Brain as a Complex System

harnessing systems neuroscience tools & notions for an empirical approach

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt
von

Shervin Safavi
aus Tehran, Iran

2021

Tag der mündlichen Prüfung:	2021-10-20
Dekan der Math.-Nat. Fakultät:	Prof. Dr. Thilo Stehle
Dekan der Medizinischen Fakultät:	Prof. Dr. Bernd Pichler
1. Berichterstatter:	Prof. Dr. Nikos K. Logothetis
2. Berichterstatter:	Prof. Dr. Anna Levina
3. Berichterstatter:	Prof. Dr. Sonja Grün
Prüfungskommission:	Prof. Dr. Nikos K. Logothetis
	Prof. Dr. Martin Giese
	Prof. Dr. Anna Levina
	Prof. Dr. Gustavo Deco

Erklärung / Declaration: Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

"Brain as a Complex System, harnessing systems neuroscience tools & notions for an empirical approach"

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I hereby declare that I have produced the work entitled *"Brain as a Complex System, harnessing systems neuroscience tools & notions for an empirical approach"*, submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, den ....

Datum / Date Unterschrift / Signature

Shervin safavi:

Brain as a Complex System
harnessing systems neuroscience tools & notions for an empirical approach

Dedicated to all the nurses, doctors, clinicians and scientists ...
who sacrifice their lives to save ours during the COVID-19 pandemic.

Dedicated to loving Farhad Meysami
who has an important contribution in shaping my mindset.


زمزمی صحنه یکتایی هنرمندی ماست
هر کسی نغمه خود خواند و او صحنه رود
صحنه پیوسته به جاست

Translation:

Life is our unique stage of performance!
Everyone sing their own song and leave ...
Stage remains ...
Remembered songs are the delighted ones.

— Poem by Zhale Esfehani
(Subjectively translated by Shervin Safavi)

The most beautiful aspect of science is that it is a collaborative enterprise —
Freeman J. Dyson

ACKNOWLEDGMENT

Science is a *collaborative endeavor* and this small piece of work could not have been accomplished without the help of many people. Indeed, that's the reason I mostly mention *we did* rather *I did* in this thesis. Probably I managed to mention a subset of those people here in this acknowledgment. Writing the acknowledgment section was one of the most pleasant parts for me as it has the sign of the collective and cooperative attitude of humankind. I hope I manage to reinforce this aspect of science, cultivate it, and ultimately do better science.

I'm grateful to Nikos Logothetis for providing me the freedom for exploration, designing my research, and sufficient resources to realize the ideas. Doing the PhD in his inspiring lab was a unique opportunity for me (from multiple perspectives). Perhaps, one of the most valuable experiences I had in his lab was doing experiments and theory (in the extreme of each) both during my PhD which already helped me to craft my neuroscientific vision and will help me throughout my career. I wish the crises related to animal activists never happened and I could have more learning opportunities from him and the members of his lab.

I'm thankful to Michel Besserve, with whom I pursued an important part of my PhD research in Nikos' lab. I'm grateful for constructive discussions and help whenever and in whichever way he could provide besides his own intensive and resource-demanding research.

I'm grateful to Anna Levina [and all of her lab members], who has always welcomed me as her "permanent guest" in their lab and provided me with scientific and moral support whenever I asked. Furthermore, I cannot imagine pursuing a scientific question smoother than how we started the project on an efficient coding criticality. I literally walked into her office, explained the idea, and she provided whatever support needed for moving it forward (and being patient with its slow progress). Matthew Chalk also joined us at the very beginning on Anna's call and supported generously and I'm very thankful to him as well.

I'm also grateful to all members of my PhD advisory board and PhD committee for sparing time for the meetings and baring with all administrative works. Indeed, all members were already mentioned above except Martin Giese, who not only helped me as a member of my PhD advisory board and PhD committee, but also he was one of my greatest teachers during my neuroscience training in Tübingen; and Gustabo Deco, who was not only was a member of my PhD committee, but also being a constant source of inspiration; and Sonja Grün who not only was the external reviewer of my thesis, but also taught me much on developing statistical methods for neural data analysis by her rigorous methodological research.

I'm thankful to Fanis (Theofanis) Panagiotaropoulos and his group whom I conducted all my experimental work with non-human primates. In a similar vein, I'm also grateful to Vishal Kapoor, from whom I learn many unwritten tips and tricks for handling non-human primates. Also, with Fanis and Vishal (and later on Abhilash Dwarakanath) I started extensively exploring and understanding the fascinating phenomenon of binocular rivalry.

All the staff at Max Planck Institute for Biological Cybernetics, department of physiology of cognitive processes were supportive and helpful. In particular thankful to Conchy Moya who solely runs the administration of the lab. No need to say, technicians of the Max Planck Institute for Biological Cybernetics (including colleagues in workshops and animal facilities) were essential for conducting experiments, and I'm grateful to them. My special thanks to Joachim Werner for his *literally* 24/7 support for the lab. Joachim helpfulness and kindness was beyond system administration support, as a friend, he generously helped whenever I asked. In a similar vein, I'm thankful to the Graduate Training Centre of Neuroscience and International Max Planck Research School for Cognitive and Systems Neuroscience for all they support, in particular for being an important bridge to let me walk from my Physics world to the world of Neuroscience, and furthermore extensively explore this new field.

In the course of my research, I have approached many scientists to ask questions or even get consulting over multiple conversations. Thanks Larissa Albantakis and Erick Hoel for helping me to understand the IIT and causal emergence; thanks Afonso Bandeira and Asad Lodhia for multiple meetings they made to discuss issues related Random Matrix Theory (RMT); thanks Uwe Ilg for his help for analysis of Optokinetic Nystagmus (OKN) responses of monkeys.

I'm thankful to many colleagues and friends for their generous scientific and moral support. Roxana Zeraati for her indispensable scientific and moral support on countless occasions; Yusuke Murayama for his altruistic support, especially moral one; Behzad Tabibian for his pleasant company during the PhD (and beyond); Hadi Hafizi, Juan F Ramirez-Villegas and Kaidi Shao for their help in various stages of the PhD; Daniel Zaldivar for sharing his valuable experiences about various stages of his scientific career. Thanks to many colleagues at Max Planck campus, University of Tübingen and elsewhere for creating a scientific and friendly atmosphere around me: Ali Danish Zaidi, Andre Marreiros, Catherine Perrodin, Fereshte Yousefi, Franziska Bröker. Georgios Keliris, Hans Kersting, Hamed Bahmani, Hamid Ramezanpour, Hao Mei, Jennifer Smuda, Leili Rabbani, Leonid Fedorov, Maryam Faramarzi, Mingyu Yang, Mojtaba Soltanlou, Oleg Vinogradov, Parvaneh Adippour, Parvin Nemati, Reza Safari, Sina Khajehabdollahi, Tanguy Fardet, Yiling Yang, and NeNa organzier team of year 2014, 2015 and 2016.

During my PhD I was lucky to be part of multiple groups and communities. The first one was the OpenCon community. I'm thankful to all the people with whom I have interacted while being active in this community (and still interacting occasionally); in particular, Ali Ghaffaari, with whom we initiated developing the idea of *d-index* in OpenCon 2018. The second one was founding the Pro-Test Deutschland together with multiple dedicated fellows. During working with this dedicated group of people I realized the value, power, and importance of science communication. Toward the end of my PhD, I had the chance to be one of the ombudspersons in Max Planck Institute for Biological Cybernetics and also PhD representatives in a different period. These contributions taught me how critical is to care about mental health in academia. Most importantly, I should say being part of these communities, beyond what I learned about open science, science communication, and mental health, I recognized the power *and the value* of collective actions by people whose fuel is pure motivation.

Last and not least (actually the most), I should express my gratitude to my extended family. Just the presence of my old close friends, Amirhossein

Ketabdar, Hadi Hafizi, Hadi Masrour, and Mohsen Soltani somewhere on the planet was already morale for me, leaving aside their support whenever I asked! I'm also grateful for new friendships after moving to Germany, Andriana Rina, and Manuel Alexantro. Coming to family I should say, *Ohana* means family, family means nobody gets left behind or forgotten (Lilo & Stitch — see the video). I wish my father (who passed the way a long ago) also could read these lines and wish I could know his thoughts and feelings about the path of science that I chose for my life ... . I doubt that I can find the words to acknowledge the unconditional support and love of my partner, nevertheless I guess the these synergistic music-poem: music 1a/ music 1b - poem 1 and music 2a/ music 2b - poem 2 express [far better than my powerless words] how she has been helping me to get my/our bearings.

The words expressed here do not give full justice to all the support I have received and the adversity I have experienced. In particular, during my PhD, the unpleasantness of the latter has dominated the former, but it's not common to write about the latter in the "acknowledgment" section. I wish I could write an articulated acknowledgment section that could reflect both pleasant and unpleasant sides in a fair manner, but . . . ! Nevertheless, perhaps related to my unpleasant experiences, I should at least mention that some of the most valuable lessons I've learned during my PhD has been practicing being patient, the dos and don'ts for conflict resolution, many coping skills. Leaving out the non-zero probability of facing similar unpleasant situations, conflicts and the like can still arise. Therefore, learning these lessons will certainly be helpful. That being said, I should also be thankful for my adversaries that taught me all that. Furthermore, by going through these adversities, I've realized how severe lack of accountability, compassion and altruism can affect people and how easy is turning adversity to joy, simply by having a little bit of accountability, compassion and altruism. Finally, I hope this acknowledgment didn't undervalue the support of whom I've mentioned. Even worse, I hope I have not forgotten the support of someone due to the lapses. If that was the case, I truly apologize.

CONTENTS

PREFACE

Finding general principles underlying brain function has been appealing to scientists. Indeed, in some branches of science like physics and chemistry (and to some degree biology) a general theory often can capture the essence of a wide range of phenomena. Whether we can find such principles in neuroscience, and [assuming they do exist] what those principles are, are important questions. Abstracting the brain as a complex system is one of the perspectives that may help us answer this question.

While it is commonly accepted that the brain is a (or even *the*) prominent example of a complex system, the far reaching implications of this fact are still arguably overlooked in our approaches to neuroscientific questions. One of the reasons for the lack of attention could be the apparent difference in foci of investigations in these two fields — neuroscience and complex systems. This thesis is an effort toward providing a bridge between systems neuroscience and complex systems by harnessing systems neuroscience tools & notions for building empirical approaches toward the brain as a complex system.

Perhaps, in the spirit of *searching for principles*, we should abstract and approach the brain as a complex *adaptive* system as the more complete perspective (rather than just a complex system). In the end, the brain, even the most "complex system", need to survive in the environment. Indeed, in the field of *complex adaptive systems*, the intention is understanding very similar questions in nature. As an outlook, we also touch on some research directions pertaining to the adaptivity of the brain as well.

SUMMARY

The brain can be conceived as a complex system, as it is made up of nested networks of interactions and moreover, demonstrates emergent-like behaviors such as oscillations. Based on this conceptualization, various tools and frameworks that stem from the field of complex systems have been adapted to answer neuroscientific questions. Certainly, using such tools for neuroscientific questions has been insightful for understanding the brain as a complex system. Nevertheless, they encounter limitations when they are adapted for the purpose of understanding the brain, or perhaps better should be stated that, developing approaches which are closer to the neuroscience side can also be instrumental for approaching the brain as a complex system. *Chapter 1*

In this thesis, after an elaboration on the motivation of this endeavor in Chapter 1, we introduce a set of complementary approaches, with the rationale of exploiting the development in the field of systems neuroscience in order to be close to the neuroscience side of the problem, but also still remain connected to the complex systems perspective. Such complementary approaches can be envisioned through different apertures. In this thesis, we introduce our complementary approaches, through the following apertures: neural data analysis (Chapter 2), neural theories (Chapter 3), and cognition (Chapter 4).

In Chapter 2, we argue that multi-scale and cross-scale analysis of neural data is one of the important aspects of the neural data analysis from the complex systems perspective toward the brain. Furthermore, we also elaborate that, investigating the brain across scales, is not only important from the abstract perspective of complex systems, but also motivating based on a variety of empirical evidence on coupling between brain activity at different scales, neural coordination and theoretical speculations on neural computation. Based on this motivation we first very briefly discuss some of the relevant cross-scale neural data analysis methodologies and then introduce two novel methodologies that have been developed as parts of this thesis (Micro-Meso relationship, , and). In Micro-Meso relationship and we introduced a multi-variate methodology for investigating spike-LFP relationship and in we introduced a methdology for detecting cooperative neural activities (neural events) in local field potentials, that can be used as a trigger to investigate simultanious activity in larger and smaller scales. A prominent example of these neural events are sharp wave-ripples that has been shown to co-occur with precise coordination in the spiking activity of individual neurons and the large-scale brain activity as well.

Chapter 2
Chapter 5/Micro-Meso relationship
Chapter 6/
Chapter 7/

In Chapter 3, we introduce a new aperture through neural theories. One way of approaching the brain as a complex system is seeking for connections between theoretical frameworks that stem from the field of complex systems and the ones established in neuroscience. On the complex systems side, we consider the *criticality hypothesis of the brain* that has strong roots in the field of complex systems, and on the neuroscience side, we consider the *efficient coding* which is one of the most important theoretical frameworks in systems neuroscience. We first briefly introduce the background on efficient coding and criticality, and elaborate further on the motivation behind our integrative approach. In , we present our interim results, which suggests the two influential, and previously disparate fields – efficient coding, and criticality – might

Chapter 3
Chapter 8/

be intimately related. We observed that, in the vicinity of the parameters that leads to optimized performance of a network implementing neural coding, the distribution of avalanche sizes follow a power-law distribution. In we also provide an extensive discussion on the implication of our interim results and its future extensions. Moreover, in we also introduce another perspective which motivates such investigations, namely seeking for potential bridges between *neural computation* and *neural dynamics*.

In Chapter 4, we argue that binocular rivalry, as a key phenomenon to investigate consciousness, is particularly relevant for a complex systems perspective toward the brain. Based on this insight, we suggest and conduct novel experimental work, namely, studying this phenomenon at a mesoscopic scale, that has not been done before. Surprisingly, in the last 30 years, almost all the previous studies on binocular rivalry were either focused on micro-scale (level of an individual neuron) or the macro-scale (level of the whole brain). Therefore, our work in this domain not only is valuable from the perspective of complex systems, but also for understanding the neural correlate of visual awareness *per se*. In , , , and we elaborate on the outcome of this investigation. and were prerequisite for the binocular rivalry experiments. In we elaborate on the importance of studying prefrontal cortex (PFC) (which was the region of interest in our investigation) for understating the neural correlate of visual awareness. In we investigate the basic aspects of neural responses (tuning curves and noise correlations) of PFC units to simple visual stimulation (in a similar setting used for our binocular rivalry experiments). In and we investigate the neural correlate of visual awareness at a mesocopic scale (which is motivating from the complex system perspective toward the brain). We show that content of visual awareness is decodable from the population activity of PFC neurons () and show oscillatory dynamics of PFC (as a reflection of collective neural activity) can be a relevant signature for perceptual switches (). I believe that this is just the very first step toward establishing a connection from a complex systems perspective to cognition and behavior. Various theoretical and experimental steps need to be taken in the future studies to build a solid bridge between cognition and complex systems perspective toward the brain.

Chapter 4
Chapter 9/
Chapter 10/
Chapter 11/
Chapter 12/

The last chapter, Chapter 13, is dedicated to an outlook, a subjective perspective on how this research line can be proceeded. In the spirit of this thesis which is *searching for principles*, I believe we are missing an important aspect of the brain which is its *adaptivity*. At the end, brain, even the most "complex system", needs to survive in the environment. Indeed, in the field of *complex adaptive systems*, the intention is understanding very similar questions in the nature. Inspired by ideas discussed in the field of complex adaptive systems, I introduce a set of new research directions which intend to incorporate the adaptivity aspect of the brain as one of the principles. These research directions also remain close to the neuroscience side, similar to the intention of the research presented in this thesis.

Chapter 13

Part I

SYNOPSIS

This part provides a general idea of this thesis. We suggest an important approach that should be taken toward understanding the brain, could be borrowed or inspired from the field of *complex systems*. In light of this perspective, new questions can be asked in various domains and moreover, old questions can be revisited based on this perspective. Contents of this thesis, pertain to three different domains, namely *methods for neural data analysis*, *neural theories*, and *cognition*. In the first domain, we introduce novel statistical methods for multi-scale investigation of neural data that we believe should be an important piece in our analysis methods for understanding the brain as a complex system. In the second domain, we first briefly introduce *criticality hypothesis of the brain*, that has been primarily developed based on statistical physics and has been suggested to explain the complex dynamics of the brain activity in different spatial and temporal scales. Then we introduce our complementary approach of investigation in this framework, and our finding regarding the hypotheses. In the third domain, we first describe the importance of investigating bistable perception phenomenon from the perspective of complex systems. Then we discuss our finding pertaining the mesoscopic neural mechanism underlying this phenomenon.

1 BRAIN AS A COMPLEX SYSTEM

1.1 COMPLEX SYSTEMS

Behavior, or better stated *collective* behavior, of wide range of system spanning the scales of movement of atoms to behavior of humans/animals can be studied under an inclusive young framework of sdudying *complex systems* [25, 250, 158, 26]. Mitchell [250, Chapter 1] introduces and defines a complex system as following:

> *"Systems in which organized behavior arises without an internal or external controller or leader are sometimes called self-organizing. Since simple rules produce complex behavior in hard-to-predict ways, the macroscopic behavior of such systems is sometimes called emergent. Here is an alternative definition of a complex system: a system that exhibits nontrivial emergent and self-organizing behaviors."*

One of the characteristic properties of complex systems are their emergent properties, or/and their coordinated dynamics. Interactions between units of the system play a crucial role in the creating its emergent properties. These two aspects (emergent properties and the underlying interactions) of complex systems is central for the development of the ideas presented in this thesis (also see Chapter 13 for the complementary ideas).

To provide an intuition for emergent properties in complex systems and how interaction lead to such emergent properties, we exploit synchronization phenomena in a system made up of coupled oscillators. Assume we have N oscillators (indexed by i), each oscillates with frequency ω_i, where oscillation frequencies are drawn from a normal distribution with mean $\overline{\omega}$ and standard deviation β,

$$\omega_i \sim \mathcal{N}(\overline{\omega}, \beta) \, .$$

In absence of interactions between oscillators, the dynamics of each oscillator (which is defined based on its phase, θ_i) is governed only by its oscillation frequency,

$$\theta_j' = \omega_j \, . \tag{1.1}$$

Whereas, in presence of interactions between oscillators, they are allowed to exert forces on each other and therefore the dynamics of each oscillator also depends on the dynamics of other oscillators. These interactions are incorporated as an interaction term in the differential equation governing the dynamics of each oscillator (second term in Equation 1.2) [1]:

$$\theta_j' = \omega_j + \kappa \frac{1}{N} \sum_{i}^{N} \sin(\theta_i - \theta_j) \, , \tag{1.2}$$

where κ indicates the strength of these interactions.

The dynamics of system of oscillators described above is illustrated in Figure 1.1 (video) and Figure 1.2 (snapshots). Each dot represents an oscillator

1 The particular choice of interaction terms is made to ease the analytical treatment and for purpose of demonstration (see [193, 194] for more elaborate discussion).

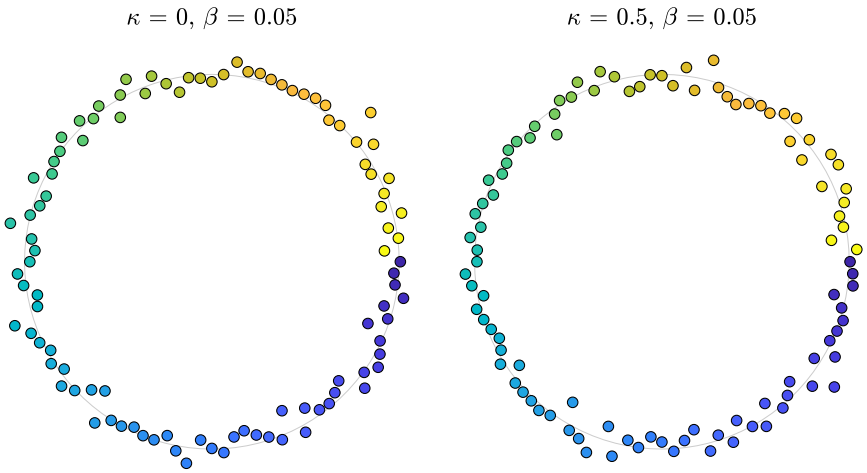


Figure 1.1: **Kuramoto model** (animation, need Adobe Acrobat Reader)
These animation demonstrate the dynamic of Kuramoto model consisting of 100 oscillators. Each dot represent an oscillator and the colors code for oscillator's intrinsic frequency. On the left, the oscillators do not interact with each other as the coupling parameter is set to zero ($\kappa = 0$). On the righ, the oscillators do interact with each other as the coupling parameter is non-zero ($\kappa = 0.5$).

and colors code for oscillator's intrinsic frequency. The oscillatory dynamics of the oscillators are represented by the circular motion of the dots. In the absence of interactions, as is evident in Equation 1.1, each oscillator, oscillates independently of the rest of the oscillators (Figure 1.1 and Figure 1.2 left). Nevertheless, in the presence of interactions and if the parameters of the system are appropriately chosen (in particular, κ, to be non-zero), the oscillators start synchronizing after a certain period (see Figure 1.2 second row, and compare simulations with and without coupling) and ultimatley all oscillators synchronize (see Figure 1.2 third row, and compare simulations with and without coupling).

Synchronization is not a genuine property of the individual units and there is no central coordinator in the system. However, oscillators tend to synchronize their activity due to the presence of interactions between the units. In this example, synchronization is considered an *emergent* property of the system.

The brain can also be conceived as a complex system, as it is made up of *nested networks of interactions* and demonstrates emergent-like behaviors such as oscillation. Different constructing units or building blocks of the brain (from molecules to networks) interact with each other [80, Chapter 1]. Indeed, this perspective toward the brain has been extensively articulated [329, 369, 337, 332, 264, 187, 58, 227, 36, 27, 250, 62, 75].

1.2 COMPLEX SYSTEM TOOLS IN NEUROSCIENCE

Inspired by perspective introduced in the previous section, various frameworks that stem from the field of complex systems has been adapted to answer neuroscientific question. Furthermore, various tools that have been developed for studying complex systems have also been customized to be applied to neural data.

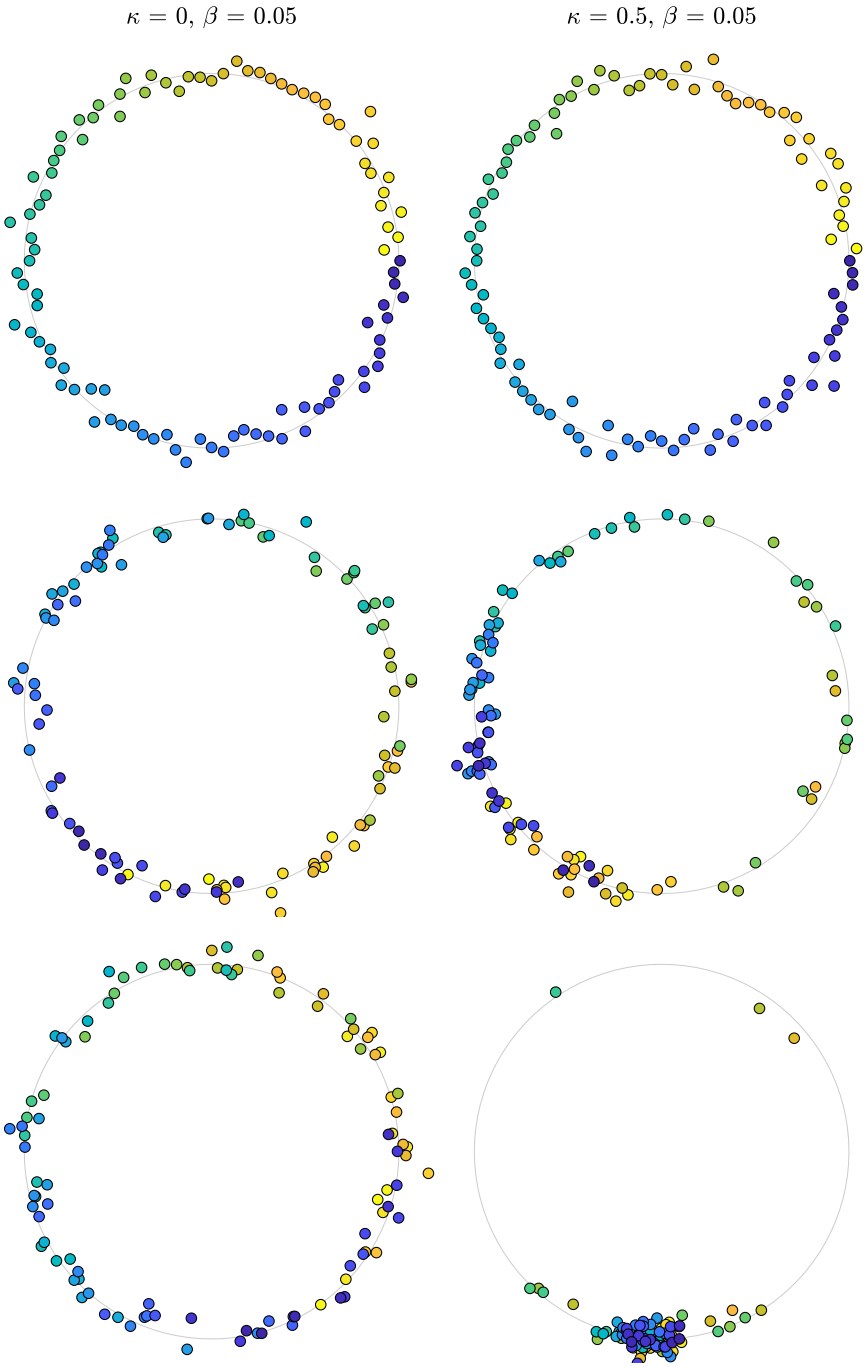


Figure 1.2: **Kuramoto model** (snapshots)
Snapshots from animations of Figure 1.1. These snapshots (each row, one snapshot) demonstrate the dynamic of Kuramoto model consisting of 100 oscillators. Each dot represent an oscillator and the colors code for oscillator's intrinsic frequency. On the left, the oscillators do not interact with each other as the coupling parameter is set to zero ($\kappa = 0$). On the righ, the oscillators do interact with each other as the coupling parameter is non-zero ($\kappa = 0.5$). The first row is a snapshot from the initial condition of the simulation, the second row is a snapshot from an intermediate state of the simulation, and the last row is the last snapshot of this simulation.

The tools and frameworks adapted from the field of complex systems to address neuroscientific questions can be divided into four categories (of courses, a subjective categorization): 1- Network science 2- Non-linear dynamics 3- Information theory and 4- Statistical physics.

NETWORK SCIENCE: Network science is perhaps the most adapted tool from the filed of complex systems to be used in neuroscience. To use tools developed in network theory, we abstract the object of interest as graphs, this includes defining the nodes and edges of the graph. Brain can also be abstracted as a graph in various levels of organization, from genes to behavior [48, 58, 357, 311, 355, 356].

NON-LINEAR DYNAMICS: Theory of dynamical systems has a broad application in neuroscience. The core idea is conceptualizing or modeling the dynamics of the brain at different scales as a [non-linear] dynamical system [243, 28, 282]. There have been various attempts to model single neuron [162], neuronal populations [137, Part 3][97], large-scale brain networks [163, 96] and even brain-environment system as dynamical systems [28].

INFORMATION THEORY: Information-theoretic tools have been extensively used in neuroscience, for purposes, as simple as studying neural coding in a single neuron [40, 358, 340, 49, 289] all the way to quantifying the level of consciousness [349, 23, 263] and providing a mathematical framework to represent the content of the conscious experience [24] (for a review see Tononi et al. [350]).

STATISTICAL PHYSICS: Statistical physics is a branch of physics which seeks for simple behaviors in systems consisting of many interacting components [315]. Such systems can be atoms of water in a glass [315], all the way to collective activity of a flock of birds [38, 37] and pattern of tweets in the Twitter network [141]. One of the phenomena that has been central in statistical physics (and other fields as well), is criticality. which has also inspired theoretical frameworks in neuroscience [256] (will be briefly discussed in Chapter 3).

1.3 NOVEL COMPLEMENTARY APPROACHES

Goal of the thesis

Certainly, using the approaches mentioned in the previous section (Section 1.2) has been tremendously insightful for understanding the brain as a complex system. This is an important achievement, given their principled and foundational nature. Nevertheless, they might also have some limitations when they are adapted for understanding the brain. For instance, information-theoretic measures are often difficult to apply to neural data in general settings due to the need for large amounts of data (but also see innovative approaches such as [377]). Such caveats become even more critical for functionally relevant information-theoretic measures such as integrated information [263]. Computing or estimating the amount of integrated information in a system for more than a handful of units is challenging [350]. There are other kinds of limitation for the mentioned approaches, but since the purpose of this thesis is introducing *complementary* (not alternative) approaches I would rather focus on these complementary approaches and the motivation behind them. In these complementary approaches, the goal is exploiting the development in the field of systems neuroscience to be

close to the neuroscience side but still remain related to the complex system perspective.

