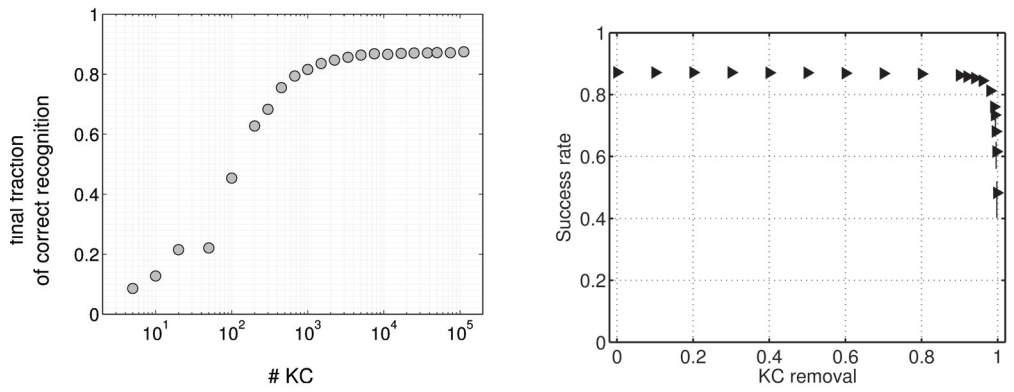


FIGURE 5. (LEFT) Accuracy in the classification of the MNIST handwritten digits for different sizes of the MB. (RIGHT) Success rate as function of random elimination of KCs.

use rules similar to [157] that are contained in the following expression:

$$\Delta w_{ij} \propto y_i\, r(US) z_j \,\text{with}\, P(US, y_i, z_j), \tag{7}$$

where the changes of the synaptic connections between the KCs and the output neurons depend on the activity of both layers and the reward signal $r(US)$ with some probability $P(US, y_i, z_j)$ that depends on the state of the MB and the nature of the unconditional stimulus, or reward signal. US denotes what behavioral experimentalists call unconditional stimulus that is our reward signal. Note that in [105] the values of $P(US, y_i, z_j)$ have a significant impact in the performance of the classifier.

Impact of MB size on accuracy and Robustness. The brain's ability to learn better is thought to be positively correlated with larger brains [158]. Larger brains consume more energy but memorize better and can survive in more complex environments. In [105] we investigated the models given by Eqs. (1,6) and apply them to the MB to solve a very well-known problem in pattern recognition: the MNIST dataset. The MNIST dataset is made of 60,000 handwritten digits for training the model and 10,000 for test [159]. Despite the digits are not obviously gases. The representation of the information in the MB is multimodal [149], so we can analyze the ability to recognize better by exploring larger brain sizes and provide a direct comparison with pattern recognition methods in machine learning. The main results are shown in Fig. 5 where we can show that the ability to have better accuracy in the recognition of digits with increasing brain sizes. The other interesting results of that investigation is the robustness of the MBs to damage or elimination of the KCs. On the right panel of Fig. 5 we can see that one has to eliminate above 99% of the KC neurons to observe a serious impairment of the performance in pattern recognition. This is another prediction. If the insect has learned and been trained previously, damage of the Calyx will not degrade its performance.

EQUIVALENCE BETWEEN THE MUSHROOM BODIES AND SUPPORT VECTOR MACHINES

Using the inhibitory SVM formalism proposed in [134], the synaptic input arriving into an output neuron, $\hat{z}_k$, can be expressed as

$$\hat{z}_k(\mathbf{x}) = \sum_j w_{kj}\Phi_j(\mathbf{x}) - \mu \sum_l \sum_j w_{lj}\Phi_j(\mathbf{x}) = \sum_j w_{kj} y_k - \mu \sum_l \sum_j w_{lj} y_j, \tag{8}$$

where $\Phi_j(\mathbf{x})$ is the nonlinear function that projects the AL code $\mathbf{x}$ into the KC neurons or $\mathbf{y}$. The response of this neuron is a threshold function on $\hat{z}_k$. For the purposes of the SVM what matters is the value of the synaptic input, $\hat{z}_k$, so we will concentrate on its analysis. To make the notation more compact let us write

$$\hat{z}_k(\mathbf{x}) = \hat{z}_k(\mathbf{y}) = \langle \mathbf{w}_k, \mathbf{y} \rangle - \mu \sum_l \langle \mathbf{w}_l, \mathbf{y} \rangle.$$

Note that to make $\hat{z}_k(\mathbf{x}) = \hat{z}_k(\mathbf{y})$ implies that during learning in the SVM there will not be learning from the projections of the AL to the MB.

For the sake of simplicity we consider that the SVM will classify a binary problem. A particular stimulus, $\mathbf{x}$, has a label $r = +1$ for positive label and $r = -1$ for negative label. Now since both the SVM and the honeybee learn by examples let us say that there are a total of N stimulus/examples, $\mathbf{y}_i$, with their corresponding labels, $r_i = +1, -1$ with $i = 1, \ldots, N$. The idea is that to have the classifier working properly then $r_i \hat{z}_k(\mathbf{y}_i) \geq 0$ for all the examples. However, a key concept in SVMs is that the SVM output needs to be above some margin such that $r_i \hat{z}_k(\mathbf{y}_i) \geq 1$. The margin value of 1 is standard although we can chose any value one likes. The most important thing to understand is that the examples belonging to different classes are sufficiently separated from each other. The next important aspect of SVMs is the loss function which is expressed as

$$\min_{\mathbf{w}_k} L = \min_{\mathbf{w}_k} \left(\frac{1}{2} ||\mathbf{w}_k||^2 + C \sum_{i=1}^{N} \max \left\{ 0, 1 - r_i \langle \mathbf{w}_k, \mathbf{y}_i \rangle \right\} \right). \tag{9}$$

The first term is called the regularization term, which is an upper bound to the generalization error [132], the second term corresponds to the measure of classification error using the hinge loss function. The hinge loss is not the most plausible error function because it is known that most of the population are risk averse and, second, the honeybees give more importance to strong odor concentrations than lower ones [160]. The implications of these two empirical observations should lead to interesting consequences that are left for further work, but the intensity of learning could be manipulated by making a a variable margin as $r_i \hat{z}_k(\mathbf{y}_i) \geq \rho(c)$ with c the gas concentration, and the hinge loss function could be replaced by another one with different weights for $r = +1$ and -1.

Following the concept of neural inhibition developed in [134], we can now write a multiclass setting for all the output neurons as

$$\min_{\mathbf{w}_1, \ldots, \mathbf{w}_O} \quad \left(\frac{1}{2} \sum_{i=1}^{O} \langle \mathbf{w}_i, \mathbf{w}_i \rangle + C \sum_{j=1}^{N} \max \left\{ 0, 1 - r_j \langle \mathbf{w}_i, \mathbf{y}_j \rangle - \mu \sum_{k=1}^{O} \langle \mathbf{w}_k, \mathbf{y}_j \rangle \right\} \right), \tag{10}$$

where O is the number of output neurons that compete with each other and the scalar value of the inhibition is optimal for $\mu = 1/O$ [134].

For the sake of simplicity and without loss of generality, we solve for $O = 2$. So we can write the loss function as

$$\min_{\mathbf{w}_1,\mathbf{w}_2} \left(\frac{1}{2}\langle \mathbf{w}_1, \mathbf{w}_1 \rangle + \frac{1}{2}\langle \mathbf{w}_2, \mathbf{w}_2 \rangle \right.$$
$$\left. + C\sum_{j=1}^{N} \max\left\{0, 1 - r_j \frac{1}{2}\langle (\mathbf{w}_1 - \mathbf{w}_2), \mathbf{y}_j \rangle\right\} + C\sum_{j=1}^{N} \max\left\{0, 1 - r_j \frac{1}{2}\langle (\mathbf{w}_2 - \mathbf{w}_1), \mathbf{y}_j \rangle\right\} \right). \tag{11}$$

As already mentioned synaptic changes do not occur in a deterministic manner [152, 153, 154]. To solve the problem (9) one may chose a gradient of the primal as in [161, 162, 163]. The gradient in this case is

$$\frac{\partial E}{\partial \mathbf{w}_k}\Big|_{\mathbf{y}_i} = \begin{cases} \mathbf{w}_k - \frac{C}{2} r_i \mathbf{y}_i & \text{if } r_i \hat{z}_k(\mathbf{y}_i) \leq 1, \\ \mathbf{w}_k & \text{otherwise,} \end{cases} \tag{12}$$

with k taking values 1 and 2. So the regularization term induces a continuous process of synaptic removal that it is well known to improve the generalization ability of the pattern recognition device. This is an important message in the sense that too much memory allows learning all of the data samples used for training but then fails on a new set of examples or stimulations. So a healthy level of memory removal boosts the ability to induce an abstract model. The second term of the gradient indicates that when the example is properly classified above some margin there is nothing to be done. On the other hand, if the stimulus is not properly classified then the synaptic changes have to be modified according to $\Delta w_{kj} \propto y_j r_i$. In other words, the activity levels of the KCs and the sign is determined by the reward signal. The main differences respect to the learning rule in Eq. (7) is that when the stimulus is properly classified above a margin no further changes are required in the connections.

CONCLUSION AND DISCUSSION

The honeybee has no more than a million neurons [150, 164]. 35% of those are in the MBs, which is the main location of learning in the insect brain. Another 20% of those are olfactory sensors, which gives a significant weight on the olfactory modality. Then, in between the olfactory receptor cells and the MBs, the AL just constitutes a 2% of the insect brain. This is the area where the information of the antenna is heavily compressed and then relayed into the MB with a significant reduction in the activity levels. Why is the AL is so important despite being son small compared to other brain areas? What is it doing with the signal: extracting dynamical features, normalizing the activity levels, decorrelating in time different stimulus? We do not know yet but our argument is that to provide an answer to these questions we first need to understand how the MBs work during learning and execution. Once we know, then we can determine aspects of the AL processing that improve the performance in pattern recognition and eventually in decision making.

Evolution and engineers. It is remarkable that when one asks engineers what problems need to be solved in pattern recognition of gases, they propose feature extraction methods to interpret the spatio-temporal signal form the sensors and a classifier and regressor to discriminate between gases and to estimate the concentrations [165, 166, 167]. The bioinspiration is not present in these arguments but yet the insect olfactory system appears to be doing just that. Preprocessing the olfactory receptor signals to extract a sparse code that will be passed to a classifier that resembles a support vector machine. In addition computational models even using this seemingly small number of neurons (a million) are extremely demanding in regular computers. Fortunately we also have alternative simulation methods based on graphics processing units (GPUs). GPUs now allow 10 to 100 fold speed-ups in simulations [168, 169] which makes the simulation of insect brains in full size and real time a possibility, removing the biases of scaled-down simplified models.

The MBs as SVMs. It is also remarkable that in contrast to the mainstream mindset that considers Artificial Neural Networks (ANN) as biologically inspired, the reality is that the paradigmatic back-propagation algorithm has yet to be found in the brain. Support Vector Machines, on the other hand, that have become the gold standard of pattern recognition due to its simplicity and nice properties during convex optimization, are actually biologically plausible, fit perfectly in the general scheme of the insect brain, and explains plasticity as a gradient of a loss function proposed by Vapnik [132]. An expert in statistical learning theory that probably thinks that insects are annoying living things rather than a fascinating puzzle of learning.

Role of Models in Neuroscience. Computational neuroscientists put incredible efforts in building computational realistic models to bridge the gap between theory and neural systems[4]. In the process of building these models they manage to reproduce a large variety of experimental observations that later are often rendered with diminished value due to the lack of predictive power, complexity of the systems and the models themselves. Our approach has been: first to understand the function, which is odor discrimination, pattern recognition and regression; second, to identify the neural architecture that solves the problem; and, third, understand the neural code if data is available. Then, taking that knowledge as constraints, we solve a pattern recognition problem and determine what minimal and simple additional key ingredients are needed to complete the task. We predicted for example strong inhibition in the output of the MB and Hebbian learning from the KCs to the output as it was later found. Another prediction derived from this type of model is robustness. As we can see in Fig. 5 the MB model can sustain heavy damage on the KCs without impairing the ability to classify incoming odors. Obviously, if the Calyx is heavily damaged the ability to learn deteriorates, but the recall power of previously stored memories is retained.

About time and Hebbian reversal. The question of how to use time effectively to better solve classification problems is still puzzling. Even though we know that training dynamical systems together with SVMs can improve performance of the classifiers, the plasticity rules are fairly unrealistic from the biological point of view. Moreover, we

[4] An inspection of ModelDB database illustrates this very clearly `http://senselab.med.yale.edu/modeldb/ListByModelName.asp?c=19&lin=-1`

still do not know whether Hebbian plasticity can actually be reversed in the presence of dopamine or octopamine [170, 157], but from the model and pattern recognition perspective the reversal of Hebbian learning needs to to be present to correct those synapses that are providing the wrong output. So a reversal of the spike timing dependent plasticity rule has to be somehow present when reinforcement signal like dopamine or octopamine is activated.

ACKNOWLEDGMENTS

The author acknowledges partial support from NIDCD R01DC011422-01, and to all the researchers that over the course of several decades have made extraordinary progress in understanding how insects solve the computational problem of olfaction.

REFERENCES

1. S. Grossberg, and W. E. Gutowski, *Psychol. Rev.* **94**, 300–318 (1987).
2. M. Hsu, I. Krajbich, C. Zhao, and C. F. Camerer, *J. Neurosci.* **29**, 2231–2237 (2009).
3. M. I. Rabinovich, R. Huerta, P. Varona, and V. S. Afraimovich, *PLoS Comput. Biol.* **4**, e1000072 (2008).
4. J. I. Gold, and M. N. Shadlen, *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
5. R. Menzel, *Nat. Rev. Neurosci.* **13**, 758–768 (2012).
6. C. R. Gallistel, S. Fairhurst, and P. Balsam, *Proc. Natl. Acad. Sci. USA* **101**, 13124–13131 (2004).
7. B. H. Smith, R. Huerta, M. Bazhenov, and I. Sinakevitch, *Distributed Plasticity for Olfactory Learning and Memory in the Honey Bee Brain*, Springer Netherlands, 2012, pp. 393–408.
8. Z. Wu, and A. Guo, *Neural Netw.* **24**, 333 – 344 (2011).
9. C. G. Galizia, S. L. McIlwrath, and R. Menzel, *Cell Tissue Res* **295**, 383–394 (1999).
10. M. Mizunami, J. M. Weibrecht, and N. J. Strausfeld, *J. Comp. Neurol.* **402**, 520–537 (1998).
11. K. Ito, K. Suzuki, P. Estes, M. Ramaswami, D. Yamamoto, and N. J. Strausfeld, *Learn Mem.* **5**, 52–77 (1998).
12. S. M. Farris, and N. J. Strausfeld, *J. Comp. Neurol.* **439**, 331–351 (2001).
13. J. D. Armstrong, K. Kaiser, A. Müller, K.-F. Fischbach, N. Merchant, and N. J. Strausfeld, *Neuron* **15**, 17–20 (1995).
14. N. J. Strausfeld, *J. Comp. Neurol.* **450**, 4–33 (2002).
15. C. G. Galizia, and B. Kimmerle, *J. Comp. Physiol. A* **190**, 21–38 (2004).
16. J. G. Hildebrand, and G. M. Shepherd, *Annu. Rev. Neurosci.* **20**, 595–631 (1997).
17. R. F. Galan, M. Weidert, R. Menzel, A. V. M. Herz, and C. G. Galizia, *Neural Comput.* **18**, 10–25 (2006).
18. G. Laurent, *Science* **286**, 723–728 (1999).
19. G. Laurent, *Nat. Rev. Neurosci.* **3**, 884–895 (2002).
20. G. Laurent, *Trends Neurosci.* **19**, 489–496 (1996).
21. T. Faber, J. Joerges, and R. Menzel, *Nature Neurosci.* **2**, 74–78 (1999).
22. R. Abel, J. Rybak, and R. Menzel, *J. Comp. Neurol.* **437**, 363–383 (2001).
23. C. G. Galizia, J. Joerges, A. Küttner, T. Faber, and R. Menzel, *J. Neurosci. Meth* **76**, 61–69 (1997).
24. C. G. Galizia, and R. Menzel, *Nature Neurosci.* **3**, 853–854 (2000).
25. N. J. Vickers, T. A. Christensen, T. C. Baker, and J. G. Hildebrand, *Nature* **410**, 466–470 (2001).
26. K. C. Daly, R. F. Galán, O. J. Peters, and E. M. Staudacher, *Front. Neuroeng.* **4** (2011).
27. T. Zars, M. Fischer, R. Schulz, and M. Heisenberg, *Science* **288**, 672–675 (2000).
28. T. Zars, R. Wolf, R. Davis, and M. Heisenberg, *Learn Mem.* **7**, 18–31 (2000).
29. T. Zars, *Curr. Opin. Neurobiol.* **10**, 790–795 (2000).
30. Y. Wang, N. J. D. Wright, H. F. Guo, Z. Xie, K. Svoboda, R. Malinow, D. P. Smith, and Y. Zhong, *Neuron* **29**, 267–276 (2001).

31. M. E. Bitterman, R. Menzel, A. Fietz, and S. Schäfer, *J. Comp. Psychol.* **97**, 107–119 (1983).
32. B. H. Smith, G. A. Wright, and K. C. Daly, "Learning-Based Recognition and Discrimination of Floral Odors," in *Biology of Floral Scent*, edited by N. Dudareva, and E. Pichersky, CRC Press, 2005, chap. 12, pp. 263–295.
33. J. S. Hosler, K. L. Buxton, and B. H. Smith, *Behav. Neurosci.* **114**, 514–525 (2000).
34. B. H. Smith, *Behav. Neurosci.* **111**, 1–13 (1997).
35. B. H. Smith, C. I. Abramson, and T. R. Tobin, *J. Comp. Psychol.* **105**, 345–356 (1991).
36. M. Heisenberg, *Nat. Rev. Neurosci.* **4**, 266–275 (2003).
37. M. Heisenberg, *Learn Mem.* **5**, 1–10 (1998).
38. R. L. Davis, *Neuron* **11**, 1–14 (1993).
39. R. F. Galán, S. Sachse, C. G. Galizia, and A. V. M. Herz, *Neural Comput.* **16**, 999–1012 (2004).
40. L. Rath, C. G. Galizia, and P. Szyszka, *Eur. J. Neurosci.* **34**, 352–360 (2011).
41. F. F. Locatelli, P. C. Fernandez, F. Villareal, K. Muezzinoglu, R. Huerta, C. G. Galizia, and B. H. Smith, *Eur. J. Neurosci.* ((in press)).
42. A. Acebes, J.-M. Devaud, M. Arnés, and A. Ferrús, *J. Neurosci.* **32**, 417–422 (2012).
43. M. Trincavelli, A. Vergara, N. Rulkov, J. Murguia, A. Lilienthal, and R. Huerta, *AIP Conference Proceedings* **1362**, 225 (2011).
44. R. Huerta, S. Vembu, M. Muezzinoglu, and A. Vergara, "Dynamical SVM for Time Series Classification," in *Pattern Recognition*, edited by A. Pinz, T. Pock, H. Bischof, and F. Leberl, Springer Berlin / Heidelberg, 2012, vol. 7476 of *Lecture Notes in Computer Science*, pp. 216–225.
45. M. Muezzinoglu, A. Vergara, N. Ghods, N. Rulkov, and R. Huerta, "Localization of remote odor sources by metal-oxide gas sensors in turbulent plumes," in *Sensors, 2010 IEEE*, IEEE, 2010, pp. 817–820.
46. J. A. Farrell, J. Murlis, X. Long, W. Li, and R. T. Cardé, *Env. Fluid Mech.* **2**, 143–169 (2002), 10.1023/A:1016283702837.
47. K. A. Justus, J. Murlis, C. Jones, and R. T. Cardé, *Env. Fluid Mech.* **2**, 115–142 (2002), 10.1023/A:1016227601019.
48. A. Vergara, S. Vembu, T. Ayhan, M. Ryan, M. Homer, and R. Huerta, *Sens. & Actuat. B: Chem.* **166**, 320–329 (2012).
49. A. Vergara, R. Calavia, R. M. Vázquez, A. Mozalev, A. Abdelghani, R. Huerta, E. H. Hines, and E. Llobet, *Analytical Chemistry* **84**, 7502–7510 (2012).
50. A. Vergara, M. Muezzinoglu, N. Rulkov, and R. Huerta, *Sens. & Actuat. B: Chem.* **148**, 298–306 (2010).
51. L. Buck, and R. Axel, *Cell* **65**, 175–187 (1991).
52. Q. Gao, B. Yuan, and A. Chess, *Nature Neurosci.* **3**, 780–785 (2000).
53. L. B. Vosshall, A. M. Wong, and R. Axel, *Cell* **102**, 147–159 (2000).
54. K. Scott, R. Brady, A. Cravchik, P. Morozov, A. Rzhetsky, C. Zuker, and R. Axel, *Cell* **104**, 661–673 (2001).
55. V. Rodrigues, *Brain Res.* **453**, 299–307 (1988).
56. P. G. Distler, B. Bausenwein, and J. Boeckh, *Brain Res.* **805**, 263–266 (1998).
57. J. Joerges, A. Küttner, C. G. Galizia, and R. Menzel, *Nature* **387**, 285–288 (1997).
58. C. G. Galizia, K. Nägler, B. Hölldobler, and R. Menzel, *Eur. J. Neurosci.* **10**, 2964–2974 (1998).
59. C. G. Galizia, S. Sachse, A. Rappert, and R. Menzel, *Nature Neurosci.* **2**, 473–478 (1999).
60. E. Av-Ron, and J. F. Vibert, *Biosystems* **39**, 241–250 (1996).
61. C. Linster, and B. H. Smith, *Behav. Brain Res.* **87**, 1–14 (1997).
62. M. Bazhenov, M. Stopfer, M. Rabinovich, R. Huerta, H. D. Abarbanel, T. J. Sejnowski, and G. Laurent, *Neuron* **30**, 553–567 (2001).
63. M. Bazhenov, M. Stopfer, M. I. Rabinovich, H. D. I. Abarbanel, T. Sejnowski, and G. Laurent, *Neuron* **30**, 569–581 (2001).
64. A. Capurro, F. Baroni, S. B. Olsson, L. S. Kuebler, S. Karout, B. S. Hansson, and T. C. Pearce, *Front. Neuroeng.* **5** (2012).
65. C. Assisi, and M. Bazhenov, *Front. Neuroeng.* **5** (2012).
66. H. Belmabrouk, T. Nowotny, J.-P. Rospars, and D. Martinez, *Proc. Natl. Acad. Sci. USA* **108**, 19790–19795 (2011).
67. S. Olsen, and R. Wilson, *Nature* **452**, 956–960 (2008).

68. C. M. Root, K. Masuyama, D. S. Green, L. E. Enell, D. R. Nässel, C.-H. Lee, and J. W. Wang, *Neuron* **59**, 311–321 (2008).
69. M. Schmuker, N. Yamagata, M. Nawrot, and R. Menzel, *Front. Neuroeng.* **4** (2011).
70. J. Kauer, *J. Physiol.* **243**, 695–715 (1974).
71. M. Meredith, *J. Neurophysiol.* **56**, 572–597 (1986).
72. J. S. Kauer, and J. White, *Annu. Rev. Neurosci.* **24**, 963–979 (2001).
73. R. W. Friedrich, and G. Laurent, *Science* **291**, 889–894 (2001).
74. M. Stopfer, and G. Laurent, *Nature* **402**, 664–668 (1999).
75. M. Stopfer, S. Bhagavan, B. H. Smith, and G. Laurent, *Nature* **390**, 70–74 (1997).
76. M. Wehr, and G. Laurent, *Nature* **384**, 162–166 (1996).
77. T. A. Christensen, B. R. Waldrop, and J. G. Hildebrand, *J. Neurosci.* **18**, 5999–6008 (1998).
78. D. P. Wellis, J. W. Scott, and T. A. Harrison, *J. Neurophysiol.* **61**, 1161–1177 (1989).
79. Y. W. Lam, L. B. Cohen, M. Wachowiak, and M. R. Zochowski, *J. Neurosci.* **20**, 749–762 (2000).
80. R. I. Wilson, G. C. Turner, and G. Laurent, *Science* **303**, 366–370 (2004).
81. S. Sachse, and C. G. Galizia, *J. Neurophysiol.* **87**, 1106–1117 (2002).
82. R. I. Wilson, and G. Laurent, *J. Neurosci.* **25**, 9069–79 (2005).
83. D. W. Wesson, R. M. Carey, J. V. Verhagen, and M. Wachowiak, *PLoS Biol.* **6**, e82 (2008).
84. G. A. Wright, M. Carlton, and B. H. Smith, *Behav. Neurosci.* **123**, 36–43 (2009).
85. M. F. Strube-Bloss, M. P. Nawrot, and R. Menzel, *J. Neurosci.* **31**, 3129–3140 (2011).
86. M. Ng, R. D. Roorda, S. Q. Lima, B. V. Zemelman, P. Morcillo, and G. Miesenbock, *Neuron* **36**, 463–474 (2002).
87. S. Sachse, and C. G. Galizia, *Eur. J. Neurosci.* **18**, 2119–2132 (2003).
88. C. E. Reisenman, T. A. Christensen, and J. G. Hildebrand, *J. Neurosci.* **25**, 8017–8026 (2005).
89. A. F. Silbering, R. Okada, K. Ito, and C. G. Galizia, *J. Neurosci.* **28**, 13075–13087 (2008).
90. A. Acebes, A. Martin-PeÃśa, V. Chevalier, and A. Ferrús, *J. Neurosci.* **31**, 2734–2745 (2011).
91. M. Stopfer, V. Jayaraman, and G. Laurent, *Neuron* **39**, 991–1004 (2003).
92. N. Yamagata, M. Schmuker, P. Szyszka, M. Mizunami, and R. Menzel, *Front. Syst. Neurosci.* **3**, 16 (2009).
93. J. E. Cometto-Muniz, W. S. Cain, and M. H. Abraham, *Chemical Senses* **28**, 467–477 (2003).
94. P. Szyszka, M. Ditzen, A. Galkin, C. G. Galizia, and R. Menzel, *J. Neurophysiol.* **94**, 3303–3313 (2005).
95. J. Perez-Orive, O. Mazor, G. C. Turner, S. Cassenaer, R. I. Wilson, and G. Laurent, *Science* **297**, 359–365 (2002).
96. M. Papadopoulou, S. Cassenaer, T. Nowotny, and G. Laurent, *Science* **332**, 721–725 (2011).
97. D. Marr, *J. Physiol.* **202**, 437–470 (1969).
98. D. Marr, *Proc. R. Soc. Lond. B Biol. Sci.* **176**, 161–234 (1970).
99. D. Marr, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **262**, 23–81 (1971).
100. D. Willshaw, and H. C. Longuet-Higgins, "Associative memory model," in *Machine Intelligence*, edited by B. Meltzer, and O. Michie, Edinburgh University Press, 1970, vol. 5.
101. S. Amari, *Neural Netw.* **2**, 451–457 (1989).
102. J.-P. Nadal, and G. Toulouse, *Network* **1**, 61–74 (1990).
103. G. Palm, and F. Sommer, *Network* **3**, 177–186 (1992).
104. R. Huerta, T. Nowotny, M. Garcia-Sanchez, H. D. I. Abarbanel, and M. I. Rabinovich, *Neural Comput.* **16**, 1601–1640 (2004).
105. R. Huerta, and T. Nowotny, *Neural Comput.* **21**, 2123–2151 (2009).
106. T. Nowotny, R. Huerta, H. D. I. Abarbanel, and M. I. Rabinovich, *Biol. Cybern.* **93**, 436–446 (2005).
107. V. Itskov, and L. F. Abbott, *Phys. Rev. Lett.* **101**, 018101 (2008).
108. S. Cassenaer, and G. Laurent, *Nature* **448**, 709–713 (2007).
109. R. Okada, J. Rybak, G. Manz, and R. Menzel, *J. Neurosci.* **27**, 11736–11747 (2007).
110. M. Heisenberg, A. Borst, S. Wagner, and D. Byers, *J. Neurogenet.* **2**, 1–30 (1985).
111. J. Mauelshagen, *J. Neurophysiol.* **69**, 609–625 (1993).
112. J. S. de Belle, and M. Heisenberg, *Science* **263**, 692–695 (1994).
113. J. B. Connolly, I. J. Roberts, J. D. Armstrong, K. Kaiser, M. Forte, T. Tully, and C. J. O'Kane, *Science* **274**, 2104–2107 (1996).
114. A. Pascual, and T. Préat, *Science* **294**, 1115–1117 (2001).
115. J. Dubnau, L. Grady, T. Kitamoto, and T. Tully, *Nature* **411**, 476–480 (2001).

116. R. Menzel, and G. Manz, *J. Exp. Biol.* **208**, 4317–4332 (2005).
117. S. Grossberg, *Math. Biosci.* **4**, 255 – 310 (1969).
118. R. O'reilly, *Neural Comput.* **13**, 1199–1241 (2001).
119. S. Cassenaer, and G. Laurent, *Nature* **482**, 47–52 (2012).
120. S. Grossberg, and N. Schmajuk, *Psychobiology* (1987).
121. B. Smith, R. Huerta, M. Bazhenov, and I. Sinakevitch, *Honeybee Neurobiology and Behavior* pp. 393–408 (2012).
122. F. B. Rodríguez, and R. Huerta, *Biol. Cybern.* **100**, 289–297 (2009).
123. T. Nowotny, M. I. Rabinovich, R. Huerta, and H. D. I. Abarbanel, *J. Comput. Neurosci.* **15**, 271–281 (2003).
124. W. S. McCulloch, and W. Pitts, *Bull. Math. Biophys.* **5**, 115–133 (1943).
125. J. Wessnitzer, J. Young, J. Armstrong, and B. Webb, *J. Comput. Neurosci.* **32**, 197–212 (2012).
126. S. Grossberg, *Proc. Natl. Acad. Sci. USA* **58**, 1329–1334 (1967).
127. Y. Cao, and S. Grossberg, *Neural Netw.* **26**, 75–98 (2012).
128. H. R. Wilson, and J. D. Cowan, *Kybernetik* **13**, 55–80 (1973).
129. S. Krofczik, R. Menzel, and M. P. Nawrot, *Front. Comput. Neurosci.* **2** (2009).
130. T. Nowotny, and R. Huerta, "On the equivalence of Hebbian learning and the SVM formalism," in *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*, IEEE, 2012, pp. 1–4.
131. K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, *IEEE Trans. Neural Netw.* **12**, 181–201 (2001).
132. V. N. Vapnik, *The nature of statistical learning theory*, Springer-Verlag New York, Inc., New York, NY, USA, 1995.
133. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, 2007, 1st ed. 2006. corr. 2nd printing edn.
134. R. Huerta, S. Vembu, J. M. Amigó, T. Nowotny, and C. Elkan, *Neural Comput.* **24**, 2473–2507 (2012).
135. Z. Li, and J. Hopfield, *Biol. Cybern.* **61**, 379–392 (1989).
136. R. Gütig, and H. Sompolinsky, *Nature Neurosci.* **9**, 420–428 (2006).
137. M. I. Rabinovich, R. Huerta, and V. Afraimovich, *Phys. Rev. Lett.* **97**, 188103 (2006).
138. M. Rabinovich, A. Volkovskii, P. Lecanda, R. Huerta, H. D. Abarbanel, and G. Laurent, *Phys. Rev. Lett.* **87**, 068102 (2001).
139. P. Arena, L. Patano, and P. S. Termini, *Neural Netw.* **32**, 35 – 45 (2012).
140. T. Cover, *IEEE Trans. Electron Comput.* **14**, 326 (1965).
141. C. Cortes, and V. Vapnik, *Mach. Learn.* **20**, 273–297 (1995).
142. S. Luo, R. Axel, and L. Abbott, *Proc. Natl. Acad. Sci. USA* **107**, 10713–10718 (2010).
143. M. Garcia-Sanchez, and R. Huerta, *J. Comput. Neurosci.* **15**, 5–17 (2003).
144. T. Nowotny, and R. Huerta, *Biol. Cybern.* **89**, 237–241 (2003).
145. T. Nowotny, ""Sloppy engineering" and the olfactory system of insects," in *Biologically Inspired Signal Processing for Chemical Sensing*, edited by S. Marco, and A. Gutierrez, Springer, 2009, vol. 188 of *Studies in Computational Intelligence*.
146. C. Assisi, M. Stopfer, G. Laurent, and M. Bazhenov, *Nature Neurosci.* **10**, 1176–1184 (2007).
147. L. A. Finelli, S. Haney, M. Bazhenov, M. Stopfer, and T. J. Sejnowski, *PLoS Comput. Biol.* **4**, e1000062 (2008).
148. M. García-Sánchez, and R. Huerta, *Neurocomputing* **58**, 337–342 (2004).
149. J. Wessnitzer, and B. Webb, *Bioinspiration & Biomimetics* **1**, 63 (2006).
150. M. Giurfa, *Curr. Opin. Neurobiol.* **13**, 726 – 735 (2003).
151. D. Hebb, *The Organization of Behavior*, Wiley, 1949.
152. C. D. Harvey, and K. Svoboda, *Nature* **450**, 1195–1200 (2007).
153. L. F. Abbott, and W. G. Regehr, *Nature* **431**, 796–803 (2004).
154. H. S. Seung, *Neuron* **40**, 1063–1073 (2003).
155. M. Hammer, *Nature* **366**, 4 (1993).
156. P. Szyszka, C. Demmler, M. Oemisch, L. Sommer, S. Biergans, B. Birnbach, A. F. Silbering, and C. G. Galizia, *J. Neurosci.* **31**, 7229–7239 (2011).
157. S. Dehaene, and J. P. Changeux, *Prog. Brain Res.* **126**, 217–229 (2000).
158. E. C. Snell-Rood, D. R. Papaj, and W. Gronenberg, *Brain Behav. Evol.* **73**, 111–128 (2009).
159. Y. LeCun, and C. Cortes, The mnist database, Available online (1998).

160. C. Pelz, B. Gerber, and R. Menzel, *J. Exp. Biol.* **200**, 837–847 (1997).
161. J. Kivinen, A. Smola, and R. Williamson, *Signal Processing, IEEE Transactions on* **52**, 2165–2176 (2004).
162. T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Proceedings of the twenty-first international conference on Machine learning*, ACM, 2004, p. 116.
163. Y. Singer, and N. Srebro, "Pegasos: Primal estimated sub-gradient solver for SVM," in *In ICML*, 2007, pp. 807–814.
164. R. Menzel, and M. Giurfa, *Trends Cogn. Sci.* **5**, 62–71 (2001).
165. R. Gutierrez-Osuna, *Sensors J., IEEE* **2**, 189–202 (2002).
166. S. Marco, and A. Gutiérrez-Gálvez, *IEEE Sens. J.* **12(11)**, 3189–3213 (2012).
167. J. Fonollosa, A. Gutierrez-Galvez, A. Lansner, D. Martinez, J. Rospars, R. Beccherelli, A. Perera, T. Pearce, P. Vershure, K. Persaud, and S. Marco, *Procedia Computer Science* **7**, 226 – 227 (2011).
168. T. Nowotny, M. K. Muezzinoglu, and R. Huerta, *Int. J. Innov. Comput.* **7**, 3825–3838 (2011).
169. T. Nowotny, *BMC Neurosci.* **12**, P239 (2011).
170. J. Houk, J. Davis, and D. Beiser, *Models of information processing in the basal ganglia*, MIT press, 1994.

Tools, flies and what to do next

A. Gomez-Marin

EMBL-CRG Systems Biology Unit, Center for Genomic Regulation, Barcelona, Spain.

Abstract. In these brief notes addressed to students and researchers, recent advances of modern neurobiology are discussed in the light of some of its challenges. I use fly larval chemotaxis as a platform to debate about how much we are able to do with the available tools as opposed to how little we actually understand what it means to decide.

Keywords: behavioral neuroscience; decision making; noise; drosophila; science and society.
PACS: 87.19.L-, 87.19.lu, 87.19.lt, 87.19.lu, 87.19.Ic, 89.20.-a, 87.85.-d, 01.70.+w, 01.75.+m

Chemotaxis is a paradigm to study decision making. Orienting in a chemical environment requires a repertoire of computational abilities to transform sensory information into motor action. In the fruit fly larva, odor-search behavior represents an active sampling process more elaborated than the biased random walks of bacteria and worms, and analogous to sniffing in rats and humans [1, 2, 3, 4, 5]. The *Drosophila* larva is capable of purposeful rich behaviors under the control of a nervous system whose general anatomical layout is similar to, but yet far simpler than, that of vertebrates. The larval brain consists of two hemispheres, each composed of approximately 1000 neurons, compared to millions in mice and billions in humans. Its olfactory system is composed of two bilaterally symmetrical nostrils, called dorsal organs, each hosting 21 olfactory sensory neurons [6]. Olfaction in the fruit fly larva represents a trade-off between numerical simplicity and behavioral complexity. Bilateral sensory input is not a necessary condition for chemotaxis behavior [7], suggesting that temporal computations in such a miniature brain can direct robust, efficient and adaptive orientation behaviors. Computer-vision tracking systems allow to measure animal posture and movement in an automated fashion and at high resolution [8]. By mapping the stimulus landscape to the position of the olfactory organs, the sensory dynamics while crawling towards an attractive odor source can then be inferred. Such accurate sensory-motor data provides a quantitative basis to examine the algorithms that determine whether a maggot will decide to turn left or right.

The study of the mechanisms underlying active sensing during orientation maneuvers represents an experimentally tractable entry point into the general problem of sensory perception and motor control [9]. Together with chemotaxis, it has been applied in other modalities such as phototaxis and thermotaxis [1, 10, 11]. Making use of the detailed knowledge about larval anatomical and physiological properties, there is great promise that testing extraordinary large collections of transgenic lines in behavioral screens will identify sparse neural substrates or critical neurons mediating the observed behaviors. Furthermore, electrophysiological recordings [12] and functional imaging via genetically encoded calcium indicators [13] represent invaluable measurements of the neural activity responsible for such computations and behavioral decisions. Incidentally, functional imaging in behaving animals is just starting to be realized [14, 15]. Transgenic

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 120-123 (2013); doi: 10.1063/1.4776508

and transparent, the larva is an optimal system in which the recent advent of optogenetic tools can be applied, allowing causal links to be established. By ectopically expressing light-activated ion channels in specific neurons, neural activity can be manipulated at single spike level without use of invasive methods [16]. Maximizing control and maintaining realism, engineers have built virtual reality arenas where real-time tracking and closed-loop stimulation gives us the opportunity to scrutinize, at an unprecedented level, animal behavior and decision making [17]. These exquisite cutting-edge tools have triggered a renaissance in genetic model organisms such as worms, flies, fish and mice.

But, what does it mean, to decide? This simple question seems to be either trivial or extremely difficult to answer. Thoughts, actions, feelings, and nearly everything we believe we are have a neural basis. As the tools to study the nervous system become more and more sophisticated, we in fact approach more elusive, slippery, and controversial topics. Concepts that in the past were circumscribed to pure speculative reasoning and excluded from serious empirical research, now for the first time, may be amenable to incisive scientific inquiry. What is consciousness? What are its neural correlates [18]? If some animals are conscious, why are others not? How could unconscious matter give rise to a conscious mind? Is it a difference of kind or degree? Is freewill an illusion [19]? An oxymoron from the start? Perhaps reformulated as a biological trait, can it be empirically revisited by modern neuroscience [20, 21]? Regardless whether I am free or not, who am I? What are the neurobiological basis of agency, volition, and the self [22, 23]? Can objective experiments be attributed to subjective experiences [24, 25]?

The answers to many of these questions require an evolutionary and developmental perspective [26]. Why did animals evolve brains? The brain is one of the most complex and remarkable information processing systems in nature. Still, it may have evolved, not only to sense or process information, but primarily to generate actions and control movement [27]. What is behavior, then? In the same way that action potentials define the principal language for neural activity [28], is there an analogous universal descriptor for behavior? Is behavior fundamentally continuous or discrete? For instance, what is the degree of arbitrariness when classifying a larval trajectory segment as a turn or a run? Are responses and actions qualitatively different? Brought to the extreme, is creativity in essence a complex reflex? Animals exhibit a set of responses that are definitely hard-wired, still behavior is variable between and within individuals. Same same but different, should we not be looking beyond the mean for discovering the origins of phenotypic variability? How is spontaneity generated [29]? Can stereotypical behavior emerge from the dynamics of behavior itself [30]? Is the input-output view able to explain how animals act, rather than react to the world [31]? Living organisms are obviously capable of generating novel actions, never performed before. Is that choice or noise?

Are we machines or like machines [32]? The difference being subtle, the consequences of confusing an analogy with its literal meaning can be bewildering, if not dangerous. If worms and flies are complex and intricate organic hardware, what about mice and men? Why would a machine then be unable to feel love, pain or guilt [33, 34]? May your next generation iPhone need a lawyer to defend its rights? Indeed, some working hypotheses are nowadays operating as scientific dogmas [35]. Regarding model organisms, to what extent is that which we learn from a fly in the lab is informative about humans? Is animal behavior in laboratories representative of that in the wild? And is it wise to concentrate 99% of the research on 1% of the species? Coming back to bio-

logical complexity, what is life [36]? A necessary process out of pure chance [37]? Is life's most inner secret really a curly thread of DNA [38]? Are the enduring mysteries of the mind hiding within a complex circuit? According to the circuit doctrine, my connectome should be who I am. But, what is a 1 to 1 scale map useful for? The whole may be understandable, not despite the fact that we do not have all the details of every single part, but precisely because of that [39]. In order to make sense of so much data, powerful frameworks such as information theory or the theory of critical phenomena are good candidates to comprise the essential and direct further experiments [40, 41, 42]. That depends on whether we are looking for exceptions or searching for principles [43].

In these times of crisis, we should ask ourselves who decides where funding goes and under what criteria [44]. What is our return to taxpayers and is economic growth the only acceptable metric to evaluate scientific impact in terms of benefits for society [45]? Since lobbying has become a common and desperate practice, we could simultaneously explore other avenues like open science. How do we reconcile competitiveness with cooperativity [46]? In a world where resources are finite, is the emphasis on certain established lines of research leaving small and creative projects under starvation? Budgets reflect choices. Have we been asking certain questions simply because we are technically capable and to appease others? Many tools that until very recently were a luxury, have become a necessity *per se*. Is the *Homo sapiens* involving into *Homo habilis*? Concerning interdisciplinary science, are we literate, or at least interested, in the problems of other research areas? What are the implications of decision making experiments on medicine or law [47]? True multidisciplinary approaches have been reasonably successful. However, integration of philosophical, sociological, ethical and historical considerations is still lacking and seems daunting or of little value to many scientists. Possibilities for progress are far from being exhausted. How many interfaces are we ready to accept [48]? Are we willing to apply the scientific method to science itself?

What to do next? The above questions are not meant to be rhetorical or pop-philosophy entertainments. They try to reflect the imbalance between how successful we are in terms of what we can do versus how little we actually know. In my opinion, many of these questions are not asked because they end up being consequential. The reality that neuroscience paints appears to be in serious conflict with the actual beliefs of the very same neuroscientists who spearhead it. And that incoherency is extensive to other areas of science and society. I believe we should be educated to be critical, rather than trained to be complacent. I also believe we can become creative individuals, beyond mere productive human resources [49]. In the face of the present situation, young researchers need the courage to freely and fearlessly explore the very many exciting paths neuroscience has to offer. As we step into the unknown [50], let us not accept what is unproven, nor deny what is yet to be disproven.

ACKNOWLEDGMENTS

I acknowledge 8 years of funding from the Spanish Ministry of Education and Science, now renamed as Ministry of Economy and Competitiveness, before dilapidating the public research system by deciding that investing in knowledge is not a priority in the 21st century. I thank Bala, Gus, Laura, Troy and Moraea for feedback and discussions.

REFERENCES

1. A. Gomez-Marin, G. J. Stephens, and M. Louis, *Nature Comm.* **2**,441 (2011).
2. H. C. Berg, and D. A. Brown, *Nature* **239**, 500–504 (1972).
3. J. T. Pierce-Shimomura, T. M. Morse, and S. R. Lockery, *J. Neurosci.* **19**, 9557–9569 (1999).
4. A. G. Khan, M. Sarangi, and U. S. Bhalla, *Nature Comm.* **3**,703 (2012)
5. J. Porter, *et al.*, *Nature Neurosci.* **10**, 27–29 (2007).
6. A. Gomez-Marin, B. J. Duistermars, M. A. Frye, and M. Louis, *Front. Cell. Neurosci.* **4**,6 (2010).
7. M. Louis, *et al.*, *Nature Neurosci.* **11**, 187–199 (2008).
8. A. Gomez-Marin, and M. Louis, *PLoS One* **7**(8), e41642 (2012).
9. A. Gomez-Marin, and M. Louis, *Curr. Opin. Neurobiol.* **22**(2), 208–15 (2012).
10. M. Busto, B. Iyengar, and A. R. Campos, *J. Neurosci.* **19**, 3337–3344 (1999).
11. L. Luo, *et al.*, *J. Neurosci.* **30**, 4261–4272 (2010).
12. D. J. Hoare, C. R. McCrohan, and M. Cobb, *J. Neurosci.* **28**, 9710–9722 (2008).
13. L. Tian, *et al.*, *Nature Methods* **6**, 875–881 (2009).
14. J. D. Seelig, *et al.*, *Nature Methods* **7**, 535–540 (2010).
15. M. B. Ahrens, *et al.*, *Nature* **485**, 471–477 (2012).
16. J. Y. Lin, M. Z. Lin, P. Steinbach, and R. Y. Tsien, *Biophys. J.* **96**, 1803–1814 (2009).
17. A. M. Leifer, *et al.*, *Nature Methods* **8**, 147–152 (2011).
18. K. Koch, *The quest for consciousness: a neurobiological approach*, Roberts & Co, Denver, 2004.
19. S. Harris, *Free Will*, Random House, 2012.
20. B. Brembs, *Proc. Royal Soc. B: Biol. Sci.*, **278**(1707), 930–939 (2010).
21. B. Gerber, *EMBO Rep.* **13**, 17–19 (2012).
22. A. Damasio, *Self Comes to Mind*, Ediciones Destino, 2010.
23. R. D. V. Glasgow, *Split Down the Sides: on the subject of laugher*, University Press of America, 1997.
24. M. Heidelberger, *Nature from within*, University of Pittsburgh Press, 2004.
25. A. J. Duggins, *Biosystems* **102**, 124–33 (2010).
26. G. L. G. Miklos, *J. Neurobiol.* **24**(6), 842–890 (1993).
27. D. W. Franklin, and D. M. Wolpert, *Neuron* **72**, 425 (2011)
28. F. Rieke, *et al.*, *Spikes: exploring the neural code*, MIT Press, Cambridge, Mass, 1999.
29. A. Maye, C. Hsieh, G. Sugihara, and B. Brembs, *PLoS One* **5**, e443 (2007).
30. G. J. Stephens, M. B. de Mesquita, W. S. Ryu, and W. Bialek, *PNAS* **108**(18), 7286–7289 (2011).
31. R. M. Costa, *Curr. Opin. Neurobiol.* **21**(4), 579–86 (2011).
32. R. Descartes, *Discours de la méthode*, Tecnos, 2008.
33. V. Braitenberg, *Vehicles: Experiments in Synthetic Psychology*, Cambridge, MA, MIT Press, 1984.
34. R. Penrose, *The Emperor's New Mind*, Oxford University Press, 1999.
35. R. Sheldrake, *The Science Delusion: Freeing the Spirit of Enquiry*, Hodder and Stoughton, 2012.
36. E. Schrodinger, *What is life?*, Tusquets, 2001.
37. J. Monod, *Chance and Necessity*, Tusquets, 2000.
38. J. D. Watson, *DNA: The secret of life*, Alfred A. Knopf, Random House, 2004.
39. Y. Lazebnik, *Biochemistry (Moscow)* **69**(12), 1403–1406 (2004).
40. A. Borst, and F. E. Theunissen, *Nat. Neurosci.* **2**, 947–957 (1999).
41. D. C. Knill, and A. Pouget, *Trends in Neurosci.* **27**, 12 (2004).
42. D. R. Chialvo, *Nature Phys.* **6**, 744–750 (2010).
43. W. Bialek, *Biophysics: Searching for Principles*, Princeton University Press, 2012.
44. P. A. Lawrence, *PLoS Biol.* **7**(9), e1000197 (2009).
45. G. Hardin, *Growth: Real and Spurious. Living Within Limits*, Oxford University Press, 1995.
46. G. Hardin, *Science, New Series* **162**(3859), 1243–1248 (1968).
47. D. M. Eagleman, *Houston Lawyer* **16**(6), 36–40 (2008).
48. E. A. Abbott, *Flatland: A Romance of Many Dimensions*, Dover Publications, 1992.
49. D. H. Pink, *Drive: The Surprising Truth About What Motivates Us*, Canongate Books, 2011.
50. J. Campbell, *The Hero with a Thousand Faces*, New World Library, California, 2008.

Modeling of spontaneous zero-lag synchronization and wave propagation in cat spinal cord

H. Kato[*], C.A. Cuellar[†], R. Delgado-Lezama[†], P. Rudomin[†], I. Jiménez[†], E. Manjarrez[**] and C. R. Mirasso[*]

[*]*Instituto de Física interdisciplinar y sistemas complejos, Universidad de las Islas Baleares, E-07122 Palma de Mallorca, Spain.*
[†]*Departamento de Fisiología, Biofísica y Neurociencias CINVESTAV, México.*
[**]*Instituto de Fisiología, Benemérita Universidad Autónoma de Puebla. 14 Sur 6301, Col. San Manuel, Apartado Postal 406, Puebla, Pue. CP 72570, México.*

Abstract. In this study, we proposed a simple but physiologically plausible network model that can reproduce both the sinusoidal electrical wave propagation and the spontaneous zero-lag synchronization experimentally observed in the cat spinal cord. Our model enhances the hypothesis of the coexistence of two alternative assemblies in the cat spinal cord.

Keywords: spinal cord network; sinusoidal wave propagation; zero-lag synchronization.

INTRODUCTION

A sinusoidal electrical wave propagation called sinusoidal cord dorsum potentials (CDPs), that travels rostrocaudally along the cat spinal cord during fictive scratching, was experimentally observed by Cuellar et al. [1]. Pérez et al. [2] suggested a theoretical model to account for the wave propagation and its possible fairules. These authors showed that the activity of the central pattern generator (CPG) and its followers represent the sinusoidal CDPs.

The spontaneous electrical activity of a neuronal assembly in the dorsal horn was observed in the lumbar spinal cord of cats. Such a characteristic electrical potential was defined with the term of negative CDPs. In a study of Mark and Gasteiger [3], it was suggested that these nCDPs were generated by internal effect of spinal cord mechanisms. Recent studies performed by Manjarrez et al. [4, 5] showed that the large amplitude of nCDPs, that was found in the dorsal grey matter by Gasteiger and Ichikawa [6], is caused by the synchronous activation of an assembly of dorsal horn neurons. Furthermore, the nCDPs lasted for 40–60 ms and had characteristic low frequency components (3–20 Hz), so it was assumed that the highly synchronous activity of the dorsal horn neural assembly induced these nCDPs. Interestingly, Cuellar et al. [1] observed that the nCDPs were superimposed over the sinusoidal electrical wave propagation and not correlated with any phase of the wave propagation. These facts allowed us to hypothesize the coexistence of two alternative assemblies in the cat spinal cord. However, there is no model accounting for this spontaneous activity.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 124-129 (2013); doi: 10.1063/1.4776509

In this study, we modeled a spinal cord network reproducing both the wave propagation and the zero-lag synchronization and unified their theories in the cat spinal cord.

MATERIALS, METHODS AND RESULTS

In order to reproduce two experimentally observed interesting phenomena, that are both the sinusoidal electrical wave propagation and the spontaneous zero-lag synchronization, we constructed a spinal cord network as depicted in Fig. 1*A*. Our network model consisted of twelve CPGs of which the set from third to tenth covered the segments L4–S1 of the spinal cord. These CPGs were connected by excitatory synapses with their neighbors in both upward and downward directions. The connectivity of assemblies in a CPG was the same as in Ref. [2] (See Ref. [2] for details of CPG). In a CPG, both excitatory and inhibitory interneuronal assemblies included twenty neurons. Half of these neurons were bursting and the others were spiking in the excitatory assemblies, while there were only bursting neurons in the inhibitory assemblies. Within an excitatory assembly, neurons were randomly connected with each other with 95% of connection probability but there were no connections within an inhibitory assembly. We allowed the neurons between the connected assemblies in a CPG to have synaptic connections with 90% connection probability. The CPGs arranged along the cat spinal cord interacted via excitatory connections with excitatory assemblies and there were two types of connections: feedforward and recurrent connections. Only bursting neurons were permitted to have feedforward connections and the connections projected to the bursting neurons in the next forward CPG. In contrast to the feedforward connections, recurrent connections were restricted to laying only between spiking neurons. The recurrent connections were able to project in both forward and backward directions until the third nearest neighbors. Introducing these two types of connections makes it possible to embed two alternative assemblies, however, the spiking and the bursting neurons were locally interacted with each other in excitatory assemblies. The connection probability of neurons between the CPGs was set to 85%. Our model had locally dense connectivity but kept globally sparse connectivity.

To simulate the neuronal activity, we used a square-bursting version of the Morris-Lecar neuron model [7] whose activity is given by

$$\begin{aligned} C\frac{dv}{dt} &= I - g_{\mathrm{Ca}} m_\infty(v)(v - V_{\mathrm{Ca}}) - g_{\mathrm{K}} w(v - V_{\mathrm{K}}) \\ &\quad - g_{\mathrm{L}} m_\infty(v)(v - V_{\mathrm{L}}) - g_{\mathrm{KCa}} z(y)(v - V_{\mathrm{K}}) - I^{\mathrm{syn}}, \qquad (1) \\ \frac{dw}{dt} &= \phi\tau(v)(w_\infty(v) - w), \qquad (2) \\ \frac{dy}{dt} &= \varepsilon(-\mu g_{\mathrm{Ca}} m_\infty(v)(v - V_{\mathrm{Ca}}) - y), \qquad (3) \end{aligned}$$

where v, w, and y describe the membrane potential, the slow recovery variable, and the calcium concentration. The variable C represents the membrane capacitance per unit of area. In this conductance-based neuron model, four ionic current is taken into account: the calcium, potassium, leakage and calcium-dependent potassium channels whose

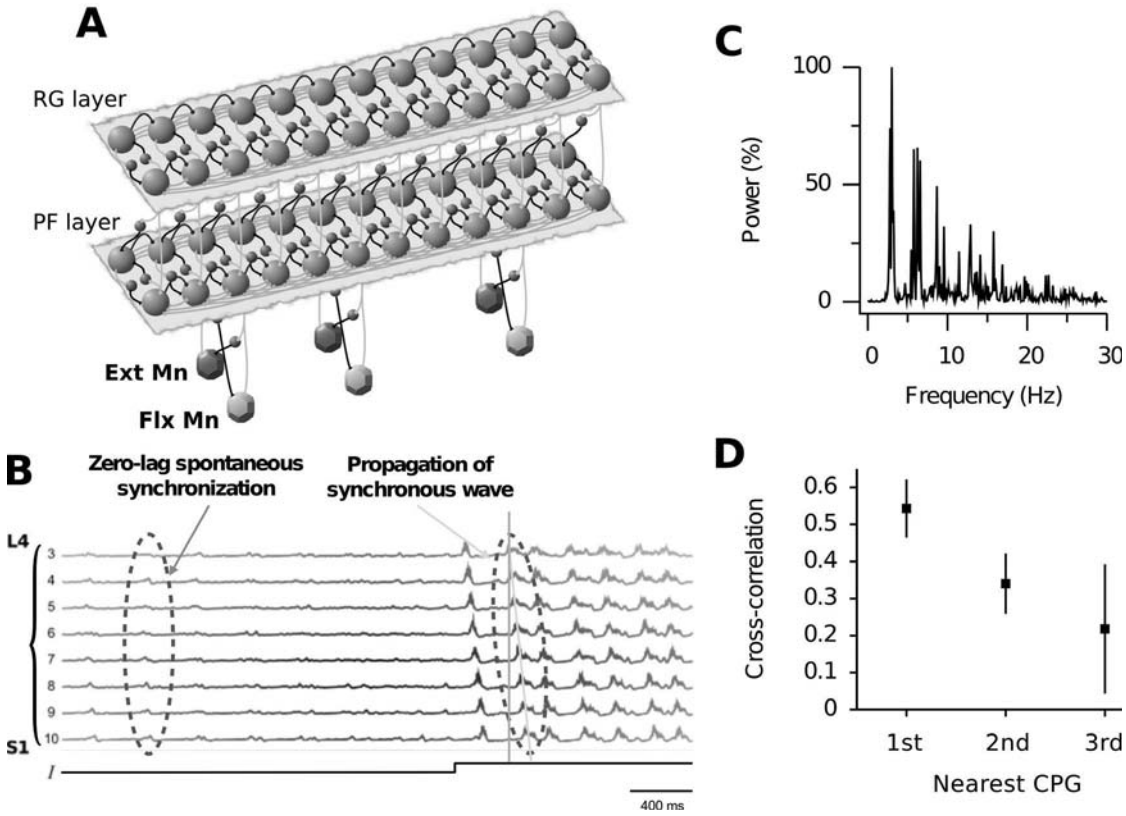


FIGURE 1. Spontaneous zero-lag and propagating synchronization in cat spinal cord. (A) Schematic of our proposal of a spinal cord network. Twelve CPGs were assumed to be arranged along the cat spinal cord. The excitatory assemblies of third, sixth, and tenth CPGs in the PF layer projected to the extensor and the flexor populations. (B) The mean potential of the excitatory populations in the RG layer from the third to tenth CPGs. (C) Power spectra of (B) in the zero-lag synchronization period. (D) The influence of distances on the correlation between intersegments in the zero-lag synchronization term. The mean cross-correlation of the mean potentials between pairs of CPGs are plotted. The mean values were calculated from all possible pairs. The vertical bars stand for the standard deviation.

corresponding conductances are g_{Ca}, g_K, g_L, and g_{KCa}. When introducing calcium-dependent potassium channel makes the model reproduces the bursting behavior. The parameters μ and ε control the dynamics of this channel. The parameter μ represents the ratio between the surface area of the neuron and the calcium volume, and the parameter ε is the product of the calcium degradation rate and the ratio of free to total calcium. The nonlinear functions governing the dynamics of the ionic current in Eqs. (1)–(3) have the

following forms:

$$m_\infty(v) = \frac{1}{2}\left(1+\tanh\left(\frac{v+V_1}{V_2}\right)\right), \tag{4}$$

$$w_\infty(v) = \frac{1}{2}\left(1+\tanh\left(\frac{v-W_1}{W_2}\right)\right), \tag{5}$$

$$\tau(v) = \cosh\left(\frac{v-W_1}{2W_2}\right) \tag{6}$$

$$z(y) = \frac{y}{1+y}. \tag{7}$$

The parameters V_1, V_2, W_1, and W_2 are constant values, and ϕ determines the relative time scale of the slow recovery variable to the membrane potential.

The synaptic current to each neuron is described as

$$I^{\text{syn}} = \sum_i g_i^{\text{syn}} r_i (v - E_{\text{s}}), \tag{8}$$

where g_i^{syn} is the synaptic conductance and E_{s} is the synaptic reversal potential [8]. The variable r_i represents the fraction of bound receptors whose kinetics are described by

$$r_i = \begin{cases} 1 - e^{-\alpha t} & t \le t_{\text{on}}, \\ (1 - e^{-\alpha t_{\text{on}}}) e^{-\beta(t - t_{\text{on}})} & t > t_{\text{on}}, \end{cases} \tag{9}$$

where α and β are the time constant determining the rise and decay of the fraction, and t_{on} represents the time to reach the peak value of the fraction of bound receptors. All parameters in Eqs. (1)–(9) are concluded in Tab. 1.

All the neurons in the network were always stimulated by weak noisy excitatory input following a Poisson distribution with the mean firing rate of 10 Hz, inducing that spiking neurons to fire asynchronously between 1 and 2 Hz in the absence of connections. Afterwards, the bias current of bursting neurons from the forward third CPG was slightly increased by 0.8 μA/cm^2 to evoke the incoming input during the fictive scratching as show in Fig. 1*B*.

Before the increase of the bias current, peaks of the mean potentials simultaneously appear, indicating that the activity of neuronal assemblies were highly synchronous between intersegments. Similar to the experimental data [4, 9, 5], however the mean potentials fluctuate and the zero-lag synchronization is irregular, therefore, the lack of the synchronization of some CPGs can be observed as well. After the bias increase, the proposed network began to exhibit propagation of electrical waves along the cat spinal cord. Our model qualitatively reproduced two characteristic phenomena experimentally observed in the cat spinal cord.

Manjarrez et al. [4] showed that the electrical potential during the zero-lag synchronization had the characteristic frequency components. In their power spectrum analysis, the largest peak was located below 10 Hz and there were some peaks until 20 Hz. Our spectrum analysis of the mean potential during the zero-lag synchronization captures these properties in [4] as shown in Fig. 1*C*.

TABLE 1. Parameters of neuron (left) and synapse (right) models.

Parameters	Bursting neurons	Spiking neurons
C	5 μF/cm^2	20 μF/cm^2
g_{Ca}	4 μS/cm^2	4 μS/cm^2
g_K	8 μS/cm^2	8 μS/cm^2
g_L	2 μS/cm^2	2 μS/cm^2
g_{KCa}	0.25 μS/cm^2	0 μS/cm^2
V_{Ca}	120 mV	120 mV
V_K	−84 mV	−84 mV
V_L	-60 ± 0.01 mV	-60 ± 0.01 mV
V_1	1.2 mV	1.2 mV
V_2	18 mV	18 mV
W_1	12 mV	12 mV
W_2	17.4 mV	17.4 mV
ϕ	0.92 s^{-1}	1/15 s^{-1}
ε	0.0175 s^{-1}	0 s^{-1}
μ	0.0149 – 0.015 s^{-1}	0 s^{-1}
I	43 μA/cm^2	39.7 – 39.9 μA/cm^2

Parameters	
α	0.33 ms^{-1}
β	0.2 ms^{-1}
t_{on}	1 ms
$g_{bur\to ex}$	0.1 μS/cm^2
$g_{spk\to ex}$	0.01 μS/cm^2
$g_{bur\to inh}$	0.12 μS/cm^2
$g_{spk\to inh}$	0.01 μS/cm^2
$g_{inh\to ex}$	0.025 – 0.05 μS/cm^2
g_{rec}	0.11 μS/cm^2
g_{ff}	0.1 μS/cm^2
g_{noise}	0.1 μS/cm^2
E_s	0, −80 mV

A sequential analysis of Manjarrez et al. [5] found that nearby neuron pairs had higher correlated activities and the correlation decreased as distance increased. Our correlation analysis in Fig. 1*D* shows the consistent results with the experimental data in Ref. [5].

CONCLUSION

In this study, we modeled not only the propagation of traveling electrical waves but also the spontaneous zero-lag synchronization experimentally observed in the cat spinal cord. From the model we observed both phenomena, and our model reproduced many properties of experimental data during the zero-lag synchronization. Our results enhanced the hypothesis of the coexistence of two alternative neuronal assemblies in the cat spinal cord, that were suggested in previous studies. Our two phenomena reproducible model still has room for improvement of its performance by introducing the bilateral monosynaptic reflex reported by Manjarrez et al. [10], which is our future work.

ACKNOWLEDGMENTS

We acknowledge the financial supports of Project No. FIS2007-60327 from MICINN (Spain) and FEDER (EU) and Comunitat Autónoma de les Illes Balears, Conacyt 153583, and Cátedra Marcos Moshinsky.

REFERENCES

1. C. A. Cuellar, J. A. Tapia, V. Juárez, J. Quevedo, P. Linares, L. Martínez, and E. Manjarrez, *J. Neurosci.* **29**, 798–810 (2009).
2. T. Pérez, J. A. Tapia, C. R. Mirasso, J. García-Ojalvo, J. Quevedo, C. A. Cuellar, and E. Manjarrez, *J. Neurosci.* **29**, 10254–10263 (2009).
3. V. H. Mark, and E. L. Gasteiger, *Electroenceph. & Clin. Neurophysiol.* **5**, 251–258 (1953).
4. E. Manjarrez, J. G. Rojas-Piloni, I. Jiménez, and P. Rudomin, *J. Physiol.* **520**, 445–460 (2000).
5. E. Manjarrez, I. Jiménez, and P. Rudomin, *Exp. Brain Res.* **148**, 401–413 (2003).
6. E. L. Gasteiger, and S. Ichikawa, *Boletín del Instituto de Estudios Médicos y Biológicos de México* **21**, 223–234 (1963).
7. J. Rinzel, and G. B. Ermentrout, *Analysis of Neural Excitability and Oscillations*, The MIT Press, Cambridge, Mass, 1989.
8. A. Destexhe, Z. F. Mainen, and T. J. Sejnowski, *Neural Comput.* **6**, 14–18 (1994).
9. E. Manjarrez, J. G. Rojas-Piloni, Méndez, L. Martinez, D. Vélez, D. Vázquez, and A. Flores, *Neurosci. Lett.* **326**, 93–96 (2002).
10. E. Manjarrez, Z. Hernández-Paxtián, and A. F. Kohn, *J. Neurophysiol.* **94**, 3199–3210 (2005).

Modelling the anesthetized brain with ensembles of neuronal and astrocytic oscillators

T. Hansard, A. C. Hale and A. Stefanovska

Department of Physics, Lancaster University, Lancaster, LA1 4YB.

Abstract. We propose a minimalistic model of the anesthetized brain in order to study the generation of rhythms observed in electroencephalograms (EEGs) recorded from anesthetized humans. We propose that non-neuronal brain cells—astrocytes—play an important role in brain dynamics and that oscillation-based models may provide a simple way to study such dynamics. The model is capable of replicating the main features (i.e. slow and alpha oscillations) observed in EEGs. In addition, this model suggests that astrocytes are integral to the generation of slow EEG ($\sim$0.7 Hz) rhythms. By including astrocytes in the model we take a first step towards investigating the interaction of the brain and cardiovasular system which are primarily connected via astrocytes. The model also illustrates that rich nonlinear dynamics can arise from basic oscillatory "building blocks" and therefore complex systems may be modelled in an uncomplicated way.

Keywords: computational neuroscience; Kuramoto; synchronization; phase dynamics; EEG; anesthesia.
PACS: 87.18.Nq, 87.19.L-, 87.19.le, 87.19.lj, 87.19.lk, 87.19.lm, 87.19.ln, 05.45.Xt

KEEPING IT SIMPLE

Scientific study commonly begins with the simple aspects of a system and only when these are well understood does one consider the necessity of additional details. As such we propose that oscillatory membrane potential (MP) dynamics underlie neuronal firing and investigate how such oscillations may produce the main features observed in the electroencephalogram (EEG) recorded from anesthetized humans. We assume that the MP is oscillating and that cellular interaction modifies these rhythms. Hence we investigate how the dynamics of different neuronal and astrocytic groups interact and how they are altered by environmental changes, as in anesthesia. We consider that action potentials (APs) are just a response to above-threshold MPs, but oscillations underlie activity at the scale of cells, cell clusters, brain regions and EEG recordings [1, 2]. Thus the MP dynamics of each cell is represented in the model by a self-sustained phase oscillator.