There are multiple examples in systems neuroscience, in which a given function is attributed to a *coordinated* activity of a group of neurons or neural units e. g. a brain circuit or an area. Just to name a few, we can mention population coding [308, 318], communication through coherence [131, 132], and memory consolidation [235, Chapter 7]. In these examples, the target function is implemented through the precise coordination of units; In population coding, by the interaction between neurons; in communication through coherence through oscillatory interaction through neural populations; And in memory consolidation through interaction between multiple regions of hippocampal formation and neocortex.

Interestingly, some of the tools and notions that system neuroscientists used to understand the coordinated phenomenon can be closely related to perspectives inspired by or related to the field of complex systems. For instance, various studies have investigated cross-scale relationships in neural activities such as relationship between spikes and local field potentials (LFP) [251] for understanding the mechanism involved in communication through coherence, or considering simultaneously two successive scales such as neural event triggered fMRI (NET-fMRI) studies to understand the memory consolidation mechanisms [220, 222, 283].

In Chapter 2, we introduce a set of methodologies for cross-scale and multi-scale analysis of neural data. Developing these tools is motivated by a perspective that results from approaching the brain as a complex system. Every system, in particular, complex systems can be described at different scales. Some systems (e. g. our solar system) can be described, to a large degree, in *isolated scales* and their behavior upon interacting with other systems can be predicted. However, many systems wherein we are interested to understand are not well described in isolated scales. To illustrate this important notion, we use a few intuitive examples adopted from Bar-Yam (2017). If we are interested in explaining the dynamics of the earth (orbits of the earth in the solar system), and how it will change when a new planet is added to the solar system, we do not need to know the details of processes happening inside the earth. Therefore, for this system, we can *separate scales* without losing our descriptive and predictive power (to a large degree). But if we are interested in the collective dynamics of a flock of birds [38], we neither can focus on the micro-scale (motion of an individual bird) as it is too fine-grained, nor the macro scale (average motion of the flock) as it is not sufficient to describe and predict the collective behaviour of the birds. Generally speaking, understanding the complex behavior which is not completely independent (random) nor it is completely coherent requires investigation across scales [26]. In Chapter 2, we further elaborate on the motivation and necessity of investigating the brain by simultaneously considering two successive scales and introduce our novel methodologies motivated by this mindset.

Approaching through neural data analysis

As mentioned earlier, the goal is establishing a bridge between systems neuroscience and a complex system perspective toward the brain. In an effort toward achieving this goal, in addition to developing analysis methods and generalizing the existing ones, we also propose two other apertures in Chapter 3 and Chapter 4. Of course these new apertures also provide us new angles to build the bridge.

In Chapter 3 we provide a potential link between one of the most important theoretical frameworks in system neuroscience, *efficient coding*, and one of

the most important theoretical framework in the field of complex systems, *criticality*. Efficient coding has different variants and many of them have been extensively investigated both experimentally and theoretically in systems neuroscience. On the other hand, the theory of critical phase transition in complex systems has been successful in explaining many phenomena in nature [240, 76], and "criticality hypothesis of the brain" [256], has been developed based on this solid foundation. In nutshell, criticality hypothesis of the brain state that, the brain operates close to a critical state. Being close to this state is beneficial for such an organ [256, 347, 253], as it has been shown that general information processing capabilities such as sensitivity to input [183, 54], dynamic range [183, 195, 262], and information transmission and storage [324, 359, 359, 224, 237], and various other computational characteristics are optimized in this state. Certainly, being in a state with such optimized capabilities are relevant for the computations in the brain, but they are too abstract to provide a concrete explanation of the computations in the brain. For instance, all the capabilities mentioned above are relevant for coding sensory information which is a relevant function for the brain and has been studied in systems neuroscience extensively, however mere adjustment for being close to criticality cannot provide a neural implementation for the coding given resource constraints. In Section 3.3.1 we provide more detail on both frameworks, efficient coding and criticality hypothesis of the brain, and provide evidence on the connection between them.

Approaching through neural theories

In Chapter 4, we introduce another aperture for establishing the mentioned connection. Perhaps, one of the most important goals of neuroscience is understating the machinery behind the cognitive capabilities of the human brain and behavior. In the first two approach we focused on method of neural data analysis and theories, and in the third approach, the focus is on cognition. We suggest bistable perception is a behavioral cognitive phenomenon that is relevant for the perspective we introduced. This approach can be motivated, based on the fact that bistable perception can be explained to some degree based on tools from complex systems (see Section 1.2). For instance, spontaneous transitory behavior that has been observed in bistable perception, to some degree, can be explained based on principles of statistical physics [39, 16] or the dynamics of the neural population can be explained by network models that are operating on the edge of a bifurcation [346, 273]. In Chapter 4, we introduce briefly the phenomenon of bistable perception, then we justify its importance from the perspective of complex systems approach to the brain. Perhaps this is the closest to one of the ultimate goals of systems and cognitive neuroscience, and the most distant from the complex systems approach. To minimize this gap we suggest and conduct novel experimental work, namely, studying the phenomena on a mesoscopic scale which has not been done before. I believe that this is just the very first step toward establishing the connection such close to cognition and behavior. Various theoretical and experimental steps need to be taken in the future studies to build a solid bridge between complex systems perspective toward the brain and cognition.

Approaching through behavior and cognition

2 APPROACHING THROUGH NEURAL DATA ANALYSIS

Based on the motivation elaborated in Chapter 1, we believe multi-scale and cross-scale analysis of neural data is one of the important aspect of neural data analysis from the complex systems prospective toward the brain and indeed is one of the apertures through which, we can seek for the complementary approaches mentioned in Section 1.3. In this chapter, after further elaboration on the need for multi-scale and cross-scale analysis of neural data, very briefly we discuss some of the relevant cross-scale neural data analysis methodologies and then introduce two novel methodologies that has been developed as part of this thesis.

2.1 NECESSITY OF INVESTIGATING ACROSS SCALES

As it was briefly discussed in Section 1.3, understanding behavior in a system whose components are neither behaving completely independent nor completely coherent, requires investigation *across scales* [26, 117]. Certainly, the brain is a prominent example of such systems [117]. Perhaps the most intuitive aspect of the brain which demonstrates this point is its oscillatory dynamics. As Chialvo (2010) pointed out,

> *"Recent work on brain rhythms at small and large brain scales showed that spontaneous healthy brain dynamics is not composed by completely random activity patterns or by periodic oscillations[62]".*

In order to investigate the brain across scales, first we need to clarify what is considered as the scale. In this thesis, we refer to different *levels of organization* as scales. Brain is organized in different *levels* (Figure 2.1).

These levels range from scale of molecules all the way to large scale brain networks [80, Chapter 1]. Different phenomenon might primarily be explained in a limited range of these levels. For instance, synaptic transmission, which is a basic form of communication in the brain, occurs at fairly small spatial scales, i. e. level of molecules, synapses, and neurons. Nevertheless, certain processes involve a broad range of levels. For instance in memory consolidation, processes from gene expressions at the level of dendrites are involved, all the way to larger-scale network **re**organization. Therefore, one expects that process happening at different levels of organization to be related to each other. It is worth to mention that, our understanding (especially from a theoretical perspective) should be consistent across the levels of organization. As elegantly described in Churchland and Sejnowski [80, Chapter 1]:

> *"... the theories on one level must mesh with the theories of levels both higher and lower, because an inconsistency or a lacuna somewhere in the tale means that some phenomenon has been misunderstood. After all, brains are assemblies of cells, and something would be seriously amiss if neurons under one description had properties incompatible with the same neurons under another description."*

Indeed, there are various empirical evidence on predictions across scales and relationships between scales: From single neurons to microcircuits [286,

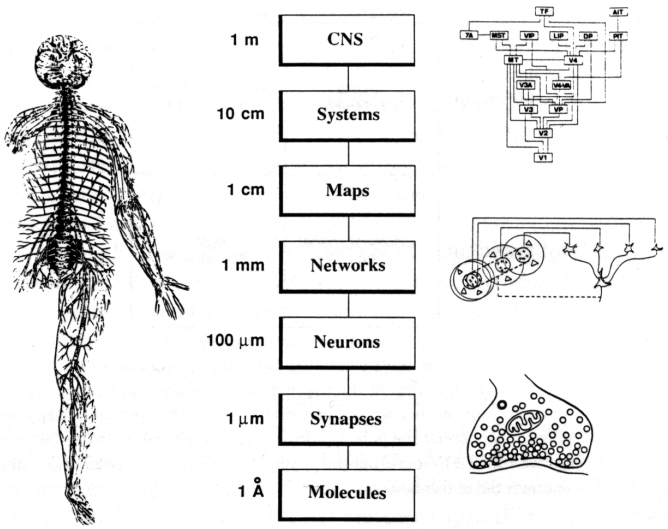


Figure 2.1: **Schematic depiction of levels of organization**
Demonstrate extremely variable spatial scales at which anatomical organizations can be identified. Icons to the right represent structures at distinct levels: (top) a subset of visual areas in visual cortex; (middle) a network model of how ganglion cells could be connected to simple cells in visual cortex, and (bottom) a chemical synapse. Figure is adopted from Churchland and Sejnowski [79] with permission.

285], from microcircuits to a single brain area [211], from a single area to the whole brain [313, 383]. In some cases, the cross-scale coupling is closely and causally related to a specific function, such as global state changes that have been shown in a study by Li et al. [211]. They showed that burst spiking of a single cortical neuron in somatosensory cortex can induce a global switch between the slow-wave sleep and Rapid-Eye-Movement (REM) sleep. In some cases, cross-scale relationships are even mechanistically interpretable as well. For instance, it has been demonstrated that spiking probability can be modulated by the underlying network oscillation. Network oscillations modulate the membrane potential of the neuron and that leads to the different levels of excitability for the given neuron. Depending on the phase of the underlying oscillation, this can lead to a higher or lower probability of spiking activity [365, 146]. Based on these simple mechanisms, *coordination by oscillation* has been hypothesized, and this lends support to various cognitive functions such as attention. The hypothesis of "Coordination by oscillation" proposes that network oscillations modulate differently the excitability of several target populations, such that a sender population can emit messages during the window of time for which a selected target is active, while unselected targets are silenced [132, 372, 131]. Overall, I believe, considering *two successive scales simultaneously*, is a principled approach for understanding collective or coordinated organizations in neural systems. Furthermore, as mentioned in Section 1.3 this approach is also justified by empirical evidence.

Investigating across scales can also be motivated from a more abstract (and perhaps more fundamental) perspectives: In dynamical systems with non-linear interaction there are various examples where activity in different scales are related [199]. One example for such non-linear dynamical systems is the Kuramoto model. As described briefly in Section 1.1, Kuramoto model

describes a system of multiple coupled oscillators [193, 194] (for an integrative review see [2]). In this model, the activity of individual oscillators is related to quantities pertaining to the average or mean-field activity of the system as a whole. More precisely, the phase of an individual oscillator can be related to the mean phase of oscillators and their phase coherence. Such core ideas from the theory of dynamical systems went beyond mere conceptual connections, but also inspired unifying formulations for neural oscillations in the brain (e. g. see [53]). For more detailed elaboration on motivations from the theory of dynamical systems for cross scales investigation of the brain see works of Le Van Quyen and colleagues [200, 198, 199].

The other abstract motivation for investigation across scales is the nature of computation in the brain. The brain is a naturally evolved biological information processing system. Therefore, the computational strategies or solutions served by the brain can be quite different from engineered information processing systems [80, Chapter 1][108]. The main difference between commonly engineered information processing systems and natural information processing systems is that the latter is constrained by the existing form of evolving organisms. As elaborately framed by Churchland and Sejnowski [80, Chapter 1]:

> *"Evolutionary modifications are always made within the context of an organization and architecture that are already in place. Quite simply, Nature is not an intelligent engineer. It cannot dismantle the existing configuration and start from scratch with a preferred design or preferred materials. It cannot mull the environmental conditions and construct an optimal device."*

Furthermore, there are other aspects that need to be taken into account in the process of thinking about the solution chosen by the brain. For instance, humans/animals are constrained by the response time (they need to be fast enough) to be able to survive in their natural environment. Finding the solution for the required computation is expected to happen in a few hundred milliseconds. This becomes even more puzzling if we take into account the computational machinery in the brain that is orders of magnitude slower than artificial information processing systems. Events in neurons happen in range of milisecond (10^{-3}) as opposed to nano second (10^{-9}) in electronic computers [108]. Other such examples are, spatial constrains (limitation by available space), energy consumption, and metabolism [80, Chapter 1]). **All being said to minimize the surprise of mentioning novel proposals (in the following) on brain computational principle that pertains to cross-scale investigation.** Bell [31, 32] proposes that, the adaptive power of biological information processing systems comes from the gating of information flows across levels, both upward and downward, as Bell [32] stated:

> *"There is thus no "functionalist cut-off level" anywhere in the biological hierarchy Nature does not seem to shield the macro from the micro in the way that a computer does."*

Although, to the best of my knowledge, this proposal is not yet formalized as a complete theoretical framework, but perhaps it gains some empirical support through recent experimental and computational studies of *ephaptic* interactions in the brain. In recent years, we have experimental [9] and modeling [9, 294, 319] on the possibility of having ephaptic interactions in the brain (for a review also see [8]). Indeed, this evidence that electrical fields in the brain can functionally modulate the activity of neurons is in line with Bell [31, 32] proposal on the computational architecture of the brain.

Overall, I believe the arguments provided above, justify the necessity of investigating brain activity across scales. In spite of the importance of this need for understating the brain, there are not sufficient methodologies for the multi-scale investigation of the brain activity In the next two sections (sections 2.2 and 2.3) we provide a brief overview of available tools and our contribution of novel methods for cross-scale investigation of brain dynamics.

2.2 AVAILABLE TOOLS FOR INVESTIGATING CROSS-SCALE RELATIONSHIPS

Brain activity can be measured using various experimental methodologies at different scales (Figure 2.2). For instance, it can be spike trains from indi-

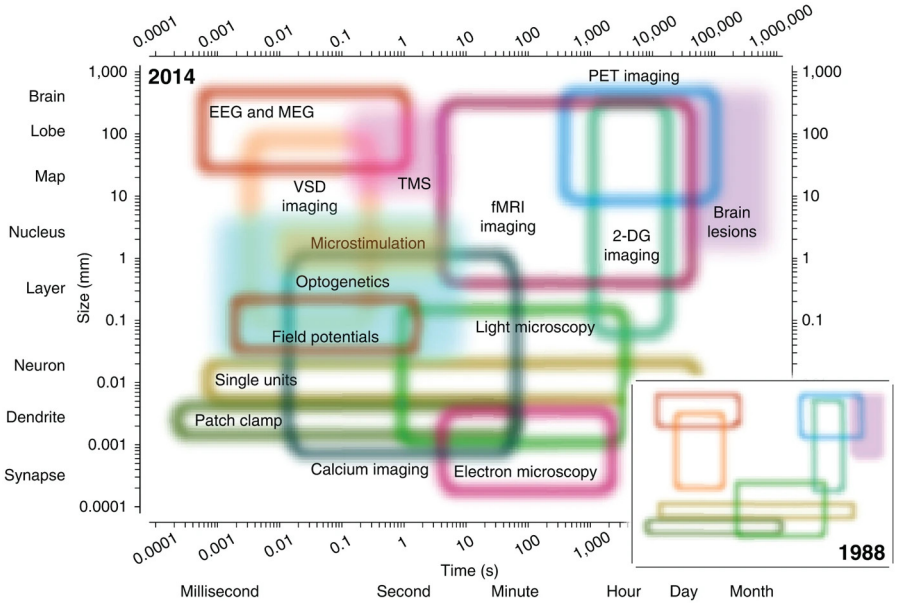


Figure 2.2: **Spatio-temporal resolution of measurement methods in neuroscience** Demonstrate the spatial and temporal resolution of measurement methods being used in neuroscience (up to 2014). Each box depict the spatial (y-axis) and temporal (x-axis) of one measurement method. Open regions represent measurement techniques and filled regions, perturbation techniques. Inset, a cartoon rendition of the methods available in 1988. The regions allocated to each domain are somewhat arbitrary and represent the estimate of Sejnowski et al. [314]. Abbreviations used in the figure: EEG, electroencephalography; MEG, magnetoencephalography; PET, positron emission tomography; VSD, voltage-sensitive dye; TMS, transcranial magnetic stimulation; 2-DG, 2-deoxyglucose. Figure is adopted from Sejnowski et al. [314] with permission.

vidual neurons, field potentials generated by small or large population of neurons or hemodynamic signals from the whole brain. Our novel development for bridging scales pertains to the relationship between, spiking activity Local Field Potentials (LFPs) and Blood-Oxygen-Level Dependent (BOLD) signals.

A number of tools have already been developed and applied to neural data, and they gave us insight into the relationship between brain activity in different scales. Here we mention very briefly a subset of such methods that are related to novel development that we introduce in the next section.

The relationship between spiking activity and LFP has been studied extensively in the context of mechanisms for coordination by oscillation in the brain. Indeed, this was one of the examples briefly discussed in Section 2.1 to motivate understanding cross-scale relationships. Various techniques has been developed for investigating the relationship between spiking activity and LFP [378, 15, 363, 362, 164, 213, 375]. Most of the approaches for investigating the spike-LFP coupling are restricted to pairwise first-order statistics of spike-LFP interactions. Given the various experimental advances, there is a growing need for conceptual and methodological frameworks to investigate this relationship in multi-variate settings (see further elaboration in Section 2.3.1).

Another line of research pertaining to cross-scale relationships, is investigating the relationship between LFP and fMRI BOLD signals. In this branch, extensive research has been done toward understanding the neural correlate or neural activity underlying the BOLD signal [223, 217, 219, 138, 374]. Methods used for exploring the relationship between these signals were conventional correlation analsysis [223], system identification [223], Canonical Correlation Analysis (CCA) and its time-resolved kernelized version [41, 257]. Certainly, the mentioned investigation shed light on the basic nature of the coupling between LFP and fMRI BOLD, but more developments needed to get into functionally relevant couplings.

Mentioned developments construct the foundations and moreover led to important methodologies for addressing questions concerning functional implications of investigating the relationship between LFP and BOLD fMRI. Along the same line of developments, Neural-Event-Triggered (NET) fMRI was also introduced recently. In NET-fMRI, characteristics neural activities of such as Sharp Wave-Ripple (SWR) are used as events to align and average the time course of large-scale brain activity to extract the global signature of the given events. Indeed, ripple-triggered activities in macaque monkeys revealed important large-scale coordination involved in the process of memory consolidation [222].

NET-fMRI can be a very informative methodology if the *event* is already well-defined. Nevertheless, there are very few such well-characterized neural activity like SWR. Therefore, we need novel methodologies to detect and characterize such distinct neural activities (see further elaboration in Section 2.3.2).

2.3 NEED FOR NEW TOOLS FOR INVESTIGATING CROSS-SCALE RELATIONSHIPS

As was motivated in the previous section (Section 2.2), novel methodologies are needed for investigating the brain dynamics across the scale. LFPs are signals at meso-scale [214], which is an intermediate scale between micro- and macro-scale, and they reflect a mesoscopic picture of the brain dynamics. LFPs result from the superposition of the electric potentials generated by ionic currents flowing across the membranes of the cells located close to the tip of recording electrodes. The LFP reflects neural cooperation due to the anisotropic cytoarchitecture of most brain regions, allowing the summation of the extracellular currents resulting from the activity of neighboring cells and potentially remote populations. As such, a number of subthreshold integrative processes (i.e. modifying the neurons' internal state without necessarily triggering spikes) contribute to the LFP signal [63, 214, 118, 149, 274]. As LFPs are rich and intermediary signals, they can be a pivotal point

for bridging the scales. We can better illustrate the importance of LFP for cross-scale analysis with an example. In LFPs, certain characteristics of neural activities, like SWRs are detectable. Interestingly, SWRs occur concurrently with well-coordinated activity at smaller scales (neurons and population of neurons), and as well as a larger scale (entire brain). For the connection to smaller scales (microscopic scale) various studies suggest SWRs emerge in the CA1 mainly due to afferent CA2- and CA3-ensemble *synchronous* discharges [89, 89, 265]. For the larger scale (macroscopic scale), as briefly mentioned earlier, concurrent recording of BOLD signal of the entire brain and SWRs, demonstrate large scale coordination of entire brain activity during SWRs [222].

Detecting characteristic activities like SWRs and finding such relationships across scales (exemplified in the previous paragraph) was the result of years of experimental work and exploration in the data. Developing new tools that allow us to find such characteristic patterns (like SWRs) in an unsupervised fashion and finding their relationship to measurement at other scales [e. g. with synchronization measures and NET-fMRI] can be of paramount importance.

Based on the ideas and motivation elaborated above, we first focus on tools that allow us to explore the relationship between spikes and LFPs (Section 2.3.1) and then, a method for the detection of neural events in an unsupervised fashion (Section 2.3.2).

2.3.1 *Tools to explore micro-meso relationsips*

A prominent example of the relationship between micro- and meso-scale activity in the brain is the spike-field coupling. Apart from its importance from the perspective discussed in Section 2.1, the synchronization between spiking activity and the phase of particular rhythms of LFP has been used as an important marker to reason about the underlying cooperative network mechanisms. Nevertheless, there is not yet a systematic way to extract the coupling information from the largely multi-variate data available to state-of-the-art recording techniques [104, 168, 167] with hundreds or even thousands of recording sites [274, 168, 60, 134]. We developed a multi-variate extension of phase-locking analysis and a statistical testing framework to assess the significance of the coupling strength. With our method (which we call Generalized Phase Locking Analysis – GPLA), we can quantify, characterize, and statistically assess the interactions between population-level spiking activity and mesoscopic network dynamics (such as global oscillations and traveling waves).

We demonstrate the capability of the GPLA by applying the method to various simulated and experimental datasets. For instance, the application of the method on simulation of hippocampal SWR can reveal various characteristics of hippocampal circuitry with minimal prior knowledge. GPLA reveals CA1 and CA3 neurons are all coupled to the field activity in the gamma and ripple band (in line with experimental and simulation results [64, 284]), suggesting this rhythm may support communication between CA1 and CA3 sub-fields during memory trace replay. Furthermore, it also allows us to tease apart the involved populations and provide hint on the communication flow from CA3 to CA1 based on label-free spike timing and LFP. As another example, the application of the method on the experimental recordings from Prefrontal Cortex (PFC) suggests a non-trivial coupling between spiking activity and LFP traveling waves in this region of the PFC. Assuming LFPs

mostly reflect local and distal input post-synaptic currents to the underlying neural population, analysis based on the GPLA accompanied by neural field simulations suggest that a connectivity structure consists of long excitatory horizontal connections and strong local recurrent inhibition as a plausible speculations for these PFC recordings (in line with previous modeling and experimental studies [298, 322, 321].

Notably, an important component of our methodological contribution for investigating the relationship between micro- and meso-scale activity is the theoretical significance test for GPLA. We describe the theoretical foundation of the test in Safavi et al. [301] (also can refer to the corresponding summary, Micro-Meso relationship) and the necessary development for practical applications on neural data is described in Safavi et al. [306] (also can refer to the corresponding summary,). In our theoretical investigation, we derive analytically the asymptotic distribution of Phase-Locking Value (a uni-variate coupling statistics which is conventionally used for quantifying spike-LFP coupling), which follows a Gaussian distribution. The implication of these results for neural data is, whitening of LFPs and normalization by the square root of the spike rate is necessary for the applicability of our theoretical results on neural data. The asymptotic distribution for the uni-variate coupling was key for the development of the statistical test for the multivariate version of phase-locking analysis. Based on Gaussianity of the uni-variate measure and random matrix theory we could derive the theoretical null distribution for the singular values of a matrix containing all pairwise coupling that we call the coupling matrix. Consequently, we show that singular values of such matrices converge to a Marchenko-Pastur distribution [236]. [1] This is a well-established asymptotic behavior in random matrix theory for matrices with independent normally distributed entries [10]. The key is Marchenko-Pastur distribution has an upper bound, meaning that, under the null condition (no coupling between spike and LFP) largest singular value of the coupling matrix should not exceed this upper limit. If the singular values resulting from data are larger than this upper limit, then there is significant coupling between the population spikes and the multi-channel LFPs. Developing a theoretical test is of paramount importance considering the constantly increasing dimensionality of modern recording techniques.

2.3.2 *Tools to explore meso-macro relationsips*

As pointed out in Section 2.3, it is important to develop tools that allow us to find characteristic patterns of LFPs (such as SWRs) in an unsupervised fashion. Such patterns are potentially very special, in the sense that, they provide us a time window that meso-scale dynamics is closely related micro and macro scale dynamics. In fact, this is of paramount importance for bridging the brain activity in different scales.

We developed an unsupervised methodology based on Non-negative Matrix Factorization (NMF) and dictionary learning to detect transient cooperative activities in a single channel LFP (see for more details). Such activities were also introduced as *neural events* in previous studies [222, 221, 283]. With this method, is not only possible to detect well-established characteristic patterns such as sharp wave-ripples, but also new characteristic neural activities that have not been identified and studied before. We demonstrate the

1 Marčenko and Pastur [236] in not written in English, but is the original publication. The reader can refer to Anderson et al. [10, Chapter 2] instead.

capability of our method by identifying neural events in Hippocampus and LGN and also and explored their brain-wide *macro-scale* signatures using concurrent fMRI recordings from anesthetized monkey. The result suggest that, similar to the previous study of Logothetis et al. [222] that was focused on sharp wave-ripples, the identified events in Hippocampus and LGN reflect a large scale coordinated dynamics, namely a competition between cortical and subcortical regions.

Furthermore, neural events can also be informative for exploring micro-scale and meso-scale relationships. By exploiting a simulation of thalamo-cortical circuitry developed by Costa et al. [86], we demonstrate that such events have the potential of even relating meso-scale dynamics to *micro-scale* dynamic, even at the cellular level. With our methodology we identified different kinds of spindles in the activity of the thalamus module of the simulation (indeed, this is another demonstration for the capability of the method), and demonstrate that different events co-occur with a characteristic activity pattern in cellular variables (such as membrane potentials and ionic currents) of the simulation.

3 APPROACHING THROUGH NEURAL THEORIES

As motivated in Chapter 1, in order to achieve the target bridge between complex systems and neuroscience, i. e. approaching the brain as a complex system by exploiting systems neuroscience tools and notions, one of the apertures through which, we can seek for the complementary approaches is neural theories (see Section 1.3). In this chapter we aim to explore two important theoretical frameworks, one closely related to the field of neuroscience, and one to complex systems. In order to establish the mentioned bridge, we explore the potential connection between them. On the neuroscience side, we consider *efficient coding* which is one of the most important theoretical frameworks in systems neuroscience, and on the complex systems side, we reflect on the *criticality hypothesis of the brain* that has strong roots in the field of complex systems. We first provide a brief overview on each of them, and then their potential connection.

3.1 CRITICALITY HYPOTHESIS OF THE BRAIN

In the course of studying the state of the matter (e. g. water, steam and ice as states of H_2O) and their phase transitions (e. g. transition from water to vapor) physicists discover some *universal* behavior in a variety of phase transitions (e. g. freezing of water and magnetization in metals [312, Chapter 5] as well as in wider ranges of natural phenomenon such as human social behavior [69] (see Mathis et al. [241] and Bar-Yam [26] for other examples). Later on, in the process of examining the relationship between microscopic variables like speed of atoms and macroscopic variables like temperature, it has been realized that, close to a critical point the usual methods fail to establish these relationships. The critical point (for water) is the point where fluctuations between liquid-like and vapor-like densities extend across the system so that the system is not smooth anymore and therefore averages are not well behaved. Furthermore, this characteristic inharmonious behavior was observable at all scales [26]. Indeed, the method of Renormalization Group (RG) has been developed to investigate mathematically such state of a system and has been applied on a wide range of systems. It turns out, in spite of differences in details of various systems (e. g. magnetic dipoles and molecules of water), their behavior can be explained based on the RG method. This important observation, led to the notion of *universality*, that allow us to explain various systems with many interacting components with a small set of variables and some scaling relations.

Based on these fundamental ideas *criticality hypothesis of the brain* has been proposed [256]. Roughly speaking, criticality hypothesis of the brain states that, brain operates close to a critical state, a state on the edge of transition between order and disorder. The first experimental evidence on scale-freeness of the brain dynamics (as one of the signatures of criticality – see Section 3.2) has been reported almost two decades ago by Beggs and Plenz [30]. Later on such scale-free dynamics have been observed in various smaller and larger scales as well. To name a few, see Bonilla-Quintana et al. [47] at the scale of actin in dendrites, Johnson et al. [166] at the scale of neuronal membranes, Varley et al. [360] at the scale of the entire brain (for

more references see [256, 3]). Moreover, being close to this state is beneficial for the brain [256, 347, 253], as it has been shown that general information processing capabilities such as sensitivity to input [183, 54], dynamic range [183, 195, 262], or information transmission and storage [324, 359, 224, 237], and various other computational characteristics has been also considered to be relevant [354, 342, 152, 151, 244, 179, 154, 247, 367, 125, 382] (also see [29, 323, 381] for a reviews).

To summarize, multiple studies have reported signatures of criticality observed in various neuronal recordings at different scales, and theoretical investigations demonstrated various aspects of information processing are optimized at the second-order phase transition (see references in [256, 3]).

3.2 SIGNATURES OF CRITICALITY IN NEURAL SYSTEMS

As motivated in the previous section, various empirical and theoretical investigations lend support to criticality hypothesis of the brain, and signify the potential functional relevance of the criticality hypothesis of the brain. Therefore, it has been motivating to search for diverse signatures of criticality in the brain. These signatures can be categorized into three groups [380]: scale-freeness neural activity (avalanche criticality), dynamical regime of the neural system (edge of bifurcation criticality), and thermodynamic of the neural data (maximum entropy criticality).

AVALANCHE CRITICALITY: Scale-free cascade of activity is a ubiquitous type of dynamics in nature: For instance in interacting tectonic plates [139], forest fires [233], nuclear chain reactions [145], threshold-crossing events that appears as one unit (e. g. a tree) exceeding a threshold (e. g. a tree fires) and because the units of the system are coupled to each other, similar threshold-crossing events *propagate* through other units of the system. Such propagating dynamics can lead to large *avalanches* of activity. Almost two decades ago Beggs and Plenz [30] observed similar cascades in activity of in-vitro neural populations and later on others reported such scale-free cascades at various other neuronal recordings in various scales (see references in [256, 3]). Truly critical systems, not only should show the mentioned scale free dynamics, but also they should follow the scaling laws introduced by Sethna et al. [316], that were observed in neural data [130] as well [1].

BIFURCATION CRITICALITY: When a dynamical system has a transition from one dynamical regime to another (such as transition from order to chaos), it experiences a *bifurcation* [162, 52, 82]. The point where the transition happens is also denoted as the critical point. There are various kinds of bifurcations (see Izhikevich [162]), but some of them have been particularly appealing for understating the dynamics of the brain as well as computation in the brain. Without getting into the theoretical details of these bifurcations and in very brief fashion, transitioning from order to chaos [33], and transitioning from an asynchronous to a synchronous state [103] have been considered as two important bifurcations for the brain (for further elaboration see Cocchi et al. [82], Muñoz [256] and references therein). Avalanche criticality and bifurcation criticality can co-occur, when there is a continuous phase transition [82] (for example see [230, 278]), nevertheless, Kanders et al.

1 Indeed, scale-free neural avalanches without following scaling laws have been observed in neural models that are not operating close to a critical point [4, 352].

[171] proposed that these two types of criticality do not necessarily co-occur and therefore should be attributed to two distinct phenomena.

THERMODYNAMIC CRITICALITY: Statistical mechanic provides a powerful framework to study collective behavior in systems consisting of interacting units with many degrees of freedom [315]. Tools from statistical mechanic have been applied in neural networks in order to understand their collective dynamics [7]. Along the same line Tkacik et al. [348] approached the activity of neurons from a thermodynamical perspective. They define a Boltzman-like distribution, derive various thermodynamic quantities such as heat capacity based on estimated Boltzman distribution, and ultimately define criticality based on thermodynamic quantities (like divergence of heat capacity). Moreover, in empirical data this novel framework is applicable and functionally relevant. This novel formulation introduces another signature or definition of criticality in neural system [348] (but also see [261]).

3.3 SEEKING FOR A BRIDGE: A COMPLEMENTARY APPROACH

As mentioned earlier, over the last two decades, multiple experimental and theoretical investigations lend support to criticality hypothesis of the brain. In particular, as it was briefly discussed in Section 3.1, closeness to criticality has been suggested to be an optimal state for information processing. To evaluate how closeness to criticality can be beneficial for the information processing in the brain, the common approach is using a model (e. g. a branching network, a recurrent neural network) that can attain various states (including critical and non-critical states), depending on control parameters (e. g. branching ratio, connection strength) of the model. Then by quantifying how general information processing capabilities such as information transmission depend on the control parameters, the advantages of being close to a critical state can be assessed. For instance, if information transmission in the model under study is optimized exclusively close to the critical state of the model (defined based on the control parameter(s)), then it can be considered as evidence for relevance of usefulness of criticality for the brain.

Indeed, one of the important reasons for the relevance of the criticality for the brain is the optimized information processing capabilities that operating close to this state offers. Nevertheless, the *optimized setting* implied by criticality hypothesis, does not imply any specific computation that the brain may need to execute, but rather *general* capabilities for computation [2]. For instance, being in a state which is optimized to have the maximum sensitivity to input [183, 54], and maximum dynamic range [183, 195, 262] are all relevant capabilities for coding sensory information, but mere adjusting for the closeness to criticality cannot provide a neural coding algorithm and its implementation for coding given resource constraints. In contrast, there are frameworks (such as efficient coding) that provide the functionally relevant objectives to be maximized or minimized (which define the optimized computation), the algorithm of computation (neural coding algorithm) and the neural implementation. Therefore, we think we need complementary approaches to

2 See also Lizier [215] (in particular chapter 6) that argue closeness to criticality is a sate where [some] computing primitives (such as information storage, transfer and modification) are optimized. Furthermore, an complementary perspective is, non-critical states can be specifically advantageous for a particular computation, and therefore brain needs to be able to flexibly switch between them [81, 380].

criticality that can bridge the gap between criticality and frameworks which focus on *functionally relevant* computations and their implementations.

3.3.1 *Efficient coding as the computational objective*

We focus on *coding*, as a functionally relevant computation (and with the ultimate purpose of establishing the bridge to criticality). Efficiency of neural coding is particularly important, as sensory systems have evolved to transmit maximal information about incoming sensory signals, given internal resource constraints (such as internal noise, and/or metabolic cost) [292, Chapeter 13][289, 281]. Indeed, models using this simple principle made various verified predictions about neural responses (e. g. receptive field in in V1 [266, 331]).

Several variants of efficient coding have been developed (for a brief overview see [72]). Depending on the answers to qualitative questions like, " *What should be encoded? What sensory information is relevant? What can be encoded given the internal constraints?* ", the suitable variant of efficient coding can be determined (see Chalk et al. [72] for a quantitative elaboration). For instance, one of the variants of efficient coding is based on *redundancy reduction*, which has the objective of encoding maximal information about *all* inputs with statistically independent responses and it is applicable in low noise regime [72]. Afterward, based on principles of efficient coding, a computational objective for a given neural system can be defined. Our choice of efficient coding computational objective is the one introduced in Boerlin et al. [45]. The objective of this coding schema is, a network of Leaky-Integrate and Fire (LIF) neurons should encode the input through a pattern of spikes, such that input stimulus can be reconstructed based on a linear readout of the spiking output. Furthermore, the network should perform the coding with minimum number of spikes and as accurate as possible. The same principle has been employed in Chalk et al. [71] in a more realistic network of LIF neurons and has been used in our investigation.

3.3.2 *Signature of criticality in efficient coding networks*

Following our motivation for the necessity of complementary approaches to criticality, we study networks that implement efficient coding (see Boerlin et al. [45] and Chalk et al. [71] for more details) and we ask if any of the criticality signatures (discussed in Section 3.2) are observable exclusively in the network that is optimized for performing efficient coding.

We investigate the scale-freeness of neuronal avalanches [30], as a potential signature of the networks operating close to criticality. A neuronal avalanche is defined as an uninterrupted cascade of spikes propagating through the network [30]. In a system operating close to criticality, the distribution of avalanche sizes (number of spikes in a cascade) follows a power law. An event is an occurrence of at least 1 spike (among all neurons) within a small window of time.

Interestingly our analysis suggests that, in the vicinity of the parameters that are optimized for efficient coding in the network the distribution of avalanche sizes follow a power-law. When the noise amplitude is considerably lower or higher for efficient coding, the network appears either super-critical or sub-critical, respectively (see for more details). Certainly, this is only a preliminary step, but indeed, it might bring us a few steps closer to bridging criticality and computational frameworks that complement the criticality.

4 APPROACHING THROUGH COGNITION

As motivated in Chapter 1, one of the apertures for approaching the brain as a complex system, that let us remain close to the neuroscience side, is through behavior and cognition. After providing a brief introduction to visual awareness and related phenomenon such as binocular rivalry, we argue that, binocular rivalry is one of the important cognitive phenomenon, that is particularly relevant for a complex system perspective toward the brain. Based on this perspective toward binocular rivalry, we suggest and conduct novel experimental works. We study the phenomena of binocular rivalry on a mesoscopic scale which has not been done before.

4.1 VISUAL AWARENESS

Consciousness is one of the most challenging problems of science [78]. However, during the last few decades, the vast technological and theoretical advancements brought consciousness research to an intense experimental phase. As a result, philosophical speculations on the nature and mechanisms of consciousness are slowly being replaced by empirical and theoretical approaches [218, 351, 186].

There are various experimental paradigms in studying consciousness. We mention two example approaches and highlight our choice. The first one is studying brain activity during various levels of consciousness, i. e. the differences between an awake, conscious state and various degrees of unconsciousness such as deep sleep, anesthesia, or coma. The second one is studying how brain activity changes when a specific visual stimulus is subjectively perceived or suppressed through experimental paradigms like Binocular Rivalry (BR), Binocular Flash Suppression (BFS), masking etc.

The first branch is about studying how brain activity changes in concert with changes in the overall level of consciousness, and indeed it is a fundamental approach. Nevertheless, it is extremely complex and it imposes a set of theoretical and experimental limitations. For example, it is technically difficult to monitor intracortical electrophysiological activity under conditions of coma. However, the second approach, i. e. studying visual awareness (a "visual form of consciousness" [87]), is an alternative approach to the problem with a more tractable framework, especially at the neuronal level. In this approach, brain activity is monitored during changes in the *content of* consciousness. For example, electrophysiological activity is studied when a visual stimulus becomes visible or invisible, while everything else, including the overall level of consciousness as well as the sensory input, remains as constant as possible. Therefore, investigating various kinds of brain activity and their relation with the perception-related events ultimately might bring us steps closer toward an understanding of the neural mechanisms involved in visual awareness.

4.1.1 *Binocular rivalry*

One prominent example of such experimental paradigms that have been exhaustively exploited for understanding the neural mechanisms involved in

visual awareness is binocular rivalry. Binocular rivalry is one of the forms of ambiguous visual stimulation. It involves simultaneous stimulation of corresponding retinal locations across the two eyes with incongruent visual stimuli. It has been shown that different species experience this kind of ambiguous stimulation with some common characteristic [68]. When the subjects are presented with such visual stimuli, they typically experience fluctuations in perception between the two visual stimuli (these fluctuations in perception are known as perceptual switches).

4.1.2 *Neural correlate of binocular rivalry*

In order to understand the neural correlate of phenomenon of binocular rivalry, brain activity can be measured using various experimental methodologies at different scales. It can be spike trains from an individual neuron, field potentials or hemodynamic signals that reflect groups of neurons etc. Each measurement technique has its own limitations [219]. For instance, non-invasive brain-imaging techniques are limited by their spatial and/or temporal resolution, and electrophysiological recordings are limited in their coverage of cell populations. Although all have their own limitations, they have provided us with a significant set of ideas about the neural mechanisms involved in conscious visual perception that we briefly review in the following (for detailed reviews, see for example Blake and Logothetis [44], Tononi and Koch [351], Panagiotaropoulos et al. [271], Koch et al. [188]).

Through single-unit recordings, we grasped a significant set of ideas and insights about the neural mechanisms underlying conscious visual perception on a local scale. Specifically, through these studies, we learned that within each stage of visual hierarchy (from Lateral Geniculate Nucleus, V1 all the way to Profrontal Cortex (PFC)) there are a number of single units whose activity reflects the content of subjective perception of the animal. The proportion of neurons which are modulated by the perception of the animal gradually increases across the visual hierarchy [271]. From no modulated cell in Lateral Geniculate Nucleues (LGN) [201], to superior temporal sulcus (STS) and inferotemporal cortex (IT) [320], and Lateral Prefrontal Cortex (LPFC) [270, 173] where 60-90% of feature selective neurons are perceptually modulated. But how does the activity of these distributed neurons relate to each other and also to other neurons (that are not involved in perception)? How do they interact within their own population? How is the activity of neuronal populations and large-scale networks organized, and how are they related to perception-related events? Single unit studies have potentially overlooked these important aspects of the underlying neural mechanisms, Perhaps, such information is hidden in dynamic patterns of activity that are distributed over larger populations of neurons.

On the other side, imaging studies to some degree characterized the global network by revealing some specific large-scale interactions. For example, frequency-specific oscillatory interactions in the fronto-parieto-occipital [153] and prefrontal-parietal networks [106] and causal interactions in prefrontal-occipital [160] network are involved in conscious perception. However, these findings could not capture the *neuronal* interactions due to their limited spatial and/or temporal resolution. Indeed, such information is potentially available to multi-electrode recordings.

4.2 WHY IS APPEALING FROM A COMPLEX SYSTEM PERSPECTIVE

An integrationist overview on the previous electrophysiology and imaging studies on the neural mechanisms involved in conscious visual perception implies that *a global network of neuronal populations that interact with each other is involved in this phenomenon* [44, 271]. Therefore, visual awareness presumably is a system property, which is associated with a set of cooperative interactions within and between highly interconnected networks of neurons. These neurons are distributed within the entire thalamo-cortical system, mainly temporal, prefrontal, occipital, parietal, lobes and thalamus [44, 271, 366, 225, 338, 153, 106, 270, 21, 351, 188, 160]. The fact that, there is a large number of *interacting* components (neurons and brain regions) involved in the phenomenon of visual awareness, is already one of the important characteristics that allows us to conceive perception as an *emergent* property of a complex system.

Given this new conceptualization for visual awareness, what are our options to tackle it experimentally – at least in terms of measuring the brain activity? Almost all the previous studies of binocular rivalry –in terms of spatial and temporal resolution– are either single-unit recordings or whole-brain imaging (EEG/MEG, fMRI). Such measurements can provide hints or evidence for the existence of such a distributed network (as indeed have been profoundly insightful), but they are not the most suitable measurement techniques to characterize the *neural interactions* [1]. Understanding the *interaction* between units of a complex system is the key for characterizing collective behaviors and therefore it is important to observe the system at scales which give the clearest picture in this regard. At first glance, we can realize that the phenomenon of binocular rivalry is poorly understood at the mesoscopic scale, which could not only reveal the phenomenon of coordinated activity within areas but also across areas in large-scale networks (see Section 2.1). Therefore, a complex system perspective motivates observation at the mesoscopic scale as the first priority and therefore motivates new experiments. Studying at this scale, not only can inform about the involved cooperative mechanisms, but also, it is the first step for bridging the studies based on single-unit recordings and imaging studies.

Conceiving perception as a system property or an emergent property resulting from interactions within a large and distributed network of neurons, is not the only reason for the glamour of binocular rivalry from a complex system perspective. Indeed, various models based on the theory of the dynamical system (which is one of the most powerful frameworks to formalize a complex system) can explain a range of characteristics of bistable perception (such as the distribution of dominance periods) [105, 51, 346, 273]. Perhaps, the most appealing theoretical explanation is provided by Pastukhov et al. [273] that showed a network model operating on the edge of a bifurcation and can explain statistical characteristics of a wide range of multi-stable phenomenon.

Overall, based on available empirical and theoretical evidence we know, we need to deal with a large and distributed network of neurons; Components of this network interact in a non-trivial way; Phenomenon of binocular rivalry seems to be inherently multi-scale; It seems, a neural network operating on an edge of bifurcation can explain various behavior-related statistical prop-

1 With EEG/MEG and fMRI we can also characterize the interaction between the component of the neural system, but due to the nature of these measurement techniques, the picture they can provide about neural interactions is more ambiguous compare to what we can get from invasive recording techniques

erties of the phenomena. Altogether, these finding make this phenomenon appealing from the perspective of complex systems. We believe one of the very first steps for understating the cooperative neural mechanism pertaining to binocular rivalry is *measuring the mesoscopic neural activity*, i. e. new experiments are needed which is the focus of the next sections.

4.3 EXPERIMENTAL CONSIDERATIONS

In the previous section (Section 4.2) we argued that meso-scale observations are necessary for understating the binocular rivalry and consequently, conducting new experiments are needed. For conducting the experimental work pertaining to binocular rivalry, in addition to considerations pertaining to the level of observation, some basic factors need to be considered as well. These factors are briefly discussed in this section.

The first consideration pertains the recording area. One of the target regions for new experiments is PFC for multiple reasons. First, PFC is a central subnetwork (in a graph-theoretic sense) [252] that play a crucial role in cognitive computations [248], especially due to an increase in the integrative aspect of information processing in higher-order cortical areas. Second, ventro-lateral PFC (vlPFC), is reciprocally connected to Inferior Temporal (IT) coretex, which contains the largest proportion of neurons that are perceptually modulated [320] and neurons in PFC have been also shown to be perceptually modulated in similar tasks [270, 150]. Third, PFC is outside of the core visual hierarchy.

For recording from PFC, we also need to be cautious with experimental design, due to the ambiguous role of PFC in perception. In a study by Frassle et al. [128], it was suggested that "frontal areas are associated with active report and introspection rather than with rivalry per se.". In Safavi et al. [299] (also can refer to the corresponding summary,), based on a broad set of evidence, we argue that evidence provided by Frassle et al. [128] is not sufficient for this conclusion, and understating the role of PFC in visual awareness needs further investigation. Due to potential confounding in activity of PFC that can happen due to behavioral report, we needed to employ a no-report paradigm (decoding the perception of the animal using optokinetic nystagmus (OKN) responses [204]).

In this experiment, we particularly needed to have the responses of neurons whose activities are modulated by features of a presented visual stimulus, and the visual stimulus had to induce OKN responses (a certain pattern of eye movement in response to moving stimuli such as moving grating). At the same time, as the core idea was monitoring the activity of neural population, the recording had to be performed with Utah array (10 x 10 array of electrodes that need to be implanted chronically). In contrast to previous similar experiments (e. g. see Panagiotaropoulos et al. [270]) that used non-chronic recording with tetrods where the experimenter could explore to find the neuron by moving the electrodes, Utah arrays are fixed and almost permanent. In Safavi et al. [298] and Kapoor et al. [173] (also can refer to the corresponding summaries, and) we reported that such neurons are accessible with this recording technique (recording with Utah arrays) and under our experimental design. Additionally, we also found that, similarly tuned neurons in this region of PFC are correlated in large distances [298] in contrast to most of sensory cortices [293, 84, 335, 336, 102] (but also see [291]). Interestingly, we also found that spatial structure of functional

connectivity in ventro-lateral PFC is generally [2] different from most sensory cortices. In most sensory cortices, noise correlation decay monotonically as a function of distance, nevertheless, in ventro-lateral PFC we observed in both anesthetized and awake monkeys noise correlation rises again after an initial decay. This observation is also compatible with anatomical differences between PFC and sensory areas [210, 6, 226, 191, 133, 343]. The finding on the spatial structure of noise correlation in vlPFC was not relevant for the binocular rivalry experiment as the spatial structures were not take into account, nevertheless, it was an important finding of the circuitry of PFC.

4.4 TOWARD A MESO-SCALE UNDERSTANDING

The very first question that can be approached based on a mesosopic-level investigation, is what can population dynamics reflect about the content of conscious perception. Second question is what can we learn about the involved neural mechanism from micro-meso relationships in PFC. Notably, both questions are approachable when we have observed the system in a mesoscopic scale (level of neural populations), and are briefly discussed in the next sections (and associated papers).

4.4.1 *Meso-scale dynamics*

The activity of the majority of PFC neurons that are responsive to visual attributes of sensory input are correlated with conscious perception of animals as well. In our case, we used vertically moving grating – upward or downward as stimuli [298, 173] (also can refer to the corresponding summaries, and) and previously it was shown this is the case for face-selective neurons as well [270]. But additionally, the content of conscious perception is decodable from the spiking activity of neural *populations* in ventro-lateral PFC. This is the first confirmation of informativity of the meso-scale observation or measurement of the neural activity. The next steps should focus on characterizing the coordinated dynamics and neural interactions (see the next section and the Part iii for further elaboration on the next steps).

4.4.2 *Micro-Meso relationship*

Given the empirical evidence on the informativeness of population spiking of PFC neurons, more specifically the fact that they reflect the content of conscious perception, it is justified to consider more intricate aspects of mesoscopic dynamics. Such aspect of mesoscopic dynamics includes signatures of neural coordination such as neural oscillation and spike-LFP relationship (also see Chapter 2 important aspect of neural coordination). Furthermore, investigating the relationship between PFC [presumed] state fluctuations conjectured based on LFP oscillatory dynamics, perceptual switches and spiking activity can hint at another aspect of the putative role of neural interactions in binocular rivalry. Indeed, one of the important findings of our study was that, spiking activity of population reflecting the dominant perception, are coupled (relatively stronger than suppressed population) to LFP in range $25-45$ Hz after the perceptual switch [114] (also can refer to the corresponding summary,).

2 By generally, it is meant in presence and absence of visual stimulation, in awake and anesthetized state of the animal.

This strong spike-LFP coupling can be a hint for an emphasized communication (or interaction) of PFC populations reflecting the conscious perception and other brain regions (see Buzsaki and Schomburg [66] for the interpretation of spike-LFP coupling as a quantity to characterize the communication channel). Further investigation is needed to characterize the interaction and functional role of this putative communication. In particular, multiple experimental evidence should be taken into account for interpreting the functional role of the mentioned neuronal interaction. First, we know that neural populations that monitor task-related activity exist in the same region of PFC in the absence of any behavioral report [172], which is important given than various studies argue that PFC is strongly involved in task monitoring [128]. Second, we know that the activity of neural populations in IT cortex is also correlated with perception in the absence of behavioral reports [150]. On the other side, from studies with causal intervention, we know that the activity of PFC is needed for difficult object recognition tasks [176]. Therefore, IT cortex might be a crucial component in this communication circuit and needed to be clarified in future studies.

Part II

MANUSCRIPTS INFORMATION

In this part of the thesis, information of all manuscripts associated to this thesis is provided, which includes the title, list of authors, status of the manuscript and statement of contributions. For statement of contributions, the standard CRediT taxonomy [50] has been used when it was available either in the published manuscript or its publicly available preprint, otherwise the "author contributions" stated in the published manuscript or its publicly available preprint has been used. A summary –with emphasis on the relevant aspects to this thesis– for each manuscript is provided as well. Summaries are written such that, redundancies between manuscripts are minimal. Furthermore, The reader is refereed to other relevant summaries or chapters of the synopsis (Part i). Therefore, summaries remain brief and at the same time, convey the coherent picture of this thesis. Summaries are ordered such that earlier summaries provide backgrounds and foundations for later ones, making it possible to be more concise as we progress through them.

5

PAPER 1

PAPER INFORMATION

TITLE: From univariate to multivariate coupling between continuous signals and point processes: A mathematical framework

AUTHORS: Shervin Safavi, Nikos K. Logothetis, Michel Besserve

STATUS: Published in Neural Computation, see Safavi et al. [302]

PRESENTATION AT SCIENTIFIC MEETINGS: NeurIPS 2019 Workshop: Learning with Temporal Point Processes [300], Bernstein 2021 [307]

AUTHOR CONTRIBUTIONS: Conceptualization, S.S., and M.B.; Methodology, S.S., and M.B.; Software, S.S. and M.B.; Formal Analysis, S.S., and M.B.; Investigation, S.S., and M.B.; Resources, N.K.L.; Data Curation, S.S., and M.B.; Writing – Original Draft, S.S., and M.B.; Writing – Review & Editing: S.S., M.B., and N.K.L.; Visualization, S.S., and M.B.; Supervision and Project administration, M.B.; Funding acquisition, N.K.L.

SUMMARY

Motivation

In various complex systems, we deal with highly multi-variate temporal point processes, that are corresponding to the activity of a large number of individuals. They can be generated by the activity of neurons in brain networks [165], such as neurons' action potentials, or by members in social networks [90, 93], such as tweets in the Twitter network. In practice, a limited number of events per unit are accessible experimentally or observable (for instance numbers of spikes generated by neurons). With such limitations, inferring the underlying dynamical properties of the studied system becomes challenging. Nevertheless, in many cases, exploiting the coupling between the point processes and aggregate measure of the complex system (such as Local Field Potentials as an aggregate measure of population neural activity) can be insightful for understanding the underlying dynamics.

Meaningful and reliable estimates of coupling between such signals can be crucial for understanding many complex systems. However, the statistical properties of many methods classically used remain poorly understood. As a consequence, statistical assessment in practice largely relies on heuristics (e. g. permutation tests). While such approaches often make intuitive sense, they are computationally expensive and may be biased by properties of the data that are unaccounted for. This is particularly relevant for quantities involving point processes and high-dimensional data, which have largely non-intuitive statistical properties, and yet are key tools for experimentalists and data analysts. In this study, we establish a principled framework for statistical analysis of coupling between multi-variate point process and continuous signal.

Material and Methods

First, we derive analytically the asymptotic distribution for a class of coupling statistics that quantify the correlation between a point process and a continuous signal. The key to this theoretical prediction is expressing coupling statistics as stochastic integrals. Indeed, a general family of coupling measures can be expressed as stochastic integrals. The Martingale Central Limit Theorem allows us to derive analytically the asymptotic Gaussian distribution of such coupling measures. We show that these coupling statistics follow a Gaussian distribution. A commonly used example of such coupling statistics is Phase Locking Value (PLV) which typically is used for quantifying spike-LFP coupling in neuroscience.

We then go beyond uni-variate coupling measures and analyze the statistical properties of a family of multi-variate coupling measures taking the form of a matrix with stochastic integral coefficients. We characterize the joint Gaussian asymptotic distribution of matrix coefficients, and exploit Random Matrix Theory (RMT) principles to show that, after appropriate normalization, the spectral distribution of such large matrices under the null hypothesis (absence of coupling between the point process and continuous signals), follows approximately the Marchenko-Pastur law [236] [1] (which is a well-characterized distribution in Random Matrix Theory), while the magnitude of the largest singular value converges to a fixed value whose simple analytic expression depends only on the shape of the matrix.

Results

We derive analytically the asymptotic distribution of Phase-Locking Value (PLV) which is a coupling statistic conventionally used for quantifying the relationship between a pair of a point process (like spikes) and an oscillatory continuous signal (like LFPs). We show that PLVs follow a Gaussian distribution with calculable mean and variance.

Based on the multi-variate extension, we show how this result provides a fast and principled procedure to detect significant singular values of the coupling matrix, reflecting an actual dependency between the underlying signals. This is of paramount importance for the analysis of empirical data given the ever-increasing dimensionality of datasets that need computationally efficient statistical tests.

Conclusion

Our results not only construct a theoretical framework, which is valuable on its own but also can have various applications for neural data analysis and beyond. For instance, based on our theoretical framework we note realistic scenarios where the PLV can be a biased estimator of spike-LFP coupling, and in light of our framework, such biases can be treated.

1 Referred paper [236], is not written in English, but it is the original publication. Reader can refer to Anderson et al. [10, Chapter 2] instead.

6

PAPER 2

PAPER INFORMATION

TITLE: Uncovering the organization of neural circuits with generalized phase locking analysis

AUTHORS: Shervin Safavi, Theofanis I. Panagiotaropoulos, Vishal Kapoor, Juan F. Ramirez-Villegas, Nikos K. Logothetis, Michel Besserve

STATUS: Preprint is available online, see Safavi et al. [306]

PRESENTATION AT SCIENTIFIC MEETINGS: ESI-SyNC 2017 [304], AREADNE 2018 [305], Cosyne 2019 [35], Cosyne 2020 [303], Bernstein 2021 [307]

AUTHOR CONTRIBUTIONS: Conceptualization, S.S., T.I.P., M.B.; Methodology, S.S., J.F.R.-V. and M.B.; Software, S.S. and M.B.; Formal Analysis, S.S. and M.B.; Investigation, S.S., T.I.P., V.K. and M.B.; Resources, N.K.L.; Data Curation, S.S., T.I.P., V.K., and M.B.; Writing – Original Draft, S.S. and M.B.; Writing – Review & Editing: S.S., T.I.P., V.K., J.F.R.-V., N.K.L. and M.B.; Visualization, S.S. and M.B.; Supervision and Project administration, T.I.P. and M.B.; Funding acquisition, N.K.L.