In addition we consider the role of astrocytes in brain dynamics. For every neuron in the human brain there may be as many as 10 astrocytes [3] however astrocytes do not extend distal processes or generate APs [3], therefore their role in generating the electrical activity of the brain is often overlooked. Recent research is elevating this position by showing that astrocytes possess many of the same neuroreceptors and neurotransmitters as neurons; extend processes which envelop synapses by which neuronal communication is modulated; alter blood flow via processes which encircle blood vessels; and control the delivery of energy substrates (primarily lactate) to neurons [3]. In addition evidence suggests that astrocytes exhibit spontaneous infra-slow oscillations at around

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 130-133 (2013); doi: 10.1063/1.4776510

0.01 Hz [4, 5] and through the numerous neuron-astrocyte interaction routes available it is likely that these oscillations impact on neuronal dynamics. Indeed, *in vitro* experiments have shown that these spontaneous astrocytic rhythms evoke inward currents in neurons which have the same time-scale as slow oscillations [4].

THE EFFECTS OF ANESTHESIA ON BRAIN RHYTHMS

The induction of unconsciousness by general anesthetics is associated with decreased neuronal activity and hence a slowing of EEG dynamics [6, 7, 8, 9]. The awake state is associated with predominant cortical oscillations at high-frequencies (i.e. $>$20 Hz), while states of unconsciousness are associated with prominent slow rhythms (i.e. $<$20 Hz) for which power increases with depth of unconsciousness [6, 9]. Furthermore, narrow-band alpha ($\sim$10 Hz) activity [10, 8] and slow oscillations ($\sim$0.5 Hz) [2] have been reported during anesthesia.

AN OSCILLATION-BASED MODEL OF THE BRAIN

The model is *explicitly* based on oscillatory dynamics by utilizing the Kuramoto model [11]. The phase velocity $\dot{\phi}_k$ is described as a function of natural frequency ω_k and the phase difference $\phi_j - \phi_k$ between the kth oscillator and all other N oscillators belonging to the same ensemble:

$$\dot{\phi}_k(t) = \omega_k + \frac{A}{N}\sum_{j=1}^{N}\sin(\phi_j(t) - \phi_k(t)) \qquad \text{for k} \in 1..\text{N and } \phi_{j,k} \in -\pi \ldots \pi, \tag{1}$$

where the strength of coupling between oscillators is given by A and the natural frequency ω_k is drawn from a normal distribution centred on a mean ω_0. The phase velocity $\dot{\phi}_k$, which has units of radians/second, may be compared to frequency characteristics in data by a conversion of units: $\dot{\phi}_k/2\pi = f_k$, where f_k is measured in Hz.

Sheeba *et al.* have previously used the Kuramoto model (1) to model the rat brain during anesthesia [12]. By including astrocytes we extend this work to model the anesthetized human brain. Different neuronal groups have been identified as exhibiting different intrinsic frequency dynamics [1, 9] and thus a given neuronal group may be represented by an ensemble of oscillators with a specified natural frequency distribution. As shown in Fig. 1, the model consists of eight ensembles representing four astrocytic groups (A1–A4), two cortical groups (C1 and C2), the thalamus (T) and the thalamic reticular nucleus (R). Coupling is global within each ensemble and between connected ensembles. The inhibitory action of thalamic reticular neurons on thalamic neurons is approximated by an addition to the equation giving frequency dynamics of T:

$$\chi = C\frac{K^R}{N^R}\sum_{j=1}^{N^R}\dot{\phi}_j^R, \tag{2}$$

where C is a *negative* constant and the subscripts *R* indicate variables of ensemble R. Thus the mean frequency of R is deducted from the frequency of each oscillator in T and hence by varying *C* the degree of thalamic slowing may be altered.

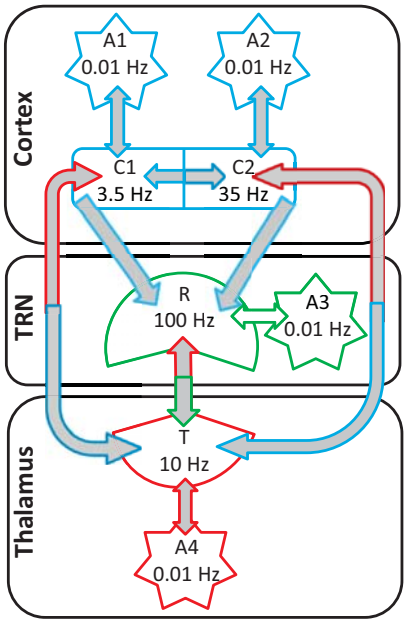


FIGURE 1. A schematic of the eight-ensemble brain model, comprising four astrocytic ensembles (A1–A4) and four neuronal ensembles (C1, C2, T and R). Frequencies indicate mean values of the natural frequency distribution for each ensemble.

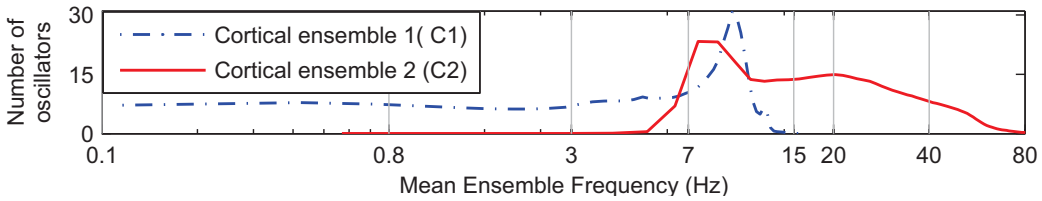


FIGURE 2. The frequency distribution averaged over time for one simulation showing that the model can exhibit dynamics in the slow (0.1-1 Hz), theta (3-7 Hz) alpha (7-15 Hz) and beta (15-20 Hz) bands.

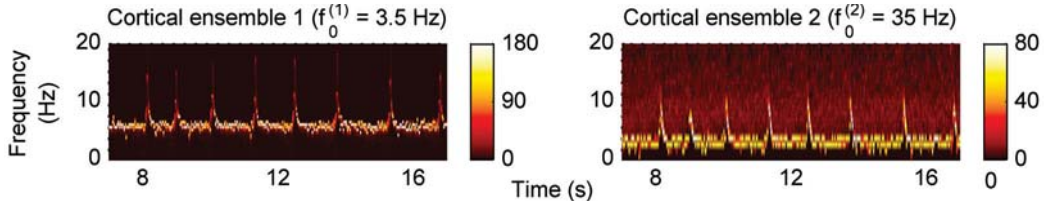


FIGURE 3. The time-frequency dynamics of the two cortical ensembles displaying distinct UP/DOWN dynamics similar to the slow oscillation. Gradient bars show the number of oscillators (out of 500 per ensemble) which express a given frequency at a given time.

This minimalistic approach produces a model by which one may focus on investigating how the different oscillatory dynamics of each ensemble interact to produce synchrony at different frequencies and thus generate the EEG. Simulations were performed for a range of coupling strengths. Parameter values that gave dynamics corresponding with EEG data and literature also corresponded with expectations (e.g. weak inter-thalamic connections but strong corticothalamic and thalamocortical connections). As shown in Figs. 2 and 3, the model exhibits the predominant characteristics of the human EEG during anesthesia: slow oscillations (0.1-0.8 Hz) and pronounced narrowband alpha oscillations (7-15 Hz). Furthermore, the presence of slow oscillations in the simulations (shown in Fig. 3) were a product of the interaction between astrocytic and neuronal ensembles. Increasingly slow oscillations resulted from of simulating an increased anesthetic depth, corresponding with experimental evidence [13].

SUMMARY

We present a model which illustrates that oscillations may be used to investigate the generation of EEGs from microscopic dynamics. The model replicates dynamics seen in EEGs recorded from anesthetized humans and simulation of increased anesthetic depth by alteration of the coupling between astrocytic and neuronal oscillators enhances the slow oscillation, thus implicating the astrocyte-neuron interaction in the generation of this rhythm. The model facilitates future investigations of the way in which the brain and cardiovascular system interact via astrocytes.

REFERENCES

1. G. Buzsáki, *Rhythms of the Brain*, Oxford University Press, New York, USA, 2006.
2. F. Amzica, and D. Neckelmann, *J. Neurophysiol.* **82**, 2731–2746 (1999).
3. P. Magistretti, and B. Ransom, "Astrocytes," in *Neuropsychopharmacology: The fifth generation of progress*, American college of neuropsychopharmacology, 2002.
4. H. Parri, and V. Crunelli, *J. Neurosci.* **120**, 979–992 (2003).
5. I. Winship, N. Plaa, and T. Murphy, *J. Neurosci.* **27**, 6268–6272 (2007).
6. E. John, and L. Prichep, *Anesthesiol.* **102**, 447–471 (2005).
7. M. Alkire, A. Hudetz, and G. Tononi, *Science* **322**, 876–880 (2008).
8. A. Cimenser, P. Purdon, E. Pierce, J. Walsh, A. Salazar-Gomez, P. Harrell, C. Tavares-Stoeckel, K. Habeeb, and E. Brown, *PNAS* **108**, 8832–8837 (2011).
9. T. Hansard, *Understanding the awake and anæsthetized brain through oscillations*, Ph.D. thesis, Lancaster University (2012).
10. V. Feshchenko, R. Veselis, and R. Reinsel, *Neuropsychobio.* **50**, 257–266 (2004).
11. Y. Kuramoto, *Chemical oscillations, waves, and turbulence*, Springer, Berlin, 1984.
12. J. Sheeba, A. Stefanovska, and P. McClintock, *Biophys. J.* **95**, 2722–2727 (2008).
13. A. Doi, M. Mizuno, T. Katafuchi, H. Furge, K. Koga, and M. Yoshimura, *Eur. J. Neurosci.* **26**, 2565–2575 (2007).

Modeling state transition of hippocampal local field potential between theta rhythm and large irregular amplitude activity by bifurcation between a limit cycle and chaotic dynamics

K. Tokuda*, Y. Katori†,** and K. Aihara**

*Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.
†FIRST, Aihara Innovative Mathematical Modelling Project, JST.
**Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.

Abstract. Here we propose a possible mathematical structure of the state transition of the hippocampal local field potential (LFP) between theta rhythm and large irregular amplitude activity (LIA) in terms of nonlinear dynamics. The basic idea is that the alternation of the state between theta rhythm and LIA can be interpreted as a bifurcation of the attractor between a limit cycle and chaotic dynamics. Tsuda et al. reported that a network composed of simple class 1 model neurons connected with gap junctions shows both synchronous periodic behavior and asynchronous chaotic behavior [1]. Here we model the network of hippocampal interneurons extending their model. The network is composed of electrically coupled simple 2-dimensional neurons with natural resonant frequency in the theta frequency. We incorporate a periodic external force representing the medial septal afferent. The system converges on a limit cycle under this external force, but shows chaotic dynamics without this external force. Furthermore, the external noise realized rapid alteration of the state obeying the change of the amplitude of the septal input.

Keywords: chaos; gap junction; diffusion; hippocampus; local field potential.
PACS: 87.19.lj

The states of the hippocampal LFP are divided into mutually exclusive states. One of them is the theta rhythm, which is a highly periodic pattern with frequency range around 4-12 Hz. The theta rhythm appears in attentive states described by the subjective terms such as "voluntary," "preparatory," "exploring," or "orienting" [2]. Yet another state of the hippocampal LFP is the large irregular amplitude activity (LIA), which occurs when the animal's behavior is characterized as immobility, sleeping, or grooming.

The medial septum-diagonal band of Broca (MS-DBB) is assumed to be the crucial structure for the generation of the hippocampal theta rhythm. Lesion or inactivation of MS-DBB abolishes the theta rhythm in the hippocampus. The periodic activity in the theta range frequency in the MS-DBB precedes the theta rhythm in the hippocampus. According to those observations, the MS-DBB is assumed to be playing the role of pacemaker of the theta rhythm. On the other hand, considerable amount of studies show that there are local rhythm generators in the hippocampus itself [3].

Recent studies show the existence and functional significance of gap junctions in the hippocampus [6, 7, 8]. Theoretical studies also suggest that electrical coupling have profound effects on the dynamics of the neural network. Tsuda et al showed that diffusive

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 134-137 (2013); doi: 10.1063/1.4776514

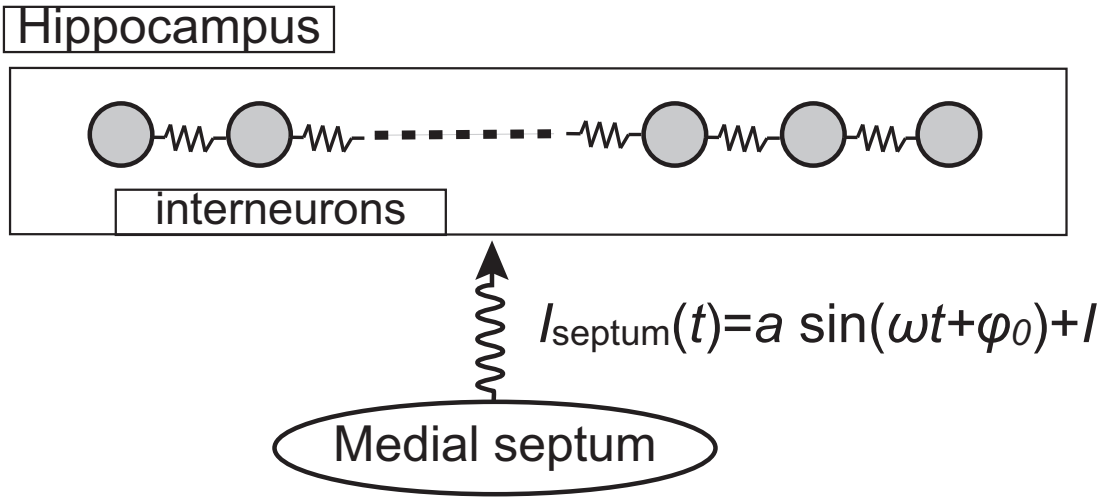


FIGURE 1. The schematic diagram of the network architecture. The main part of the model is the network of interneurons coupled by gap junctions with each neighbors.

coupling realized by gap junctions between class 1 neurons induces chaotic dynamics [1]. In this model, both periodic dynamics and chaotic dynamics are realized with a set of fixed parameters, and these two states appear alternately. Katori et al showed that alternations between synchronous and asynchronous oscillatory state can be realized with gap junction-coupled simple conductance-based model neurons [4]. We proposed [5] a network model of interneurons of the hippocampus using class 1 neurons mutually coupled with electrical synapses, extending the model proposed by Tsuda et al.. We incorporated the input from the MS-DBB as a periodic external current and showed that the dynamics of the hippocampal inhibitory network is controlled by entrainment, which is in accordance with experimental observations.

In the present study, we show that the incorporation of additional external noise realizes rapid alteration obeying the change of external septal input. Figure 1 illustrates the schematic diagram of the network architecture. The model of the network is given by the following equations:

$$\frac{dx_i}{dt} = -y - \mu x_i^2 (x_i - \frac{3}{2}) + J_i + I_{\text{septum}}(t) + D\xi(t), \tag{1}$$

$$\frac{dy_i}{dt} = -y + \mu x_i^2, \tag{2}$$

$$I_{\text{septum}}(t) = a \cdot \sin(\omega t + \phi_0) + I, \tag{3}$$

$$J_i = \begin{cases} g_{\text{GJ}}(x_2 - x_1) & (\text{for} \quad i = 1), \\ g_{\text{GJ}}(x_{i+1} + x_{i-1} - 2x_i) & (\text{for} \quad i = 2, ..., N-1), \\ g_{\text{GJ}}(x_{N-1} - x_N) & (\text{for} \quad i = N), \end{cases} \tag{4}$$

where x_i is the membrane potential, and y_i is the inactivating variable of the ith neuron, J_i is the sum of the currents from neighboring neurons through the gap junctions, μ is a parameter of the model, and the $I_{septum}(t)$ is the external input representing the septal afferent, $D\xi(t)$ is Gaussian white noise with zero mean and standard deviation of D.

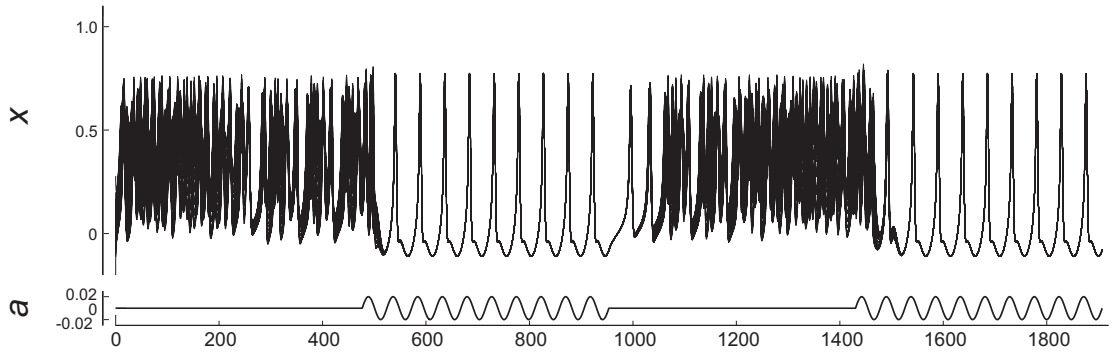


FIGURE 2. Bifurcation induced by changing the amplitude of periodic external force. The upper traces show the time evolution of variable x of all the neurons in the model. The lower trace shows the external force. The model behavior changes into synchronous state following the change of external force, as is observed in the physiological brain. $\mu = 1.65, N = 30, g_{\mathrm{GJ}} = 0.8, D = 0.005, a = 0.02, I = 0.00385, \omega = 0.1318$.

Figure 2 shows the simulation result of the model, when the amplitude of the septal afferent is changed as in the real brain. The system shows asynchronous chaotic dynamics without the septal input. By increasing the amplitude of the septal afferent, the system undergoes a bifurcation and converges on a limit cycle. When the amplitude of the septal input decreases, the orbit escapes from the limit cycle rapidly because the small noise perturbs the orbit slightly from the limit cycle.

We incorporated external noise to the model network of the hippocampal interneurons with class 1 neuron models connected with the diffusive couplings and reproduced the transition between synchronous state and asynchronous state. The weak diffusive couplings between the oscillators induce instability on the all-synchronized periodic orbit, but the periodic external force entrains the system and realizes a stable limit cycle. The external noise realized rapid alteration of the state obeying the amplitude of the external input. We suggest that this corresponds to the experimental observation that the septal periodic afferent entrains hippocampus to the theta rhythm. With our model, the transition of the state of hippocampal LFP can be naturally interpreted as a bifurcation between a limit cycle and chaotic dynamics in terms of dynamical systems theory.

ACKNOWLEDGMENTS

This research is partially supported by the Aihara Innovative mathematical Modelling Project the Japan Society for the Promotion of Science (JSPS) through the "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)," initiated by the Council for Science and Technology Policy (CSTP) and Grant-in-Aid for Scientific Research (A) (20246026) from JSPS.

REFERENCES

1. I. Tsuda, H. Fujii, S. Tadokoro, T. Yasuoka, and Y. Yamaguti, *J. Integr. Neurosci.* **3**, 159–182 (2004).
2. G. Buzsáki, *Neuron* **33**, 325–340 (2002).
3. R. Goutagny, J. Jackson, and S. Williams, *Nature Neurosci.* **12**, 1491–1493 (2009).
4. Y. Katori, E. J. Lang, M. Onizuka, M. Kawato, and K. Aihara, *Int. J. Bifur. Chaos* **20**, 583–603 (2010).
5. K. Tokuda, Y. Katori, and K. Aihara, *BMC Neurosci* **13** (Suppl 1), 189 (2012).
6. T. Fukuda, and T. Kosaka, *J. Neurosci.* **20**, 1519–1528 (2000).
7. J. Konopacki, T. Kowalczyk, and H. Golebiewski, *Brain. Res.* **1019**, 270–274 (2004).
8. K. Allen, E. C. Fuchs, H. Jaschonek, D. M. Bannerman, and H. Monyer, *J. Neurosci.* **31**, 6542–6552 (2011).

The resemblance of an autocorrelation function to a power spectrum density for a spike train of an auditory model

Y. V. Ushakov[1,*], A. A. Dubkov[2,*] and B. Spagnolo[†]

[*]*Lobachevsky State University, Radiophysics department, 23, Gagarin ave., Nizhniy Novgorod, 603950, Russia.*
[†]*Dipartimento di Fisica, Group of Interdisciplinary Physics, Viale delle Scienze, Ed. 18, Universitá di Palermo, I-90128, Italy.*

Abstract. In this work we develop an analytical approach for calculation of the all-order interspike interval density (AOISID), show its connection with the autocorrelation function, and try to explain the discovered resemblance of AOISID to the power spectrum of the same spike train.

Keywords: power spectrum; autocorrelation; inter-spike interval density.
PACS: 87.19.lc, 87.19.ll, 43.80.+p

INTRODUCTION

As it is well-known, the Fourier transformation allows imaging a signal as a sum of sinusoidal components. In case of a spike train, it seems to be not consequent to consider a sequence of sharp pulses as a sum of smooth sinusoids. Apparently, this is one of the reasons why neurophysiologists use the histogram of interspike intervals, in particular, all-order interspike interval density (AOISID), more often, then the power spectrum density (PSD). The power spectrum of a short in time pulse inevitably contains high frequency components, which do not have anything to do with interspike intervals. So, the spectrum provides just the redundant information about a pulse shape, which does not play a role in an inter-neuron communication. Actually, in the past, before the fast Fourier transformation algorithm invention and its wide applications, the usual tool for the signal analysis was a correlometer providing the autocorrelation function (ACF), which is stated [1] to be directly connected with AOISID.

In the presented work we describe a quite unexpected connection between ACF and PSD observed in the auditory system models. The connection between ACF and AOISID is also rigorously derived here.

The presented study has been motivated by the discovery of the resemblance of the PSD at the output of a simple neural model [2] to the AOISID at the output of much more complex model [1] of the same auditory system of mammals with similar parameters of input signals (Fig. 1A). The problem of the analysis of this resemblance was that we

[1] Contact e-mail: ushakov@rf.unn.ru
[2] Contact e-mail: dubkov@rf.unn.ru

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 138-141 (2013); doi: 10.1063/1.4776515

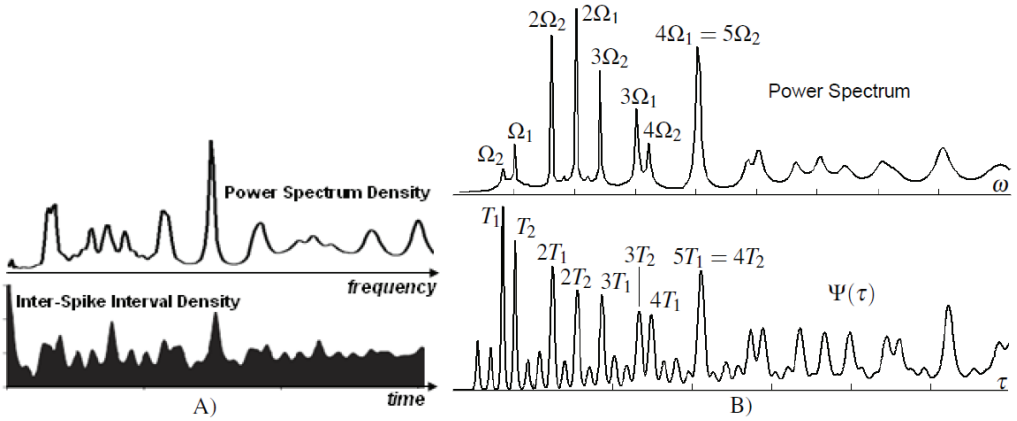


FIGURE 1. Resemblance of PSD to AOISID: A) PSD from Ref. [3] *vs.* AOISID from Ref. [1]; B) PSD and analytical AOISID from Eq. (2) for the same spike train.

had analytical expressions for the PSD of the simple model, but did not have ones for the AOISID. As per the complex model, here we had only AOISID plots and a not clear enough statement about the direct correspondence between AOISID and ACF.

MODEL DESCRIPTION

The mentioned PSD has been calculated for the model, which is described in details in the paper [2]. It consists of three Leaky Integrate-and-Fire neurons, two of which are the input (sensory) elements and are driven by sinusoidal signals, and the third one is the output element receiving spikes from the sensory ones. Each neuron is also influenced by white Gaussian noise. In the previous studies, the characteristics of an output spike train of the model were analyzed at various combinations of input sinusoids' frequencies Ω_1, Ω_2, namely, the combinations, which are typical for harmonious and dissonant musical chords. It was shown [2, 3] that in a case of commensurable input frequencies, the system may be described by the hidden Markov chain with the finite number of states and the transition matrix $\{\pi_{ij}\}$. At the moment of the output spike generation, the system switches between the states. If the target i^{th} state is known, then the distribution $\rho^{(i)}(t)$ of a time interval until the next output spike is also known. In the paper [3] the deriving procedure for the PSD formula is provided, given the matrix $\{\pi_{ij}\}$ and the distributions $\rho^{(i)}(t)$.

AOISID AND ACF

In the paper [1], the following AOISID calculation procedure is proposed. First, all the first-order interspike intervals (ISIs) are extracted from the set of a number of parallel output spike trains. Then, the second-order ISIs are extracted, i.e., the sums of all pairs of consequent intervals. These second-order ISIs are added to the same array as the first-order ones. In the same way, the third-, fourth-, etc. order ISIs are collected in one place

and then distributed into histogram bins. Despite the name of the characteristics, surely, this is not the "All-order" ISI distribution, because the maximal order of a considered ISI is always limited.

Let us consider some sequence of random intervals. The rigorous mathematical approach to the All-Order ISI Distribution requires understanding of the nature of the random quantity, which is being distributed. It is not hard to make sure that in our case the random quantity is the sum of random number N of random intervals $t_1, t_2, t_3, \ldots$:

$$\tau = \sum_{n=1}^{N} t_n, \tag{1}$$

where $N \in \{1, 2, 3, \ldots, N_{max}\}$ and $t_n \in (0, +\infty)$. For example, defining $N_{max} = 2$ and using the formula of the total probability, one can obtain the probability density of τ as $\Psi_2(t) = P(H_1)\rho_1(t) + P(H_2)\rho_2(t)$, where H_1, H_2 are the mutually exclusive events of having one or two addends in the sum Eq. (1), respectively; $P(H_1), P(H_2)$ are their probabilities; $\rho_1(t)$ is the probability density of one interspike interval t_1 to be in the delta-neighbourhood of t; and $\rho_2(t)$ is the same probability density for the sum of two consequent intervals $t_1 + t_2$. Assuming $P(H_1) = P(H_2)$, what, at least, is not in contradiction with the procedure of Ref. [1], we obtain $\Psi_2(t) = [\rho_1(t) + \rho_2(t)]/2$. In the general case, this formula allows inducing the following one:

$$\Psi_{N_{max}}(t) = \frac{1}{N_{max}} \sum_{n=1}^{N_{max}} \rho_n(t). \tag{2}$$

Here, the probability density $\rho_2(t)$ is calculated as the integral: $\int_0^t \rho(t_1, t - t_1) dt_1$, where $\rho(t_1, t_2)$ is the joint probability density for the consequent intervals t_1 and t_2. Analogously, $\rho_3(t) = \int_0^t dt_1 \int_0^{t-t_1} \rho(t_1, t_2, t - t_1 - t_2) dt_2$, etc.

In order to find a connection with the autocorrelation function, one should refer to the paper [4], where the following expression is proposed for ACF:

$$K(\tau) = f^2 \left[\delta(\tau) + \sum_{n=1}^{\infty} W_{t_n}(|\tau|) - 1/T\right]/T. \tag{3}$$

Here, f and T are some constants; $\delta(t)$ is the Dirac delta-function; and the quantity $W_{t_n}(|\tau|)$ is just the same as $\rho_n(\tau)$ in the Eq. (2). Hence, one may assert the proportionality between ACF and AOISID, but not the identity.

AOISID AND PSD

The problem of $\rho_n(t)$ calculation in the Eq. (2) has been solved in the aforementioned case (see "Model description") with the same approach as in Refs. [3, 5], resulting in the AOISID depicted in Fig. 1B. Obviously, there is a resemblance between the top panel and bottom panel plots. Actually, generation of input frequencies' multiples (here $\Omega_1/\Omega_2 = 5/4$) in the power spectrum of the output signal of a nonlinear system is a well-known phenomenon in physics. So, it is interesting to understand, why the temporal characteristics $\Psi(\tau)$ behaves in the same manner with respect to the periods $T_{1,2} = 2\pi/\Omega_{1,2}$? The other interesting question arises, if one remembers about the

direct connection between ACF and PSD through the Fourier transformation. Indeed, is there any advantage in usage of a system with a signal having similar PSD and ACF? However, first of all: is this similarity typical for the brain subsystems, or this is just the negligible particular case? The theoretical research is in progress, and the experimental contribution is welcome and will be much appreciated.

ACKNOWLEDGMENTS

The presented study is supported by the Russian Foundation for Basic Research (grant number 11-02-01418-a).

REFERENCES

1. P. Cariani, *J. New Music Res.* **30**, 107–135 (2001).
2. Y. V. Ushakov, A. A. Dubkov, and B. Spagnolo, *Phys. Rev. E* **81**, 041911–1–13 (2010).
3. Y. V. Ushakov, and A. A. Dubkov, *Moscow University Physics Bulletin* **65**, 372–377 (2010).
4. O. A. Chichigina, A. A. Dubkov, D. Valenti, and B. Spagnolo, *Phys. Rev. E* **84**, 021134 (2011).
5. Y. V. Ushakov, A. A. Dubkov, and B. Spagnolo, *Phys. Rev. Lett.* **107**, 108103 (2011).

Towards holographic "brain" memory based on randomization and Walsh-Hadamard transformation

S. Dolev[1,*], S. Frenkel[2,†] and A. Hanemann[*]

[*]*Dept. of Computer Science, Ben-Gurion University of the Negev, Israel.*
[†]*Institute of Informatics Problems, Russia.*

Abstract. The holographic conceptual approach for cognitive processes in human brain is investigated by neuroscientists due to the ability of holography to describe sophisticated phenomena of human perception and cognition. In this work we suggest a new mathematical description for Pribram's holographic or "holonomic" representation approach for the mind. Namely, we consider: (i) randomization of information, and (ii) Walsh-Hadamard spectral representation of holograms, rather than the well-known Fourier transform representation. The randomization reflects the belief that perceptual processes are not direct, but depend on the perceiver's expectations and previous knowledge as well as the information available in the stimulus itself. The use of Fourier transform and in our case Walsh-Hadamard transform reflects the possibility that each neuron or group of neurons encode some information about the entire image rather than the whole information about a part of the image. We demonstrate that the Walsh-Hadamard transform has benefits over the general Fourier transform. The encoding is performed on randomized information that is then represented by a set of spectral Wash-Hadamard coefficients that have holographic properties. Namely, any portion of the set of coefficients defines a "blurry image" of the original data. The values of the coefficients of the Walsh-Hadamard transformation are distributed approximately normally when the information is randomized, ensuring, with high probability, that growing sets of coefficients implies a monotonic gain of information. Moreover the randomization of the original information yields robust code that is able to cope with missing coefficients. The bridge between the randomization and holographic encoding with the well-known holographic human brain assumption may bring an interesting interpretation of the perception phenomena. In particular, holographic encoding fits the mystery of the human memory encoding, where damage of portions leaves a blurred image and memories. Finally, we give an example of a simple implementation of our approach using neural networks.

Keywords: holographic memory; Walsh-Hadamard; brain; neural network.
PACS: 87.19.lv

HOLOGRAPHIC BRAIN

The Holographic brain theory by Karl H. Pribram [1, 2], suggests that the brain holds memories in a Holographic manner. In a hologram, the data is not localized but distributed. Each part of the holographic recording film contains some information about

[1] Partially supported by Deutsche Telekom, Rita Altura Trust Chair in Computer Sciences, Lynne and William Frankel Center for Computer Sciences, Israel Science Foundation (grant number 428/11) and Cabarnit Cyber Security MAGNET Consortium.
[2] Partially supported by the Russian Foundation for Basic Research under grant RFBR 12-07-00109.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 142-150 (2013); doi: 10.1063/1.4776518

the entire image. Thus, reconstruction by a small piece of the holographic recording film results in a noisy version of the entire original image [3], as opposed to a classic photograph where holding a small piece of the film results in a sharp image of the specific piece. In the brain, local damage in a small part of the cortex may result in "graceful degradation" of memories rather than the complete loss of a specific memory. Experiments demonstrating the fact that the data in the brain is not localized go back to the 1920's where Karl S. Lashley was not able to completely abolish rats maze-habits by removing parts of their brain. The degradation was the same without correlation to the location of the lesion [4].

HOLOGRAPHIC MEMORY USING WALSH-HADAMARD TRANSFORM

A Hadamard matrix is a $n \times n$ orthogonal matrix with entries in $-1, 1$. The Walsh-Hadamard matrix is a specific Hadamard matrix [5]. The Walsh-Hadamard transform (the process of multiplication by a the Walsh-Hadamard matrix) is a holographic transform in a sense that each coefficient (entry in the output vector) is a result of a simple computation (additions and subtractions) that includes *all* the entries of the original input vector (each entry in the matrix is either 1 or -1 and multiplying by the matrix is equivalent to additions and substructions). Thus, each coefficient contains some information about all the entries in the input vector. Moreover, each entry in the input vector has the same weighted effect on each coefficient. Due to the fact that the Walsh-Hadamard transform can be calculated with additions and subtractions only, it can be very easily implemented by the simplest neural model whereas other holographic transforms such as discrete fourier transform require more complicated neurons or a more complicated network of simple neurons, for example, Velik [6] used linear threshold neurons. Figure 1 shows the Walsh-Hadamard transform on a $\{-1, 1\}^8$ vector. The output vector is the result of multiplication of the input vector by the Hadamard matrix. The input vector can be reconstructed from the output vector by multiplying the output vector by the transposed matrix and dividing each value by 8. In case a coefficient is corrupted, the resulting vector after reconstruction of the input vector from all the coefficients will contain minor errors (small change in the value) in all the entries rather than major errors (large change in the value) in part of the entries. This can be regarded as graceful degradation or blurriness. This example works for an input vector with integer or real values. For a binary input vector, there is no distinction between a minor and a major error because the only possible error is a bit flip. In this case we define blurriness as uniform distribution of the errors, namely, errors in the reconstructed input vector are uniformly distributed over the entries.

The well ordered charactarisics of the Walsh-Hadamard matrix may, in some cases, cause a pure distribution issue. A low entropy input vector in $-1, 1^n$ will result in an coefficients vector with poor coefficient distribution, namely, the coefficients vector will contain a few relatively large coefficients and a lot of zeros. in this case, erasure of one of the few large coefficient will result in a major change in the reconstructed original vector. This may occur because a well ordered input vector will have a high correlation with one of well ordered columns in the Walsh-Hadamard matrix, namely,

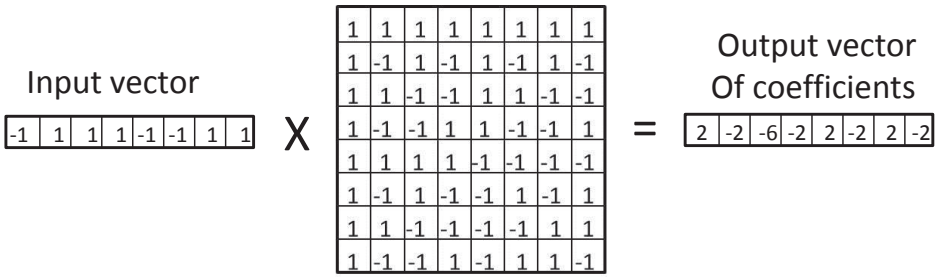


FIGURE 1. Walsh-Hadamard transform on a $\{-1,1\}^8$ vector.

the input vector will have a high amount of indices that have a common value with the corresponding indices of a specific column. This will result in a large coefficient because a large amount of multiplication result will be positive. Due to the orthogonality of the columns in the Walsh-Hadamard matrix, the higher the correlation with one column is, the multiplication with the other columns will be close to zero. Assuming that real data has low entropy and noise has high entropy, we suggest normalizing a binary input vector by xoring it with a random input vector. In case of an input vector in $\{-1,1\}^n$ the normalization can be done by multiplying each entry by the corresponding entry in a random vector in $\{-1,1\}^n$. After reconstruction, the result will be xored or multiplied again to receive the original vector. The randomization process is good for coefficient erasure. this is not the case for errors such as bit flips where the weight of an error is not connected to the original coefficient size, for example, in binary representation, a bit flip may add or subtract 2^k to the coefficient with probability that is not affected by the coefficient's original value.

The holographic model we suggest starts with xoring the original data with a random binary vector (or multiplying by the entries of a vector in $\{-1,1\}^n$) where the probability of a bit value 0 or bit value 1 is 0.5. Many researches consider various types of sparse coding of images projected to retina, which produce decorelated neurons firing signals in cortex [7], thus we assume that some kind of randomization of the sensor information (visual, in particular) of this type may occur in the brain. After the said normalization, we apply Walsh-Hadamard transform on the xored vector and save the output vector in memory. In order to reconstruct the original vector, we simply apply Walsh-Hadamard transform again, divide every entry of the result by the original length of the output vector, round the entries' values to the closest binary or $-1,1$ value and xor (or multiply) the result with the same random vector. The use of an erroneous coefficient fetched from the memory results in blurred version of the original vector.

The redundancy of neurons in the brain's memory is well-known [8]. One of the benefits of redundancy in the brain may be the ability to correct errors. Death of neurons may cause errors in stored memories. Thus, we suggest to incorporate an error correction technique in the holographic memory, expanding the work of Dolev et al. [9] that concentrated on erasure. The suggested technique corrects errors in the memory in a "single step" manner. In case of an error, there is no need of finding the closest codeword

before decoding, instead, the decoding procedure includes the correcting procedure. This error correction is done by replacing the original $n \times n$ Walsh-Hadamard matrix H used for the output vector calculation with a $n \times m, m > n$ rectangular Hadamard matrix [10], *error correcting Hadamard matrix* H_{ec} which is a matrix with entries in $-1, 1$, all rows of which are orthogonal to each other.

Figure 2 shows an example of the error correcting Walsh-Hadamard transform with $m = 32$ on a $\{-1,1\}^8$ vector. The output vector is the result of multiplication of the input vector by the matrix. The input vector can be reconstructed from the output vector by multiplying the output vector by the transposed matrix and dividing each value by 32. The output is a vector of size m. This shows the redundancy because a vector of size n is sufficient for reconstruction. The reconstruction is done by multiplying the saved output vector with the transposed $n \times m$ matrix H_{ec}^T, dividing the value of each entry in the reconstructed vector by m and rounding the entries values to the closest integer for an integer input vector. In case the input vector is in $\{-1,1\}^n$, the division and rounding can be replaced by setting the final value to 1, if the resulting value is greater than or equal to 0, and -1, otherwise. In case there are less than $m/2n$ errors for of an input vector in $\{-1,1\}^n, n = 2^k$ (as we prove in Theorem 1), the errors are automatically fixed. In case there are $m/2n$ errors or more for an input vector in $\{-1,1\}^n$, the result will have the same graceful degradation as before.

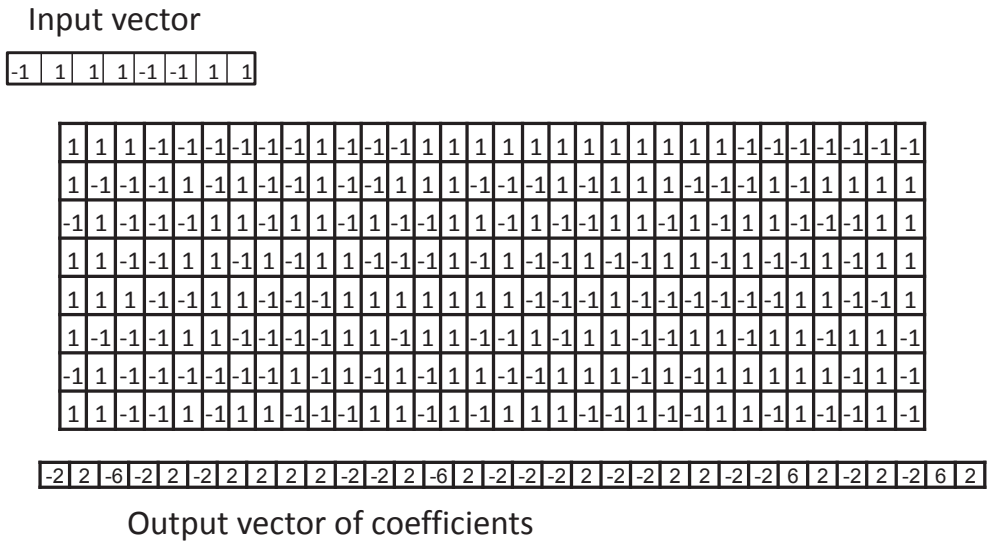


FIGURE 2. Error correcting Walsh-Hadamard transform with $m = 32$ on a $\{-1,1\}^8$ vector.

Theorem 1 *Let H_{ec} be an error correcting Walsh-Hadamard matrix which is a $n \times m$ Hadamard $n = 2^k, m > n$. H_{ec} can correctly reconstruct an input vector in $\{-1,1\}^n$ given that there are up to $m/2n - 1$ erroneous bits in the coefficients vector and that the coefficient vector is represented by binary code.*

Proof. When there is no error, the value of each coefficient can range between n and $-n$ (or between 0 and $2n$). Thus each coefficient can be represented with $log(2n)$ bits. The maximal change that one bit can cause in a coefficient is adding or subtracting $2n$ by

flipping the most significant bit. The reconstruction of each entry of the original vector is done by a series of additions and subtractions that correspond to one of the columns in H_{ec}^T and setting the final value to 1 if the result is greater than 0 and -1 otherwise. In case the result is 1 when there are no errors in the coefficient vector, the value can be reconstructed correctly when the result is any number greater than 0. This means that if the total sum of the errors is less than m, the value will be reconstructed correctly. In the worst case , a change of size m can be caused by $m/2n$ bits because each bit causes a $2n$ error. The same proof also apply to the case where the reconstructed value is supposed to be -1.

NEURAL IMPLEMENTATION

In order to show the feasibility of our brain coding model, we provide a neural implementation for the coding procedure. We use simple"McCulloch-Pitts" neurons [11] with real neurons weights in order to show that the procedure is feasible even with very simple neurons. The randomization part is done by an array of randomly distributed *signal inverting units* (neural units that output the opposite of their input value) and *signal conserving units* (neural units that output the same as their input value). For convenience, we use in some parts signal conserving or inverting units that also change the input form $\{0,1\}^n$ (firing and non-firing neurons) to be a vector in $\{-1,1\}^n$ (all firing neurons with positive or negative post-synaptic weights). The error correcting Walsh-Hadamard transform is performed by additions and subtractions of the randomized input in a manner that fits the additions and subtractions done by multiplying the input by the $n \times (k \cdot n)$ error correcting Hadamard matrix. Through all the process subtractions are done by signal inverting units. The coefficients vector is a vector of integers rather than binary units. For convenience we represent the coefficients with a binary representation of neurons (firing and non-firing) although other representations may also be considered without changing the essence of the procedure.

Figure 3 shows the High level view of the entire procedure for an input in $\{0,1\}^8$ and $m = 32$. The first step is the normalization of the input vector done by connecting each neuron to a signal conserving or signal inverting unit. The signal inverting unit used in this part appears in the bottom of Fig. 4. This unit maintain the vector space (keeping it in $\{0,1\}^n$). In this case signal conserving is done by not using a signal inverting unit. The second step is the error correcting Walsh-Hadamard with $m = 32$. The vector of 32 coefficients is built from 32, $5-$*neuron* units. The use of 5 neurons for each coefficient is due to the use of binary representation of numbers between -8 and 8. For convenience we add 8 to each coefficient to get values between 0 and 16 which are represented by 5-firing/non-firing neurons. Figure 5 shows the calculation of the second coefficient. The positive and negative weights are implemented by the signal conserving and signal inverting units (the two top units in Fig. 4) respectfully. The addition of 8 is done by a constantly firing neuron with a post-synaptic weight of 8.

In order to obtain binary representation, each neuron represents a bit. The upper neuron represents the least significant bit and the lowest one represents the most significant bit. The least significant bit neuron has a threshold 1, the next neuron has a threshold 2, the next neuron has a threshold 4 etc. All the neurons receive the same inputs. The most

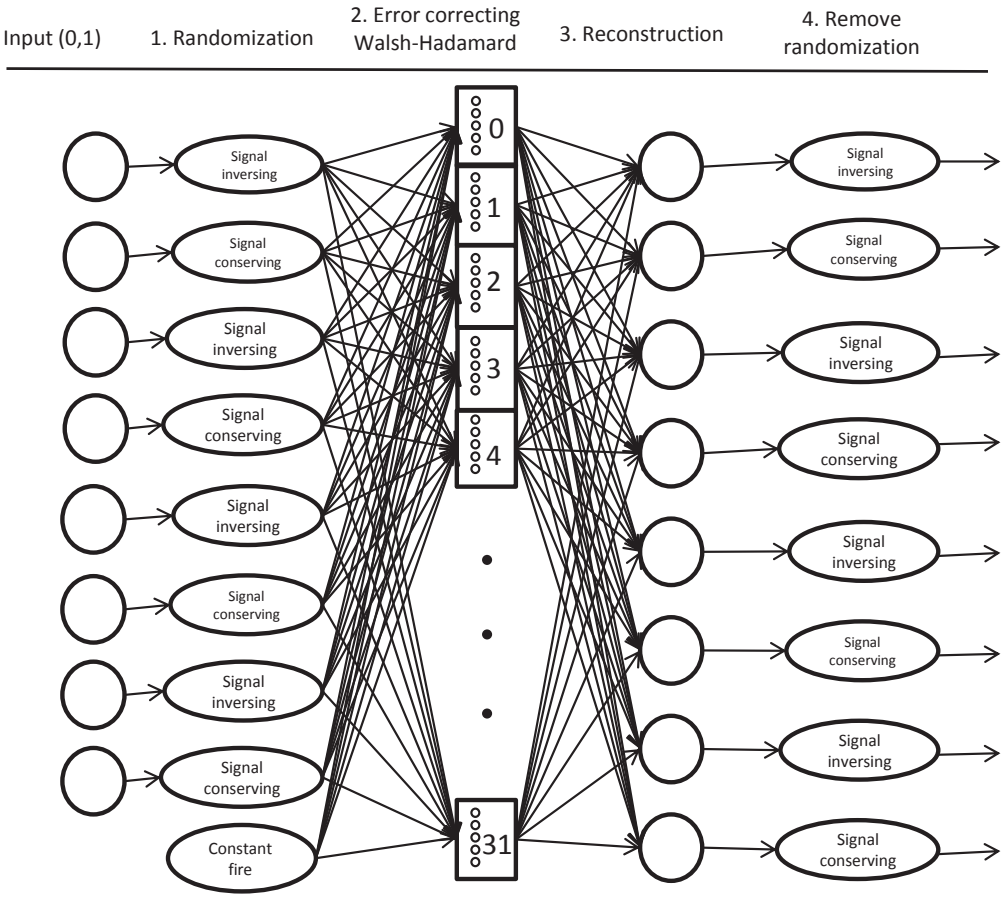


FIGURE 3. Neural implementation of the entire procedure.

significant bit (which in our case has a threshold 16) sends synapses of weight -16 to all the less significant bit neurons. Thus if the most significant bit is on, the rest bits are off because in our case 16 is the largest value that a coefficient can have. The neuron with threshold 8 sends synapses of weight -8 to all neurons that represent less significant bits. The neuron with threshold 4 sends synapses of weight -4 to all neurons that represent less significant bits and so on. Thus, the coding is done "recursively", if the coded number is 16, the neuron representing 16 fires and no other neuron fires. If the coded number is less than 16 but greater than or equal to 8, the neuron representing 8 fires and the less significant neurons code the rest of the number (following the subtraction of 8) and so on.

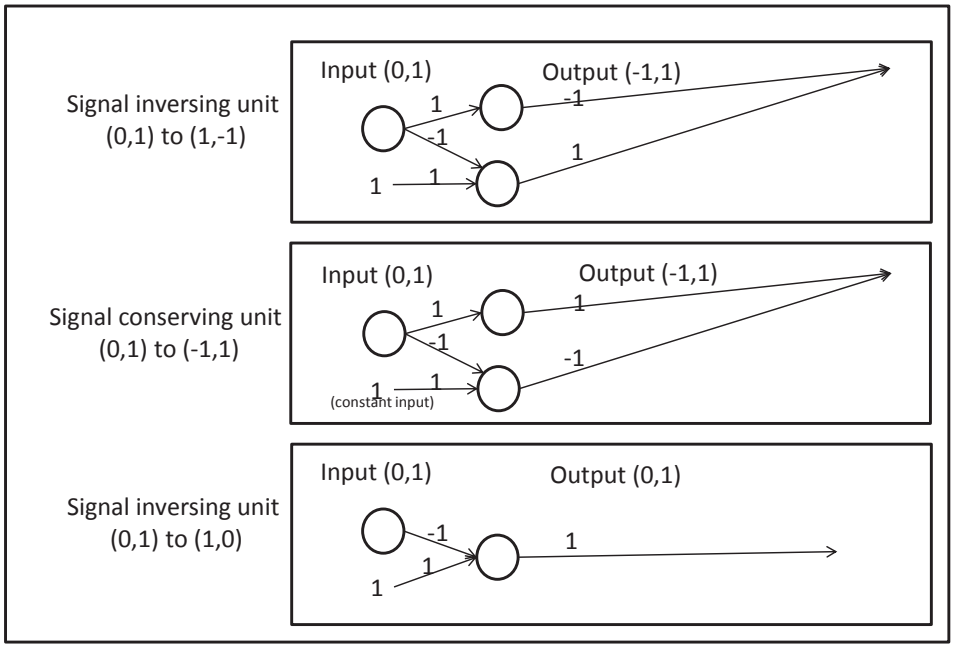


FIGURE 4. Signal inversing and conserving units.

The next step is the reconstruction of the normalized vector (Fig. 6 shows the reconstruction of the second entry of the normalized vector) by multiplying the coefficient vector by the transposed error correcting Walsh-Hadamard matrix H_{ec}^{T} which is again a combination of additions and subtractions. Addition of a binary represented number is done by sending a synapse of weight 16 from the most significant bit neuron, then a synapse of weight 8 from the next neuron etc' while subtracting the 8 that was added for convenience in the encoding step. Instead of dividing each value in the resulting vector by 32 and rounding it to the closest integer, it is possible to set the value to 1 if the value is greater than 0 and -1 otherwise, because the original vector was in $\{-1,1\}^{8}$. In order to do so, the resulting vector is represented by neurons with threshold 0.

CONCLUSIONS AND FUTURE PLANS

Holographic memory has some benefits over standard memory in cases of errors in the saved data or availability to only parts of the data, in applications that prefer some knowledge on the entire data over whole information on part of the data. The ability of the brain to retrieve some information about memories after some cases of local damage might be explained using such memory.

In this work we presented a simple modification to the Walsh-Hadamard transform which allows storing data in an error correcting holographic manner. A neural implementation was also given. The implementation is given to show feasibility of such model

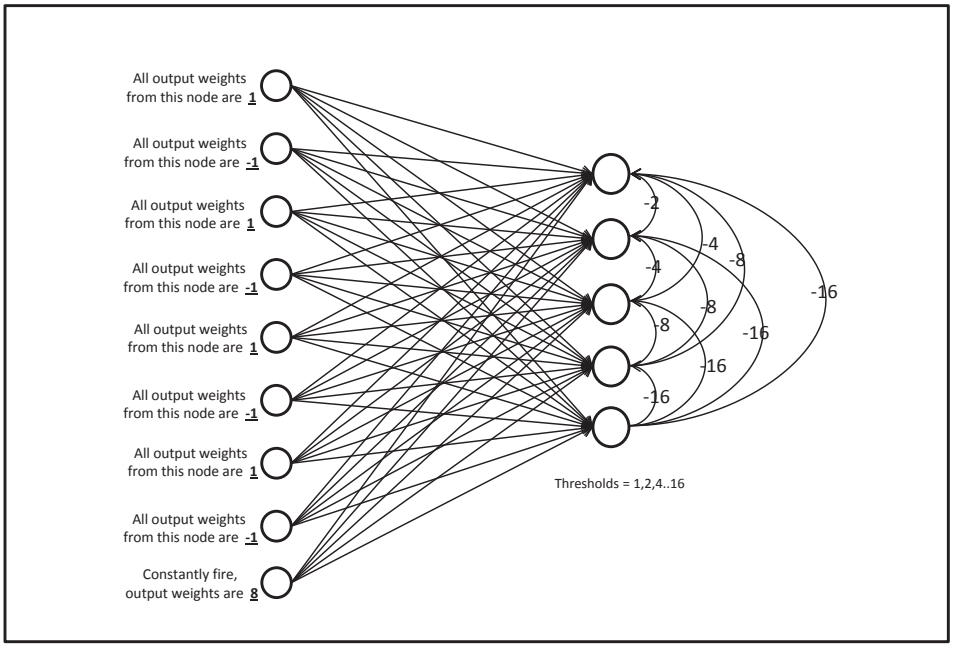


FIGURE 5. Calculation of the second coefficient of the error correcting Walsh-Hadamard transform.

in the brain. The simple "McCulloch-Pitts" neuron model was used in order to show that this type of memory can be built easily without the use of more complicated models. The well ordered binary representation in the model was given as an example and does not imply the belief that the brain holds information in this manner. Our implementation of the normalization part requires a type of random vector of signal inverting and conserving units. This type of arrangement might be available in the brain from birth or after some learning procedure that takes place in early stage of life.

In this paper we did not supply the actual memory saving part but only the data manipulation that allows saving data in a holographic manner. Future research may include adding the part in the implementation that will save the transformed data. One of the options for this part is to use the Hopfield network [12]. In the current implementation of the Holographic memory, the magnitude of the damage may change according to the specific damaged bit. This property is not holographic because it means that some bits are riskier than others. This can be improved by adding an error correction method that protects significant bits more than it protects less significant bits. Testing the applicability of the procedure to other fields such as communication is also part of possible future research, for example sending some coefficients of the transformed data when the communication is expensive in order to immediately convey information on the entire data to the receiver.

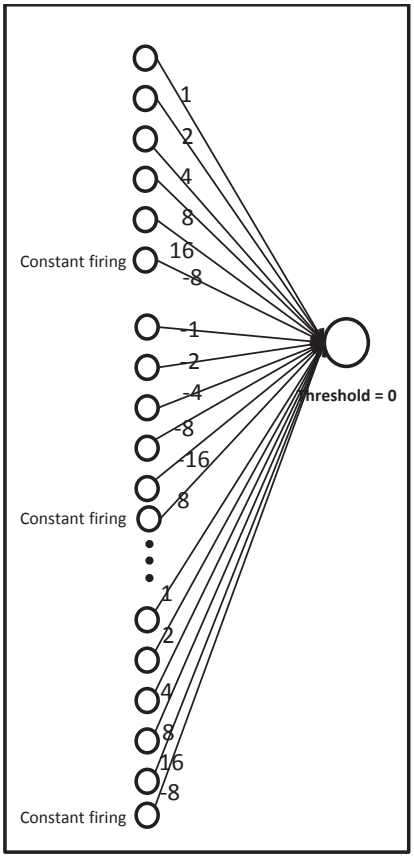


FIGURE 6. Reconstruction of the second entry in the normalized input vector.

REFERENCES

1. K. H. Pribram, *Languages of the brain: experimental paradoxes and principles in neuropychology*, Prentice-Hall, 1971.
2. K. H. Pribram, *Brain and perception: holonomy and structure in figural processing*, Lawrence Erlbaum Associates, 1991.
3. D. Gabor, *Proceedings of the IEEE* **60**, 655–668 (1972).
4. K. S. Lashley, *Brain Mechanisms and Intelligence. A Quantitative Study of Injuries to the Brain*, University of Chicago Press, 1929.
5. S. Agaian, H. Sarukhanayan, K. Egiazarian, and J. Astola, *Hadamard transforms*, SPIE press, 2011.
6. R. Velik, "Discrete Fourier Transform Computation Using Neural Networks," in *International Conference on Computational Intelligence and Security, 2008. CIS '08.*, 2008, vol. 1, pp. 120–123.
7. W. E. Vinje, and J. L. Gallant, *Science* **287**, 1273–1276 (2000).
8. J. M. Fuster, *Memory in the Cerebral Cortex An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*, MIT press, 1999.
9. S. Dolev, and S. Frenkel, "Multiplication free holographic coding," in *2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, 2010, pp. 146–150.
10. S. Kishore, *Stat. & Prob. Lett.* **6**, 141–142 (1988).
11. W. McCulloch, and W. Pitts, *Bull. Math. Biol.* **5**, 115–133 (1943).
12. J. J. Hopfield, *Proc. Nat. Acad. Sci. USA* **79**, 2554–2558 (1982).

Scaling and intermittency of brain events as a manifestation of consciousness

P. Paradisi[*], P. Allegrini[†,**], A. Gemignani[‡,**], M. Laurino[†,**], D. Menicucci[†,**] and A. Piarulli[§,**]

[*]*Istituto di Scienza e Tecnologie dell'Informazione (ISTI-CNR), Via Moruzzi 1, 56124 Pisa, Italy.*
[†]*Istituto di Fisiologia Clinica (IFC-CNR), Via Moruzzi 1, 56124 Pisa, Italy.*
[**]*Centro EXTREME, Scuola Superiore Sant'Anna, P.zza Martiri della Libertà 7, 56127 Pisa, Italy.*
[‡]*Dipartmento di Scienze Fisiologiche, Università di Pisa, Via San Zeno 31, 56127 Pisa, Italy.*
[§]*PERCeptual RObotics laboratory (PERCRO), Scuola Superiore Sant'Anna, P.zza Martiri della Libertà 7, 56127 Pisa, Italy.*

Abstract. We discuss the critical brain hypothesis and its relationship with intermittent renewal processes displaying power-law decay in the distribution of waiting times between two consecutive renewal events. In particular, studies on complex systems in a "critical" condition show that macroscopic variables, integrating the activities of many individual functional units, undergo fluctuations with an intermittent serial structure characterized by avalanches with inverse-power-law (scale-free) distribution densities of sizes and inter-event times. This condition, which is denoted as "fractal intermittency", was found in the electroencephalograms of subjects observed during a resting state wake condition. It remained unsolved whether fractal intermittency correlates with the stream of consciousness or with a non-task-driven default mode activity, also present in non-conscious states, like deep sleep. After reviewing a method of scaling analysis of intermittent systems based of event-driven random walks, we show that during deep sleep fractal intermittency breaks down, and re-establishes during REM (Rapid Eye Movement) sleep, with essentially the same anomalous scaling of the pre-sleep wake condition. From the comparison of the pre-sleep wake, deep sleep and REM conditions we argue that the scaling features of intermittent brain events are related to the level of consciousness and, consequently, could be exploited as a possible indicator of consciousness in clinical applications.

Keywords: critical brain; fractal intermittency; renewal point processes; diffusion scaling; consciousness.
PACS: 05.40.-a, 87.10.-e, 89.75.Da, 87.19.le

INTRODUCTION

Information processing in the brain is driven by highly nonlinear interactions among neurons, with a high tendency to generate collective behavior, self-organized structures and clustering at several time and space scales. Clusters, or neural assemblies, that emerge at some scale, interact with clusters formed at some other scales with a continuous dynamical interactions among different scales. This is associated with a very rich dynamics that is thought to be associated with the emergence of consciousness [1, 2]. This scale-to-scale interaction in brain dynamics is nowadays recognized to involve universal mechanisms and features of emergent complexity and critical phenomena [3, 4]. Avalanches, scale-free or self-similar behavior, long-range correlations and burstiness are found, both experimentally and theoretically, in many complex systems and, in particular, in neuronal networks [5, 6, 7, 8, 9].

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 151-161 (2013); doi: 10.1063/1.4776519

The hypothesis of a critical brain is attracting the interest of the scientific community, with particular attention to the emergence of consciousness. Many neurobiologically plausible models of brain dynamics that are focused on explaining the emergence of consciousness include typical features of critical systems. Among others, it is worth citing the Global Workspace [10, 11], the Dynamic Core [11, 12] and the Operational Architectonics theory [1, 2].

Criticality in the brain. The emergence of collective, self-organized behavior, associated with scale-free or power-law behavior and long-range correlations, is typically observed in dynamical systems posed near a *critical point*. In critical phenomena [13] and self-organized criticality [14, 15], the system typically moves towards a critical point, corresponding to a phase transition from an uncorrelated to a correlated condition. This transition is characterized by means of the critical value of a cooperation parameter driving the non-linear coupling among many individual units. The phase-transition hypothesis for the brain dynamics was already discussed by Turing in his pioneering work [16], where he conjectured that an intelligent system cannot "live" neither in a too much correlated condition (order, super-critical), nor in a too chaotic one (disorder, sub-critical). Consequently, brain is expected to operate in the intermediate region between these two extremes, i.e., in a critical condition. This is confirmed by recent neurophysiological literature (see, e.g., [17] or, more recently, [4, 18]). The robust behavior and the great plasticity of the brain dynamics are argued to be strictly related with this critical condition [19].

Many authors investigated the spatial and/or structural complexity of neuronal network models, *in vitro* data and functional Magnetic Resonance Imaging (fMRI) of the human brain and found features in agreement with criticality. In particular, the authors of Refs. [5, 6, 7] found a scale-free distribution of avalanche (or cluster) sizes, which is a signature of spatial and structural long-range correlations, in network models and *in vitro* data. The authors of Refs. [3, 4] studied the functional connectivity of the brain defined through the network of above-threshold cross-correlations derived from fMRI data, which is again a structural property. They evaluated the degree distribution, being the degree of a node the number of links of that node with other nodes, and found a scale-free degree distribution similar to that of the Ising magnetization model at the critical point, corresponding to a second-order phase transition in the magnetization field.

Intermittency in critical systems. The above cited studies about criticality in the brain are focused on the spatial or structural complexity. A often overlooked property is the temporal complexity, where the focus is on the *time* long-range correlations with power-law decay (equivalent to $1/f$ noise) and on *time intermittency*, which is defined by the presence of *crucial events* in the complex/critical system.

The authors of Refs. [8, 20, 21] found that the fluctuations of a random field at the critical point, i.e., the "order parameter" averaging microscopic fluctuations, are described in terms of a *Type-I intermittent* dynamical map similar to the well-known Manneville map [22], which mimics turbulent bursting. This kind of dynamical systems is characterized by the presence of a marginally unstable point determining an alternation between long time intervals with calm motion and short-time bursting events. These events, occurring in the temporal evolution of the order parameter, are described by a serial fractal

point process, i.e., a sequence of intermittent events that: (a) occur randomly in time and (b) display a slow (power-law) decay in the distribution of inter-event or Waiting Times (WTs). Notice that this point process has to be interpreted as a birth-death process of cooperation, where the cooperation is here represented by the intermittent formation and decay of neural assemblies in the brain dynamics. Type-I intermittency is in agreement with a fast decay of memory in correspondence of event occurrences. In the language of stochastic processes, this is described by a *renewal point process* [23], which is defined by the condition of mutual statistical independence of the events and, consequently, of the WTs. In other words, the macroscopic fluctuations of a critical system are driven by a renewal point process, which is the mathematical tool used here to describe intermittency and, in the case of a self-similar of fractal distribution of WTs, *fractal intermittency*, The renewal condition is related to burstiness with fast memory decay and it was found to characterize the intermittency features of several complex systems, from blinking quantum dots [24, 25] to turbulence [26, 27, 28] and brain dynamics [9]. The renewal property seems to play a crucial role in the perturbation of complex systems [29, 30, 31, 32, 33] and it is a fundamental assumption in the derivation of a new Fluctuation-Dissipation Theorem (FDT) based on renewal events [34, 35], whose main prediction is that two complex systems have a maximum interaction when they have similar complexities. The power-law relaxation foreseen by this new FDT was also experimentally validated in the weak turbulence regime of a liquid crystal [36]. Regarding biological systems and, in particular, brain dynamics, we can roughly say that the renewal condition allows to reduce the disorder of the system, as the entropy is significantly increased only in correspondence of event occurrences, while a long memory characterizes the system's evolution in between two events, being this related to the observed long-range correlations. At the same time, the dynamics of events, associated with metastable states, allows for a greater capability of adaptation to external stimuli [2]. Thus, the renewal condition seems to be the only one that can deal, at the same time, with the need of a slow disorder increase (long-range correlations) and a sufficiently rapid adaptability to environmental stimuli (memory erasing events).

Intermittency and consciousness. The existence of crucial events in the brain is well-established, as spontaneous neuronal activity exhibits relatively quiet periods in alternation with chaotic or bursty periods. Such brain events can be extracted from ElectroEncephaloGram (EEG) data with detection algorithms. Events are here defined as abrupt transitions or Rapid Transition Processes (RTPs) [1, 37, 38]) to and from metastable states, via multichannel EEGs [9]. On short time scales brain events typically display a complex structure in terms of neuronal avalanches [39].

Exploiting the concept of RTP events, we investigated the temporal complexity of brain dynamics in terms of intermittency features [9, 39]. We found that a serial renewal process of global integration exists in the human brain during a *resting state* wake condition and that this renewal process has well-defined scaling exponents in both distributions of avalanche sizes and inter-event times [9, 39]. These scaling exponents, being a signature of Type-I fractal intermittency, confirm the critical brain hypothesis [19]. The scaling exponents were evaluated through the diffusion scaling of different random walks driven by the RTP events (see details in the next section). This approach based on diffusion scaling allowed to get a robust estimation of the intermittency exponent or

complexity index μ, i.e., the exponent of the inverse power-law tail in the WT distribution: $\psi(\tau) \sim 1/\tau^{\mu}$. It is worth noting that similar approaches, based on brain events and point processes, have been recently applied, confirming the robust and universal critical behavior of brain dynamics and neuronal networks [40, 41].

All the above findings lead to the idea that consciousness is related with the emergence of criticality and fractal intermittency. However, this is just a hypothesis as it is not yet clear if this renewal fractal process is uniquely associated with consciousness or with a non-task-driven default mode activity [42], also present in non-conscious states like deep sleep.

In this paper we clarify this point by evaluating the event-driven diffusion scaling of EEG data collected from the observation of healthy human subjects during sleep. The statistical analysis we use is essentially the same as in Ref. [9]. In "Data Description and Methods of Analysis" section we describe the dataset and the methods of data analysis. In particular, we will introduce the diffusion scaling method. In "Results and Discussion" section we show our results and we discuss the hypothesis that the emergence of intermittent events described by a (serial) renewal fractal process and of anomalous diffusion is a signature of consciousness, while the lack of fractal features and the emergence of normal diffusion could characterize non-conscious states.

DATA DESCRIPTION AND METHODS OF ANALYSIS

A normal night's sleep consists of a few (from 4 to 6) cycles, each cycle consisting of different phases, defined on the presence of different "waves", or graphoelements, and specific rhythms. After a pre-sleep wakefulness, the first cycle begins with a shallow sleep called N1. As the sleep deepens, due to the diminished presence of various neurotransmitters, sleep phases N2 and N3 (or Slow Wave Sleep, SWS) are visited one or more times, till the Rapid Eye Movement (REM) phase (typically a dreaming phase) occurs, that marks the end of the cycle. The phases N1, N2 and N3 (or SWS) are globally referred to as Non-REM (NREM) phase. At variance with NREM phase, REM is characterized by a high level of the acetilcholine (AC) neurotransmitter. At the end of the first cycle, a second cycle begins, with or without N1 or wakefulness episodes (Wakefulness After Sleep Onset, WASO), with the presence of NREM sleep (AC again drops to low values), again ending with a REM phases, and so on.