SUMMARY

Motivation

The synchronization between spiking activity and the phase of particular rhythms of LFP has been suggested as an important marker to reason about the underlying cooperative network mechanisms; nevertheless, there is not yet a systematic way to extract concise coupling information from the largely multi-variate data available in current recording techniques. We introduce Generalized Phase Locking Analysis (GPLA) which is a multi-variate extension of phase-locking analysis. Phase-locking analysis is a common uni-variate method of quantifying the spike-LFP relationship. With GPLA, we can quantify, characterize and statistically assess the interactions between population-level spiking activity and mesoscopic network dynamics (such as global oscillations and traveling waves).

Material and Methods

We collect the coupling information between spikes and LFP in a coupling matrix. The coupling matrix, constructed by all the pairwise complex-value spike-field coupling coefficients, represents the population-level spiking activity and all LFP channels. We use Singular Value Decomposition (SVD) to provide a low-rank representation of the coupling matrix. Therefore, we summarize the information of the coupling matrix with the largest singular value and the corresponding singular vectors. Singular vectors represent the dominant LFP and spiking patterns and the singular value, called generalized Phase Locking Value (gPLV), characterizes the strength of the coupling between LFP and spike patterns.

We further investigate the statistical properties of the gPLV and develop an empirical and theoretical statistical testing framework for assessing the significance of the coupling measure gPLV. For the empirical test, we synthesize surrogate data with spike jittering for the generation of the null hypothesis and use it to estimate the p-value for the gPLV calculated from the data. For the theoretical test, we used Martingale theory and [1] Random Matrix Theory (RMT) [10] to approximate the distribution of singular values under the null hypothesis (see Safavi et al. [301] for the details and Chapter 5 for a summary). This allows us to derive a computationally efficient significance test in comparison to the empirical one.

Results

Firstly, if both GPLA and its uni-variate counterpart are applicable, GPLA is superior as it can extract a more reliable coupling structure in the presence of an excessive amount of noise in LFP. Furthermore, to demonstrate the capability of GPLA for mechanistic interpretation of the neural data, we apply GPLA to various simulated and experimental data. Application of GPLA on simulation of hippocampal Sharp-Wave-Ripples (SWR) can reveal various characteristics of hippocampal circuitry with minimal prior knowledge. For instance, with GPLA we can show CA1 and CA3 neurons are all coupled to the field activity in the gamma and ripple band (in line with experimental and simulation results [64, 284]), suggesting this rhythm may support communication between CA1 and CA3 sub-fields during memory trace replay. Furthermore, it also allows us to tease apart the involved populations based on the label-free spike timing and LFP. GPLA can also provide hints on the propagation of activity between the populations (propagation from CA3 to CA1). Application of the method on the experimental recordings from monkey PFC suggests a *global* coupling between spiking activity and LFP traveling waves in this region of PFC. Overall, exploiting the phase distributions across space and frequencies captured by GPLA combined with neural field modeling help to untangle the contribution of inhibitory and excitatory recurrent interactions to the observed spatio-temporal dynamics.

Conclusion

GPLA is a multi-variate method to quantify, characterize and statistically assess the interactions between population-level spiking activity and mesoscopic network dynamics such as global oscillations, traveling waves, and transient neural events. Spike and LFP vectors compactly represent the dominant LFP and spiking patterns and generalized Phase Locking Value (gPLV), characterizes the strength of the coupling between LFP and spike patterns. Our theoretical statistical testing framework allows a computationally efficient assessment of the significance of coupling measure gPLV. This is of paramount importance for neural data analysis given the ever-increasing dimensionality of modern recording techniques that need computationally efficient statistical tests.

7

PAPER 3

PAPER INFORMATION

TITLE: The complex spectral structure of transient LFPs reveals subtle aspects of network coordination across scales and structures

AUTHORS: Michel Besserve, Shervin Safavi, Bernhard Schölkopf, Nikos Logothetis

STATUS: Work-in-progress; a preliminary manuscript is available in the appendix, see Paper 3.

PRESENTATION AT SCIENTIFIC MEETINGS: Machine Learning Summer School [34]

AUTHOR CONTRIBUTIONS: Conceptualization, M.B. and N.K.L; Methodology, M.B. and S.S.; Software, S.S. and M.B.; Formal Analysis, M.B.; Investigation, S.S. and M.B.; Resources, B.S. and N.K.L.; Data Curation, M.B. and N.K.L; Writing - Original Draft, M.B. and S.S.; Writing - Review & Editing: M.B., S.S., B.S. and N.K.L; Visualization, M.B. and S.S.; Supervision and Project administration, M.B.; Funding acquisition, B.S. and N.K.L.

SUMMARY

Motivation

LFPs are intermediary signals, and as such, they reflect a mesoscopic picture of the brain dynamics [214]. As LFPs are rich signals [63, 214, 118], they can be a pivotal point for bringing the brain dynamics at different scales together. In particular, certain transient activities of LFPs reflect cooperative dynamics (we call them *neural events*). A prominent example of such neural events are sharp wave-ripples (SWRs), and it has been observed they co-occur with well-coordinated activity at smaller scales (neurons and populations of neurons) [89, 89, 265], as well as larger scale (entire brain) [222, 177]. In spite of the importance of such characteristic neural activities (neural events), there are not many principled methods for identifying them in a single channel LFP. We introduce a principled method for identifying neural events in a single channel LFP.

Material and Methods

We detect the neural events by isolating transient characteristic neural activities. We first compute the spectrograms of the LFP signals by applying short-term Fourier transform (STFT) on LFPs in order to exploit the spectral content of the LFPs. To identify the frequent transient neural activity with similar spectral content we apply non-negative Matrix Factorization (NMF). Notably, due to scale-invariant nature of LFPs (similar to other extracellular field potential [63]) [129, 147], we used Itakura-Saito divergence in the

optimization procedure of NMF [124] in order to avoid under-weighting of high-frequency components due to their low power in the spectrum. The components result from NMF, provide the information on the spectral content of the neural events. In order to temporally isolate the neural events and characterize their temporal profile, we apply a shift-invariant dictionary learning (a modified version of dictionary learning provided by Mailhé et al. [232]). The latter step, allows us to temporally locate the neural events and also identify the time-domain profiles of events that their spectral content are characterized by the NMF step.

We demonstrate the capability of our method by identifying neural events and their brain-wide signatures in Hippocampus and LGN recorded from anesthetized monkeys. Furthermore, in order to demonstrate that neural events have the potential of relating the meso-scale dynamics even to cellular dynamics, we investigate the neural events in the simulation of thalamocortical circuitry developed by Costa et al. [86] where allow us to access both meso-scale dynamics and also some level of cellular dynamics. The simulation consists of neural mass models with two modules, one for the thalamus and one for the cortex, and mimics the behavior of these circuits during different stages of sleep.

Results

We developed a novel methodology for detecting neural events (transient cooperative neural activities) such as sharp wave-ripples. With our method, neural events can be detected with minimal prior knowledge about the structure under study. Namely, the spectral content is automatically identified by the method, and various other attributes of neural events such as the number of neural event clusters can also be identified by the method in an unsupervised fashion.

Furthermore, we demonstrate the capability of the method by identifying neural events in Hippocampus and LGN and also explore their brain-wide *macro-scale* signatures using concurrent fMRI recordings from anesthetized monkeys. The results suggest that similar to the previous study of Logothetis et al. [222] that was focused on sharp wave-ripples, the identified events in Hippocampus and LGN reflect a large-scale coordinated dynamics. Indeed, this demonstrates the insightfulness of neural events for bridging the meso-scale and macro-scale brain dynamics.

Our results also suggest that neural events can be insightful for establishing a bridge between meso-scale and micro-scale brain dynamics, even at the cellular level. We demonstrate this aspect, by investigating a simulation of the thalamocortical system developed by Costa et al. [86]. With our methodology, we identified different kinds of spindles in the activity of the thalamus module of the simulation, and demonstrate that different events co-occur with characteristic activity patterns in the cellular variables (such as membrane potentials and ionic currents) of the simulation.

Conclusion

With this method, we can find characteristic patterns of LFPs in an unsupervised fashion. This methodology not only allows us to detect well established neural events such as SWRs in a principled fashion, it also identifies characteristic patterns in a single channel LFP that have not been explored, and they can be insightful about cooperative and multi-scale dynamics of the brain.

Such patterns are potentially very special in the sense that, they provide us a time window at which meso-scale dynamics are closely related to micro- and macro-scale dynamics. In fact, as pointed out in Section 2.1 and Section 2.3, this is of paramount importance for bridging the scales of neural dynamics, in particular when combined with GPLA introduced in and NET-fMRI [221].

8

PAPER 4

PAPER INFORMATION

TITLE: Signatures of criticality in efficient coding networks

AUTHORS: Shervin Safavi, Matthew Chalk, Nikos K. Logothetis, Anna Levina

STATUS: Work-in-progress; a preliminary manuscript is available in the appendix, see Paper 4.

PRESENTATION AT SCIENTIFIC MEETINGS: Conference on Complex Systems (CCS 2018) Satellite: Complexity from Cells to Consciousness: Free Energy, Integrated Information, and Epsilon Machines [295], DPG-Frühjahrstagung 2019 [296], Cosyne 2020 [209]

AUTHOR CONTRIBUTIONS: Conceptualization, S.S., and A.L.; Methodology, S.S., M.C., A.L.; Software, S.S. and M.C; Formal Analysis, S.S., M.C and A.L.; Investigation, S.S., M.C and A.L.; Resources, N.K.L. and A.L.; Data Curation, S.S., M.C and A.L.; Writing – Original Draft, S.S.; Writing – Review & Editing, not applicable (this letter has not been communicated with other co-authors so far); Visualization, S.S.; Supervision and Project administration, A.L.; Funding acquisition, N.K.L. and A.L.

SUMMARY

Motivation

Understanding the computations that the brain needs to implement (neural computation) and the dynamics of the brain activity (neural dynamics) are two important goals of computational neuroscience [80, Chapter 1]. Ideally, we need a framework that can accommodate both aspects of the brain in one framework [80, 122]. Nevertheless, to the best of my knowledge, no framework has been developed to satisfy this important need.

An intermediate step toward developing such a framework is exploiting the frameworks and models that are either centered around neural computation or neural dynamics *with implications for the other aspect*. Indeed, there are normative models that have implications for neural dynamics [202, 99, 100, 342, 56, 45, 43, 71, 379, 116] and also models of neural dynamics with implications for neural computation [33, 119, 341, 152, 328, 229, 182, 74, 247, 125] We suggest seeking for "bridges" between such frameworks can be a first step. Neural coding is of particular interest for building such bridges as there have been various studies that suggest potential connections between neural coding and neural dynamics [121, 45, 328, 71, 5, 169, 290, 116]. In particular, multiple recent studies provide qualitative or quantitative evidence on the usefulness of operating close to a phase transition for coding [328, 71, 169, 290]. Interestingly, the phase transition is also one of the pillars of the criticality hypothesis of the brain [256, 347, 253]. In spite of this apparent and exciting connection, networks implementing neural coding

have never been investigated for signatures of criticality. In this study, we investigate networks that can be optimized for neural coding for signatures of criticality.

Material and Methods

In this study, we investigate a network of Leaky-Integrate and Fire (LIF) neurons whose connectivity and dynamics can be optimized for coding a one-dimensional sensory input [71]. This network can be optimized to encode the input efficiently (i. e. with a minimal number of spikes) and accurately (i. e. with minimal reconstruction error). The input is reconstructed by performing a linear readout of spike trains (see [45]). Given an idealized network with instantaneous synapses, the optimal network could be derived analytically from first principles [45]. In this case, neurons that receive a common input avoid communicating redundant information via instantaneous recurrent inhibition. However, adding realistic synaptic delays leads to network synchronization, which impairs coding efficiency. Chalk et al. [71] demonstrated that, in the presence of synaptic delays, a network of LIF neurons can nonetheless be optimized for efficient coding by adding noise to the network. The network's performance depends non-monotonically on the noise amplitude, with the optimal performance achieved for an intermediate noise level. We investigate potential signatures of criticality such as the scale-freeness of neuronal avalanches [30] in the spiking activity of the network.

Results

In this study, we introduce a new approach to better connect neural dynamics and neural computation. Here we search for a potential connection between models of neural dynamics with implications on neural computation, and normative models of neural computation with implications for neural dynamics. We search for signatures of criticality in neuronal networks that can be optimized based on objectives of efficient coding. We investigate efficient coding networks for signatures of criticality. Interestingly, almost exclusively in the optimized network, we observe the signatures of criticality and when the noise amplitude is too low or too high for efficient coding, the network appears either super-critical or sub-critical, respectively. In both cases, the noise level that was optimal for coding also resulted in a scale-free avalanche behavior.

Conclusion

Our results suggest that coding-based optimality might co-occur with closeness to criticality. This result has important implications, as it shows how two influential, and previously disparate fields — efficient coding, and criticality — might be intimately related. This work proposes several promising avenues for future research on the computation and dynamics of the neural system.

9

PAPER 5

PAPER INFORMATION

TITLE: Is the frontal lobe involved in conscious perception?

AUTHORS: Shervin Safavi*, Vishal Kapoor*, Nikos K. Logothetis, Theofanis I. Panagiotaropoulos (* indicate equal contribution)

STATUS: Published in Frontiers in Psychology, see Safavi et al. [299]

AUTHOR CONTRIBUTIONS: Conceptualization, S.S., V.K., N.K.L. and T.I.P.; Methodology, not applicable; Software, not applicable; Formal Analysis, not applicable; Investigation, S.S., V.K. and T.I.P.; Resources, N.K.L.; Data Curation, not applicable; Writing – Original Draft, S.S., V.K. and T.I.P.; Writing – Review & Editing, S.S., V.K., N.K.L. and T.I.P.; Visualization, not applicable; Supervision and Project administration, T.I.P.; Funding acquisition, N.K.L.

SUMMARY

PFC as part of the subsystem that serves the goal-directed character of behavior [216], needs to closely interact with two other subsystems. One is responsible for sensory representation and the other reflects the internal states of the organism, such as arousal or motivation [216]. Moreover, PFC is also a central sub-network [in a graph-theoretic sense] [252] that plays a crucial role in various cognitive functions [248]. Therefore, it is expected to behave differently compared to sensory-related networks in various tasks (e. g. binocular rivalry).

In recent years, novel paradigms have been used to dissociate the activity related to conscious perception from the activity reflecting its prerequisites and consequences [14, 95, 353]. In particular, one of these studies focused on resolving the role of frontal lobe in conscious perception [128]. In this study, Frassle et al. [128] through a novel experimental design, concluded that "frontal areas are associated with active report and introspection rather than with rivalry per se." Therefore, activity in prefrontal regions could be considered as a consequence rather than a neural correlate of conscious perception.

However, based on both fMRI and electrophysiological studies we suspect that PFC is indeed involved in conscious visual perception. Regarding the fMRI studies, Zaretskaya and Narinyan [376], in response to Frassle et al. [128], reviewed the experimental evidence based on fMRI BOLD activity in frontal lobe which suggests even with contrastive analysis (similar to Frassle et al. [128]), some regions of frontal lobe are engaged and therefore play a role in conscious perception. Electrophysiological studies also provided evidence on involvement of some regions of frontal lobe in the absence of behavioral reports (i. e. using no-report paradigms), namely lateral PFC, in visual awareness [270, 173, 114]. In particular, two recent studies [173, 114], (which were carried out as a part of this thesis, see Chapter 4) used a similar paradigm to the one used in Frassle et al. [128]. Moreover, a recent study

by Kapoor et al. [172] based on analysis of a wider range of single units in vlPFC (not just feature selective neurons) suggests that, both task-related and perception-related neurons co-exist in the same region of PFC.

Last but not least, the last decade witnessed a similar disagreement but on the role of primary visual cortex instead of frontal lobe [206, 231, 178, 205]. Ultimately, measuring both electrophysiological activity and the BOLD signal in the same macaques engaged in an identical task of perceptual suppression settled the debate [231, 205]. Therefore, to address such discrepancies we can benefit from multiple measurement techniques simultaneously or in the same animal along with a careful experimental design.

In this opinion paper, we advocate that formulating our conclusions related to prerequisites, consequences and true correlates of conscious experiences, we need to have an *integrative* view on the in hand collection of new evidence. Our investigations and conclusions about the neural correlates of consciousness must not only entail better designed experiments but also diverse experimental techniques (e.g., BOLD fMRI, electrophysiology) that could measure brain activity at different spatial and temporal scales. Moreover, different measurement techniques can reflect complementary information on the brain activity. Therefor, such a multi-modal approach holds great promise in refining our current understanding of conscious processing (and understating the brain in a broader sense).

10

PAPER 6

PAPER INFORMATION

TITLE: Nonmonotonic spatial structure of interneuronal correlations in prefrontal microcircuits

AUTHORS: Shervin Safavi*, Abhilash Dwarakanath*, Vishal Kapoor, Werner Joachim, Nicholas Hatsopoulos, Nikos K. Logothetis, Theofanis I. Panagiotaropoulos (* indicate equal contributions)

STATUS: Published in PNAS, see Safavi et al. [298]

PRESENTATION AT SCIENTIFIC MEETINGS: NeNa 2015 [115], AREADNE 2016 [297]

AUTHOR CONTRIBUTIONS: Conceptualization, T.I.P.; Methodology, S.S., A.D., V.K. and T.I.P.; Software, S.S., A.D., T.I.P. and J.W.; Formal Analysis, S.S., A.D. and T.I.P.; Investigation, V.K., A.D., T.I.P., S.S. and N.G.H.; Resources, N.K.L.; Data Curation, A.D., T.I.P., V.K., and S.S.; Writing – Original Draft, T.I.P., S.S., and A.D.; Writing – Review & Editing: V.K., A.D., T.I.P., N.G.H., and N.K.L.; Visualization, S.S., A.D, V.K. and T.I.P.; Supervision and Project administration, T.I.P.; Funding acquisition, N.K.L.

SUMMARY

Motivation

It has been suggested that mammalian's neocortex follow certain canonical features [109, 107, 110, 144] . One of the features is in the spatial pattern of connectivity. Indeed, there is a large body of evidence suggesting that functional connectivity, inferred based on spike count correlations [84], rapidly decay as a function of lateral distance in most of the sensory areas of the brain [85, 293, 84, 335, 336, 102]. Nevertheless, there are functional and anatomical evidence, that hint at deviations from these canonical features in PFC. PFC is a central sub-network [in a graph-theoretic sense] [252] that play a crucial role in cognitive computations [248], especially due to an increase in the integrative aspect of information processing in higher-order cortical areas. Moreover, anatomical studies have shown that in contrast to early visual cortical areas where we have a limited spread of lateral connections, in later stages of cortical hierarchy like PFC [6, 191, 11, 343, 364] lateral connections are considerably expanded [210, 6, 226, 191, 133, 343] . In this study, we investigate the functional connectivity ventro-lateral PFC (vlPFC) as a function of lateral distance.

Material and Methods

In this study, we investigate the correlated fluctuations of single-neuron discharges in a mesoscopic scale. Electrophysiology data was recorded from

4 macaque monkeys, two in anesthetized state, and two in awake state. Spiking activity was recorded from a Utah array chronically implanted in vlPFC. For the awake experiments, monkeys were trained to fixate for 1000 ms on moving grating in 8 different directions distributed randomly across multiple trials. Tasks were started with the appearance of a red dot as a fixation point (with the size of 0.2°) on the screen for ~300 ms (followed by a moving grating in one of the 8 directions). The moving grating was only presented if the monkey maintains the fixation for the ~300 ms period. Moving grating had the size of 8°, speed of 12-13 degrees per second, and spatial frequency of 0.5 cycles per degree.

In anesthetized experiments, monkeys were exposed with 10 s of stimulation with natural movies. Both awake and anesthetized experiments also included, spontaneous sessions where neural activities recorded in the absence of any behavioral task.

Tuning curves were computed based on conventional procedures [84] by averaging the firing rate across trials for each of the eight presented directions of motion. Signal correlations were defined as the correlation coefficient between the tuning curves of a neuronal pair.

Noise correlations for anesthetized data were computed by dividing the period of visual stimulation into 10 periods, each being 1000 ms long, and considered these periods as different successive stimuli. The same procedure was used for the intertrial periods as well. In the awake data, visual stimulation and intertrial periods were 1000 ms long each; therefore, no additional procedure was required. In the spontaneous data (both anesthetized and awake), the entire length of the recording period was divided into periods of 1000 ms bins and they were treated as a trial.

The spike count correlation coefficients were computed similarly to previous classical studies [22] First, for each condition (either presentation of each moving grating in awake experiment or a single bin of movie clip in the anesthetized experiment), we normalized the spike counts across all trials by converting them into z scores. For each pair, we computed the Pearson's correlation coefficient for normalized spike counts and averaged across conditions to obtain the correlation value.

Results

We found that the spatial structure of functional connectivity (measured based on noise correlations) in vlPFC is different from most of the sensory cortices. In most sensory cortices, noise correlations decay monotonically as a function of distance; nevertheless, in vlPFC we observed in both anesthetized and awake monkeys noise correlation rises again after an initial decay. Moreover, we showed that the characteristic non-monotonic spatial structure in vlPFC, is pronounced with structured visual stimulation.

Conclusion

Our results suggest that spatial inhomogeneities in the functional architecture of the PFC arise from strong local and long-range lateral interactions between neurons. These characteristic patterns of interactions among PFC neurons lead to a non-monotonic spatial structure of correlations in vlPFC. Moreover, the mentioned spatial inhomogeneities are pronounced during structured visual stimulation in the awake state which can be instrumental for distributed information processing in PFC.

11

PAPER 7

PAPER INFORMATION

TITLE: Decoding the contents of consciousness from prefrontal ensembles

AUTHORS: Vishal Kapoor*, Abhilash Dwarakanath*, Shervin Safavi, Joachim Werner, Michel Besserve, Theofanis I. Panagiotaropoulos, Nikos K. Logothetis (* indicate equal contributions)

STATUS: Accepted for publication in Nature Communication (preprint is available online, see Kapoor et al. [173])

PRESENTATION AT SCIENTIFIC MEETINGS: FFRM 2015 [12], SfN 2018 [269], FENS 2018 [174], ASSC 2019 [175]

AUTHOR CONTRIBUTIONS: V.K., A.D. and T.I.P. designed the study. V.K., A.D. and S.S. trained animals. V.K. and A.D. performed experiments and collected data, with occasional help from S.S. V.K. and A.D. analyzed the data. S.S. contributed to spike sorting and selectivity analysis of control experiments. M.B. contributed to the decoding analysis. V.K. prepared and arranged the figures in the final format. S.S. provided the MATLAB generated version of the figures displayed in figure 3D, S12, S13 and S14 A. T.I.P. and N.K.L. supervised the study. N.K.L. and J.W. contributed unpublished reagents/analytical tools. N.K.L. provided the support to the group. V.K. and T.I.P. wrote the original manuscript draft. All authors participated in discussion and interpretation of the results and editing the manuscript.

SUMMARY

Motivation

The role of prefrontal cortex (PFC) has been controversial in recent consciousness studies. Different frameworks of consciousness attribute different, even contradictory roles for PFC in generation of conscious experience. Several frameworks, namely, frontal lobe hypothesis [88], higher order theory [196] and global neuronal workspace framework [18, 98] consider PFC play a mechanistic role in generation of conscious experience. On the opposite side, another important framework of studying consciousness, integrated theory of consciousness [349, 23, 24, 263] (for a review see Tononi et al. [350]), does not consider a similar role for PFC in generation of conscious experience, rather attribute the role of PFC to prerequisites and consequences of consciousness [14, 95].

There are various differences between the aforementioned studies that support each of the two hypothesis. For instance, studies that support attributing the role of PFC to prerequisites and consequences of consciousness, used fMRI as the primary measurement technique, which can potentially lead to discrepancies. In contrast, studies that support the opposite conclusion use electrophysiology (see Chapter 9 for a short discussion). Second, a large portion of studies that support a mechanistic role for PFC in conscious

perception, use externally induced perceptual switches such as Binocular Flash Suppression (BFS) [270]. Third, the majority of the experiments used behavioral reports by the subject in order to know the content of conscious experience (for a review see [353, 188]). This study was an effort, to bring this controversy one step closer to the resolution by recording the neural activity from monkey ventro-lateral PFC (vlPFC) during a no-report Binocular Rivalry (BR) paradigm.

Focus of investigations on phenomenon of BR, in terms of spatio-temporal scales of measurements, was mainly micro-scale (level of individual neurons) and macro-scale (level of large-scale networks) Almost all the previous studies either focus on the activity of feature selective neurons measured based on single unit recordings [201, 320, 178, 21, 270], or the whole-brain dynamics measured with imaging techniques (EEG/MEG, fMRI) [366, 225, 338, 153, 106, 351, 160] (for reviews see [44, 271, 188]). A complex system perspective to binocular rivalry phenomenon, motivates observation of the system in a mesoscopic scale as a very first step to understand the role of neural interactions (see Section 4.2 for further elaboration). In this study, we address this need, by measuring spiking activity of neural populations in vlPFC with multi-electrode recording techniques.

Material and Methods

In this study, we investigate the neural correlate of visual awareness in mesoscopic scale. Recording procedure is similar to awake experiments of explained earlier (see "Material and Methods"of Chapter 10). The core behavioral paradigm used in this study was a passive ambiguous stimulation, and consist of two tasks, Binocular Rivalry (BR) and Physical Alternation (PA). Both tasks consist of fixation period similar to fixation task explained earlier in Chapter 10, and followed by presentation of 1 or 2 seconds upward or downward moving gratings (presented only to one eye – half of the trials for each eye). After the phase of stimulus presentation, in PA trials, the first stimulus was removed and a moving grating in the contralateral eye was presented in the opposite direction. BR trials had the identical structure of the stimulus presentation, but with the difference that, the second stimulus was presented without removing the first stimulus. In BR trials that two opposite moving grating were presented simultaneously, the perception of the monkey spontaneously switches between the stimulus (i. e. upward and downward grating) across the the entire length of trial (8-10 seconds). Whereas, in PA trials, there are no perceptual switches, but perception of the animal changes by the alternation of the presented stimuli (upward and downward grating). Parameters of the visual stimulus (moving gratings) are identical to the experiment explained in Chapter 10. Furthermore, Optokinetic Nystagmus (OKN) reflexes [1] has been used to determine the perception of the animal.

In addition to the main experiment that consist of BR and PA tasks, we additionally have a control experiment for controlling eye movement as a confounding factor. Given that determining the animal perception is based on eye movements (OKN reflexes), to rule out the eye movement as a confounding factor, we perform a passive fixation experiment similar to the awake experiment of explained earlier (see "Material and Methods"of Chapter 10), but without eye movement. In this experiment, the eye movement during presentation of moving grating were suppressed by instructing the animal

1 OKN reflexes are characteristic patterns of eye movements in response to moving stimuli, that consist of smooth pursuit and fast saccadic eye movements.

to maintain the fixation during the task (by overlaying a fixation point with size of 1-2° on top of the moving grating).

Results

Firstly, the perpetual dominance periods detected based on OKN reflexes follow a gamma distribution which is compatible with previous studies [208]. This indicates that using no-report paradigms of BR lead to compatible results with human studies. Given the availability of neurons [recorded by Utah array] that respond to direction of motion of moving grating stimuli in PFC (see), we can quantify the proportion of perceptual modulation of neurons in our experiment that use upward and downward moving gratings as rivaling patterns. Interestingly, compatible with previous studies that used different tasks and visual stimuli [270], majority of sensory modulated units were also perceptually modulated. Moreover, in the population level, the content of conscious perception of the animals was decodable from spiking activity of neural populations in vlPFC. Lastly, the decoding algorithm that we used for decoding the content of the perception [246], could also reliably decode the content of the presented visual stimulus (in the passive fixation experiment) both in presence and absence of eye movement i. e. training the decoder with responses in presence of eye movement, and test when the eye movement are suppressed (fixation-on task) and vice versa. Therefore, our control analysis suggest that eye movements are not a confounding factor for our perceptual modulation.

Conclusion

In this study, we showed that activity of the majority of sensory modulated neurons of vlPFC is correlated with conscious perception in a no-report binocular rivalry task, and the content of conscious experience is decodable from mesoscopic dynamics of PFC. Moreover, this study has an important implication for the neural correlate of visual awareness. This study adds another piece of evidence for the involvement of PFC in conscious perception which has been an important debate in the field of consciousness research in the last few years (also see).