Data set. Our data set is composed of 29 whole-night high-density (128 channel, 4ms sampling time) EEG recordings. Subjects slept two nights with the same experimental setup, namely after an adaptation night the second one was recorded. All subjects signed informed consent according to local ethical committees. Through visual inspection of the polygraphic traces, namely a selection of few EEG channel plus miogram (muscle tone intensity) and oculogram (eye movements) all recordings were segmented into different cycles and phases. For the purpose of the present paper, however, we will focus on global properties of sleep in the various phases and we will freely make recourse to grand averages over the 29 whole-night recordings. Artifacts were semiautomatically removed, and only artifact-free segments of time duration longer than 3 min-

utes were kept for the RTP detection. We use only segments of the first cycle, as signal quality decreased in subsequent cycles.

Rapid Transition Processes. Herein, for each EEG channel, pass-band filtered between 0.3 and 40 Hz (Chebyshev II filter algorithm), RTPs are extracted as a "significant" selection of intersection between two different moving averages of the Hilbert transform of the signal modulus. Moving averages have windows of 5 and 125 ms, respectively. By significant we mean that we select only the points where the intersection between the two curves is above a threshold angle. To do this we select a 125 ms window surrounding the intersection and compute the sum of the modulus of the difference between the two curves. For each channel significant RTPs are those in the highest decile (the ones higher than 90% are chosen). This method is inspired and similar to that introduced in Ref. [37], but with slightly differences. We however proved in [43] that our subsequent analysis is robust with respect to a variation of event definition.

We are here interested on global events, i.e., on the "simultaneous" occurrences of RTP in different EEG channels. For each EEG recording, the sequence of coincidences, or (concurrent) Multi-Channel RTPs (MC-RTPs), is obtained from single-channel RTPs via the introduction of two thresholds: The first one, Δt_c, defines the maximum time distance for two single-channel RTPs (from different channels) to be considered concurrent; the second one, N_t, defines the minimum number of concurrent single-channel RTPs required for a MC-RTPs to be recorded as a global event. Since events that have a distance less than Δt_c are considered to be simultaneous, Δt_c must be small. We herein use $\Delta t_c = 4\,\mathrm{ms}$, equal to the instrumental sampling time, and $N_t = 5$.

Event-driven random walks and diffusion scaling. The random walks driven by renewal events [9, 28] are inspired to the Continuous Time Random Walk (CTRW) of Montroll and co-workers [44, 45]. In CTRW it is allowed to have random time steps, corresponding to a sequence of WTs from a renewal process. Here, the WT sequences derived from the EEG recordings are used to define two different CTRWs driven by the same RTP global events. Firstly, we introduce a discrete artificial signal $\xi(t)$, i.e., a kind of random discontinuous velocity that changes value only in correspondence of event occurrences. In Figs. 1 and 2 a sketch of the two signals $\xi(t)$ is reported. The times $t_0,\ t_1,\ t_2, \ldots$ correspond to the occurrence of the events $0,\ 1,\ 2, \ldots$, while $\tau_1,\ \tau_2, \ldots$ are the WTs, i.e., the time interval between the events 0 and 1, the events 1 and 2 and so on. In particular, we have:

(a) Asymmetric Jump (AJ) rule:
the walker makes a positive jump ($\xi(t_n) = 1$) in correspondence of each event n, otherwise it stands ($\xi(t) = 0$). Then, $\xi(t)$ is a sequence of pulses of constant intensity.

(b) Symmetric Jump (SJ) rule:
as in the AJ rule, but the walker can make positive or negative jumps in correspondence of an event: $\xi(t_n) = \pm 1$. The sign $\pm$ is chosen with a coin tossing prescription.

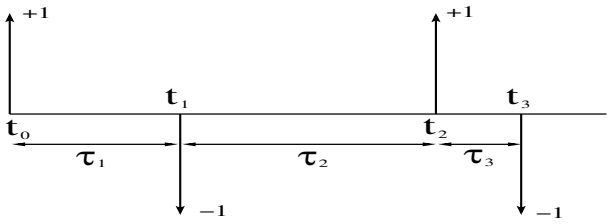


FIGURE 1. The SJ walking rules for the "velocity signal" $\xi(t)$.

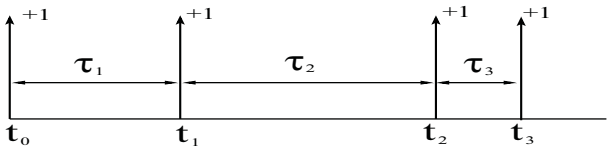


FIGURE 2. The AJ walking rules for the "velocity signal" $\xi(t)$.

Then, from the artificial signal $\xi(t)$ the diffusion variable of the CTRW is defined as follows:

$$X(t) = X_0 + \sum_{j=0}^{j=t} \xi(j)\,\Delta t\ , \tag{1}$$

being Δt the sampling time of the experimental time series.

The scaling properties of these random walks were extensively investigated in several papers (see [9, 27, 28] for a brief review) by applying the analytical methods of CTRW. Here we are interested in the scaling exponent H of the second moment

$$\sigma^2(t) = \left\langle \left(X(t) - \overline{X}\right)^2 \right\rangle \sim t^{2H}\ , \tag{2}$$

where $\overline{X}$ is the mean value of $X(t)$.

Analytical expressions of the scaling H as a function of the complexity index μ were determined in the case of renewal WTs with inverse power-law tail: $\psi(\tau) \sim 1/\tau^{\mu}$. These

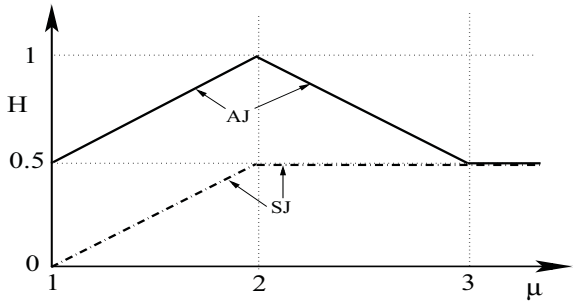


FIGURE 3. Diffusion scaling H vs. complexity index μ for SJ and AJ walking rules: AJ (continuous line), SJ (dotted-dashed line).

expressions $H = H(\mu)$ are reported in Fig. 3 and summarized in the following:

(AJ)

$$H_{AJ} = \begin{cases} \mu/2\,; & 1 < \mu < 2 \\ 2 - \mu/2\,; & 2 \leq \mu < 3 \\ 1/2\,; & \mu \geq 3 \end{cases} \tag{3}$$

(SJ)

$$H_{SJ} = \begin{cases} (\mu - 1)/2\,; & 1 < \mu < 2 \\ 1/2\,; & \mu \geq 2 \end{cases} \tag{4}$$

Both rules give a normal scaling $H = 1/2$ for $\mu \geq 3$, corresponding to normal (Gaussian) diffusion. For the SJ rule this is true also in the range $2 < \mu \leq 3$, while AJ rule is super-diffusive ($H > 1/2$) in all the interval $1 < \mu < 3$. On the contrary, the SJ rule is sub-diffusive ($H < 1/2$) for $1 < \mu < 2$. We note that, if the WTs comes from a Poisson process, the value of H is again $1/2$ and, in the long-time, we have a Gaussian diffusion.

The joint use of these walking rules can be used to evaluate the value of the μ by inverting the expressions given in Eqs. (3-4). It can be seen from Fig. 3 that $H_{AJ}(\mu)$ is not an invertible function, as the same value of H corresponds to two distinct values of μ, one smaller and the other greater than 2. When $H_{SJ} < 1/2$ it results $\mu < 2$ and both rules, i.e., the associated values of μ derived from AJ and SJ rules, could be compared to each other. On the contrary, for $H_{SJ} = 1/2$, a value of μ cannot be derived from the SJ rule, but we can assume $\mu > 2$. For this reason, the SJ rule could be used to discriminate between $\mu < 2$ and $\mu > 2$, overcoming the ambiguity of AJ rule.

Detrended Fluctuation Analysis. The diffusion scaling H of the two random walks introduced above is estimated by means of Detrended Fluctuation Analysis (DFA) [46]. We briefly recall the main steps of this method:

- For a discrete time $L = 4, 5, ...$, the time series of the diffusion process $X(t)$ is split into not-overlapping time windows of length L: $[kL+1, kL+L]$. The window number is given by $[M/L]$, i.e., the integer part of M/L, being M the total length of the time series.
- For each time window $[kL+1, kL+L]$ ($k = 0, 1, ..., [M/L]$), the local trend is evaluated with a least-squares straight line fit: $\overline{X}_{k,L}(t) = a_{k,L}t + b_{k,L}$; $kL < t \leq (k+1)L$.
- The fluctuation is derived in the usual way: $\widetilde{X}_{k,L}(t) = X(t) - \overline{X}_{k,L}(t) = X(t) - a_{k,L}t - b_{k,L}$; $kL < t \leq (k+1)L$.
- For a given time scale L, the mean-square deviation of the fluctuation is calculated over every window:

$$F^2(k,L) = \frac{1}{L}\sum_{t=kL+1}^{(k+1)L} \widetilde{X}^2_{k,L}(t) = \frac{1}{L}\sum_{t=kL+1}^{(k+1)L} \left(X(t) - \overline{X}_{k,L}(t)\right)^2 \tag{5}$$

- Finally, an average over the windows is performed:

$$F^2(L) = \frac{1}{[M/L]} \sum_{k=0}^{[M/L]} F^2(k,L) \tag{6}$$

In the case of a self-similar process, it results: $F(L) \sim L^H$. Then, by defining $z = \log(F(L))$ and $y = \log(L)$, it is possible to apply a least-squares straight line fit:

$$z = Hy + C \,, \tag{7}$$

where C is a constant.

Improvement of statistical accuracy in DFA. Given a time series of total length L, the DFA evaluation is reliable up to about $L/10$ and this is due to the lack of statistics in the long-time regime. However, we do not have only one time series, but several independent time segments, each one separated from the others by at least one artifact or phase shift in the original EEG recording. Several DFA curves can be obtained, one for each time segment, and then averaged to get a mean DFA curve. In this way, we are able to compute DFA up to a time given by the maximum among the values $L_i/10$, that is $\max_i(L_i/10)$, where i runs over all time segments and L_i is the total duration time of the i-th time segment. Actually, the statistical accuracy remains stable up to a time given by $\min_i(L_i/10)$ and then decreases for longer time scales. In fact, the number of segments entering the average decreases very rapidly when approaching the time scale $\max_i(L_i/10)$.

We improved the statistical accuracy on longer time scales, without the risk of making the running window explore segments that belong to different segments. Firstly, for each sleep phase, we evaluated the minimal duration time: $L_m = \min_i(L_i)$; then, for each segment, we computed the DFA up to time L_m; finally, we performed the average over all the segments. Note that L_m is not only 10 times greater than $\min_i(L_i/10)$, but it is also greater than $\max_i(L_i/10)$. With this approach, a much better accuracy on long time scales is obtained. In fact, even if the statistical accuracy is low for the segments with the shortest duration times, the number of segments entering the average is greatly increased in the time range between $\min_i(L_i/10)$ and $L_m = \min_i(L_i)$, as all segments always enter in the average operation.

RESULTS AND DISCUSSION

Criticality has been found both in neuronal networks (models and *in vitro*, see Refs. [5, 6, 7]) and human brain [3, 4] by investigating the spatial and structural complexity, while temporal complexity, i.e., time intermittency, in brain EEG was investigated in our previous papers (see Refs. [9, 39]). In particular, from the analysis of EEG data in resting state (wakefulness) condition we found that the brain (RTP) events introduced by the authors of Refs. [1, 37] are driven by an underlying renewal fractal point process with well-defined scaling properties (**fractal intermittency**). As already said in "Introduction" section, it is not clear if fractal intermittency is uniquely associated with

consciousness or with a non-task-driven default mode activity [42], also present in non-conscious states like deep (NREM) sleep. To clarify this point, let us summarize some observations about consciousness:

1. the conscious brain is associated with an emerging "giant cluster" or Global Workspace [10]) that co-exist with clusters of any size having scale-free size distribution, in analogy with what happens in critical systems [6, 7];
2. Conscious scenes are unitary and occur serially: only one scene at a time takes place [10];
3. consciousness is a sequence of metastable states (giant clusters), which reflect rapidly adaptive selection mechanisms in perception and memory; in the consciousness theory of Baars [10], the Global Workspace is an emerging serial process that, in some way, selects only one scene at a time from an underlying set of parallel scenes, and only this selected scene comes into consciousness;
4. In conscious states, there's a competition among cooperative global integration and autonomous fragmentation; the interplay of these two components constitutes the metastable regime of brain dynamics and determines the complex intermittent behavior in the EEG field [2, 12, 47].
5. the renewal fractal process derived from EEG data, which is defined by the sequence of renewal RTP events, is a particular serial process, as only a global metastable state (giant cluster) at a time takes place and the short-time RTP events mark the death of a metastable state and the birth of a new one [9, 39].

From the above observations, we are then lead to make the following assumption:
The **renewal point process** describing **fractal intermittency**, which is experimentally defined in EEG data by the sequence of **global RTP events** with inverse power-law distributed WTs, is a **correlate of consciousness**.
We validate this assumption by comparing different states of consciousness in healthy subjects during sleep. In "Data Description and Methods of Analysis" section we have already given a description of the dataset and of the methods used to analyze the EEG data, which can be summarized as follows: (a) segmentation and artifact removal; (b) RTP detection, global brain events (c) computation of event-driven random walks (SJ and AJ) and estimation of second moment scaling H by applying DFA. The diffusion scaling H of the SJ rule is definitively $H = 0.5$ for all time segments and subjects and, then, also for the mean DFA. This is a signature that the complexity index μ is greater than 2. In Fig. 4 we show the square root of the second moment $\sigma(t)$ for the AJ rule, averaged over all subjects and nights and over the time segments of sleep cycle I as explained at the end of "Data description and Methods of Analysis" section. The second moment scaling H switches from an anomalous diffusion scaling ($H = 0.75$) in the case of (pre-sleep) wake and REM phases to a normal diffusion scaling ($H = 0.5$) in deep (SWS) sleep. Inverting Eq. 3, this means that in wake and REM phases, which are conscious states, we get an average value $\mu = 2.5$, thus giving fractal intermittency and long-range correlations, whereas in the deep (SWS) sleep phase we get $\mu > 3$. We recall that normal diffusion ($H = 0.5$) is also in agreement with a Poisson condition, i.e., with exponentially distributed WTs or, more realistically, with an exponential cut-off emerging at relatively short WTs and, thus, with short-time correlations.

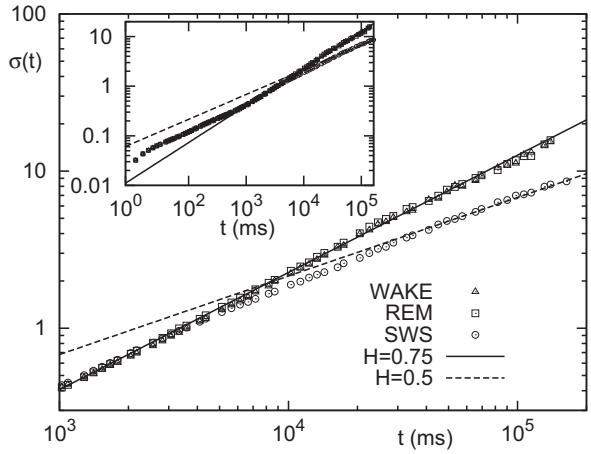


FIGURE 4. Asymptotic time range in the DFA computed for AJ rule applied to different sleep phases (cycle I). Continuous and dashed lines are a guide to the eye for the slopes $H = 0.75$ and $H = 0.5$, respectively. In the inset we report the entire time range over which the DFA has been computed. Notice that, in the short-time range, the DFA of the three phases (WAKE, REM and SWS) are essentially superposed, all displaying normal diffusion.

The normal diffusion regime during SWS phase could be explained in terms of the fragmentation of the Global Workspace into local, independent, functional units working in parallel, which is a condition known to be associated with the lack of consciousness. Notice that the fragmentation is related to the large number of Sleep Slow Oscillations (SSOs) during SWS [48], which determine a reset of the neuronal activity by means of a hyper-polarizing wave putting most neurons in a down-state, i.e., a state far from the activation threshold of the membrane potential. This is also called "electrical silence". Finally, from a purely descriptive point of view, we can conclude that the result of Fig. 4 demonstrates that the scaling H, and the associated complexity index μ, could be proposed as a reliable indicator of conscious states. The interpretation of these results deserve further investigations that, however, will be the focus of future research work.

REFERENCES

1. A. A. Fingelkurts, and A. A. Fingelkurts, *B*rain and Mind **2**, 261–296 (2001).
2. A. A. Fingelkurts, A. A. Fingelkurts, and C. F. H. Neves, *Phys. Life Rev.* **7**,195–249 (2010).
3. D. Fraiman, P. Balenzuela, J. Foss, and D. R. Chialvo, *Phys. Rev. E* **79**, 061922 (2009).
4. D. R. Chialvo, *Nature Phys.* **6**, 744–750 (2010).
5. J. M. Beggs, and D. Plenz, *J. Neurosci.* **23**, 11167–11177 (2003).
6. D. Plenz, and T. C. Thiagarjan, *Trends Neurosci.* **30**, 101–110 (2007).
7. G. L. Pellegrini, L. de Arcangelis, H. J. Hermann, and C. Perrone-Capano, *Phys. Rev. E* **76**, 016107 (2007).
8. M. Turalska, B. J. West, and P. Grigolini, *Phys. Rev. E*, **83**, 061142 (2011).
9. P. Allegrini, D. Menicucci, R. Bedini, L. Fronzoni, A. Gemignani, P. Grigolini, B.J. West, and P. Paradisi, *Phys. Rev. E* **80**, 061914 (2009).
10. B. J. Baars, *A cognitive theory of consciousness*, Cambridge University Press, Cambridge, UK, 1998.
11. G. M. Edelman, J. A. Gally, and B. J. Baars, *Front. Psychol.* **2**, 4 (2011).

12. G. Tononi, and G.M. Edelman, *Science* **282**, 1846–1851 (1998).
13. D. Sornette, *Critical phenomena in natural sciences*, Springer-Verlag, Berlin, 2006.
14. H. J. Jensen, *Self-organized criticality*, Cambridge University Press, Cambridge, 1998.
15. P. Bak, C. Tang, and K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987); P. Bak, C. Tang and K. Wiesenfeld, *J. Phys. A* **38**, 364 (1988).
16. A. M. Turing, *Mind* **59**, 433–460 (1950).
17. D. R. Chialvo, and P. Bak: *Neuroscience* **90**, 1137–1148 (1999).
18. G. Werner, *Biosystems* **96**, 114–119 (2009).
19. E. Tagliazucchi, and D. R. Chialvo, arXiv: 1103.2070v1 [q-bio.NC] (2011).
20. Y. F. Contoyiannis, and F. K. Diakonos, *Phys. Lett. A* **268**, 286–292 (2000).
21. Y. F. Contoyiannis, F. K. Diakonos, and A. Malakis *Phys. Rev. Lett.* **89**(3), 035701 (2002).
22. P. Manneville: *J. Physique (France)* **41**(11), 1235–1243, (1980).
23. D. R. Cox, *Renewal Theory*, Methuen, London, 1962.
24. P. Paradisi, P. Allegrini, F. Barbi, S. Bianco, and P. Grigolini, *AIP Conf. Proc.* **800**(1), 92–97, 2005.
25. S. Bianco, P. Grigolini, and P. Paradisi, *J. Chem. Phys.* **123**(17), 174704, 2005.
26. P. Paradisi, R. Cesari, D. Contini, A. Donateo, and L. Palatella, *Eur. Phys. J. Special Topics* **174**, 207–218, (2009).
27. P. Paradisi, R. Cesari, A. Donateo, D. Contini, and P. Allegrini, *Nonl. Proc. Geoph.* **19**, 113–126 (2012).
28. P. Paradisi, R. Cesari, A. Donateo, D. Contini, and P. Allegrini, *Rep. Math. Phys.* **70**, 205–220 (2012).
29. O. C. Akin, P. Paradisi, and P. Grigolini, *Physica A* **371**(2), 157–170 (2006).
30. P. Allegrini, F. Barbi, P. Grigolini, and P. Paradisi, *Chaos, Solit. & Fract.* **34**(1), 11–18 (2007).
31. S. Bianco, P. Grigolini, and P. Paradisi, *Chem. Phys. Lett.* **438** (4-6), 336–340 (2007).
32. P. Paradisi, R. Cesari, and P. Grigolini, *Cent. Eur. J. Phys.* **7**(3), 421–431 (2009).
33. O. C. Akin, P. Paradisi, and P. Grigolini, *J. Stat. Mech.: Theory Exp.*, P01013, (2009).
34. P. Allegrini, M. Bologna, P. Grigolini, and B. J. West, *Phys. Rev. Lett.* **99**(1), 010603 (2007).
35. P. Allegrini, P. Paradisi, D. Menicucci, R. Bedini, A. Gemignani, and L. Fronzoni, *J. Phys.: Conf. Series* **306**, 012027 (2011).
36. L. Silvestri, L. Fronzoni, P. Grigolini, and P. Allegrini, *Phys. Rev. Lett.* **102**, 014502 (2009).
37. A. Y. Kaplan, A. A. Fingelkurts, A. A. Fingelkurts, B. S. Borisov, and B. S. Darkhovsky, *Signal Process.* **85**, 2190–2212 (2005).
38. A. A. Fingelkurts, and A. A. Fingelkurts, *Open Neuroimag. J.* **2**, 73–93 (2008).
39. P. Allegrini, P. Paradisi, D. Menicucci, and A. Gemignani, *Front. Physiol.* **1**, 128 (2010).
40. F. Lombardi, H. J. Herrmann, C. Perrone-Capano, D. Plenz, and L. de Arcangelis, *Phys. Rev. Lett.* **108**, 228703 (2012).
41. E. Tagliazucchi, P. Balenzuela, D. Fraiman, and D. R. Chialvo, *Front. Physiol.* **3**, 15 (2012).
42. M. E. Raichle, A. M. Mac Leod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, *Proc. Natl. Acad. Sci. USA* **98**, 676–682 (2001).
43. P. Allegrini, D. Menicucci, R. Bedini, A. Gemignani, and P. Paradisi, *Phys. Rev. E* **82**, 015103 (2010).
44. E. W. Montroll, *Proc. Symp. Appl. Math., Am. Math. Soc.* **16**, 193–220 (1964).
45. G. H. Weiss, and R. J. Rubin, *Adv. Chem. Phys.* **52**, 363–505, 1983.
46. C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, *Phys. Rev. E* **49**, 1685 (1994).
47. G. Tononi, *Biol. Bull.* **215**, 216–242 (2008);
48. D. Menicucci, A. Piarulli, U. Debarnot, P. d'Ascanio, A. Landi, and A. Gemignani, *PLoS One* **4**, e7601 (2009).

Pure state consciousness and its local reduction to neuronal space

A. J. Duggins[1]

Sydney Medical School, University of Sydney; Western Clinical School, Westmead Hospital, Hawkesbury Road, Westmead, NSW 2145, Australia.
Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, United Kingdom.

Abstract. The single neuronal state can be represented as a vector in a complex space, spanned by an orthonormal basis of integer spike counts. In this model a scalar element of experience is associated with the instantaneous firing rate of a single sensory neuron over repeated stimulus presentations. Here the model is extended to composite neural systems that are tensor products of single neuronal vector spaces. Depiction of the mental state as a vector on this tensor product space is intended to capture the unity of consciousness. The density operator is introduced as its local reduction to the single neuron level, from which the firing rate can again be derived as the objective correlate of a subjective element. However, the relational structure of perceptual experience only emerges when the non-local mental state is considered. A metric of phenomenal proximity between neuronal elements of experience is proposed, based on the cross-correlation function of neurophysiology, but constrained by the association of theoretical extremes of correlation / anticorrelation in inseparable 2-neuron states with identical and opponent elements respectively.

Keywords: binding; unity; emergence; neural synchrony; quantum mechanics.
PACS: 87.19.lj, 87.19.ll, 87.19.lm, 87.19.ls, 87.19.lt

INTRODUCTION

In a previous paper [1], an element of experience was introduced at the level of a single neuron: a neuronal state, inaccessible to direct measurement, quantified experience as a uni-dimensional 'perception value' proportional to that neuron's instantaneous firing rate. A mathematical formalism borrowed from quantum mechanics was used to model the neural state as a vector in a complex vector space, spanned by an orthonormal basis of integer 'action potential' states[2]. These basis states were eigenvectors of a Hermitian number operator N, an inevitable n spikes belonging to the n-action potential state $|n\rangle$. The more general normalised neural state $|\psi\rangle$ was resolved onto each basis vector to give a complex amplitude $\langle n|\psi\rangle$. The squared modulus of this amplitude was the prior probability of n spikes, the expected spike count $\langle\psi|N|\psi\rangle$ being the sum of spike counts weighted by such probabilities.

---

[1] Corresponding author, present address: Department of Neurology, Level 1, Westmead Hospital, Hawkesbury Road, Westmead, NSW 2145, Australia; email: andrew.duggins@sydney.edu.au; phone +61 2 98456793, fax: +61 2 96356684

[2] An 'action potential' here is not the depolarisation event itself but the propensity for a neuron eventually to produce a single spike. In the conventional neuron doctrine, action potential and spike are synonymous.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 162-181 (2013); doi: 10.1063/1.4776505

This description was limiting in two major respects. Firstly, it provided only a vague hint as to the dynamics of the neural state over time. An assumption that neural firing is 'memoryless', consistent with the exponential decay of evoked responses to brief sensory stimuli, lead to the conclusion that the expected spike count of an initial single neuronal state vector must be proportional to the initial instantaneous firing rate and perception value (with proportionality constants w and r respectively). Yet the quantum mechanical parallel would imply that such a coherent state could only persist in a closed system. In a model of consciousness at a neural level, it may have been reasonable to neglect quantum physical interactions between subneural structures and the extraneural environment [2]. But clearly there is internal inconsistency in discussing spikes and evoked responses while neglecting synaptic interaction between neurons.

A formal account of the evolution of the neural state through synaptic interaction will be held over to a subsequent paper. Here an attempt will be made to overcome a second major limitation of the model that stems from considering the neuron to be isolated: it fails to capture emergent properties of consciousness in neural systems. In the model as it stands, the experiences attributable to multiple single neurons would be no more than a loose federation of uni-dimensional 'microconsciousnesses'[3]. There could only be a 'summative atomism' of consciousness, as suggested by Leibniz and elaborated by Mill [4], Wundt [5] and Titchener [6] (see [7]). From this perspective, the attempt to ascribe subjective experience to single neurons seems hopelessly naive.

Yet there was good reason to introduce the idea of a 'pure' state vector at the level of the single neuron. The mathematical structures required to reveal the relation between mind and brain have their own fundamental elements that need to be introduced early in discussion, elements that in interacting neural systems do not necessarily have single neuronal correlates. In fact, it is only when the pure state concept is expanded to a system of interacting neurons that a formal mathematical characterisation of emergence and reduction becomes apparent.

So what are the fundamentals of consciousness that emerge? It seems likely, given the common neuro-anatomy and physiology between humans, that the essential qualities of experience are shared. But it is difficult to describe these qualities without straying into specific interpretation reflecting a particular cultural influence. Nevertheless, it seems important to state the author's view, if only to establish terms of reference for this theoretical inquiry.

In the author's previous work [1], two features were nominated as fundamental: subjectivity and smooth evolution. Here the unity of consciousness is added to the list: "a sub-mind is an atrocious monstrosity, just as is a plural-mind - neither having any counterpart in anybody's experience, neither being in any way imaginable" [8]. A unified consciousness is singular and not plural; complete in itself, not a fraction of a greater whole. In contrast, sensation, cognition, emotion and volition are not complete experiences but aspects of the subject's consciousness. Indeed, whatever level of perceptual

---

[3] The 'microconsciousness' theory, proposed by Zeki, held that consciousness was (neuro-anatomically) locally explicable: "Activity in each separate processing node generates a microconsciousness for the attribute for which that node is specialised. Consequently, there are several microconsciousnesses, corresponding to the activity of cells at different nodes within different processing systems" [3].

or conceptual experience one introspects upon, it always seems possible to consider a higher level at which one experiences the relation between simultaneous percepts or concepts. In this sense, unity subsumes the phenomenon of binding, through which elementary visual experiences of motion, colour and form are bound in the percept of a falling red ball. Unity then does not imply uni-dimensionality. Just as the bound percept might be resolved along multiple perceptual dimensions, perhaps a unified consciousness might usefully be represented as a 'mental state vector' in a 'subjective space' spanned by all of the dimensions of perceptual experience:

> "It seems inevitable that psychophysical explanation will apply first at the level of some kind of elements of experience; but if these elements come together in a single consciousness, they must also be components of a single point of view" [9].

To make progress, it will be necessary to develop formal rules for combining neural spaces so that a state of a composite system quantifies experience along multiple perceptual dimensions, while also defining the qualitative relationship between elements of experience. For the most part, this will be a process of assimilation of the quantum mechanical principles for combining state spaces. First the concept of a tensor product of single neuron states and state spaces will be introduced, considering a system of just 2 neurons. An expression that satisfactorily separates the experience of a product state into two single neuronal elements will be extended to inseparable pure states whose existence is implied by the superposition principle on the composite 2-neuron space. The reduced single neuron state in this more general case will be shown to be mixed rather than pure, mathematical description of this mixed state demanding introduction of a density matrix rather than vector formalism. Finally it will be shown that what is lost in the process of reduction of consciousness to mixed neural elements, and regained through consideration of pure states of neural systems, is a characteristation of the relational aspects of unified experience.

TENSOR PRODUCTS OF SINGLE NEURONAL STATES

In a 2-neuron system, call the state space of the first neuron A, spanned by the states of 0, 1, 2... action potentials $|0^A\rangle$, $|1^A\rangle$, $|2^A\rangle$..., the dimension of the space limited by the maximal potential to spike. Similarly the space of neuron B is spanned by $|0^B\rangle$, $|1^B\rangle$, $|2^B\rangle$.... Now it may be that a state of the 2-neuron system can be completely described by the states of the individual neurons considered separately[4]. If these states of A and B are the vectors $|a\rangle$ and $|b\rangle$ respectively, then this 'separable' state of the composite 2-neuron system will be denoted $|a\rangle|b\rangle$ or for simplicity just $|ab\rangle$. For any particular state $|a_1\rangle$ of neuron A, there is then a range of composite states $|a_1 b\rangle$, corresponding to the vector space B of possible states $|b\rangle$. Consider two such states of neuron B, perhaps $|b_1\rangle$ and $|b_2\rangle$. By the superposition principle the state $c_1|b_1\rangle + c_2|b_2\rangle$ must also

[4] Objectively, this is a requirement that the joint probability density of simultaneous spikes in the two neurons is the product of single neuron firing rates (see equation (13)).

describe a valid state of neuron B, where c_1 and c_2 are arbitrary complex numbers. Hence $|a_1\rangle(c_1|b_1\rangle + c_2|b_2\rangle)$ is a valid state of the composite system. We might expect this to be equivalent to the same linear combination of composite states ([10], 2.2.8), such that[5]

$$|a_1\rangle(c_1|b_1\rangle + c_2|b_2\rangle) = c_1|a_1b_1\rangle + c_2|a_1b_2\rangle.$$

This is like a distributive axiom of multiplication, suggesting that the composite state $|ab\rangle$ is a type of product of single neuron states $|a\rangle$ and $|b\rangle$.

Now consider the product states $|0^A0^B\rangle$ and $|0^A1^B\rangle$. In the former state it is certain that neuron B will remain silent, whereas in the latter it is certain that neuron B will spike once. Just as are the integer action potential states of the single neuron, so must these product states be orthogonal. This is consistent with an expression for the scalar product

$$\langle a_1b_1|a_2b_2\rangle = \langle a_1|a_2\rangle\langle b_1|b_2\rangle. \tag{1}$$

Since the single neuron integer action potential states $|0^A\rangle$, $|1^A\rangle$,... and $|0^B\rangle$, $|1^B\rangle$,... are all normalised, $|0^A0^B\rangle$ and $|0^A1^B\rangle$ are also orthonormal[6]. It is possible then to describe a *tensor product* space $A \otimes B$ spanned by the orthonormal basis

$$\{|0^A0^B\rangle, |0^A1^B\rangle, |0^A2^B\rangle..., |1^A0^B\rangle|1^A1^B\rangle|1^A2^B\rangle..., |2^A0^B\rangle, |2^A1^B\rangle, |2^A2^B\rangle...\}.$$

The tensor product of vectors $|ab\rangle$ could be represented with respect to this basis as

$$|ab\rangle = \begin{pmatrix} \langle 0^A0^B|ab\rangle \\ \langle 0^A1^B|ab\rangle \\ \vdots \\ \langle 1^A0^B|ab\rangle \\ \langle 1^A1^B|ab\rangle \\ \vdots \\ \langle 2^A0^B|ab\rangle \\ \langle 2^A1^B|ab\rangle \\ \vdots \end{pmatrix} \tag{2}$$

or with respect to the $\{|0^A\rangle, |1^A\rangle, |2^A\rangle...\}$ basis of A as

$$|ab\rangle = \begin{pmatrix} \langle 0^A|a\rangle|b\rangle \\ \langle 1^A|a\rangle|b\rangle \\ \langle 2^A|a\rangle|b\rangle \\ \vdots \end{pmatrix} \tag{3}$$

---

[5] We will further assume that the composite product state has the properties (see [11])

$$\begin{aligned} c(|ab\rangle) &= (c|a\rangle)|b\rangle = |a\rangle(c|b\rangle), \\ |a_1\rangle(|b_1\rangle + |b_2\rangle) &= |a_1b_1\rangle + |a_1b_2\rangle, \\ (|a_1\rangle + |a_2\rangle)|b_1\rangle &= |a_1b_1\rangle + |a_2b_1\rangle. \end{aligned}$$

[6] In an orthonormal basis set $\{|n\rangle\}$, scalar products between members of the set satisfy $\langle n|n\rangle = 1$ and $\langle m|n\rangle = 0$.

where each element is a vector in the neural space B.

PURE STATES AS UNIDIMENSIONAL PROJECTORS

It would be intuitively appealing for any weighted superposition of these basis states also to be a valid state of the composite system. If instead the superposition principle were to hold only within single neuron spaces, this would be to make the neural level privileged in relating subjective to objective properties. This would seem contrary to the philosophy, espoused here and in the previous paper [1], that a neural theory of consciousness should be merely a convenient starting point for future theoretical development.

However, if the superposition principle were to be adopted, it would be necessary to concede that valid states of the composite system exist that are inseparable, in the sense that they are not simple products of single neuron states. For example, the distributive axiom would require that

$$c_0^2|0^A 0^B\rangle + c_1^2|1^A 1^B\rangle \neq (c_0|0^A\rangle + c_1|1^A\rangle) \otimes (c_0|0^B\rangle + c_1|1^B\rangle).$$

In one sense this is a positive development for the theory of consciousness, in that a priori it would be unlikely that the bound percept could be reduced to elements of experience in simple combination. This is not to abandon the practice of attributing experience to single neurons. In fact, the process of specifying the perception value of neuron A when the neural system is in a product state, when applied to an inseparable state of the composite system, reveals a more general expression for the reduced state of the single neuron. This will, however, require an alternative description of neural states, which is introduced next.

In the original formulation [1], the pure state of the neuron was portrayed as the state vector $|\psi\rangle$. The entity $|1\rangle\langle 1|$ 'projects' $|\psi\rangle$ onto the 1-action potential state, in that

$$|1\rangle\langle 1|\psi\rangle = \langle 1|\psi\rangle|1\rangle.$$

The sum of projectors onto integer action potential states (or onto any other complete set of orthonormal basis states for the neural space) must equal the identity operator I, in that for example

$$(|0\rangle\langle 0| + |1\rangle\langle 1| + ...)|\psi\rangle = \langle 0|\psi\rangle|0\rangle + \langle 1|\psi\rangle|1\rangle + ...$$

which is just a resolution of $|\psi\rangle$ onto a complete set of orthonormal basis vectors.

The expected spike count can then be expanded

$$\langle\psi|N|\psi\rangle \;=\; \sum_n \langle\psi|n\rangle\langle n|N|\psi\rangle = \sum_n \langle n|N|\psi\rangle\langle\psi|n\rangle \tag{4}$$

which is the sum of diagonal elements of a matrix representation of $N|\psi\rangle\langle\psi|$ with respect to the $\{|n\rangle\}$ basis[7], known as the trace of the matrix (see [10], 2.1.8). The

[7] Just as the nth element in a representation of $|\psi\rangle$ with respect to the $\{|n\rangle\}$ basis is $\langle n|\psi\rangle$ (equation (2)), the element in the mth row and nth column in a representation of an operator Q is $\langle m|Q|n\rangle$.

choice of basis was of course arbitrary, so the expected spike count (equation (4)) can be considered the trace of any matrix representation of the operator $N|\psi\rangle\langle\psi|$,

$$\langle\psi|N|\psi\rangle = \mathrm{tr}(N|\psi\rangle\langle\psi|).$$

Description of the pure neural state ψ as the projector onto $|\psi\rangle$ is then completely equivalent to its description as the vector. The requirement that $|\psi\rangle$ is normalised:

$$\sum_n \langle n|\psi\rangle\langle\psi|n\rangle = 1$$

is equivalent to

$$\mathrm{tr}(|\psi\rangle\langle\psi|) = 1. \tag{5}$$

A pure state of the composite 2-neuron system ψ^{AB} could equally well be considered a vector $|\psi^{AB}\rangle$ or a unidimensional projector on $A \otimes B$:

$$\rho^{AB} = |\psi^{AB}\rangle\langle\psi^{AB}| \tag{6}$$

This new notation for states of neural systems is intentionally chosen to leave open the possibility that the reduced state of neuron A might also be expressed as a generalization of a unidimensional projector on neural space A, ρ^A.

A DENSITY OPERATOR FORMALISM OF EMERGENCE

Returning to the perception value concept, for this to have any meaning, it must be possible to take a pure state of the 2-neuron system, such as ρ^{AB} (equation (6)) and from it derive a level of perceptual experience attributable to neuron A that does not depend on the experience attributable to neuron B. In previous work [1], a perceptual variable operating on a single neuronal space, the 'nervous energy' H, related the perception value to a pure state of the neuron. Considering neuron A specifically:

$$H^A = rN^A$$

perception value $A = \langle a|H^A|a\rangle = \mathrm{tr}(H^A|a\rangle\langle a|)$. r is a constant so that the perception value is proportional to the expected spike count.

Taking a product state $|ab\rangle$ of the 2-neuron system, and extending the rules governing the order of terms in the scalar product (equation (1)) to matrix multiplication generally

$$\langle ab|H^A \otimes I^B|ab\rangle = \langle a|H^A|a\rangle\langle b|I^B|b\rangle = \langle a|H^A|a\rangle,$$

implying that $H^A \otimes I^B$ is the extension of H^A to the composite neural space (see *Appendix*). It will be assumed that the expectation value of the operator $H^A \otimes I^B$ also gives the perception value A in the more general (possibly inseparable) state of the composite system, ψ^{AB}:

$$\langle H^A\rangle = \langle\psi^{AB}|H^A \otimes I^B|\psi^{AB}\rangle = \mathrm{tr}(H^A \otimes I^B\rho^{AB}). \tag{7}$$

A good candidate for the reduced neural state A would then be some state ρ^A, which could be expressed in terms of basis states of A alone, such that

$$\langle H^A \rangle = \mathrm{tr}(H^A \rho^A). \tag{8}$$

Indicating integer action potential basis states of A as $|p^A\rangle$ and integer action potential basis states of B as $|m^B\rangle$, the pure state ψ^{AB} could be described by the normalised vector

$$|\psi^{AB}\rangle = \sum_{pm} c_{pm} |p^A\rangle |m^B\rangle \tag{9}$$

or equivalently as the projector

$$\rho^{AB} = \sum_{pqmn} c_{pm} c^*_{qn} |p^A\rangle\langle q^A| m^B\rangle\langle n^B|. \tag{10}$$

When the latter is substituted into equation (7), the expression for the element of experience attributable to neuron A can be simplified, utilising properties of the trace and orthonormalisty of basis states (see *Appendix*), to give

$$\langle H^A \rangle = \mathrm{tr}\left[H^A \sum_{pqm} c_{pm} c^*_{qm} |p^A\rangle\langle q^A| \right]$$

where

$$\rho^A = \sum_{pqm} c_{pm} c^*_{qm} |p^A\rangle\langle q^A| \tag{11}$$

satisfies the criterion of equation (8) for the reduced neural state A.

Comparison with equation (10) reveals that

$$\rho^A = \sum_m \langle m^B | \rho^{AB} | m^B \rangle.$$

Similarly we could write

$$\rho^B = \sum_p \langle p^A | \rho^{AB} | p^A \rangle \tag{12}$$

which is known as the partial trace over A (see *Appendix*):

$$\rho^B = \mathrm{tr}^A(\rho_{AB}).$$

If, consistent with our interpretation of $|\psi\rangle$ in the single neuron, we interpret $|\psi^{AB}\rangle$ or ρ^{AB} as the subjective state of the composite system, then taking each partial trace of ρ^{AB} to yield ρ^A and ρ^B is the mathematical counterpart of reducing consciousness to its neural correlates (in this simple example of a 2-neuron system). Taking the partial trace yields a local 'reduction' of consciousness in the sense that the experience attributable to the system is inadequately described by ρ^A and ρ^B alone. Mathematically, the property

of ρ^{AB} that it does not in general equal the tensor product of these neural correlates, corresponds to the 'emergence' of consciousness in the combined system.

Many of the results of the previous paper [1] relied upon a representation of the neural state as a vector, or a projector onto a single dimension of the neural space. It is now evident that this was an oversimplification, applying only when the combined system exists in a product state. While we still assume that the state of the composite system is a projector onto a single dimension of the composite space, our description of the state of the neural subsystem must now be more general.

As with the perception value, it is possible to reformulate, in terms of the reduced neural state ρ^A on neuron A, expressions for the prior probability of 1 eventual spike

$$P(1) = \mathrm{tr}(|1^A\rangle\langle 1^A|\rho^A)$$

for the expected spike count (p)

$$\langle p\rangle = \mathrm{tr}(N^A\rho^A), \quad N^A = p|p^A\rangle\langle p^A|$$

and for the instantaneous firing rate

$$f_A(t) = w\,\mathrm{tr}(N^A\rho^A).$$

Like the number operator (see [1]), ρ^A (equation (11)) is Hermitian and positive (see *Appendix*). Unlike N^A, ρ^A is normalised (as in equation (5)). These three conditions, hermiticity, positivity and normalisation, characterise a 'density operator'. They imply that ρ^A has a spectral decomposition

$$\rho^A = \sum_k p_k|\phi_k\rangle\langle\phi_k|$$

where $\{|\phi_k\rangle\}$ is some orthonormal basis for the neural space A, and the eigenvalues p_k are positive real numbers which sum to one (a representation of ρ^A with respect to the $\{|\phi_k\rangle\}$ basis would be a diagonal matrix, with elements p_k along the diagonal). Conceptually, the neural state A could be considered a weighted mixture of orthogonal projectors $|\phi_k\rangle\langle\phi_k|$ corresponding to the neural states ϕ_k, each occurring with probability p_k . The unidimensional projector ρ^{AB} also satisfies the criteria for a density operator on the combined space AB, but in that there is a single term in the spectral decomposition (equation (6)), it is a 'pure' as opposed to 'mixed' state.

It is postulated that there is a direct correspondence between the unity of experience, a property of the system as a whole, and the purity of the density operator that describes the state [12]. Extending this principle to the ultimate multi-partite system, the human brain, one would want the unified mental state to be described by a mental state vector or a pure uni-dimensional projector on the tensor product of neural spaces: a combined system of vast dimension. The reduced state of any subsystem of one or many neurons would be a mixed-state density operator. Were this not so, the completeness privilege of consciousness described above would be lost. Imagine a situation in which it were possible to define several neural subsystems, whose subjective correlates could be accurately and completely described by pure states. In that case, presumably the overall mental state would be described as a separable product of microconsciousnesses. It would be intrinsically composite, and there would be no unity.

THE CORRESPONDENCE HYPOTHESIS

So far little understanding has been gained of how the structure of experience emerges moving from single neuron reduced states ρ^A and ρ^B to the pure state description ρ_{AB} on the tensor product space. In previous work [1] it was suggested that the contribution of a single sensory neuron to the neural code for the sensory environment parallels the relationship between a neuronal element of experience and the overall sensory percept. This 'correspondence hypothesis' is now extended to address the structure of perceptual experience:

> *However the similarity between environmental features specified by a pair of neurons is encoded in their joint activity, this objective association parallels the qualitative proximity of respective neuronal elements of experience.*

Note that a neuroscientist is able to establish a neuron's functional specificity experimentally by correlating its activity with aspects of the sensory environment. In contrast, the brain has no other access to the environment than the sensory neurons themselves. A definition of neural coding that assists a formulation of consciousness cannot depend on such external validation.

For example, how is it possible for the brain to interpret, from the activity of neurons in early visual cortices, the geometry of physical space? An initial possibility is that a sensory neuron somehow has access to its own location in a neuroanatomical map of the environment. Optical principles dictate that there is an inverted but spatially veridical mapping from the location of a source within the visual field to the retinal location of photoreceptor activation. However, this cannot be the source of phenomenal spatial relationships unless the retina itself supports consciousness. Although there is a retinotopic map in primary visual cortex (V1), a magnification factor is introduced to allocate a proportionately greater cortical area to the central region of the visual field. Higher visual and association cortices receive inputs from V1, but presumably have no access to the topography of the original retinal image, nor any other map of spatial relationships[8] that would allow the reverse transformation. Nor does it seem that any such transformation is 'hardwired': rod achromats, who lack photoreceptors in the central foveal region of the retina, develop a primary visual cortex without the usual magnification factor [14], yet negotiate their environment as if they encode spatial relationships in the same way as normals. These observations suggest that that there exists no absolute spatial reference frame to which sensory input is co-registered.

---

[8] Applying an 'enactive theory' of perception to visuo-spatial relationships, one would contend that we learn veridical spatial relationships by acting upon our environment. This would seem to require a primary veridical sense of proprioception from which we might establish order in our visual experience. The magnification factor for oral and manual inputs in the primary somatosensory homunculus (reminiscent of V1) makes this seem unlikely [13].

If indeed only relative spatial locations are encoded[9], there is support for the correspondence hypothesis in the relativity of spatial experience. A transformation that inverts the visual input relative to the orientation of the retina, but preserves relative retinal location, does not alter visuospatial experience[10]. The term 'subjective symmetry' is coined for any such transformation of the sensory input that leaves experience unchanged.

Further evidence for the correspondence hypothesis is found in the representation and experience of colour. The organisation of wavelength information into opponent channels in the retina and thalamus imposes a 2-dimensional encoding space on an input that has a single degree of freedom (the visual spectrum). The activity of a V1 opponent cell in one such channel reflects the difference in activity between long (*L*, peak 558 *nm*) and medium (*M*, peak 531 *nm*) wavelength cones in the retina, whose spectral response functions overlap considerably[11]. V1 neurons in a second opponent channel compare the sum of activities in *M* and *L* cones with the activity of short wavelength cones, whose spectral response function (peak 420 *nm*) is relatively distinct.

Using the simple criterion that minimally perceptually different colours should lie adjacent and separated by a uniform distance, it is possible also to define a 2-dimensional vector space of chromatic experience in which opposing hues are reflected through the origin (CIE 1976 L*a*b*, see [17]). If each opponent channel could be associated with one of the principal red-green (a*) and yellow-blue (b*) axes of this chromatic plane, then the antagonistic inputs from photoreceptors to wavelength opponent cells would provide a satisfying explanation of phenomenal colour opposition. Corresponding to its instantaneous firing rate, the perception value of one such wavelength-opponent cell would be the chroma of colour experience along a single qualitative dimension from saturated green, through grey, to saturated red. Corresponding to the 'orthogonality' of red/green and blue/yellow qualities of colour experience, the neuroscientist should discover independence of firing of neurons in each channel.

In fact this model is too simplistic. Consistent with the relativity of spatial experience, the attribution of red colour actually depends not on the absolute intensity of low frequency light reflected from a surface, but on the relative intensity compared to all other surfaces within the visual field (known as a 'lightness record', see [16]). The subjective symmetry in the case of colour vision is the alteration of frequency components in illuminant light that preserves phenomenal colour constancy. If there is dependence in firing that corresponds to qualitative proximity between V1 neuronal elements of experience, one surmises that it will be found not only within but also between retinotopic modules.

---

[9] Perhaps relative spatial relationships are learned by applying a primary concept of size constancy to define isometric regions as objects move through egocentric space. Alternatively, since the spatial frequency of natural scenes does not vary with retinal eccentricity, perhaps the statistics of neural firing across a retinotopic map, over a critical period of immersion in the natural world, are themselves sufficient to establish the proportions of space.

[10] Normal visual experience is restored after several days to a subject who wears inverting lenses [15]. Presumably the delay reflects gradual adaptation to inverted 'sensorimotor contingencies'.

[11] Actually, V1 neurons demonstrate 'double opponency', performing a comparison of wavelength comparisons between the centre and surround of the receptive field [16]. Whilst relevant to the ensuing discussion of the relativity of colour experience, this complication can be neglected at this stage.

DEPENDENT FIRING AND PHENOMENAL PROXIMITY

What objective measure of paired association would complement the relationship between perception values and instantaneous firing rates that has been proposed, but would remain consistent with neurophysiologic evidence? Certainly the joint probability density of spikes between neurons A and B, $f_{A,B}(t)$, does not satisfy the former criterion. This measure is affected not only by the dependence in spike timing between neurons, but also by single neuron firing rates $f_A(t)$, $f_B(t)$. If single neuron firing rates are to be allowed a role in the quantity but not the quality of experience, $f_{A,B}(t)$ needs to be 'corrected' to produce a purer measure of dependence in spike timing:

$$Q_{AB} = \frac{f_{A,B}(t)}{f_A(t) f_B(t)}. \tag{13}$$

Instantaneous firing rates have been hypothesised to be directly proportional to perception values. Q_{AB} can then be reformulated in terms of expectation values of operators in a tensor product neural space:

$$Q_{AB} = \frac{\langle H^A \otimes H^B \rangle}{\langle H^A \rangle \langle H^B \rangle}.$$

Q_{AB} is postulated to be the phenomenal proximity between elements of experience A and B quantified by perception values $\langle H^A \rangle$ and $\langle H^B \rangle$.

Happily the same idea, of paired association between two neurons as a deviation from independent firing, is behind the cross-correlation function [18] of cellular neurophysiologic studies. This is a plot of the conditional firing rate of neuron A in the period surrounding spike occurrence in B (defined as $t = 0$), calculated by summing over consecutive spikes at site B. 'Synchrony' is said to occur when the conditional firing rate at $t = 0$ significantly exceeds the average firing over an epoch of measurement. Figure 1 presents a fictitious example of responses evoked in two neurons by recurrent stereotyped stimulus presentation, to illustrate the connection between dependence in firing probability and another measure of paired association: the Pearson correlation, commonly displayed in the normalised joint peri-stimulus time histogram (see [19]).

Stimulus parameters that modulate association in neural firing have been extensively studied, but rarely the association between single units in remote regions that is envisaged here. In fact, most invasive studies in animals and humans have employed multi-unit recordings that count spikes from multiple neurons within the vicinity of an electrode. The success of such multi-unit recording depends on being able to isolate neuronal subsets that are homogenous in their firing patterns. Cerebral cortex has a modular organisation in which neurons that share feature specificities tend to reside within a column (extending inwards from the cortical surface), such that they are jointly accessible to a suitably placed electrode. Synchrony seems to depend on this joint feature specificity, evident as significant 'auto-correlation' of pooled neuronal firing at $t = 0$ ([20]; explained in [21], Figure 2.5).

Cross-correlation also depends on the similarity of neuronal feature specificity between recording sites. For example, when stimulating two remote orientation-specific regions of cat visual cortex with moving light bars of optimal orientation for neurons in

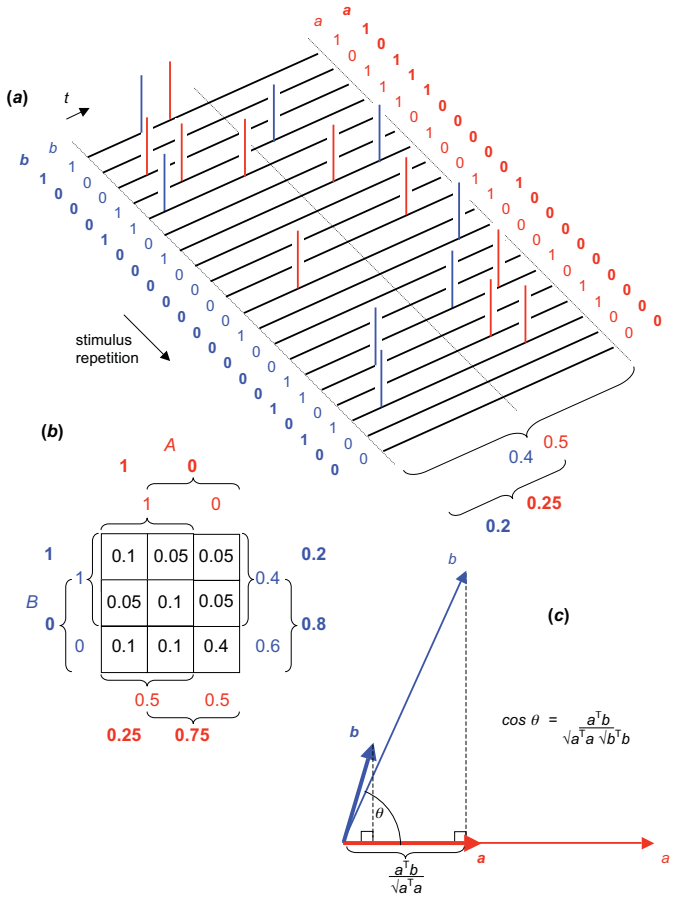


FIGURE 1. Dependent firing of two neurons A and B evoked by recurrent stereotyped stimulus presentation. It is assumed that this small sample is representative of an infinite ensemble of stimulus repetitions in three ways. Firstly, the time window is sufficiently short that neither neuron ever fires more than once. Secondly, the frequency of spikes over these 20 trials is the actual spike frequency. Thirdly, just as the frequency of spikes is the same in both the first and second half of the window, it is assumed that the instantaneous firing rate remains constant throughout the window. **(a)** The response of neuron A (in red) and B (in blue) within a time window of fixed latency relative to 20 stimulus presentations. Unit elements of response vectors identify trials in which a spike occurs in the first half (bold font) or at any time (normal font) during the window. Frequencies of these outcomes are indicated at bottom right. **(b)** Table of joint probabilities of spike outcomes. Cells in the first column (row) correspond to outcomes in which neuron A (B) fires in the first half, in the second column (row) the second half of the window, and in the third column (row) outcomes in which neuron A (B) does not fire. Marginal probabilities are written below and to the right of the table. The joint probability of spikes in both neurons at any time during the window is 1.5x the product of marginal probabilities from each, whereas the probability of a spike in neuron B in the first half doubles, given a spike in A during this shorter period. Dependence in spike probability will remain stable when the temporal resolution of measurement exceeds the fundamental precision of correlations, so that the joint probability density of spikes and single neuronal instantaneous firing rates can be substituted for spike probabilities, as in the text. **(c)** Depiction of a 2D subspace of the vector space that includes mean-corrected response vectors a and b from (a). The Pearson correlation coefficient between spike counts is the cosine of the angle between vectors, which reduces as the time window narrows (vectors from the first half-window are in bold).

the respective columns, synchrony between sites only occurs when the two regions share the same preferred orientation. Synchrony of firing between these remote sites is amplified still further when a single light bar extending across both receptive fields replaces separate light bars of the same orientation [22].

Following from the correspondence hypothesis, it is assumed that neurons united in maximally correlated and anti-correlated states contribute identical and opponent elements of experience respectively. When firing is 'uncorrelated'

$$Q_{AB} = 1$$

but this is also a firing pattern that defines a 'separable' product state of a two neuron system (see footnote 4, also [23]).

The extreme example of an inseparable state of a 2-neuron system in which firing is maximally correlated is

$$|\psi^{AB}\rangle = \sqrt{1/2}\,|0^A 0^B\rangle + \sqrt{1/2}\,|1^A 1^B\rangle.$$

The equivalent inseparable state in which firing is maximally anti-correlated is

$$|\psi^{AB}\rangle = \sqrt{1/2}\,|0^A 1^B\rangle + \sqrt{1/2}\,|1^A 0^B\rangle.$$

What will be the conditional firing rate of B given a spike in A when a 2-neuron composite system is in such an inseparable state? In the spirit of the previous discussion of the single neuron [1], such posterior probabilities should be consistent with an objective interpretation of a spike as action potential annihilation. Applying such an interpretation to the maximally correlated state, if a spike occurred in A then this would imply that an action potential had been present. *A posteriori* a spike in B would become certain, **as if** 1 action potential also existed in neuron B. When Q_{AB} is calculated (see *Appendix*), the conditional instantaneous firing rate of B doubles, given a simultaneous spike in A. In the equivalent anticorrelated state, $\langle H^A \otimes H^B \rangle$ and Q_{AB} vanish, suggesting that the qualitative distance (conceived as an angle in radians) between elements of experience A and B, quantified by perception values $\langle H^A \rangle$ and $\langle H^B \rangle$, follows the anti-linear relation

$$\theta(A,B) = \pi - (\pi/2)Q_{AB}. \tag{14}$$

DISCUSSION

The unity of consciousness

The mental state has been modeled as a vector on a tensor product space in order to capture the unity of consciousness. Mental states are in general 'inseparable'. They reduce to but are not entirely constituted in semi-classical phenomenal correlates of firing rate at the single neuron level. In the formalism that has been chosen, the locally irreducible aspect of consciousness is the relation between elements of experience, manifest objectively as a correlation in firing between anatomically remote neurons.

Not all theorists of consciousness acknowledge that the unity constraint exists. According to MacLennan [24], "subjective qualities come from the structure of the interdependencies of protophenomena": elements of experience not unlike perception values. He identifies such interdependencies with firing correlations established through synaptic transmission: "Under the assumption that ...protophenomenal intensity corresponds to the membrane potential of the cell body, protophenomenal dependencies are mediated by the axons, synapses, and dendrites". But for MacLennan "there is no need to postulate a reified 'phenomenon' to integrate the coherently changing intensities of masses of protophenomena". According to this view, consciousness is merely a list of qualitative relationships between neural elements.

On the other hand, the unity constraint compels some theorists to conclude that areas that are only indirectly connected, via a pathway of several synaptic steps, cannot simultaneously be contributing to consciousness. For example, Crick and Koch [25] exclude primary visual cortex partly on the basis of its lack of direct connectivity to frontal cortex, an anatomic region to which they are more confident in attributing experience. But this stringent requirement for direct connectivity, between neurons participating in consciousness, would seem severely to limit the possibility of a rich visual percept. Others see a semblance of unity among neurons whose activities are coordinated by polysynaptic thalamocortical feedback that seems to justify a slackening of the direct connectivity criterion [26]. In such models, the richness of experience is further accommodated by the rapid confederation and dissolution of such groupings across cortices of varying functional specificity.

Other authors who allow recurrent projections a pivotal role in consciousness are motivated to do so by the idea that self-reference is a model for subjectivity [27]. Recurrent projections from higher to lower-order visual cortices, which seem vital to allow report of visual experience [28], and which are equated with consciousness by Lamme [29], are abundant but remain modular. Recurrent projections from frontoparietal to visual areas might seem to have more prospect of binding visual experience, but do not issue from a single source. The unity of experience still eludes explanation.

An interesting way out of the conundrum, proposed by Edwards [30], is to claim that a complete copy of consciousness belongs to each individual neuron. Here, the relevance of neural connectivity to the unity of consciousness is in providing each neuron "simultaneous (cotemporal) access to many elements (of information) in defined interrelationships". We experience unity, according to this model, not because we have a single mental state, but because our billions of mental states are all the same! The obvious problem is that the model requires that each neuron have access to the same inputs, a contention that is patently false: the firing of a neuron in V5 in response to motion in a specific direction, and a neuron in V4 to wavelength information relative to a lightness record, surely reflects access of these cells to differing information in their dendritic inputs.

In conclusion, it seems inevitable that any classical model grounded in realistic neural connectivity will be unable to capture both the unity and complexity of consciousness. One of the reasons to pursue a quantum mechanical formalism was the potential for a better characterisation of these aspects of consciousness than a classical model would allow. Whereas classical neuroscience would accept only objective local influences upon neural transmembrane potential, in the quantum formalism the occurrence of a spike in

one neuron instantaneously influences the posterior probability of a spike in a remote neuron with which it is united in an inseparable pure state. Regrettably, the model has not been developed sufficiently to predict the remote modulation of spike probabilities in a realistic brain. However, the 2-neuron model does suggest that the need to invoke non-local influences upon spike probability might be used as an indication that reductive explanation of brain has failed.

The function of a unified consciousness

Evolution provides a powerful argument in favour of taking such a leap of credibility: consciousness must have a function, yet classical neuroscience appears causally closed. An influential recent account sees sensory experience as a process of distillation of order from the chaos of unregulated interaction with the environment: "Out of this melee of sensory events only a few privileged events make it into phenomenal feelings, while the rest are discarded into an experiential limbo. Natural selection pursued a strategy that amounts to summarizing most of the pertinent facts about the outside world compactly and sending this description to the planning stages to consider the organism's optimal course of action" ([31], 14.1). It remains to be explored how these putative processes might be facilitated in a brain endowed with a pure, inseparable mental state.

Considering initially this synthesis of "phenomenal feelings" from "sensory events", the idea that a metric of paired association between neurons might express the qualitative distance between elements of experience within a visual sensory modality could easily be extended to binding of motion, form or colour experiences. "To say an element is 'bound' to another is simply another way of saying that they are represented in awareness dependently and are nested together... Furthermore, the degree to which lower order features are bound into a higher order feature is directly related to the extent to which lower order features lose their independence from each other" [32].

In fact, it must be conceded that the relation between firing dependencies and the structure of sensory experience is likely to be more subtle than has been portrayed (equation (14)). It seems likely, given the vast numbers of neurons involved in visual representation, that the firing patterns of neuronal arrays are higher dimensional than the perceptual experiences that they afford. Yet somehow we **do** perceive the two intrinsic degrees of freedom of random dot motion across a screen, or the two degrees of freedom imposed by the organisation of wavelength information into opponent channels in the retina and thalamus.

In principle, "overlapping local neighbourhoods-collectively analysed-can provide information about global geometry" [33]. If a linear relationship between some measure of association in firing and phenomenal proximity were to hold only for qualitatively similar elements of experience (for example V5 neurons of similar direction specificity), then it might still be possible to define the local geometry of perceptual experience. This would imply that within the higher dimensional space defined by V5 neuronal firing patterns, there is embedded a (locally linear) 2-dimensional curved 'surface' or manifold that captures the 2 modes of variation in the experience of motion on a screen. More generally, perhaps genetically encoded patterns of neural connectivity, refined and

pruned through exposure to the sensory environment, constrain the population activity in early visual cortices to lie on low-dimensional manifolds. Perhaps it is on such a manifold that $\theta(A,B)$ (equation (14)) expresses the qualitative distance in a particular visual modality between two neuronal elements of experience. Still, it would seem necessary to extract the information from overlapping or intersecting manifolds in early visual cortices, in order to capture the perceptually meaningful structure of the sensory environment. Perhaps this is the function of the ascending dorsal and ventral visual streams.

The suggestion that some form of dimensionality reduction is necessary for experience to arise prompts speculation on the neural criteria for consciousness. It is tentatively proposed that dependence in firing between neurons is a necessary (but insufficient) criterion for a neural representation to be considered 'conscious': satisfied when the combined system exists in an inseparable state. Similarly MacLennan [24] proposes, as a correlate of "unconscious neural activity", "incoherent protophenomenal activity, which is unconscious because it does not cohere into conscious phenomena".

Whatever is the dimension of the neural representation of the sensory enviroment, this summary of "pertinent facts about the outside world" ([31], 14.1) would be exponentially more compact in the current formalism than a purely classical neuronal register. A linear increase in the number of neurons would be associated with an exponential increase in the dimension of the tensor product of neural spaces. Just as a single neuron could be in a superposition of 0- and 1- action potential states, a pure state of n neurons could be in a superposition of at least 2^n states of definite spike potential.

Turning now to the translation of perception into action, one putative benefit of a pure mental state might be an improved efficiency of spike generation afforded by synchronous dendritic inputs. "100 fast excitatory inputs, distributed over the dendritic tree of a large pyramidal neuron, are sufficient to generate ...(a spike)... if they are activated within a millisecond of each other. If the pre-synaptic spikes arrive smeared out over a 25 msec window, however, twice as many synapses are needed to fire the cell" ([31], 2.3). The objective manifestation of an inseparable pure mental state would be correlations between neurons in anatomically remote higher-order visual cortices that have no direct synaptic connection with each other, perhaps correlations that could not have been established classically by common synaptic input from some lower level of the visual hierarchy. Not only would the subject experience the bound percept of a falling red ball, but synchronous convergent input to neurons in motor cortex might allow him to catch it!

Conclusion

The unified mental state has been proposed to exist at the apex of a hierarchy of bound percepts and concepts. This structure of experience would be expected to be manifest in specific correlation patterns between neuronal sub-systems, and even remote brain regions. The question is, if not constrained by this notion of unity of consciousness, would it have been possible to come up with a different set of neuronal elements of experience that would be manifest as an identical pattern of objective neural activity?

Or could there be something inexplicable in these terms, something irreducible about the brain that parallels the emergence of a unified consciousness? It is hoped that future work will explore just how demanding are the conditions imposed on neural mutual information by the existence of a pure mental state in the tensor product space, presenting an opportunity for experimental test of the scheme.

APPENDIX

Just as in the representation of a vector product (equation (3)), the representation of $H^A \otimes I^B$ with respect to the $\{|0^A\rangle, |1^A\rangle, |2^A\rangle ...\}$ basis of A would be

$$H^A \otimes I^B = r \begin{pmatrix} 0\,I^B & 0\,I^B & 0\,I^B & .. \\ 0\,I^B & 1\,I^B & 0\,I^B & .. \\ 0\,I^B & 0\,I^B & 2\,I^B & .. \\ : & : & : & :: \end{pmatrix}$$

where each element is an operator on neural space B.

When the expression of the pure state of the 2-neuron composite system as a projector (equation (10)) is inserted into in equation (7), the perception value A then becomes

$$\langle H^A \rangle = \mathrm{tr} \left[\sum_{pqmn} c_{pm} c^*_{qn}\, H^A |p^A\rangle\langle q^A| \otimes |m^B\rangle\langle n^B| \right].$$

Each term in the sum inside the brackets is a weighted tensor product of an operator $H^A|p^A\rangle\langle q^A|$ on A with an operator $|m^B\rangle\langle n^B|$ on B. In a matrix representation of the sum of such terms, each element would be the sum of corresponding elements of the tensor product matrices. Clearly then the trace of the sum of terms must be the sum of traces of each tensor product. Consideration of matrix representations also reveals that the trace of a tensor product is the product of traces of each operator[12]. Moreover, in a representation of $|m^B\rangle\langle n^B|$ with respect to the orthonormal $\{|m^B\rangle\}$ basis we find that the trace vanishes unless $m = n$ (see footnotes 6 and 7) , in which case it equals 1, so

$$\langle H^A \rangle = \mathrm{tr} \left[\sum_{pqm} c_{pm} c^*_{qm}\, H^A |p^A\rangle\langle q^A| \right].$$

---

[12] For example, the trace of $H^A \otimes I^B$ is $r(0+1+2+...)\mathrm{tr}(I^B) = \mathrm{tr}(H^A)\mathrm{tr}(I^B)$.