12

PAPER 8

PAPER INFORMATION

TITLE: Prefrontal state fluctuations gate access to consciousness

AUTHORS: Abhilash Dwarakanath*, Vishal Kapoor*, Joachim Werner, Shervin Safavi Leonid A. Fedorov, Nikos K. Logothetis, Theofanis I. Panagiotaropoulos (* indicate equal contributions)

STATUS: Preprint is available online, see Dwarakanath et al. [114]

PRESENTATION AT SCIENTIFIC MEETINGS: FFRM 2015 [12], SfN 2018 [269], AREADNE 2018 [113]

AUTHOR CONTRIBUTIONS: Conceptualisation: A.D., V.K., T.I.P. (lead), N.K.L.; Data curation: A.D. (lead), V.K. and J.W.; Formal analysis: A.D. (lead), V.K., J.W., L.A.F.; Funding acquisition: N.K.L.; Investigation: A.D. (equal), V.K. (equal), T.I.P. (supporting); Methodology: A.D. (equal), V.K. (equal), J.W. & S.S. (supporting), T.I.P. (equal); Project administration: T.I.P.; Resources: J.W., N.K.L. (lead); Software: A.D. (lead), V.K., J.W., L.A.F. & S.S. (supporting); Supervision: T.I.P.; Visualisation: A.D. (lead), T.I.P. (supporting); Writing – original draft: A.D., T.I.P. (lead); Writing – review & editing: A.D., V.K., L.A.F., S.S., T.I.P. (lead), N.K.L.

SUMMARY

Motivation

In Section 4.2 we elaborated on the motivations for studying the phenomenon of binocular rivalry (BR) in a mesoscopic scale and in we showed that content of conscious experience is decodable from mesoscopic dynamics of PFC. This was the first confirmation on the usefulness of the meso-scale observation. This allows us to go one step further in studying the mesoscopic dynamics of PFC. One of the most important markers of coordination in mesoscopic dynamics of the brain, is neural oscillations [62, 65]. In this study we investigate oscillatory dynamics in ventro-lateral PFC (vlPFC) and its connection to conscious visual perception.

Material and Methods

Most of the experimental details for this study was explained in summaries of the other papers (, and). Recording procedure is similar to awake experiment of explained earlier (see "Material and Methods"of Chapter 10). The behavioral paradigm used in this study is also explained earlier (see "Material and Methods"of Chapter 11). In this study, Continuous Wavelet Transform (CWT) [234] has been used to extract spectral content of LFPs and Chronux toolbox [46] for quantifying spike-LFP coupling by computing Spike-Field-Coherence (SFC).

Results

This study reveals various characteristic oscillatory activities which are happening in the vicinity of the perceptual switches detected based on Optokinetic Nystagmus (OKN) reflexes. The frequency of these transient oscillatory activities are covering low and intermediate ranges (namely 1-9 Hz and 20-40 Hz). In addition to presence of these coordinated dynamics in the mesoscopic activity of PFC neural populations and their relationship to perceptual events, the statistics and spatio-temporal patterns of some of these transitory events lend support to important frameworks of studying the consciousness.

Conclusion

This study adds another piece of evidence for the involvement of PFC in conscious perception, in addition to the one discussed earlier in . In particular, it reveals signatures of neural coordination reflected in the oscillatory dynamics (see Section 2.3.2) of neural populations involved in conscious visual perception. Revealing these signatures could not be possible without investigating the system in meso-scale (see more elaborating in Section 4.2). Lastly similar to , this study has an important implication for the neural correlate of visual awareness. This study highlights the involvement of PFC in conscious perception which has been an important debate in the field of consciousness research in the last few years (also see).

Part III

OUTLOOK

This part is dedicated to a subjective perspective on how the research line of this thesis can or should be extended. In this thesis, we sought for *principled* ways of approaching the brain. Although this thesis touched on various such aspects, but I believe it misses an important aspect of the brain which is its *adaptivity*. In the end, brain, presumably the most "complex system", needs to survive in the environment. Indeed, in the field of *complex adaptive systems*, the endeavor is understanding very similar questions in the nature. Inspired by some ideas discussed in the field of complex adaptive systems, we suggest a set of new research directions that intend to incorporate the adaptivity aspect of the brain as one of the principles. Of course, these research directions, remain close to the neuroscience side, similar to the intention of the research presented in previous parts.

13 BRAIN AS A COMPLEX & ADAPTIVE SYSTEM

In Chapter 1, we argue that brain can be approached as a complex system. Certainly, this is a valuable perspective toward the brain and was the pivotal idea of this thesis. Nevertheless, an important aspect of the brain, as a biological information processing system, is not taken into account in the approach we followed and discussed in this thesis. This important aspect is *adaptivity* of humans/animals. They need to be *adaptive* in order to survive. That being said, perhaps we should consider humans/animals as *adaptive agents* and the brains as a complex *and* adaptive system. Indeed, Complex Adaptive Systems (CAS) have been an independent field of research (see Holland (2006) for a brief review).

*Approaching the brain as a complex **and** adaptive system*

Inspired by general properties and mechanisms introduced for CAS (that are briefly discussed in Section 13.1), again, new questions can be asked in various domains of neuroscience, and moreover, even old questions can be revisited based on this perspective. In this chapter, we introduce a set of new research directions that we believe are complementary to the ideas that motivated and shaped this thesis.

Conceiving the brain as a CAS implies that certain computations are needed to satisfy the adaptivity of the agent (see Section 13.2 for further elaboration). Moreover, as we discussed earlier (see Chapter 1), conceiving the brain as a complex system has implications on the dynamics of the brain. More generally, on one hand, behavior is a rich source for seeking and understanding the computational objectives (pertaining to adaptivity of humans and animals) On the other hand, multi-scale dynamics of the brain, as briefly discussed in Chapter 2, is a rich source for understanding the biophysical machinery of this adaptive agent implementing the computation. For instance, concerning the adaptivity of the humans and animals, focusing on behavior have led us to various developments in ecological psychology [288], reinforcement learning [259], and even understanding the emotion [19] that all inform us about the brain computations [260]. Concerning the multi-scale dynamics, studying the brain across scales, has helped us to understand the emergent properties of this biophysical machinery (for further elaboration, see Pesenson [275, Chapter 1] and Siettos and Starke [330]).

Through behavior we can understand computation needed to be adaptive and through multi-scale dynamics of the brain we can understand the brain's biophysical machinery

From a broader perspective, particularly in terms of Marr's levels of understating [238], it can be argued that, understanding the brain dynamics, brings us closer to the implementation level and perhaps to some degree to the algorithmic level; and understating the behavior brings us closer to understanding the computation and more explicitly the algorithm. With no doubt, both of these aspects are utterly important for understating the brain. Therefore, it is import to establish a connection between these two, in order gain an *integrative* understating of the brain (see Churchland and Sejnowski [80, Chapter 2, Section 2] for a broad perspective on the importance of this bridge and Stephan et al. [339] and Forstmann and Wagenmakers [126, Chapter 8] for showcases of their importance in translational neuroscience). Motivated by the importance of establishing this bridge, in Section 13.3 we outline various approaches we can take for relating behavior to multi-scales brain dynamics.

*An **integrative** understating of the brain need a bridge*

13.1 COMPLEX ADAPTIVE SYSTEMS

Complex adaptive systems (CAS) can be broadly defined as a system composed of multiple elements, called agents, *"that learn or adapt in response to other agents"* [158, Chapter 3]. CAS have been studied for decades (see Morowitz and Singer [255] for historical note), and there have been efforts to explain the behavior of various natural and artificial systems based on the CAS formalism; They include adaptive behavior of the immune system [77], finical market [158] and even language [120].

Different sets of properties and mechanisms which are considered to be common between different CAS have been suggested [55]. We outline the 4 features proposed by Holland [156]. Although, some of the core ideas are common among most of the other proposals and indeed those commonalities are the foundations for ideas presented in the following, but readers are also encouraged to refer to properties and mechanism proposed by others as well (for example see Gell-Mann [135] and Arthur et al. [13, Chapter 1]).

Holland [156] introduces 4 major features or characteristics that CAS have in common in spite of their substantial differences:

1. Parallelism: Complex systems (also briefly discussed in Chapter 1) are constructed with many *intently interacting* components. Due to the need for tight coordination, simultaneous communications between components of the system are inevitable.

2. Conditional actions: In CAS, agents need to act conditionally as the required action is defined by the agent's internal state (condition) and actions of external agents.

3. Modules and hierarchies: CAS are often organized in a modular and hierarchical fashion (for the latter see [158, Chapter 7] and [157, Chapter 8]).

4. Adaptation and evolution: Agents in CAS need to change over time in order to gain a better performance. Adaptation requires solutions to two important problems, namely *credit assignment* and *rule discovery*.

Features or characteristics mentioned in the number two and four of Holland's idea are particularly pertaining to *computations* that CAS need to perform. Interestingly, some of these computations are already a focus of research in the field of neuroscience as well (but not necessarily based on a similar foundation we motivate by CAS ideas). In section Section 13.2 we briefly discuss some of these computational objectives that can be closely connected to the brain.

13.2 BRAIN COMPUTATIONAL OBJECTIVES

As briefly discussed earlier, humans/animals as information processing systems, are adaptive agents, and need to interact with a complex environment. We can conceive the brain as a CAS, and based on CAS notions introduced earlier, we can argue that due to their adaptivity they need to perform certain computations. Indeed, Mitchell [250, Chapter 12] argue that,

> *"At a very general level, one might say that computation is what a complex system does with information in order to succeed or adapt in its environment."*

To emphasize conceiving the brain as a CAS and the computations it implies, we highlight some of the computational objectives of the brain that are under active investigation *and* and are closely related to general properties of CAS discussed in Section 13.1. The need for *conditional actions*, solving the *credit assignment* problem and *discovering rules* in the environment that were mentioned in Section 13.1 as general properties of CAS, are closely related to *representation*, *decision making* and *reinforcement learning* which are actively investigated in neuroscience.

One of these computational objectives is efficient representations. The ability of an agent to act upon actions and states of external agents relies on *efficient representation* of information pertaining to external agents. The other computational objective is credit assignment and rule discovery that are both premises of reinforcement learning [370].

Certainly, this section, by no means, provides a comprehensive list of computational objectives of the brain that have been already studied in neuroscience. Rather, it highlights examples that are closely related to the ones CAS should have in a general sense. In the next step, we need to find the connections between these computational objectives and their biophysical machinery by investigating the relationship between behavior and multi-scale dynamics of the brain.

13.3 RELATING BEHAVIOR TO MULTI-SCALE BRAIN DYNAMICS

As argued earlier, behavior is a rich source for understating such computational objectives in human/animals and multi-scale dynamics is a rich source for understating the biophysical machinery behind it. This is the motivation for relating the behavior to multi-scale brain dynamics. In this section, we introduce potential approaches that we think can relate these two facets of the brain.

Certainly, establishing this connection is challenging. Therefore, we need to decompose it into smaller but complementary steps that can be supported by the existing models and/or empirical evidence. In the next sections (Section 13.3.1, Section 13.3.2, and Section 13.3.3), we propose various approaches that are more or less accessible and can potentially bring us a few steps closer to establishing a bridge between behavior and multi-scale brain dynamics.

13.3.1 *Relating neural dynamics and neural computation*

As discussed earlier, neural computation and dynamics are both important aspects of the brain. There are various frameworks and models in neuroscience which are either centered around neural computation [202, 99, 100, 342, 56, 45, 43, 71, 379, 116] or neural dynamics [33, 119, 341, 152, 328, 229, 182, 74, 247, 125] but also have some implications for the other one (also see Maass [229] for a brief review). These models are not necessarily well connected to *behavior* and *multi-scale* dynamics of the brain, but still can fill some space in this large gap between behavior and multi-scale. Further investigation in such frameworks and models, that are outlined in the next sections, can potentially help us to accomplish the mentioned goal, which is relating behavior to multi-scale brain dynamics.

13.3.1.1 *Normative models with implications for neural dynamics*

There have been various efforts to relate neural computation to neural dynamics by introducing normative models of neural computation (e. g. based on sampling theories, Bayesian inference algorithms) which can explain some aspects of observed dynamics of the brain such as irregular spiking and neural oscillations [121, 56, 45, 43, 71, 239, 116, 112]. More generally there have been efforts to relate the state of the machinery implementing a given neural computation to a putative dynamical regime of the neural circuits. For instance, Echeveste et al. [116] and Lengyel et al. [202] have developed neuronal networks which implement Bayesian inference that are attractor networks as well. Neural coding, in particular, is one of the well established computations that brain needs to accomplish [281] and there have been various efforts to connect neural coding and neural dynamics [121, 45, 71, 116]. In most of such normative models, we optimize or train a network of neurons based on a specific computational objective (such as reconstruction error), and the features of the neural dynamics appear in the resulting network activity automatically.

All the features of neural dynamics that have been explained by the previous normative models are among the important ones and some of them are even considered computationally relevant (like oscillations [71, 276]). Nevertheless, the brain dynamics has been shown to be more complex than the reach of normative models so far [97, 52]. Not only in terms of complexity of the observed dynamics, but also in terms of scale, particularly large scale dynamics and multi-scale dynamics [129, 3]. Next steps should include developing normative models with richer neural dynamics, in particular, the large scale and multi-scale dynamics.

13.3.1.2 *Models of neural dynamics with implications for neural computation*

One of the frameworks for explaining the neural dynamics with connection to neural computation is the "criticality hypothesis of the brain" (for a review see [256] – also briefly discussed in Section 3.1). Certainly, frameworks like criticality are insightful for brain dynamics [256] in particular because they provide explanations for observed multi-scale dynamics of the brain [3].

One approach to better connect the criticality hypothesis of the brain to neural computation could be the one we used in Chapter 3, which is searching for signatures of criticality in neuronal networks that can be optimized based functionally relevant computational objectives (in Chapter 3, we used efficient coding objectives). Of course, this is not necessarily informative on a mechanistic level, rather is an indication of *potential* connections. Presence of signatures of criticality may or may not hint for more mechanistic approaches. Nevertheless, some clues can guide us toward more formal investigations. For instance, for the particular case discussed in Chapter 3, Fisher information can be a candidate quantity that both frameworks – efficient coding [368] and criticality [280, 91, 170, 192] – use to assess the closeness to their optimal point.

Another potential approach is seeking for other kinds of functionally relevant attributes for notions established in criticality hypothesis of the brain. For instance, it has been suggested that neural avalanches are related to cell assemblies [279] and indeed the notion of cell assemblies are closely connected to computations implemented in the brain [333, 143, 142, 61, 345].

13.3.2 *Exploiting models of pivotal tasks*

For the purpose expressed in Section 13.3, we can also exploit behavioral tasks which have been comprehended from a wide range of perspectives. To the best of my knowledge, not so many such tasks are identified and exhaustively explored. Nevertheless, we believe this small number is sufficient to make further exploration in this direction justified, given the potential insight that we can get from them. For instance, Cavanagh et al. [70] studied perceptual decision-making through interventional experimentation, and multi-scale computational modeling. Indeed, such theory-experiment hybrid approaches can be insightful, both for understanding the multi-scale dynamics of the phenomenon (in this case from synapse to behavior) and also the computations involved in the task (in this case evidence accumulation process). Frank [127] and colleagues also studied the decision making and cognitive control through reinforcement learning models and biophysical modeling of a single cortico-basal ganglia circuit and similarly, they could gain an integrative understating of the involved computation and also biophysical and dynamical characteristics that have been observed during such tasks. A key in both examples was exploiting the tasks that have been comprehended from a wide range of perspectives (normative modeling, biophysical modeling, measuring electrophysiological activity of involved circuits).

One example of such tasks that has been studied from a wide range of perspectives and wide range of tools is the *bistable perception*. On one hand, a large body of computational studies focus on explaining the dynamics of bistable perception [254, 327, 273, 361, 83]; On the other hand, another class of computational models which tried to explain the phenomenon with normative approaches centered around the computation that the brain might need to perform pertaining to perception [39, 92, 155, 17, 136]. Notably, most of these studies are centered around Bayesian model of the brain [185, 111].

Next to this extensive computational models (which include both normative and biophysical models) there is a large body of psychophysical (for review see [184]), electrophysiological and imaging (for review see [44, 271]), pharmacological [67, 245], and genetic studies [249, 258, 197, 73]. Particularly, as briefly discussed in Chapter 4, from electrophysiological and imaging we learn that a distributed network of neurons is involved in the phenomenon and therefore this is inherently a multi-scale problem.

We believe a wide range of perspectives toward the phenomenon of bistable perception, that led to this immense range of studies and their resulting insight, justify bistable perception as one of the ideal tasks to be studied with the purpose of relating behavior (and their accompanied computation) to multi-scales brain dynamics. In this thesis, we approach the phenomenon of binocular rivalry differently from the conventional approaches (see Chapter 4), and our initial results (see and) justified the usefulness of our proposed mesoscopic scale observation of the brain during a binocular rivalry task. Indeed, a meso-scale observation can also be the first step for understanding the multi-scale dynamics of binocular rivalry. In Chapter 2 we introduced a set of novel methodologies for cross-scale and multi-scale analysis of neural data, in particular mesoscopic signals like LFPs. Transient and cooperative neural activities in hippocampus (such as sharp wave-ripples) have been studied extensively. As exemplified in Section 2.3, such characteristic events can co-occur with well-coordinated activity in smaller scales (scale of neurons and population of neurons), and a larger scale (whole brain) as well. Therefore, investigating the presence of such events in the mesoscopic

activity of neurons during binocular rivalry [assuming their existence] and the relationship between these neural events and behavior can potentially bridge the multi-scale dynamics of the brain and behavior (which is binocular rivalry in this case).

Indeed, recent electrophysiological studies in the cortex also revealed neural activities with cooperative and transient nature that are involved in cognitive functions other than memory consolidation. For instance, Womelsdorf et al. [371] reported burst firing events in Prefrontal Cortex accompanied with particular large-scale synchronization patterns and attention switches.

What has been discussed can be a potential road map to bridge the multi-scale dynamics of the brain and behavior in binocular rivalry, but still the connection to computation remains elusive. Regarding the computations that brain presumably needs to perform, as mentioned earlier, there are already computational models [39, 92, 155, 17, 136]. Some of these models can even explain many aspects of binocular rivalry psychophysics and some aspects of neural dynamics [207, Chapter 3]. Certainly, bridging the multi-scale dynamics and computations explicitly, should be investigated in the next steps.

13.3.3 *A principled framework for data fusion*

One of the core components of the proposed goal, *relating behavior to multi-scale brain dynamics*, is relating dynamics of the brain across scales even independent of behavior and computation. Indeed, in Chapter 2, we introduced novel methodologies for the very same purpose – bridging the scales. Nevertheless, most of such methodologies (including the ones introduced in this thesis) are designed for particular choices of data modalities (e. g. spike-LFP coupling, LFP-BOLD relationship). This implies, for each pair of modalities, we tend to develop a set of tools accustomed to the nature of that particular type of data (which is a reasonable choice for the first try). Of course, such modality-specific methodologies have been insightful and certainly will be, but having a general framework which is capable of embedding or allowing the investigation of different datasets in a common space can potentially bring a wider range of opportunities for investigating brain dynamics across scales and ultimately relate them to the behavior and computation.

Indeed, a few frameworks exploiting kernel-based methods [41, 257, 42, 123] and topological data analysis [384] have been proposed, that are potentially capable of fusing multi-modal data in a principled fashion. Next steps should include broad investigation of such frameworks for various modalities including the ones accessible via invasive recording techniques such as spikes and extracellular field potentials (as they are less explored compared to non-invasive ones). In particular, data modalities that can be better represented by point processes (such as spike trains) are more challenging to be fused with the other kinds of neural data which are continuous in nature (should be noted that there have been some efforts in this direction based on kernel-based methods [326, 267, 268, 212], and for a review see Park et al. [272]).

13.4 UNDERSTATING THE NEURO-PRINCIPLES THROUGH DYSFUNCTIONS

Understanding the brain dysfunctions, in addition to its humanistic aspects and potential societal impacts can also be insightful for gaining a mecha-

nistic understanding of the brain. In particular, understanding cognition and behavior is one of the most important goals of the brain science, and among brain dysfunctions, psychiatric disorders are specifically connected to the malfunctioned cognition and disorders of behavior [159]. A window for understanding the machinery behind cognitive capabilities and neural correlates of behavior can happen through the understanding of when and why they malfunction i. e. *mechanistically* to understand the syndromes we observe in psychiatric disorders.

Furthermore, Psychiatry is unique from various other perspectives. Approaches used for understanding the psychiatric disorders are extremely diverse. In terms of scales or levels of organization [80, Chapter 1], psychiatric disorders have been studied from their genetic basis [59, 161, 334] all the way to their roots in the social interactions [309, 203, 317] In terms of [Marr] levels of understanding [238], psychiatric disorders have been attacked in all three levels [287, Chapter 5][159].

The mentioned diversity of approaches goes beyond the conventional research in the systems neuroscience. As the last example, it is worth mentioning the research on psychiatric disorders for establishing the connection between the nervous system and the immune system. Recently, a peculiar connection between psychiatric disorders (in particular depression and schizophrenia) and dysfunctions of the immune system has been established [180, 57, 344, 373, 242, 310, 140, 181] and more generally the interaction between the immune system and the brain has been receiving more attention and support recently ([57, 94, 20, 277, 325, 228, 148, 190, 189]).

Despite this diversity, there are also potential connections and bridges between them. For instance, in many brain dysfunctions we have clues about both impaired computation and brain dynamics. Whether there is a connection between them, it needs to be thoroughly investigated. However, at least the current state of [Computational] Psychiatry is not clueless about integration of neural computation and neural dynamics. For instance, [101], based on their implementation of circular inference, have suggested that pathological inference attributed in schizophrenia can be mapped into excitation-inhibition imbalance in the neural circuit implementing the inference.

Overall, we believe, understating the brain dysfunction is an intriguing window for gaining an integrative understating of the brain function given the richness and diversity of the empirical data in the field.

ACRONYMS

BOLD Blood-Oxygen-Level Dependent

BFS Binocular Flash Suppression

CAS Complex Adaptive System

CCA Canonical Correspondence Analysis

fMRI functional Magnetic Resonance Imaging

LFP Local Field Potential

LGN Lateral Geniculate Nucleues

LIF Leaky-Integrate and Fire

LPFC Lateral Prefrontal Cortex

vlPFC ventro lateral Prefrontal Cortex

PFC Prefrontal Cortex

PLV Phase Locking Value

MUA Multi Unit Activity

NET-fMRI Neural-Event-Triggered functional Magnetic Resonance Imaging

NMF Non-negative Matrix Factorization

OKN Optokinetic Nystagmus

REM Rapid-Eye-Movement

RG Renormalization Group

STFT Short-Term Fourier Transform

SUA Single Unit Activity

SFC Spike Field Coherence

SNR Signal to Noise Ratio

SVD Singular Value Decomposition

BIBLIOGRAPHY

[1] Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and Event History Analysis: A Process Point of View*. Statistics for Biology and Health. Springer, New York, NY. (Cited on page 32.)

[2] Acebrón, J. A., Bonilla, L. L., Pérez Vicente, C. J., Ritort, F., and Spigler, R. (2005). The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.*, 77(1):137–185. (Cited on page 11.)

[3] Agrawal, V., Chakraborty, S., Knöpfel, T., and Shew, W. L. (2019). Scale-Change Symmetry in the Rules Governing Neural Systems. *iScience*, 12:121–131. (Cited on pages 18 and 54.)

[4] Aitchison, L., Corradi, N., and Latham, P. E. (2016). Zipf's Law Arises Naturally When There Are Underlying, Unobserved Variables. *PLOS Computational Biology*, 12(12):e1005110. (Cited on page 18.)

[5] Alamia, A. and VanRullen, R. (2019). Alpha oscillations and traveling waves: Signatures of predictive coding? *PLOS Biology*, 17(10):e3000487. (Cited on page 37.)

[6] Amir, Y., Harel, M., and Malach, R. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J. Comp. Neurol.*, 334(1):19–46. (Cited on pages 25 and 41.)

[7] Amit, D. J. and Amit, D. J. (1992). *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press. (Cited on page 19.)

[8] Anastassiou, C. A. and Koch, C. (2014). Ephaptic coupling to endogenous electric field activity: Why bother? *Curr. Opin. Neurobiol.*, 31C:95–103. (Cited on page 11.)

[9] Anastassiou, C. A., Perin, R., Markram, H., and Koch, C. (2011). Ephaptic coupling of cortical neurons. *Nat. Neurosci.*, 14(2):217–223. (Cited on page 11.)

[10] Anderson, G. W., Guionnet, A., and Zeitouni, O. (2010). *An Introduction to Random Matrices*. Cambridge University Press, Cambridge; New York. (Cited on pages 15, 30, and 32.)

[11] Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupé, J.-M., Bullier, J., and Lund, J. S. (2002). Circuits for Local and Global Signal Integration in Primary Visual Cortex. *J. Neurosci.*, 22(19):8633–8646. (Cited on page 41.)

[12] Antoniou, R., Safavi, S., Kapoor, V., Logothetis, N. K., and Panagiotaropoulos, T. (2015). Perceptual modulation of pupillary reflex in macaque monkeys. In *Federation of European Neuroscience Society Featured Regional Meeting (FFRM 2015)*. (Cited on pages 43 and 47.)

[13] Arthur, W. B., Durlauf, S. N., Lane, D. A., and Program, S. E. (1997). *The Economy as an Evolving Complex System II*. Addison-Wesley. (Cited on page 52.)

[14] Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience & Biobehavioral Reviews*, 36(2):737–746. (Cited on pages 39 and 43.)

[15] Ashida, G., Wagner, H., and Carr, C. E. (2010). Processing of Phase-Locked Spikes and Periodic Signals. In *Analysis of Parallel Spike Trains*, Springer Series in Computational Neuroscience, pages 59–74. Springer, Boston, MA. (Cited on page 13.)

[16] Atwal, G. S. (2014a). Statistical mechanics of multistable perception. *bioRxiv*. (Cited on page 8.)

[17] Atwal, G. S. (2014b). Statistical mechanics of multistable perception. *BioRxiv*. (Cited on pages 55 and 56.)

[18] Baars, B. J. (2005). Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Prog Brain Res*, 150:45–53. (Cited on page 43.)

[19] Bach, D. R. and Dayan, P. (2017). Algorithms for survival: A comparative perspective on emotions. *Nat Rev Neurosci*, 18(5):311–319. (Cited on page 51.)

[20] Badimon, A., Strasburger, H. J., Ayata, P., Chen, X., Nair, A., Ikegami, A., Hwang, P., Chan, A. T., Graves, S. M., Uweru, J. O., Ledderose, C., Kutlu, M. G., Wheeler, M. A., Kahan, A., Ishikawa, M., Wang, Y.-C., Loh, Y.-H. E., Jiang, J. X., Surmeier, D. J., Robson, S. C., Junger, W. G., Sebra, R., Calipari, E. S., Kenny, P. J., Eyo, U. B., Colonna, M., Quintana, F. J., Wake, H., Gradinaru, V., and Schaefer, A. (2020). Negative feedback control of neuronal activity by microglia. *Nature*, pages 1–7. (Cited on page 57.)

[21] Bahmani, H., Logothetis, N., and Keliris, G. (2011). Neural correlates of binocular rivalry in parietal cortex. (Cited on pages 23 and 44.)

[22] Bair, W., Zohary, E., and Newsome, W. T. (2001). Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(5):1676–97. (Cited on page 42.)

[23] Balduzzi, D. and Tononi, G. (2008). Integrated information in discrete dynamical systems: Motivation and theoretical framework. *PLoS computational biology*, 4(6):e1000091. (Cited on pages 6 and 43.)

[24] Balduzzi, D. and Tononi, G. (2009). Qualia: The geometry of integrated information. *PLoS computational biology*, 5(8):e1000462. (Cited on pages 6 and 43.)

[25] Bar-Yam, Y. (2003). *Dynamics of Complex Systems*. Studies in Nonlinearity. Westview Press, Boulder, CO. (Cited on page 3.)

[26] Bar-Yam, Y. (2017). Why Complexity is Different. (Cited on pages 3, 7, 9, and 17.)

[27] Bassett, D. S. and Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in cognitive sciences*, 15:200–9. (Cited on page 4.)

[28] Beer, R. D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72(1):173–215. (Cited on page 6.)

[29] Beggs, J. M. (2008). The criticality hypothesis: How local cortical networks might optimize information processing. *Philos T R Soc A*, 366:329–343. (Cited on page 18.)

[30] Beggs, J. M. and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23:11167–77. (Cited on pages 17, 18, 20, and 38.)

[31] Bell, A. J. (1999). Levels and loops: The future of artificial intelligence and neuroscience. *PhilTrans R Soc LondB*, page 8. (Cited on page 11.)

[32] Bell, A. J. (2007). Towards a Cross-Level Theory of Neural Learning. *AIP Conference Proceedings*, 954(1):56–73. (Cited on page 11.)

[33] Bertschinger, N. and Natschlager, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural computation*, 16:1413–1436. (Cited on pages 18, 37, and 53.)

[34] Besserve, M. and Safavi, S. (2016). Practical on Machine Learning for Neuroscience. In *Machine Learning Summer School (MLSS 2016)*. (Cited on page 33.)

[35] Besserve, M., Safavi, S., Kapoor, V., Panagiotaropoulos, T. I., and Logothetis, N. K. (2019). Generalized phase locking analysis of electrophysiology data. In *Computational and Systems Neuroscience Meeting (COSYNE 2019)*, pages 184–185. (Cited on page 31.)

[36] Betzel, R. F. and Bassett, D. S. (2017). Multi-scale brain networks. *NeuroImage*, 160:73–83. (Cited on page 4.)

[37] Bialek, W., Cavagna, A., Giardina, I., Mora, T., Pohl, O., Silvestri, E., Viale, M., and Walczak, A. M. (2014). Social interactions dominate speed control in poising natural flocks near criticality. *Proceedings of the National Academy of Sciences of the United States of America*, 111:7212–7. (Cited on page 6.)

[38] Bialek, W., Cavagna, A., Giardina, I., Mora, T., Silvestri, E., Viale, M., and Walczak, A. M. (2012). Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences of the United States of America*, 109:4786–91. (Cited on pages 6 and 7.)

[39] Bialek, W. and DeWeese, M. (1995). Random switching and optimal processing in the perception of ambiguous signals. *Physical review letters*, 74:3077–3080. (Cited on pages 8, 55, and 56.)

[40] Bialek, W., Rieke, F., van Steveninck, R. d. R., and Warland, D. (1991). Reading a neural code. *Science*, 252(5014):1854–1857. (Cited on page 6.)

[41] Bießmann, F., Meinecke, F. C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N. K., and Müller, K.-R. (2009). Temporal kernel CCA and its application in multimodal neuronal data analysis. *Mach. Learn.*, 79(1-2):5–27. (Cited on pages 13 and 56.)

[42] Biessmann, F., Plis, S., Meinecke, F. C., Eichele, T., and Muller, K. R. (2011). Analysis of multimodal neuroimaging data. *IEEE Rev Biomed Eng*, 4:26–58. (Cited on page 56.)

[43] Bill, J., Buesing, L., Habenschuss, S., Nessler, B., Maass, W., and Legenstein, R. (2015). Distributed Bayesian Computation and Self-Organized Learning in Sheets of Spiking Neurons with Local Lateral Inhibition. *PLOS ONE*, 10(8):e0134356. (Cited on pages 37, 53, and 54.)

[44] Blake, R. and Logothetis, N. (2002). Visual competition. *Nat Rev Neurosci*, 3(1):13–21. (Cited on pages 22, 23, 44, and 55.)

[45] Boerlin, M., Machens, C. K., and Deneve, S. (2013). Predictive coding of dynamical variables in balanced spiking networks. *PLoS computational biology*, 9:e1003258. (Cited on pages 20, 37, 38, 53, and 54.)

[46] Bokil, H., Andrews, P., Kulkarni, J. E., Mehta, S., and Mitra, P. P. (2010). Chronux: A platform for analyzing neural signals. *Journal of neuroscience methods*, 192(1):146–51. (Cited on page 47.)

[47] Bonilla-Quintana, M., Wörgötter, F., D'Este, E., Tetzlaff, C., and Fauth, M. (2020). Actin in Dendritic Spines Self-Organizes into a Critical State. *bioRxiv*, page 2020.04.22.054577. (Cited on page 17.)

[48] Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., and Waldorp, L. J. (2011). The Small World of Psychopathology. *PLOS ONE*, 6(11):e27407. (Cited on page 6.)

[49] Borst, A. and Theunissen, F. E. (1999). Information theory and neural coding. *Nat. Neurosci.*, 2(11):947–957. (Cited on page 6.)

[50] Brand, A., Allen, L., Altman, M., Hlava, M., and Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learn. Publ.*, 28(2):151–155. (Cited on page 27.)

[51] Braun, J. and Mattia, M. (2010). Attractors and noise: Twin drivers of decisions and multistability. *NeuroImage*, 52:740–751. (Cited on page 23.)

[52] Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature neuroscience*, 20(3):340–352. (Cited on pages 18 and 54.)

[53] Breakspear, M., Heitmann, S., and Daffertshofer, A. (2010). Generative Models of Cortical Oscillations: Neurobiological Implications of the Kuramoto Model. *Front. Hum. Neurosci.*, 4. (Cited on page 11.)

[54] Brochini, L., de Andrade Costa, A., Abadi, M., Roque, A. C., Stolfi, J., and Kinouchi, O. (2016). Phase transitions and self-organized criticality in networks of stochastic spiking neurons. *Sci. Rep.*, 6:35831. (Cited on pages 8, 18, and 19.)

[55] Brownlee, J. (2007). Complex Adaptive Systems. Technical Report 070302A. (Cited on page 52.)

[56] Buesing, L., Bill, J., Nessler, B., and Maass, W. (2011). Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS computational biology*, 7:e1002211. (Cited on pages 37, 53, and 54.)

[57] Bullmore, E. (2018). *The Inflamed Mind: A Radical New Approach to Depression*. (Cited on page 57.)

[58] Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, 10(3):186–98. (Cited on pages 4 and 6.)

[59] Burmeister, M., McInnis, M. G., and Zöllner, S. (2008). Psychiatric genetics: Progress amid controversy. *Nat. Rev. Genet.*, 9(7):527–540. (Cited on page 57.)

[60] Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nat. Neurosci.*, 7(5):446–451. (Cited on page 14.)

[61] Buzsaki, G. (2010). Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362–85. (Cited on page 54.)

[62] Buzsaki, G. (2011). *Rhythms of the Brain*. Oxford University Press, New York, USA. (Cited on pages 4, 9, and 47.)

[63] Buzsaki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. *Nature reviews. Neuroscience*, 13(6):407–20. (Cited on pages 13 and 33.)

[64] Buzsaki, G., Horvath, Z., Urioste, R., Hetke, J., and Wise, K. (1992). High-frequency network oscillation in the hippocampus. (Cited on pages 14 and 32.)

[65] Buzsaki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms. *Neuron*, 80(3):751–64. (Cited on page 47.)

[66] Buzsaki, G. and Schomburg, E. W. (2015). What does gamma coherence tell us about inter-regional neural communication? *Nature neuroscience*, 18:484–9. (Cited on page 26.)

[67] Carter, O., Pettigrew, J., Hasler, F., Wallis, G., and Vollenweider, F. Psilocybin slows binocular rivalry switching through serotonin modulation. page 1. (Cited on page 55.)

[68] Carter, O., van Swinderen, B., Leopold, D., Collin, S., and Maier, A. (2020). Perceptual rivalry across animal species. *J. Comp. Neurol.*, n/a(n/a). (Cited on page 22.)

[69] Castellano, C., Marsili, M., and Vespignani, A. (2000). Nonequilibrium Phase Transition in a Model for Social Influence. *Phys. Rev. Lett.*, 85(16):3536–3539. (Cited on page 17.)

[70] Cavanagh, S. E., Lam, N. H., Murray, J. D., Hunt, L. T., and Kennerley, S. W. (2019). A circuit mechanism for irrationalities in decision-making and NMDA receptor hypofunction: Behaviour, computational modelling, and pharmacology. *bioRxiv*, page 826214. (Cited on page 55.)

[71] Chalk, M., Gutkin, B., and Deneve, S. (2016). Neural oscillations as a signature of efficient coding in the presence of synaptic delays. *eLife*, 5. (Cited on pages 20, 37, 38, 53, and 54.)

[72] Chalk, M., Marre, O., and Tkačik, G. (2018). Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U.S.A.*, 115(1):186–191. (Cited on page 20.)

[73] Chen, B., Zhu, Z., Na, R., Fang, W., Zhang, W., Zhou, Q., Zhou, S., Lei, H., Huang, A., Chen, T., Ni, D., Gu, Y., Liu, J., Fang, F., and Rao, Y. (2018). Genomic Analyses of Visual Cognition: Perceptual Rivalry and Top-Down Control. *J. Neurosci.*, 38(45):9668–9678. (Cited on page 55.)

[74] Chen, G. and Gong, P. (2019). Computing by modulating spontaneous cortical activity patterns as a mechanism of active visual processing. *Nat Commun*, 10(1):1–15. (Cited on pages 37 and 53.)

[75] Chialvo, D. R. (2010). Emergent complex neural dynamics. *Nat Phys*, 6(10):744–750. (Cited on pages 4 and 9.)

[76] Chialvo, D. R. (2018). Life at the edge: Complexity and criticality in biological function. *ArXiv181011737 Q-Bio*. (Cited on page 8.)

[77] Chowdhury, D. (1999). Immune Network: An Example of Complex Adaptive Systems. In Dasgupta, D., editor, *Artificial Immune Systems and Their Applications*, pages 89–104. Springer, Berlin, Heidelberg. (Cited on page 52.)

[78] Christof, K. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Roberts and Company Publishers, Denver, Colo., 1st edition edition. (Cited on page 21.)

[79] Churchland, P. S. and Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science*, 242(4879):741–745. (Cited on pages 10 and 59.)

[80] Churchland, P. S. and Sejnowski, T. J. (1992). *The Computational Brain*. Computational Neuroscience. MIT Press, Cambridge, Mass. (Cited on pages 4, 9, 11, 37, 51, and 57.)

[81] Clawson, W. P., Wright, N. C., Wessel, R., and Shew, W. L. (2017). Adaptation towards scale-free dynamics improves cortical stimulus discrimination at the cost of reduced detection. *PLoS computational biology*, 13:e1005574. (Cited on page 19.)

[82] Cocchi, L., Gollo, L. L., Zalesky, A., and Breakspear, M. (2017). Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in Neurobiology*, 158:132–152. (Cited on page 18.)

[83] Cohen, B. P., Chow, C. C., and Vattikuti, S. (2019). Dynamical modeling of multi-scale variability in neuronal competition. *Commun Biol*, 2(1):1–11. (Cited on page 55.)

[84] Cohen, M. R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature neuroscience*, 14(7):811–9. (Cited on pages 24, 41, and 42.)

[85] Constantinidis, C. and Goldman-Rakic, P. S. (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of neurophysiology*, 88(6):3487–3497. (Cited on page 41.)

[86] Costa, M. S., Weigenand, A., Ngo, H.-V. V., Marshall, L., Born, J., Martinetz, T., and Claussen, J. C. (2016). A Thalamocortical Neural Mass Model of the EEG during NREM Sleep and Its Response to Auditory Stimulation. *PLOS Computational Biology*, 12(9):e1005022. (Cited on pages 16 and 34.)

[87] Crick, F. (1996). Visual perception: Rivalry and consciousness. *Nature*, 379(6565):485–6. (Cited on page 21.)

[88] Crick, F. and Koch, C. (1998). Consciousness and neuroscience. *Cerebral cortex*, 8(2):97–107. (Cited on page 43.)

[89] Csicsvari, J., Hirase, H., Mamiya, A., and Buzsáki, G. (2000). Ensemble Patterns of Hippocampal CA3-CA1 Neurons during Sharp Wave–Associated Population Events. *Neuron*, 28(2):585–594. (Cited on pages 14 and 33.)

[90] Dai, H., Wang, Y., Trivedi, R., and Song, L. (2016). Recurrent Coevolutionary Latent Feature Processes for Continuous-Time Recommendation. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, DLRS 2016, pages 29–34, New York, NY, USA. Association for Computing Machinery. (Cited on page 29.)

[91] Daniels, B. C., Ellison, C. J., Krakauer, D. C., and Flack, J. C. (2016). Quantifying collectivity. *Curr Opin Neurobiol*, 37:106–113. (Cited on page 54.)

[92] Dayan, P. (1998). A Hierarchical Model of Binocular Rivalry. *Neural Comput.*, 10(5):1119–1135. (Cited on pages 55 and 56.)

[93] De, A., Valera, I., Ganguly, N., Bhattacharya, S., and Gomez-Rodriguez, M. (2016). Learning and forecasting opinion dynamics in social networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, pages 397–405, Red Hook, NY, USA. Curran Associates Inc. (Cited on page 29.)

[94] de Abreu, M. S., Giacomini, A. C. V. V., Zanandrea, R., dos Santos, B. E., Genario, R., de Oliveira, G. G., Friend, A. J., Amstislavskaya, T. G., and Kalueff, A. V. (2018). Psychoneuroimmunology and immunopsychiatry of zebrafish. *Psychoneuroendocrinology*, 92:1–12. (Cited on page 57.)

[95] de Graaf, T. A., Hsieh, P.-J., and Sack, A. T. (2012). The 'correlates' in neural correlates of consciousness. *Neurosci Biobehav Rev*, 36(1):191–197. (Cited on pages 39 and 43.)

[96] Deco, G., Jirsa, V. K., and McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature reviews. Neuroscience*, 12:43–56. (Cited on page 6.)

[97] Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS computational biology*, 4(8):e1000092. (Cited on pages 6 and 54.)

[98] Dehaene, S. and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–27. (Cited on page 43.)

[99] Deneve, S. (2008a). Bayesian spiking neurons I: Inference. *Neural computation*, 20:91–117. (Cited on pages 37 and 53.)

[100] Deneve, S. (2008b). Bayesian spiking neurons II: Learning. *Neural computation*, 20:118–45. (Cited on pages 37 and 53.)

[101] Deneve, S. and Jardri, R. (2016). Circular inference: Mistaken belief, misplaced trust. *Current Opinion in Behavioral Sciences*, 11:40–48. (Cited on page 57.)

[102] Denman, D. J. and Contreras, D. (2014). The Structure of Pairwise Correlation in Mouse Primary Visual Cortex Reveals Functional Organization in the Absence of an Orientation Map. *Cereb Cortex*, 24(10):2707–2720. (Cited on pages 24 and 41.)

[103] di Santo, S., Villegas, P., Burioni, R., and Muñoz, M. A. (2018). Landau–Ginzburg theory of cortex dynamics: Scale-free avalanches emerge at the edge of synchronization. *PNAS*, page 201712989. (Cited on page 18.)

[104] Dickey, A. S., Suminski, A., Amit, Y., and Hatsopoulos, N. G. (2009). Single-Unit Stability Using Chronically Implanted Multielectrode Arrays. *J. Neurophysiol.*, 102(2):1331–1339. (Cited on page 14.)

[105] Ditzinger, T. and Haken, H. (1989). Oscillations in the Perception of Ambiguous Patterns - a Model Based on Synergetics. *Biological cybernetics*, 61(4):279–287. (Cited on page 23.)

[106] Doesburg, S. M., Green, J. J., McDonald, J. J., and Ward, L. M. (2009). Rhythms of consciousness: Binocular rivalry reveals large-scale oscillatory network dynamics mediating visual perception. *PLoS One*, 4(7):e6142. (Cited on pages 22, 23, and 44.)

[107] Douglas, R. J. and Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27(1):419–451. (Cited on page 41.)

[108] Douglas, R. J. and Martin, K. A. (2007a). Recurrent neuronal circuits in the neocortex. *Current biology : CB*, 17:R496–500. (Cited on page 11.)

[109] Douglas, R. J., Martin, K. A., and Whitteridge, D. (1989). A Canonical Microcircuit for Neocortex. *Neural Comput.*, 1(4):480–488. (Cited on page 41.)

[110] Douglas, R. J. and Martin, K. A. C. (2007b). Mapping the Matrix: The Ways of Neocortex. *Neuron*, 56(2):226–238. (Cited on page 41.)

[111] Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N. (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press. (Cited on page 55.)

[112] Dubreuil, A. M., Valente, A., Beiran, M., Mastrogiuseppe, F., and Ostojic, S. (2020). Complementary roles of dimensionality and population structure in neural computations. *bioRxiv*, page 2020.07.03.185942. (Cited on page 54.)

[113] Dwarakanath, A., Kapoor, V., Safavi, S., Logothetis, N. K., and Eschenko, O. (2018). Perisynaptic activity in the prefrontal cortex reflects spontaneous transitions in conscious visual perception. In *AREADNE 2018: Research in Encoding And Decoding of Neural Ensembles*, page 58. AREADNE Foundation. (Cited on page 47.)

[114] Dwarakanath, A., Kapoor, V., Werner, J., Safavi, S., Fedorov, L. A., Logothetis, N. K., and Panagiotaropoulos, T. I. (2020). Prefrontal state fluctuations control access to consciousness. *bioRxiv*, page 2020.01.29.924928. (Cited on pages 25, 39, 47, and 91.)

[115] Dwarakanath, A., Safavi, S., Kapoor, V., Logothetis, N. K., and Panagiotaropoulos, T. I. (2015). Temporal Regimes of State-Dependent Correlated Variability in the Macaque Ventrolateral Prefrontal Cortex. page 18. (Cited on page 41.)

[116] Echeveste, R., Aitchison, L., Hennequin, G., and Lengyel, M. (2020). Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nat. Neurosci.*, 23(9):1138–1149. (Cited on pages 37, 53, and 54.)

[117] Einevoll, G. T., Destexhe, A., Diesmann, M., Grün, S., Jirsa, V., de Kamps, M., Migliore, M., Ness, T. V., Plesser, H. E., and Schürmann, F. (2019). The Scientific Case for Brain Simulations. *Neuron*, 102(4):735–744. (Cited on page 9.)

[118] Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nature reviews. Neuroscience*, 14(11):770–85. (Cited on pages 13 and 33.)

[119] Eliasmith, C. (2005). A Unified Approach to Building and Controlling Spiking Attractor Networks. *Neural Computation*, 17(6):1276–1314. (Cited on pages 37 and 53.)

[120] Ellis, N. C. and Larsen-Freeman, D. (2009). *Language as a Complex Adaptive System*. John Wiley & Sons. (Cited on page 52.)

[121] Ermentrout, G. B., Galán, R. F., and Urban, N. N. (2007). Relating Neural Dynamics to Neural Coding. *Phys. Rev. Lett.*, 99(24):248103. (Cited on pages 37 and 54.)

[122] Eurich, C. W. (2003). Neural Dynamics and Neural Coding Two Complementary Approaches. Technical report. (Cited on page 37.)

[123] Fazli, S., Dahne, S., Samek, W., Biessmann, F., and Muller, K. R. (2015). Learning From More Than One Data Source: Data Fusion Techniques for Sensorimotor Rhythm-Based Brain-Computer Interfaces. *P Ieee*, 103:891–906. (Cited on page 56.)

[124] Févotte, C., Bertin, N., and Durrieu, J. (2009). Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. *Neural Comput.*, 21(3):793–830. (Cited on page 34.)

[125] Finlinson, K., Shew, W. L., Larremore, D. B., and Restrepo, J. G. (2020). Optimal control of excitable systems near criticality. *Phys. Rev. Research*, 2(3):033450. (Cited on pages 18, 37, and 53.)

[126] Forstmann, B. U. and Wagenmakers, E.-J., editors (2015). *An Introduction to Model-Based Cognitive Neuroscience*. Springer, New York. (Cited on page 51.)

[127] Frank, M. J. (2015). Linking Across Levels of Computation in Model-Based Cognitive Neuroscience. In Forstmann, B. U. and Wagenmakers, E.-J., editors, *An Introduction to Model-Based Cognitive Neuroscience*, pages 159–177. Springer, New York, NY. (Cited on page 55.)

[128] Frassle, S., Sommer, J., Jansen, A., Naber, M., and Einhauser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34(5):1738–47. (Cited on pages 24, 26, and 39.)

[129] Freeman, W. J. and Breakspear, M. (2007). Scale-free neocortical dynamics. *Scholarpedia*, 2(2):1357. (Cited on pages 33 and 54.)

[130] Friedman, N., Ito, S., Brinkman, B. A., Shimono, M., DeVille, R. E., Dahmen, K. A., Beggs, J. M., and Butler, T. C. (2012). Universal critical dynamics in high resolution neuronal avalanche data. *Physical review letters*, 108:208102. (Cited on page 18.)

[131] Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9:474–480. (Cited on pages 7 and 10.)

[132] Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, 88:220–35. (Cited on pages 7 and 10.)

[133] Fujita, I. and Fujita, T. (1996). Intrinsic connections in the macaque inferior temporal cortex. *J. Comp. Neurol.*, 368(4):467–486. (Cited on pages 25 and 41.)

[134] Fukushima, M., Chao, Z. C., and Fujii, N. (2015). Studying brain functions with mesoscopic measurements: Advances in electrocorticography for non-human primates. *Current Opinion in Neurobiology*, 32:124–131. (Cited on page 14.)

[135] Gell-Mann, M. (1994). Complex Adaptive Systems. In Cowan, G., Pines, D., and Meltzer, D., editors, *Complexity: Metaphors, Models, and Reality*, number 19, pages 17–45. Addison-Wesley, Reading, MA. (Cited on page 52.)

[136] Gershman, S., Vul, E., and Tenenbaum, J. B. (2014). Perceptual Multistability as Markov Chain Monte Carlo Inference. In *Advances in Neural Information Processing Systems*, pages 611–619. (Cited on pages 55 and 56.)

[137] Gerstner, W., Kistler, W. M., Naud, R., and Paninsk, L. (2014). *Neuronal Dynamics, From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, University Printing House, Cambridge CB2 8BS, United Kingdom. (Cited on page 6.)

[138] Goense, J. B. and Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current biology : CB*, 18:631–40. (Cited on page 13.)

[139] Gutenberg, B. and Richter, C. F. (1956). Earthquake magnitude, intensity, energy, and acceleration(Second paper). *Bulletin of the Seismological Society of America*, 46(2):105–145. (Cited on page 18.)

[140] Haddad, F., Patel, S., and Schmid, S. (2020). Maternal Immune Activation by Poly I:c as a preclinical Model for Neurodevelopmental Disorders: A focus on Autism and Schizophrenia. *Neuroscience & Biobehavioral Reviews*. (Cited on page 57.)

[141] Hall, G. and Bialek, W. (2019). The statistical mechanics of Twitter communities. *J. Stat. Mech.*, 2019(9):093406. (Cited on page 6.)

[142] Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nature reviews. Neuroscience*, 6(5):399–407. (Cited on page 54.)

[143] Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., and Buzsaki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424:552–6. (Cited on page 54.)

[144] Harris, K. D. and Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, 503(7474):51–8. (Cited on page 41.)

[145] Harris, T. E. (1963). *The Theory of Branching Processes*. Grundlehren Der Mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg. (Cited on page 18.)

[146] Hasenstaub, A., Shu, Y., Haider, B., Kraushaar, U., Duque, A., and McCormick, D. A. (2005). Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron*, 47(3):423–35. (Cited on page 10.)

[147] He, B. J. (2014). Scale-free brain activity: Past, present, and future. *Trends in cognitive sciences*, 18:480–7. (Cited on page 33.)

[148] Heine, J., Prüß, H., Scheel, M., Brandt, A. U., Gold, S. M., Bartsch, T., Paul, F., and Finke, C. (2020). Transdiagnostic hippocampal damage patterns in neuroimmunological disorders. *NeuroImage: Clinical*, 28:102515. (Cited on page 57.)

[149] Herreras, O. (2016). Local Field Potentials: Myths and Misunderstandings. *Front Neural Circuit*, 10:101. (Cited on page 13.)

[150] Hesse, J. K. and Tsao, D. Y. (2020). A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli. *eLife*, 9:e58360. (Cited on pages 24 and 26.)

[151] Hidalgo, J., Grilli, J., Suweis, S., Maritan, A., and Muñoz, M. A. (2016). Cooperation, competition and the emergence of criticality in communities of adaptive systems. *J. Stat. Mech.*, 2016(3):033203. (Cited on page 18.)

[152] Hidalgo, J., Grilli, J., Suweis, S., Munoz, M. A., Banavar, J. R., and Maritan, A. (2014). Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences of the United States of America*, 111:10095–100. (Cited on pages 18, 37, and 53.)

[153] Hipp, J. F., Engel, A. K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*, 69:387–96. (Cited on pages 22, 23, and 44.)

[154] Hoffmann, H. and Payton, D. W. (2018). Optimization by Self-Organized Criticality. *Sci. Rep.*, 8(1):2358. (Cited on page 18.)

[155] Hohwy, J., Roepstorff, A., and Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3):687–701. (Cited on pages 55 and 56.)

[156] Holland, J. H. (2006). Studying Complex Adaptive Systems. *Jrl Syst Sci & Complex*, 19(1):1–8. (Cited on pages 51 and 52.)

[157] Holland, J. H. (2012). *Signals and Boundaries: Building Blocks for Complex Adaptive Systems*. The MIT Press, illustrated edition edition. (Cited on page 52.)

[158] Holland, J. H. (2014). *Complexity: A Very Short Introduction*. Oxford University Press, Oxford, United Kingdom, 1 edition edition. (Cited on pages 3 and 52.)

[159] Huys, Q. J. M., Browning, M., Paulus, M. P., and Frank, M. J. (2020). Advances in the computational understanding of mental illness. *Neuropsychopharmacology*, pages 1–17. (Cited on page 57.)

[160] Imamoglu, F., Kahnt, T., Koch, C., and Haynes, J. D. (2012). Changes in functional connectivity support conscious object recognition. *NeuroImage*, 63(4):1909–17. (Cited on pages 22, 23, and 44.)

[161] Issler, O. and Chen, A. (2015). Determining the role of microRNAs in psychiatric disorders. *Nat. Rev. Neurosci.*, 16(4):201–212. (Cited on page 57.)

[162] Izhikevich, E. M. (2010). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting (Computational Neuroscience)*. The MIT Press, Cambrige, Massachusetts, USA. (Cited on pages 6 and 18.)

[163] Izhikevich, E. M. and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9):3593–8. (Cited on page 6.)

[164] Jiang, H., Bahramisharif, A., van Gerven, M. A. J., and Jensen, O. (2015). Measuring directionality between neuronal oscillations of different frequencies. *NeuroImage*, 118:359–367. (Cited on page 13.)

[165] Johnson, D. H. (1996). Point process models of single-neuron discharges. *J Comput Neurosci*, 3(4):275–299. (Cited on page 29.)

[166] Johnson, J. K., Wright, N. C., Xia, J., and Wessel, R. (2019). Single-cell membrane potential fluctuations evince network scale-freeness and quasicriticality. *J. Neurosci.*, pages 3163–18. (Cited on page 17.)

[167] Juavinett, A. L., Bekheet, G., and Churchland, A. K. (2019). Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *eLife*, 8:e47188. (Cited on page 14.)

[168] Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydin, C., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Hausser, M., Karsh, B., Ledochowitsch, P., Lopez, C. M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P. D., Rossant, C., Sun, W. L., Svoboda, K., Carandini, M., Harris, K. D., Koch, C., O'Keefe, J., and Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551:232–236. (Cited on page 14.)

[169] Kadmon, J., Timcheck, J., and Ganguli, S. (2020). Predictive coding in balanced neural networks with noise, chaos and delays. *ArXiv200614178 Cond-Mat Q-Bio Stat*. (Cited on page 37.)

[170] Kalloniatis, A. C., Zuparic, M. L., and Prokopenko, M. (2018). Fisher information and criticality in the Kuramoto model of nonidentical oscillators. *Phys. Rev. E*, 98(2):022302. (Cited on page 54.)

[171] Kanders, K., Lorimer, T., and Stoop, R. (2017). Avalanche and edge-of-chaos criticality do not necessarily co-occur in neural networks. *Chaos*, 27(4):047408. (Cited on page 19.)

[172] Kapoor, V., Besserve, M., Logothetis, N. K., and Panagiotaropoulos, T. I. (2018a). Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. *Commun. Biol.*, 1(1):1–12. (Cited on pages 26 and 40.)

[173] Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T. I., and Logothetis, N. K. (2020). Decoding the contents of consciousness from prefrontal ensembles. *bioRxiv*, page 2020.01.28.921841. (Cited on pages 22, 24, 25, 39, 43, and 91.)

[174] Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Nicholas, H., Logothetis, N. K., and Panagiotaropoulos, T. I. (2018b). Spiking activity in the prefrontal cortex reflects spontaneous perceptual transitions during a no report binocular rivalry paradigm. In *11th FENS Forum of Neuroscience*. (Cited on page 43.)

[175] Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Panagiotaropoulos, T. I., and Logothetis, N. K. (2019). Neuronal discharges in the prefrontal cortex reflect changes in conscious perception during a no report binocular rivalry paradigm. (Cited on page 43.)

[176] Kar, K. and DiCarlo, J. J. (2020). Fast recurrent processing via ventral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. *bioRxiv*, page 2020.05.10.086959. (Cited on page 26.)

[177] Karimi Abadchi, J., Nazari-Ahangarkolaee, M., Gattas, S., Bermudez-Contreras, E., Luczak, A., McNaughton, B. L., and Mohajerani, M. H. (2020). Spatiotemporal patterns of neocortical activity around hippocampal sharp-wave ripples. *eLife*, 9:e51972. (Cited on page 33.)

[178] Keliris, G. A., Logothetis, N. K., and Tolias, A. S. (2010). The role of the primary visual cortex in perceptual suppression of salient visual stimuli. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(37):12353–65. (Cited on pages 40 and 44.)

[179] Khajehabdollahi, S., Abeyasinghe, P., Owen, A., and Soddu, A. (2019). The emergence of integrated information, complexity, and consciousness at criticality. *bioRxiv*, page 521567. (Cited on page 18.)

[180] Khandaker, G. M., Cousins, L., Deakin, J., Lennox, B. R., Yolken, R., and Jones, P. B. (2015). Inflammation and immunity in schizophrenia: Implications for pathophysiology and treatment. *The Lancet Psychiatry*, 2(3):258–270. (Cited on page 57.)