A representation of ρ^{AB} with respect to the $\{|p^A m^B\rangle\}$ basis is (see footnote 7)

$$\begin{pmatrix}
\langle 0^A 0^B|\rho^{AB}|0^A 0^B\rangle & \langle 0^A 0^B|\rho^{AB}|0^A 1^B\rangle & . & \langle 0^A 0^B|\rho^{AB}|1^A 0^B\rangle & . & . \\
\langle 0^A 1^B|\rho^{AB}|0^A 0^B\rangle & \langle 0^A 1^B|\rho^{AB}|0^A 1^B\rangle & . & . & \langle 0^A 1^B|\rho^{AB}|1^A 1^B\rangle & . \quad . \\
. & . & : \; : & . & . & : \; : \\
\langle 1^A 0^B|\rho^{AB}|0^A 0^B\rangle & . & . & \langle 1^A 0^B|\rho^{AB}|1^A 0^B\rangle & . & . \\
. & \langle 1^A 1^B|\rho^{AB}|0^A 1^B\rangle & . & . & \langle 1^A 1^B|\rho^{AB}|1^A 1^B\rangle & . \quad . \\
. & . & : \; : & . & . & : \; : \\
 & & & & & . \; . \\
 & . & & & . & . \; .
\end{pmatrix}$$

or equivalently ([34], 3.2.1) with respect to the $\{|p^A\rangle\}$ basis of A (compare with equation (3))

$$\begin{pmatrix}
\langle 0^A|\rho^{AB}|0^A\rangle & \langle 0^A|\rho^{AB}|1^A\rangle & . \\
\langle 1^A|\rho^{AB}|0^A\rangle & \langle 1^A|\rho^{AB}|1^A\rangle & . \\
. & . & : \; :
\end{pmatrix}$$

where each element in the latter matrix corresponds to a submatrix of the former and is an operator on neural space B alone (compare with $H^A \otimes I^B$ above).

Equation (12) is then seen to be a kind of trace, a sum of diagonal submatrices.

Like ρ^{AB}, ρ^A is Hermitian, since for each term in the sum (equation (11)) in which $p = q$

$$c_{pm}c^*_{pm}|p^A\rangle\langle p^A| = (c_{pm}c^*_{pm}|p^A\rangle\langle p^A|)^\dagger$$

and for each term $p \neq q$, there is a corresponding term in the sum that is adjoint

$$c_{pm}c^*_{qm}|p^A\rangle\langle q^A| = (c_{qm}c^*_{pm}|q^A\rangle\langle p^A|)^\dagger.$$

If $|\psi^{AB}\rangle$ is normalised, then from equation (9)

$$\sum_{pm} c_{pm}c^*_{pm} = 1,$$

implying that the trace of both ρ^{AB} (equation (10)) and ρ^A (equation (11)) is 1.

ρ^{AB} is a positive operator since for any vector $|\varphi^{AB}\rangle$

$$\langle\varphi^{AB}|\psi^{AB}\rangle\langle\psi^{AB}|\varphi^{AB}\rangle = |\langle\varphi^{AB}|\psi^{AB}\rangle|^2 \geq 0.$$

But like H^A, ρ^A could be considered the product of adjoint operators[13]

$$\sum_{pm} c_{pm}|p^A\rangle\langle m^B|, \quad \sum_{pm} c^*_{pm}|m^B\rangle\langle p^A|$$

---

[13] The latter operator takes a vector from A to B: $|b\rangle = \left(\sum_{pm} c^*_{pm}|m^B\rangle\langle p^A|\right)|a\rangle$. In that the squared length of $|b\rangle$ must be positive or zero, the operator ρ^A is positive: $\langle b|b\rangle = \langle a|\rho^A|a\rangle \geq 0$.

so it must also be positive.

The inseparable state of a 2-neuron system in which firing is maximally correlated is written in density operator format

$$\rho^{AB} = 1/2|0^A 0^B\rangle\langle 0^A 0^B| + 1/2|0^A 0^B\rangle\langle 1^A 1^B| + 1/2|1^A 1^B\rangle\langle 0^A 0^B| + 1/2|1^A 1^B\rangle\langle 1^A 1^B|.$$

With respect to the basis $\{|0^A 0^B\rangle, |0^A 1^B\rangle, |1^A 0^B\rangle, |1^A 1^B\rangle\}$ of $A \otimes B$:

$$\rho^{AB} = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}.$$

The equivalent inseparable state in which firing is maximally anti-correlated is

$$\rho^{AB} = 1/2|0^A 1^B\rangle\langle 0^A 1^B| + 1/2|0^A 1^B\rangle\langle 1^A 0^B| + 1/2|1^A 0^B\rangle\langle 1^A 0^B| + 1/2|1^A 0^B\rangle\langle 0^A 1^B|$$

or

$$\rho^{AB} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In either inseparable state, taking the partial trace as in equation (12), the reduced states of neurons A and B are each even mixtures of 0- and 1-action potential states:

$$\rho^A = 1/2|0^A\rangle\langle 0^A| + 1/2|1^A\rangle\langle 1^A|,$$
$$\rho^B = 1/2|0^B\rangle\langle 0^B| + 1/2|1^B\rangle\langle 1^B|.$$

In a representation with respect to the integer action potential basis:

$$\rho^B = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \quad H^B = \begin{pmatrix} 0 & 0 \\ 0 & r \end{pmatrix}, \quad H^A \otimes H^B = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r^2 \end{pmatrix}$$

so that the perception value

$$\langle H^B \rangle = \mathrm{tr}(H^B \rho^B) = r/2.$$

The expectation value of $H^A \otimes H^B$ is similarly the trace of the matrix product $\mathrm{tr}(H^A \otimes H^B \rho^{AB})$, which in the case of the correlated inseparable state equals $r^2/2$. Correcting by perception values $r/2$, as in equation (13), yields the dependence measure Q_{AB}: the conditional instantaneous firing rate of B doubles, given a simultaneous spike in A.

REFERENCES

1. A. Duggins, *Biosystems* **102**, 124–33 (2010).
2. M. Tegmark, *Phys. Rev. E* **61**, 4194–206 (2000).
3. S. Zeki, and A. Bartels, *Conscious Cogn.* **8**, 225–59 (1999).
4. J. Mill, *An Analysis of the Phenomena of the Human Mind*, Longmans, Green, Reader, and Dyer Publisher, London, 1829.
5. W. Wundt, *Grundz§ge der Physiologischen Psychologie*, Wilhelm Engelmann, Leipzig, 1911.
6. E. Titchener, *A Textbook of Psychology*, Scholars' Facsimiles and Reprints, New York, 1910/1980.
7. M. Wertheimer, *A Brief History of Psychology*, Holt, Rinehart and Winston, New York, 1987.
8. E. Schrödinger, *Mind and matter*, Cambridge University Press, Cambridge, 1958.
9. T. Nagel, "What is the mind body problem?" in *Experimental and theoretical studies of consciousness; Ciba Foundation symposium*, Vol. 174. Wiley, New York pp. 316, 1993.
10. M. A. Nielsen, and I. L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, 2000.
11. R. Penrose, *Shadows of the mind: a search for the missing science of consciousness*, Random House, London, 1995.
12. C. J. S. Clarke, *Mind and Matter* **5**, 45–81 (2007).
13. J. K. O'Regan, and A. Noe, *Behav. Brain Sci.* **24**, 939–73; discussion 973–1031 (2001).
14. H. A. Baseler, A. A. Brewer, L. T. Sharpe, A. B. Morland, H. Jagle, and B. A. Wandell, *Nat. Neurosci.* **5**, 364–70 (2002).
15. I. Kohler, *Scientific American* **206**, 62–86 (1961).
16. S. Zeki, *A vision of the brain*, Blackwell Scientific Publications, Oxford, 1993.
17. F. W. Billmeyer, and M. Saltzman, *Principles of color technology*, Wiley, New York, 1981.
18. D. H. Perkel, G. L. Gerstein, and G. P. Moore, *Biophys. J.* **7**, 419–40 (1967).
19. E. N. Brown, R. E. Kass, and P. P. Mitra, *Nat. Neurosci.* **7**, 456–61 (2004).
20. D. H. Perkel, G. L. Gerstein, and G. P. Moore, *Biophys. J.* **7**, 391–418 (1967).
21. F. Rieke, D. Warland, R. de Ruyter van Stevenick, and W. Bialek, *Spikes: exploring the neural code*, MIT Press, Cambridge, Mass, 1999.
22. C. M. Gray, P. Konig, A. K. Engel, and W. Singer, *Nature* **338**, 334–337 (1989).
23. C. J. S. Clarke, *Found. Phys. Lett.* **15**, 495–500 (2002).
24. B. MacLennan, "Protophenomena: The Elements of Consciousness", in *Irreducibly Conscious: Towards a Non-Reductionist Account of the Mind,* A. Batthyány, et al. (Eds.), Universitätsverlag Winter, Heidelberg, pp. 189–214, 2007.
25. F. Crick, and C. Koch, *Nature* **375**, 121–123 (1995).
26. G. Tononi, and G. M. Edelman, *Science* **282**, 1846–1851 (1998).
27. V. A. F. Lamme, and P. R. Roelfsema, *Trends in Neurosci.* **23**, 571–579 (2000).
28. A. Pascual-Leone, and V. Walsh, *Science* **292**, 510–512 (2001).
29. V. A. Lamme, *Trends Cogn. Sci.* **10**, 494–501 (2006).
30. J. C. W. Edwards, *J. Cons. Studies* **12**, 60–76 (2005).
31. C. Koch, *The quest for consciousness: a neurobiological approach*, Roberts and Co., Denver, Colo, 2004.
32. T. Feinberg, *Neurocase* **6**, 75–81 (2000).
33. S. T. Roweis, and L. K. Saul, *Science* **290**, 2323–2326 (2000).
34. J. Preskill, *Lecture Notes for Physics 229: Foundations II: Measurement and Evolution, Advanced Mathematical Methods for Physics*, 1998.
    `http://theory.caltech.edu/~preskill/ph219/`

Fractal characterization of neural correlates of consciousness

A. J. Ibañez-Molina and S. Iglesias-Parro

Dpto. Psicología, Facultad de Humanidades y CCEE, Universidad de Jaén, Paraje las Lagunillas s/n 23071 Jaén, Spain.

Abstract. In this work we present a novel experimental paradigm, based on binocular rivalry, to address the study of internally and externally generated conscious percepts. Assuming the nonlinear nature of the EEG signals, we propose the use of fractal dimension to characterize the complexity of the EEG associated with each percept. Data analysis showed significant differences in complexity between the internally and externally generated percepts. Moreover, EEG complexity of auditory and visual percepts was unequal. These results support fractal dimension analyses as a new tool to characterize conscious perception.

Keywords: EEG; fractal; consciousness.
PACS: 87.19.le, 87.17.-h

NEURAL CORRELATES OF CONSCIOUS EXPERIENCES

Conscious experiences could be internally or externally produced. We will refer to the former as those subjective percepts caused directly by sensory inputs. The latter type arises when attention is focused on the own train of thoughts (mind wandering-MW). One reliable method to study conscious perception is to consider sensory inputs that elicit alternating subjective experiences (e.g., binocular rivalry, binaural rivalry). In binocular rivalry procedures, two visual inputs are presented independently to each eye producing an alternating perception of each one. In binaural rivalry, two dissimilar stimuli are presented one to each ear, only one being perceived by participants. Since switches between each input occur during fixed physical presentation, changes at any physiological measure are attributed to the conscious experience [1]. Although there is not a consensus, research suggests that Neural Correlates of Consciousness (NCC) depends on sensory modality. Thus, some areas involved in visual awareness (e.g., inferior temporal cortex) are not related with auditory NCC. On the contrary, areas like the medial temporal gyrus that have never been found linked with visual NCC are related with auditory percepts [2].

Conscious perception also arises from self-generated inputs. This type of conscious "mode" has been called MW and may appear during in-attention of an ongoing task or when an individual is not engaged at any task [3]. Smallwood and colleagues [4] proposed that MW arises with a combination of default mode network and the fronto-parietal control network that would protect internally generated trains of thought from disruption. MW has been characterized by the examination of oscillatory properties of the EEG. A study conducted by Braboszcz and Delorme [5] recorded the EEG of participants while doing a simple breath cycles counting task. MW states were located

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 182-184 (2013); doi: 10.1063/1.4776511

during the task by online reports of participants. Obtained results showed an increased power in the theta and delta activity and a decrease power in the alpha and beta bands.

NONLINEAR FRACTAL ANALYSIS

The fundamental assumption of non-linear methods is that EEG signals are generated by deterministic processes reflecting nonlinear associations between neuron assemblies [6]. One of the applications of the theory of non-linear dynamics to the study of EEG has relied in the use of the fractal dimension (FD) to characterize chaos variations under different conditions. FD of EEG series has been already applied to a variety of behavioral and cognitive tasks and populations (see, for example ref [7]). Although these studies suggest that EEG complexity is indicative of important properties of the functional organization of cortical structures, it has not been used to characterize NCC. Among all algorithms developed to calculate FD of time series data, HiguchiâĂŹs algorithm (HFD) [8] produces the more accurate estimation [9]. In addition, HFD may be computed with relatively short time series from the EEG.

GOALS AND PREDICTIONS

We wanted to study whether EEG complexity, measured with HFD can be used to characterize internally and externally conscious perceptions. In addition, we wanted to explore whether binocular rivalry and MW procedures can be successfully combined to extend previous investigations about NCC. In order to combine binocular rivalry and MW procedures, participants watched a video in which the visual and auditory outputs did not match. Hence, video and audio were mutually exclusive as binocular inputs in binocular rivalry experiments. As in experiments designed to study MW, EEG segments were assigned to visual, auditory and MW conditions according to introspective judgments of participants. We calculated HFD from EEG associated to each of conscious percepts. Given that externally generated percepts and MW have been related with different neural networks [4], it was reasonable to expect that our method would show complexity variations depending on the type of percept. Specifically, variations between MW and the rest of conditions are expected to be preponderant across the majority of electrodes while visual and auditory differences will be less distributed on the scalp.

RESULTS AND DISCUSSION

For each of 11 participants, 10 relevant EEG segments of 50s were extracted and labeled according with verbal self-reports as: visual content, auditory content or MW. Epochs were submitted to Infomax ICA [10] for EEG denoising. HFD means obtained for each epoch were submitted to a Multivariate Analysis of Variance with electrode as dependent variables and conscious percepts (visual, auditory or MW) as between-participant factor. Data analyses revealed global differences in signal complexity depending on type of conscious percepts. Specifically, Bonferroni post-hoc comparisons showed that complexity

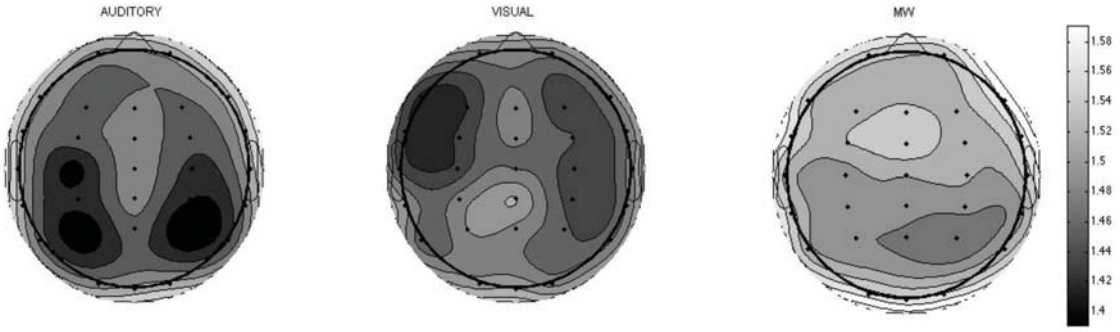


FIGURE 1. Topological distribution maps of HFD means for each conscious percept.

was significantly higher in fronto-temporal electrodes for auditory than visual percepts. Moreover, parietal electrodes showed the higher complexity for auditory percepts when compared with visual ones. Regarding MW, the HDF was higher when compared with visual and auditory percepts across the majority of channels. This result suggests more distributed neural generators during MW when compared with internally generated experiences (See Fig. 1).

This study compares the non-linear complexity of the NCC of internally and externally generated experiences. It was shown that the NCC of MW consisted of a higher complexity of the EEG signature across the entire scalp. This finding may indicate that neural networks supporting MW are more distinct and distributed in the cortex. The second important result in our experiment is that visual and auditory percepts differ in complexity at frontal and central-parietal sites, which may reflect unequal distribution of neural generators on sensory cortical areas for the two types of experiences. Our study clearly indicates that non-linear EEG complexity can be a fine measure of the NCC.

REFERENCES

1. R. Blake, and N. K. Logothetis, *Nat. Neurosci. Rev.* **3**, 13–21 (2002).
2. A. Brancucci, P. Franciotti, A. D'Anselmo, S. Della Penna, and L. Tommasi, *J. Neurosci.* **31**, 16611–16618 (2011).
3. M. Gruberger, Y. Levkovitz, A. Zangen, A. Ben-Simon, and T. Hendler, *Front. Hum. Neurosci.* **5**, 56 (2011).
4. J. Smallwood, K. Brown, B. Baird, and J. W. Schooler, *Brain Res.* **1428**, 60–70 (2012).
5. C. Braboszcz, and A. Delorme, *Neuroimage* **54**, 3040–3047 (2011).
6. E. Pereda, R. Quian Quiroga, and B. Bhattacharya, *Prog. Neurobiol.* **77**, 1–37 (2005).
7. J. Jeong, *Clin. Neurophysiol.* **115**, 1490–1505 (2004).
8. T. Higuchi, *Physica D* **31**, 277–283 (1988).
9. C. Gómez, A. Mediavilla, R. Hornero, D. Abásolo, and A. Fernández, *Med. Eng. Phys.* **31**, 306–13 (2009).
10. A. J. Bell, and T. J. Sejnowski, *Neural Comput.* **7**, 1129–1159 (1995).

Short-term synaptic plasticity and heterogeneity in neural systems

J. F. Mejias*, H. J. Kappen†, A. Longtin* and J. J. Torres**

*Dept. of Physics and Centre for Neural Dynamics, University of Ottawa, Canada.
†Donders Institute for Brain, Cognition and Behaviour, Radboud Univ. Nijmegen, The Netherlands.
**Institute "Carlos I" for Theoretical and Computational Physics, and Dept. of Electromagnetism and Physics of the Matter, University of Granada, Spain.

Abstract. We review some recent results on neural dynamics and information processing which arise when considering several biophysical factors of interest, in particular, short-term synaptic plasticity and neural heterogeneity. The inclusion of short-term synaptic plasticity leads to enhanced long-term memory capacities, a higher robustness of memory to noise, and irregularity in the duration of the so-called up cortical states. On the other hand, considering some level of neural heterogeneity in neuron models allows neural systems to optimize information transmission in rate coding and temporal coding, two strategies commonly used by neurons to codify information in many brain areas. In all these studies, analytical approximations can be made to explain the underlying dynamics of these neural systems.

Keywords: short-term depression and facilitation; storage capacity; Up and Down states; heterogeneity; neural coding.
PACS: 87.19.La

INTRODUCTION

Theoretical and computational modeling has become a powerful tool to deepen our understanding of neural systems. This is especially important when dealing with mechanisms or neural circuits that are not easily accessible experimentally, or when experimental data by itself is not enough to provide a clear picture of the phenomena under study. A prominent biophysical mechanism that fits in such a framework is short-term synaptic plasticity, also known as dynamic synapses [1]. This type of fast, activity-dependent variation of the synaptic strength has been shown to have a strong impact on a number of important neural tasks, such as cortical gain control [2], coincidence detection [3, 4], sound localization [5], broadband coding [6], up and down cortical transitions [7, 8], working memory [9] or network storage capacity [10, 11, 12]. There are still a number of open questions, and thus short-term synaptic plasticity deserves further attention from the theoretical and computational point of view.

Another important feature that can be studied in an efficient manner by using theoretical and computational modeling is the effect of cellular heterogeneity on the performance of neural networks. When investigating neural coding strategies used by brain circuits, most theoretical and computational studies do not take into account the intrinsic variability found among actual neurons. In addition, such variability is also usually difficult to control experimentally, making neural heterogeneity a typically underrated factor in neural information processing studies. In recent years, however, a number of

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 185-194 (2013); doi: 10.1063/1.4776513

studies have suggested a functional role of the intrinsic heterogeneity found in real neural systems, and in particular its possible influence on network synchronization [13, 14], signal transmission [15, 16], neural coding [17] and sensory processing [18, 19].

In this work, we review some of our recent results on these topics. First, we consider the effect of short-term synaptic plasticity on the storage abilities of attractor neural networks, showing that it leads to enhanced memory capacities [11, 12] and a higher robustness to noise [12]. Then, we consider a simplified model of a neural population to investigate the effect of short-term synaptic plasticity on the irregularity of the duration of up states in cortical networks [8], which has been observed in experiments [20, 21, 22]. Finally, we address the effect of some level of heterogeneity among neurons in a network on the optimization in the information transmission for rate coding and temporal coding, two strategies commonly used by neurons to codify information in many brain areas [17]. In the three studies presented, analytical approximations can be made to explain the underlying dynamics of these neural systems.

MAXIMUM STORAGE CAPACITY

We first study the influence of dynamic synapses on the storage abilities of neural networks (for more details, see [11, 12] and references therein). We consider a fully connected network of N binary neurons whose state $\mathbf{s} \equiv \{s_i = 0,1; \forall i = 1,\ldots,N\}$ follows a probabilistic parallel dynamics

$$P[s_i(t+1) = 1] = \frac{1}{2}\{1 + \tanh[2\beta(h_i(\mathbf{s},t) - \theta_i)]\} \quad \forall i = 1 \ldots N, \tag{1}$$

where $h_i(\mathbf{s},t)$ is the local field or the total input synaptic current to neuron i, namely

$$h_i(\mathbf{s},t) = \sum_{j \neq i} \omega_{ij} x_j(t) u_j(t) s_j(t). \tag{2}$$

Also, $\beta \equiv T^{-1}$ is a temperature or noise parameter (i.e., for $\beta \to \infty$ we have a deterministic dynamics), and θ_i represents the neuron firing threshold. The coefficients ω_{ij} are fixed synaptic weights, consequence of the slow learning process of M memory patterns of activity. In the following we choose the following learning rule

$$\omega_{ij} = \frac{1}{Nf(1-f)} \sum_{\mu=1}^{M} (\xi_i^\mu - f)(\xi_j^\mu - f), \tag{3}$$

where $\left\{\xi_i^\mu = 0,1; i = 1 \ldots N\right\}$ represents the M stored random patterns with mean activity over the patterns $\langle \xi_i^\mu \rangle = f = 1/2$. On the other hand, the variables x_j, u_j appearing in h_i describe the short-term depression and facilitation synaptic mechanisms, respectively. We assume that these variables evolve according to the discrete dynamics [23, 24]

$$x_j(t+1) = x_j(t) + \frac{1 - x_j(t)}{\tau_{rec}} - U_{SE} u_j(t) x_j(t) s_j(t) \tag{4}$$

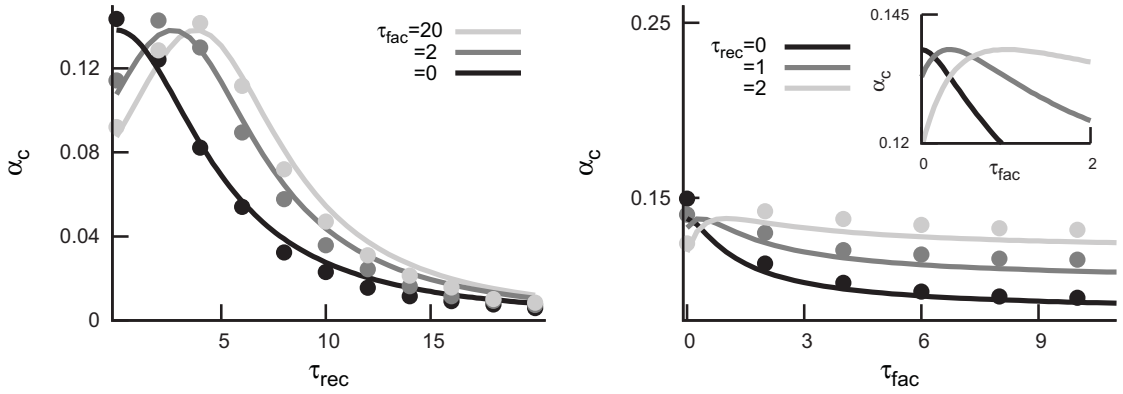


FIGURE 1. Critical network load in the presence of short-term synaptic plasticity for $U_{SE} = 0.2$. Left: network load vs depression time constant, for different values of the facilitation time constant. Right: critical network load vs facilitation time constant, for different values of the depression time constant (the inset shows the low τ_{fac} regime, where maxima appear). As one can see, the presence of facilitation allows optimal retrieval abilities in the network while preserving the dynamic nature of synapses. Mean-field predictions (lines) agree with simulations (symbols), which were done with a network of $N = 3000$ neurons.

$$u_j(t+1) = u_j(t) + \frac{1-u_j(t)}{\tau_{fac}} + (1-U_{SE}u_j(t))s_j(t). \tag{5}$$

Here, τ_{rec}, τ_{fac} are the time constants for depressing and facilitating processes respectively, and U_{SE} is a parameter related with synaptic unreliability. The original Hopfield model is recovered when $x_i = u_i = 1,\ \forall\ i,t$ (i.e. static synapses). By simple inspection of Eqs. (4-5), this corresponds to the case of $\tau_{rec}, \tau_{fac} \ll 1$ which makes x_j and $u_j\ \forall j$ quickly reach their maximum values, $x_j = u_j = 1\ \forall j,t$ (see [11] for a careful explanation of this limit). We also choose

$$\theta_i = \frac{1}{2}\sum_{j\neq i}\omega_{ij}. \tag{6}$$

The network load is defined as $\alpha \equiv M/N$. One can obtain a mean field solution of the network at the limit of zero temperature (see [11] for details) and obtain the following expression for the critical network load α_c (defined as the maximum number of patterns per neuron that the network is able to retrieve), simplified here for clarity purposes:

$$\alpha_c \sim \frac{0.138}{1+\left(\frac{1+\gamma\gamma'-\gamma'}{\gamma'}\right)^2}, \tag{7}$$

where $\gamma \equiv U_{SE}\tau_{rec}$ and $\gamma' \equiv \frac{1+\tau_{fac}}{1+U_{SE}\tau_{fac}}$. We can employ such a mean field solution to compare with the numerical simulations of the model and study the effect of the time constants $\tau_{rec},\ \tau_{fac}$ on the critical network load. As Fig. 1 shows, the presence of purely depressing synapses (that is, $\tau_{fac} = 0$) leads to low critical network loads; this effect is stronger for larger τ_{rec} (see also [24]). On the other hand, the presence of a certain level of facilitation allows an optimal network load (and therefore optimal retrieval abilities) for certain finite nonzero values of $\tau_{rec},\ \tau_{fac}$. This is highly desirable from both

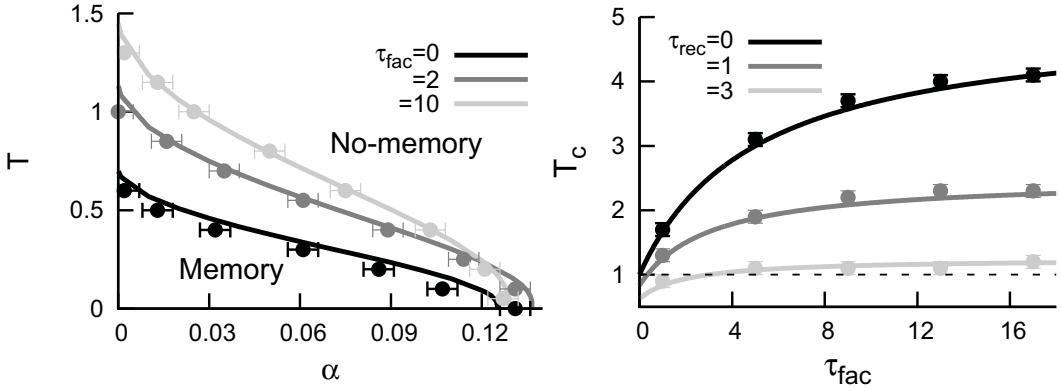


FIGURE 2. Effect of short-term plasticity on the retrieval abilities of noisy networks. Left: phase diagram of the network for $U_{SE} = 0.2$, $\tau_{rec} = 2$ and different values of τ_{fac}. Right: critical temperature as a function of τ_{fac}, for $U_{SE} = 0.2$ and different values of τ_{rec}. Mean-field predictions (lines) agree with simulations (symbols), which were performed with a network of $N = 3000$ neurons.

computational and biophysical points of view, since one would want to have a neural network with good retrieval abilities while preserving the dynamic nature of synapses, which endow it with further information processing and coding capabilities.

One can also extend the mean-field solution found in [11] to the case of nonzero temperatures, in order to study the behavior of the network in more general conditions (see [12] for details). In particular, we can compute the phase diagrams of the model as a function of the network load and temperature, and then investigate the effect of short-term plasticity on these diagrams. As the left panel of Fig. 2 shows, the presence of short-term facilitation increases the area of the memory phase, even in the presence of short-term depression [12]. The effect on the critical temperature T_c, which is the maximum temperature allowing good retrieval abilities and corresponds to the $\alpha \to 0$ case, is worth mentioning. We can observe clearly in the right panel of Fig. 2 that increasing τ_{fac} leads to large values of T_c, for different values of the depression time constant. Such dependence can be found analytically [12], and it is $T_c = \gamma'/(1+\gamma\gamma')$. These large T_c values comfortably surpasses the critical temperature of the standard Hopfield model (marked as a dashed line in the figure), indicating that the presence of facilitation allows the network to perform optimally in retrieval tasks even in strong noise conditions.

IRREGULARITY OF CORTICAL UP STATES

We have seen so far that, from a purely theoretical standpoint, short-term synaptic plasticity, and in particular short-term facilitation, may be highly beneficial for a neural network to improve information retrieval properties. However, short-term depression by itself is also crucial to explain certain features observed in actual neural systems. A good example of this is the spontaneous transitions between activity states observed in cortical areas in the brain, a phenomenon which is referred to as *up* and *down* transitions. Such behavior may provide a framework for neural computations [25], and could also

coordinate some sleep rhythms into a coherent rhythmic oscillatory behavior in cortical and thalamocortical areas [26, 27, 28].

A prominent feature of up and down transitions is that, for certain experimental conditions, the times between transitions seem to be highly irregular, and in particular the duration of up states is found to range from a scale of miliseconds to seconds [20, 21, 22]. While previous modeling studies (which consider up and down transitions as a phenomenon induced by synaptic depression) are not able to explain such irregularity [29, 7, 30], a combination of short-term depression and other biophysical factors could still be able to explain such erratic behavior. Here we briefly discuss a simplified model of a neural population, with short-term depression and synaptic noise, which is able to explain the irregularity found in the duration of up states (for more details, see [8]).

We assume that the dynamics of the firing rate of the neural population is described by

$$\frac{d\nu(t)}{dt} = -\nu(t) + \mathscr{F}[J\,x(t)\nu(t) - \theta] + \zeta(t), \tag{8}$$

where $\nu(t)$ is the mean firing rate or activity of the neural population, J is the synaptic coupling strength in absence of short-term depression, and θ is a firing threshold. The variable $\zeta(t)$ is a Gaussian white noise of zero mean and standard deviation δ, which takes into account the inner stochasticity of neurons. The term $\mathscr{F}[z] \equiv \frac{\nu_{max}}{2}(1+\tanh(z))$ is a sigmoidal function, which sets the up and down activity levels to $\nu = \nu_{max}$ and $\nu = 0$, respectively.

The variable $x(t)$ takes into account the synaptic variations due to short-term depression, and evolves according to

$$\frac{dx(t)}{dt} = \frac{1-x(t)}{\tau_{rec}} - U_{SE}\,x(t)\nu(t) + \frac{D}{\tau_{rec}}\xi(t), \tag{9}$$

where τ_{rec} is again the depression time constant, and U_{SE} is a parameter related with the synaptic unreliability (we do not consider short-term facilitation here). The last term (with $\xi(t)$ being a Gaussian white noise of zero mean and unitary variance, and D a constant representing the strength of the noise) takes into account any source of uncontrolled noise that is not included in $\zeta(t)$ and that could directly influence the synaptic strength, such as fluctuations in neurotransmitter release or receptor unreliability [31, 32].

A typical temporal evolution of this model is shown in Fig. 3A, while Fig. 3B corresponds to the histogram of population activity values, which reflects the bimodal nature found in experiments. The activity of up and down states can be easily identified as 5 spikes/s and 0.5 spikes/s, respectively.

One can develop a theoretical estimation of the probability distribution $P(T)$ of a certain up state duration T (such as the one displayed in Fig. 3A). Briefly, by carefully analyzing the conditions for the existence of a double well in the dynamics (8), one can find that only a limited window of values of $x(t)$ is compatible with the existence of such a double well. When $x(t)$ is above (below) this window, the double well turns into a single well centered at the up (down) activity state, respectively. If the synaptic noise is strong enough, the variable $x(t)$ will be constantly pulled away from this window, and as a result the synaptic strength will drive the transitions between up and down states. In

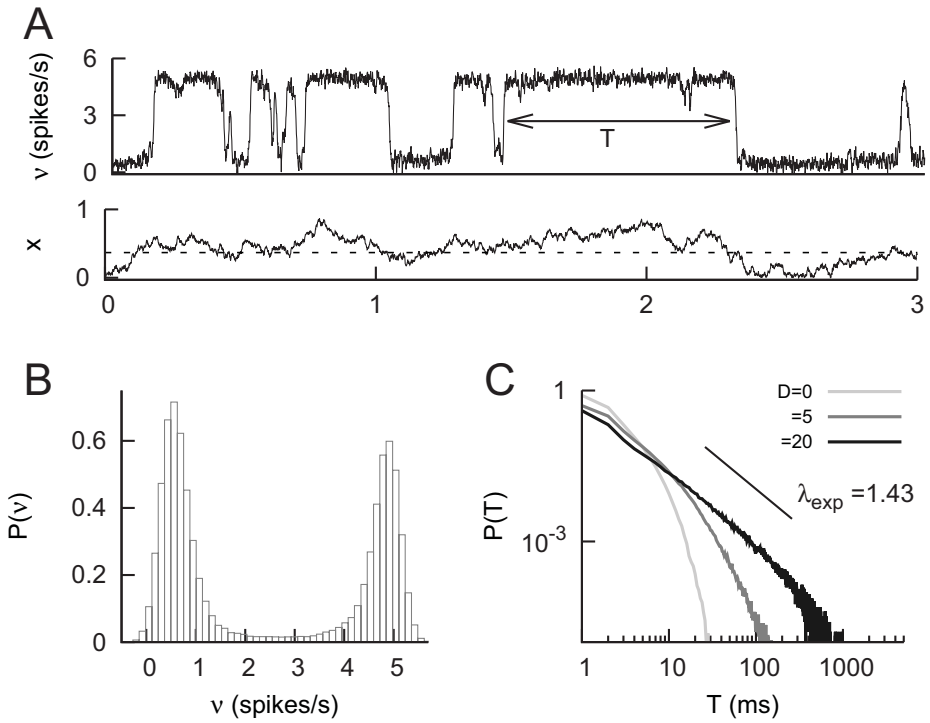


FIGURE 3. (A) Transitions between up and down states in a neural population with depressing synapses. In the model, the transition between the two possible activity levels is driven by the average strength of the synapses at a given time, namely $x(t)$. Population activity (top) as well as averaged synaptic strength (bottom) are shown, with the mean value of $x(t)$ denoted by a dashed line. (B) Histogram of the population firing rate, displaying the bistable nature of the population dynamics. (C) Distribution of duration times of the up state, for different levels of synaptic noise. The slope $\lambda_{exp} = 1.43$ corresponds to the value found in *in vitro* experiments (see main text for details). Parameters are $J = 1.1$, $U_{SE} = 0.6$, $\tau_{rec} = 1000\ ms$, $\delta = 0.3$ and $\nu_{max} = 5$ spikes/s.

practice, the double well window for $x(t)$ will be small and centered on its mean value x_0. One can see this effect in Fig. 3A, where large deviations of $x(t)$ from its mean value (dashed line) drive the population activity towards up or down states.

In this context, the presence of temporal correlations in $x(t)$ (reflected by the short-term depression time constant τ_{rec}) will increase the duration of the excursions of $x(t)$ out of the double well window, which in turn will imply an increment in the duration of the up states. Thus, calculating the distribution of *ruin times* of $x(t)$ (that is, the mean duration of the excursions of $x(t)$ away from its mean value) will give us the distribution of up state durations, which results in $P(T) \sim T^{-\lambda}$ with $\lambda = 1.5$.

The numerical simulations of our model indicate, as can be seen in Fig. 3C, that the distribution of up state durations $P(T)$ tends to a power-law distribution with slope ~ 1.5 for sufficiently large values of the synaptic noise D, as our theoretical estimations predict. The figure also shows, for a direct comparison, the slope of $\lambda_{exp} \sim 1.43$ found in *in vitro* experiments [21], in a very good agreement with our theoretical and numerical predictions.

HETEROGENEITY AND NEURAL CODING

In this last section, we focus on the effect of neural heterogeneity on the coding properties of neural networks (a detailed description of this section can be found in [17]). We start by considering a fully connected network of N excitatory neurons, where the dynamics of each neuron i is described by

$$\tau_m \frac{dV_i(t)}{dt} = -V_i(t) + S(t) + \mu + \sqrt{\tau_m}\sigma\xi_i(t) + \frac{\tau_m}{N}\sum_j \sum_k J\ \delta(t - t_{ij}^k), \qquad (10)$$

where τ_m is the neuron membrane time constant, V_i is its membrane potential, $S(t)$ is an external input signal to be determined, μ is a constant input bias, $\xi_i(t)$ is a gaussian white noise of zero mean and unitary variance, σ is the noise strength, J is the synaptic coupling strength, and the $k-th$ spike from neuron j arrives at neuron i at t_{ij}^k; the effect of this spike on the neuron is modeled as a delta-like pulse. Each neuron i is assumed to fire an action potential (AP) every time V_i reaches a certain firing threshold, and after that the membrane potential is reset to V_r for a time period τ_{ref}. In addition, we assume here that each neuron i has a firing threshold θ_i which is randomly distributed following a gaussian profile $P(\theta)$ with mean $\overline{\theta}$ and standard deviation w. Such heterogeneity reflects some of the variability in the individual excitability properties of neurons found in actual neural systems. The network will be homogeneous for $w = 0$, when the distribution of thresholds becomes a delta centered at $\overline{\theta}$.

We now consider that the external signal $S(t)$ is a weak, low-frequency sinusoidal signal. Assuming a homogeneous network (that is, $w = 0$) that remains in an asynchronous state (see [33]), the signal is able to slowly modulate the mean firing rate of the network (see Fig. 4A). Such modulation of the firing rate, which captures and transmits the information of the signal, is referred to as *rate coding* [34, 35]. In order to investigate the effects of neural heterogeneity on information transmission and neural coding, we now assume a certain level of heterogeneity in the network (so $w > 0$) and compute the modulation of the mean firing rate due to $S(t)$ in this case. We observe, both analytically and numerically, that the strength of the modulation depends on the level of heterogeneity in a non-trivial way (Fig. 4B). In particular, we found a non-monotonic behavior of the modulation (measured as an input-output correlation) with w, suggesting that there is a certain neural heterogeneity level which optimizes information transmission under rate coding. This optimization is found to be caused by the nonlinear effect of heterogeneity on the baseline mean firing rate of the network [17].

On the other hand, if we assume that the working point of our homogeneous network ($w = 0$) is close to the stability line of the system, a small external perturbation may be able to destabilize the network and make the neurons synchronize briefly, producing a population spike [33, 36]. Networks can use this high sensitivity to small external perturbations to process well time-located incoming signals, as seen in Fig. 4C. Such a detection strategy, which strongly relies on the generation of sharp responses precisely located in time, is known as temporal coding [34, 35]. When considering the effect of neural heterogeneity in information transmission under temporal coding (for which we use here the positive predictive value, PPV, a widely used measure [37]), we observe a nonlinear dependence as well (see Fig. 4D). More precisely, we find a non-monotonic

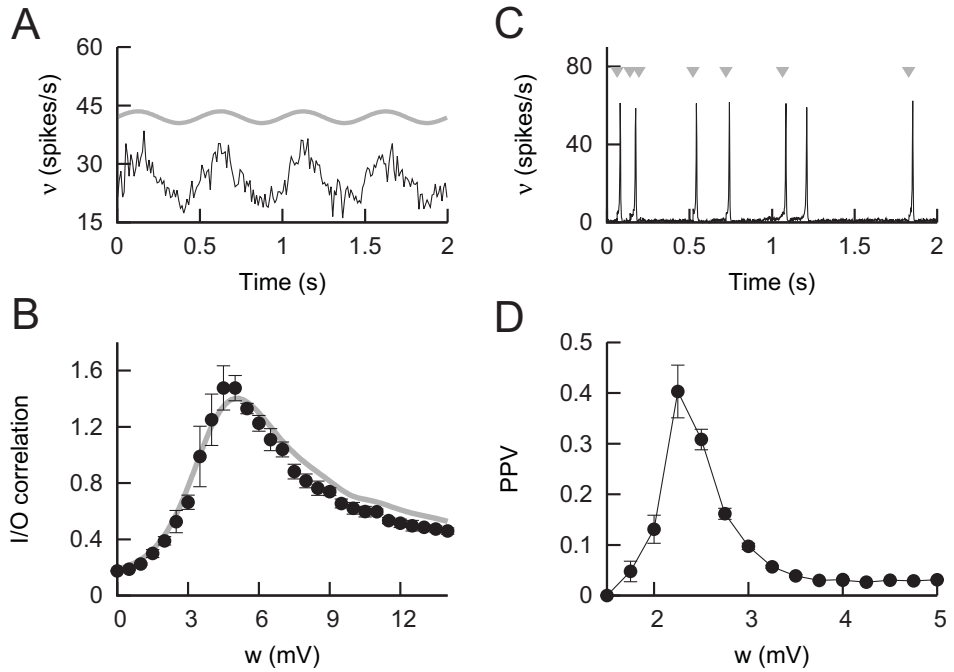


FIGURE 4. Effect of neural heterogeneity on the coding properties of spiking neural networks. (A) Modulation of the network mean firing rate (black) due to a weak sinusoidal signal (grey); the network is in an asynchronous working regime and $w = 0$. (B) Signal transmission, measured as the input-output correlation of the system, is optimized for a certain nonzero level of neural heterogeneity under rate coding. Theoretical estimation (line) agrees with numerical simulations (points). (C) Detection of sharp inputs (triangles) by the network through population-spike generation. The network is close to the stability line, and $w = 0$. (D) Signal transmission, measured as the PPV, is optimized for a certain nonzero level of neural heterogeneity under temporal coding. Parameter values are $N = 1500$ (for simulations), $\tau_m = 20\ ms$, $\mu = 14\ mV$, $V_r = 10\ mV$, $\tau_{ref} = 5\ ms$, $\overline{\theta} = 20\ mV$ and $\sigma = 3\ mV$. Coupling strength J is $10\ mV$ (A, B) or $20\ mV$ (C, D).

shape of the PPV with w, suggesting that a certain neural heterogeneity value is able to optimize the detection of signals under temporal coding. This phenomenon is due to a diversity-induced synchronization of the neurons which leads to a resonance-like behavior of the system [15, 16, 17].

CONCLUSION

In this work, we have reviewed some recent results concerning the role of several factors, such as short-term synaptic plasticity or neural heterogeneity, on the dynamics of neural networks. In the case of short-term synaptic plasticity, we have shown that short-term facilitation has a positive impact (with respect to purely depressing synapses) in the retrieval abilities of attractor neural networks [11]. In particular, the presence of facilitation allows to have a neural network with good retrieval abilities while preserving the natural fast dynamics of synaptic weights, convenient for information processing and coding. Furthermore, the storage properties of attractor neural networks are improved

with facilitation also in high-noise conditions, yielding a larger area of good memory retrieval and higher critical temperatures [12]. This suggests that facilitation may have an important role in helping actual neural circuits to access and maintain previously stored information in the presence of strong noise, which would occur, for instance, in working memory tasks [9].

Short-term depression, on the other hand, has been found to have a major impact on several tasks, such as gain control [2] or sound localization [5]. As we reviewed here, the combination of short-term synaptic depression and synaptic noise may be responsible for the irregularity of the duration of up states [8]. Such irregularity has been observed in several experimental conditions [20, 21, 22], although there is also experimental evidence of a much lower irregularity in different conditions [22, 38, 39]. Therefore, further theoretical and experimental work is needed to understand the origin of the irregularity of up state duration, and why it is absent in some situations. This absence could involve, for instance, underlying mechanisms modulating the level of short-term depression or the strength of synaptic noise.

Finally, we have analyzed the role of intrinsic neural heterogeneity on the coding properties of spiking neural networks. Contrary to what is traditionally assumed, heterogeneity on neural systems does not only allow networks to process information properly, but a certain level of such heterogeneity may even boost the abilities of the network to process and transmit information under several neural strategies, namely, rate coding and temporal coding. Such results are found to be robust in more realistic structures, such as in sparsely connected networks of excitatory and inhibitory neurons [17]. The way in which short-term synaptic plasticity and neural heterogeneity could interact in the processing of information constitutes an open question which is currently under study.

ACKNOWLEDGMENTS

This work was supported by *Ministerio de Ciencia y Tecnología* and *Junta de Andalucía* of Spain, and NSERC Canada. We thank J. Marro for useful discussions.

REFERENCES

1. R. S. Zucker, and W. G. Regehr, *Ann. Rev. Physiol.* **64**, 355–405 (2002).
2. L. F. Abbott, J. A. Varela, K. Sen, and S. B. Nelson, *Science* **275**, 220–224 (1997).
3. H. Kuba, K. Koyano, and H. Ohmori, *Eur. J. Neurosci.* **15**, 984–990 (2002).
4. J. F. Mejias, and J. J. Torres, *J. Comput. Neurosci.* **24**, 222–234 (2008).
5. D. L. Cook, P. C. Schwindtt, L. A. Grande, and W. J. Spain, *Nature* **421**, 66–70 (2003).
6. B. Lindner, D. Gangloff, A. Longtin, and J. E. Lewis, *J. Neurosci.* **29**, 2076–2088 (2009).
7. D. Holcman, and M. Tsodyks, *PLoS Comput. Biol.* **2**, 174–181 (2006).
8. J. F. Mejias, H. J. Kappen, and J. J. Torres, *PLoS One* **5**, e13651 (2010).
9. S. Romani, D. J. Amit, and G. Mongillo, *J. Comput. Neurosci.* **20**, 201–217 (2006).
10. D. Bibitchkov, J. M. Herrmann, and T. Geisel, *Network: Comput. Neural Syst.* **13**, 115–129 (2002).
11. J. F. Mejias, and J. J. Torres, *Neural Comput.* **21**, 851–871 (2009).
12. J. F. Mejias, B. Hernandez-Gomez, and J. J. Torres, *EPL* **97**, 48008 (2012).
13. L. Neltner, D. Hansel, G. Mato, and C. Meunier, *Neural Comput.* **12**, 1607–1641 (2000).
14. E. Persi, D. Horn, V. Volman, R. Segev, and E. Ben-Jacob, *Neural Comput.* **16**, 2577–2595 (2004).
15. C. J. Tessone, C. R. Mirasso, R. Toral, and J. D. Gunton, *Phys. Rev. Lett.* **97 (19)**, 194101 (2006).

16. T. Perez, C. R. Mirasso, R. Toral, and J. D. Gunton, *Phil. Trans. R. Soc. A* **368**, 5619–5632 (2010).
17. J. F. Mejias, and A. Longtin, *Phys. Rev. Lett.* **108** (2012).
18. G. Marsat, and L. Maler, *J. Neurophysiol.* **104**, 2543–2555 (2010).
19. K. Angelo, E. A. Rancz, D. Pimentel, C. Hundahl, J. Hannibal, A. Fleischmann, B. Pichler, and T. W. Margrie, *Nature* **488**, 375–378 (2012).
20. I. Lampl, I. Reichova, and D. Ferster, *Neuron* **22**, 361–374 (1999).
21. S. Fujisawa, N. Matsuki, and Y. Ikegaya, *Cereb. Cortex* **16**, 639–654 (2006).
22. G. Deco, D. Marti, A. Ledberg, R. Reig, and M. V. Sanchez-Vives, *PLoS Comput. Biol.* **5**, e1000587 (2009).
23. M. V. Tsodyks, and H. Markram, *Proc. Natl. Acad. Sci. USA* **94**, 719–723 (1997).
24. J. J. Torres, L. Pantic, and H. J. Kappen, *Phys. Rev. E.* **66**, 061910 (2002).
25. D. A. McCormick, *Curr. Biol.* **15**, R294–R296 (2005).
26. M. V. Sanchez-Vives, and D. A. McCormick, *Nat. Neurosci.* **3**, 1027–1034 (2000).
27. M. Steriade, A. Nunez, and F. Amzica, *J. Neurosci.* **13**, 3252–3265 (1993).
28. M. Steriade, A. Nunez, and F. Amzica, *J. Neurosci.* **13**, 3266–3283 (1993).
29. L. Pantic, J. J. Torres, H. J. Kappen, and S. C. A. M. Gielen, *Neural Comput.* **14**, 2903–2923 (2002).
30. J. J. Torres, J. Cortes, J. Marro, and H. Kappen, *Neural Comp.* **19**, 2739–2755 (2007).
31. L. E. Dobrunz, and C. F. Stevens, *Neuron* **18**, 995–1008 (1997).
32. A. Zador, *J. Neurophysiol.* **79**, 1219–1229 (1998).
33. N. Brunel, *J. Comp. Neurosci.* **8**, 183–208 (2000).
34. A. Borst, and F. E. Theunissen, *Nat. Neurosci.* **2**, 947–957 (1999).
35. A. Kumar, S. Rotter, and A. Aertsen, *Nat. Rev. Neurosci.* **11**, 615–627 (2010).
36. A. Loebel, and M. Tsodyks, *J. Comput. Neurosci.* **13**, 111–124 (2002).
37. X. H. Zhou, N. A. Obuchowski, and D. M. McClish, *Statistical Methods in Diagnostic Medicine*, Wiley and Sons, 2002.
38. M. Mattia, and M. V. Sanchez-Vives, *Cognit. Neurodyn.* **6**, 239–250 (2012).
39. J. M. Benita, A. Guillamon, G. Deco, and M. V. Sanchez-Vives, *Front. Comput. Neurosci.* **6**, 64 (2012).

Neural networks with dynamical synapses: from mixed-mode oscillations and spindles to chaos

K. Lee*, A. V. Goltsev*,†, M. A. Lopes* and J. F. F. Mendes*

*Departamento de Física da Universidade de Aveiro, I3N, 3810-193 Aveiro, Portugal.
†A.F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia.

Abstract. Understanding of short-term synaptic depression (STSD) and other forms of synaptic plasticity is a topical problem in neuroscience. Here we study the role of STSD in the formation of complex patterns of brain rhythms. We use a cortical circuit model of neural networks composed of irregular spiking excitatory and inhibitory neurons having type 1 and 2 excitability and stochastic dynamics. In the model, neurons form a sparsely connected network and their spontaneous activity is driven by random spikes representing synaptic noise. Using simulations and analytical calculations, we found that if the STSD is absent, the neural network shows either asynchronous behavior or regular network oscillations depending on the noise level. In networks with STSD, changing parameters of synaptic plasticity and the noise level, we observed transitions to complex patters of collective activity: mixed-mode and spindle oscillations, bursts of collective activity, and chaotic behavior. Interestingly, these patterns are stable in a certain range of the parameters and separated by critical boundaries. Thus, the parameters of synaptic plasticity can play a role of control parameters or switchers between different network states. However, changes of the parameters caused by a disease may lead to dramatic impairment of ongoing neural activity. We analyze the chaotic neural activity by use of the 0-1 test for chaos (Gottwald, G. & Melbourne, I., 2004) and show that it has a collective nature.

Keywords: short-term synaptic depression; chaotic neural activity; brain rhythms; stochastic neural network.
PACS: 87.18.Sn, 87.19.lg, 87.19.lw, 87.19.lj, 87.10.Mn

INTRODUCTION

Short-term synaptic depression (STSD) is an important form of short-term plasticity that provides a dynamic gain-control mechanism enhancing the sensitivity of cortical neurons to afferent firing patterns and expanding the range of possible coding strategies for cortical neurons [1, 2]. Recent experimental studies and phenomenological model of STSD (so-called Tsodyks-Markram (TM) model, [3]) have reported that transmission across neocortical synapses depends on the frequency of presynaptic activity (spike-timing dependent plasticity, STDP). Hence, synaptic efficacy is changed and adapted according to the dynamics of presynaptic and postsynaptic neurons. In turn, changes in synaptic efficacies influence activity of neurons. Thus, interplay between STSD and neuronal activity is an underlying mechanism that influences collective dynamics of neural network, in particular, brain rhythms. At the present time, understanding of this influence is elusive.

The brain is always surrounded by noise and also it is noisy. Noise leads to stochastic processes that are important ingredients of dynamics of neural networks. Intuitively, noise is damaging. However, in neural networks, noise can play a positive role support-

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 195-201 (2013); doi: 10.1063/1.4776517

ing oscillations and synchrony [4, 5].

In the present work, we study the role of STSD in the formation of complex patterns of neuronal activities. We use a cortical circuit model of neural networks with structure of a sparsely connected random network and with excitatory and inhibitory neurons that have stochastic dynamics [6]. The advantage of this model is that it can be studied analytically and does not need time-consuming simulation. For describing STSD, we apply the TM model to excitatory-excitatory synapses. In our simulations of the cortical model, we observed that STSD yields a rich repertoire of neuronal activities, such as mixed-mode oscillations (MMOs), spindles, and chaotic behavior. The parameters of synaptic plasticity can play a role of control parameters or switchers between different network states. We analyze the chaotic neural activity by use of the 0-1 test for chaos [7, 8] and show that it has a collective nature.

CORTICAL NETWORK MODEL

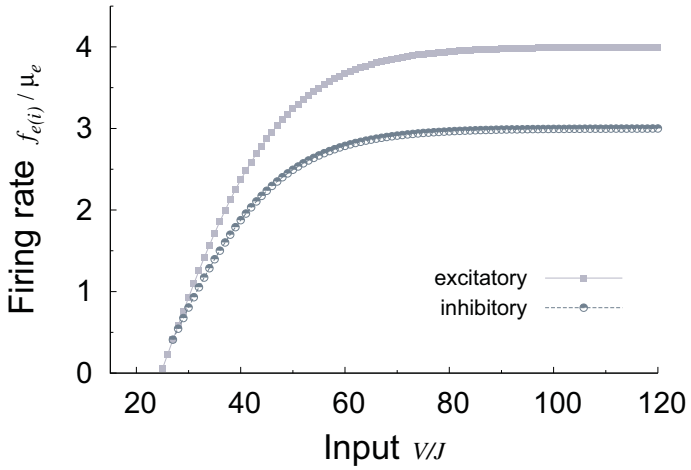


FIGURE 1. Spike frequencies of excitatory (squares) and inhibitory (circles) neurons versus input V.

Network structure. Let us consider two types of neurons: excitatory and inhibitory neurons. The total number of neuron is N. The fractions of excitatory and inhibitory neurons are g_e and $g_i = 1 - g_e$, respectively. Neurons are linked at random by directed edges and form a directed classical random graph with an adjacency matrix a_{nm}, where $m, n = 1, 2, ..., N$. An entry a_{nm} is equal to 1 if there is an edge directed from neuron n to neuron m, otherwise, $a_{nm} = 0$. If neuron m is connected to presynaptic neuron n, the synaptic efficacy of this synapses is $J_{nm}(t)$. Each neuron can be in either an active or inactive state. Active neurons fire random trains of spikes with a Poisson inter-spike intervals distribution [4, 9]. We assume that there is no phase correlation between trains of random spikes generated by different neurons.

Noise. Neurons are bombarded by random spikes that represent synaptic noise and random spikes from other areas of the brain. This is the only source of noise in our model. In simulations, we used a Gaussian external noise with a mean rate of random spikes in the range from 0 to 20 kHz.

Input-output relationships. Neurons demonstrate various types of spiking behavior in response to stimulus [10, 11]. There are two types of threshold behavior: neurons that show continuous (type 1) or discontinuous (type 2) input-output curves [12, 13]. Excitatory neurons (pyramidal cells) show type 1 behavior, whereas inhibitory neurons (interneurons) show type 2 behavior (see Fig. 1).

The total input at neuron m is a sum of spikes that arrive from presynaptic neurons and random spikes from noise during the time interval $[t-\tau, t]$ where τ is the integration time:

$$V_m(t) = \sum_{n=1}^{N} k_n(t) a_{nm} J_{nm} + \xi(t), \tag{1}$$

where $k_n(t)$ is the number of spikes that arrive from presynaptic neuron n and ξ is the number of random spikes from noise. Here we consider the case $\tau f < 1$, where f is the frequency of presynaptic neuron. In this case, a postsynaptic neuron receives only one spike ($k_n = 1$) or none ($k_n = 0$) from an active presynaptic neuron during the integration time τ. The value τf has a meaning of the probability that a postsynaptic neuron receives a spike from an active presynaptic neuron during time τ.

Activation-deactivation. If the total input $V_m(t)$ at an inactive excitatory or inhibitory neuron m at time t is at least a certain threshold Ω_a (i.e. $V_m(t) \geq \Omega_a$), then neuron m is activated with a rate μ_a where $a = e$ for excitatory and $a = i$ for inhibitory neurons, respectively. If $V_m(t) < \Omega_a$, then active neuron m is inactivated at a rate μ_a. For simplicity, we assume that the rate μ_a does not depend on the input. The relation between excitatory and inhibitory activation rates is $\mu_i = \alpha \mu_e$ where α is a parameter in the model.

Short-term synaptic depression. The model of short-term synaptic depression takes into account depletion of a pool of vesicles containing neurotransmitters [3]). In the present paper, we assume that only excitatory-excitatory synapses are dynamical. We use the model proposed in [3]. The strength of synapse depends on time as follows $J_e(t) = J_0 y(t)$ where J_0 is the absolute strength and $y(t)$ is the fraction of releasable synaptic resources. The rate equation for evolution of y is

$$\frac{dy}{dt} = \frac{1-y}{\tau_R} - P_d y \delta(t - t_s), \tag{2}$$

where P_d is the fraction of available synaptic resources ($0 \leq P_d \leq 1$), τ_R is the recovery time, and t_s is the arrival time of a presynaptic spike.

In numerical simulations, we studied sparsely connected networks and used the following algorithm. We divided time t into intervals of width $\Delta t = \tau$. At each time step, for each neuron, we calculated the input Eq. (1), taking into account that each active presynaptic neuron contributes with a spike with probability $\Delta t f$ where f is a firing rate of presynaptic neuron. The number of random spikes from noise in this input is generated according to the Gaussian distribution with the mean number $< \xi >$ of random spikes. Then, with the probability $\Delta t \mu_a$ states of neurons are updated by use of the stochastic rules formulated above.

In simulations, we used the following parameters: the network size $N = 10^3 - 10^5$; the mean degree 10^3; the fraction of inhibitory neurons $g_i = 0.25$; the absolute strength of

excitatory efficacies $J_e(0) = 1$; the inhibitory efficacies $J_i = -3$; $\Omega_e = 25$ and $\Omega_i = 27$ are excitatory and inhibitory activation thresholds, respectively; the maximum mean firing rates of excitatory and inhibitory neurons are $f_e^{max} = 4$ and $f_i^{max} = 3$, respectively; the activation rate of excitatory neurons $\mu_e = 1$; the deviation of the Gaussian random noise $\sigma = 10$; the ratio $\mu_i/\mu_e = \alpha$ was in interval $[0, 1]$; the fraction of available synaptic resources $P_d = [0, 1]$. These parameters agree with measurements in cortex.

RESULTS

In order to describe dynamics of a neural network, we introduce the fractions $\rho_e(t)$ and $\rho_i(t)$ of active excitatory and inhibitory neurons, respectively. $\rho_e(t)$ and $\rho_i(t)$ are so called "activities"of the excitatory and inhibitory populations. These activities can be related with neuronal activity in EEG measurements in the brain. To clarify the effects of depressing synapses, first we study the cortical model in the absence of synaptic plasticity. This model has a complex phase diagram and demonstrates first-order phase transitions, hysteresis phenomena, and neural avalanches in activation processes, and damped and sustained network oscillations [6]. The phase diagram in $\alpha - < \xi >$ plane is shown in Fig. 2. On can see that, in neural networks without synaptic plasticity, neuronal activity shows different forms of collective behavior: (I) an asynchronous state with a weak neural activity; (II) an active state with damped oscillations at $\alpha < 1$; (III) sustained network oscillations. We analyzed the phase diagram by use of an analytical approach and simulations and found a very good agreement at $N = 100000$. To investigate the effect of STSD on dynamics of neural networks, we chose representative points on the phase diagram in regions II and III, and near the phase boundary between II and III (see the points A, B, C in Fig. 2). The points, A, B, and C correspond to the same noise level $< \xi >= 15$ but different values of the parameter α, α = 0.2, 0.6, and 0.8, respectively.

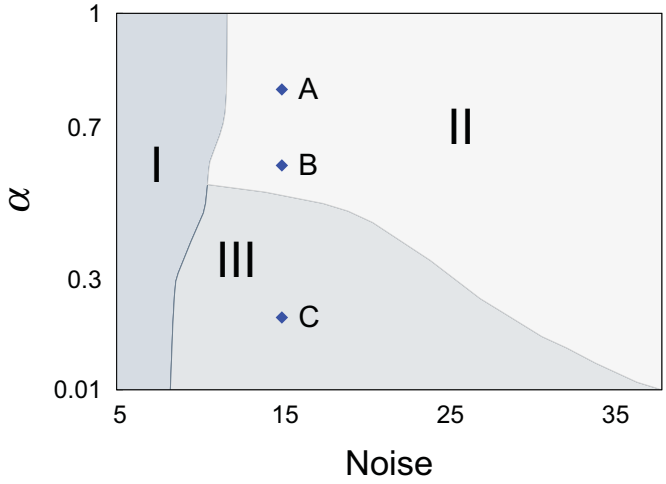


FIGURE 2. α–$< \xi >$ plane of the phase diagram of the considered cortical model: (I) an asynchronous state with a weak neural activity; (II) an active state; (III) sustained network oscillations. Points A, B, and C show the parameters chosen to study the role of STSD: (A) $\alpha = 0.8$ and $< \xi >= 15$; (B) $\alpha = 0.6$ and $< \xi >= 15$; (C) $\alpha = 0.2$ and $< \xi >= 15$.

Our simulation showed that the model with dynamical (STSD) excitatory-excitatory synapses demonstrates various patterns of collective neuronal activities that were experimentally observed in vitro and in vivo mammalian brain.

At point C in Fig. 2, we observed that sustained oscillations of neuronal activity reveal amazing robustness against noise level $< \xi >$ and STSD parameters P_d and τ_R. With increasing the parameter P_d and/or the recovery time τ_R, new patterns of sustained oscillations appear, so-called mixed-mode oscillations (MMOs) (see Fig. 3).

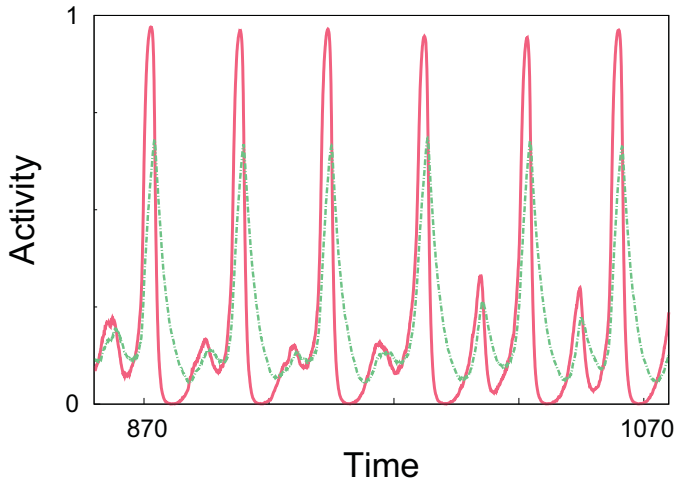


FIGURE 3. Activity of excitatory (solid line) and inhibitory (dashed-dotted line) populations versus time in the case of mixed-mode oscillations. Parameters: $\alpha = 0.2$, $< \xi >= 15$, $P_d = 0.01$, $\tau_R = 51$.

In the case of mixed state without plasticity (point B in Fig. 2), in a certain range of STSD parameters $\{P_d, \tau_R\}$, a new phase with a so-called spindle-like oscillations appears. For example, at parameters $P_d = 0.008$ and $\tau_R = 9$, we found $4-5$ Hz spindle-like oscillations (about 24 spindles/min) (see Fig. 4b). It is interesting to note that in the brain, spindles are generated during early stages of sleep. The frequency of spindles characterizes the quality of sleep.

At the point A in Fig. 2, in a certain range of parameters P_d and τ_R, the model with STSD demonstrated a new phase with chaotic neuronal activities as shown in Fig. 4c. Emergence of chaos is very intriguing phenomenon which was found in EEG measurements [14, 15]. Usually, MMOs and spindle oscillations are considered as examples of chaotic behavior.

To check whether, in our model, the observed activity is chaotic or not, we used a recently proposed method, so called the 0-1 test for chaos introduced by Gottwald and Melbourne [7, 8]. If the parameter K_c calculated by this method is equal to 1, then the behavior is chaotic. For a chaotic behavior in Fig. 4c, we obtained $K_c = 0.971950$. This confirms that the observed behavior is chaotic. In the regime with spindles in Fig. 4b, we obtained $K_c = 0.764383$. For sustained network oscillation in Fig. 4a, we found a much smaller value, $K_c \simeq 0$.

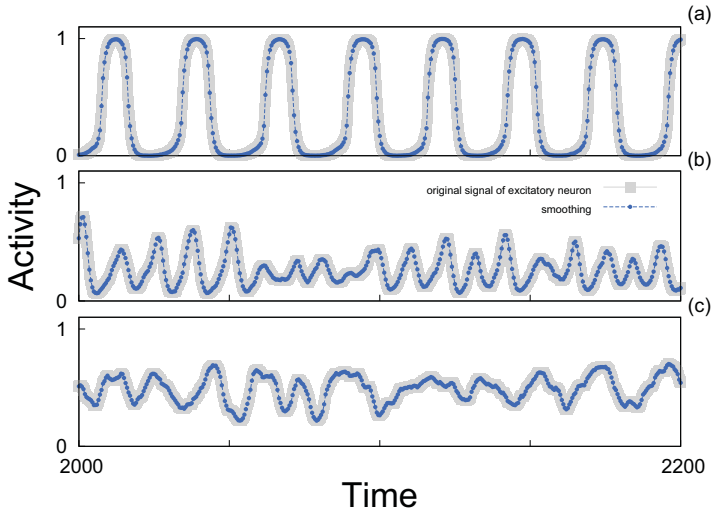


FIGURE 4. Three patterns of neuronal activity versus time and the parameter K_c from the 0-1 test for chaos. (a) $K_c = 0.17$ for sustained oscillations ($\alpha = 0.2$, $< \xi >= 15$, $P_d = 0.002$, $\tau_R = 4$). (b) $K_c = 0.76$ for spindle oscillations ($\alpha = 0.6$, $< \xi >= 15$, $P_d = 0.008$, $\tau_R = 9$). (c) $K_c = 0.97$ for chaotic behavior ($\alpha = 0.8$, $< \xi >= 15$, $P_d = 0.004$, $\tau_R = 2.5$). Squares and solid lines (gray color online) represent the original activity of excitatory neurons. Circles and dashed lines (blue color online) show smoothing points (time interval is equal to 5 integration times, 5τ).

CONCLUSION

In conclusion, our investigation of the cortical circuit model with excitatory and inhibitory neurons have showed that the short-term synaptic depression (STSD) plays a very important role in the formation of various patterns of collective behavior in neuronal networks. We have performed intensive simulations and showed that if the STSD is absent, then, in dependence on the noise level, the neural network exhibits either asynchronous behavior or sustained network oscillations. Taking into account excitatory synapses with STSD leads to a rich repertoire of network oscillations such as mixed-mode oscillations, spindles, and chaotic activity. We have analyzed the chaotic neural activity by use of the 0-1 test for chaos [7, 8] and have showed it has a collective nature. These complex patterns of collective activity are stable in a certain range of the STSD parameters and are separated by critical boundaries. Thus, the parameters of synaptic plasticity can play a role of control parameters or switchers between different network states. On the other hand, if changes of the parameters are caused by a disease, then this may lead to dramatic dysfunction of ongoing neuronal activity.

ACKNOWLEDGMENTS

This work was partially supported by the PTDC projects SAU-NEU/103904/2008, FIS/108476/2008, MAT/114515/2009, and PEst-C/CTM/LA0025/2011. K. E. Lee was supported by FCT under grant SFRH/BPD/ 71883/2010, M. A. L. was supported by FCT under Grant No. SFRH/BD/ 68743/2010.

REFERENCES

1. L. F. Abbott, J. A. Varela, K. Sen, and S. B. Nelson, *Science* **275**, 221–224 (1997).
2. M. Tsodyks, and H. Markram, *Proc. Nat. Acad. Sci. USA* **94**, 719–723 (1997).
3. M. Tsodyks, K. Pawelzik, and H. Markram, *Neural Comput.* **10**, 821–835 (1998).
4. G. B. Ermentrout, R. F. Galán, and N. N. Urban, *Trends Neurosci.* **31**, 428–434 (2008).
5. A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, *Nat. Rev. Neurosci.* **9**, 292–303 (2008).
6. A. V. Goltsev, F. V. de Abreu, S. N. Dorogovtsev, and J. F. F. Mendes, *Phys. Rev. E* **81**, 061921 (2010).
7. G. A. Gottwald, and I. Melbourne, *Physica D* **212**, 100 – 110 (2005).
8. G. Gottwald, and I. Melbourne, *SIAM J. Appl. Dyn. Syst.* **8**, 129–145 (2009).
9. C. Bédard, H. Kröger, and A. Destexhe, *Phys. Rev. Lett.* **97**, 118102 (2006).
10. E. M. Izhikevich, *IEEE Trans. Neur. Netw.* **15**, 1063–1070 (2004).
11. E. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*, Computational Neuroscience, MIT Press, 2006.
12. T. Tateno, A. Harsch, and H. P. C. Robinson, *J. Neurophysiol.* **92**, 2283–2294 (2004).
13. A. Bogaard, J. Parent, M. Zochowski, and V. Booth, *J. Neurosci.* **29**, 1677–1687 (2009).
14. A. Babloyantz, *Behav. & Brain Sci.* **10**, 173–174 (1987).
15. C. A. Skarda, and W. J. Freeman, *Behav. & Brain Sci.* **10**, 161–173 (1987).

Stochastic resonance as an emergent property of neural networks

M. A. Lopes*, A. V. Goltsev*,†, K.-E. Lee* and J. F. F. Mendes*

*Department of Physics & I3N, University of Aveiro, 3810-193 Aveiro, Portugal.
†Ioffe Physico-Technical Institute, 194021, St. Petersburg, Russia.

Abstract. In biological sensory systems, a presence of noise can actually enhance detection of weak signals. This phenomenon is called stochastic resonance (SR). We show that SR can emerge as a collective phenomenon in neural networks. We consider a cortical circuit model composed by stochastic excitatory and inhibitory neurons that form a sparsely connected network. We find that SR appears due to nonlinear dynamics in a region near the critical point of a dynamical phase transition to network oscillations. The critical point is actually an emergent threshold in the collective dynamics. Using the cortical model, we mimic experiments of Gluckman *et al.* [B. J. Gluckman *et al.*, PRL **77**, 4098 (1996)] that observed stochastic resonance in a response of CA1 networks from mammalian brain on periodic electric stimuli. Results of our numerical calculations are in agreement both qualitatively and quantitatively with these experiments.

Keywords: neural networks; stochastic resonance; signal recognition; brain rhythms.
PACS: 87.19.ll, 87.10.Mn, 87.19.lt

INTRODUCTION

The nervous-system is noisy [1]. How noise influences brain function is an open question in neuroscience. Here, we are interested in studying how sensorial systems process information regardless of noise. In fact, there are two different possibilities [1], the system either tries to compensate somehow noise, or it can use noise for its own benefit by means of a phenomenon called stochastic resonance (SR). This phenomenon allows nonlinear systems to recognize weak signals using noise. The fact that the signal-to-noise ratio has a maximum at a non-zero level of noise is a fingerprint of SR [2]. Stochastic resonance has been observed in biological sensory systems and it may explain the sensitivity of some animals to weak signals in noisy environments [3, 4, 5, 6, 7].

Sensory neurons operate as nonlinear threshold systems in the presence of noise. Several models based on nonlinear dynamics of single neurons were proposed to explain SR in the brain. However, these models neglect interactions between neurons. In the present paper, in contrast to these models, we consider a cortical circuit model composed by stochastic excitatory and inhibitory neurons which interact with each other and form a sparsely connected network [8]. We are motivated by the fact that SR can occur at the network level [9]. In particular, Gluckman *et al.* [10] observed SR in a response of CA1 networks from mammalian brain on periodic electric stimuli. We model these experiments using the cortical model and we show that SR is actually an emergent property of neural networks.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 202-206 (2013); doi: 10.1063/1.4776520

CORTICAL MODEL OF NEURAL NETWORKS

We consider a simple cortical circuit model introduced in [8]. The circuit is composed by N neurons connected by directed edges (synapses) at random (directed Erdős-Rényi network). The probability that two randomly chosen neurons are connected is c/N where c is the mean number of synaptic connections. There are two populations of neurons, g_eN excitatory and g_iN inhibitory neurons ($g_e + g_i = 1$). Neurons are active if they fire bursts of spikes and inactive if they do not. For simplicity, we consider that active neurons fire with a constant firing rate ν. A neuron receives signals from its presynaptic active neurons (a positive contribution from excitatory neurons and a negative contribution from inhibitory neurons) and random spikes. These random spikes represent synaptic noise and random spikes arriving from other areas of the brain. Within a certain time window τ, a neuron n integrates all inputs and compares the total input V_n with a threshold Ω. If $V_n \geq \Omega$ then inactive neuron n is activated at a rate μ_{a1} ($a = e$ for excitatory and $a = i$ for inhibitory neurons). Similarly, active neuron n is inactivated at a rate μ_{a2} if $V_n < \Omega$. For simplicity, we consider the case $\mu_{a1} = \mu_{a2} = \mu_a$. Note that μ_e^{-1} and μ_i^{-1} are the first spike latencies of excitatory and inhibitory neurons, respectively.

In order to study collective dynamics of neural networks, we introduce the quantities $\rho_e(t)$ and $\rho_i(t)$ as the fractions of active excitatory and inhibitory neurons at time t. We call them "activities". Using the stochastic rules for single neurons formulated above, we derived rate equations for the activities $\rho_e(t)$ and $\rho_i(t)$,

$$\frac{d\rho_a(t)}{\mu_a dt} = F(t)(1-\rho_a(t)) - \rho_a(t) + \Psi(\rho_e(t) + A_e(t), \rho_i(t)) \tag{1}$$

where $a = e, i$ and

$$\Psi(\rho_e + A_e, \rho_i) = \sum_{\xi,k,l \geq 0} \Theta(J_e k + J_i l + \xi - \Omega) G(\xi) P_k(c g_e \nu \tau [\rho_e + A_e]) P_l(c g_i \nu \tau \rho_i). \tag{2}$$

$F(t)$ is a stochastic force acting at the network level and producing stochastic fluctuations of neural activity. We define the sensory input $A_e(t) = x(t) g_s / g_e$ where g_s is a fraction of the excitatory neural population stimulated by a stimulus $x(t)$. The external stimulation $x(t)$ can come from sensory neurons or from an applied electric field. $\Psi(\rho_e, \rho_i)$ is the probability that a randomly chosen neuron has a total input at least the threshold Ω. $\Theta(x)$ is the Heaviside step function. ξ is the random number of spikes due to noise. ξ follows the Gaussian distribution $G(\xi)$. $P_k(g_e \rho_e c \nu \tau)$ and $P_l(g_i \rho_i c \nu \tau)$ are the probabilities that during a time τ a neuron receives k spikes from active excitatory neurons and l spikes from active inhibitory neurons. $P_n(c)$ is the Poisson distribution function. J_e and J_i are the efficacies (synaptic strengths) of connections with presynaptic excitatory and inhibitory neurons, respectively, see [8] for details. We believe that Eq. (1) and Eq. (2) are exact in the thermodynamic limit, $N \to \infty$. In our numerical solution of Eq. (1), we used the following parameters: the mean degree $c = 1000$; the threshold $\Omega = 30$; the frequency $\nu = 100$ Hz; the integration time $\tau = 10$ ms; the excitatory and inhibitory efficacies $J_e = 1$, $J_i = -3$, respectively; the fraction of excitatory neurons $g_e = 0.75$; the fraction of excitatory neurons which receive a stimulus $g_s = 0.1$;

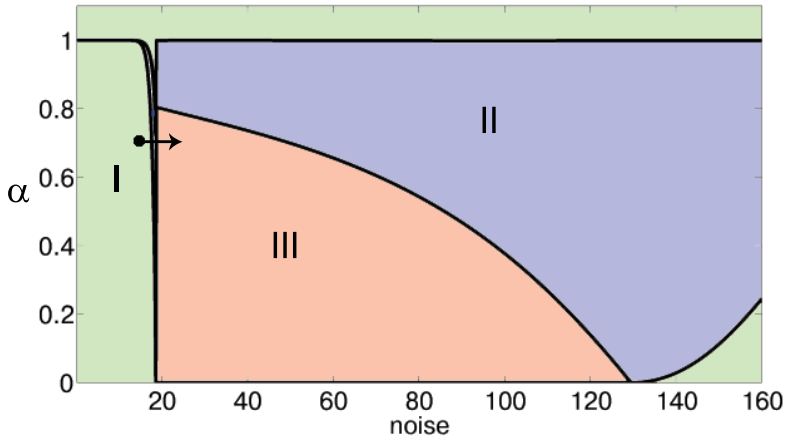


FIGURE 1. Phase diagram of the cortical circuit model. There are three regions of collective neural activity: (I) exponential relaxation to a steady state; (II) decaying oscillations; (III) network oscillations. The point shows the parameters used in our numerical calculations for SR: $\alpha = 0.7$ and $\langle \xi \rangle = 15$. The arrow represents the effect of a stimulus $x(t)$ that moves the system from region (I) to region (III).

the first spike latency of excitatory neurons $\mu_e^{-1} = 20$ ms; and the variance of the random number of spikes due to noise $\sigma_\xi^2 = 10$.

Solving the rate equations, Eq. (1), at $d\rho_a/dt = 0$, we determined the stationary states [8]. Then, we studied how the activities relax to a steady state at different levels of noise (the noise level is characterized by the mean number of random spikes, $\langle \xi \rangle$, that a neuron receives during the time window τ) and at different values of the ratio $\mu_i/\mu_e \equiv \alpha$. Fig. 1 shows the phase diagram of the model in the plane $\alpha - \langle \xi \rangle$. There are three different regions of relaxation: (I) a region of exponential relaxation to a steady state; (II) a region of decaying oscillations; (III) a region of sustained network oscillations. The emergence of network oscillations occurs as a dynamical phase transition at a critical boundary that separates region (I) from region (III).

STOCHASTIC RESONANCE

Let us consider the case when our neural network stays in region (I) (a region of asynchronous dynamics and weak neural activity), on the left of region (III) (a region of sustained network oscillations) (see the point in Fig. 1). In order to move neural dynamics from region (I) to region (III), one can either increase noise level $\langle \xi \rangle$ or alternatively apply a stimulus $x(t)$ to the neural network (see the arrow in Fig. 1). We would like to outline that the critical boundary is an emergent threshold in the collective dynamics. In our model, it is due to this emergent property that it is possible to find SR.

Now we discuss the experiments in Ref. [10]. Gluckman *et al.* prepared hippocampal slices from the rat temporal lobe. Then, they applied a time varying electric field to deliver both a sinusoidal signal and noise directly to the neuronal network. They demonstrated stochastic resonance in the response of the neuronal network finding an optimal

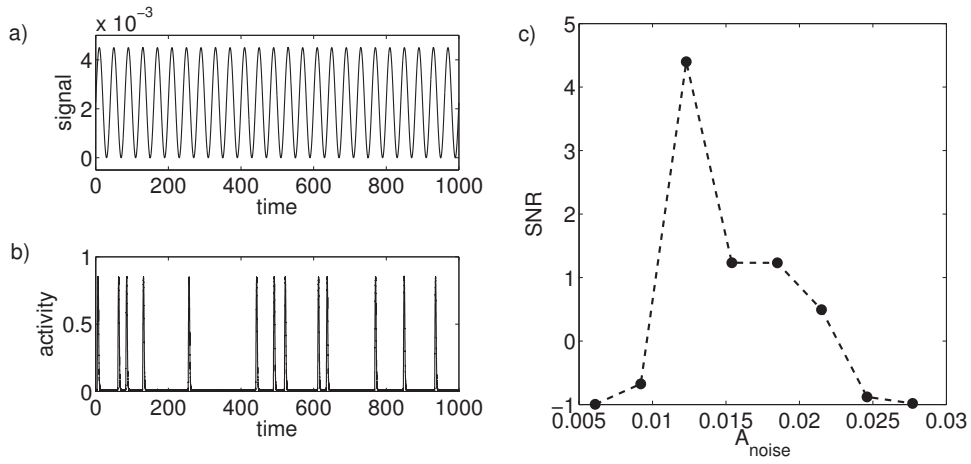


FIGURE 2. (a) Signal applied to the neural network. (b) Response of the neural network to stimulus composed by the signal (a) and noise with the mean amplitude $A_{noise} = 1.23 \times 10^{-2}$. (c) Signal-to-noise ratio (SNR) as a function of A_{noise} for a neural network. A maximum of the SNR at a non-zero level of noise is a fingerprint of stochastic resonance. Time is in the units μ_e^{-1}. Parameters: $\alpha = 0.7$ and $\langle \xi \rangle = 15$.

non-zero magnitude of the stochastic component for which the signal-to-noise ratio had a maximum. In order to understand the mechanism of the observed SR, we mimic the conditions present in these experiments.

First, we placed our neural network in region (I) (see the point in Fig. 1). For simplicity, we set $F(t) = 0$ in Eq. (1) and approximated the expected stochastic fluctuations due to the stochastic force by adding white noise with standard uniform distribution on the open interval $[0, 10^{-2}]$ to the activities $\rho_a(t)$.

Second, we applied a stimulus which was the sum of a weak sinusoidal signal and noise to the neural network. In Eq. (1), we assumed that the sensory input $A_e(t)$ was proportional to the stimulus $x(t)$ where $x(t) = x[\sin(2\pi ft) + 1]/2 + \zeta(t)$ was the sum of a weak signal and noise, respectively. We used $x = 4.5 \times 10^{-3}$ and $f = 1.25$ Hz, see Fig. 2 (a) (like in Ref. [10] we chose the stimulus frequency smaller than the frequency of network oscillations which was about 4 Hz). The noise input $\zeta(t)$ was generated by a Gaussian process characterized by the mean amplitude A_{noise} and the standard deviation $\sigma_n = 1.5A_{noise}$. We chose a sufficiently weak signal, i.e., the signal that could not move the network through the critical boundary, from region (I) to region (III).

Third, we calculated the response of the neural network to the stimulus using Eq. (1) for different mean amplitudes of the input noise A_{noise}, see Fig. 2 (b). We chose the amplitude A_{noise} such that the ratio x/A_{noise} was close to that used by Gluckman *et al.*.

Finally, we calculated the signal-to-noise ratio (SNR), using the same method as Gluckman *et al.*: (i) we calculated the response of activities of the neural network to the stimulus for each amplitude of noise A_{noise}; (ii) we obtained the power spectrum of each response from the respective Fourier transform; (iii) in the power spectrum, we measured the amplitude A of the peak at the frequency of the input sinusoidal signal, and the background averaged amplitude B of the spectrum; (iv) we calculated the SNR$= (A-B)/B$.

Our results are represented in Fig. 2 (c). In this figure we plot the signal-to-noise ratio versus the noise amplitude A_{noise} applied to the neural network. The figure shows that there is an optimal level of noise at which the SNR is maximum. This maximum of the signal-to-noise ratio at a non-zero level of noise is a fingerprint of stochastic resonance. Interestingly, the maximum of the SNR takes place at an amplitude A_{noise} such that the ratio x/A_{noise} is close to the ratio found in Ref. [10]. Thus, the proposed model reproduced well the observed SR and our results are both qualitatively and quantitatively in agreement with the experiments of Gluckman *et al.*.

CONCLUSION

In the present paper, using a cortical circuit model, we have investigated the mechanism of stochastic resonance in neural networks. We have studied response of neural networks on periodic stimulus plus noise. We have showed that, due to the interaction between neurons and nonlinear dynamics near the critical point of a dynamical phase transition to network oscillations, the model demonstrates a resonance phenomena similar to the SR phenomenon observed in neural networks from mammalian brain [10]. The critical point is actually an emergent threshold in the collective dynamics. Results of our numerical calculations agree both qualitatively and quantitatively with the experiments of Gluckman *et al.* [10].

ACKNOWLEDGMENTS

This work was partially supported by the following projects PTDC: SAU-NEU/ 103904/2008, FIS/108476/2008, MAT/114515/2009, and PEst-C/CTM/LA0025/2011. K. E. Lee was supported by FCT under grant SFRH/BPD/ 71883/2010, M. A. L. was supported by FCT under Grant No. SFRH/BD/ 68743/2010.

REFERENCES

1. A. Faisal, L. Selen, and D. M. Wolpert, *Nat. Rev. Neurosci.* **9**, 292–303 (2008).
2. L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, *Rev. Mod. Phys.* **70**, 223–287 (1998).
3. J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss, *Nature* **365**, 337 (1993).
4. K. Wiesenfeld, and F. Moss, *Nature* **373**, 33 (1995).
5. J. E. Levin, and J. P. Miller, *Nature* **380**, 165–168 (1996).
6. D. F. Russell, L. A. Wilkens, and F. Moss, *Nature* **402**, 291–294 (1999).
7. F. Moss, L. M. Ward, and W. G. Sannita, *Clin. Neurophysiol.* **115**, 267 – 281 (2004).
8. A. V. Goltsev, F. V. de Abreu, S. N. Dorogovtsev, and J. F. F. Mendes, *Phys. Rev. E* **81**, 061921 (2010).
9. M. D. McDonnell, and L. M. Ward, *Nat. Rev. Neurosci.* **12**, 415–426 (2011).
10. B. J. Gluckman, T. I. Netoff, E. J. Neel, W. L. Ditto, M. L. Spano, and S. J. Schiff, *Phys. Rev. Lett.* **77**, 4098–4101 (1996).

Spike-time reliability of layered neural oscillator networks

K. K. Lin*, E. Shea-Brown† and L.-S. Young**

*Department of Mathematics and Program in Applied Mathematics, the University of Arizona, USA.
†Department of Applied Mathematics, the University of Washington, USA.
**Courant Institute of Mathematical Sciences, New York University, USA.

Abstract. If a network of neurons is repeatedly driven by the same fluctuating signal, will it give the same response each time? If so, the network is said to be *reliable*. Reliability is of interest in computational neuroscience because the degree to which a network is reliable constrains its ability to encode information in precise temporal patterns of spikes. This note outlines how the question of reliability may be fruitfully formulated and studied within the framework of random dynamical systems theory. A specific network architecture, that of a single-layer network, is examined. For the type of single-neuron dynamics and coupling considered here, single-layer networks are found to be very reliable. A qualitative explanation is proposed for this phenomenon.

Keywords: reliability; spike-time precision; coupled oscillators; random dynamical systems; neuronal networks; Lyapunov exponents.
PACS: 87.19.lj, 87.19.lm, 87.19.ls, 05.45.Jn, 05.45.Xt

Introduction. If we repeatedly drive a network of neurons with the same complicated signal, will the network's response be the same each time? A network for which the answer is "yes" is said to be *reliable*. This property is of interest in computational neuroscience because neurons communicate information via brief electrical impulses, or *spikes*, and how reliable a system is may affect its ability to encode information via precise temporal patterns of spikes.

A variety of experimental and theoretical studies have found that single neurons are reliable under a broad range of conditions (see, e.g., [3, 11, 4]); much less is known about the reliability of networks, both experimentally and theoretically. The work described in this brief note is part of a program aimed at understanding the types of networks and network conditions (e.g., network architecture, type of stimulus, etc.) that may enhance or disrupt the reliability of a network. This work combines qualitative theoretical ideas from dynamical systems theory and numerical simulations.

The results mentioned here are summarized in [10] and explained in detail in [9]; interested readers are referred to those papers. In addition to layered networks, we have also carried out a general study of reliability of feed-forward and recurrent networks. This work is described in [8], and some of the main results are surveyed in [7].

Models. Our networks are composed of so-called "theta neuron" models. These are idealized models for neurons that, in the absence of external forcing, spike periodically

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 207-209 (2013); doi: 10.1063/1.4776521

at a fixed frequency. The network equations take the form

$$\dot{\theta}_i = \omega_i + z(\theta_i)\left(\sum_{j\neq i} a_{ji}\, g(\theta_j) + \varepsilon I(t)\right), \quad i = 1,2,\cdots,N, \tag{1}$$

where the state of the ith neuron is given by an angle θ_i; $\omega_i > 0$ is the intrinsic frequency of the ith neuron; $I(t)$ represents an external stimulus, here modeled as white noise; g is an approximate delta function, i.e., it is a smooth function supported in a small interval around 0 satisfying $\int_0^{2\pi} g(\theta)\, d\theta = 1$ (this is a simple model for relatively fast synapses); and z is the *phase response curve* (PRC), modeling state-dependent neuronal response to inputs. The angle θ_i represents the fraction of its cycle that a neuron has completed; the neuron is viewed as spiking at $\theta = 0$. A nonzero input $I(t)$ modulates the firing rate of the neuron, and the phase response z captures the state-dependent response of the neuron to stimuli. Here, we assume $z(\theta) = 1 - \cos(\theta)$, a so-called "Type I" PRC [2].

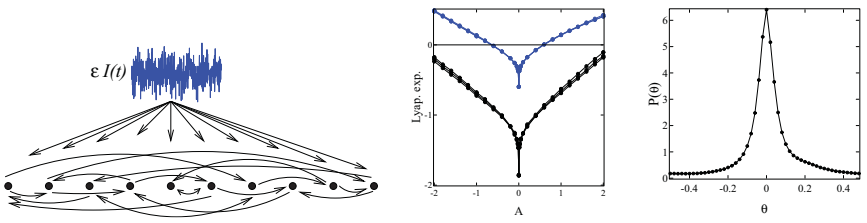
In Eq. (1), the coupling matrix $A = (a_{ji})$ encodes the network structure. In this note, we consider only *single-layer networks,* i.e., the neurons are sparsely and randomly coupled to each other, and receive a single common input (Fig. 1(a)); see [10] for details.

Reliability formulation and a mathematical framework. In this context, we can define reliability to mean that the network state $\Theta(t) = (\theta_1(t), \cdots, \theta_N(t))$ is reproducible across repeated trials, where each new "trial" means generating a new set of initial conditions (but keeping all else — network structure, stimulus, oscillator frequencies — fixed). Note that this notion of reliability is not the only one relevant to applications; some other types of reliability are discussed in [9].

A mathematically equivalent way to view reliability is in terms of how the flow defined by Eq. (1) acts on an *ensemble* of initial conditions. That is, let ω denote a specific realization of the stimulus, and let $\Phi_\omega^{s,t}$ be the solution map of Eq. (1), so that for any solution Θ we have $\Phi_\omega^{s,t}(\Theta(s)) = \Theta(t)$. (For properties of such flow maps, see [1].) Then reliability is equivalent to the statement that an ensemble of initial conditions, when transported in time by the flow $\Phi_\omega^{0,t}$, collapses to a single, distinguished trajectory.

Because the mappings $\Phi_\omega^{s,t}$ depend on the white noise stimulus, it defines a *random dynamical system.* A useful tool that enables us to study reliability efficiently is the *maximum Lyapunov exponent* $\lambda_{\max}$ of a system. Roughly speaking, the maximum Lyapunov exponent measures the rate of separation of nearby trajectories in state space: $\lambda_{\max} > 0$ means nearby trajectories diverge exponentially fast, whereas $\lambda_{\max} < 0$ mean exponential convergence. The relation of $\lambda_{\max}$ to reliability is encapsulated in two theorems: the first, due to Le Jan [6] (with extensions by Baxendale [1]), states that subject to certain non-degeneracy conditions, if $\lambda_{\max} < 0$, then the ensemble will collapse to a single trajectory. The second theorem, due to Ledrappier and Young [5], states that if $\lambda_{\max} > 0$, then the ensemble does not collapse to a single trajectory, but rather converges to a "random strange attractor." The latter are complicated geometric objects that wind around phase space in a complicated way, and for our systems are usually not localized in phase space. Taken together, these theorems tell us that $\lambda_{\max}$ is a useful indicator of reliability.

Single layer networks. Fig. 1(b) shows $\lambda_{\max}$ for a single-layer network, plotted against a quantity A measuring the total synaptic input received by each neuron (precise definitions are in [10, 9]). Observe first that when $A = 0$, which corresponds to a "network"



(a) Network schematic (b) Lyap. exp. (c) Phase distribution

FIGURE 1. Single-layer network. Panel (a) shows an example. In (b), we plot $\lambda_{\max}$ for vs. mean total input A for 2 stimulus amplitudes ε. Panel (c) shows the phase distribution of neurons at pulse arrival.

where the oscillators are uncoupled, we have $\lambda_{\max} < 0$. This is because single theta neurons can be shown to be always reliable. As $|A|$ increases, $\lambda_{\max}$ increases, suggesting that the coupling is destabilizing. But even when $A = 2$ (which is quite strong [10]), the system remains reliable. Why is this the case? The following explanation is proposed in [9]: if all neurons have the same frequency and there are no couplings, then the neurons would be synchronized by the common input. But if we now perturb the frequencies and couplings away from 0, then we would expect the neurons to remain *nearly* synchronized. But the phase response $z(\theta) = O(\theta^2)$ for $\theta \sim 0$, so when the neurons are spiking, they are not listening to their inputs. Thus, the couplings are effectively attenuated.

Among other things, this proposed mechanism predicts that the neurons would have phases that are highly clustered, e.g., if we look at the distribution of phases conditioned on the arrival of an incoming pulse, the distribution should be highly peaked. This is indeed the case (Fig. 1(c)). Also, any factor that makes it harder for the neurons to synchronize would be detrimental to reliability; this has also been shown numerically.

REFERENCES

1. P. H. Baxendale, "Statistical equilibrium and two-point motion for a stochastic flow of diffeomorphisms", in *Spatial stochastic processes, Progr. Probab.* **19**, *pp.* 189–218, Birkhäuser, Boston, 1991.
2. E. Brown, P. Holmes, and J. Moehlis. Globally coupled oscillator networks. In *Problems and Perspectives in Nonlinear Science: A celebratory volume in honor of Lawrence Sirovich*, E. Kaplan, J.E. Marsden, and K.R. Sreenivasan, editors, pages 183–215. Springer, New York, 2003.
3. H. L. Bryant, and J. P. Segundo. *J. Physiol.* **260**, 279–314 (1976).
4. J. Hunter, J. Milton, P. Thomas, and J. Cowan, *J. Neurophysiol.* **80**, 1427–1438 (1998).
5. F. Ledrappier, and L.-S. Young, *Probab. Th. and Rel. Fields* **80**, 217–240 (1988).
6. Y. Le Jan, *Ann. Inst. H. Poincaré Probab. Statist.* **23**, 111–120 (1987).
7. K. K. Lin, "Stimulus-response reliability of biological networks," to appear in *Nonautonomous and Random Dynamical Systems in the Life Sciences,* P. Kloeden and C. Poetzsche, eds. (2012).
8. K. K. Lin, E. Shea-Brown, and L.-S. Young, *J. Nonlin. Sci.* **19**, 497–545 (2009).
9. K. K. Lin, E. Shea-Brown, and L.-S. Young, *J. Comp. Neurosci.* **27**, 135–160 (2009).
10. K. K. Lin, E. Shea-Brown, and L.-S. Young, *Commun. Math. Sci.* **7**, 239–247 (2009).
11. Z. Mainen and T. Sejnowski, *Science* **268**(5216), 1503–1506 (1995).

Anticipated synchronization in neuronal network motifs

F. S. Matias[*,†], L. L. Gollo[†], P. V. Carelli[*], M. Copelli[*] and C. R. Mirasso[†]

[*]*Departamento de Física, Universidade Federal de Pernambuco, Recife, PE 50670-901 Brazil.*
[†]*Instituto de Fisica Interdisciplinar y Sistemas Complejos, CSIC-UIB, Campus Universitat de les Illes Balears E-07122 Palma de Mallorca, Spain.*

Abstract. Two identical dynamical systems coupled unidirectionally (in a so called master-slave configuration) exhibit anticipated synchronization (AS) if the one which receives the coupling (the slave) also receives a negative delayed self-feedback. In oscillatory neuronal systems AS is characterized by a phase-locking with negative time delay τ between the spikes of the master and of the slave (slave fires before the master), while in the usual delayed synchronization (DS) regime τ is positive (slave fires after the master). A 3-neuron motif in which the slave self-feedback is replaced by a feedback loop mediated by an interneuron can exhibits both AS and DS regimes. Here we show that AS is robust in the presence of noise in a 3 Hodgkin-Huxley type neuronal motif. We also show that AS is stable for large values of τ in a chain of connected slaves-interneurons.

Keywords: Hodgkin-Huxley; synchronization; phase-locking.
PACS: 87.19.lj, 87.19.ll, 87.19.lm

WHAT IS ANTICIPATED SYNCHRONIZATION?

Synchronization of oscillators was initially studied by Huygens with two pendulum clocks. Since then an extensive study of synchronization of nonlinear systems has been done on a variety of physical and biological systems [1]. Here we are interested in a new scheme of synchronization called "anticipated synchronization" (AS) [2]. It consists of the stable state reached by two identical dynamical systems coupled in a master-slave (MS) configuration if the slave is also subjected to a negative delayed self-feedback. Such systems are described by the following set of equations:

$$\begin{aligned} \dot{M} &= f(M(t)), \\ \dot{S} &= f(S(t)) + K[M(t) - S(t - t_d)]. \end{aligned} \tag{1}$$

The solution $S(t) = M(t + t_d)$, which characterizes AS, means that at time t the slave (S) is in the same state as the master (M) will be in a future time $t + t_d$.

3-NEURON MOTIF: MASTER-SLAVE-INTERNEURON

The first verification of AS in a neuronal model was done in a system of two FitzHugh-Nagumo models, coupled as MS, forced by the same random external current and subjected to a recurrent negative delayed connection in the slave neuron [3]. However this recurrent feedback has no obvious biological correlate. In a previous work, we have

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 210-212 (2013); doi: 10.1063/1.4776534

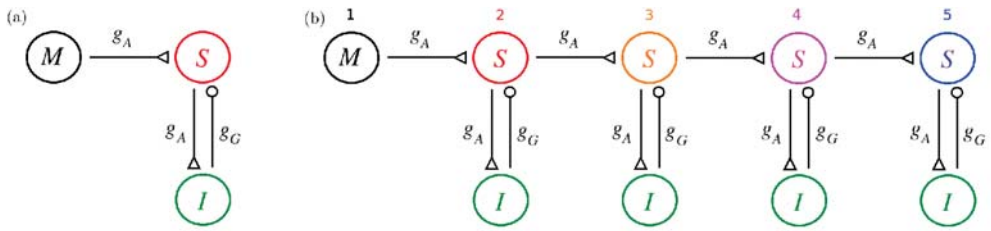


FIGURE 1. Neuronal network motifs that exhibit anticipated synchronization. (a) 3 neurons in a master-slave-interneuon (MSI) configuration. (b) Chain of slave-interneurons.

proposed a more biophysically plausible system in which the delayed self-feedback is replaced by an inhibitory feedback loop mediated by a third neuron [4]. It consists of the master-slave-interneuron (MSI) motif shown in Fig. 1(a) which is composed of 3 neurons coupled by two excitatory synapses (from M to S and from S to I) and one inhibitory synapse from I to S.

In Fig. 1 each neuron is a Hodgkin-Huxley (HH) type model [5]. It consists of four coupled ordinary differential equations associated to the membrane potential V and the ionic currents flowing across the axonal membrane corresponding to the Na, K and leakage currents. Each link is an excitatory (or inhibitory) synapse mediate by AMPA (GABA_A) with synaptic conductance g_A (g_G). Each synaptic current is also described by a dynamical equation [4].

It has been shown that 3 HH neuron models coupled as MSI and driven by either constant or periodical external current exhibits both DS and AS regimes for a large set of parameters [4]. Since DS and AS are phase-locking regimes (master and slave fire with the same period), it is possible to define the time delay τ as the difference between the spike timing of the master and the slave in each cycle: $\tau = t^S - t^M$. τ turns out to be a nonlinear function of g_G. In the absence of the recurrent inhibitory feedback ($g_G = 0$) the slave fires a little later than the master: $\tau > 0$, which characterizes the DS regime. As we increase the value of g_G, τ decreases: master and slave fire spikes closer and closer until they reach a regime of complete (or zero-lag) synchronization. Increase g_G even more makes the slave fire spikes before the master: $\tau < 0$, which characterizes the AS regime. The transition from AS to DS is smooth and continuous. For larger values of g_G the system reaches a phase drift regime (master and slave fire with different periods).

An MSI motif of modified HH model neurons including a slow K+ current [6] and colored noise added to the constant external current also exhibits both DS and AS regimes depending on the strengths g_G of the inhibitory synapse (see Fig. 2(a)-(d)). This is a more realistic model for cells in the cortex. Due to the noise, spiking is not periodic, as can be seen in Fig. 2(a). τ varies in each cycle but maintains a well defined sign, as shown in Fig. 2(b). The mean value of τ is a well behaved function of g_G.

LARGER MOTIFS: CHAIN NETWORKS

It has been show that a chain network can propagate stable activity with temporal precision in songbirds [7], and that a chain of chaotic slave neural network can exhibit

AS [8]. As the brain exhibits well defined sequences of neuronal processes during complex behaviors, such as cognitive tasks, we wondered if a chain of slave-interneurons (see Fig. 1(b)) can exhibit AS and control the temporal precision between spikes of different neurons. Such a chain of connected standard HH neurons driven by a constant current can provide a mechanism for obtaining larger anticipation time between the first master and the last slave as shown in Fig. 2(e) and (f). Furthermore, the chain network motif has precise time differences among the spikes that depend on the synaptic conductances and the external current.

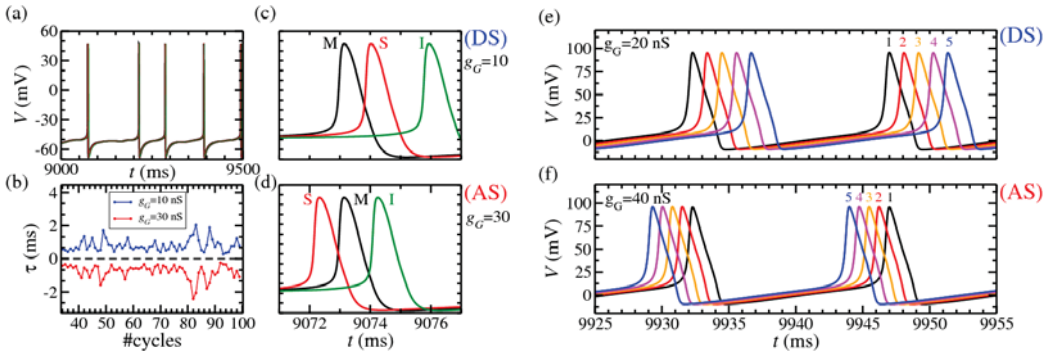


FIGURE 2. Characterizing DS and AS regimes. (a)-(d) Results for the 3-neuron motif. (a) With colored noise added to the input current, spiking is not periodic. (b) Under noisy dynamics, the time delay τ in each cycle fluctuates around a mean with well-defined sign (for the DS and AS regimes) (c)-(d) g_G controls the relative timing of the master and slave spikes, leading to DS or AS. The chain in Fig. 1(b) exhibits (e) DS for weaker inhibition and (f) AS for stronger inhibition. Note that the anticipation time is largest between neurons 1 and 5.

ACKNOWLEDGMENTS

We thank CNPq, FACEPE, CAPES and special programs PRONEX, INCeMaq and PRONEM, the Ministerio de Educación y Ciencia (Spain) and Fondo Europeo de Desarrollo Regional (FEDER) under project FIS2007-60327 (FISICOS) for financial support.

REFERENCES

1. A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge, UK, 2001.
2. H. U. Voss, *Phys. Rev. E* **61**, 5115 (2000).
3. R. Toral, C. Masoller, C.R. Mirasso, M. Ciszak, and O. Calvo, *Physica A* **325**, 192 (2003).
4. F. S. Matias, P. V. Carelli, C. R. Mirasso, and M. Copelli, *Phys. Rev. E* **84**, 021922 (2011).
5. A. L. Hodgkin, and A. F. Huxley, *J. Physiol.* **117**, 500 (1952).
6. M. Pospischil, *et al.*, *Biol. Cybern.* **99**, 427 (2008).
7. M. A. Long, D. Z. Jin, and M. S. Fee, *Nature* **468**, 394 (2010).
8. M. Ciszak, *et al.*, *Phys. Rev. E* **72**, 046218 (2005).

Spike timing analysis in neural networks with unsupervised synaptic plasticity

B. E. P. Mizusaki, E. J. Agnes, L. G. Brunnet and R. Erichsen Jr.

Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.

Abstract. The synaptic plasticity rules that sculpt a neural network architecture are key elements to understand cortical processing, as they may explain the emergence of stable, functional activity, while avoiding runaway excitation. For an associative memory framework, they should be built in a way as to enable the network to reproduce a robust spatio-temporal trajectory in response to an external stimulus. Still, how these rules may be implemented in recurrent networks and the way they relate to their capacity of pattern recognition remains unclear. We studied the effects of three phenomenological unsupervised rules in sparsely connected recurrent networks for associative memory: spike-timing-dependent-plasticity, short-term-plasticity and an homeostatic scaling. The system stability is monitored during the learning process of the network, as the mean firing rate converges to a value determined by the homeostatic scaling. Afterwards, it is possible to measure the recovery efficiency of the activity following each initial stimulus. This is evaluated by a measure of the correlation between spike fire timings, and we analysed the full memory separation capacity and limitations of this system.

Keywords: unsupervised synaptic plasticity; spiking neurons; homeostasis; STDP; STP.
PACS: 87.18.Sn, 87.19.lg, 87.19.lj, 87.19.lp, 87.19.lv, 87.19.lw

INTRODUCTION

The ever-going changes in synaptic connections are thought to play a major role in a neural network's memory capacity. In this work we investigate how experimentally derived plasticity rules may affect their dynamics within a recurrent architecture, and whether they might give rise to some kind of associative memory. More specifically, we focused on homeostatic scaling[1, 2, 3] with pre-synaptic dependency[4] and spike-timing-dependent-plasticity[5, 6, 7] and their relation to the recovery of patterns of precise timing of spikes, as observed for example in some layers of the visual cortex[8].

SIMULATION DETAILS

We used two models for neural dynamics: an analytically solvable integrate-and-fire with Dirac's delta interactions and an Izhikevich model for pyramidal neurons[9] with conductance-based synapses, and added short-term-plasticity (as in [9]) in both for stability. The system consists of $N = 500$ neurons, being 80% excitatory and 20% inhibitory, randomly connected with probabilities taken from the literature [10].

The simulation starts with a quiescent network to which is imposed a pattern of neuronal spikes. The synaptic weights are updated after the transient activity ceases, in a trial scheme (such as [4]). As the synaptic weights evolve, the network starts to

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 213-215 (2013); doi: 10.1063/1.4776522

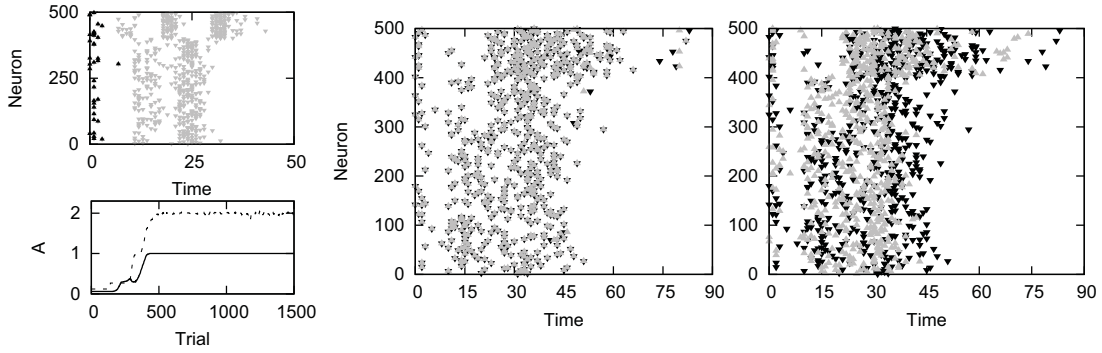


FIGURE 1. On the left: development of activity during learning time on the bottom (full line is the mean A for excitatory neurons, dotted is for inhibitories), and on top, raster plot of the untrained input (black) and subsequent response after learning (grey). On the centre and on the right, juxtaposition of raster plots from two different trials (black and grey symbols): same learned input and different learned inputs respectively. Note that neurons 1-400 are excitatories and 401-500 are inhibitories.

develop activity following the input (Fig. 1), and it converges to a mean value A_{Goal} defined by the homeostatic scaling rule,

$$W_{ij}^{\nu+1} = W_{ij}^{\nu} + \alpha_W A_j^{\nu}(A_{Goal} - A_i^{\nu})W_{ij}^{\nu}$$

where W_{ij}^{ν} is the synaptic weight from the neuron j to the neuron i and A_i^{ν} is an accumulated mean value of the number of spikes S_i^{ν} of the neuron i in the trial ν:

$$A_i^{\nu+1} = A_i^{\nu} + \alpha_A(S_i^{\nu} - A_i^{\nu}) \quad .$$

A_j is the pre-synaptic dependence factor, used as a mechanism to further stabilize the activity. A_{Goal} was set to 1 spike per trial for excitatories and 2 for inhibitories.

Correlation measure

The correlation measured between the activities of two trials

$$C^{\nu\nu'} = \frac{1}{\gamma}\sum_i \sum_{k,l=1}^{S_i^{\nu}, S_i^{\nu'}} e^{-\left(\frac{t_{i,k}^{\nu} - t_{i,l}^{\nu'}}{2\sigma^2}\right)^2}$$

$$\gamma = \left[MAX(S^{\nu}, S^{\nu'}, N)\right]$$

compares the k^{th} spike time of the neuron i in the trial ν with the l^{th} time, the closest one, in the trial ν', with a tolerance of $\sigma = 1$ms. $C^{\nu\nu'}$ is normalized by γ, the largest value within either the number of spikes of one of the trials, or N. This is to take into account that the activity should obey the homeostasis constraint of 1 spike for each excitatory neuron in one trial. Fig. 2 shows the simulation results up to 15 trained patterns. It is possible to see that it recovers the activity if a trained pattern is presented again ($C^{\nu\nu'} \approx 1$), and that it is different from the other trained ones ($C^{\nu\nu'} \approx 0$).

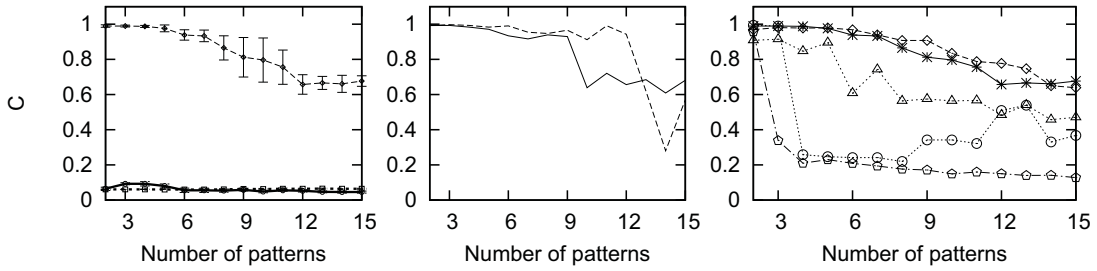


FIGURE 2. On the left, averages for correlation between the same input (dashed line), two different trained inputs (full) and a trained and an untrained one (dotted). On the centre, correlation measures for two presentations of the same pattern with (full line) and without (dashed) STDP. On the right, averages for correlation with the same pattern with varying network connectivity values. The stars are the same of the other graphs, excitatory→excitatory: 0.12, excitatory→inhibitory: 0.2 and inhibitory→excitatory: 0.2. For the others, connectivities are scaled by: 0.75 (diamonds), 0.5 (triangles), 1.5 (circles) and 2 (pentagons).

CONCLUSIONS

We searched for associative memory of spike patterns and presented results of timing correlation as a function of increasing trained patterns, which also enhances the network level of recurrence. There were no major differences of activity behaviour or correlation capacities between the two models of synapses used (static charges for integrate-and-fire or conductance-based for Izhikevich model). The pre-synaptic dependence factor A_j introduces a Hebb-like relation to the connections, making the homeostatic scaling sufficient for the recovery of spike timings. The system without STDP actually yields similar activity and correlation results for a small number of learned inputs. We also checked that the network connectivity sparseness heavily influences this capacity, and there is an optimal range for it (Fig. 2).

ACKNOWLEDGMENTS

This work received financial support from Brazilian agencies CNPq and CAPES and used computational resources from the Centro de Física Computacional - IF (CFCIF).

REFERENCES

1. G. G. Turrigiano, and S. B. Nelson, *Nat. Rev. Neurosci.* **5**, 97–107 (2007).
2. G. LeMasson, E. Marder, and L. F. Abbott, *Science* **259**, 1915–1917 (1993).
3. G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. Nelson, *Nature* **391**, 892–896 (1998).
4. J. K. Liu, and D. V. Buonomano, *J. Neurosci.* **29**(42), 13172–13181 (2009).
5. A. Morrisson, M. Diesmann, and W. Gerstner, *Biol. Cyber.* **98**, 459–478 (2008).
6. W. Gerstner, R. Kempter, J. L. Van Hemmen, and H. Wagner, *Nature* **383**, 76–78 (1996).
7. C. Q. Bi, and M. M. Poo, *J. Neurosci.* **18**, 10464–10472 (1998).
8. P. Tiesinga, J.-M. Fellous, and T. J. Sejnowski, *Nat. Rev. Neurosci.* **9**, 97–107 (2008).
9. E. M. Izhikevich, and G. M., Edelman, *Proc. Natl. Acad. Sci. USA* **105**(9), 3593–3598 (2008).
10. S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, *PLoS Biol.* **3**, 507–519 (2005).

Introducing time-varying parameters in the kuramoto model for brain dynamics

S. Petkoski and A. Stefanovska

Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom.

Abstract. The mean field dynamics of the Kuramoto model (KM) with time-dependent (TD) parameters is described, and the response in the adiabatic and non-adiabatic limits is explained. The observed low-frequency, homogeneity-dependent filtering is discussed, together with the possible implications to the modelling of the brain dynamics.

Keywords: Kuramoto model; time-dependent variables; non-autonomicity.
PACS: 05.45.Xt, 87.10.Mn

INTRODUCTION

The KM results from the phase reduction in dealing with large populations of interacting oscillating subsystems [1], and was largely motivated by biological examples [2]. It assumes that in the limit of many single units, the time of the occurrence of the events matters more than the actual magnitude of each of them. In that sense, e.g. for populations of firing neurons, the measured EEG signals represent the mean field of many thousands or millions of oscillators. When they are synchronized, an amplitude peak at the entrainment frequency will be observed.

Although the model itself only represents idealized scenario, its analytical tractability makes it the prevailing approach in tackling a wide variety of important problems. Still, neither the original model, nor most of its extensions (for a review of all generalizations and the problems they address see [3]), have incorporated a fundamental property of living systems – their inherent time-variability.

A recent generalization of the KM [4] introduced explicit consideration of deterministically time-varying parameters. As a result the dynamics of the collective rhythms consists of the external system superimposed on the autonomous one, a characteristic feature of many thermodynamically open systems. In this way, many important characteristics of open systems that stem from their time-dependent parameters can be better described. Namely, experimentally reported results for the anaesthetized brain [5] provide a strong motivation to model it using time-varying couplings [6], whereas variability in neuron firing rates [7] can be also deterministically encompassed with this model.

KURAMOTO MODEL WITH TIME-DEPENDENT PARAMETERS

The original KM consists of phase oscillators running at arbitrary intrinsic frequencies and coupled through the sine of their phase differences. The oscillators' natural frequencies and/or couplings are influenced by identical external forces with constant or

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 216-218 (2013); doi: 10.1063/1.4776523

distributed strengths. The generalization [4] additionally introduces an external, explicitly time-dependent, bounded function $x(t)$, that modulates the frequencies or couplings of the original model. In the most general case, the strengths of the interactions I_i are distributed according to a probability density function $h(I)$, and likewise the distribution $g(\omega)$ of the natural frequencies ω_i. Thus, depending on which parameter is influenced two models emerge

$$A \quad : \quad \dot{\theta}_i = \omega_i + I_i x(t) + K\, r(t) \sin[\psi(t) - \theta_i], \tag{1}$$

$$B \quad : \quad \dot{\theta}_i = \omega_i + [K + I_i x(t)]\, r(t) \sin[\psi(t) - \theta_i]. \tag{2}$$

Here, a TD complex order parameter is introduced

$$z(t) = r(t) e^{i\psi(t)} = \frac{1}{N} \sum_{j=1}^{N} e^{i\theta_j}, \tag{3}$$

where $r(t)$ and $\psi(t)$ are the TD mean-field amplitude and phase respectively.

In the thermodynamic limit $N \to \infty$ the state of the systems (1, 2) is described by a continuous PDF $\rho(\theta, \omega, I, t)$ which gives the proportion of oscillators with phase θ at time t, for fixed ω and I. A recent ansatz by Ott and Antonsen [8] gives a particular solution to the continuity equation for $\rho(\theta, \omega, I, t) = \frac{1}{2\pi}\{1 + \{\sum_{n=1}^{\infty} [\alpha(\omega, I, t)]^n e^{in\theta} + \text{c.c.}\}\}$, as long as $\alpha(\omega, I, t)$ evolves with

$$A \quad : \quad \frac{\partial \alpha}{\partial t} + i[\omega + Ix(t)]\alpha + \frac{K}{2}(z\alpha^2 - z^*) = 0, \tag{4}$$

$$B \quad : \quad \frac{\partial \alpha}{\partial t} + i\omega\alpha + \frac{K + Ix(t)}{2}(z\alpha^2 - z^*) = 0. \tag{5}$$

Using the same ansatz the complex order parameter of the system becomes

$$z^* = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \alpha(\omega, I, t) g(\omega) h(I) d\omega dI. \tag{6}$$

Eqs. (4, 5) hold for any distributions of ω and I, and for any forcing $x(t)$. They describe the evolution of the parameter α which is related to the complex mean field through the integral equation (6). However, for polynomial or multimodal δ-like distributions of ω and I, the integral (6) can be solved, thus leading to low-dimensional evolution of the order parameter $z(t)$. Several different scenarios are shown in Fig. 1.

Slow/Fast Reduction

The plots in Fig. 1 (a)-(b) show that the oscillations of the mean field follow the frequency of the external forcing. Another obvious feature of the same results is the low-frequency filtering of the external fields, i.e. the only difference between plots (a) and (b) is the frequency of the external forcing, whilst its influence is much more pronounced in the latter. In that sense, the magnitude of the mean field's oscillations depends solely on the system's inherent transient time, compared to the period of the external force.

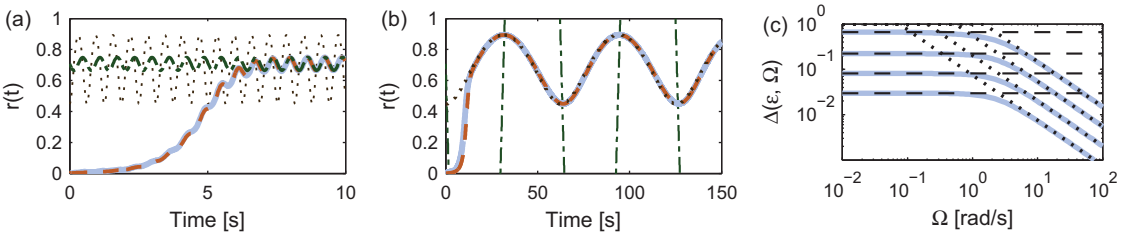


FIGURE 1. (a-b) The time-varying mean field and (c) the magnitude of the response, $\Delta(\varepsilon,\Omega)$, of model A. (a-b) The full system's dynamics (light blue) are in agreement with the low-dimensional (dashed red). Adiabatic (dotted brown), and non-adiabatic evolutions (dashed-dotted green), are confirmed in their limits (see [4] for details). The distribution $h(I) = (\varepsilon\omega)$ with $K = 3.5$, $\varepsilon = 0.6$, $\Omega = 10$ and $\Omega = 0.1$ respectively. (c) Non-adiabatic (dotted black) and adiabatic (dashed black) evolution for $\varepsilon \in \{0.05, 0.0629, 0.0791, 0.0995\}$, compared with the full dynamics (light blue). $K = 4.5$ and $\Omega \in [10^{-2}, 10^{2}]$

CONCLUSION

The system's response to the external forcing depends solely on its heterogeneity. Hence, the mean behaviour of homogeneous ensembles is more easily influenced by changes in the parameters. Similarly, heterogeneous ensembles are more resistant to fast changes. This means that the ensembles behave as low-frequency filters, with their heterogeneity and coherence being the only factors determining whether certain changes are going to be adiabatically mapped in the mean field.

Many possibilities arise for applying these results to the brain dynamics. For example, they could explain how slow-varying signals from the cardio-vascular system [9] could modulate membrane potentials of the populations, leading to modulated spiking activity. Or analogously, the same slow signals would be better captured by the neurons, than the faster signals that would mostly influence more homogeneous and more synchronized neurons.

ACKNOWLEDGMENTS

The work was supported by the EPSRC (UK) and by a Lancaster University PhD grant.

REFERENCES

1. Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*, Springer-Verlag, Berlin, 1984.
2. S. Strogatz, *Sync: The Emerging Science of Spontaneous Order*, Hyperion, New York, 2003.
3. J. A. Acebrón, *et al.*, *Rev. Mod. Phys.* **77**, 137 (2005); A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization – A Universal Concept in Nonlinear Sciences*, CUP, Cambridge, 2001.
4. S. Petkoski, and A. Stefanovska, *Phys. Rev. E* **86**, 046212 (2012).
5. B. Musizza, *et al.*, *J. Physiol.* **580**, 315 (2007).
6. J. H. Sheeba, A. Stefanovska, and P. V. E. McClintock, *Biophys. J.* **95**, 2722 (2008).
7. D. Rudrauf, *et al.*, *Neuroimage* **31**, 209 (2006).
8. E. Ott, and T. M. Antonsen, *Chaos* **18**, 037113 (2008).
9. A. Stefanovska, and M. Bračič, *Contemp. Phys.* **40**, 31 (1999).

Oscillatory dynamics in an attractor neural network with firing rate adaptation

S. Rathore[*,**,†], D. Bush[‡,**], P. Latham[1,§] and N. Burgess[2,‡,**]

[*]*CoMPLEX, UCL, Physics Building, Gower Street, London, WC1E 6BT.*
[**]*UCL Institute of Cognitive Neuroscience, Queen Square, London WC1N 3AR.*
[†]*UCL Institute of Behavioural Neuroscience, 26 Bedford Way, London WC1H 0AP.*
[‡]*UCL Institute of Neurology, Queen Square, London, WC1N 3BG.*
[§]*Gatsby Computational Neuroscience Unit, London WC1N 3AR.*

Abstract. We develop a framework for generating oscillations in ring attractor networks with firing rate adaptation. We show the relationship between the frequency of rotation around the ring of the shifting bump of activity, the adaptation variable and other model parameters using perturbation theory. The analytic solutions are validated against simulations of such networks. Further preliminary findings indicate that the frequency of these networks can be simply controlled using an external stimulus. The mechanism developed here could potentially be used for temporal coding of position through interference of oscillators of different frequencies.

Keywords: oscillations; attractor networks; adaptation.

INTRODUCTION

Oscillations occur ubiquitously in nature and throughout mammalian physiology, from cardiac pacemaker cells to circadian rhythms marking phases of wakefulness and sleep. Oscillations in the brain are thought to be involved in many processes, such as input selection, synaptic plasticity, communication between ensembles of neurons and temporal codes for spatial position [1].

The hippocampal formation of freely moving rodents shows a prominent theta rhythm (4-12Hz) [2], whose frequency is modified by the running speed of the animal [3, 4]. Complementing this temporal periodicity, 'grid cells' in the medial entorhinal cortex display strikingly periodic spatially modulated firing. The oscillatory interference model of grid cell firing [5] posits that the periodic spatial firing pattern of grid cells is formed by the thresholded sum of a baseline oscillation and one or more active oscillations whose frequencies vary from the baseline frequency according to the running speed and direction of the animal. We wanted to investigate the possible mechanisms which could generate such variable frequency oscillations.

More specifically we investigated ring attractor networks as a possible mechanism for generating oscillations, inspired by the work of Zhang [6] who demonstrated that one-dimensional ring attractors could underlie a mental compass where each neuron in the network represents a specific head-direction [7]. The pattern of activity in this type of

[1] E-mail: pel@gatsby.ucl.ac.uk
[2] E-mail: n.burgess@ucl.ac.uk

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 219-223 (2013); doi: 10.1063/1.4776524

attractor network converges in time to a family of possible stable activity patterns. Such networks require recurrent connectivity with a symmetric center-surround pattern of excitation and inhibition, i.e., each neuron locally exciting their near neighbors and distally inhibiting the rest in both directions. This connectivity pattern has rotation invariance: each neuron has the same connectivity regardless of their position in the ring of neurons. Such a network can generate a family of activation patterns corresponding to approximate cosinusodial bumps of activity centered anywhere around the ring. These bumps can be shifted around the ring by introducing an asymmetric component to the connectivity matrix which effectively skews the connectivity pattern of excitatory/inhibitory connections of each neuron either in the clockwise or anticlockwise direction. If a constant asymmetric component is added the bump of activity rotates around the ring therefore causing phase-offset oscillations in each of the neurons in the network, see also [4, 8]. A change in frequency in such a network requires a change in the magnitude of the asymmetric component.

Another possible method of destabilizing and shifting the bump of activity in an attractor network is through firing rate adaptation. In real neurons one often sees that during input of a persistent tonic stimulus the firing rate of the neuron decreases. A potential mechanism for this effect is an increased potassium conductance, from calcium gated potassium channels responding to small increases in intracellular calcium following an action potential , which hyperpolarizes the cell [9], although a separate non-calcium dependent potassium current may also contribute to hyperpolarization [10]. Recently active neurons will therefore fire less than recently inactive neurons, given the same input. This would cause a bump of activity to move, as neurons on the trailing edge of the bump show firing rate adaptation compared to those on the leading edge. Thus a stationary bump will become unstable and a moving bump will speed up to reach a stable oscillation of constant frequency.

We investigated the effect of firing rate adaptation in a ring attractor network, and whether it could generate stable oscillations of activity around the network. Existing work has examined adaptation in attractor network models of hippocampal place cells, to model mental exploration of possible trajectories [11]. A similar use of rate adaptation was adopted by Itskov et al [12] to generate fast drift of the activation pattern to give a read-out of time elapsed during straight trajectories. Work by Melamed et al [13] used combined formulations of synaptic depression (effectively similar to adaptation where an increase in activity is marked by a depletion in synaptic resources) and facilitation (increase in synaptic resources) to generate slow oscillations during 'up-states' in recurrent networks.

Here we aimed to characterize analytically the relationship between the model parameters and the frequency of rotation of the activity bump.

ANALYSIS

We analyzed a standard rate coded attractor network with rate adaptation as described by the coupled pair of differential equations in 1,2.

$$\tau_m \dot{v}_i = \phi(\sum_{j=1}^{N} W_{ij} v_j - a_i + h_i) - v_i \tag{1}$$

$$\tau_a \dot{a}_i = c v_i - a_i \tag{2}$$

In the coupled equations v_i denotes the rate of neuron i, W_{ij} the connectivity matrix, ϕ a sigmoidal activation function, a_i the adaptation variable of each neuron, c the adaptation constant (the fraction of the rate which is integrated following neuron activation) and τ_m and τ_a indicate the membrane and adaptation time constants respectively.

To begin our analysis we ignore the external input term h_i and we assume that the symmetric rotation invariant weights permit a stable family of solutions such that:

$$f(\theta - \theta_i) = \phi(\sum_{j=1}^{N} (W_{ij} - c\delta_{ij}) f(\theta - \theta_j)) \tag{3}$$

Where θ indicates location around the ring of neurons, θ_i being the location of neuron i, and δ_{ij} is the Kronecker delta.

From this stage we look at the effect of a small perturbation δv_i about this attractor state by substituting $v_i = v_i + \delta v_i$ in Equation 1 and Taylor expanding about the stable solution. Following this we carry out a few stages of eigenvector manipulation followed by cancelling terms which are sufficiently small and rearrangement to reach a final expression in Equation 4

$$\dot{\theta} = \frac{1}{\tau_m \sum_i \frac{f'(\theta - \theta_i)^2}{\phi_i'}} (-\sum_i f'(\theta - \theta_i) a_i) \tag{4}$$

We can confirm that the form of this equation is a requirement for displacement of the bump; as the function $-f(\theta - \theta_i)$ is an odd function which is positive on the back of the bump and negative on the leading edge. The second step of the analysis proceeds by examining Equation 2, with the substitution $v_i = f(\theta - \theta_i) + \delta v_i$ (we ignore the term $c\delta v_i$ as both factors are assumed to be sufficiently small). Giving the following equation:

$$\tau_a \dot{a}_i = c f(\theta - \theta_i) - a_i \tag{5}$$

Intuitively we see that when $c = 0$ there is a stationary bump and as c is increased the perturbation grows relative to the rate of relaxation of a_i and therefore one can destabilize the bump.

Equation 5 is of the form $\frac{da}{dt} = Pa + Q$ where P and Q are functions of t and is therefore solvable using an integrating factor. Doing the integration and Taylor expanding θ in terms of a shift in time, permits a simple solution differential equation which is

again Taylor expanded assuming smallness of $\tau_a\dot{\theta}$ to give Equation 6 (where $f(\Delta\theta_i) = f(\theta(t) - \theta_i)$)

$$a_i = c\sum_{n=0}^{\infty} f^n(\Delta\theta_i)(-\tau_a\dot{\theta})^n \tag{6}$$

In order to find a closed form solution we can assume that the bump can be characterized with a cosinusodial profile, which is fairly valid within the non-saturating regime of the activation function. Following this assumption we can write Equation 6 as a summation of an infinite geometric series of sines and cosines, which after simplification gives us the following expression for a_i:

$$a_i = c\frac{\cos(\theta - \theta_i) + \sin(\theta - \theta_i)(\tau_a\dot{\theta})}{1 + (\tau_a\dot{\theta})^2} \tag{7}$$

Combining this equation and the equation for frequency (Equation 4) derived in the first step, cancelling orthogonal terms and rearranging gives us an expression for the frequency of the bump, where the scalar constant k is a term dependent of the shape of the bump of activity and the derivative of activation function about the stable solution:

$$\dot{\theta} = \frac{1}{\tau_a}\sqrt{\frac{k\tau_a c}{\tau_m} - 1} \tag{8}$$

It is important to note that Equation 8 is a particular solution for the cosinusodial activity profile. A more general form can be written as $v_0 = \frac{1}{\tau_a}\sqrt{\frac{kc}{c_0} - 1}$. Where we have a constant frequency v_0 for some factor c_0 depending on the bump profile.

SIMULATION RESULTS

Following the analysis we simulated the network specified by Equations 1 and 2 with $N = 100$ neurons and integrated these equations using Eulers method with a step size $\delta t = 0.25ms$ and set the time constants of the membrane and adaptation as $\tau_m = 1ms$ and $\tau_a = 50ms$ respectively.

We integrated the simulations for 20000 steps and calculated the phase of the activity bump by computing the weighted circular mean of the activity profile at each time step and computing the change of phase over time giving the instantaneous frequency ($\dot{\theta}$). To verify our analysis we ran the simulation for different values of the adaptation constant c, measuring the instantaneous frequency and averaging it over the time where we have oscillations (as opposed to stationary attractors), see Figure 1.

There is strong agreement here between the results of the simulation and the analytic solution, excepting values just greater than $c = 1$ (where we start to see oscillations); this is due to the fact that the smaller c becomes, the longer it takes for the negative feedback provided by a_i to get the bump moving. Theoretically therefore at the point where $\frac{kc}{c_0}$ is infinitesimally greater than 1, it will take an infinite amount of time for the stationary attractor to move into a steady oscillation. Therefore small values of c do not match accurately. However we can see the general square-root singularity characteristic is preserved showing the effectiveness of the approach.

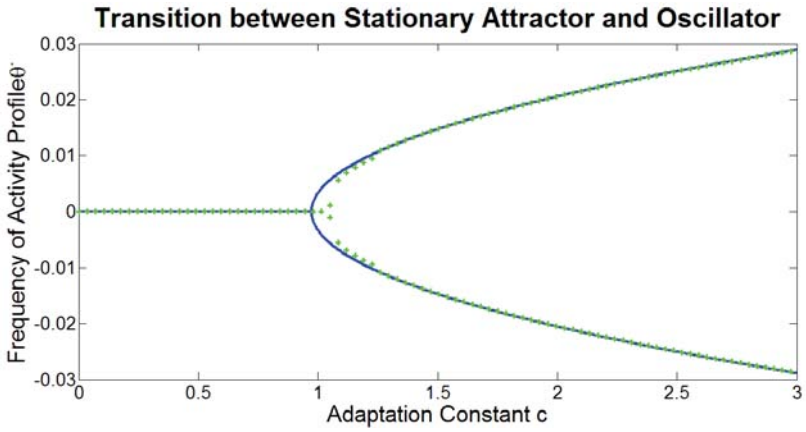


FIGURE 1. This figure shows a bifurcation diagram of the frequency of the bump rotation $\dot{\theta}$ against the adaptation constant c. The crossed markers indicate results from simulation and the line shows the analytic solution. We can clearly see the transition from stationary attractor (where $\dot{\theta} = 0$) to an oscillator (at $c \geq 1$).