[181] Khandaker, G. M., Meyer, U., and Jones, P. B. (2020). *Neuroinflammation and Schizophrenia*. (Cited on page 57.)

[182] Kim, C. M. and Chow, C. C. (2018). Learning recurrent dynamics in spiking networks. *eLife*, 7:e37124. (Cited on pages 37 and 53.)

[183] Kinouchi, O. and Copelli, M. (2006). Optimal dynamical range of excitable networks at criticality. *Nat Phys*, 2:348–352. (Cited on pages 8, 18, and 19.)

[184] Klink, P. C., van Wezel, R. J. A., and van Ee, R. (2012). United we sense, divided we fail: Context-driven perception of ambiguous visual stimuli. *Philos Trans R Soc Lond B Biol Sci*, 367(1591):932–941. (Cited on page 55.)

[185] Knill, D. C. and Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in neurosciences*, 27:712–9. (Cited on page 55.)

[186] Koch, C. (2012a). *Consciousness: Confessions of a Romantic Reductionist*. The MIT Press. (Cited on page 21.)

[187] Koch, C. (2012b). Systems biology. Modular biological complexity. *Science*, 337:531–2. (Cited on page 4.)

[188] Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nat. Rev. Neurosci.*, 17(5):307–321. (Cited on pages 22, 23, and 44.)

[189] Kol, A. and Goshen, I. (2021). The memory orchestra: The role of astrocytes and oligodendrocytes in parallel to neurons. *Current Opinion in Neurobiology*, 67:131–137. (Cited on page 57.)

[190] Koren, T., Krot, M., Boshnak, N. T., Amer, M., Ben-Shaanan, T., Azulay-Debby, H., Hajjo, H., Avishai, E., Schiller, M., Haykin, H., Korin, B., Farfara, D., Hakim, F., Rosenblum, K., and Rolls, A. (2020). Remembering immunity: Neuronal ensembles in the insular cortex encode and retrieve specific immune responses. *bioRxiv*, page 2020.12.03.409813. (Cited on page 57.)

[191] Kritzer, M. F. and Goldman-Rakic, P. S. (1995). Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.*, 359(1):131–143. (Cited on pages 25 and 41.)

[192] Kuebler, E. S., Calderini, M., Lambert, P., and Thivierge, J.-P. (2019). Optimal Fisher Decoding of Neural Activity Near Criticality. In Tomen, N., Herrmann, J. M., and Ernst, U., editors, *The Functional Role of Critical Dynamics in Neural Systems*, Springer Series on Bio- and Neurosystems, pages 159–177. Springer International Publishing, Cham. (Cited on page 54.)

[193] Kuramoto, Y. (1975). Self-entrainment of a population of coupled non-linear oscillators. In Araki, H., editor, *International Symposium on Mathematical Problems in Theoretical Physics*, Lecture Notes in Physics, pages 420–422, Berlin, Heidelberg. Springer. (Cited on pages 3 and 11.)

[194] Kuramoto, Y. (2003). *Chemical Oscillations, Waves, and Turbulence*. Courier Corporation. (Cited on pages 3 and 11.)

[195] Larremore, D. B., Shew, W. L., and Restrepo, J. G. (2011). Predicting Criticality and Dynamic Range in Complex Networks: Effects of Topology. *Phys. Rev. Lett.*, 106(5):058101. (Cited on pages 8, 18, and 19.)

[196] Lau, H. and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences*, 15:365–73. (Cited on page 43.)

[197] Law, P. C., Miller, S. M., and Ngo, T. T. (2017). The effect of stimulus strength on binocular rivalry rate in healthy individuals: Implications for genetic, clinical and individual differences studies. *Physiology & Behavior*, 181:127–136. (Cited on page 55.)

[198] Le Van Quyen, M. (2003). Disentangling the dynamic core: A research program for a neurodynamics at the large-scale. *Biol. Res.*, 36(1):67–88. (Cited on page 11.)

[199] Le Van Quyen, M. (2011). The brainweb of cross-scale interactions. *New Ideas in Psychology*, 29(2):57–63. (Cited on pages 10 and 11.)

[200] Le Van Quyen, M., Chavez, M., Rudrauf, D., and Martinerie, J. (2003). Exploring the nonlinear dynamics of the brain. *Journal of Physiology-Paris*, 97(4):629–639. (Cited on page 11.)

[201] Lehky, S. R. and Maunsell, J. H. R. (1996). No binocular rivalry in the LGN of alert macaque monkeys. *Vision research*, 36(9):1225–1234. (Cited on pages 22 and 44.)

[202] Lengyel, M., Kwag, J., Paulsen, O., and Dayan, P. (2005). Matching storage and recall: Hippocampal spike timing–dependent plasticity and phase response curves. *Nat. Neurosci.*, 8(12):1677–1683. (Cited on pages 37, 53, and 54.)

[203] Leong, V. and Schilbach, L. (2019). The promise of two-person neuroscience for developmental psychiatry: Using interaction-based sociometrics to identify disorders of social interaction. *Br. J. Psychiatry*, 215(5):636–638. (Cited on page 57.)

[204] Leopold, D., Maier, A., and Logothetis, N. (2003). Measuring Subjective Visual Perception in the Nonhuman Primate. *Journal of Consciousness Studies*, 10(9-10):115–130. (Cited on page 24.)

[205] Leopold, D. A. (2012). Primary Visual Cortex: Awareness and Blindsight. *Annu. Rev. Neurosci.*, 35(1):91–109. (Cited on page 40.)

[206] Leopold, D. A. and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, 379(6565):549–53. (Cited on page 40.)

[207] Leptourgos, P. (2018). *Dynamical Circular Inference in the General Population and the Psychosis Spectrum : Insights from Perceptual Decision Making*. Thesis, Paris Sciences et Lettres. (Cited on page 56.)

[208] Levelt, W. J. M. (1967). Note on the Distribution of Dominance Times in Binocular Rivalry. *Br. J. Psychol.*, 58(1-2):143–145. (Cited on page 45.)

[209] Levina, A., Safavi, S., Logothetis, N. K., and Chalk, M. (2020). Signatures of criticality observed in efficient coding networks. In *Computational and Systems Neuroscience Meeting (COSYNE 2020)*, page 109. (Cited on page 37.)

[210] Levitt, J. B., Lewis, D. A., Yoshioka, T., and Lund, J. S. (1993). Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *J. Comp. Neurol.*, 338(3):360–376. (Cited on pages 25 and 41.)

[211] Li, C. Y., Poo, M. M., and Dan, Y. (2009). Burst spiking of a single cortical neuron modifies global brain state. *Science*, 324(5927):643–6. (Cited on page 10.)

[212] Li, L., Brockmeier, A. J., Choi, J. S., Francis, J. T., Sanchez, J. C., and Principe, J. C. (2014). A tensor-product-kernel framework for multiscale neural activity decoding and control. *Computational intelligence and neuroscience*, 2014:870160. (Cited on page 56.)

[213] Li, Z., Cui, D., and Li, X. (2016). Unbiased and robust quantification of synchronization between spikes and local field potential. *Journal of neuroscience methods*, 269:33–8. (Cited on page 13.)

[214] Liljenstroem, H. (2012). Mesoscopic brain dynamics. *Scholarpedia*, 7(9):4601. (Cited on pages 13 and 33.)

[215] Lizier, J. T. (2013). *The Local Information Dynamics of Distributed Computation in Complex Systems*. Springer Theses. Springer, Berlin. (Cited on page 19.)

[216] Logothetis, N. (2014a). Studies of Large-Scale Networks with DES- & NET-fMRI. Technical report, Max Planck Institute for Biological Cybernetics. (Cited on page 39.)

[217] Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23:3963–71. (Cited on page 13.)

[218] Logothetis, N. K. (2006). Vision: A Window into Consciousness. *Sci Am*, 16(3):4–11. (Cited on page 21.)

[219] Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453:869–78. (Cited on pages 13 and 22.)

[220] Logothetis, N. K. (2012). Intracortical recordings and fMRI: An attempt to study operational modules and networks simultaneously. *NeuroImage*, 62(2):962–9. (Cited on page 7.)

[221] Logothetis, N. K. (2014b). Neural-Event-Triggered fMRI of large-scale neural networks. *Curr. Opin. Neurobiol.*, 31C:214–222. (Cited on pages 15 and 35.)

[222] Logothetis, N. K., Eschenko, O., Murayama, Y., Augath, M., Steudel, T., Evrard, H. C., Besserve, M., and Oeltermann, A. (2012). Hippocampal–cortical interaction during periods of subcortical silence. *Nature*, 491(7425):547–553. (Cited on pages 7, 13, 14, 15, 16, 33, and 34.)

[223] Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–7. (Cited on page 13.)

[224] Lukovic, M., Vanni, F., Svenkeson, A., and Grigolini, P. (2014). Transmission of information at criticality. *Physica A*, 416:430–438. (Cited on pages 8 and 18.)

[225] Lumer, E. D., Friston, K. J., and Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, 280(5371):1930–4. (Cited on pages 23 and 44.)

[226] Lund, J. S., Yoshioka, T., and Levitt, J. B. (1993). Comparison of Intrinsic Connectivity in Different Areas of Macaque Monkey Cerebral Cortex. *Cereb Cortex*, 3(2):148–162. (Cited on pages 25 and 41.)

[227] Lynn, C. W., Papadopoulos, L., Kahn, A. E., and Bassett, D. S. (2020). Human information processing in complex networks. *Nat. Phys.*, pages 1–9. (Cited on page 4.)

[228] M, D., Cs, M., Gn, P., J, R., An, B., Dj, T., Y, L., Km, J., and Zx, W. (2020). Social isolation alters behavior, the gut-immune-brain axis, and neurochemical circuits in male and female prairie voles. *Neurobiology of Stress*, page 100278. (Cited on page 57.)

[229] Maass, W. (2016). Searching for principles of brain computation. *Curr. Opin. Behav. Sci.*, 11:81–92. (Cited on pages 37 and 53.)

[230] Magnasco, M. O., Piro, O., and Cecchi, G. A. (2009). Self-tuned critical anti-Hebbian networks. *Physical review letters*, 102:258102. (Cited on page 18.)

[231] Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., and Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature neuroscience*, 11(10):1193–200. (Cited on page 40.)

[232] Mailhé, B., Lesage, S., Gribonval, R., Bimbot, F., and Vandergheynst, P. (2008). Shift-invariant dictionary learning for sparse representations: Extending K-SVD. In *2008 16th European Signal Processing Conference*, pages 1–5. (Cited on page 34.)

[233] Malamud, B. D., Morein, G., and Turcotte, D. L. (1998). Forest Fires: An Example of Self-Organized Critical Behavior. *Science*, 281(5384):1840–1842. (Cited on page 18.)

[234] Mallat, S. G. (1999). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 2nd ed edition. (Cited on page 47.)

[235] Malsburg, C. V. D., Phillips, W. A., and Singer, W. (2010). *Malsburg, C: Dynamic Coordination in the Brain - From Neuron: From Neurons to Mind*. The MIT Press, Cambridge, Mass, illustrated edition edition. (Cited on page 7.)

[236] Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR Sb.*, 1(4):457–483. (Cited on pages 15 and 30.)

[237] Marinazzo, D., Pellicoro, M., Wu, G., Angelini, L., Cortes, J. M., and Stramaglia, S. (2014). Information transfer and criticality in the Ising model on the human connectome. *PloS one*, 9:e93616. (Cited on pages 8 and 18.)

[238] Marr, D. and Poggio, T. (1979). From Understanding Computation to Understanding Neural Circuitry. *Neurosci. Res. Program Bull.*, 15(3):470–488. (Cited on pages 51 and 57.)

[239] Mastrogiuseppe, F. and Ostojic, S. (2018). Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks. *Neuron*, 99(3):609–623.e29. (Cited on page 54.)

[240] Mathis, C., Bhattacharya, T., and Walker, S. I. (2017a). The Emergence of Life as a First-Order Phase Transition. *Astrobiology*, 17(3):266–276. (Cited on page 8.)

[241] Mathis, C., Bhattacharya, T., and Walker, S. I. (2017b). The Emergence of Life as a First-Order Phase Transition. *Astrobiology*, 17(3):266–276. (Cited on page 17.)

[242] Mayer, A. (2017). *Optimal Immune Systems : A Ressource Allocation and Information Processing View of Immune Defense*. Theses, PSL Research University. (Cited on page 57.)

[243] McKenna, T. M., McMullen, T. A., and Shlesinger, M. F. (1994). The brain as a dynamic physical system. *Neuroscience*, 60(3):587–605. (Cited on page 6.)

[244] Mediano, P. A. M., Farah, J. C., and Shanahan, M. (2016). Integrated Information and Metastability in Systems of Coupled Oscillators. *ArXiv160608313 Q-Bio*. (Cited on page 18.)

[245] Mentch, J., Spiegel, A., Ricciardi, C., and Robertson, C. E. (2019). GABAergic Inhibition Gates Perceptual Awareness During Binocular Rivalry. *J. Neurosci.*, 39(42):8398–8407. (Cited on page 55.)

[246] Meyers, E. M. (2013). The neural decoding toolbox. *Frontiers in neuroinformatics*, 7:8. (Cited on page 45.)

[247] Michiels van Kessenich, L., Berger, D., de Arcangelis, L., and Herrmann, H. J. (2019). Pattern recognition with neuronal avalanche dynamics. *Phys. Rev. E*, 99(1):010302. (Cited on pages 18, 37, and 53.)

[248] Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24:167–202. (Cited on pages 24, 39, and 41.)

[249] Miller, S. M., Hansell, N. K., Ngo, T. T., Liu, G. B., Pettigrew, J. D., Martin, N. G., and Wright, M. J. (2010). Genetic contribution to individual variation in binocular rivalry rate. *Proceedings of the National Academy of Sciences*, 107(6):2664–2668. (Cited on page 55.)

[250] Mitchell, M. (2011). *Complexity: A Guided Tour*. Oxford University Press, Oxford, 1 edition edition. (Cited on pages 3, 4, and 52.)

[251] Mitra, P. and Bokil, H. (2007). *Observed Brain Dynamics*. Oxford University Press, USA. (Cited on page 7.)

[252] Modha, D. S. and Singh, R. (2010). Network architecture of the long-distance pathways in the macaque brain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30):13485–90. (Cited on pages 24, 39, and 41.)

[253] Mora, T. and Bialek, W. (2011). Are Biological Systems Poised at Criticality? *J Stat Phys*, 144:268–302. (Cited on pages 8, 18, and 37.)

[254] Moreno-Bote, R., Rinzel, J., and Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of neurophysiology*, 98(3):1125–39. (Cited on page 55.)

[255] Morowitz, H. J. and Singer, J. L. (1995). *The Mind, The Brain And Complex Adaptive Systems*. Westview Press, Reading, Mass. (Cited on page 52.)

[256] Muñoz, M. A. (2018). Colloquium: Criticality and dynamical scaling in living systems. *Rev. Mod. Phys.*, 90(3):031001. (Cited on pages 6, 8, 17, 18, 37, and 54.)

[257] Murayama, Y., Biessmann, F., Meinecke, F. C., Muller, K. R., Augath, M., Oeltermann, A., and Logothetis, N. K. (2010). Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA. *Magnetic resonance imaging*, 28:1095–103. (Cited on pages 13 and 56.)

[258] Ngo, T. T., Mitchell, P. B., Martin, N. G., and Miller, S. M. (2011). Psychiatric and genetic studies of binocular rivalry: An endophenotype for bipolar disorder? *Acta Neuropsychiatr.*, 23(1):37–42. (Cited on page 55.)

[259] Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154. (Cited on page 51.)

[260] Niv, Y. (2020). The primacy of behavioral research for understanding the brain. Technical report, PsyArXiv. (Cited on page 51.)

[261] Nonnenmacher, M., Behrens, C., Berens, P., Bethge, M., and Macke, J. H. (2017). Signatures of criticality arise from random subsampling in simple population models. *PLOS Computational Biology*, 13(10):e1005718. (Cited on page 19.)

[262] Nur, T., Gautam, S. H., Stenken, J. A., and Shew, W. L. (2019). Probing spatial inhomogeneity of cholinergic changes in cortical state in rat. *Sci. Rep.*, 9(1):9387. (Cited on pages 8, 18, and 19.)

[263] Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS computational biology*, 10:e1003588. (Cited on pages 6 and 43.)

[264] Olbrich, E., Achermann, P., and Wennekers, T. (2011). The sleeping brain as a complex system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1952):3697–3707. (Cited on page 4.)

[265] Oliva, A., Fernandez-Ruiz, A., Buzsaki, G., and Berenyi, A. (2016). Role of Hippocampal CA2 Region in Triggering Sharp-Wave Ripples. *Neuron*, 91:1342–55. (Cited on pages 14 and 33.)

[266] Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–9. (Cited on page 20.)

[267] Paiva, A. R., Park, I., and Principe, J. C. (2009). A reproducing kernel Hilbert space framework for spike train signal processing. *Neural computation*, 21(2):424–49. (Cited on page 56.)

[268] Paiva, A. R. C., Park, I., and Principe, J. C. (2010). Inner Products for Representation and Learning in the Spike Train Domain. *Stat. Signal Process. Neurosci. Neurotechnology*, pages 265–309. (Cited on page 56.)

[269] Panagiotaropoulos, F., Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Hatsopoulos, N. G., and Logothetis, N. K. (2018). Modulation of neural discharges and local field potentials in the macaque prefrontal cortex during binocular rivalry. In *48th Annual Meeting of the Society for Neuroscience (Neuroscience 2018)*. (Cited on pages 43 and 47.)

[270] Panagiotaropoulos, T. I., Deco, G., Kapoor, V., and Logothetis, N. K. (2012). Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron*, 74(5):924–35. (Cited on pages 22, 23, 24, 25, 39, 44, and 45.)

[271] Panagiotaropoulos, T. I., Kapoor, V., and Logothetis, N. K. (2014). Subjective visual perception: From local processing to emergent phenomena of brain activity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1641):20130534. (Cited on pages 22, 23, 44, and 55.)

[272] Park, I. M., Seth, S., Paiva, A. R. C., Li, L., and Principe, J. C. (2013). Kernel methods on spike train space for neuroscience: A tutorial. *arXiv*. (Cited on page 56.)

[273] Pastukhov, A., Garcia-Rodriguez, P. E., Haenicke, J., Guillamon, A., Deco, G., and Braun, J. (2013). Multi-stable perception balances stability and sensitivity. *Frontiers in computational neuroscience*, 7:17. (Cited on pages 8, 23, and 55.)

[274] Pesaran, B., Vinck, M., Einevoll, G. T., Sirota, A., Fries, P., Siegel, M., Truccolo, W., Schroeder, C. E., and Srinivasan, R. (2018). Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation. *Nat. Neurosci.*, page 1. (Cited on pages 13 and 14.)

[275] Pesenson, M. Z., editor (2013). *Multiscale Analysis and Nonlinear Dynamics: From Genes to the Brain*. Reviews of Nonlinear Dynamics and Complexity. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. (Cited on page 51.)

[276] Peterson, E. J. and Voytek, B. (2018). Healthy oscillatory coordination is bounded by single-unit computation. *bioRxiv*, page 309427. (Cited on page 54.)

[277] Pfeiffer, T. and Attwell, D. (2020). Brain's immune cells put the brakes on neurons. *Nature*. (Cited on page 57.)

[278] Pittorino, F., Ibáñez-Berganza, M., di Volo, M., Vezzani, A., and Burioni, R. (2017). Chaos and Correlated Avalanches in Excitatory Neural Networks with Synaptic Plasticity. *Phys. Rev. Lett.*, 118(9):098102. (Cited on page 18.)

[279] Plenz, D. and Thiagarajan, T. C. (2007). The organizing principles of neuronal avalanches: Cell assemblies in the cortex? *Trends in neurosciences*, 30:101–10. (Cited on page 54.)

[280] Prokopenko, M., Lizier, J. T., Obst, O., and Wang, X. R. (2011). Relating Fisher information to order parameters. *Phys. Rev. E*, 84(4):041116. (Cited on page 54.)

[281] Quian Quiroga, R. and Panzeri, S., editors (2013). *Principles of Neural Coding*. CRC Press, Boca Raton. (Cited on pages 20 and 54.)

[282] Rabinovich, M. I., Varona, P., Selverston, A. I., and Abarbanel, H. D. I. (2006). Dynamical principles in neuroscience. *Rev. Mod. Phys.*, 78(4):1213–1265. (Cited on page 6.)

[283] Ramirez-Villegas, J. F., Logothetis, N. K., and Besserve, M. (2015). Diversity of sharp-wave-ripple LFP signatures reveals differentiated brain-wide dynamical events. *Proceedings of the National Academy of Sciences of the United States of America*, 112:E6379–87. (Cited on pages 7 and 15.)

[284] Ramirez-Villegas, J. F., Willeke, K. F., Logothetis, N. K., and Besserve, M. (2018). Dissecting the Synapse- and Frequency-Dependent Network Mechanisms of In Vivo Hippocampal Sharp Wave-Ripples. *Neuron*, 100(5):1224–1240.e13. (Cited on pages 14 and 32.)

[285] Rasch, M., Logothetis, N. K., and Kreiman, G. (2009). From neurons to circuits: Linear estimation of local field potentials. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29:13785–96. (Cited on page 10.)

[286] Rasch, M. J., Gretton, A., Murayama, Y., Maass, W., and Logothetis, N. K. (2008). Inferring spike trains from local field potentials. *Journal of neurophysiology*, 99(3):1461–76. (Cited on page 9.)

[287] Redish, A. D. and Gordon, J. A., editors (2016). *Computational Psychiatry: New Perspectives on Mental Illness*. Strüngmann Forum Reports. The MIT Press, Cambridge, Massachusetts. (Cited on page 57.)

[288] Reed, E. S. (1996). *Encountering the World: Toward an Ecological Psychology*. Oxford University Press, New York, 1 edition edition. (Cited on page 51.)

[289] Rieke, F., Warland, D., Rob de de Ruyter van Steveninck, and Bialek, W. (1999). *Spikes: Exploring the Neural Code*. A Bradford Book. (Cited on pages 6 and 20.)

[290] Roeth, K., Shao, S., and Gjorgjieva, J. (2020). Efficient population coding depends on stimulus convergence and source of noise. *bioRxiv*, page 2020.06.15.151795. (Cited on page 37.)

[291] Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. (2017). The spatial structure of correlated neuronal variability. *Nature neuroscience*, 20:107–114. (Cited on page 24.)

[292] Rosenblith, W. A., editor (2012). *Sensory Communication*. The MIT Press. (Cited on page 20.)

[293] Rothschild, G., Nelken, I., and Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nat. Neurosci.*, 13(3):353–360. (Cited on pages 24 and 41.)

[294] Ruffini, G., Salvador, R., Tadayon, E., Sanchez-Todo, R., Pascual-Leone, A., and Santarnecchi, E. (2020). Realistic modeling of mesoscopic ephaptic coupling in the human brain. *PLOS Computational Biology*, 16(6):e1007923. (Cited on page 11.)

[295] Safavi, S., Chalk, M., Logothetis, N. K., and Levina, A. (2018a). From optimal efficient coding to criticality. In *Conference on Complex Systems (CCS 2018) Satellite: Complexity from Cells to Consciousness: Free Energy, Integrated Information, and Epsilon Machines*. (Cited on page 37.)

[296] Safavi, S., Chalk, M., Logothetis, N. K., and Levina, A. (2019a). Signatures of criticality in efficient coding networks. In *DPG-Frühjahrstagung 2019*. (Cited on page 37.)

[297] Safavi, S., Dwarakanath, A., Besserve, M., Kapoor, V., Logothetis, N. K., and Panagiotaropoulos, T. I. (2016). A Non-Monotonic Correlation Structure in the Macaque Ventrolateral Prefrontal Cortex. page 53. (Cited on page 41.)

[298] Safavi, S., Dwarakanath, A., Kapoor, V., Werner, J., Hatsopoulos, N. G., Logothetis, N. K., and Panagiotaropoulos, T. I. (2018b). Nonmonotonic spatial structure of interneuronal correlations in prefrontal microcircuits. *PNAS*, page 201802356. (Cited on pages 15, 24, 25, 41, and 91.)

[299] Safavi, S., Kapoor, V., Logothetis, N. K., and Panagiotaropoulos, T. I. (2014). Is the frontal lobe involved in conscious perception? *Front. Psychol.*, 5. (Cited on pages 24, 39, and 91.)

[300] Safavi, S., Logothetis, N. K., and Besserve, M. (2019b). Multivariate coupling estimation between continuous signals and point processes. In *NeurIPS 2019 Workshop: Learning with Temporal Point Processes*. (Cited on page 29.)

[301] Safavi, S., Logothetis, N. K., and Besserve, M. (2020a). From univariate to multivariate coupling between continuous signals and point processes: A mathematical framework. *ArXiv200504034 Q-Bio Stat*. (Cited on pages 15, 32, and 91.)

[302] Safavi, S., Logothetis, N. K., and Besserve, M. (2021a). From Univariate to Multivariate Coupling between Continuous Signals and Point Processes: A Mathematical Framework. *Neural Computation*, pages 1–67. (Cited on page 29.)

[303] Safavi, S., Panagiotaropoulos, F., Kapoor, V., Ramirez-Villegas, J. F., Logothetis, N. K., and Besserve, M. (2020b). Uncovering the organization of neural circuits with generalized phase-locking analysis. In *Computational and Systems Neuroscience Meeting (COSYNE 2020)*, pages 150–151. (Cited on page 31.)

[304] Safavi, S., Panagiotaropoulos, T., Kapoor, T., Logothetis, N. K., and Besserve, M. (2017). Generalized phase locking analysis of electrophysiology data. In *ESI Systems Neuroscience Conference (ESI-SyNC 2017): Principles of Structural and Functional Connectivity*. (Cited on page 31.)

[305] Safavi, S., Panagiotaropoulos, T. I., Kapoor, V., Logothetis, N. K., and Besserve, M. (2018c). Generalized phase locking analysis of electrophysiology data. In *AREADNE 2018: Research in Encoding And Decoding of Neural Ensembles*, page 88. AREADNE Foundation. (Cited on page 31.)

[306] Safavi, S., Panagiotaropoulos, T. I., Kapoor, V., Ramirez-Villegas, J. F., Logothetis, N. K., and Besserve, M. (2020c). Uncovering the Organization of Neural Circuits with Generalized Phase Locking Analysis. *bioRxiv*, page 2020.12.09.413401. (Cited on pages 15, 31, and 91.)

[307] Safavi, S., Panagiotaropoulos, T. I., Kapoor, V., Ramirez-Villegas, J. F., Logothetis, N. K., and Besserve, M. (2021b). Generalized phase locking analysis: A multivariate technique for investigating spike-field coupling. (Cited on pages 29 and 31.)

[308] Sanger, T. D. (2003). Neural population codes. *Curr. Opin. Neurobiol.*, 13:238–49. (Cited on page 7.)

[309] Schilbach, L. (2016). Towards a second-person neuropsychiatry. *Philos Trans R Soc Lond B Biol Sci*, 371(1686). (Cited on page 57.)

[310] Schiller, M., Ben-Shaanan, T. L., and Rolls, A. (2020). Neuronal regulation of immunity: Why, how and where? *Nat. Rev. Immunol.*, pages 1–17. (Cited on page 57.)

[311] Scholtens, L. H. and van den Heuvel, M. P. (2018). Multimodal Connectomics in Psychiatry: Bridging Scales From Micro to Macro. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(9):767–776. (Cited on page 6.)

[312] Schroeder, D. V. (1999). *An Introduction to Thermal Physics*. Pearson, San Francisco, CA, 1st edition edition. (Cited on page 17.)

[313] Schwalm, M., Schmid, F., Wachsmuth, L., Backhaus, H., Kronfeld, A., Aedo Jury, F., Prouvot, P. H., Fois, C., Albers, F., van Alst, T., Faber, C., and Stroh, A. (2017). Cortex-wide BOLD fMRI activity reflects locally-recorded slow oscillation-associated calcium waves. In *eLife*, volume 6. (Cited on page 10.)

[314] Sejnowski, T. J., Churchland, P. S., and Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nat. Neurosci.*, 17(11):1440–1441. (Cited on pages 12 and 60.)

[315] Sethna, J. and Sethna, L. o. A. a. S. S. P. J. P. (2006). *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. OUP Oxford. (Cited on pages 6 and 19.)

[316] Sethna, J. P., Dahmen, K. A., and Myers, C. R. (2001). Crackling noise. *Nature*, 410(6825):242–50. (Cited on page 18.)

[317] Sevgi, M., Diaconescu, A. O., Henco, L., Tittgemeyer, M., and Schilbach, L. (2020). Social Bayes: Using Bayesian Modeling to Study Autistic Trait–Related Differences in Social Cognition. *Biological Psychiatry*, 87(2):185–193. (Cited on page 57.)