DISCUSSION

We have demonstrated analytically and through simulation a possible role of adaptation in generating oscillations in ring attractor networks. Preliminary work suggests that oscillation frequency can be tuned by varying the net external excitatory input to the network however further work is needed. This mechanism could provide a candidate for generating velocity modified theta as observed in the hippocampal circuit [3, 4].

One prominent criticism of attractor networks [14], which could be addressed by future work, is how they are affected by heterogeneities in synaptic connectivity and neural responses. An experimental verification of the ring attractor architecture could be validated by future advances in optogenetics and connectomics.

REFERENCES

1. J. O'Keefe, and M. Recce, *Hippocampus* **3**, 317–330 (1993).
2. G. Buzsáki, *Neuron* **33**, 325–340 (2002).
3. A. Jeewajee, C. Barry, J. O'Keefe, and N. Burgess, *Hippocampus* **18**, 1175–1185 (2008).
4. A. Welday, I. Shlifer, M. Bloom, K. Zhang, and H. Blair, *J. Neurosci.* **31**, 16157–16176 (2011).
5. N. Burgess, C. Barry, and J. O'Keefe, *Hippocampus* **17**, 801–812 (2007).
6. K. Zhang, *J. Neurosci.* **16**, 2112–2126 (1996).
7. J. Taube, R. Muller, and J. Ranck Jr, *J. Neurosci.* **10**, 420–435 (1990).
8. H. Blair, K. Gupta, and K. Zhang, *Hippocampus* **18**, 1239–1255 (2008).
9. R. Meech, *Ann. Rev. Biophys. & Bioengin.* **7**, 1–18 (1978).
10. L. Donald Partridge, *J. Neurobiol.* **11**, 613–622 (1980).
11. J. Hopfield, *Proc. Nat. Acad. Sci. USA* **107**, 1648–1653 (2010).
12. V. Itskov, C. Curto, E. Pastalkova, and G. Buzsáki, *J. Neurosci.* **31**, 2828–2834 (2011).
13. O. Melamed, O. Barak, G. Silberberg, H. Markram, and M. Tsodyks, *J. Comput. Neurosci.* **25**, 308–316 (2008).
14. J. Knierim, and K. Zhang, *Ann. Rev. Neurosci.* **35**, 267–285 (2012).

Exploring the future with anticipatory networks

A. M. J. Skulimowski[*,†]

[*]*Department of Decision Science,Chair of Automatic Control and Biomedical Engineering, AGH University of Science and Technology, al. Mickiewicza 30, 30-050 Kraków, Poland.*
[†]*International Centre for Decision Sciences and Forecasting Progress & Business Foundation, Kraków, Poland.*

Abstract. This paper presents a theory of anticipatory networks that originates from anticipatory models of consequences in multicriteria decision problems. When making a decision, the decision maker takes into account the anticipated outcomes of each future decision problem linked by the causal relations with the present one. In a network of linked decision problems, the causal relations are defined between time-ordered nodes. The scenarios of future consequences of each decision are modeled by multiple vertices starting from an appropriate node. The network is supplemented by one or more relations of anticipation, or future feedback, which describe a situation where decision makers take into account the anticipated results of some future optimization problems while making their choice. So arises a multigraph of decision problems linked causally and by one or more anticipation relation, termed here the anticipatory network. We will present the properties of anticipatory networks and propose a method of reducing, transforming and using them to solve current decision problems. Furthermore, it will be shown that most anticipatory networks can be regarded as superanticipatory systems, i.e. systems that are anticipatory in the Rosen sense and contain a future model of at least one other anticipatory system. The anticipatory networks can also be applied to filter the set of future scenarios in a foresight exercise.

Keywords: anticipatory networks; superanticipatory systems; causal fields; scenarios; foresight.
PACS: 02.50.Le, 07.05.Mh , 89.75.Fb

INTRODUCTION

This paper introduces the reader to the theory of anticipatory networks, which generalizes the ideas related to anticipatory models of consequences in multicriteria optimization problems presented in [1,2,3]. We assume that when making a decision, the decision maker takes into account the anticipated outcomes of each future decision problem linked by the causal relations with the present one. In a network of linked decision problems, the causal relations can be defined between the time-ordered nodes only. The future scenarios of the causal consequences of each decision are modeled by multiple edges starting from an appropriate node. This network is supplemented by one or more relations of anticipation, or *anticipatory feedback*, which models the situation where decision makers take into account the anticipated results of some future decisions modeled by optimization problems while making their choice. They then explore the causal dependences of future constraints and preferences on the choice just made so that future decision outcomes fulfill the conditions specified as the anticipatory feedback relations.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 224-233 (2013); doi: 10.1063/1.4776525

Both types of relations as well as forecasts and scenarios regarding the future model parameters form an information model, which is called the *anticipatory network*. We will present the basic properties of anticipatory networks and the methods of using them to computing the solutions to current decision problems.

The theory outlined above resulted from the need to create an alternative approach to selecting a solution to multicriteria optimization problems that takes into account direct multi-stage models of future consequences of the decision made [1]. The anticipatory behavior of decision makers corresponds to the definition of anticipatory systems proposed by Rosen [4] and developed further in [5,6]. Namely, a system is called *anticipatory* if it makes its decisions taking into account anticipated future states of its outer environment and of itself. A bibliographic survey of these ideas can be found in [7]. The ability to create a model of the future that characterizes an anticipatory system is also a prerequisite for an anticipatory network, where each node models an anticipatory system and they are able to influence each other according to causal order. Here, we restrict the study of anticipatory networks to model decisions made in networked optimization or gaming problems. If each decision node models an optimization problem, then a network of optimizers arises − a class of information processing systems studied in [2]. In a similar way, one can construct networks with nodes modeling Nash equilibria, set choice problems, random or irrational decision makers, or hybrid networks containing nodes of all types [3].

The study of anticipatory networks starts from its simplest form – the chains of anticipatory units, after which anticipatory trees and general networks will be analyzed. One can note that anticipatory networks with loops correspond to strong anticipatory systems in the Dubois sense [5], while the acyclic networks correspond to weak ones. Then, motivated by the properties of the anticipatory networks, we will present the basic properties of superanticipatory systems. By definition, a *superanticipatory system* is anticipatory in the sense of Rosen [4] or Dubois [5] (weak or strong) and contains a future model of at least one other anticipatory system whose outcomes may influence its current decisions when taking into account anticipatory feedback relations. This notion is idempotent, i.e. the inclusion of other superanticipatory systems into the model of the future of a superanticipatory system does not yield an extended class of systems. Moreover, they can be classified according to a grade that counts the number of nested superanticipation. Most anticipatory networks can be regarded as superanticipatory systems because future decisions can be based on similar anticipatory principles with respect to the subsequent nodes in the network.

Among real-life applications of anticipatory networks, we can include the selection of compromise solutions to multicriteria strategic planning problems applying scenarios of anticipated consequences provided by foresight exercises. Specifically, the models presented in [2] and in this paper have been applied to solve scenario filtering problems that occurred in an IT foresight research project. Reducing the number of plausible scenarios is made possible due to the elimination of scenarios that correspond to irrational or contradictory future decisions. Another relevant class of applications is the coordinated cooperation of autonomous robots that are capable of mutually anticipating team members' actions and planning their own operations taking into account a collective performance criterion.

ANTICIPATORY NETWORKS: BASIC IDEAS

The original idea behind introducing anticipatory networks as consequence models can be formulated [1,2] as follows "*To use anticipated future consequences of a solution selected in a decision problem as additional preference information*". The exploration of anticipatory feedback in multicriteria decision making is possible owing to the following assumptions:

- There exist estimates (forecasts or foresight scenarios) of future decision problem formulations, their solution rules and preference structures.
- The decision making unit responsible for solving any decision problem included in the network knows whether and how the parameters of future problems are influenced by the solutions to preceding problems. This allows us to model and control the consequences of a decision to be made in any problem modeled by the network.
- The assessment of the anticipated outcomes of some future problems can be merged with the preference structures of the causally preceding problems.

Thus an anticipatory network is a synthetic structure that incorporates decision-making units that can be living intelligent beings, groups or organizations of them, as well as artificial intelligent agents and even animals. The question as to whether the anticipatory mechanisms occur at a neural level [8] is the subject of neurophysiological research. The nature of such mechanisms could be based on learning causal sequences that may activate the anticipatory neural response if there is an associated consequence in the internal memory in response to an input signal. Similar mechanisms can even be exhibited by non-specialized cells [9]. However, it is not clear how anticipatory reactions at the cellular level can explain the informed anticipation and model building at human intelligence level, and whether it is possible at all. To address this question, in [10] a formal anticipatory neuron and a model of an artificial anticipatory neural network has been proposed. The internal memory makes this neuron capable of forming artificial anticipatory neural networks that can simulate reinforcement learning and exhibit interesting properties, but the future model building capability remains unexplained.

An even more intriguing issue related to anticipatory systems and networks is the physical realization of anticipation beyond causal model building. Recently, quantum entanglement across time has been investigated a. o. in [11,12] showing a way to constructing strong anticipatory systems [5], i.e. systems, where anticipation is based on a direct functional influence of the future on the present rather than on a modeled (predicted) future. However, in this paper we will not apply a feedback with future events: the anticipatory feedback defined further in this section uses the predicted model of the future to influence the present assuming that that the three assumptions above and the causality principle hold. This complies with the definition of weak anticipatory systems [5].

Anticipatory networks model causally-linked decision problems and their environment. According to a slightly more general definition given in [2], a decision maker A acts as an optimizer on a set of feasible decisions U and on the preference structure P and selects a subset $X \subset U$ according to P and to the fixed set of optimization criteria F that are characteristic for this decision problem.

We will assume that the decision units at each node of the network solve the multicriteria optimization [13] problems of the form

$$(F:U\rightarrow E)\rightarrow min(P), \tag{1}$$

where P is an arbitrary preference structure, i.e. $P:=\{\pi(u)\subset U:\ u\in U\}$ such that if $v\in \pi(u)$ and $w\in \pi(v)$ then $w\in \pi(u)$. Usually E is a vector space with a partial order introduced by a convex cone θ, and

$$\pi(u):=\pi(u,\theta)=\{v\in U:\ F(v)\leq_\theta F(u)\}.$$

A *free decision maker A* [14] may select any solution u_0 from U that is *nondominated* with respect to P and F in (1), i.e. if u_0 belongs to the set

$$\Pi(U,F,P):=\{u\in U:\ [\forall v\in U:\ F(v)\leq_\theta F(u)\Rightarrow v=u]\} \tag{2}$$

A is then uniquely characterized by U, F, and P and may be denoted as a 3-tuple $A:=(U,F,P)$. If the admissible solution set for A may be different from $\Pi(U,F,P)$ and equal to $X\subset\Pi(U,F,P)$, we will denote it as $A:=(U,F,P,X)$. X will be interpreted as the set of those solutions to the problem (1) which are capable of being actually selected.

Besides their optimizing capabilities, the solutions made by any decision maker may influence the parameters of some future decision problems, thus forming networks with some new properties compared to the theory of multi-level multicriteria problems [15]. In particular, in feed-forward anticipatory networks of multicriteria decision problems, constraints and preference structures in some problems are causally linked to the results of solving other problems and may depend on their preference structures. Thus, in a network of multicriteria decision problems the parameters of the actual instances of problems to be solved vary depending on the solutions of other problems in the network.

An influence relation r that describes how the decision maker A_1 affects the scope of admissible decisions in the subordinated decision problem A_2 may be defined as

$$A_1:=(U_1,F_1,P_1,X_1)\ r\ A_2:=(U_2,F_2,P_2,X_2)\Leftrightarrow\exists\varphi:X_1\rightarrow 2^{U_2}:\ X_2=\varphi(X_1), \tag{3}$$

where φ is a multifunction that defines the restriction of the admissible decision scope at A_2, $U_1\subset E_1$, $U_2\subset E_2$. Influence relations linking preference structures may be defined analogously. If r is acyclic it will be termed a *causal relation*. From this point on, the term *causal network* will refer to the graph of a causal relation. In a causal network of decision problems, the function φ influences the constraints in A_i transforming the outputs from the problem preceding A_i into additional constraints. One can also consider influence relations with multifunctions φ capable of creating additional admissible solutions in E_2 beyond U_2.

To complete the definition of anticipatory networks, we will define the anticipatory feedback relation.

Definition 1. Suppose that G is a causal network consisting of decision problems and that a decision problem A_j in G precedes another one, A_i, in causal order r. Then the *anticipatory feedback* between A_i and A_j in G is an information flow concerning the anticipated output from A_i regarded as an input information to the decision to be made at the node A_j. ■

As in the case of causal relations, there may also exist multiple types of anticipatory information feedback in a network, each one related to the different way the anticipated future decisions, usually the optimization results, are considered at a decision node. Both relations, the causal influence and the anticipatory feedback, when considered jointly and expressed in a diagram, form an anticipatory network of decision problems.

Definition 2. An *anticipatory network* (of decision problems) is a causal network of decision problems with at least one additional anticipatory feedback relation. ■

In general, for given decision problems O_k, O_n and O_m, there may exist different ways of influencing O_m by O_n, so a causal diagram of an anticipatory network could be a multigraph. This is discussed further in the next sections, where we will also solve the anticipatory networks, according to the following definition.

Definition 3. An anticipatory network is termed *solvable* if the process of restricting the sets of admissible decisions at all problems represented in the network by considering the anticipatory feedbacks results in selecting a unique non-empty solution set at the starting problem. ■

A simple causal graph of decision problems that can be embedded in a straight line will be called a chain. The general underlying idea behind anticipatory network solution procedures is to analyze chains of decision problems linked by a causal influence relation, then to identify in a network of decision problems elementary cycles consisting of causal influence along chains and future information feedback relations, i.e. cycles which do not contain other cycles. For chains and trees of decision problems we have proposed numerical solution procedures [2] based on an analysis of the above-defined elementary cycles, starting from those most distant in the network. The procedures are also based on replacing a solved elementary cycle by a synthetic decision unit and updated links to the remaining elements of the network. The process is repeated iteratively until all cycles are solved. A general network can be decomposed into chains, which makes it possible to apply aggregated chain rules iteratively, gradually eliminating solved chains.

SOLVING ANTICIPATORY NETWORKS

We will refer to anticipatory networks of multicriteria optimizers (1)-(3) with the causal influence relation defined by linking multifunctions

$$Y_i: F_{i-1}(U_{i-1}) \to 2^{U_i},\ \varphi(i) := Y_i \circ F_{i-1} \tag{4}$$

imposing additional constraints in sets U_i. The dependence of preference relations P_i on the outcomes of previous problems is defined by the functions

$$\psi: X(U_{i-1}, F_{i-1}, P_{i-1}) \ni f \to P_i. \tag{5}$$

In [2] the anticipatory information feedback in causal networks of decision problems has been applied to selecting a solution to networked discrete choice problems with respect to multiple criteria. Specifically, while making a decision, the decision maker takes into account forecasts concerning the parameters of future decision

problems, the anticipation concerning the behavior of future decision makers as well as the forecasted causal dependence relations linking the parameters of nodes in the network. Also taken into account are the anticipatory relations that point out relevant future outcomes to particular decisions to be made at nodes preceding them in the causal order.

A large class of anticipatory networks can be reduced to a subsequent analysis of all chains in a network. A formal background for solving the initial decision problem $A:=(U,F,P)$ in an anticipatory chain of decision problems A_i, $i=0,1,...,N$, with discrete admissible solution sets U_i, is given below.

Causal relations between decision problems are given in the form of restrictions in the scope of admissible decisions defined as multifunctions $\varphi(i)$ that depend on solutions of previous problems modeled by A_{i-1}, for $i=1,...,N$. Future information feedback is defined as information about anticipated fulfillment (or not) of certain conditions by the values of criteria in future optimization problems. The following definitions will be helpful for describing the solution procedure in a more rigid manner.

Definition 4. For a chain of decision problems A_k $k=0,1,...,N$, let us define the (weak) *anticipatory feedback* condition at A_i, $0 \leq i < N$, as a requirement that

$$\forall j \in J(i) \text{ for a given family of subsets } \{V_{ij}\}_{j \in J(i)} \; u_j \in V_{ij} \subset U_j \text{ or } u_j \text{ is closest to } V_{ij} \qquad (6)$$

where the sets V_{ij} represent the most preferred decisions to be made at the j-th decision problem from the point of view of the decision maker that is responsible for the outcomes from the i-th decision problem and the satisfaction of this preference is measured by the proximity of the solutions of the j-th problem to V_{ij}. ■

The satisfaction of the anticipatory feedback condition (6) means that the decision maker at A_i strives to select the solution which guarantees that the results of future decision problems A_{ij} are as close as possible to the specified solution sets V_{ij}. The criteria values $F_j(V_{ij})$ are of special importance to the decision makers and can be defined as reference sets [16].

In an anticipatory network with the starting node (U,F,P) the following decision problem can be formulated:

Problem 1. For all chains in the network find all the sequences of decisions starting at an $u_0 \in U$ that additionally fulfil anticipatory feedback condition (6). ■

The solution approach for anticipatory chains can easily be generalized for the case where solutions of each decision problem can influence multiple decisions to be made in the future and that do not depend on each other. If such a situation occurs, it can be represented as a *causal tree*. An example of a simple causal tree is given in Fig. 1.

The reduction of the analysis of anticipatory trees to the subsequent analysis of anticipatory chains in the tree is possible due to the following property:

Theorem 1. Assume that the decision made at the decision making unit A_i influences two causally independent decision problems A_k and A_m in an anticipatory tree T and let A_t be the (unique) bifurcation decision problem for A_i,A_k and A_m. , i.e. such that no other problem that is causally influenced by A_t can influence both A_k and A_m. Furthermore, let C_k and C_m be the sets of admissible decision chains starting at A_i and ending at A_k and A_m, respectively, and let C_k' and C_m' contain the elements of C_k and

C_m, respectively, starting at A_i and truncated at the bifurcation problem A_t. Then the set of all admissible decision chains with respect to both A_k and A_m starting at A_i can be generated as extensions of the elements of the intersection of C_k' and C_m'. More precisely, an extension of any such sequence of decisions starting at A_i and ending at A_t is to be concatenated with an arbitrary subsequence starting at A_t of an admissible decision chain that was truncated at prior to the intersection of C_k' and C_m'. ■

Proof of the above theorem is given in [2] (Proposition 2).

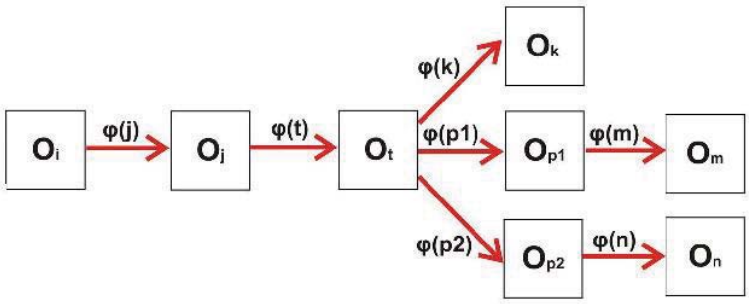


FIGURE. 1. An example of a causal tree, where O_t is the bifurcation problem (cf. Thm.1) for O_m, O_n. O_k, O_{p1} and O_{p2}. Causal relations are defined by the multifunctions $\varphi(i):=Y_i \circ F_{i-1}$. Anticipatory feedback relations can occur between any pair of causally linked decision problems (not shown in this figure).

In general, in a network of decision problems there may exist units that are influenced causally by two or more predecessors. Such problems may emerge in practice when, for example, an input to a production function comes from two independent technological processes, which are both optimized with respect to quality and price. In order to deal with such a situation, observe first that the causal dependences in the form of constraints on the set of admissible decisions in a subsequent problem A_k that comes from two or more causally independent decision problems $A_i=(U_i,F_i,P_i)$ and $A_j=(U_j,F_j,P_j)$ as the multifunctions Y_i and Y_j, *respectively,* yield, in fact, just an intersection of constraints that can be represented as a new multifunction Y defined on the Cartesian product of $F(U_i)$ and $F(U_j)$ in the following way

$$Y(u_{ip},u_{jr}):= Y_i(u_{ip}) \cap Y_j(u_{jr}).$$

Based on this observation, in the case of arbitrary networks, the calculations can again be reduced to an analysis of chains and elementary loops in the network, i.e. loops which may consist of both causal relations and anticipatory feedbacks, and do not contain other loops. Analogously to surveying the bifurcation decision problems and 'cutting the branches' of an anticipatory tree, all decision problems which are causally influenced by two or more predecessors must be surveyed. If an elementary loop is detected, it can be replaced by a synthetic decision unit with a reduced set of admissible chains and updated links to the remaining elements of the network. The process can be repeated iteratively until the 0^{th} decision problem has been reached.

To solve Problem 1, information about future optimization problems and their mutual relations is required. If the time horizon of anticipatory planning is large compared to the time allotted to modeling and computing the decision, usually the changes in the modeled environment also proceed slowly enough allowing an analyst to rely on the information gathered prior to performing all computations. This is the case of foresight applications, where the time horizon is usually between 10 and 20 years, the analytic phase can be stretched over several months and the resources available allow us to explore the future to a sufficient extent.

Anticipatory Networks as Superanticipatory Systems

Let us observe that in the above presented approach to solving anticipatory networks we have assumed that anticipation is a universal principle governing the solution of decision problems at all stages. In particular, future decision makers modeled at the starting decision node A_0 can in the same way take into account the network of their relative future decision problems when making decisions. Thus, the model of the future of the decision maker at A_0 contains models of future agents including their respective future models. This has motivated us to introduce the notion of superanticipatory systems, which are direct generalizations of anticipatory systems in the sense of Rosen [4]:

Definition 5. *A superanticipatory system is an anticipatory system that contains at least one model of another future anticipatory system and both are linked either by a causal or by an anticipatory feedback relation.* ■

Since, by definition, every superanticipatory system is also anticipatory, the class of superanticipatory systems remains closed when an anticipatory system contains a model of a superanticipatory one. However, the notion of a superanticipatory system grade can be introduced, namely a superanticipatory system *is of grade n* if it contains a model of a superanticipatory system of grade *n-1*. By definition, an anticipatory system that does not contain any model of another such system is superanticipatory of grade *0*. It can be observed that an anticipatory network containing a chain on *n* decision problems, each one linked with A_0 and with all its causal predecessors by an anticipatory feedback, is an example of a superanticipatory system of grade *n*.

FINAL REMARKS

This paper presented the main ideas concerning anticipatory networks, the basic methods of solving them, and their extension, so-called superanticipatory systems.

Anticipatory networks may be applied to model and solve a broad range of problems, both real-life and theoretical. The above approach focuses on the problem of finding feasible foresight scenarios based on the identification of future decision-making processes and on anticipating their outcomes. Scenarios, such as those defined and used in foresight and strategic planning [17], can depend on the choice of a decision in one of the networked optimization problems as well as be external-event driven. When included in a causal network of decision problems, the anticipation of future decisions and alternative external events would allow us to generate alternative structures of de-

cision problems in the network. Assuming that at each decision problem in the causal network the decision makers strive to select their decisions in a rational way and applying multicriteria optimization methods to find all potential variants of anticipated future problem outcomes, a set of potential elementary scenarios [18] of future trends and events modeled by the network can be found. In addition, considering the anticipatory feedbacks in the network that are defined in the next section, a filtering of plausible outcomes from each problem is made possible, as well as a reduction of the set of plausible elementary scenarios. This application can be very supportive when building foresight scenarios by clustering elementary scenarios. For further potential fields of application of anticipatory networks, the reader is referred to [2].

Anticipatory networks should be regarded as a new class of world models that can describe decision making processes in a clear formal way. The *forecasting* of events in causal systems can thus be complemented by *anticipation* of rational decisions. Furthermore, the ideas behind anticipatory networks naturally led us to introduce superanticipatory systems.

Future research on further extensions of the theory and potential applications of anticipatory networks will include relations to anticipatory neural mechanisms that would make it possible to build neural structures which fit forecasting and anticipation tasks in an optimal way. Moreover, observing that the anticipatory capabilities of an autonomous system go beyond the abilities attributed usually to artificial consciousness, further studies of superanticipatory systems should provide clues as regards the structure of artificial consciousness.

ACKNOWLEDGEMENT

The author is grateful for the support of research project No. WND-POIG.01.01.01-00-021/09: "Scenarios and development trends of selected information society technologies until 2025" funded by the ERDF within the Innovative Economy Operational Programme, 2006-2013, www.ict.foresight.pl.

REFERENCES

1. A. M. J. Skulimowski, *Found. Control Engrg*. **10**(1), 25–38 (1985).
2. A. M. J. Skulimowski, *Int. J. Sys. Sci.,* iFirst, 1-21 (2012).
3. A. M. J.:Skulimowski, "Hybrid anticipatory networks", in *Proc. ICAISC 2012*, edited by L. Rutkowski, *et al.*, *Lecture Notes in Artificial Intelligence* **7268**, Springer, Berlin-Heidelberg, 2012, pp.706–715
4. R. Rosen, *Anticipatory Systems - Philosophical, Mathematical and Methodological Foundations*. Pergamon Press, London, 1985.
5. D. M. Dubois, "Mathematical Foundations of Discrete and Functional Systems with Strong and Weak Anticipations", in: *Anticipatory Behavior in Adaptive Learning Systems, State-of-the-Art Survey*, edited by M. V. Butz, O. Sigaud, and P. Gérard, *Lecture Notes in Artificial Intelligence* **2684**, pp. 110–132, Springer, Berlin-Heidelberg, 2003.
6. M. V. Butz, O. Sigaud, G. Pezzulo, and G. Baldassare, "Anticipations, Brain, Individual and Social Behavior", in *Anticipatory Behavior in Adaptive Learning Systems: From Brains to Individual and Social Behavior* edited by G. Pezzulo, *Lecture Notes in Artificial Intelligence* **4520**, pp. 1–18, Springer, Berlin-Heidelberg, 2007.
7. M. Nadin, *Int. J. General Sys.*, **39**(1), 35–133 (2010).

8. M. R. Mehta, *Neuroscientist* **7**(6) (2001), 490–495.
9. N. S. Baliga, *Science* **320**, 1297–1298 (2008).
10. G. Terenzi, "Anticipation in Biological and Cognitive Systems: The Need for a Physical Theory of Biological Organization", in *Physics of Emergence and Organization*, pp.371–388, edited by I. Licata, and A. Sakaji, World Scientific Publishing Co., Singapore, 2008.
11. S. J. Olson, and T. C. Ralph, *Phys. Rev. Lett.* **106**, 110404 (2011).
12. R. Wiegner, C. Thiel, J. von Zanthier, and G. S. Agarwal, *Optics Lett.* **36**(8), 1512–1514 (2011).
13. I. Kaliszewski, *Soft Computing for Complex Multicriteria Decision Making*, Springer, New York, 2006.
14. A. M. J. Skulimowski, "Freedom of Choice and Creativity in Multicriteria Decision Making", in *Knowledge, Information, and Creativity Support Systems: KICSS2010 Revised Selected Papers*, edited by T. Theeramunkong, S. Kunifuji, V. Sornlertlamvanich, and C. Nattee, *Lecture Notes in Artificial Intelligence* **6746**, pp. 190–203, Springer, Berlin, Heidelberg, 2011.
15. I. Nishizaki, and M. Sakawa, *Cooperative and Noncooperative Multi-Level Programming*. Operations Research/Computer Science Interfaces Series Vol. 48, Springer, Dordrecht, Heidelberg, London, New York, 2009.
16. A. M. J. Skulimowski, "Methods of Multicriteria Decision Support Based on Reference Sets", in *Advances in Multiple Objective and Goal Programming*, edited by R. Caballero, F. Ruiz, and R. E. Steuer, *Lecture Notes in Economics and Mathematical Systems* **455**, pp. 282–290, Springer, Berlin-Heidelberg-New York, 1997.
17. M. Godet, *Creating Futures – Scenario Planning as a Strategic Management Tool*. Economica, London, 2001.
18. A. M. J. Skulimowski, "Discovering Complex System Dynamics with Intelligent Data Retrieval Tools", in *Sino-foreign-interchange workshop on Intelligent Science and Intelligent Data Engineering IScIDE 2011, Xi'an, China, Oct. 23-26, 2011*, edited by Y. Zhang, *et al.*, *Lecture Notes in Computer Science* **7202**, pp. 614–626, Springer, Berlin–Heidelberg, 2012.

In vitro closed loop optical network electrophysiology: An introduction

A. El Hady[*,†,‡,§] and W. Stühmer[†,‡,§]

[*] *Max Planck Institute for Dynamics and Self Organization, Goettingen, Germany.*
[†] *Max Planck Institute of Experimental Medicine, Goettingen, Germany.*
[‡] *Bernstein Focus for Neurotechnology, Goettingen, Germany.*
[§] *Bernstein Center for Computational Neuroscience, Goettingen, Germany.*

Abstract. We present a novel experimental paradigm "In vitro closed loop optical network electrophysiology (ivCLONE)". This seminar note gives an overview of the basics of optical neurostimulation, network electrophysiology and closed loop electrophysiology. Moreover, the notes discuss how combination of aforementioned techniques would help us to address network-level phenomenon and how single neuron properties are related to collective network dynamics.

Keywords: optogenetics; network electrophysiology; closed loop stimulation.
PACS: 87.19.lj, 87.18.Sn, 87.19.lr, 87.19.lv, 87.19.lw

OPTICAL NEUROSTIMULATION

Recent advances in genetics, chemistry and optics have provided unprecedented opportunities to use light to stimulate, inhibit or control neuronal activity with molecular specificity and a high temporal and spatial resolution. Optical simulation offers a non-invasive method to control neuronal activity. It has the advantage of producing fewer artifacts than electrical stimulation. It also allows stimulating neurons in a high spatial and temporal resolution taking into consideration advances in optical technologies such as digital micromirror devices (DMD) and holographic photostimulation.

The optical control strategies can be broadly divided into non-genetic and genetic methods for optical control. The non genetic methods comprise: caged neurotransmitters, reversibly caged signaling molecules that can be released by a flash of light allowing the liberated compound to act on endogenous or exogenous neuronal targets before diffusing away (e. g. Bis-Q, GluAzo, XAQs), photoswitched tethered ligands for native channels (QBr and a Quaternary Ammonium PAL) and nanoparticles that stimulate neurons upon magnetic-field mediated heating. The genetic methods include genetically engineered light gated channels and receptors (SPARK light gated K^+ channels and LiGluR: light gated kainate type glutamate receptor) and Opsin based control of neuronal activity. The opsin-based tools are a large class of channels that are genetically targeted. It comprises the ChARGe channel, Melanopsin, Channelrhodopsin 1 (ChR1), Channelrhodopsin 2 (ChR2), Volvox Channelrhodopsin 1 (VChR1), Volvox Channelrhodopsin 2 (VChR2) and Halorhodopsin [1]. The genetic methods have the advantage that constructs can be expressed and targeted to specific neuronal compartments thus avoiding unspecific effects.

Channelrhodopsin 2 (ChR2) is our optogenetic tool of choice. Channelrhodopsin 2 is

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 234-243 (2013); doi: 10.1063/1.4776526

an inwardly rectifying non-selective cation channel. At neutral pH, ChR2 is permeable to physiologically relevant cations such as H^+, Na^+, K^+, and Ca^{2+} [2, 3]. The single ion channel conductance of ChR2 has been estimated at 50 fS [4]. Channelrhodopsin 2 is a membrane spanning retinylidene protein. It has a 7 transmembrane domain structure. The chromophore is an all *trans* retinal that undergo isomerization to 13-cis-retinal upon absorption of a blue photon.

The typical photocurrent of channelrhodopsin 2 consists of a large transient peak that has onset of around 4–10 ms [2, 5, 6]; this transient peak quickly decays to a stationary component that is typically <20–50% of the initial peak photocurrent [2] Upon removing the light, ChR2 closes with a time constant of 10–20 ms [2]. After switching off the light, the photocurrent decays in a biexponential manner [7].

Although wild type ChR2 is the gold standard in optogenetics, it suffers from few shortcomings that prevent it from being used for some neuroscience applications. The shortcomings are: non-selectivity, fast inactivation, slow recovery and low conductance. Variants and mutants are of ChR2 are designed to circumvent the aforementioned shortcomings. The E90Q mutation [8] has increased sodium selectivity and much reduced proton permeability vs. wild-type ChR2. Another channel called CatCh with the mutation L132C had increased calcium permeability [9]. The variant K132A and Q95A show strong photocurrents and increased potassium selectivity thus enabling suppressing of neuronal activity. On the other hand, the H134R mutant [10] demonstrates increased conductance by approximately twofold. Step function opsins (SFO) are built using mutations to C128 [11] drastically slowing down the rate of ChR2 closure from the open state, thus effectively creating a bistable open P520 state until illuminated with green light. The SFO mutations are designed to stabilize the active retinal isomer, which results in the prolongation of the active state of the channel even after light-off. Another SFO with the mutation D156A have even longer inactivation time-constant that can reach eight minutes [12]. A new class of channels called Stabilized Step Function Opsins (SSFOs) was constructed by combining both the D156 and C128 mutations that led to spontaneous deactivation times of around 30 minutes [13].

In order to allow high frequency stimulation, the E123T mutant, combined with the H134R mutation, speeds channel closure and increases the precision of neural action-potential firing at the expense of photocurrent and light sensitivity [14], resulting in a mutant called ChETA. The E123T mutation was combined with T159C mutation to produce a channel that can drive neurons at high frequencies and have a high light sensitivity [15]. On the other hand, chimeras of ChR1 and ChR2 have been constructed by several researchers [3, 16], one of which was that composed of ChR1 helices A-E and ChR2 helices F-G (called ChEF). These chimeras displayed the small inactivation of ChR1, but the large photocurrents of ChR2 on account of improved membrane localization and light sensitivity. An I190 V substitution to ChEF led to the molecule, "ChIEF", capable of driving more reliable fast spiking due to the much larger stationary current and faster channel closing kinetics after light offset [16]. Another chimera called C1V1 was constructed. C1V1 is composed of the first two and one half helices of ChR1 and the last four and one half helices of VChR1, which led to a red shifted activation spectrum for the chimeric channel and nanoampere currents. All of the above variants and mutants of ChR2 provide a versatile toolbox to control neuronal activity.

Optogenetic tools have been proposed and implemented for advancing the analysis of

neuronal systems on all levels from single cells through circuit's structure and function up to the level of behaviour. Optogenetics has been used in many in-vitro studies. It was used in investigating synaptic physiology and plasticity at single synapses level [17]. It was also used to induce homeostatic synaptic depression and to understand the molecular machinery underpinning it [18]. It was also used to study oscillations in hippocampal slices in vitro [19]. It is also contributing to understand receptor pharmacology [20, 21, 22]. It was also used to probe various aspects of astrocytes functioning [23]. Recent developments have led to use of optogenetics tools to control signaling pathway [24].

Optogenetics have been used in many in vivo animal models. It was used to control the c.elegans muscle wall motor neuron and mechanosensory neuron activity [10]. It was also used in flies to investigate the neuronal basis of the nociceptive response [25]. On the other hand, it was also used in Zebrafish to examine cardiac function and development [26], transduction of sensory neuron mechanoreception [27] command of swim behaviour [28] and saccade generation [29]. In the mouse, ChR2 was used to investigate the contribution of the hypothalamic hypocretin neurons to sleep and wakefulness [30]. Optogenetic stimulation was also used to stimulate axonal terminals in the nucleus accumbens, which lead to the discovery that dopamine neurons co-release glutamate [31, 32]. Reports on the functions of parvalbumin expressing fast spiking interneurons demonstrated directly their involvement in gamma oscillations and information processing in mouse prefrontal [33] and somatosensory cortex [34, 35]. It also enabled rapid functional mapping of motor control across the motor cortex [36].

Optogenetics is also being used to discern the possible therapeutic mechanism of cortical intervention in mouse models of depression [37] and to develop novel strategies for control of peripheral neurons [38]. It will increase our understanding for disease states and the development of novel therapeutics as it has been used for example [39] to optically control symptoms of Parkinson's disease and also to control of epileptiform activity [40].

There has also been some work on optogenetic modulation of primate neurons [41, 42, 43, 44]. Optogenetics will have great impact on the development of neuroprothetics specially retinal prosthetics that are now reaching a mature and advanced stage that might allow it to be translated for use in human beings [45].

NETWORK ELECTROPHYSIOLOGY

There is a growing consensus that individual elements of information are encoded by populations or clusters of cells, and not by individual cells. This encoding strategy is named "Population coding". Visual features for example such as orientation, color, direction of motion and depth are encoded with population codes in visual cortical areas [46, 47] Motor commands in the motor cortex rely also on population codes [48]. Thus, it became clearer that sensory processing in our brain and memory and learning processes are coordinated by the activity of many neurons in a network. Another crucial aspect that is crucial for neuronal information processing is the topology and connectivity of the networks. Over the past decades experimental and theoretical studies have revealed candidate connectivity architectures that are expected to enable networks of neurons to operate as memory storage devices, as sensory modules that can track rapidly changing

sensory inputs or as discrimination devices that can support e. g. categorical perception. It further highlight the crucial role of networks to perform computations that are relevant to the brain cognitive functions. In order to study the problem of distributed network processing and the network structure-function relationship, it is important to develop experimental tools that address neurons on the network level. These experimental tools will help us to understand network-level phenomenon and the relationship between single neuron properties and population activity.

Neurons spontaneously form functional synapses when cultured in vitro and develop complex patterns of activity that closely resemble those recorded from developing brains of animals [49]. Neurons retain their morphological and pharmacological identities in culture but there are likely to be numerous subtle changes in their properties due to the unnatural environment in which they have been placed. Many techniques have been developed recently in order to track the activity of neurons grown in vitro and to tackle these network level activities. These techniques can be divided into electrophysiological or optical methods. Optical methods either use population calcium imaging [50, 51, 52] or voltage sensitive imaging [53, 54] in order to track the activity of multiple neurons simultaneously. Electrical methods include planar titanium nitride based multielectrode arrays, CMOS based microelectrode arrays [55], field effect transistor arrays [56], vertical nanowire arrays [57] and gold mushroom shaped microelectrodes [58]. Of particular interest are the nanoelectrode arrays, which promise to provide intracellular like recordings and stimulation of many individual neurons while the electrodes maintain an extracellular position. Micha Spira coined the term “In cell recording” [59], reflecting the fact that interfacing neurons with these arrays of nanoelectrodes will allow recordings of individual action potentials and sub-threshold potentials with matching quality and signal to noise ratio of conventional intracellular sharp glass microelectrodes or patch electrodes. Moreover, it will ultimately offer a high spatial resolution and might achieve the single synapses resolution so that one can monitor several synapses simultaneously.

For our purpose, we used the conventional commercially available titanium nitride based multielectrode arrays (MEA). MEAs are produced with variable layouts, number of electrodes, electrode materials, electrode size and interelectrode distances specially for slice recording where a specific geometry is required to monitor activity of different brain regions: retina [60, 61], spinal cord [62] and Hippocampus [63, 64]. Multielectrode arrays are able to gather data from multiple sites in parallel, and to avoid the need to place all electrodes individually by hand thus allowing multi-unit neuronal recordings. It also provides the opportunity to perform long-term recordings of cultured neuronal networks. On the other hand, they have the following limitations: Smaller amplitude recordings as compared to traditional instrumentation such as intracellular recordings because the electrodes are not inserted inside the cells or the tissue and the electrodes cannot be moved independently because they are arranged in fixed patterns.

On the application side, MEAs have been used in neuronal and cardiac electrophysiological applications. They were used for multisite slice recordings on hypothalamic slices to investigate the effect of Gherlin on hypothalamic network activity, on the activity of dissociated root ganglia cell cultures and on acute hippocampal slice investigating oscillations and rhythmic activity [65] and to monitor synchronized cardiac muscle and stem cell culture activity. It can also be used for studying learning and memory on the network level [66, 67] and to study of development of network electrical

activity and population bursting dynamics [68]. MEAs were also used to study retinal information processing and the role of correlations in the retinal circuitry [69, 70]. MEAs might also be used to establish high-throughput systems to perform drug screenings and toxicology studies [71].

CLOSED LOOP ELECTROPHYSIOLOGY

The basic paradigm of closed loop electrophysiology involves recording neural activity or behavior and delivering activity-dependent stimulation in real time. Closing the loop around neural systems offers advantages over traditional open-loop feedforward neurophysiological approaches by providing the ability to stimulate neural systems contingent on their behavior. Closed loop neurophysiology has been advanced by recent software and hardware developments and by the emergence of novel tools to control neuronal activity with spatial and temporal precision. Real-time stimulation feedback enables a wide range of innovative studies of information processing and plasticity in neuronal networks under realistic conditions [72, 73, 74].

In addition to advancing basic neuroscience, bidirectional neural interfaces provide novel adaptive neuroprosthetic devices that incorporate artificial sensory feedback. Activity-dependent stimulation also promises innovative paradigms for effective treatment of neurological diseases. Using closed loop stimulation, it is possible to program an artificial feedback with defined rules and constrains. Closed loop electrical stimulation has been successfully used beforehand to clamp network activity [75], to control bursting activity [76] and realize embodiment by using the network represented on the network to control a robotic arm [77].

Feedback closed loop stimulation will further increase our system versatility by providing a mean to photostimulate neurons depending on their current state. We termed our experimental system, combining optical neurostimulation and closed loop network electrophysiology, *"In Vitro Closed loop Optical Network Electrophysiology"*.

IN VITRO CLOSED LOOP OPTICAL NETWORK ELECTROPHYSIOLOGY (IVCLONE)

System description

The system [78] is composed of a 60 channel MEA amplifier that records from multielectrode arrays on which ChR2 transfected neurons are grown and a high power blue LED used for whole field illumination (photostimulation) (Figure 1).

Cell preparation

Cell cultures were prepared according to Brewer et al. [79]. Hippocampal neurons were obtained from Wisteria WU rat embryos at 18 days of gestation (E18). Cells were then cultured on multielectrode arrays (Standard MEA; type TiN-200-30iR from

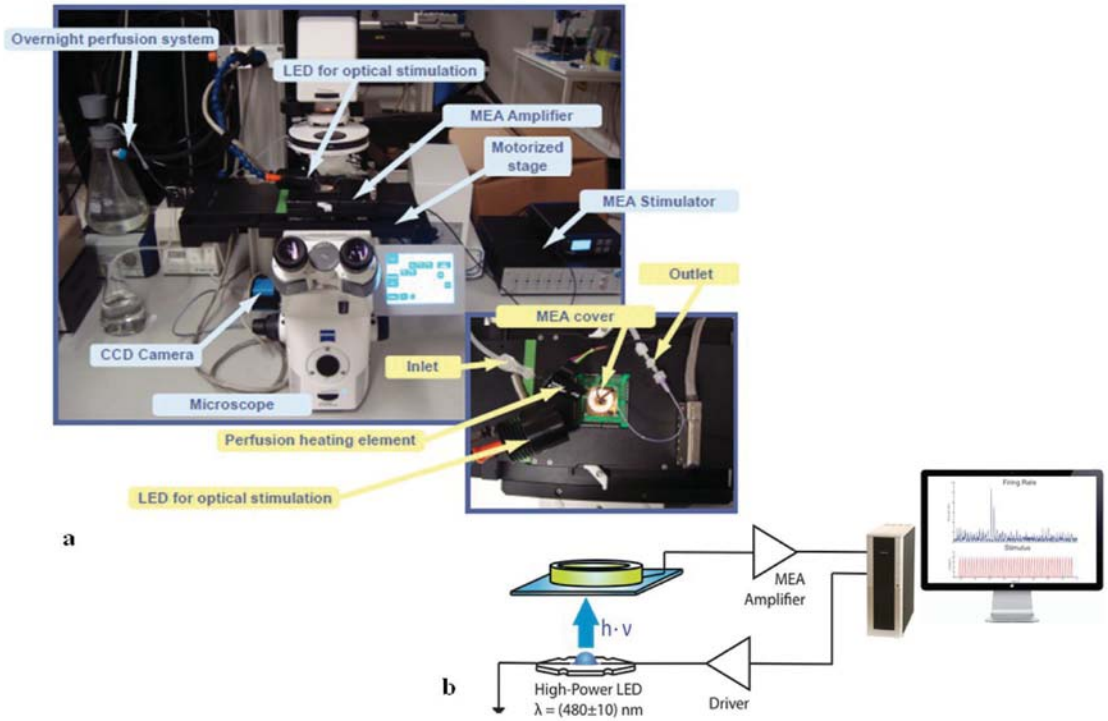


FIGURE 1. ivCLONE setup: (a) Picture depicting the setup with its different components. (b) Sketch showing the main components of the system.

Multichannel Systems) coated with a mixture of poly-D-lysine and laminin at a density of 1000 cells per mm^2. A droplet of approximately 100 μl cell suspension was added in the middle of the multielectrode array to cover the recording area. The arrays were then filled with 1 ml of the aforementioned serum free B27/Neurobasal medium .The cells were kept in an incubator at 37°C and a mixture of 5% CO2 + 95% O2. Half of the medium was changed every two days. The cultures were kept till 40 DIV. MEAs were sealed with gas permeable membranes, which allowed the long-term culture. The culture dishes are sealed with a Teflon membrane, fluorinated ethelyene-propylene. Although the membrane has no pores (thus preventing infection), it is quite permeable to some small molecules notably oxygen and carbon dioxide. It is hydrophobic and thus relatively impermeable to water and water vapor. The membrane slows the shift in pH of carbonate buffered media caused by removal from an incubator with 5% CO2 atmosphere, by about a factor of two compared to a standard culture disch with an air gap [80]. All animals were kept and bred in the animal house of the Max Planck Institute for Experimental Medicine according to the German guidelines for experimental animals. Animal experiments were carried out with authorization of the responsible federal state authority.

Multielectrode arrays measurements

Recordings were done on 21 DIV hippocampal neuronal cultures (transfected at 14 DIV with AAV-CAG-CHOP2 virus). Data from MEAs were captured at 25 kHz using a 64-channel A/D converter and MC_Rack software (Multichannel Systems, Reutlingen). The MCS measurement card has 64 analogue input channels, with a resolution of 16 bit, a programmable gain and an input voltage range of $\pm$400 mV to $\pm$4 V, depending on the gain level specified. Sixty of these serve as input channels for the multi electrode array (MEA), three serve as analogue inputs, and one of which serves as a combined channel for 16 digital inputs, which each set a single bit. After high pass filtering (Butterworth 2nd order, 100 Hz) action potentials are detected in a cutout recorded 1ms before and 2 ms after crossing a threshold of -20 μV, which was $>$ 3 times standard deviations of the baseline activity. Routinely, it was made sure that the amplifier noise does not exceed $\pm$10 μV, which was indicated by manufacturer to be the acceptable noise level. Standard recording were performed for a maximum of 30 minutes. Longer recordings for many hours were performed under continuous perfusion.

Whole field photostimulation

The key requirements for the light source used for illumination were high light power at around 480 nm, fast and well controllable modulation of the light power and stable illumination over several hours. An additional requirement for the illumination of the spatially extended MEAs is homogeneous light power density over an area of 1 $\times$ 1 mm. All these requirements were met by a blue light emitting diode (LED, Luxeon rebel color with Lambertian dome, Philips Lumileds) with 5 W maximal power consumption, placed 25 mm below the illuminated hippocampal neuronal cultures grown on multielectrode arrays. The absorption spectra of the LEDs have a large overlap with the absorption spectra of channelrhodopsin 2. The light output was controlled via the voltage of STG 2008 stimulus generator (Multichannel systems, Reutlingen) , converted to current in a custom made analog driver circuit, resulting in a input of 1 W at the LED for each Volt at the D/A-board. Rise-time to maximum Light power was $<$ 20 μs (Figure 2).

System's applications

Our optical network electrophysiology was used to induce network level plasticity and modify the intrinsic collective dynamics of a cultured neuronal network [81]. Towards this aim, we designed a photostimulation paradigm that aims to drive neurons in a more naturalistic in vivo like fashion [82, 83]. The possibility to detect action potentials over long periods of time and from many individual neurons in parallel combined with non-invasive photostimulation will enable us to address new questions e. g. screening for the effect of mutations or protein knockdown on the dynamical properties of neurons and also on their ability to be potentiated. It can also help us compare among individual

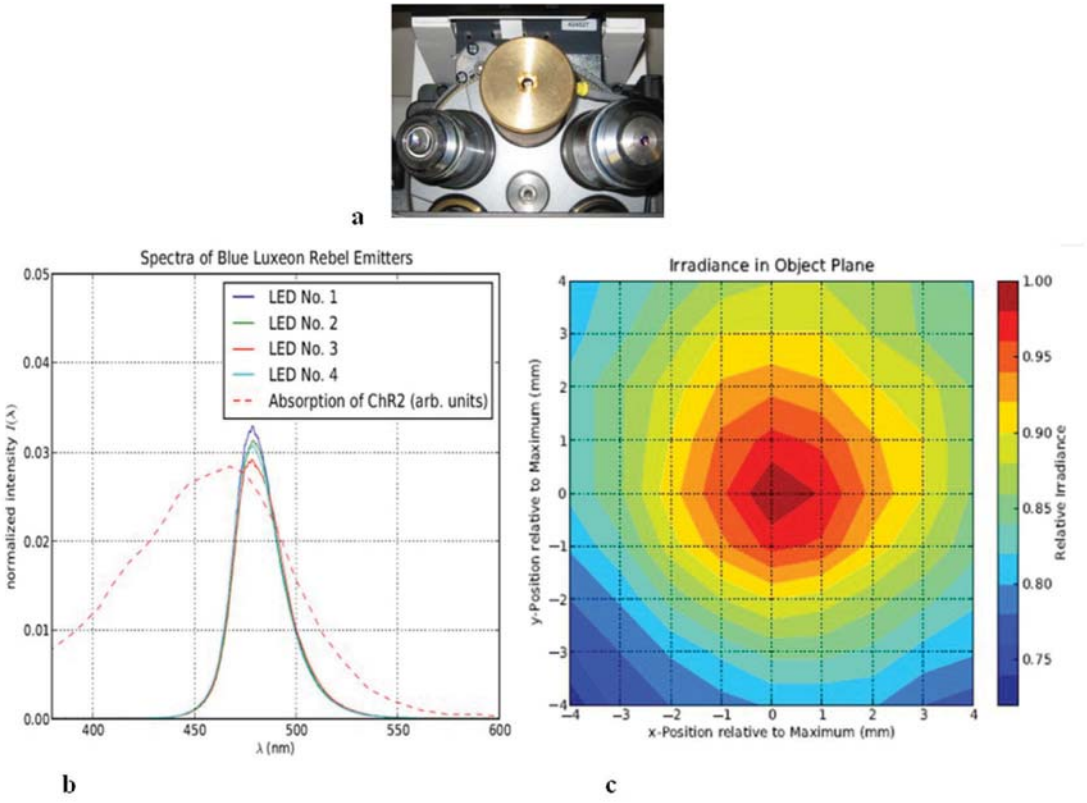


FIGURE 2. Whole field blue light illumination: (a) The LED holder which screws into objective turret under microscope stage. (b) Spectra of four blue Luxeon Rebel Color emitters compared with each other overlapped with action spectrum of ChR2. (c) Relative variation on the irradiance respective to the point of maximum power density, measured approximately 25 mm above the bare LED. The active area of the MEA is only 1×1 mm^2 in size.

neurons with respect to their encoding diversity thus contributing to understand the biophysical basis of such diversity.

Closing the loop using our optical stimulation system would allow us to futher extend the questions to be addressed as the photostimulation can be adjusted depending on the response of neurons thus offering a better control over neuronal dynamics. In the context of learning and memory, closing the loop would help to stabilize a learned sequence over long time and most interestingly, one can address the cellular and molecular mechanisms underlying such long term network level memory.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of Fred Wolf, Theo Geisel, Ragnar Fleischmann, Ghazaleh Afshar and Kai Bröking to this project. The authors acknowledge the BMBF for funding (Grant number 01GQ0811, 01GQ01005B).

REFERENCES

1. S. Szobota, and E. Y. Isacoff, *Annu. Rev. Biophys.***39**, 329-348 (2010).
2. G. Nagel, *et al.*, *Proc. Natl. Acad. Sci. USA* **100**, 13940-13945 (2003).
3. H. Wang, *et al.*, *J. Biol. Chem.* **284**, 5685-5696 (2009).
4. K. Feldbauer, *et al.*, *Proc. Natl. Acad. Sci. USA* **106**(30), 12317–12322 (2009).
5. E. S. Boyden, *et al.*, *Nat. Neurosci.* **8**(9), 1263–1268(2005).
6. C. Bamann, *et al.*, *J. Mol. Biol.* **375**, 686–694 (2008).
7. K. Nikolic, *et al.*, *Photochem. Photobiol.* **85**(1), 400–11 (2009).
8. E. Ritter, *et al.*, *J. Biol. Chem.* **283**, 35033–35041 (2008).
9. S. Kleinlogel, *et al.*, *Nat. Neurosci.* **14**(4), 513–518 (2011).
10. G. Nagel, *et al.*, *Curr. Biol.* **15**(24), 2279–2284 (2005).
11. A. Berndt, *et al.*, *Nat. Neurosci.* **12**, 229–234 (2009).
12. C. Bamann, *et al.*, *Biochemistry* **49**(2), 267–278 (2010).
13. O. Yizhar, *et al.*, *Nature* **477**(7363), 171–178 (2011).
14. J. L. Harwood, and I. A. Guschina, *Biochimie* **91**,679–684 (2009).
15. A. Berndt, *et al.*,*Proc. Natl. Acad. Sci. USA* **108**(18), 7595–7600 (2011).
16. J. Y. Lin, *et al.*, *Biophys. J.* **96**, 1803–1814(2009).
17. P. Schoenenberger, *et al.*, *Exp. Physiol.* **96**(1), 34–39 (2010).
18. C. P. Goold, and R. A. Nicoll, *Neuron* **68**(3), 512–528 (2010).
19. T. Akam, *et al.*, *Nat. Neurosci.* **15**,763–768 (2012).
20. K. A. Bell, *et al. Neuropharmacology* **61**(8), 1379–1388 (2011).
21. M. J. Highley, and B. L. Sabatani, *Nat. Neurosci.* **13**(8), 958–966 (2010).
22. C. Schöne, *et al.*, *J. Neurosci.* **32**(36), 12437–12443 (2012).
23. M. Figueiredo, *et al.*, *Exp. Physiol.* **96**(1), 40–50 (2011).
24. J. E. Toettcher, *et al.*, *Nat. Methods.* **8**(1), 35–38 (2011).
25. R. Y. Hwang, *et al.*, *Curr. Biol.* **17**(24), 2105–2116 (2007).
26. A. B. Arrenberg, *et al.*, *Science* **330**(6006), 971–974 (2010).
27. S. E. Low, *et al.*, *J. Neurosci.* **30**,9359–9367 (2010).
28. A. B. Arrenberg, *et al.*, *Proc. Natl. Acad. Sci. USA* **106**,17968–17973(2009).
29. P. J. Schoonheim, *et al.*, *J. Neurosci.***30**,7111–7120 (2010).
30. A. R. Adamantidis, *et al.*, *Nature* **450**,420–424 (2007).
31. G. D. Stuber, *et al.*, *J. Neurosci.* **30**,8229–8233 (2010).
32. F. Tecuapetla, *et al.*, *J. Neurosci.* **30**, 7105–7110 (2010).
33. V. S. Sohal, *et al.*, *Nature* **459**, 698–702 (2009).
34. J. A. Cardin, *et al.*, *Nat. Protoc.* **5**, 247–54 (2010).
35. J. A. Cardin, *et al.*, *Nature* **459**, 663–67 (2009).
36. O. G. Ayling, *et al.*, *Nat. Methods* **6**, 219–24 (2009).
37. H. E. Covington, *et al.*, *J. Neurosci.* **30**, 16082–90 (2010).
38. M. E. Llewellyn, *et al. Nat.Med.***16**, 1161–1165 (2010).
39. V. Gradinaru, *et al.*, *Science* **324**, 354–59 (2009).
40. J. TÃÿnnesen, *et al.*, *Proc. Natl. Acad. Sci. USA* **106**(29), 12162–12167 (2009).
41. X. Han, *et al.*, *Neuron* **62**, 191–198 (2009).
42. I. Diester, *et al.*, *Nat. Neurosci.* **14**, 387–397 (2011).
43. M. Jazayeri, *et al.*, *Nat. Neurosci.* **15**(10), 1368–1370 (2012).
44. A. Gerits, *et al.*, *Curr. Biol.* **22**(18), 1722–1726 (2012).
45. V. Busskamp, and B. Roska, *Curr. Opin. Neurobiol.*. **21**(6), 942–946. (2011).
46. W. M. Usreyand, and R. C. Reid, *Annu. Rev. Physiol.* **61**, 435–456 (1999).
47. A. Pouget, *et al.*, *Nat. Rev. Neurosci.* **1**(2), 125–132 (2000).
48. D. J. Tolhurst, *et al.*, *Vision Res.* **23**, 775–785 (1983).
49. Y. Ben-Ari, *Trends Neurosci.* **24**(6), 353–60 (2001).
50. M. Minderer, *et al.*, *J. Physiol.* **590**(1), 99–107 (2012).
51. K. Schulz, *et al.*, *Nat. Methods.* **9**(6), 597–602 (2012).
52. B. M. Kampa, *et al.*, *Cold Spring Harb. Protoc.* **2011**(11), 1340–1349 (2011).
53. W. Akemann, *et al.*, *J. Neurophysiol.* **108**(8), 2323–2337 (2012).
54. L. Barnett, *et al.*, *PLoS One* **7**(9), e43454 (2012).

55. S. Hafizovic, *et al.*, *J. Neurosci. Methods* **164**, 93–106 (2007).
56. P. Fromherz, *Ann. N. Y. Acad. Sci.* **1093**, 143–160 (2006).
57. J. T. Robinson, *et al.*,*Nat. Nanotechnol.*. **7**(3), 180–184 (2012).
58. A. Hai, *et al.*, *J. R. Soc. Interface.***6**(41), 1153–1165 (2009).
59. A. Hai, *et al.*, *Nat. Methods***7**(3), 200–202 (2010).
60. A. E. Grumet, *et al.*, *J. Neurosci. Methods* **101**(1), 31–42 (2000).
61. M. Meister, *et al.*, *Science* **252**(5008), 939–943 (1991).
62. D. A. Borkholder, *et al.*, *J. Neurosci. Methods.***77**(1), 61–66 (1997).
63. S. A. Boppart, *et al.*, *IEEE Trans. Biomed. Eng.* **39**(1), 37–42 (1992).
64. U. Egert,, *et al.*, *Brain Res. Brain Res. Protoc.* **2**(4), 229–242 (1998).
65. K. Shimono, *et al.*, *J. Neurosci.* **20**(22), 8462–8473 (2000).
66. G. Shahaf, and S. Marom, *J. Neurosci.* **21**(22), 8782–8788 (2001).
67. Y. Jimbo, *et al.*, *Biophys. J.* **76**(2), 670–678 (1999).
68. D. A. Wagenaar, *et al.*, *BMC Neurosci.* **7**, 11 (2006).
69. E. Ganmor, *et al.*, *J. Neurosci.* **31**(8), 3044–3054 (2011).
70. E. Schneidman, *et al.*, *Nature* **440**(7087), 1007–1012 (2006).
71. D. A. Stenger, *et al.*, *Trends Biotechnol.***19**(8), 304–309 (2001).
72. J. D. Rolston, *et al.*, *Front. Neurosci.* **4**, 31 (2010).
73. M. Arsiero, *et al.*, *Arch. Ital. Biol.* **145**(3-4), 193–209 (2007).
74. P. Chamorro, *et al.*, *PLoS One* **7**(7), e40887 (2012).
75. A. Wallach, *et al.*, *Front. Neuroeng.* **4**, 3 (2011).
76. D. A. Wagenaar, *et al.*, *J. Neurosci.* **25**(3), 680–688 (2005).
77. D. J. Bakkum, *et al.*, *Front. Neurorobot.* **1**, 5 (2007).
78. A. El Hady, *et al.* "In Vitro Closed loop Optical Electrophysiology of Networks I: Whole field illumination Paradigm" ,Proceedings of 7thInt. Meeting on substrate integrated multlelectrode array meeting 2010 , pp. 253-254.
79. G. J. Brewer, *J. Neurosci. Res.* **42**(5), 674–683 (1995).
80. S. M. Potter, and T.B. DeMarse, *J. Neurosci. Methods* **110**(1-2), 17–24 (2001).
81. G. Afshar, *et al.*, "Optogenetic modification of network burst structure", *Society for Neuroscience conference 2011 abstract*.
82. A. Neef, *et al.*, "Continuous Dynamic Photostimulation - delivering defined, in-vivo-like fluctuating stimuli with Channelrhodopsins", *Society for Neuroscience conference 2011 abstract*.
83. A. Neef, *et al.*, "Continuous Dynamic Photostimulation - inducing defined, in-vivo-like fluctuating stimuli with Channelrhodopsins", *Bernstein Center for computational neuroscience 2011 conference*.

Experiments on clustered neuronal networks

S. Teller and J. Soriano

Departament d'Estructura i Constituents de la Matèria. Facultat de Física. Universitat de Barcelona, Spain.

Abstract. Neuronal cultures show a rich repertoire of spontaneous activity. However, the mechanisms that relate a particular network architecture with a specific dynamic behavior are still not well understood. In order to investigate the dependence of neuronal network dynamics on architecture we study spontaneous activity in networks formed by interconnected aggregates of neurons (*clustered neuronal networks*). In the experiments we monitor the spontaneous activity using calcium fluorescence imaging. Network's firing is characterized by bursts of activity, in which the clusters fire sequentially in a short time window, remaining silent until the next bursting episode. We also investigate perturbations on the connectivity of the network. We mainly focus in physical damage. In some cases we observe important changes in the collective activity of the network, while in other cases some dynamic motifs are preserved, hinting at the existence of dynamic robustness.

Keywords: neuronal cultures; clustered networks; spontaneous activity; motifs.
PACS: 87.85.Wc, 87.18.Sn, 87.19.lh, 87.19.lj, 87.19.ll

INTRODUCTION

One of the major challenges of modern neuroscience is the understanding of the interplay activity–connectivity, i.e. the relationship between a particular neuronal network architecture and the activity patterns or dynamic scenarios that it exhibits. Recent studies [1, 2] have focused on the dependence of network's activity on connectivity, and studied the stability and synchronization of neuronal groups as well as the efficiency of these networks to propagate, process and store information.

Many of these studies are motivated by the importance of spontaneous activity in neuronal networks, which plays a pivotal role in processes as complex as development, learning, memory, and synchronization [3]. Spontaneous activity is often characterized by episodes of intense collective activity of several hundred of milliseconds (*network bursts*), separated by quieter inter–burst intervals of several seconds in duration [4]. Despite substantial efforts, the mechanisms behind the origin, maintenance and regulation of spontaneous activity are still unclear.

The aim of our work is study how changes in the connectivity of a neuronal network influence its spontaneous dynamics. We focus on a particular neuronal network design formed by densely packed aggregates of neurons (clusters) [5, 6] connected to one another. Clustered cultures *in vitro* are constituted by interconnected aggregates of neurons and glia (Fig. 1a). When neurons are deposited on a substrate and are free to move, the inherent motility of cells together with the tension forces exerted by neurite bundles results in a fast aggregation process that leads to the formation of the clusters. Clustered neuronal networks are a versatile experimental tool that has proven very successful for the development of patterned neuronal cultures [6, 7], and

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 244-246 (2013); doi: 10.1063/1.4776529

the investigation of several problems, including synchronized oscillations [5], neuronal collective activity [8] , and self–organization mechanisms [9].

In our experiments we study the richness and variability of the clusters' spontaneous activity. Among other aspects, we investigate the initiation of activity (for instance by identifying hubs or specially active clusters). We also study which are the fundamental topological and dynamical ingredients that maintain network's collective activity.

PREPARATION OF CLUSTERED CULTURES

The neurons in our experiments are harvested from embryonic Sprague–Dawley rats. We use cortical neurons of $18-19$ day old embryos. The cortices are isolated from the rest of the brain and dissociated by repeated pipetting. Finally, the neurons are plated on a 13 mm glass cover slips in the presence of supporting medium, 5% CO_2 and 95% humidity. Further details in the culturing procedure can be found in [10].

An important aspect of the culture preparation process is the absence of adhesive proteins in the glass substrate. The free substrate facilitates cell motility and a preferential attachment of a neurons with its neighbors. The resulting clusters have a characteristic spherical shape that minimizes the surface contact with the substrate (Fig. 1a).

Activity measurements are carried out $2-3$ weeks after plating. At this stage the network is considered mature, i.e. the clusters and their interconnectivity is stable and the whole network is spontaneously active. Neuronal activity is monitored through calcium fluorescence imaging [11], and we use Fluo–4 as fluorescence calcium indicator.

Cultures are observed in a Zeiss inverted microscope connected to a CCD camera. Images of clusters' spontaneous activity are recorded at 60 frames per second and a size of 700×600 pixels that cover about 4×3 (width$\times$height) mm^2. In a typical experiment we record spontaneous activity for about 1 h in networks containing ~ 20 clusters.

RESULTS AND DISCUSSION

Activity in our clustered networks is characterized by a series of bursting episodes in which clusters fire sequentially. The sequence of activation and the number of clusters participating may vary from burst to burst, as shown in the raster plot of Fig. 1b.

A first and simple strategy to identify bursts that share a similar activation sequence (dynamic *motif*) consists in calculating the average firing onset time of the clusters within a burst, and then compare the obtained value from burst to burst. Since the connectivity of the culture is not changed during the experiment, this analysis provides a general overview of the statistical richness of the activation sequences, as well as their similarity. The analysis is shown in Fig. 1c for the 7 first bursts of a typical experiment.

We also studied the loss of actual connections in the network by ablating connections through a pulsed laser. Preliminary results show that some dynamic motifs are preserved upon network damage, hinting at the existence of dynamic robustness. However, in other cases the spontaneous activity is highly affected, but in a non–trivial manner. Indeed, it seems that secondary connections are reinforced when the primary ones are lost, suggesting that the network regulates itself to maintain the overall activity.

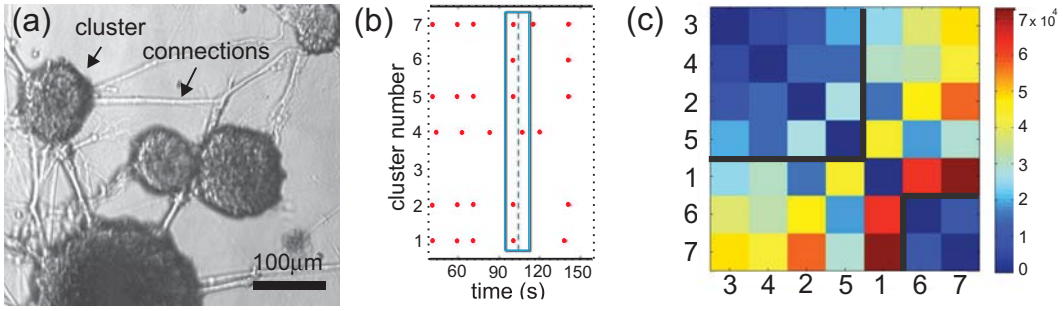


FIGURE 1. Experiments on clustered neuronal cultures. (a) Bright field image of a region of a culture. (b) Raster plot of clusters activity, with the dots marking the firing of a given cluster. The blue outline depicts a network burst, and the dashed line the average firing time of the clusters within the burst. (c) Classification of the first 7 bursts recorded in the experiment. Three distinct groups are observed (bursts 2 to 5; 1; and 6–7), each group corresponding to a similar clusters' firing sequence.

Our experiments, and the results by others [12] reveal that clustered cultures exhibit a rich repertoire of motifs and suggest different explanations for burst repetition in a propagation profile. For instance, repetition may arise from a particular balance between inhibitory and excitatory connections, noise, a favorable network architecture, or the existence of special neurons that initiate the spontaneous activity. Our work and new experiments may help clarifying in the future these and other aspects.

ACKNOWLEDGMENTS

We thank J. Orlandi, E. Tibau, T. Coll, and P. Loza-Álvarez for fruitful discussions and assistance. We also thank the Institute of Photonic Sciences (ICFO, Castelldefels, Spain) for its support. We acknowledge financial support from Ministerio de Ciencia e Innovación (Spain) under projects FIS2008-06335-C02-01, FIS2009-07523, and FIS2010-21924-C02-02, and the Generalitat de Catalunya under project 2009-SGR-00014.

REFERENCES

1. S. Feldt, P. Bonifazi, and R. Cossart, *Cell Press Rev.* **34**, 225–236 (2011).
2. E. Marconi, *et al.*, *PloS ONE* **7**, e34648 (2012).
3. L. Sirovich, A. Omurtag , and K. Lubliner, *Comp. Neur. Science* **102** 17, 3–29 (2006).
4. J. P. Eckman, *et al.*, *Phys. Rep.* **449**, 54–76 (2007).
5. M. Shein Idelson, E. Ben-Jacob, and Y. Hanein, *PloS ONE* **5**, el4443 (2010).
6. R. Sorkin, T. Gabay, P. Blinder, D. Baranes, and E. Ben-Jacob, *J. Neural Eng.* **3**, 95–101 (2006).
7. M. Shein, *et al.*, *Biomed. Microdevices* **11**, 495–501 (2009).
8. M. Shein-Idelson, E. Ben-Jacob, and Y. Hanein, *Front. Neuroeng.* **4**, 10 (2011).
9. R. Segev, M. Benveniste, Y. Shapira, and E. Ben-Jacob, *Phys. Rev. Lett* **90**, 168101 (2003).
10. J. Soriano, M. Rodríguez-Martínez, T. Tlusty, and E. Moses, *Proc. Nat. Acad. Sci. USA* **105**, 13758–13763 (2008).
11. N. Rochefort, H. Jia, and A. Konnerth, *Cell Press Rev.* **14**, 389–399 (2008).
12. N. Raichman, and E. Ben-Jacob, *J. Neurosci. Methods* **170**, 96–110 (2008).

Effect of input noise on neuronal firing rate

S. Gonzalo-Cogno and I. Samengo

Centro Atómico Bariloche and Instituto Balseiro, Bariloche, Argentina.

Abstract. When neurons are driven with a noisy input, the mean and the variance of the stimulus modulate the firing rate. Previous studies have shown that in linear-nonlinear model neurons the mean firing rate obtained in response to a noisy input is the average rate that would be obtained from an ensemble of constant currents. In this work, we study the firing rate of several neuron models, focusing on its dependence on the amount of input noise. We find that for models with monotonic activation curves, the theory provides a good qualitative approximation of the firing rate. For neurons with non-monotonic activation curves, however, the theory fails. The discrepancies between the theory and the simulations appear because rapidly fluctuating stimuli involve intrinsically dynamical processes that cannot be interpreted as an ensemble of constant stimuli.

Keywords: firing rate; LN models; activation curve; receptive field.
PACS: 87.16.ad, 87.17.Aa, 87.18.Tt, 87.19.ll

INTRODUCTION

Neurons in the brain process unsynchronized signals arriving to their thousands of dendrites. These signals add up to an irregular input current. Therefore, to understand how neurons process information in realistic conditions, researchers must not only analyze the responses to stereotyped stimuli as sinusoidal signals or step currents, but also explore noisy inputs. In presence of noise, the firing rate of a neuron depends both on the variance and the mean of the stimulus. In this work we focus on linear-nonlinear (LN) models. These models are based on the concept of the *receptive field*, that is, the direction in stimulus space that maximizes neuronal output. This direction is inherent to each neuron, so different neurons have different receptive fields.

For neurons with a single receptive field, LN models assume that the firing probability depends on the similarity (the scalar product) between the stimulus and the receptive field. Mathematically, the probability of having a spike at time t_0 in response to a stimulus *s(t)* is

$$P[\text{spike at } t_0|\mathbf{s}] = g\left[\int_{-\infty}^{t_0} s(t)\ f(t-t_0)\ \mathrm{d}t\right], \tag{1}$$

where g is typically a nonlinear function and $f(t)$ is the receptive field of the neuron, which can be estimated from the Spike Triggered Average (STA),

$$STA(t) = \frac{1}{N}\sum_{t_i} s(t-t_i), \tag{2}$$

where N is the total number of spikes, and t_i are the spiking times.

Consider a neuron that has a single receptive field and is driven with a stimulus $s(t) = I_0 + \eta\,\xi(t)$, where I_0 and η^2 represent the mean and the variance of the signal,

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 247-250 (2013); doi: 10.1063/1.4776516

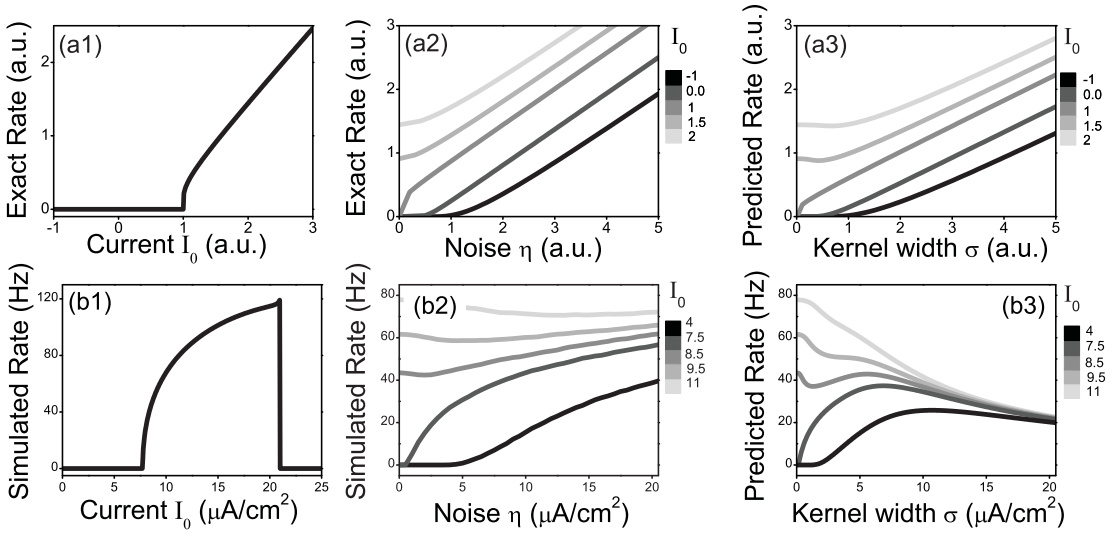


FIGURE 1. Mean firing rate of different types of models. (a) Integrate and firing neuron. **(b)** Type 1 Morris-Lecar neuron. (a1) and (b1): Firing rate as a function of I_0 for $\sigma = 0$. (a2) and (b2): Firing rate as a function of the input noise η for different I_0 values. (a3) and (b3): Predicted firing rate calculated with Eq. 3 as a function of the kernel width σ.

and $\xi(t)$ is Gaussian white noise of zero mean and unit variance. In this case, the STA is proportional to the receptive field [1]. Hence, once the STA has been estimated, the firing probability of Eq. (1) is easily obtained. With this result, if the noiseless activation curve $g(I_0)$ is known, by averaging Eq. (1) one can derive the mean firing rate also for $\eta > 0$ [2]. The LN prediction of the firing rate reads

$$f_p(I_0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} e^{-(x-I_0)^2/2\sigma^2} g(x) dx, \tag{3}$$

where $\sigma = \eta/\varepsilon$ is the kernel width of the gaussian and $\varepsilon = \int_{-\infty}^{0} f(t)\mathrm{d}t$.

RESULTS

Here we test the validity of Eq. (3) in two different types of models.

Models with monotonic activation curves

In these models, in the absence of noise, the firing rate is a monotonic function of the input current I_0. In Fig. 1 (a) we show the leaky integrate-and-fire model as an example. In this model, the exact firing rate can be found analytically [3]. In (a1) the firing rate as a function of I_0 is shown for the noiseless case ($\eta = 0$). The exact firing rate appears in (a2) as a function of η, and can be compared with (a3), where the predicted rate is shown as a function of the kernel width σ. The two families of curves are similar, since after an

initial transient, in both of them the firing rate grows linearly with the amount of noise. Hence, the LN approximation provides a good description of the mean firing rate. The same conclusion is reached when Eq. (3) is tested with other models with monotonic firing rates, for example for Hodgkin-Huxley [4] or Wang Buszaki [5] neurons.

Models with non-monotonic activation curves

We now move to neuron models with non-monotonic activation curves, more specifically, to models that in the absence of noise only fire for a finite range of I_0 values, bounded from above and from below. One such example is a Type I Morris Lecar model [6]. In this case the noiseless firing rate cannot be obtained analytically, so we compute it numerically (b1). When driven with constant currents, this model has two bifurcations. The neuron only fires for I_0 values lying between the two critical currents. In (b2) and (b3) the simulated and the predicted firing rates are displayed as a function of the amount of noise. We see that as noise grows, the simulated firing rate increases whereas the predicted firing rate decreases. We conclude that the LN description is not successful in this case. Other models with non-monotonic activation curves show the same discrepancy, as for example the Type 2 Morris Lecar neurons, and the slow currents of transiently firing thalamic neurons [7].

DISCUSSION

LN models assume that fluctuating input currents are equivalent to an ensemble of constant input currents: In Eq. (1), increasing σ can cause the same effect as modifying I_0, if the scalar product remains unchanged. As a consequence, Eq. (3) states that the firing rate in response to noisy stimuli can be obtained by averaging the firing rate in response to constant stimuli. This picture works well when applied to neural models with monotonic activation curves. However, when applied to transiently firing models, Eq. (3) predicts that for large σ firing rates should decay, contradicting analytical results and numerical simulations. The discrepancy appears because large σ values are considered equivalent to a broad ensemble of I_0 values—some of them large—and for large I_0 these models stop firing. Ceasing to fire for large I_0, hence, is traduced in diminished firing for large σ. When transient models stop firing, the phase portrait no longer shows a closed limit cycle. However, the system still has excitable trajectories. Therefore, if noise forces the voltage to jump away from its resting value, the system requires a long detour along a spike-shaped trajectory to return to the fixed point. Therefore, although periodic firing is not possible in the noiseless case, spike-like excursions are still possible when $\eta > 0$.

REFERENCES

1. E. J. Chichilnisky, *Network* **12**, 199–213 (2001).
2. S. Hong, B. N. Lundstrom, and A. L. Fairhall, *PLoS Comput. Biol.* **4**, e1000119 (2008).
3. A. N. Burkitt, *Biol. Cybern.* **95**, 1–19 (2006).
4. A. L. Hodgkin, and A. F. Huxley, *J. Physiol. (London)* **117**, 500–544 (1952).

5. X. J. Wang, and G. Buzsáki, *J. Neurosci.* **16**, 6402–6413 (1996).
6. C. Morris, and H. Lecar, *Biophys. J.* **35**, 193–213 (1981).
7. H. Yin, C. Cox, P. Mehta, and U. Shanbhag, "Bifurcation analysis of a thalamic relay neuron model," in *American Control Conference, 2009. ACC '09.*, 2009, pp. 337 –342, ISSN 0743-1619.

Unsupervised learning in neural networks with short range synapses

L. G. Brunnet, E. J. Agnes, B. E. P. Mizusaki and R. Erichsen Jr.

Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051, 9150-970 Porto Alegre, RS, Brazil.

Abstract. Different areas of the brain are involved in specific aspects of the information being processed both in learning and in memory formation. For example, the hippocampus is important in the consolidation of information from short-term memory to long-term memory, while emotional memory seems to be dealt by the amygdala. On the microscopic scale the underlying structures in these areas differ in the kind of neurons involved, in their connectivity, or in their clustering degree but, at this level, learning and memory are attributed to neuronal synapses mediated by long-term potentiation and long-term depression. In this work we explore the properties of a short range synaptic connection network, a nearest neighbor lattice composed mostly by excitatory neurons and a fraction of inhibitory ones. The mechanism of synaptic modification responsible for the emergence of memory is Spike-Timing-Dependent Plasticity (STDP), a Hebbian-like rule, where potentiation/depression is acquired when causal/non-causal spikes happen in a synapse involving two neurons. The system is intended to store and recognize memories associated to spatial external inputs presented as simple geometrical forms. The synaptic modifications are continuously applied to excitatory connections, including a homeostasis rule and STDP. In this work we explore the different scenarios under which a network with short range connections can accomplish the task of storing and recognizing simple connected patterns.

Keywords: pattern formation; theoretical neuroscience; synapses.
PACS: 87.19.lg, 87.19.lp, 87.10.Hk, 87.19.lv, 87.19.lw

INTRODUCTION

Since the original works on artificial neuron networks in the last century [1] the scientific contact among physicists and biologists has increased considerably. These attempts converged to a more realistic description of the phenomena and has enriched the knowledge of this field. Versions of model neurons [2] adapted to specific needs and conditions have been proposed along the last fifty years, but just rather recently experiments have advanced to allow for a detailed description of Hebbian like synapses [3, 4] and on mechanisms to regulate network homeostasis [6, 7]. Detailing the connections is also a hard task and it is frequently supposed that they happen involve many neurons. Network models then usually assume a fraction of random connections among neurons and search for properties related to pattern learning, associative memory and storage capacity. This random connections construction is quite artificial since neurons are physical entities that will contact their closest neighbors with a greater probability than farther ones. Here we ask how far should neurons be connected in order to reproduce the expected network properties. The aim of this work is to construct a model neuron lattice with local synaptic connections and to search for the conditions under which spatially induced pattern memories may be recovered.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 251-254 (2013); doi: 10.1063/1.4776532

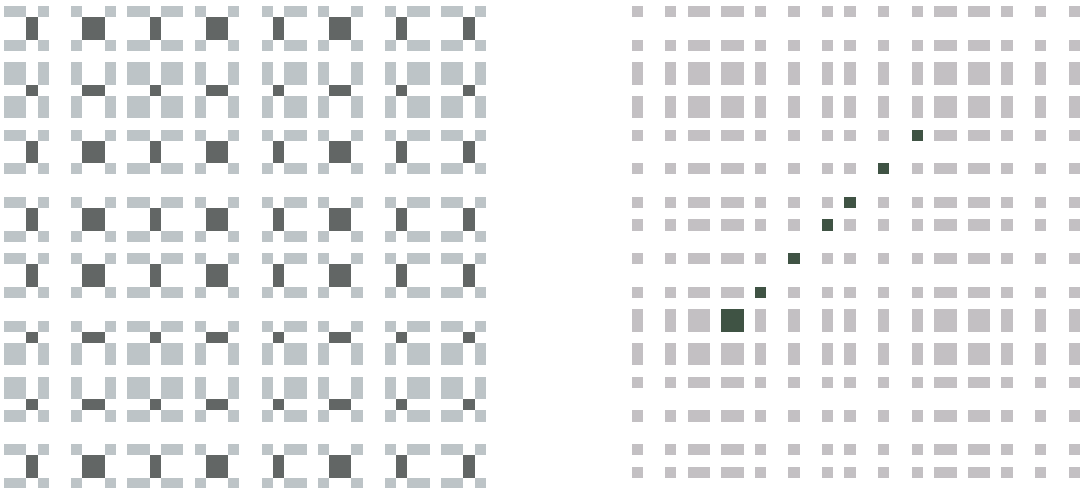


FIGURE 1. Left: lattice topology. Excitatory neurons (pale gray) and inhibitory neurons (dark gray). Right: Representation of excitatory neurons, typical pattern marked on excitatory neurons (dark gray)

THE NEURON LATTICE

The lattice nodes are composed by integrate and fire Izhikevich [8] neurons which are either regular spiking excitatory neurons or fast spiking inhibitory neurons in a proportion of 1 inhibitory to 4 excitatory. Excitatory synaptic currents include both AMPA and NMDA modeling terms and the inhibitory ones contain $GABA_A$ and $GABA_B$ terms. The regular lattice used in the simulations (Fig. 1) has 256 excitatory and 64 inhibitory neurons. In this work we explore two types of connections: i) first neighbors; ii) first and second neighbors.

Presynaptic dependent scaling (PSD) [7] is one of the mechanisms used for synaptic modification. Here it is presented in a continuous version. First we define activity A_i for a neuron i;

$$\tau_A \frac{dA_i}{dt} = (S_i - A_i) \tag{1}$$

where $S_i \rightarrow \sum_k \delta(t - t^k)/t_{max}$ is related to the number of spikes of neuron i in the interval t_{max}. With this definition the weight between the presynaptic neuron i and the postsynaptic neuron j are modified by

$$\tau_w \frac{dW_{ij}}{dt} = \frac{A_i}{A_{GOAL}} \frac{(A_{GOAL} - A_j)}{A_{GOAL}} W_{ij} \ . \tag{2}$$

The second mechanism used for synaptic modification is spike time dependent plasticity (STDP) which increases (decreases) synaptic intensity when the presynaptic neuron fires before (after) the postsynaptic one. This can be modeled [7] (also in a continuous version) by the expression:

$$\frac{dW_{ij}}{dt} = 1 + \sum_{k=1}^{K} \sum_{l=1}^{L} F(t_l^j - t_k^i - \delta_{ij}) \tag{3}$$

where t_l^j is the l^{th} spike of neuron j, δ_{ij} is the synaptic delay and

$$F(\Delta t) = \begin{cases} c_p \exp(-\Delta t/\tau_p), & \Delta t > 0 \\ -c_d \exp(\Delta t/\tau_d), & \Delta t \leq 0 \end{cases} \quad (4)$$

RESULTS AND CONCLUSIONS

We analyzed the two mechanisms for synaptic modification separately. The protocol for the input is to use low frequency (0.2 Hz) injection of current to a line of eight neighboring neurons. We have chosen 4 different directions for these lines: vertical, horizontal and two diagonal lines with $\pm 45^\circ$ (cross pattern). When studying PSD only these currents are presented simultaneously; in the case of STDP they are separated by 6 ms. Yet for STDP, the procedure is repeated until a first neuron reaches the maximal synaptic weight. After that only the first neuron of the line is excited to test if the pattern has been learned. In both cases the initial synaptic weighs were fixed do half the maximal value. The inhibitory weights were kept constant during all simulations.

In the case of PSD being the only mechanism the input protocol is applied constantly at 0.2 Hz and we observe the lattice response. Two distinct behaviors are found for first neighbor synapses: either the initial wave produced by the neurons line induces spiking in some near by neurons and fades out or it produces a wave that propagates by whole system. The parameter governing the transition is the maximal weight. This behavior happens for any of the four excitation lines.

In the case of STDP only the lattice response depends on geometric details: only vertical or horizontal patterns are learned with first neighbor synapses. Obviously diagonal lines cannot be memorized since their neighboring neurons lines are not connected. When considering second neighbors only for excitatory neurons not just these patterns are memorized but also some nearby ones are incorrectly excited. The correct pattern reproduction for some diagonal lines (not all of them) only happens when second neighbors are considered for inhibitory neurons and the patterns are not simultaneously presented. If simultaneous patterns crossing at some point are presented, the part after the crossing is not retrieved. Future work should consider both PSD and STDP and also extend the synapses neighborhood and gradually test the effect of long range connections.

ACKNOWLEDGMENTS

We acknowledge the Brazilian agencies CNPq, CAPES, FAPERGS, PROPESQ/UFRGS for the for the financial support and the CFCIF-UFRGS for the computational resources.

REFERENCES

1. W. McCulloch, and W. Pitts, *B*ull. Math. Biophys. **7**, 115-133 (1943); S. Grossberger, *J*. Stat. Phys. **1**, 319-350 (1969); J.J. Hopfield, *P*roc. Nat. Acad. Sci. USA **8**, 2554-2558 (1982).
2. A. L. Hodgkin, and A. F. Huxley, *J. Physiol.* **117** 500-544 (1952); J. L. Hindmarsh, and R. M. Rose, *P*roc. R. Soc. London, Ser. B **221** 87-102 (1984);
3. C. C. Bell,V.Z. Han, Y. Sugawara, and K. Grant, *Nature* **387** 278-281 (1997).

4. Bi G-Q, and Poo M-M, *J Neurosci* **18** 10464-10472 (1998).
5. W. Gerstner, et al., *Nature* **383**, 76–78 (1996).
6. G. G. Turrigiano, *Trends Neurosci.* **22**, 221–227 (1999).
7. J. K. Liu, and D. V. Buonomano, *J. Neurosci.* **29**, 13172–13181 (2009).
8. E. M. Izhikevich, and G. M. Edelman, *Proc. Natl. Acad. Sci. USA* **105**, 3593–3598 (2008).

Strategies to associate memories by unsupervised learning in neural networks

E. J. Agnes, B. E. P. Mizusaki, R. Erichsen Jr. and L. G. Brunnet

Instituto de Física, Universidade Federal do Rio Grande do Sul, Brazil.

Abstract. In this work we study the effects of three different strategies to associate memories in a neural network composed by both excitatory and inhibitory spiking neurons, which are randomly connected through recurrent excitatory and inhibitory synapses. The system is intended to store a number of memories, associated to spatial external inputs. The strategies consist in the presentation of the input patterns through trials in: i) ordered sequence; ii) random sequence; iii) clustered sequences. In addition, an order parameter indicating the correlation between the trials' activities is introduced to compute associative memory capacities and the quality of memory retrieval.

Keywords: unsupervised learning; spiking neurons; homeostasis; STDP.
PACS: 87.18.Sn, 87.19.lg, 87.19.lj, 87.19.lv, 87.19.lw

INTRODUCTION

Memories, and everything that is processed by the brain, are associated with the connections among neurons. It is well established that learning includes mechanisms based on Hebb's hypothesis [1], which consists basically on potentiating and depressing connections between neurons with correlated and uncorrelated activities, respectively. The main mechanism of synaptic modification that is responsible for the emergence of associative memory in an unsupervised way is Spike-Timing-Dependent Plasticity (STDP) [2, 3], which is a hebbian-like rule. Addionally to the associative memory mechanism, the cells present homeostasis, which has been studied in theoretical works [4, 5].

It is known that, with some predefined connections between neurons and without synaptic plasticity [6, 7], a neural network of spiking neurons has a certain capacity to store memories. But how these specific connection matrices are acquired in an unsupervised way is yet unknown, and here we introduce three distinct methods to do so.

LEARNING

We have used the Izhikevich model [8] with Regular Spiking (RS) parameters for the 320 excitatory neurons and Fast Spiking (FS) for the 80 inhibitory neurons. The learning process was based on the work by Liu and Buonomano [5], where a trial, τ, is defined as the network response after a spatial input. All the synaptic modifications are applied after each trial, since the time window of a trial is less than 150 ms, which could match synaptic plasticity time scales.

As used in ref. [5], the probability of connection from an excitatory to another excitatory neuron, $P(exc. \rightarrow exc.)$, was set to 0.12. For the other connections, we used

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 255-257 (2013); doi: 10.1063/1.4776533

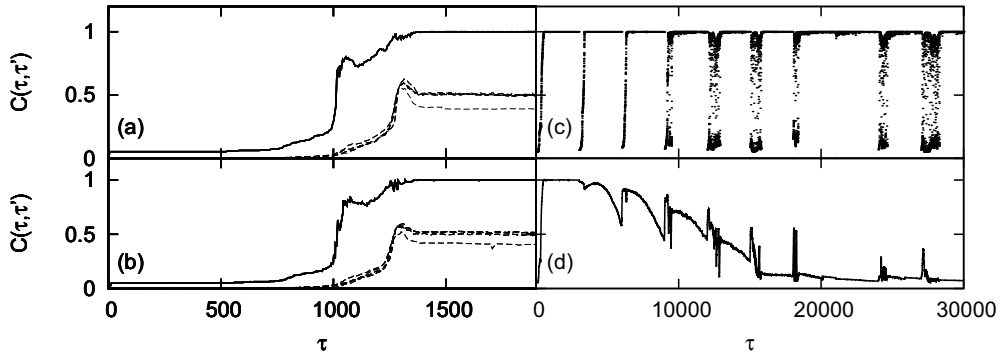


FIGURE 1. Order parameter, $C(\tau,\tau')$, versus learning trials, τ. In (a) and (b), the ordered sequence and the random sequence, respectively. The correlation in the clustered order in (c) and (d). Details of τ' on the text.

$P(exc. \rightarrow inh.) = 0.2$ and $P(inh. \rightarrow exc.) = 0.2$. Each input pattern consists on a randomly chosen set of 20 excitatory and 10 inhibitory neurons that fire within the first 10 ms of the trial. Initial synaptic weights were set to $W_{EE} = 0.04$ nS, $W_{EI} = 0.008$ nS and $W_{IE} = 0.1$ nS. Excitatory synapses were modified according to both homeostatic and STDP rules used in ref. [5], with the same parameter values. Short-term plasticity (STP) was implemented as described in ref. [8], with depression in all synapses.

The analysis was done with an order parameter which correlates the spike times of the excitatory neurons in two distinct trials. It is defined as

$$C(\tau,\tau') = \gamma^{-1} \sum_{i=1}^{320} \sum_{\{k,j\}} \exp\left[\frac{-(t_{i_k}^{\tau} - t_{i_j}^{\tau'})^2}{100}\right], \tag{1}$$

where $\gamma = MAX(320, S^{\tau}, S^{\tau'})$ is the normalization for the correlation value and $S^{\tau} = (1/320)\sum_i^{320} S_i^{\tau}$, a sum over only the 320 excitatory neurons. The sum over $\{k,j\}$ indicates that kth and jth spikes are close in time.

We used three different learning methods of spatial input presentation: i) ordered sequence; ii) random sequence; iii) clustered sequences. In the first strategy, each spatial pattern is presented in a ordered way through the trials. The second strategy consists in presenting the patterns in a random sequence and, in the third one, each pattern is presented successively for a defined number of trials, which was set to 3000 trials. The synaptic modifications are applied to excitatory connections, including a homeostatic plasticity and STDP, as described in ref. [5]. The homeostasis rule is used to increase the synaptic weights until the network presents a desired activity and thereafter to maintain a stable activity.

Figure 1 shows the evolution of the correlation $C(\tau,\tau')$, where $\tau' = \tau_{\mu=1}$ in (a) and (b) and $\tau' = \tau - 1$ in (c). A trial $\tau_{\mu=1}$ is the last trial - before τ - that the pattern $\mu = 1$ was the input pattern. In Fig. 1(d), the plot is the correlation of a trial between τ and $\tau + 1$ with the presentation of the pattern $\mu = 1$ without synaptic plasticity from homeostasis and STDP rules and the last trial with the presentation of pattern $\mu = 1$ with synaptic

plasticity. For $\tau < 3000$, the curve is the same as in Fig. 1(c) and for $\tau > 3000$, pattern $\mu = 2$ begins to be trained and the curve informs how much information about pattern $\mu = 1$ is being forgotten.

Full lines in graphs 1(a) and (b) correspond to τ which has as spatial input the pattern $\mu = 1$, indicating correlation when the same input is presented at different times. These lines converge to $C = 1$, showing that the spatiotemporal response from the network is equivalent when the same input is used. We tested the same with the other 4 patterns and the result is equivalent. Dashed lines correspond to correlation between trials with different input patterns and they do not converge to $C = 1$, indicating that different trained inputs evoke distinct spatiotemporal responses.

The plot in Fig. 1(c) shows that, using 30 patterns, the correlation converges to $C = 1$ for each one of them within a learning window of 3000 trials, but, as shown in Fig. 1(d), the network response to a learned memory vanishes as new memories begin to be stored.

CONCLUSIONS

We presented a neural network with unsupervised learning, simulated using the Izhikevich model and synaptic plasticity applied in trials. We defined three different ways of learning through trials - ordered, random and clustered sequences - and showed that the two first methods are equivalent for 5 patterns while the last method is robust for a large number of patterns but it presents a forgetting curve, which means that the memories are forgotten when new ones are trained. When the number of stored patterns increase in random and ordered sequences, some inputs do not develop spatiotemporal response (not shown). A detailed study of the network's behavior when increasing the number of input patterns is needed for a more accurate understanding of the problem.

ACKNOWLEDGMENTS

The authors acknowledge Brazilian funding agencies CNPq, Capes and Fapergs, also the Centro de Física Computacional - IF - UFRGS (CFCIF).

REFERENCES

1. J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the theory of neural computation.* Addison-Wesley, Boston, 1991.
2. W. Gerstner, *et al.*, *Nature* **383**, 76–78 (1996).
3. C. Q. Bi, and M. M. Poo, *J. Neurosci.* **18**, 10464–10472 (1998).
4. G. G. Turrigiano, *Trends Neurosci.* **22**, 221–227 (1999).
5. J. K. Liu, and D. V. Buonomano, *J. Neurosci.* **29**, 13172–13181 (2009).
6. E. J. Agnes, R. Erichsen Jr., and L. G. Brunnet, *Phys. A* **391**, 843–848 (2012).
7. E. J. Agnes, R. Erichsen Jr., and L. G. Brunnet, *LNCS* **7552**, 145–152 (2012).
8. E. M. Izhikevich, and G. M. Edelman, *Proc. Natl. Acad. Sci. USA* **105**, 3593–3598 (2008).

Invariance of covariances arises out of noise

D. Grytskyy*, T. Tetzlaff*, M. Diesmann*,† and M. Helias*

*Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Jülich Research Center, Germany.
†Medical Faculty, RWTH Aachen University, Germany.

Abstract. Correlated neural activity is a known feature of the brain [1] and evidence increases that it is closely linked to information processing [2]. The temporal shape of covariances has early been related to synaptic interactions and to common input shared by pairs of neurons [3]. Recent theoretical work explains the small magnitude of covariances in inhibition dominated recurrent networks by active decorrelation [4, 5, 6]. For binary neurons the mean-field approach takes random fluctuations into account to accurately predict the average activity in such networks [7] and expressions for covariances follow from a master equation [8], both briefly reviewed here for completeness. In our recent work we have shown how to map different network models, including binary networks, onto linear dynamics [9]. Binary neurons with a strong non-linear Heaviside gain function are inaccessible to the classical treatment [8]. Here we show how random fluctuations generated by the network effectively linearize the system and implement a self-regulating mechanism, that renders population-averaged covariances independent of the interaction strength and keeps the system away from instability.

Keywords: covariances; linearization by noise; spectral radius; chaos.
PACS: 87.19lj, 87.19ll, 87.19ln

A binary neuron has two states, 0 and 1, representing inactivity and activity, respectively. The model used here has stochastic transitions between these two states happening at random points in time controlled by the transition rates. The state space of a network of N such neurons is described by a binary vector $\mathbf{n} = (n_1, \ldots, n_N) \in \{0,1\}^N$, illustrated for $N = 2$ in Fig. 1B. The rate of an up transition in the i-th neuron is given by the gain function $\frac{1}{\tau}F_i(\mathbf{n})$ depending on the activity of all neurons providing synaptic input to neuron i, for a down transition it is $\frac{1}{\tau}(1 - F_i(\mathbf{n}))$ as shown in Fig. 1A. We denote as $\mathbf{n}_{i+} = (n_1, \ldots, n_i = 1, \ldots, n_N)$ the state with the active i-th neuron ($n_i = 1$), and as $\mathbf{n}_{i-}$ if it is inactive ($n_i = 0$). In a stationary state, the rates entering and leaving each state must sum to zero, leading to the master equation

$$0 \;=\; \sum_{i=1}^{N} \underbrace{(2n_i - 1)}_{\text{direction of flux}} \left(p(\mathbf{n}_{i-})F_i(\mathbf{n}_{i-}) - p(\mathbf{n}_{i+})(1 - F_i(\mathbf{n}_{i+}))\right) \quad \forall \mathbf{n} \in \{0,1\}^N, \quad (1)$$

illustrated in Fig. 1B. From (1) follows the expected activity, the first moment $\langle n_k \rangle$ of neuron k by multiplying both sides with n_k and summing over all possible states $\mathbf{n} \in \{0,1\}^N$. Only the term with $n_i = n_k$ remains, because for all other terms in the sum the configuration with $n_k = 0$, so that $2n_k - 1 = -1$, cancels the another configuration

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 258-262 (2013); doi: 10.1063/1.4776531

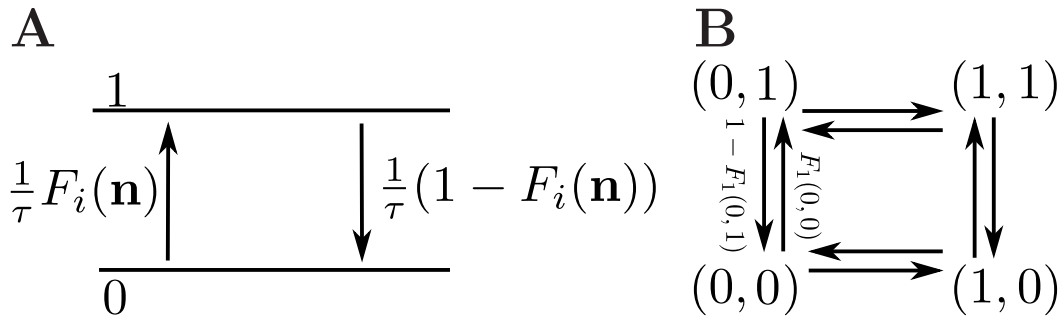


Figure 1. **(A)** Binary states of a neuron i. A transition from state 0 to state 1 happens with rate $\frac{1}{\tau}F_i(\mathbf{n})$, a transition from state 1 to state 0 takes place with rate $\frac{1}{\tau}(1-F_i(\mathbf{n}))$. Transition rates depend on the state all neurons that provide incoming synaptic connections to neuron i. **(B)** State space of a network of two neurons described by two numbers $\mathbf{n}=(n_1,n_2)$ that take values $\{0,1\}$ each. Stationarity requires a vanishing sum of fluxes (arrows) entering and leaving each state, leading to constant occupation probability $p(\mathbf{n})$ formally expressed by (1).

with $n_k = 1$, so that $2n_k - 1 = 1$, leaving us with

$$0 = \sum_{\mathbf{n}} n_k \left(p(\mathbf{n}_{k-}) F_k(\mathbf{n}_{k-}) - p(\mathbf{n}_{k+})(1 - F_k(\mathbf{n}_{k+})) \right) \tag{2}$$

$$\langle n_k \rangle = \sum_{\mathbf{n}\setminus n_k} p(\mathbf{n}_{k+}) = \sum_{\mathbf{n}\setminus n_k} p(\mathbf{n}_{k-}) F_k(\mathbf{n}_{k-}) + p(\mathbf{n}_{k+}) F_k(\mathbf{n}_{k+}) = \sum_{\mathbf{n}} F_k(n)\, p(\mathbf{n}) = \langle F_k(\mathbf{n}) \rangle.$$

The correlation between neuron k and neuron l is defined as $\bar{c}_{kl} = \langle n_k n_l \rangle = \sum_{\mathbf{n}} p(\mathbf{n}) n_k n_l$. It is large if both neurons are frequently activated together, the neurons are said to be positively correlated. If activated independently, the correlation equals $\bar{c}_{kl} = \langle n_k \rangle \langle n_l \rangle$. We determine the correlations in the network from (1), multiplying both sides by $n_k n_l$ and summing over all possible states $\mathbf{n} \in \{0,1\}^N$ to obtain

$$\begin{aligned} 0 &= \sum_{\mathbf{n}\setminus n_k} \left(p(\mathbf{n}_{k-}) F_k(\mathbf{n}_{k-}) - p(\mathbf{n}_{k+})(1 - F_k(\mathbf{n}_{k+})) \right) n_l \\ &+ \sum_{\mathbf{n}\setminus n_l} \left(p(\mathbf{n}_{l-}) F_l(\mathbf{n}_{l-}) - p(\mathbf{n}_{l+})(1 - F_l(\mathbf{n}_{l+})) \right) n_k. \end{aligned}$$

As before, only the terms containing n_k or n_l remain, because all other terms in the sum appear twice with opposite signs. With $\bar{c}_{kl} = \sum_{\mathbf{n}\setminus n_k} p(\mathbf{n}_{k+}) n_l$ rearranging terms results in

$$2\bar{c}_{kl} = \langle F_k(\mathbf{n}) n_l \rangle + \langle F_l(\mathbf{n}) n_k \rangle. \tag{3}$$

Often just the fluctuations around the mean value are of interest, motivating the definition of the covariance c as

$$c_{kl} = \bar{c}_{kl} - \langle n_k \rangle \langle n_l \rangle = \frac{1}{2} \langle F_k(\mathbf{n}) \delta n_l \rangle + \frac{1}{2} \langle F_l(\mathbf{n}) \delta n_k \rangle, \tag{4}$$

where $\delta n_i = n_i - \langle n_i \rangle$. This equation has a simple interpretation. The right hand side measures the influence of neuron l's fluctuations around the mean activity on the transition probability of neuron k and vice versa.

MEAN FIELD SOLUTION

We will now consider a randomly connected recurrent network of N_E excitatory and N_I inhibitory neurons with $N_I/N_E = \gamma$ and $N = N_E + N_I$. Each neuron has K excitatory and γK inhibitory randomly drawn input connections, each with amplitude J and $-gJ$ respectively. We assume $F_k(\mathbf{n}) = H(h_k - \theta)$, where H is the Heaviside-function and $h_k = \sum_{l=1}^{N} J_{kl} n_l$ the summed input to neuron k. J_{kl} is the synaptic weight from neuron l to neuron k and θ the threshold. A binary state is either 1 or 0, so $n^2 = n$ and the second moment that describes the strength of fluctuations is $\langle n_k^2 \rangle = \sum_{\mathbf{n}} p(\mathbf{n}) n_k^2 = \sum_{\mathbf{n}} p(\mathbf{n}) n_k = \langle n_k \rangle$. The variance hence is $a_k \stackrel{\text{def}}{=} \langle n_k^2 \rangle - \langle n_k \rangle^2 = (1 - \langle n_k \rangle)\langle n_k \rangle$ determined by the mean. For homogeneous connectivity we can assume that all neuron's average activities are well described by a single mean $\langle n \rangle$. The average input to each neuron then is

$$\langle h_k \rangle \;=\; \sum_l J_{kl} \langle n \rangle = KJ(1 - \gamma g)\langle n \rangle \stackrel{\text{def}}{=} \mu.$$

Under the assumption of sufficiently irregular network activity, we can further assume that the neurons approximately act independently, so that their variances add up to the variance of the total input h_k

$$\langle h_k^2 \rangle - \langle h_k \rangle^2 \;=\; \sum_l J_{kl}^2 (1 - \langle n \rangle)\langle n \rangle = KJ^2(1 + \gamma g^2)(1 - \langle n \rangle)\langle n \rangle \stackrel{\text{def}}{=} \sigma^2. \qquad (5)$$

As h_k is a sum of typically thousands of synaptic inputs, to good approximation it follows as Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2. We are now ready to calculate the mean activity in the network [10]

$$\langle n \rangle \;=\; \langle F(\mathbf{n}) \rangle \simeq \int_{-\infty}^{\infty} H(x - \theta)\, \mathcal{N}(\mu, \sigma^2, x)\, dx = \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{\theta - \mu(\langle n \rangle)}{\sqrt{2}\sigma(\langle n \rangle)}\right)\right). \qquad (6)$$

This equation needs to be solved self-consistently, because μ and σ depend on $\langle n \rangle$ themselves. Figure 2A illustrates the graphical solution.

COVARIANCES IN THE RECURRENT NETWORK

Next we need to calculate the covariances in the recurrent network. We proceed along similar lines as before. We start with equation (4) and apply a linearization to the two terms of the form $\langle F_k(\mathbf{n}) \delta n_l \rangle$. In the recurrent network, the activities of pairs of neurons may be correlated. Therefore, the input h_k to neuron k not only depends on n_l directly, but also indirectly through the covariances of n_l with any of the other neuron n_i that projects to k. Taking this dependence into account in the linearization we obtain

$$\begin{aligned}
\langle F_k(\mathbf{n}) \delta n_l \rangle \;&=\; \langle H(h_k) \delta n_l \rangle = \sum_i \langle H(h_{k \setminus n_i} + J_{ki} - \theta) n_i \delta n_l + H(h_{k \setminus n_i} - \theta)(1 - n_i)\delta n_l \rangle \\
&\simeq\; \sum_i \langle H(x + J_{ki}) - H(x) \rangle_x \underbrace{\langle n_i \delta n_l \rangle_{\mathbf{n}}}_{=c_{il}} + \langle H(x) \rangle_x \underbrace{\langle \delta n_l \rangle_{\mathbf{n}}}_{=0} \simeq S(\mu, \sigma) \sum_j J_{ki} c_{il},
\end{aligned}$$

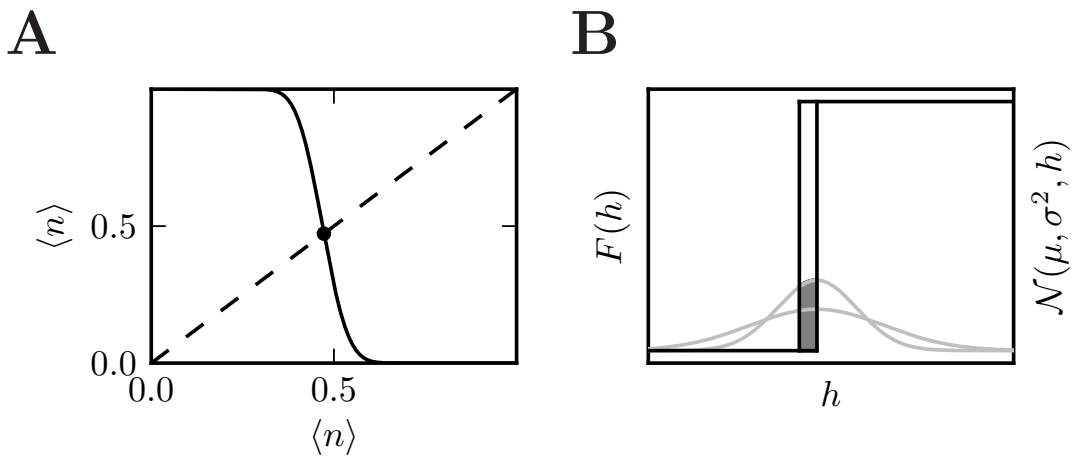


Figure 2. **A** Graphical solution of the mean field equation (6) for $\langle n \rangle$. **B** Distributions (light gray) of input h for different J, and corresponding σ . An additional input causes a left-shift of the gain function F (black), the gray area is proportional to the susceptibility S.

where we introduced the susceptibility $S(\mu,\sigma) = \frac{\partial}{\partial \varepsilon}|_{\varepsilon=0}\langle H(x+\varepsilon) - H(x)\rangle_x = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mu-\theta)^2}{2\sigma^2}}$ given the synaptic weight J_{ki} is small compared to the total fluctuations of h_k. In the second line we separated $h_k - \theta = x + Jn_j$ into the fluctuations x distributed as $\mathcal{N}(\mu-\theta,\sigma^2)$, and assumed to be independent of the state of n_l and those fluctuations that either depend on or covary with n_l . The linearized (4) turns into

$$c_{kl} \;=\; \frac{S(\mu,\sigma)}{2} \sum_j \left(J_{kj} c_{jl} + J_{lj} c_{jk} \right).$$

We now take advantage of the random structure of the network and replace each pairwise covariance c_{ij} by the average value over many pairs of neurons. Here we distinguish the excitatory ($\mathscr{E}$) and inhibitory ($\mathscr{I}$) neuron types ()and define $c_{\mathscr{E}\mathscr{E}} = \frac{1}{N_e^2}\sum_{i\neq j\in\mathscr{E}} c_{ij}$, $c_{\mathscr{I}\mathscr{I}} = \frac{1}{N_i^2}\sum_{i\neq j\in\mathscr{I}} c_{ij}$, $c_{\mathscr{E}\mathscr{I}} = c_{\mathscr{I}\mathscr{E}} = \frac{1}{N_e N_i}\sum_{i\in\mathscr{E}, j\in\mathscr{I}} c_{ij}$. The variances c_{ii} of the binary variables are fixed by their mean $c_{ii} = a = \langle n\rangle(1-\langle n\rangle)$, as shown before. Replacing the individual covariances by the respective mean and counting the number of connections between neurons we arrive at a set of linear equations

$$\left[\mathbf{1} - \frac{1}{2} q \begin{pmatrix} 2-\gamma g & -\gamma g \\ 1 & 1-2\gamma g \end{pmatrix}\right] \begin{pmatrix} c_{\mathscr{E}\mathscr{E}} \\ c_{\mathscr{I}\mathscr{I}} \end{pmatrix} \;=\; \frac{qa}{N} \begin{pmatrix} 1 \\ -g \end{pmatrix} \tag{7}$$

$$c_{\mathscr{E}\mathscr{I}} \;=\; c_{\mathscr{I}\mathscr{E}} = \frac{1}{2}\left(c_{\mathscr{E}\mathscr{E}} + c_{\mathscr{I}\mathscr{I}}\right),$$

which can be solved for c with elementary methods. Figure 3B compares this solution to direct simulation. The parameter $q = KJS(\mu,\sigma)$ scales the effective linearized cou-

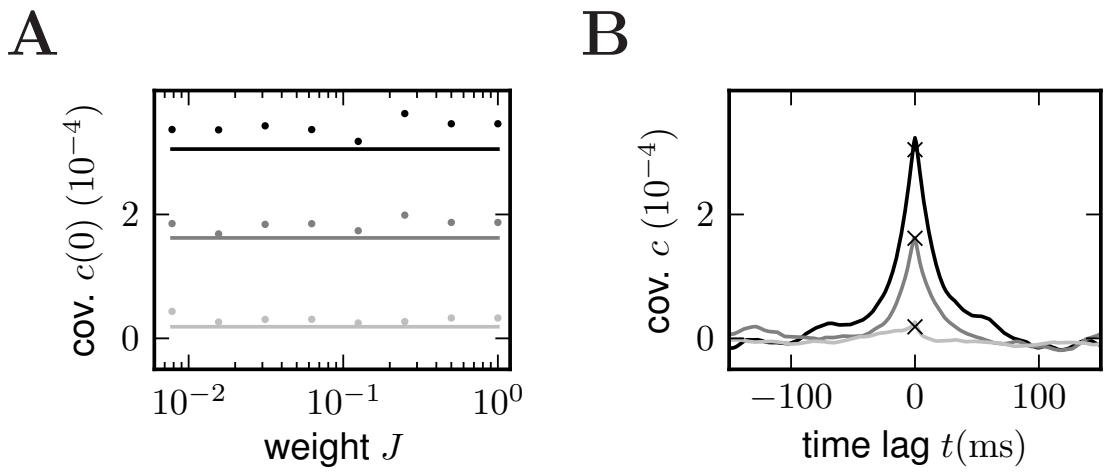


Figure 3. **A** Cross covariance $c_{\mathscr{E}\mathscr{E}}$ (black), $c_{\mathscr{E}\mathscr{I}}$ (gray) and $c_{\mathscr{I}\mathscr{I}}$ (light gray) for zero time-lag simulated (dots) and predicted (lines) by theory (7). **B** Cross covariance functions averaged over 10^6 pairs of excitatory neurons (black), inhibitory neurons (light gray) and pairs of one excitatory, one inhibitory neuron (lgray). The crosses mark the analytical expectation (7).

pling and $q = KJ\frac{1}{\sqrt{2\pi}\sigma}\underbrace{e^{-\frac{(\mu-\theta)^2}{2\sigma^2}}}_{\leq 1} \leq \sqrt{\frac{K}{2\pi(1+\gamma g^2)(1-\langle n\rangle)\langle n\rangle}}$ is bounded. If $\langle n\rangle$ is held constant at 0.5, choosing the threshold appropriately, $q \leq \frac{\sqrt{2K}}{\sqrt{\pi(1+\gamma g^2)}}$. For non-saturated activity $0 < \langle n\rangle < 1$ it follows that $\mu \simeq \theta$ so q is close to the maximum allowed by the inequality. Moreover, the covariance is almost independent of J, as shown in Fig. 3A, because $\sigma \propto J$ (5) and the peak of the input distribution decreases with σ (Figure 2B), so $S \propto \sigma^{-1}$. This self-regulating mechanism also preserves the system from the instability induced by modes corresponding to eigenvalues of the effective connectivity matrix with real part more than one: the spectral radius ρ containing the eigenvalues of the linearized random connectivity matrix [11, 12] is bounded as $\rho = \frac{q}{K}\sqrt{Np(1-p)(1+\gamma g^2)} \leq \sqrt{\frac{2(1-p)}{\pi}} < 1$.

REFERENCES

1. G. L. Gerstein, and D. H. Perkel, *Science* **881**, 828–830 (1969).
2. M. R. Cohen, and A. Kohn, *Nat. Rev. Neurosci.* **14**, 811–819 (2011).
3. D. H. Perkel, G. L. Gerstein, and G. P. Moore, *Biophys. J.* **7**, 419–440 (1967).
4. J. Hertz, *Neural Comput.* **22**, 427–447 (2010).
5. A. Renart, J. De La Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, *Science* **327**, 587–590 (2010).
6. T. Tetzlaff, M. Helias, G. Einevoll, and M. Diesmann, *PLoS Comput. Biol.* **8**, e1002596 (2012).
7. C. van Vreeswijk, and H. Sompolinsky, *Science* **274**, 1724–1726 (1996).
8. I. Ginzburg, and H. Sompolinsky, *Phys. Rev. E* **50**, 3171–3191 (1994).
9. D. Grytskyy, T. Tetzlaff, M. Diesmann, and M. Helias, *BMC Neurosci.* **13**, P147 (2012).
10. C. van Vreeswijk, and H. Sompolinsky, *Neural Comput.* **10**, 1321–1371 (1998).
11. K. Rajan, and L. Abbott, *Phys. Rev. Lett.* **97**, 188104 (2006).
12. B. Kriener, T. Tetzlaff, A. Aertsen, M. Diesmann, and S. Rotter, *Neural Comput.* **20**, 2185–2226 (2008).

Dimensionality reduction of dynamical systems with parameters

Ch. Welshman* and J. Brooke†

*CICADA, School of Mathematics, University of Manchester, Oxford Road, Manchester, M13 PL, UK.
†School of Computer Science, University of Manchester, Oxford Road, Manchester, M13 PL, UK.

Abstract. We describe a method for reproducing the dynamical behaviour observed in systems of very high dimension in a state space of much lower dimension. The method is designed for systems where the solution evolves onto an attractor of dimension m which is much lower than that of the state space of the full system, n. Whitney's embedding theorem guarantees that the attractor can be embedded in a space of dimension $d = 2m+1$. We describe how such methods can be extended to reproducing the vector field on the attractor so that the dynamics of a parameterized family of attractors can be explored in the low dimensional space $\mathbb{R}^d$.

Keywords: dynamical systems; bifurcations; dimension reduction.
PACS: 05.45Gg, 05.45Jn, 05.45Pq

INTRODUCTION

There is considerable interest in the idea that networks of very large numbers of neurons can exhibit dynamics that can be described by a manifold of very low dimension. We present a mathematical method that involves obtaining a low dimensional description of a system that is originally specified in high dimensional terms. Unlike classic methods of data reduction, which consider data alone, we are attempting to reproduce the dynamics of the high dimensional system in a much lower dimensional ambient space so that it can be more easily studied. There are mathematical theorems (principally Whitney's embedding theorem) which guarantee that, if the high dimensional system produces dynamics which can be represented on a low dimensional manifold, the dynamics can be reproduced in a space of a dimension comparable to this manifold [1]. Models of dynamics often involve parameters representing physical processes. Thus we consider dynamical systems with parameters, where the parameters index a family of vector fields which produce a corresponding family of attractors. Existing methods of dimensionality reduction typically either do not consider the inclusion of parameters, or they deal with control inputs, where the focus is on preserving input-output behaviour, rather than geometric structures in the state space. Although the attractors can change significantly with respect to the parameters, the underlying vector field is often smoothly dependent on the parameters. We can take advantage of this, to produce a low-dimensional family of vector fields that reproduce the corresponding attractors which are indexed by a parameterization from the original parameter space.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 263-266 (2013); doi: 10.1063/1.4776530

STANDARD METHODS OF DIMENSIONALITY REDUCTION

The standard approach to this problem is to obtain a projection onto a suitable linear subspace of the state space, which is then used to determine a corresponding 'reduced' ODE that describes the dynamics in the subspace. A sample of points from an orbit of the system (referred to as 'snapshots') are often used to describe the attractor, which allows for numerical methods to be used in finding a suitable projection. Once a projection has been found, the reduced dynamic is obtained by algebraically manipulating the original differential equation to produce a reduced ODE. Let $x \in X$ be the state variable with dynamics given by $\dot{x} = f(x)$, and $P : X \to X$ be a projection. We can write $x = Px + Rx$, where $R = \mathrm{id}_X - P$. Applying this to the original dynamic gives $P\dot{x} = Pf(Px + Rx)$. The *Galerkin* approach is to choose a projection such that $Rx = 0$ on the attractor, i.e. to project onto the subspace which the orbit explores, $Px(t) = x(t)$. In practice one seeks the lowest-dimensional subspace such that this is approximately true. A popular method of finding such a projection is the proper orthogonal decomposition (POD) (also known as the Karhunen-Loève decomposition), which finds an orthogonal projection that minimises the mean of $\|Rx\|^2$ over the data set. This can be performed by use of a singular value decomposition (SVD), which has the benefit of also providing the information necessary to evaluate how many dimensions are required for the subspace via the singular values [1, 2]. For a general subspace, the residual may be non-zero and the inverse projection is nonlinear. Methods that attempt to describe this nonlinear inverse are sometimes called nonlinear Galerkin methods, or approximate inertial manifolds [3, 4].

A subspace can be described as the image of a linear embedding of a vector space, $\hat{X}$, of appropriate dimension, given by $W : \hat{X} \to X$. The inner product on X can be pulled-back to $\hat{X}$, giving $W^\dagger W = \mathrm{id}_{\hat{X}}$, where $\dagger$ is the adjoint. The orthogonal projection onto the subspace $W(\hat{X})$ is then given by $P = WW^\dagger$. This allows the dynamics in the subspace to be described in $\hat{X}$ as $\dot{\hat{x}} = \hat{f}(\hat{x}) := W^\dagger f(W\hat{x})$, where $\hat{x} = W^\dagger x$. However, although this equation specifies a low-dimensional vector field on $\hat{X}$, in general it does so in high-dimensional terms: it requires evaluation of the high-dimensional nonlinear vector field. In contrast, a bottom-up approach can be taken by constructing an approximation of $\hat{f}$ directly, which bypasses the need for restrictions on the inverse and the form of f. A recent method of this type is the 'discrete Empirical Interpolation Method' [5] (discrete EIM), which is an adaptation of EIM [6] for the finite-dimensional case, which is used to approximate the (nonlinear part of) $\hat{f}$.

GEOMETRIC APPROACHES

However we now consider methods can be regarded as geometrical in inspiration because they utilise the geometrical properties of the attractor, e.g. the set of all secants between points of the attractor, as the basis for the projection methods. Projections of this type can be found in Broomhead and Kirby [7][8], who also use radial basis functions to reproduce $\hat{f}$ and its derivatives on the attractor.

We wish to determine a suitable dimension-reducing map, $W^\dagger$, that can be used to reconstruct, in a low dimensional space, the dynamics occurring in the high dimensional space. It is critical to this method that the dynamics in the high dimensional space

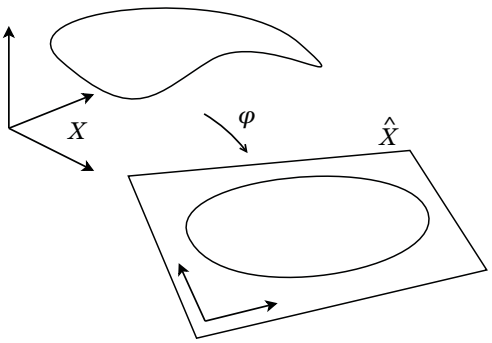


FIGURE 1. The projection of a limit cycle from a high to a low dimensional state space.

is actually evolving on an attractor of much lower dimension embedded in the high dimensional space. We show this in Fig. 1 . We determine the orthogonal projection described by W using data points from the attractor by performing optimization over the Grassman manifold. To construct a cost function for the optimisation, we consider the set of all pairs of points on the attractor. Each distinct pair of points (x,y) generate a *secant*, $(x-y)$, which we normalise to get a unit secant, $(x-y)/\|x-y\|$. In a realisation of the method on a computer we can only consider a subset of the set of all pairs of points, determining what is a sufficiently representative subset is empirically determined and depends on the complexity of the structure of the attractor. Let $\mathscr{K}$ be the set of unit secants generated by points on the attractor. The cost function is then

$$\mathscr{F}(W) = \frac{1}{|\mathscr{K}|}\sum_{k\in\mathscr{K}}\left\|W^{\dagger}k\right\|^{-1}. \tag{1}$$

The cost function aims to preserve the lengths of the projected secants, so that distinct points are not projected on top of each other. This makes the inverse well-conditioned.

The basis of this method is the Whitney embedding theorem, which states that an m-dimensional manifold can always be embedded into $\mathbb{R}^{2m+1}$. Thus in the case of our limit cycle ($m=1$), this can be embedded in a three dimensional space, no matter how high the dimension of the original state space in which the limit cycle evolves. A consequence of the Whitney embedding theorem is that the set of bad projections (those resulting in null projected secants) is *nowhere dense* in the Grassman manifold, $Gr_d(\mathbb{R}^n)$, as long as the dimension d is 'large enough' to contain the attractor ($d \geq 2m+1$). This guarantee gives the method robustness. In practice we may be able to do better than $2m+1$, depending on the example.

INCLUDING PARAMETER DEPENDENCE

A significant complication to this problem is the introduction of parameters, $\dot{x} = f(x;\lambda)$. In control engineering, dimensionality reduction goes by the name *model order reduction*; however, in this field the parameters are usually dynamically-varying control inputs, and the objective of the reduction is to preserve input-output behaviour [9], rather

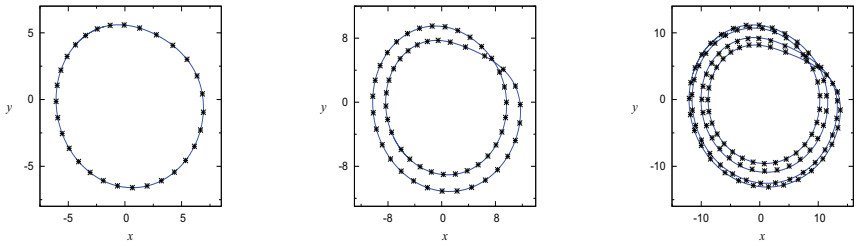


FIGURE 2. Dynamics of the Rössler equations reconstructed using the methods described in the text, showing two period doubling bifurcations of a limit cycle as a control parameter is varied.

than particular geometric objects in the state space. In contrast to this, we consider static parameters, where the objective is to describe a family of attractors in the state space; an autonomous view is taken without any inputs or outputs. Such parameters are studied in the theory of bifurcations and they are also used to steer numerical experiments to direct the computation to regions of parameter space where the solutions are of physical interest (e.g. [10]. An early result of our work is shown in Fig. 2 where we show a sequence containing two period-doubling bifurcations in the solutions of the Rössler system [11].

ACKNOWLEDGMENTS

Chris Welshman was supported under EPSRC grant EP/C536452/1 by a CICADA PhD studentship. This research was funded in part by EPSRC grant EP/E050441/1.

REFERENCES

1. M. Hinze, and S. Volkwein, "Proper Orthogonal Decomposition Surrogate Models for Nonlinear Dynamical Systems: Error Estimates and Suboptimal Control," in *Dimension Reduction of Large-Scale Systems*, edited by P. Benner, D. C. Sorensen, V. Mehrmann, T. J. Barth, M. Griebel, D. E. Keyes, R. M. Nieminen, D. Roose, and T. Schlick, Springer Berlin Heidelberg, 2005, vol. 45 of *Lect. Not. Comput. Sci. & Eng.*, pp. 261–306.
2. B. Noack, M. Schlegel, M. Morzynski, and G. Tadmor, "Galerkin Method for Nonlinear Dynamics," in *Reduced-Order Modelling for Flow Control*, edited by B. Noack, M. Morzynski, and G. Tadmor, Springer Vienna, 2011, vol. 528 of *CISM Courses and Lectures*, pp. 111–149, ISBN 978-3-7091-0757-7.
3. G. Rega, and H. Troger, *Nonl. Dynamics* **41**, 1–15 (2005).
4. A. Steindl, and H. Troger, *Int. J. Sol. & Struct.* **38**, 2131–2147 (2001).
5. S. Chaturantabut, and D. Sorensen, *SIAM J. Scient. Comp.* **32**, 2737–2764 (2010).
6. M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, *Comptes Rendus Mathematique* **339**, 667–672 (2004).
7. D. S. Broomhead, M. Kirby, *SIAM J. Appl. Math.* **60**, 2114–2142 (2000).
8. D. S. Broomhead, M. Kirby, *Nonl. Dynamics* **41**, 47–67 (2005).
9. A. C. Antoulas, D. C. Soresnsen, S. Gugercin, *Contemp. Math.* **280**, 193–219 (2001).
10. A. Salhi, and J. M. Brooke, *J. Phys.: Conf. Ser.* **216**, 012003 (2010).
11. O. E. Rössler, *Phys. Lett. A* **57**, 397–398 (1976).

The Neurona@Home project: Simulating a large-scale cellular automata brain in a distributed computing environment

L. Acedo*, J. Villanueva-Oller†, J. A. Moraño** and R.-J. Villanueva**

*Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Building 8G, 2° Floor, 46022, Valencia, Spain, e-mail: luiacrod@imm.upv.es .
†CES Felipe II, Universidad Complutense de Madrid, Aranjuez, Spain.
**Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, Building 8G, 2° Floor, 46022, Valencia, Spain.

Abstract. The Berkeley Open Infrastructure for Network Computing (BOINC) has become the standard open source solution for grid computing in the Internet. Volunteers use their computers to complete an small part of the task assigned by a dedicated server. We have developed a BOINC project called Neurona@Home whose objective is to simulate a cellular automata random network with, at least, one million neurons. We consider a cellular automata version of the integrate-and-fire model in which excitatory and inhibitory nodes can activate or deactivate neighbor nodes according to a set of probabilistic rules. Our aim is to determine the phase diagram of the model and its behaviour and to compare it with the electroencephalographic signals measured in real brains.

Keywords: BOINC project; cellular Automata; electroencephalography.
PACS: 87.18.Sn, 87.19.le, 89.20.Hh

INTRODUCTION

The idea that the brain is a network dates back to the foundational works on Neuroscience by Santiago Ramón y Cajal in the XIXth century and the early XXth century [1]. Although this concept was not developed at that time, it has become clear since the 70s of the past century that the brain is the most complex example of the network paradigm. In the human brain, neurons project an average of 10,000 synapses to their neighbours and this fact only illustrates the topological complexity of this biological system [2].

On the other hand, we have the Cellular Automata family of mathematical models. These models are composed by a set of units called automatons, each of them can be found in different states. The dynamical evolution of the state of every automaton is determined by rules that usually depend on the states of the automatons in their immediate neighbourhood. The patterns arisen from these models are very complex in many cases despite the simplicity of the underlying rules [3]. Usually, the rules are deterministic but we can define stochastic cellular automata as well.

In our project we have combined the structural complexity of random networks with stochastic evolution rules in order to build a Cellular Automata Model of the brain. In order to simulate a sufficiently large number of automata neurons, around 1,000,000, we need a distributed computing solution because the amount of computational work is

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 267-270 (2013); doi: 10.1063/1.4776528

vast.

This number of neurons is still very small in comparison with the number of neurons in the human brain but it could serve as a model for the complex mini-brains of some insects: honeybee brains contains around 960,000 neurons which enable them with a memory and behavior repertoire astounding for their size. In these insect brains, the synchronization in the firing of large assemblies of neurons are related to odour discrimination [4] while in humans they have been related to higher-order functions such as attention, memory and conscious awareness [5]. It is remarkable that we find both synchronous oscillations and quasi-gaussian noise in our model. In the next section we describe the BOINC approach to grid computing and some preliminary results are given in the Results section.

THE BOINC ENVIRONMENT

The BOINC open source grid computing environment is a software developed at the University of Berkeley in 2002 [6] to tap into the immense computing capacity of the Internet. It can be considered an evolution of the famous SETI@Home project in which radio signals from many stars in our galaxy are analyzed in the hope of detecting intelligent activity.

BOINC is particularly concerned with security issues, specially after security breaches were exploited by users of SETI@Home. The BOINC concept is classical in grid computing and it is described by the following elements:

- A client software is installed in every individual computer. This client request tasks from the server and manages the completion of them for the project using the local CPU and even GPU (graphic processor unit).
- The scheduling server sends the task and takes into account that a particular computer is capable of managing them by considering its amount of RAM, etc
- The PC reports the completed task to the server, the output files are finally sent and it receives more tasks.

A system of credits was also developed to control the amount of work performed by every computer connected to the project server. In the last decade many researchers have developed their BOINC projects for a variety of purposes from Medicine to Astrophysics and nowadays there are more than forty active projects for volunteers to join.

The Neurona@Home project was developed at the Falúa Laboratory for distributed computing at the Campus of Aranjuez, Complutense University of Madrid [7]. Initially a total of 80 computers from the University Laboratory were included in the project test (low connectivity degree networks) achieving an initial performance of 225 GFLOPS.

The tasks requires huge amount of RAM memory for the largest value of the average degree of the network. In particular, for $k = 2000$ a computer with, at least, 10 Gigabytes RAM is required. This set a stringent limit on the number of volunteers with computers capable of managing the tasks.

Currently we have more than 300 hosts running the client for the Neurona project and the average floating point operations per second has reached 378 GFLOPS.

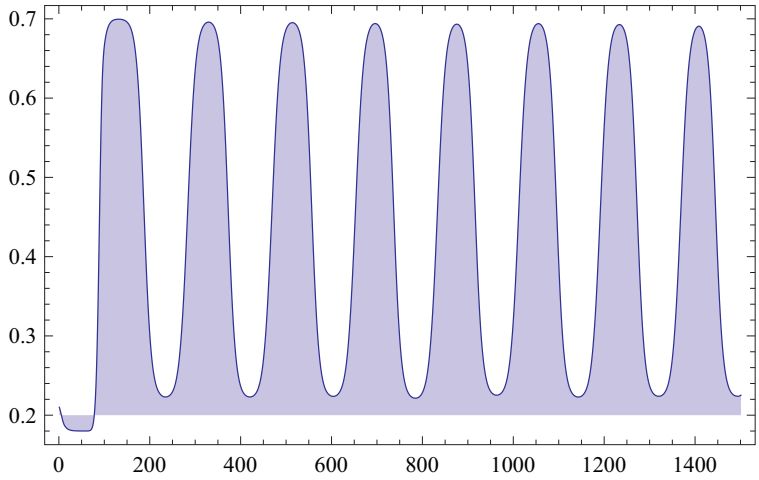


FIGURE 1. Fraction of firing neurons versus discrete time for a cellular automaton brain with 1,000,000 neurons. Parameters are $\nu = 1/10$, $\gamma = 1/98$, $\alpha = \beta = 0.000904$ in a random network with $k = 300$ average number of links per node.

RESULTS

In our model we consider both excitatory and inhibitory neurons. It is a well-known fact that about a thirty per cent of cortical neurons are inhibitory. Cellular automaton neurons are found in one of three states: Resting, Firing and Refractory. The transition from firing to refractory is measured by a Poisson stochastic process with probability ν. The return of a neuron from the refractory state to the resting state proceeds with probability γ per unit time. The network dynamics enter in the activation of resting neurons from firing neurons. Every excitatory neuron connected with a resting neuron in the network can induce this transition from resting to firing with probability α. Similarly, firing inhibitory neurons deactivate firing neurons (or inhibit the activation of a resting one) with probability β. The substrate is a random network characterized by an average number of links per node, k. We expect to explore the region in the range $k = 100$-2000.

In Fig. 1 we show a typical example in which self-sustained oscillatory behavior is found for the number of firing neurons in the network. It is important to remark that this oscillation is generated by the network itself as a self-organized behavior and no need of external forcing input is necessary. This could help to understand the different regimes in which the neuronal networks generate collective patterns and their role in the evolutionary history of the brain from insects to humans.

After the completion of the project we hope to obtain a whole phase diagram of behaviors in terms of the model parameters and to fit the data to real electroencephalograms EEG and electrode measures of activity.

ACKNOWLEDGMENTS

L. Acedo, J. Moraño and R. J. Villanueva acknowledge support from the grant FIS PI-10/01433. L. Acedo and J. Moraño acknowledge support from the Universitat Politècnica de Valéncia grant PAID-06-11 Ref: 2087. We also gratefully acknowledge the support from the BOINC community and the volunteers contributing to this project worldwide.

REFERENCES

1. S. Ramón y Cajal, *Proc. R. Soc. Lond.* **55**, 444–468 (1894).
2. M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge University Press, Cambridge, 1991.
3. S. Wolfram, *A New Kind of Science*, Wolfram Media, Champaign, Illinois, 2002.
4. M. Stopfer, S. Bhabavan, B. H. Smith, and G. Lauren, *Nature* **390**, 70–74 (1997).
5. L. M. Ward, *Trends Cogn. Sci.* **7**(12), 553–559 (2003).
6. BOINC main page: boinc.berkeley.edu
7. You can join the Neurona@Home project at the Falua webpage: `http://falua.cesfelipesegundo.com/Neurona/`

Can brains generate random numbers?

V. Chvátal[1] and M. Goldsmith

Concordia University, Montreal, Canada.

Abstract. Motivated by EEG recordings of normal brain activity, we construct arbitrarily large McCulloch-Pitts neural networks that, without any external input, make every subset of their neurons fire in some iteration (and therefore in infinitely many iterations).

Keywords: epilepsy; EEG; neural networks; random number generation.

Epilepsy is a group of neurologic conditions, the common and fundamental characteristic of which is recurrent, unprovoked epileptic seizures. These seizures are transient changes in attention or behavior, often accompanied by convulsions; they result from excessive, abnormal firing patterns of neurons that are located predominantly in the cerebral cortex. There are a number of different types of seizures and these different types of seizures manifest themselves differently in electroencephalogram (EEG) recordings of the electrical activity in the brain. One frequent occurrence is a transition from an irregular, disorderly EEG before the seizure (the pre-ictal state) to a more organized sustained rhythm of spikes or sharp waves during the seizure (the ictal state).

A *linear threshold function* is a function $f : \mathbf{R}^n \to \{0,1\}$ such that, for some real numbers $w_1, w_2, \ldots, w_n$ (mnemonic for "weights") and θ (mnemonic for "threshold"), $f(x_1, x_2, \ldots, x_n) = 1$ if and only if $\sum_{j=1}^{n} w_j x_j \geq \theta$. A *McCulloch-Pitts neural network* (with no peripheral afferents) is a mapping $\Phi : \{0,1\}^n \to \{0,1\}^n$ defined by $\Phi(x) = (f_1(x), f_2(x), \ldots, f_n(x))$ for some linear threshold functions $f_i : \{0,1\}^n \to \{0,1\}$ $(i = 1,2,\ldots,n)$. *States* of this network are zero-one vectors with n components; given an initial state s, the network computes the sequence of states $s, \Phi(s), \Phi^2(s), \ldots$, which is called its *trajectory.*

Warren Sturgis McCulloch and Walter Pitts [4] proposed these networks as a model of the central nervous system. Here, each f_i represents a neuron and variable t marks discrete time; the bits of each state $\Phi^t(s)$ tell us which neurons are firing at time t. This model, now superseded by more realistic models of the brain, played a seminal role in the development of artificial neural networks and even today is routinely referenced in medical literature.

Motivated by electroencephalograph recordings of the pre-ictal brain activity, we asked whether there are McCulloch-Pitts networks whose trajectories are random-like in the sense of following no readily discernible scheme. An essential prerequisite of every such network is that, starting from any state in its domain, it eventually produce many, if not all, states in this domain as points of the trajectory. This means that the *period of* Φ, defined as the smallest t such that $\Phi^{t+1}(s) = s$ for some s in its domain, is

[1] Canada Research Chair in Discrete Mathematics

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 271-273 (2013); doi: 10.1063/1.4776527

reasonably close, if not equal, to the size of the domain.

For every positive integer n, we have constructed a McCulloch-Pitts network $\Phi_n : \{0,1\}^n \to \{0,1\}^n$ with period 2^n. However, the trajectories of these networks are far from being random-like. To point out two of their blatant blemishes, let us consider a trajectory $s, \Phi_n(s), \Phi_n^2(s), \ldots$ and let us write $(x_1(t), x_2(t), \ldots, x_n(t))$ for $\Phi_n^t(s)$. It turns out that

$$x_1(t) = x_1(t+1) \Rightarrow x_1(t+1) \neq x_1(t+2)$$

and, for all $i = 2, 3, \ldots, n$,

$$x_i(t) \neq x_i(t+1) \Rightarrow x_i(t+1) = x_i(t+2).$$

However, if Φ_n were a random permutation of $\{0,1\}^n$, then we would expect $x_1(t) = x_1(t+1) = x_1(t+2)$ for about about 25% of the values of t and, for each $i = 2, 3, \ldots, n$, we would expect $x_i(t) \neq x_i(t+1) \neq x_i(t+2)$ for about about 25% of the values of t.

The statement that trajectories of a mapping $\Phi : \{0,1\}^n \to \{0,1\}^n$ are random-like can be given a more rigorous meaning as follows. Define $X_n = \{k/2^n : k = 0, 1, \ldots, 2^n - 1\}$ and note that there is a natural bijection $f : \{0,1\}^n \to X_n$: explicitly, $f(s_0, s_1, \ldots, s_{n-1}) = \sum_{i=0}^{n-1} 2^{-1} s_i$. Now Φ can be interpreted as a mapping $g_\Phi : X_n \to X_n$ (explicitly, $g_\Phi(x) = f(\Phi(f^{-1}(x)))$) and saying that trajectories of Φ are random-like can be interpreted as saying that g_Φ is a satisfactory uniform pseudorandom number generator.

To be considered satisfactory, a pseudorandom number generator has to pass a number of statistical tests. A number of these tests is commonly agreed on; our favourite ones are implemented in the software library TestU01 of L'Ecuyer and Simard [2, 3]. In particular, TestU01 includes batteries of statistical tests for sequences of uniform random numbers in the interval $[0,1)$. The least stringent of them, `SmallCrush`, consists of ten tests. The pseudorandom number generator g_Φ with Φ our Φ_{32} fails all of them.

Is there a McCulloch-Pitts network $\Phi : \{0,1\}^{32} \to \{0,1\}^{32}$ such that g_Φ passes all ten tests of `SmallCrush`? The promising candidates seem to be those with long periods. McCulloch-Pitts networks $\Phi : \{0,1\}^n \to \{0,1\}^n$ have period at most 2^n; how many of them attain this bound? Computer search shows that there are just two such networks when $n = 2$ (this is easy to determine without a computer), that there are precisely 24 of them when $n = 3$, and that there are precisely 9984 of them when $n = 4$. (Two distinct McCulloch-Pitts networks may be isomorphic through permuting subscripts and flipping bits. The two networks of $n = 2$ are isomorphic, the 24 networks of $n = 3$ come in two isomorphism classes, and the 9984 networks of $n = 4$ come in 56 isomorphism classes.)

ACKNOWLEDGMENTS

This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program. A detailed account of its results can be found in [1].

REFERENCES

1. V. Chvátal, and M. Goldsmith, "Can brains generate random numbers?" `arXiv:1208.6451 [math.DS]`

2. P. L'Ecuyer, and R. Simard, *ACM Trans. Math. Soft.* **33**, article 22 (2007).
3. P. L'Ecuyer, and R. Simard, TestU01: A software library in ANSI C for empirical testing of random number generators. User's guide, compact version. (Version: August 17, 2009). `http://www.iro.umontreal.ca/~simardr/testu01/guideshorttestu01.pdf`
4. W. S. McCulloch, and W. Pitts, *Bull. Math. Biophys.* **5**, 115–133 (1943).

Harmony perception and regularity of spike trains in a simple auditory model

B. Spagnolo*, Y. V. Ushakov† and A. A. Dubkov†

*Dipartimento di Fisica, Group of Interdisciplinary Physics and CNISM, Viale delle Scienze, ed. 18, I-90128 Palermo, Italy.
†Lobachevsky State University, Radiophysics Faculty, 23 Gagarin Ave., 603950 Nizhni Novgorod, Russia.

Abstract. A probabilistic approach for investigating the phenomena of *dissonance* and *consonance* in a simple auditory sensory model, composed by two sensory neurons and one interneuron, is presented. We calculated the interneuron's firing statistics, that is the interspike interval statistics of the spike train at the output of the interneuron, for consonant and dissonant inputs in the presence of additional "noise", representing random signals from other, nearby neurons and from the environment. We find that blurry interspike interval distributions (ISIDs) characterize *dissonant* accords, while quite regular ISIDs characterize *consonant* accords. The informational entropy of the non-Markov spike train at the output of the interneuron and its dependence on the frequency ratio of input sinusoidal signals is estimated. We introduce the regularity of spike train and suggested the high or low regularity level of the auditory system's spike trains as an indicator of feeling of *harmony* during sound perception or *disharmony*, respectively.

Keywords: auditory system; consonant and dissonant accords; environmental noise; hidden Markov chain; informational entropy; regularity.
PACS: 87.19.lc, 87.19.lt, 87.10.Ca, 02.50.Ga, 89.70.-a, 05.40.-a

INTRODUCTION

The perception and processing of environmental complex signals resulting from the combination of two or more input periodical signals are still an open problem for physicists and physiologists. In particular, the precise neural and physiological bases for our perception of musical consonance and dissonance are still largely unknown [1] – [3]. Although there is no single musical definition, *consonance* is usually referred to as the pleasant stable sound sensation produced by certain combinations of two tones played simultaneously. Conversely, *dissonance* is the unpleasant unstable sound heard with other sound combinations [4]. The dominant and the oldest theory of consonance and dissonance is that of Pythagoras ($570-495$ BC). He observed that the simpler the frequency ratio between two tones [1], the more consonant they will be perceived. Example: the consonant octave is characterized by a $1/2$ frequency ratio between two tones, while the dissonant semitone is characterized by a $15/16$ ratio. In 1843 Georg Ohm first proposed that the ear works as a Fourier analyzer [5]. In the same period,

[1] *Pure tone* is a single frequency tone with no harmonic components, or *overtones*. *Complex tone* is a combination of the fundamental frequency tone together with its harmonic components. Sounds produced from musical instruments are complex tones.

Physics, Computation, and the Mind - Advances and Challenges at Interfaces
AIP Conf. Proc. 1510, 274-289 (2013); doi: 10.1063/1.4776512

August Seebeck noticed the "*missing fundamental*" pitch perception [6]: a stimulus with a severely attenuated lowest component is subjectively assigned the same pitch as one with the lowest component at full strength [2]. In this work, after shortly reviewing the first physical theory on *consonance* and *dissonance* of von Helmholtz, two recent theoretical approaches (ghost stocastic resonance and nonlinear synchronization of oscillators), and the *pitch shift effect* related to the fundamental experiment on *virtual pitch* perception, we review our theoretical probabilistic approach to the statistics of *consonance* and *dissonance* musical accords by a simple auditory sensory model.

Helmholtz's theory and pitch perception

In 1877, Helmholtz analyzed the phenomenon of consonance and dissonance in the more general context of complex tones and proposed the "beat theory" [7]. When two complex tones are played together as an interval [3], the harmonics of each tone are present in the stimulus arriving at the ear of the listener. For some combinations (simple ratio n/m) the harmonic frequencies match, for others (complicated ratio n/m) they do not. As the frequency ratio n/m becomes more "complicated", the two tones share fewer common harmonics and there is an increase in harmonics pair slightly mismatched in frequency which give unpleasant beating sensation. In other words, the dissonance is proportional to the number of frequency components present in the two complex tones that produce beats. In Table 1 is shown the ordering of consonances for two tone intervals as accepted in the Western musical culture in decreasing order of "perfection" from most consonant to most dissonant [7]. The third column lists the frequency ratios of the two tones, and the fourth column lists $\Delta\Omega$, the width of the stability interval.

TABLE 1. Ordering of consonances for two–tone intervals from most dissonant (down) to most consonant (up)

interval	**name**	**interval ratio**	$\Delta\Omega$	**Consonance**
absolute consonances	unison	1 : 1	0.075	↑
	octave	1 : 2	0.023	↑
perfect consonances	fifth	2 : 3	0.022	↑
	fourth	3 : 4	0.012	↑
medial consonances	major sixth	3 : 5	0.010	↑
	major third	4 : 5	0.010	↑
imperfect consonances	minor third	5 : 6	0.010	↑
	minor sixth	5 : 8	0.007	↑
dissonances	major second	8 : 9	0.006	↑
	major seventh	8 : 15	0.005	↑
	minor seventh	9 : 16	0.003	↑
	minor second	15 : 16	-	↑
				Dissonance

[2] *Pitch* is the perceived fundamental frequency of a tone. *Pitch salience* is the strenght of tone sensation.
[3] *Interval* in music theory is the difference in pitch between the fundamental frequencies of two tones.

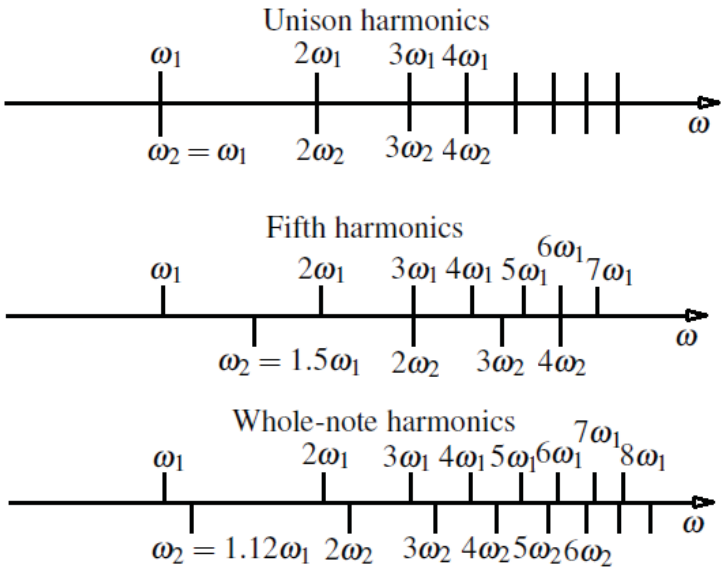


FIGURE 1. Three different intervals, namely whole–tone, perfect $5th$, unison (from bottom to top).

In the following Fig. 1, three different intervals are shown, namely whole–tone, perfect $5th$, unison (from bottom to top). We can see that the unison matches exactly, and this interval is considered to be the most consonant. Next, the perfect $5th$ shows some matched and some mismatched frequencies. The whole tone shows a mismatch for all frequencies. These frequencies can be close enough together so that discernible beats can result. As a result, the whole–tone interval is more dissonant than the perfect $5th$, which in turn is more dissonant than the octave that is more dissonant than the unison. As one proceeds down the Fig. 1, the number of mismatched harmonics increases and so does the dissonances. In beat theory of Helmholtz therefore, mismatched harmonics are considered the cause of the dissonance. Intervals were consonant if there were no or few beats between the partials. For dissonant intervals, the partials of different tones were so close together in frequency that the beating between them was perceived as dissonance [7].

Pitch is a very fundamental concept in music. In fact, music is essentially a variation in loudnesses, pitches, and timbres as a function of time [8]. The official definition of pitch is "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low" [9]. Pitch is a subjective place of a perceived complex sound on the frequency scale: pitch represents the perceived frequency of a sound. Pitch may be quantified as a frequency, but pitch is not a purely objective physical property, it is a subjective psychoacoustic attribute of sound. A high pitch ($> 2kHz$) will be perceived to be getting higher if its loudness is increased. A low pitch ($< 2kHz$) will be perceived to be going lower with increasing loudness. This is called Stevens's rule [10], the pitch of a pure sinusoidal tone depends not only on its frequency but also on its intensity. Complex tones evoke pitch sensations which are often determined exclusively by overtones. However, how the brain estimates the pitch of complex sounds, formed by a combination of pure tones, remains a controversial issue [11] – [14]. Another important

quantity in music is the *pitch salience*, which represents the evidence of the periodicity of some spike train, being the pitch value the periodicity itself. In other words, pitch salience is the probability of noticing a tone, the clarity or strength of tone sensation. The estimated salience, or relative strength, of the strongest pitch of complex tones, that is the maximum salience, is an estimation of the perceived consonance.

A complex tone composed of two sine waves of 900 and 1200 Hz gives rise to three pitches: two spectral pitches at 900 and 1200 Hz, due to the physical frequencies of the pure tones, and the combination tone at 300 Hz, corresponding to the repetition rate of the waveform. This is the so called *missing fundamental* frequency, which is the greatest common divisor of the frequencies present in the input sound.

Perception of concurrent combinations of tones is central to physiological theories of musical harmony and melody. In fact, perception of consonance in music involves *sensory* and *perceptual* processes that are relatively independent of context, as well as *cognitive* processes depend on musical context [15]. When a harmonic interval is played, neurons throughout the auditory system that are sensitive to one or more frequencies (partials) contained in the interval respond by firing action potentials. For consonant intervals, the fine timing of auditory nerve fiber responses contains strong representations of harmonically related pitches implied by the interval and all or most of the partials can be resolved by finely tuned neurons throughout the auditory system. By contrast, dissonant intervals evoke auditory nerve fiber activity that does not contain strong representations of constituent notes or related bass notes. Moreover, many partials are too close together to be resolved. Consequently, they interfere with one another, cause coarse fluctuations in the firing of peripheral and central auditory neurons, and give rise to perception of roughness and dissonance [11].

It is important to distinguish between musical *consonance/dissonance*: a given sound evaluated within a *musical context*, and psychoacoustic, or *sensory consonance/dissonance*: a given sound evaluated in isolation. *Musical consonance/dissonance* is culturally determined: variation across cultures and historical periods. Judgments of *sensory consonance/dissonance* are culturally invariant and largely independent of musical training, involving basic auditory processing mechanisms. Moreover, rodents, birds, monkeys, and human infants discriminate isolated musical chords on the basis of sensory consonance and dissonance similarly to expert human listeners and experienced musicians [1]. We will consider in this work the *just intonation* [4] musical accords, that is the *sensory consonance/dissonance*.

Ghost Stochastic Resonance

For harmonic complex sound signals, whose constituent frequencies are multiple integers of a fundamental frequency, the perceived pitch is the fundamental, even if that frequency is not spectrally present in the input signal. This is known as *missing fundamental illusion*. Recently, a mechanism for the perception of pitch has been proposed on

[4] *The just intonation* tuning is the basic scaling method in which the frequencies of notes are related by ratios of integers.

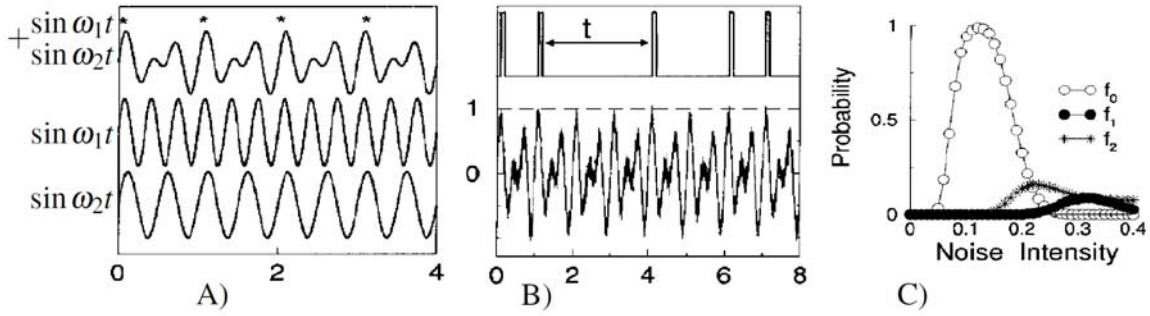


FIGURE 2. A) A complex sound s_c obtained by adding two sinusoidal signals with frequencies $\omega_1 = (k+1)\omega_0$ and $\omega_2 = k\omega_0$, namely $s_c(t) = s_1(t) + s_2(t) = a_1 sin(\omega_1 t) + a_2 sin(\omega_2 t)$. Here $a_1 = a_2 = 1, k = 2, \omega_0 = 1$. The peaks (asterisks) exhibited by s_c result from constructive interference between $s_1(t)$ and $s_2(t)$. B) The peaks of $s_c(t)$ shown in A) can be detected by a nonlinear threshold by adding a noise signal, generating interspike intervals "t" close to, or to integer multiples of, the fundamental period. C) The most probable interspike interval corresponds with the *missing fundamental* (here $f_0 = \omega_0/2\pi, f_1 = \omega_1/2\pi, f_2 = \omega_2/2\pi$).

the basis of the so called *ghost stochastic resonance* (GSR) [16] – [20]. According to the proposed mechanism, a neuron responds optimally to the missing fundamental of a harmonic complex signal for an appropriate level of noise. The main ingredients are: (i) a linear interference between the individual tones, producing peaks of constructive interference at the fundamental frequency (ghost frequency), whose amplitude is not suitable to trigger the neuron; (ii) a nonlinear threshold that detects those peaks with the help of a suitable amount of noise.

In the following Fig. 2 it is shown a complex sound $s_c(t)$ obtained by adding two sinusoidal signals $s_1(t)$ and $s_2(t)$. The constructive interference between $s_1(t)$ and $s_2(t)$ gives rise to the peaks (asterisks in the figure) in $s_c(t)$ at the period of the missing fundamental ω_0. These peaks together with a noise signal can be detected by a nonlinear threshold (see Fig. 2B). In fact, the complex tone s_c is the input to a neuron which produces a membrane potential excursion that, because of its low amplitude, cannot fire a spike. When noise is added to s_c, it induces spikes with high probability at the interference preaks. Moreover, peak detection is optimized at some noise intensity [16, 17].

The GSR mechanism was extended to describe a higher level of perception processing: the binaural pitch perception in Refs. [18, 19]. Two different neurons, at a different auditory channel, receive one single component of the complex signal each, and their output spike trains drive a third neuron that processes the information. This processing neuron responds preferentially at the ghost frequency and the response is optimized by synaptic noise.

Nonlinear synchronization theory of musical consonance

A nonlinear synchronization theory of *consonance* that goes beyond the linear beating theory of Helmholtz was recently proposed in Ref. [2]. This theory is based on the mode locking properties of simple dynamical models of pulse-coupled neurons. The mode locking describes the phenomenon where the frequencies of two oscillators remain in a given ratio for some finite range of parameters. When the oscillators, that is the periodically firing neurons, adjust their frequency to maintain the same ratio, this is a signature of nonlinear synchronization. For example we have one–to–one (1 : 1) mode locking if one neuron fires at a frequency which is synchronized with that of the second neuron. If the first neuron fires only once for every two firing of the second neuron, we have a 1 : 2 mode locking and so on, in general we have n : m mode locking (with n and m integers). By using a simple scheme of two mutually coupled neural oscillators, the authors show that the mode-locked states ordering gives precisely the standard ordering of *consonance* [3].

The authors of Ref. [2] analyze the dynamics of two coupled leaky integrate-and-fire neuron models, with mutual excitatory coupling, by finding that the mode locking ratios n/m are ordered according to the “Farey sequence", which orders all rational fractions n/m in the interval [0,1] according to their increasing denominators m [21]. By plotting the ratio of actual firing frequencies as a function of the ratio of natural intrinsic frequencies of the two coupled oscillators, they reproduce the so-called “Devil’s Staircase", with flat steps corresponding to different mode-locked states. This is a universal feature of driven coupled oscillators [22]. The width of each step, that is of the mode-locked interval, is an indicator of the structural stability of the synchronization. It is therefore possible to order the mode-locked states by their stability index, by finding a correspondence with the theoretical ordering of musical intervals according to their *consonance* evaluation. The steps decrease in width as higher integers occur in their fractional representation of the mode locking (see Fig. 3). Heffernan and Longtin in Ref. [3] analyzed in detail the same model of Ref. [2] by considering different values of coupling between the oscillators. They found that the ordering of mode locked states is not universal, but depends on the coupling strength. Moreover, the noise jitters the spike times and mode locked patterns, but the overall shape of the firing mode lockings is preserved.

Pitch shift effect

Almost all musical sounds are complex tones that consist of a lowest frequency component, or fundamental, together with higher frequency overtones. The fundamental plus the overtones are together called partials. The first perceptual theories considered pitch to arise at a peripheral level in the auditory system [5, 6, 7, 23, 24], while experiments have shown that pitch processing of complex tones is carried out before the primary auditory cortex [25]. The ability of the auditory system to perceive the fundamental frequency of a sound even when this frequency is removed from the stimulus is an interesting phenomenon related to the pitch of complex sounds. This capability is known as "*residue perception*", "*virtual pitch*" or *missing fundamental*, and consists of the per-

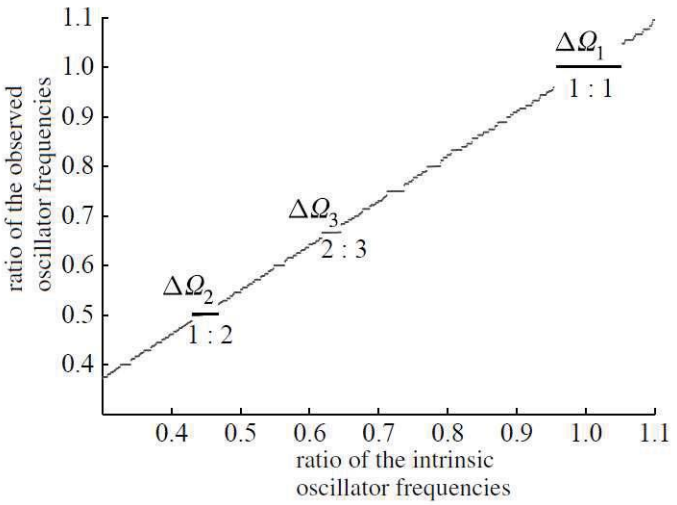


FIGURE 3. The ratio of the observed oscillator frequencies when coupled as a function of the ratio of the oscillator's natural intrinsic frequencies.

ception of a pitch that cannot be mapped to any frequency component of the stimulus (see Fig. 4). According to the Helmholtz theory, the missing fundamental can be obtained by the difference combination tone between two sources with two frequencies. However, Schouten *et al.* [23] found in their crucial experiment that the behavior of the residue cannot be described by a difference combination tone. By shifting all the partials by the same amount Δf (see Fig. 4c), the complex is no longer harmonic, the difference combination tone remains unchanged, and the same should thus be true of the residue. Instead, it is found that the perceived pitch also shifts, showing a linear dependence on Δf (Fig. 4d). This phenomenon is known as the first *pitch-shift effect*, and has been accurately measured in many psychoacoustic experiments [26]. The fundamental experiment of Ref. [23] was accurately described in terms of generic attractors of nonlinear dynamical systems, by modeling the auditory system as a generic nonlinear forced oscillator [27].

PROBABILISTIC APPROACH

The key element of the cochlea in the inner ear of mammals is the basilar membrane, which performs the sound Fourier transform with a good precision [28, 29]. As a result, different spectral components of the input signal, i.e., different oscillating parts of the basilar membrane, act upon different sensory neurons (sensors) , which send their output of spike trains to the interneurons. Because we restrict our analysis by two spectral harmonics (simple chords of tone pairs), it is sufficient to consider the model with two sensors at the input (see Fig. 5). The sensors N_1, N_2 are subjected to the mixture of subthreshold sinusoidal signals with different frequencies and statistically independent additional white Gaussian noises. The sum of weighted sensors' spike trains summed with the third statistically independent white Gaussian noise is sent to the interneuron

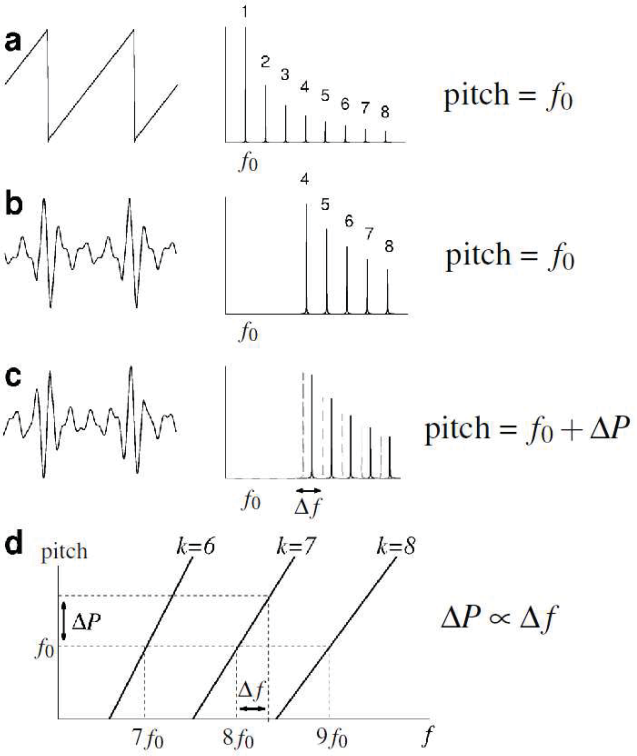


FIGURE 4. (a) A harmonic complex tone. The overtones are successive integer multiples $k = 2, 3, 4, ..$ of the fundamental f_0 that determines the pitch. (b) Another harmonic complex tone with the fundamental and the first few higher harmonics removed. The pitch remains the same and equal to the *missing fundamental*. This pitch is known as *virtual* or *residue pitch*. (c) An anharmonic complex tone, where the partials, which are no longer harmonics, are obtained by a uniform shift Δf of the previous harmonic case (shown as dashed line). Although the difference combination tones between successive partials remain unchanged and equal to the missing fundamental, the pitch shifts by a quantity ΔP that depends linearly on Δf. (d) Pitch as a function of the central frequency $f_c = (k+1)f_0 + \Delta f$ of a three component complex tone, namely $kf_0 + \Delta f$, $(k+1)f_0 + \Delta f$ and $(k+2)f_0 + \Delta f$. This is the *pitch shift effect*, shown here for $k = 6, 7$, and 8 (see Ref. [27]).

N_3, which is an internal neuron connecting sensory neurons to other neurons within the same region of the brain. The output spike train of the interneuron is the main object of investigation [30, 31].

Each neuron is modeled by the simple nonlinear model referred to as the noisy leaky integrate-and-fire neuron [32]. We analyze the probability distribution of interspike intervals (ISIDs) of the output signal of the interneuron by assuming to know the ISIDs of the output signals of the two sensory neurons $\rho_1(t)$ and $\rho_2(t)$. We reduce the number of events for which the interneuron can fire to four main scenarios because all other events have a very negligible probability to happen in comparison with the previous four. In this way we are able to calculate the first passage time distribution at the output of the interneuron $\rho_3(t)$, using conditional probabilities and first passage time distributions at the output of sensory neurons. Moreover, for periodical input signal at the sensors with frequency ratio m/n we obtain (m+n-1) different patterns of input spike

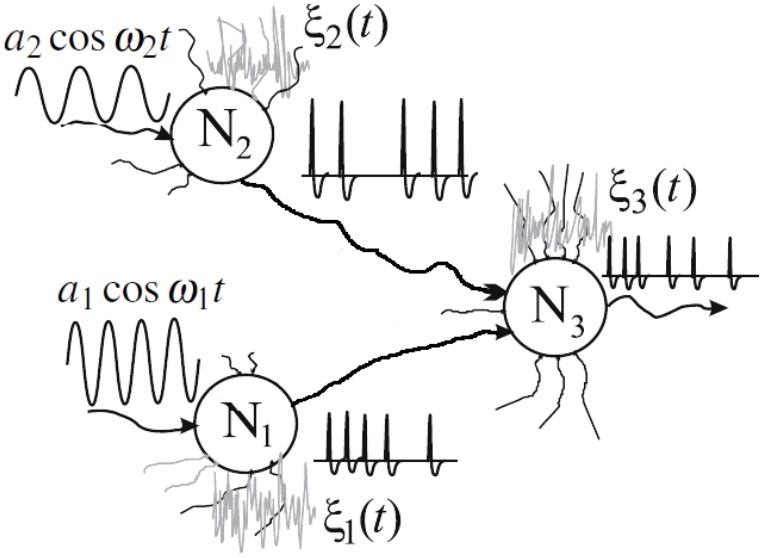


FIGURE 5. The investigated model. N_1 and N_2 are the sensory neurons, driven by subtreshold sinusoidal signals with different frequencies. Spike trains of sensors are received by the interneuron N_3. $\xi_1(t)$, $\xi_2(t)$, and $\xi_3(t)$ are the statistically independent white Gaussian noises.

trains for the interneuron, with different ISIDs at its output. The final interspike interval density of the interneuron $\rho_{out}(T)$ is obtained by averaging the first passage time density $\rho_3(t)$ over all different states (m+n-1) of the interneuron. We show how a complex input composed of two harmonic signals is transformed by the proposed simple sensory system into different types of spike trains, depending on the ratio of input frequencies. Looking for the differences in the statistical sense, we find out that the output ISIDs for some combinations of frequencies, corresponding to consonant accords, have more regular pattern, while inharmonious signals, corresponding to dissonant accords, show less regular spike trains and blurry ISIDs. This difference indicates that consonant accords are higher stable, with respect to the noise environment, in comparison with the dissonant accords in the processing of information throughout the auditory system.

Model

As a neuron model for our sensory system (see Fig. 5) we consider the Leaky Integrate-and-Fire (LIF) model. Therefore, the set of stochastic differential equations describing our system is

$$\begin{cases} \dot{v}_1 = -\mu_1 v_1 + A_1 \cos(\Omega_1 t) + \sqrt{D_1}\xi_1(t), \\ \dot{v}_2 = -\mu_2 v_2 + A_2 \cos(\Omega_2 t) + \sqrt{D_2}\xi_2(t), \\ \dot{v}_3 = -\mu_3 v_3 + k_1 s_1(t) + k_2 s_2(t) + \sqrt{D_3}\xi_3(t), \end{cases} \tag{1}$$

where $v_i(t)$ and μ_i stand for the membrane potential and the relaxation parameter, respectively, and the subscript i labels the different neurons, with $i = 1, 2$ representing the

two input sensory neurons (N_1 and N_2) and $i = 3$ (N_3) denoting the processing interneuron. A_i and Ω_i (with $i = 1,2$) are the amplitude and the frequency of the corresponding harmonic input of the sensors. We consider that the three neurons have different synaptic connections, they are not subject to the same background noise and the three noise sources $\xi_i(t)$ are independent of each other. Therefore, in Eqs. (1), the three white Gaussian noise terms $\xi_i(t)$ ($i = 1,2,3$) are uncorrelated and with the usual statistical properties $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(t') \rangle = \delta(t-t')\delta_{ij}$. D_i is the noise intensity in each neuron. In Eq. (1) $s_i(t) = \sum_{j=0}^{N_i(t)} \delta(t - t_{ij}), i = 1,2$ are the spike trains generated by the sensors and received by the interneuron as input, k_i ($i = 1,2$) are the coupling coefficients. Spikes are modelled by Dirac δ-functions. The LIF model doesn't comprise any mechanism of spike generation. When the membrane potential v_i reaches the threshold value v_{th}, the neuron is said to fire a spike, and v_i is reset to its initial value v_i^0. In particular, the input spikes at the interneuron, coming from the sensory neurons, can produce spikes or jumps in the membrane potential of the interneuron, depending on whether or not they are suitable to fire the interneuron.

All simulation and theoretical results presented in this work are obtained using the following set of values of system parameters, namely $\mu_1 = \mu_2 = 1$, $\mu_3 = 0.3665$, $D_1 = D_2 = D_3 = 1.6 \cdot 10^{-3}$, $k_1 = k_2 = 0.98$, $v_1^0 = v_2^0 = 0$, $v_3^0 = -1$, and $v_{th} = 1$, unless stated otherwise. The *refractory period* T_{ref} of the output interneuron is introduced explicitly as the time at which the membrane potential reaches the level $v_3 = -0.1$, that is $T_{ref} = 6.28$. The first two equations of system (1) describe the Ornstein-Uhlenbeck processes with harmonic driving forces. For the Ornstein-Uhlenbeck neuronal model, the ISID was obtained analytically with different approaches in Refs. [33, 34]. This distribution, which coincides with the first passage time probability distribution related to the firing event of sensory neurons, is our starting point to obtain the ISID at the output of the interneuron.

It is important to note here that the ISIDs at the output of two sensors are non-Poissonian (see Fig. 6b). These spike trains are the input of the third neuron, and as a consequence the dynamics of the membrane potential of the interneuron is non-Markovian. The output of the interneuron is shown in Fig. 6c. In order to perform this analysis we use three main assumptions: (i) The input harmonic signals are *subthreshold* for the sensors, that is the signal $A_i \cos(\Omega_i t)$ is not able to bring the membrane potential of the i^{th} sensor above the threshold in the absence of noise ($D_i = 0$). This means absence of spikes at the output of the sensors. (ii) Only one spike can be generated at each period of the harmonic driving force, and, at the same time, the spiking on each period is the most probable situation (see Fig. 6a). This means that the relaxation times of sensors are smaller than the periods of the sinusoidal signals. (iii) Each of coupling coefficients k_i is less than the threshold value of the membrane potential v_{th}. It means that any separate incoming spike (see Fig. 6c) evokes a subthreshold impulse of the membrane potential of the interneuron $v(t)$, i.e. spike generation is impossible without noise. At the same time, the sum of the two coupling coefficients is greater than v_{th}.

Therefore, we can evaluate the probability $\Delta P_3(t) = \rho_3(t)\Delta t$ that the interneuron N_3 fires in the short time interval $(t, t+\Delta t)$, by considering the occurrence of the following events:

1. receiving a *separate firing spike* from the sensory neuron N_1;

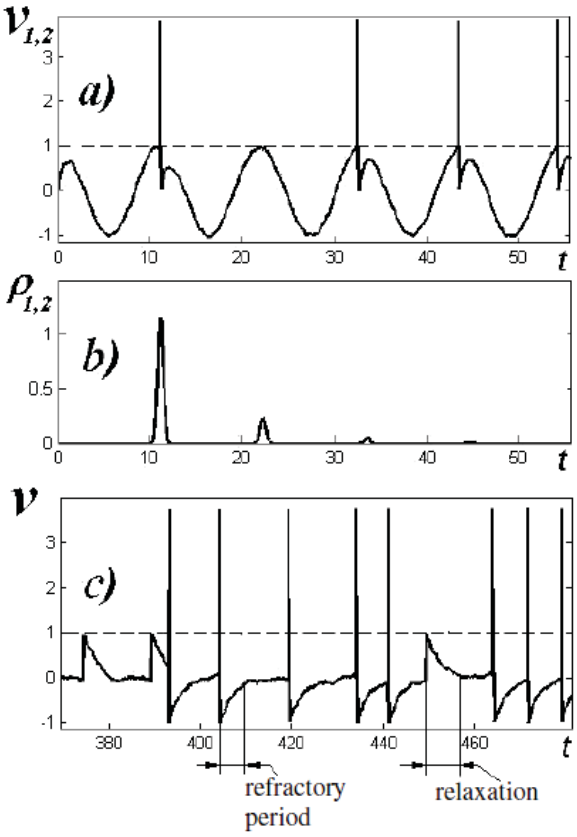


FIGURE 6. (a) Typical behavior of the membrane potential $v_i(t)$ of sensory neurons versus time for a noise realization. (b) ISI distribution of the sensory neurons. The highest probability of a spike after $t = 0$ is near one period of external force ($t = 10.47$). The probability of firing after two, three, etc. periods decreases exponentially. (c) Typical behavior of the membrane potential $v_3(t)$ of the interneuron versus time for the same noise realization. Here are well visible the refractory state (*ref*), characterized by the refractory time T_{ref}, and the noisy background (*bg*) during the relaxation time T_{relax}.

2. receiving a *separate firing spike* from the sensory neuron N_2;
3. receiving a firing spike from the neuron N_1 on the *background* of the membrane potential relaxing, after the jump due to the spike from the N_2 neuron, towards the zero value; in other words, sensor N_2 causes the jump and then sensor N_1 the spike;
4. receiving a firing spike from the neuron N_2 on the *background* of the membrane potential relaxing, after the jump due to the spike from the N_1 neuron, towards the zero value; in other words sensor N_1 causes the jump and then sensor N_2 the spike.

We neglect the contribution of multiple jump events to fire the interneuron and the noise-induced spike events occurring during the relaxation of the membrane potential after a jump, because they have very negligible probability to happen in comparison with the previous four, with the chosen range of system parameters. The four described scenarios exclude each other, so they are mutually exclusive events. As a result, according to the formula of total probability we have to add up all probabilities of the above

mentioned events.

Interspike interval distributions

Now we calculate numerically the interspike interval distributions of the interneuron for two groups of *consonant* and *dissonant* accords by numerical simulations of Eqs. (1) (see Figs. 7 and Fig. 8).

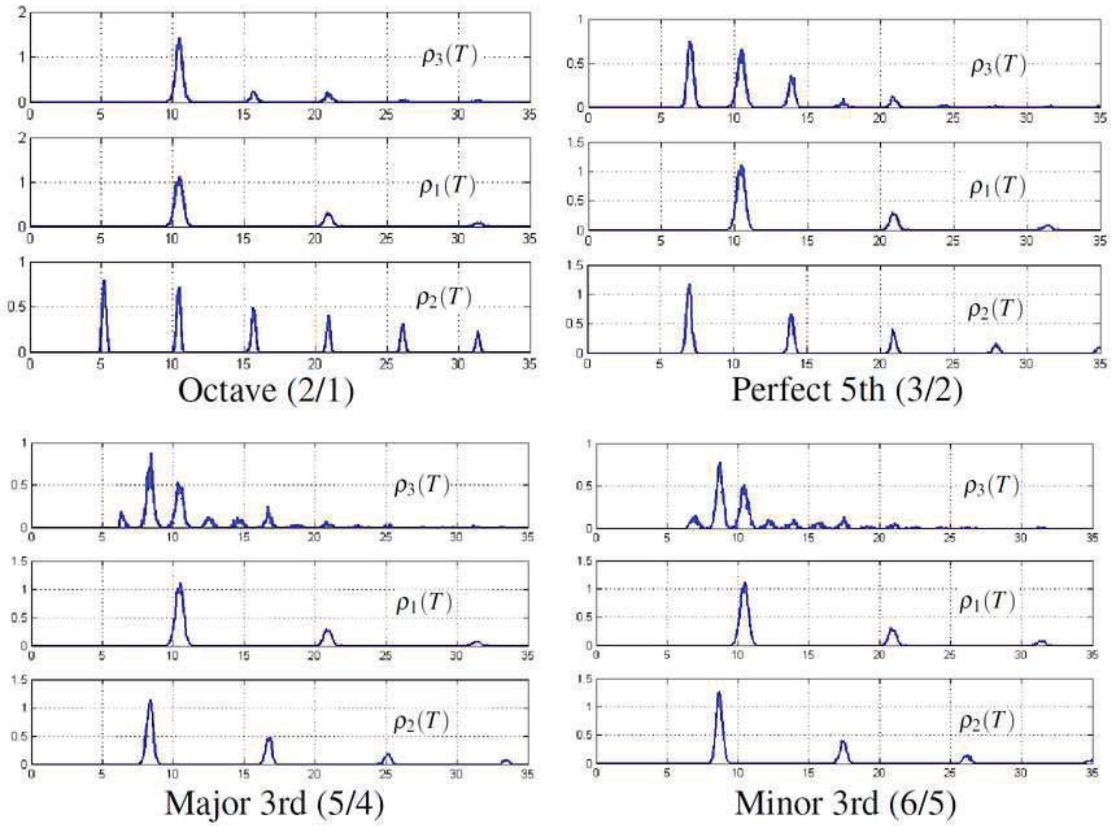


FIGURE 7. ISI distributions of the *consonant* accords: octave $(2/1)$, perfect $5th$ $(3/2)$, major $3rd$ $(5/4)$, and minor $3rd$ $(6/5)$. All curves are obtained through the direct numerical simulation of the Eqs. (1).

We note the very regular behavior of the patterns of $\rho_{out}(T)$ in all the *consonant* accords considered, and the very rich pattern with many peaks in the major $3rd$ $(5/4)$ and minor $3rd$ $(6/5)$ accords.

The ISIDs of *dissonant* accords are blurry with respect to the ISIDs of the *consonant* accords. This means that we can consider the ISID as an investigative tool to discriminate between *consonant* and *dissonant* accords. In fact higher are the integers m, n less regular and blurry are ISIDs, while lower are the integers more regular are the ISIDs.

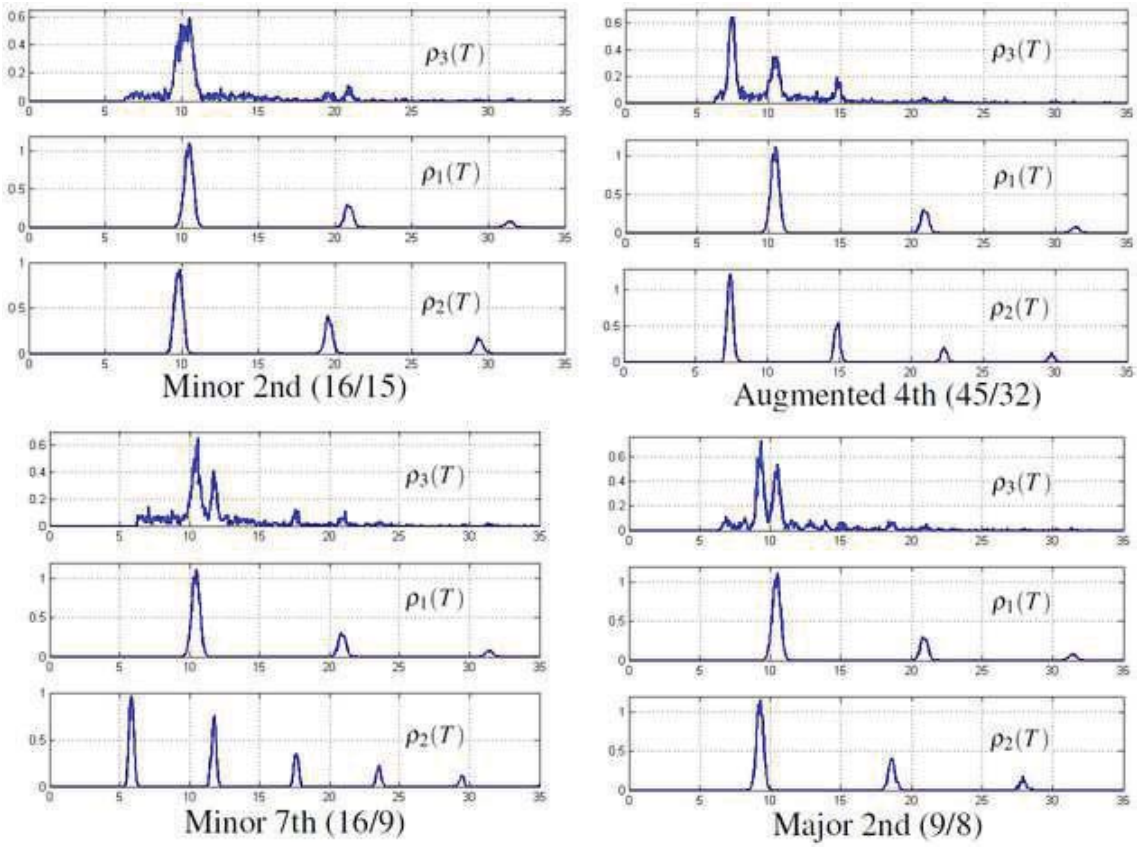


FIGURE 8. ISI distributions of the *dissonant* accords: major 2*nd* (9/8), minor 7*th* (16/9), minor 2*rd* (16/15), and augmented 4*th* (45/32). All curves are obtained through the direct numerical simulation of the Eqs. (1).

REGULARITY

Each different state of the interneuron belongs to a hidden Markov chain (HMC). For each state of the HMC we are able to calculate the First Passage Time Probability Density (FPTPD) for the passage of the interneuron's threshold of spike generation by the theoretical approach presented in the previous section (see Ref. [30] for details). For input frequencies with ratio ($\Omega_1/\Omega_2 = m/n$), all FPTPDs consist of peaks, and each peak corresponds to switching into some existing state of the HMC. Thus, the element of the HMC's transition matrix is obtained as follows: $\pi_{ij} = \int\limits_{(i\to j)} \rho^{(i)}(t)dt$, where $\rho^{(i)}(t)$ is the FPTPD of the interneuron in the i-th state, and $(i \to j)$ is the interval, in which the peak of $\rho^{(i)}(t)$, corresponding to switching into a state j, is situated.

Starting from the HMC's transition matrix we calculate the specific informational entropy H of the interneuron's spike train using the Shannon's formula [31]

$$H = -\sum_{i=0}^{M-1} p_i \sum_{j=0}^{M-1} \pi_{ij} \log_2 \pi_{ij}, \tag{2}$$

where p_i is the probability of state i, which can be obtained from the $\{\pi_{ij}\}$ matrix, and M is the whole number of states of the HMC.

To characterize the regularity of the spike trains we introduce the *spike regularity measure* R as

$$R(m/n) = H_{max} - H(m/n), \tag{3}$$

where H_{max} is the maximal entropy value over all considered m/n ratios. Obviously, R is defined up to a multiplicative constant [31]. In Fig. 9 the dependence $R(m/n)$ corroborates the hypothesis of the connection between the harmony perception and highly regular spike trains in neural ensembles of the auditory system [30]. Indeed, the regularity R (the entropy H) is high (low) for small integers m,n (namely, $m,n < 10$), i.e. the investigated system produces a regular output spike train under influence of consonant accords at the input. R grows linearly with increasing ratio m/n at fixed difference $(m-n)$ (Fig. 9, bold solid lines).

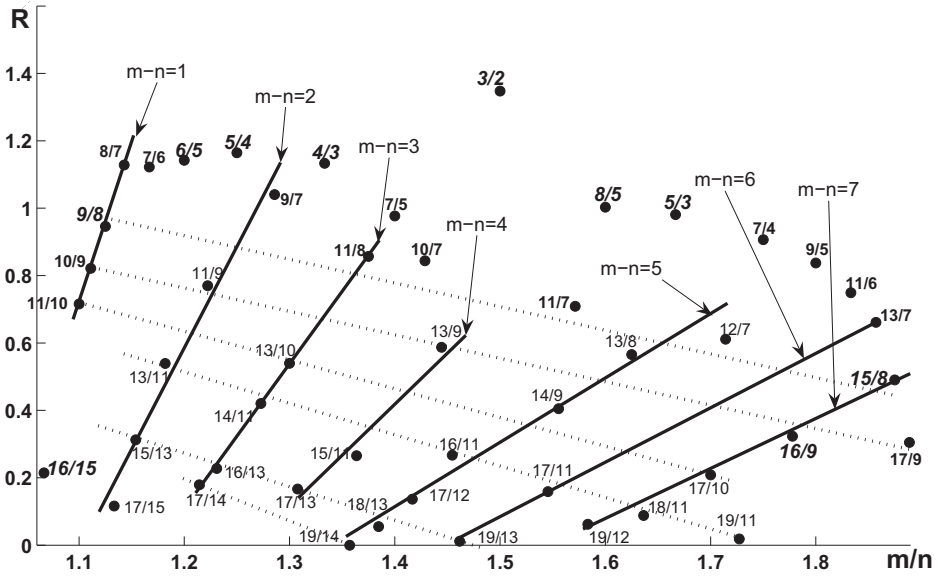


FIGURE 9. Regularity of the interneuron's spike train depending on the frequency ratio of input sinusoidal signals $m/n = \Omega_1/\Omega_2$. The bold solid lines approximate the locus of the $R(m/n)$ points for constant differences $(m-n)$. The noise intensity is the same for all three noise sources.

We note that this behavior of the regularity is very similar to the well-known first pitch-shift effect [23] in the psychoacoustics: the linear growth of pitch for the linear upward shift of frequencies of sounding tones at a given difference between the frequencies. Therefore, because the pitch is a proxy of the regularity, the observed qualitative correspondence between the obtained dependence $R(m/n)$ and the dependence $f_p(m/n)$, confirmed in experiments [23] (see also Fig. 4), proves the feasibility of the model under investigation. In some sense, the regularity embraces both the pitch value (periodicity of a spike train) and the pitch salience (evidence of the periodicity). Thus, the use of the *regularity* value R as a measure of the "*consonance level*" may have a number of advantages in comparison with the use of the pitch salience. Firstly, regularity is a clear

physical quantity of a concrete spike train. Secondly, an R value can be obtained directly from a spike train by calculation of specific informational entropy [31]. Thirdly, obtaining of a regularity value does not require determination of a pitch value, which is a problem in a case of unknown or too complex input sound, e.g., a voice of a human. We suppose also that an experimental confirmation of the plots shown in the Fig. 9 can be very fruitful for neurophysiological applications. For example, the discovery of brain regions where property of the spike train regularity could help to understand how pleasant or unpleasant are perceived by a mammal sounds, which are more complex than simple musical accords.

CONCLUSIONS

With our simple model of the auditory system, we are able to discriminate between *consonant* and *dissonant* accords by analyzing the first passage time probability distributions at the output of the interneuron. Blurry ISIDs characterize *dissonant* accords, while quite regular ISIDs characterize *consonant* accords (Figs. 7, 8). We have calculated the informational entropy for the non-Markov spike train at the output of the auditory system model, and introduced the regularity of spike train. The high or low regularity level of auditory system's spike trains has been suggested as an indicator of feeling of harmony during sound perception or disharmony, respectively. By considering an extension of this simple model to a more complex realistic auditory system, composed of many sensory neurons and different layers, we should be able to know at which extent the *dissonant* accords will "survive", against the *consonant* ones, in the noisy neural environment of the brain.

ACKNOWLEDGMENTS

The presented study is supported by MIUR and the Russian Foundation for Basic Research (grant number 11-02-01418-a).

REFERENCES

1. Y. I. Fishman *et al.*, *J. Neurophysiol.* **86**, 2761–2788 (2001).
2. I. Shapira Lots, and Lewi Stone, *J. R. Soc. Interface* **5**, 1429–1434 (2008).
3. B. Heffernan, and A. Longtin, *J. Neurosci. Meth.* **183**, 95–106 (2009).
4. R. Plomp, and W. J. M. Levelt, *J. Acoust. Soc. Am.* **38**, 548–560 (1965).
5. G. S. Ohm, *Ann. Phys. Chem.* **59**, 513–565 (1843).
6. A. Seebeck, *Ann. Phys. Chem.* **60**, 449–481 (1843).
7. H. L. F. Helmholtz, *On the sensations of tone as a physiological basis for the theory of music*, Dover, New York, 1954. Translation of Fourth German edition (1877) by Alexander J. Ellis, originally published in 1885.
8. A. J. M. Houtsma, *Music Perception* **1**(3), 296–307 (1984).
9. *American Standards Association: Acoustic Terminology*, S1.1, New York 1960.
10. S. S. Stevens, *J. Acoust. Soc. Am.* **6**, 150–154 (1935).
11. M. J. Tramo, P. A. Cariani, B. Delgutte, and L. D. Braida, *Ann. N. Y. Acad. Sci.* **930**, 92–116 (2001).

12. C. J. Plack, and A. J. Oxenham, "The psychophysics of pitch", in *Pitch: Neural coding and perception*, C. J. Plack, A. J. Oxenham, R. R. Fay and A. N. Popper eds., pp. 7–55, New York, Springer-Verlag, 2005.
13. A. de Cheveigné, "Pitch perception models", in *Pitch: Neural coding and perception*, C. J. Plack, A. J. Oxenham, R. R. Fay and A. N. Popper eds., pp. 169–233, New York, Springer-Verlag, 2005.
14. D. Wile, and E. Balaban, *PLoS ONE* **2**(4), e369 (2007).
15. P. Cariani, in *Proc. 8th Int. Conf. on Music Perception and Cognition, Evanston, IL*, edited by S. D. Lipscomb *et al.* pp. 310–314, Causal Productions, Adelaide, Australia, 2004.
16. D. R. Chialvo, O. Calvo, D. L. Gonzalez, O. Piro, and G. V. Savino, *Phys. Rev. E* **65**, 050902(R)(4) (2002).
17. D. R. Chialvo, *Chaos* **13**, 1226–1230 (2003).
18. P. Balenzuela and J. García-Ojalvo, *Chaos* **15**, 023903(8) (2005).
19. A. Lopera, J. M. Buldu, M. C. Torrent, D. R. Chialvo, and J. Garcia-Ojalvo, *Phys. Rev. E* **73**, 021101(6) (2006).
20. P. Balenzuelaa, H. Braun, and D. R. Chialvo, *Contemporary Physics* **53**(1), 17–38 (2012).
21. L. Glass, and M. C. Mackey, *From clocks to chaos*, Princeton University Press, Princeton, 1988.
22. H. G. Schuster, *Deterministic Chaos*, VCH-Verlag, Weinheim, 1994.
23. J. F. Schouten, R. J. Ritsma, and B. L. Cardozo, *J. Acoust. Soc. Am.* **34**, 1418–1424 (1962).
24. R. Plomp, *J. Acoust. Soc. Am.* **41**, 1526–1533 (1967).
25. C. Pantev, M. Hoke, B. Lütkenhöner, and K. Lehnertz, *Science* **246**, 486–488 (1989).
26. W. M. Hartmann, *J. Acoust. Soc. Am.* **100**, 3491–3502 (1996).
27. J. H. E. Cartwright, D. L. Gonzalez, and O. Piro, *Phys. Rev. Lett.* **82**, 5389–5392 (1999).
28. D. Benson, *Music: a Mathematical Offering*, Cambridge University Press, Cambridge, 2006.
29. V. M. Eguiluz, M. Ospeck, Y. Choe, A. J. Hudspeth, and M. O. Magnasco, *Phys. Rev. Lett.* **84**, 5232–5235 (2000).
30. Y. V. Ushakov, A. A. Dubkov, and B. Spagnolo, *Phys. Rev. E* **81**, 041911(13) (2010).
31. Y. V. Ushakov, A. A. Dubkov, and B. Spagnolo, *Phys. Rev. Lett.* **107**, 108103(4) (2011).
32. A. N. Burkitt, *Biol. Cybern.* **95**, 97–112 (2006).
33. H. E. Plesser, and S. Tanaka, *Phys Lett A* **225**, 228–234 (1997); H. E. Plesser, and T. Geisel, *Phys. Rev. E* **63**(6), 031916 (2001).
34. C.F. Lo, and T.K. Chung, *First Passage Time Problem for the Ornstein-Uhlenbeck Neuronal Model*, King *et al.* (Eds.), ICONIP 2006, Part I, Springer-Verlag Berlin Heidelberg, *Lecture Notes in Computer Science* **4232**, 324–331 (2006).

List of Participants

12th Granada Seminar

L.	Acedo	*Spain*
E.J.	Agnes	*Brazil*
J.	Aguilá	*Spain*
D.G.	Almeida-Filho	*Brazil*
M.L.	Antunes	*Portugal*
A.H.	Azizi	*Germany*
D.	Battaglia	*France*
G.	Bolat	*Romania*
G.	Bolat	*Romania*
J.A.	Bonachela	*U.S.A.*
J.	Brooke	*United Kingdom*
R.R.	Carrillo	*Spain*
D.	Chialvo	*U.S.A.*
V.	Chvátal	*Canada*
J.	Cortés	*Spain*
L.	de Arcangelis	*Italy*
S.	de Franciscis	*Italy*
M.J.	de Linares	*Spain*
F.	de los Santos	*Spain*
B.	de Sancristobal	*Spain*
V.	de Souza	*Spain*
J.J.	del Pozo	*Spain*
M.	di Volo	*Italy*
S.	Dolev	*Israel*
V.	Domínguez	*Spain*
A.J.	Duggins	*Australia*
M.	España-Acebal	*Spain*
J.F.	Fernández	*Spain*
M.M.	Fernández-Martínez	*Spain*
P.L.	Garrido	*Spain*
J.A.	Garrido	*Italy*
M.	Goldsmith	*Canada*
A.V.	Goltsev	*Portugal*
A.	Gómez-Marin	*Spain*
X.S.	Gonzalo-Cogno	*Argentina*
J.	Goñi	*U.S.A.*
L.	Gregory-Brunnet	*Brazil*
J.	Grollier	*France*
C.	Gros	*Germany*
S.	Grossberg	*U.S.A.*
D.	Grytskyy	*Germany*
S.D.	Gunay	*Turkey*
S.	Hajiji	*France*
A.	Hanemann	*Israel*
T.	Hansard	*United Kingdom*
J.	Hidalgo	*Spain*
R.	Huerta	*U.S.A.*
P.I.	Hurtado	*Spain*
A.	Ibañez-Molina	*Spain*
S.	Iglesias-Parro	*Spain*
J.J.	Jaramillo	*Germany*
S.	Johnson	*United Kingdom*
N.J.	Joshi	*U.S.A.*
H.J.	Kappen	*Holanda*
H.	Kato	*Spain*
F.	Kayadibi	*Turkey*
A.	Kolchinsky	*U.S.A.*
A.A.	Kostikov	*Ukraine*
K.E.	Lee	*Portugal*
K.K.	Lin	*U.S.A.*
R.	Livi	*Italy*
N.R.	Luque-Sola	*Spain*
L.	Lyra-Gollo	*Spain*
E.	Madrid	*Spain*
J.A.	Magro	*Spain*
L.P.	Maia	*Brazil*
D.	Manzano-Diosdado	*Spain*
D.	Marinazzo	*Belgium*
J.	Marro	*Spain*
J.M.	Martín	*Spain*
H.	Matsueda	*Japan*
B.L.	McNaughton	*Canada*
J.	Medeiros de Araujo	*Brazil*
J.	Mejías	*Canada*
A.F.	Meyer	*Germany*
C.	Mirasso	*Spain*
R.	Molina	*Spain*
A.	Montakhab	*Iran*

P.	Moretti	*Spain*
M.A.	Muñoz	*Spain*
F.	Naveros	*Spain*
A.	Negrean	*The Netherlands*
J.G.	Orlandi	*Spain*
P.	Paradisi	*Italy*
F.	Pérez-Díaz	*United Kingdom*
S.	Petkioski	*United Kingdom*
B.E.	Pimentel-Mizusaki	*Brazil*
S.	Rathore	*United Kingdom*
D.	Risso	*Chile*
E.	Ros	*Spain*
M.	Ruipérez-Alonso	*Nehterlands*
T.	Sako	*Japan*
C.	Sánches-Ferreira	*Spain*
M.V.	Sánchez-Vives	*Spain*
S.	Scarpetta	*Italy*
F.	Selingardi Matias	*Spain*
A.M.	Skulimowski	*Poland*
J.	Soriano	*Spain*
B.	Spagnolo	*Italy*
W.	Stühmer	*Spain*
S.	Teller	*Spain*
J.	Tobochnik	*U.S.A.*
K.	Tokuda	*Japan*
J.J.	Torres	*Spain*
B.	Ulrich	*Germany*
Y.V.	Ushakov	*Russia*
P.	Villa-Martín	*Spain*
J.	Wyller	*Norway*
L-S	Young	*U.S.A.*

List of Selected Contributions Presented at the Conference

NEUROSCIENCE

- **A computational model for preplay in the hippocampus**
  A. H. Azizi and S. Cheng
- **Synchronization phenomena in networks of spiking neurons with a correlated scale-free topology**
  S. de Franciscis, J. Mejias , S. Johnson and J. J. Torres
- **Crossover from synchronous firing to population burst statistics**
  M. di Volo and R. Livi
- **Homeostatic mechanisms at cerebellar parallel fiber-Purkinje cell connections through deep cerebellar nuclei LTD and LTP**
  J. A. Garrido, N. R. Luque, E. Ros and E. D'Angelo
- **Signal integration shapes the dynamics and enhances the dynamic range of neural networks**
  L. L. Gollo, C. Mirasso and V. M. Eguíluz
- **Stochastic amplification of fluctuations in cortical UP-states**
  J. Hidalgo, L. F. Seoane, J. M. Cortés and M. A. Muñoz
- **Dimensional analysis of EEG plots**
  A. J. Ibáñez-Molina, F. J. Esteban and S. Iglesias-Parro
- **The minimal complexity of adapting agents increases with fitness**
  N. J. Joshi, G. Tononi and Ch. Koch
- **Potential control implications of the inferior olive→deep cerebellar nuclei pathway in a distributed plasticity cerebellar model**
  N. R. Luque-Sola, J. A. Garrido, R. R. Carrillo and E. Ros
- **Robust and efficient receptive field inference from binary responses with stochastic gradient descent**
  F. Meyer and J. Anemüller
- **dHAN model of neural networks in the light of experimental neuroscience**
  A. Montakhab and S. Sepehri
- **The glutamate receptor interacting protein Shisa-9 alters hippocampal neuronal network synchronization**
  M. Ruipérez-Alonso, R. Klaassen, A. B. Smit and H. D. Mansvelder
- **Prefrontal theta oscillations track the time course of interference during selective memory retrieval**
  C. Sánchez-Ferreira, A. Marful, T. Staudigl, T. Bajo and S. Hanslmayr
- **Stochastic coherence in UP and DOWN states. Experiments and modeling**
  B. Sancristóbal, M. V. Sanchez-Vives and J. Garcia-Ojalvo

OTHER TOPICS

- **Excess molar volumes of binary mixtures (alkyl benzene + an alkyl propionate) at 298.15 K**
  G. Bolat and S. C. I. Strugaru

- **Dynamic electrochemical impedance spectroscopy study for NiTi and NiTiNb in physiological solution**
  G. Bolat , D. Mareci, S. Iacoban and C. Munteanu
- **Volume swelling of by irradiation of alpha particles and lattice recovery by molecular dynamics simulation**
  S. D. Günay and Ç. Taşseven
- **Kinetic study in Ar-N2-H2 plasma discharge at low pressure:density profiles of NHx radicals**
  S. Hajiji, I. Jauberteau, J. L. Jauberteau, J. Aubreton and M. F. Elchinger
- **Radiation Damage in LiO2 and Recovery During Thermal Annealing**
  F. Kayadibi and S. D. Günay
- **Vortex configuration in granular thin superconducting film: Monte Carlo study**
  L.V. Belevtsov and A. A. Kostikov
- **Blue-violet photonic emission from human bodies by will in healing Qi practice**
  H. Matsueda
- **Hydrodynamics modes in shallow granular fluids**
  D. Risso, R. Brito and R. Soto
- **Origin of Hund's multiplicity rule in singly-excited two-electron systems**
  T. Sako

AUTHOR INDEX