[318] Shamir, M. (2014). Emerging principles of population coding: In search for the neural code. *Curr. Opin. Neurobiol.*, 25C:140–148. (Cited on page 7.)

[319] Sheheitli, H. and Jirsa, V. K. (2020). A mathematical model of ephaptic interactions in neuronal fiber pathways: Could there be more than transmission along the tracts? *Netw. Neurosci.*, 4(3):595–610. (Cited on page 11.)

[320] Sheinberg, D. L. and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences of the United States of America*, 94(7):3408–13. (Cited on pages 22, 24, and 44.)

[321] Sherfey, J., Ardid, S., Miller, E. K., Hasselmo, M. E., and Kopell, N. J. (2020). Prefrontal oscillations modulate the propagation of neuronal activity required for working memory. *Neurobiology of Learning and Memory*, 173:107228. (Cited on page 15.)

[322] Sherfey, J. S., Ardid, S., Hass, J., Hasselmo, M. E., and Kopell, N. J. (2018). Flexible resonance in prefrontal networks with strong feedback inhibition. *PLOS Computational Biology*, 14(8):e1006357. (Cited on page 15.)

[323] Shew, W. L. and Plenz, D. (2013). The functional benefits of criticality in the cortex. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 19:88–100. (Cited on page 18.)

[324] Shew, W. L., Yang, H., Yu, S., Roy, R., and Plenz, D. (2011). Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(1):55–63. (Cited on pages 8 and 18.)

[325] Shields, G. S., Spahr, C. M., and Slavich, G. M. (2020). Psychosocial Interventions and Immune System Function: A Systematic Review and Meta-analysis of Randomized Clinical Trials. *JAMA Psychiatry*, 77(10):1031–1043. (Cited on page 57.)

[326] Shpigelman, L., Singer, Y., Paz, R., and Vaadia, E. (2005). Spikernels: Predicting arm movements by embedding population spike rate patterns in inner-product spaces. *Neural computation*, 17(3):671–90. (Cited on page 56.)

[327] Shpiro, A., Moreno-Bote, R., Rubin, N., and Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *Journal of computational neuroscience*, 27(1):37–54. (Cited on page 55.)

[328] Shriki, O. and Yellin, D. (2016). Optimal Information Representation and Criticality in an Adaptive Sensory Recurrent Neuronal Network. *PLOS Computational Biology*, 12(2):e1004698. (Cited on pages 37 and 53.)

[329] Siegelmann, H. T. (2010). Complex systems science and brain dynamics. *Frontiers in computational neuroscience*, 4. (Cited on page 4.)

[330] Siettos, C. and Starke, J. (2016). Multiscale modeling of brain dynamics: From single neurons and networks to mathematical tools. *Wiley Interdiscip Rev Syst Biol Med*, 8(5):438–458. (Cited on page 51.)

[331] Simoncelli, E. P. and Olshausen, B. A. (2001). Natural Image Statistics and Neural Representation. *Annu. Rev. Neurosci.*, 24(1):1193–1216. (Cited on page 20.)

[332] Singer, W. (2009). The Brain, a Complex Self-organizing System. *Eur. Rev.*, 17(2):321–329. (Cited on page 4.)

[333] Singer, W., Gray, C., Engel, A., Konig, P., Artola, A., and Brocher, S. (1990). Formation of cortical cell assemblies. *Cold Spring Harb Sym*, 55:939–52. (Cited on page 54.)

[334] Smeland, O. B., Frei, O., Dale, A. M., and Andreassen, O. A. (2020). The polygenic architecture of schizophrenia — rethinking pathogenesis and nosology. *Nat. Rev. Neurol.*, pages 1–14. (Cited on page 57.)

[335] Smith, M. A. and Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(48):12591–603. (Cited on pages 24 and 41.)

[336] Smith, M. A. and Sommer, M. A. (2013). Spatial and temporal scales of neuronal correlation in visual area V4. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(12):5422–32. (Cited on pages 24 and 41.)

[337] Sporns, O., Tononi, G., and Edelman, G. M. (2000). Connectivity and complexity: The relationship between neuroanatomy and brain dynamics. *Neural Networks*, 13:909–922. (Cited on page 4.)

[338] Srinivasan, R., Russell, D. P., Edelman, G. M., and Tononi, G. (1999). Increased synchronization of neuromagnetic responses during conscious perception. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19(13):5435–48. (Cited on pages 23 and 44.)

[339] Stephan, K. E., Iglesias, S., Heinzle, J., and Diaconescu, A. O. (2015). Translational Perspectives for Computational Neuroimaging. *Neuron*, 87(4):716–732. (Cited on page 51.)

[340] Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., and Bialek, W. (1998). Entropy and Information in Neural Spike Trains. *Phys. Rev. Lett.*, 80(1):197–200. (Cited on page 6.)

[341] Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Curr Opin Neurobiol*, 25:156–63. (Cited on pages 37 and 53.)

[342] Tanaka, T., Kaneko, T., and Aoyagi, T. (2008). Recurrent Infomax Generates Cell Assemblies, Neuronal Avalanches, and Simple Cell-Like Selectivity. *Neural Computation*, 21(4):1038–1067. (Cited on pages 18, 37, and 53.)

[343] Tanigawa, H., Wang, Q., and Fujita, I. (2005). Organization of Horizontal Axons in the Inferior Temporal Cortex and Primary Visual Cortex of the Macaque Monkey. *Cerebral Cortex*, 15(12):1887–1899. (Cited on pages 25 and 41.)

[344] Teixeira, A. L. and Bauer, M. E. (2019). *Immunopsychiatry: A Clinician's Introduction to the Immune Basis of Mental Disorders*. Oxford University Press. (Cited on page 57.)

[345] Tetzlaff, C., Dasgupta, S., Kulvicius, T., and Wörgötter, F. (2015). The Use of Hebbian Cell Assemblies for Nonlinear Computation. *Sci. Rep.*, 5(1):12866. (Cited on page 54.)

[346] Theodoni, P., Panagiotaropoulos, T. I., Kapoor, V., Logothetis, N. K., and Deco, G. (2011). Cortical microcircuit dynamics mediating binocular rivalry: The role of adaptation in inhibition. *Front Hum Neurosci*, 5:145. (Cited on pages 8 and 23.)

[347] Tkacik, G. and Bialek, W. (2016). Information Processing in Living Systems. *Annu Rev Conden Ma P*, 7:89–117. (Cited on pages 8, 18, and 37.)

[348] Tkacik, G., Mora, T., Marre, O., Amodei, D., Palmer, S. E., Berry, M. J., and Bialek, W. (2015). Thermodynamics and signatures of criticality in a network of neurons. *Proceedings of the National Academy of Sciences of the United States of America*. (Cited on page 19.)

[349] Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.*, 5:42. (Cited on pages 6 and 43.)

[350] Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature reviews. Neuroscience*. (Cited on pages 6 and 43.)

[351] Tononi, G. and Koch, C. (2008). The neural correlates of consciousness: An update. *Annals of the New York Academy of Sciences*, 1124(1):239–61. (Cited on pages 21, 22, 23, and 44.)

[352] Touboul, J. and Destexhe, A. (2017). Power-law statistics and universal scaling in the absence of criticality. *Phys. Rev. E*, 95(1):012413. (Cited on page 18.)

[353] Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V. A. F. (2015). No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences*, 19(12):757–770. (Cited on pages 39 and 44.)

[354] Turing, A. M. (1950). I.—Computing Machinery and Intelligence. *Mind*, LIX:433–460. (Cited on page 18.)

[355] van den Heuvel, M. P., Scholtens, L. H., and Kahn, R. S. (2019). Multiscale neuroscience of psychiatric disorders. *Biological Psychiatry*. (Cited on page 6.)

[356] van den Heuvel, M. P. and Sporns, O. (2019). A cross-disorder connectome landscape of brain dysconnectivity. *Nat. Rev. Neurosci.*, 20(7):435. (Cited on page 6.)

[357] van den Heuvel, M. P. and Yeo, B. T. T. (2017). A Spotlight on Bridging Microscale and Macroscale Human Brain Architecture. *Neuron*, 93(6):1248–1251. (Cited on page 6.)

[358] van Steveninck, R. R. d. R., Lewen, G. D., Strong, S. P., Koberle, R., and Bialek, W. (1997). Reproducibility and Variability in Neural Spike Trains. *Science*, 275(5307):1805–1808. (Cited on page 6.)

[359] Vanni, F., Lukovic, M., and Grigolini, P. (2011). Criticality and transmission of information in a swarm of cooperative units. *Physical review letters*, 107:078103. (Cited on pages 8 and 18.)

[360] Varley, T. F., Sporns, O., Puce, A., and Beggs, J. (2020). Differential Effects of Propofol and Ketamine on Critical Brain Dynamics. *bioRxiv*, page 2020.03.27.012070. (Cited on page 17.)

[361] Vattikuti, S., Thangaraj, P., Xie, H. W., Gotts, S. J., Martin, A., and Chow, C. C. (2016). Canonical Cortical Circuit Model Explains Rivalry, Intermittent Rivalry, and Rivalry Memory. *PLOS Computational Biology*, 12(5):e1004903. (Cited on page 55.)

[362] Vinck, M., Battaglia, F. P., Womelsdorf, T., and Pennartz, C. (2012). Improved measures of phase-coupling between spikes and the Local Field Potential. *Journal of computational neuroscience*, 33:53–75. (Cited on page 13.)

[363] Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., and Pennartz, C. M. (2010). The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. *NeuroImage*, 51:112–22. (Cited on page 13.)

[364] Voges, N., Schüz, A., Aertsen, A., and Rotter, S. (2010). A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Progress in Neurobiology*, 92(3):277–292. (Cited on page 41.)

[365] Volgushev, M., Chauvette, S., and Timofeev, I. (2011). Long-range correlation of the membrane potential in neocortical neurons during slow oscillation. *Progress in brain research*, 193:181–99. (Cited on page 10.)

[366] Wang, M., Arteaga, D., and He, B. J. (2013). Brain mechanisms for simple perception and bistable perception. *Proceedings of the National Academy of Sciences of the United States of America*, 110(35):E3350–9. (Cited on pages 23 and 44.)

[367] Wang, R., Lin, P., Liu, M., Wu, Y., Zhou, T., and Zhou, C. (2019). Hierarchical Connectome Modes and Critical State Jointly Maximize Human Brain Functional Diversity. *Phys. Rev. Lett.*, 123(3):038301. (Cited on page 18.)

[368] Wei, X.-X. and Stocker, A. A. (2015). Mutual Information, Fisher Information, and Efficient Coding. *Neural Computation*, 28(2):305–326. (Cited on page 54.)

[369] Werner, G. (2013). Consciousness viewed in the framework of brain phase space dynamics, criticality, and the Renormalization Group. *Chaos Soliton Fract*, 55:3–12. (Cited on page 4.)

[370] Woergoetter, F. and Porr, B. (2008). Reinforcement learning. *Scholarpedia*, 3(3):1448. (Cited on page 53.)

[371] Womelsdorf, T., Ardid, S., Everling, S., and Valiante, T. A. (2014). Burst firing synchronizes prefrontal and anterior cingulate cortex during attentional control. *Current biology : CB*, 24:2613–21. (Cited on page 56.)

[372] Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., and Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, 316:1609–12. (Cited on page 10.)

[373] Yuan, N., Chen, Y., Xia, Y., Dai, J., and Liu, C. (2019). Inflammation-related biomarkers in major psychiatric disorders: A cross-disorder assessment of reproducibility and specificity in 43 meta-analyses. *Transl Psychiatry*, 9(1):1–13. (Cited on page 57.)

[374] Zaldivar, D., Rauch, A., Whittingstall, K., Logothetis, N. K., and Goense, J. (2014). Dopamine-induced dissociation of BOLD and neural activity in macaque visual cortex. *Current biology : CB*, 24:2805–11. (Cited on page 13.)

[375] Zarei, M., Jahed, M., and Daliri, M. R. (2018). Introducing a Comprehensive Framework to Measure Spike-LFP Coupling. *Front. Comput. Neurosci.*, 12. (Cited on page 13.)

[376] Zaretskaya, N. and Narinyan, M. (2014). Introspection, attention or awareness? the role of the frontal lobe in binocular rivalry. *Frontiers in human neuroscience*, 8:527. (Cited on page 39.)

[377] Zbili, M. and Rama, S. (2020). A quick and easy way to estimate entropy and mutual information for neuroscience. *bioRxiv*, page 2020.08.04.236174. (Cited on page 6.)

[378] Zeitler, M., Fries, P., and Gielen, S. (2006). Assessing neuronal coherence with single-unit, multi-unit, and local field potentials. *Neural computation*, 18:2256–81. (Cited on page 13.)

[379] Zeldenrust, F., Gutkin, B., and Denéve, S. (2019). Efficient and robust coding in heterogeneous recurrent networks. *bioRxiv*, page 804864. (Cited on pages 37 and 53.)

[380] Zeraati, R. (2017). Studying criticality and its different measures in neuroscience. Technical report, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany. (Cited on pages 18 and 19.)

[381] Zeraati, R., Priesemann, V., and Levina, A. (2021a). Self-Organization Toward Criticality by Synaptic Plasticity. *Front. Phys.*, 9:103. (Cited on page 18.)

[382] Zeraati, R., Shi, Y.-L., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., Levina, A., and Engel, T. A. (2021b). Attentional modulation of intrinsic timescales in visual cortex and spatial networks. (Cited on page 18.)

[383] Zerbi, V., Floriou-Servou, A., Markicevic, M., Vermeiren, Y., Sturman, O., Privitera, M., von Ziegler, L., Ferrari, K. D., Weber, B., De Deyn, P. P., Wenderoth, N., and Bohacek, J. (2019). Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation. *Neuron*, 103(4):702–718.e5. (Cited on page 10.)

[384] Zhang, M., Kalies, W. D., Kelso, J. A. S., and Tognoli, E. (2020). Topological portraits of multiscale coordination dynamics. *Journal of Neuroscience Methods*, 339:108672. (Cited on page 56.)

Part IV

MANUSCRIPTS

This appendix includes the PDF of all the published papers, preprints and in-preparation manuscripts. They appear as they appeared in Part ii, with the following order:

1. Safavi et al. [302, Neural Computation 2021]
2. Safavi et al. [306, bioRxiv 2020]
3. Besserve et al.; preliminary manuscript is available in the appendix, (Paper 3)
4. Safavi et al.; preliminary manuscript is available in the appendix, (Paper 4)
5. Safavi et al. [299, Front. Psychol. 2014]
6. Safavi et al. [298, PNAS 2018]
7. Kapoor et al. [173, bioRxiv 2020]
8. Dwarakanath et al. [114, bioRxiv 2020]

LETTER **Communicated by Debashis Paul**

From Univariate to Multivariate Coupling Between Continuous Signals and Point Processes: A Mathematical Framework

Shervin Safavi
shervin.safavi@tuebingen.mpg.de
MPI for Biological Cybernetics, and IMPRS for Cognitive and Systems Neuroscience, University of Tübingen, 72076 Tübingen, Germany

Nikos K. Logothetis
nikos.logothetis@tuebingen.mpg.de
MPI for Biological Cybernetics, 72076 Tübingen, Germany; International Center for Primate Brain Research, Songjiang, Shanghai 200031, China; and University of Manchester, Manchester M13 9PL, U.K.

Michel Besserve
michel.besserve@tuebingen.mpg.de
MPI for Biological Cybernetics and MPI for Intelligent Systems, 72076 Tübingen, Germany

Time series data sets often contain heterogeneous signals, composed of both continuously changing quantities and discretely occurring events. The coupling between these measurements may provide insights into key underlying mechanisms of the systems under study. To better extract this information, we investigate the asymptotic statistical properties of coupling measures between continuous signals and point processes. We first introduce martingale stochastic integration theory as a mathematical model for a family of statistical quantities that include the phase locking value, a classical coupling measure to characterize complex dynamics. Based on the martingale central limit theorem, we can then derive the asymptotic gaussian distribution of estimates of such coupling measure that can be exploited for statistical testing. Second, based on multivariate extensions of this result and random matrix theory, we establish a principled way to analyze the low-rank coupling between a large number of point processes and continuous signals. For a null hypothesis of no coupling, we establish sufficient conditions for the empirical distribution of squared singular values of the matrix to converge, as the number of measured signals increases, to the well-known Marchenko-Pastur (MP) law, and the largest squared singular value converges to the upper end of the MP support. This justifies a simple thresholding approach to assess the significance of multivariate coupling. Finally, we illustrate with

Neural Computation **33, 1751–1817** (2021)

https://doi.org/10.1162/neco_a_01389

simulations the relevance of our univariate and multivariate results in the context of neural time series, addressing how to reliably quantify the interplay between multichannel local field potential signals and the spiking activity of a large population of neurons.

1 Introduction

The observation of highly multivariate temporal point processes, corresponding to the activity of a large number of individuals or units, is pervasive in many applications (for example, neurons in brain networks; Johnson, 1996) and members in social networks (Dai, Wang, Trivedi, & Song, 2016; De, Valera, Ganguly, Bhattacharya, & Rodriguez, 2016). As the number of observed events per unit may remain small, inferring the underlying dynamical properties of the studied system from such observations is challenging. However, in many cases, it is possible to observe continuous signals whose coupling with the events can offer key insights.

In neuroscience, this is the case of the extracellular electrical field, which provides information complementary to spiking activity. Local field potentials (LFP) are mesoscopic (Liljenstroem, 2012) signals resulting from the superposition of the electric potentials generated by ionic currents flowing across the membranes of the cells located close to the tip of recording electrodes. The LFP reflects neural cooperation due to the anisotropic cytoarchitecture of most brain regions, allowing the summation of the extracellular currents resulting from the activity of neighboring cells. As such, a number of subthreshold integrative processes (i.e., modifying the neurons' internal state without necessarily triggering spikes) contribute to the LFP signal (Buzsaki, Anastassiou, & Koch, 2012; Buzsaki, Logothetis, & Singer, 2013; Einevoll, Kayser, Logothetis, & Panzeri, 2013; Pesaran et al., 2018; Herreras, 2016).

Reliably quantifying the coupling between activities of individual units (e.g., spikes generated by individual neurons) in a circuit and the aggregated measures (such as the LFP) may provide insights into underlying network mechanisms, as illustrated in the electrophysiology literature. At the single neuron level, the relationship of spiking activity to subthreshold activity has broad implications for the underlying cellular and network mechanisms at play. For instance, it has been suggested that synaptic plasticity triggers changes in the coupling between spikes and LFPs (Grosmark, Mizuseki, Pastalkova, Diba, & Buzsáki, 2012; Grosmark & Buzsáki, 2016). Regarding the putative functional role of such observed couplings, it has been hypothesized to support cognitive functions such as attention. Such coordination by oscillations hypothesis proposes that network oscillations modulate differentially the excitability of several target populations, such that a sender population can emit messages during the window of time for which a selected target is active, while unselected targets are silenced (Fries, 2005, 2015; Womelsdorf et al., 2007).

In the case of two continuous signals, coupling measures such as coherence and phase locking value (PLV) (Rosenblum, Pikovsky, Kurths, Schäfer, & Tass, 2001; Pereda, Quiroga, & Bhattacharya, 2005) are widely used, and their statistical properties have been investigated, in particular in the stationary gaussian case (Brillinger, 1981; Aydore, Pantazis, & Leahy, 2013). In a similar way, PLV (Ashida, Wagner, & Carr, 2010) and spike-field coherence (SFC) (Mitra, 2007) can measure spike-LFP coupling (see among others: Vinck, Battaglia, Womelsdorf, & Pennartz, 2012; Vinck, van Wingerden, Womelsdorf, Fries, & Pennartz, 2010; Jiang, Bahramisharif, van Gerven, & Jensen, 2015; Zarei, Jahed, & Daliri, 2018; Li, Cui, & Li, 2016) and are broadly used to makes sense of the role played by neurons in coordinated network activity (Buzsaki & Schomburg, 2015). There are notable contributions investigating potential biases of those measures when both point processes and continuous signals are involved (Lepage, Kramer, & Eden, 2011; Kovach, 2017). However, two issues relevant for practical applications remain: (1) the effect of the intrinsic variability of spike occurrence on key statistical properties of the estimates, such as the variance, have not yet been thoroughly described, and (2) how to extend the rigorous statistical analysis of spike-field coupling in the context of the highly multivariate signals available with modern recording techniques remains largely unaddressed.

We address these two issues by using continuous time martingale theory (see Liptser & Shiryaev, 2013a), the related concept of stochastic integration (see Protter, 2005) and random matrix theory (Anderson, Guionnet, & Zeitouni, 2010; Capitaine & Donati-Martin, 2016). The martingale central limit theorem (CLT) allows us to derive analytically the asymptotic gaussian distribution of a general family of coupling measures that can be expressed as stochastic integrals. We exploit this general result to show that the classical univariate PLV estimator is also asymptotically normally distributed and provide the analytical expression for its mean and variance. Furthermore, we study potential sources of bias for the commonly used von Mises coupling model (Ashida et al., 2010). We then go beyond univariate coupling measures and analyze the statistical properties of a family of multivariate coupling measures taking the form of a matrix with stochastic integral coefficients. We characterize the jointly gaussian asymptotic distribution of matrix coefficients, and exploit random matrix theory (RMT) principles to show that after appropriate normalization, the spectral distribution of such large matrices under the null hypothesis (of absence of coupling), follows approximately the Marchenko-Pastur (MP) law (Marchenko & Pastur, 1967), while the magnitude of the largest singular value converges to a fixed value whose simple analytic expression depends only on the shape of the matrix. We finally show how this result provides a fast and principled procedure to detect significant singular values of the coupling matrix, reflecting an actual dependency between the underlying signals. In the appendixes, we included detailed proofs and background material on RMT

and stochastic integration, such that nonexpert readers can further apply these tools in neuroscience.

2 Background

2.1 Spike-Field Coupling in Neuroscience. Although our results are relevant to a broad range of applications within and beyond neuroscience, we will use the estimation of spike-LFP coupling introduced above as the guiding example of this letter. Spikes convey information communicated between individual neurons. This information is believed to be encoded in the occurrence times of successive spike events, which are typically modeled with point processes—for example, Poisson (Softky & Koch, 1993) or Hawkes process (Truccolo, 2016; Krumin, Reutsky, & Shoham, 2010).

While oscillatory dynamics is ubiquitous in the brain and instrumental to its coordinated activity (Buzsaki, 2006; Buzsaki et al., 2013; Peterson & Voytek, 2018), it is often challenging to uncover based solely on the sparse spiking activity of recorded neurons. On the other hand, LFPs often exhibit oscillatory components that can be isolated with signal processing tools (typically bandpass filtering or template matching), such that pairing the temporal information from LFPs and spiking activity can help extract reliable markers of neural coordination.

An example of a coupling measure achieving such pairing is the phase locking value (PLV). Given, on the one hand, event (spike) times $\{t_j\}$ where $j \in \{1, 2, \ldots, N\}$ (with N the number of spikes in the spike train), and on the other hand, $\phi(t)$ the time-varying phase of an oscillatory continuous signal, which is typically a bandpassed filtered LFP, phase locking between these two signals is estimated by the complex number

$$\widehat{\mathrm{PLV}} = \frac{1}{N}\sum_{j=1}^{N} e^{i\phi(t_j)}, \text{ with } i^2 = -1. \tag{2.1}$$

We use a "hat" notation to reflect that this quantity is empirical: indeed, even if we assume a fixed ϕ, the PLV depends on the specific values of event times t_j. In this work, we assume these points are drawn from a Poisson process, with a possibly time-varying rate (inhomogeneous Poisson process), such that we can define a population statistics that is a function of the point process population distribution instead of its empirical counterpart. We then address under which conditions the empirical PLV reflects a true coupling between the rate of the underlying point process and ϕ.

2.2 Counting Process Martingales. We use a continuous time framework leading to powerful results based on concise deterministic and stochastic integral expressions, which can trivially be approximated using discrete time signals in practice. A (continuous time) stochastic process

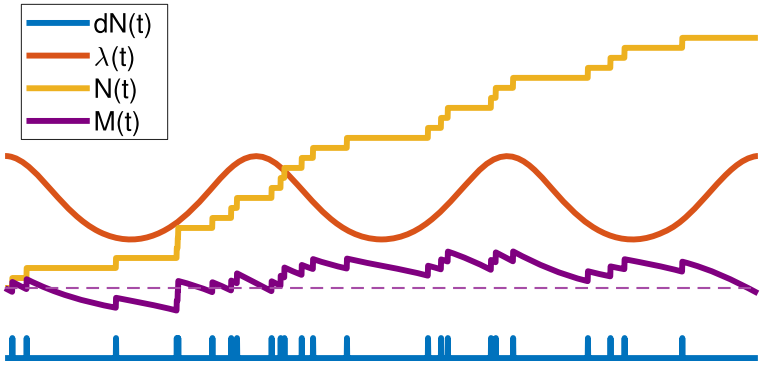


Figure 1: Doob-Meyer decomposition for an example inhomogeneous Poisson process with oscillatory of rate $\lambda(t)$ of frequency $f = 1$ Hz; average firing rate $\lambda_0 = 5$ Hz (dashed line indicates the reference 0). See section 3.3 for details of the simulation.

$M = \{M(t); t \in [0, \tau]\}$ is a zero-mean martingale relative[1] to the filtration $\{\mathcal{F}_t\}$ (which represents the past information accumulated up to time t) if (1) $M(0) = 0$, (2) it is adapted to $\{\mathcal{F}_t\}$ (informally the law of M up to time t "uses" only past information up to t), and (3) it satisfies the martingale property:

$$E\left[M(t)|\mathcal{F}_s\right] = M(s), \quad \text{for all} \quad t > s. \tag{2.2}$$

Consider now a (univariate) counting process $\left\{(N(t), \mathcal{F}_t);\ t \geq 0\right\}$, counting the number of events that occurred up to time t, adapted to filtration $\{\mathcal{F}_t\}$ (Aalen, Borgan, & Gjessing, 2008, chap. 2). Under mild assumptions, it has a Doob-Meyer decomposition,

$$N(t) = M(t) + \int_0^t \lambda(t)dt, \tag{2.3}$$

where $\lambda(t)$ is a predictable process with respect to $\{\mathcal{F}_t\}$ called the intensity function and $M(t)$ is a martingale, called the compensated counting process. Figure 1 shows an illustration of this decomposition for a Poisson process with sinusoidal intensity.

Consider now an empirical coupling measure c between a (real or complex) predictable process $x(t)$ and $N(t)$ observed during time interval $[0,\ T]$, which takes the form of the stochastic integral (see Protter, 2005),

$$\widehat{c} = \sum_{t_k < T} x(t_k) = \int_0^T x(t)dN(t), \tag{2.4}$$

[1] Any martingale in this paper is zero-mean.

where $\{t_k\}$ denote the jump times of the counting process (note that the PLV defined in equation 2.1 is a normalized version of such coupling). The empirical coupling measure, $\widehat{c}$, can then be decomposed as

$$\widehat{c} = \int_0^T x(t)\lambda(t)dt + \int_0^T x(t)dM(t). \tag{2.5}$$

Interestingly, it can be shown that the second integral on the right-hand side is also a martingale (see Liptser & Shiryaev, 2013b, theorem 18.7).

In order to keep our results concise, we assume the following deterministic setting in the remainder of this letter (see section 5 for potential extensions).

Assumption 1. Assume the intensity function, $\lambda(t) = \lambda(t|\mathcal{F}_t)$ of $N(t)$ and the signal $x(t)$ are deterministic bounded left-continuous and adapted to $\mathcal{F}_t$ over $[0,\ T]$.

Note this entails that $N(t)$ is a (possibly inhomogeneous) Poisson process (Liptser & Shiryaev, 2013b, theorem 18.10). Under assumption 1, the terms of equation 2.5 separate the deterministic part from the (zero-mean) random fluctuations of the measure that are integrally due to the martingale term. Using martingale properties, the statistics of the coupling measure are[2]

$$c^* \triangleq \mathbb{E}[\widehat{c}] = \int_0^T x(t)\lambda(t)dt \quad \text{and}$$

$$\mathrm{Var}[\widehat{c}] = \mathbb{E}\left[|\widehat{c} - c^*|^2\right] = \int_0^T \left|x^2(t)\right| \lambda(t)dt. \tag{2.6}$$

In case $x(t)$ integrates to zero, the expected coupling c^* thus reflects the covariation across time between $x(t)$ and the intensity of the point process up to random fluctuations.

2.3 Random Matrix Theory. As data sets get increasingly high dimensional, it becomes important to replace the above univariate measure $\widehat{c}$ by a quantity that summarizes the coupling between a large number of units and continuous signals. This extension leads to assessing the spectral properties of a coupling matrix $\widehat{\mathbf{C}}$ that gathers all pairwise measurements. However, such task is nontrivial due to the martingale fluctuations affecting $\widehat{\mathbf{C}}$, leading to spurious nonzero coupling coefficients, and can also hide the deterministic structure of the matrix associated with significant coupling.

Random matrix theory allows investigating the spectral properties of some matrices in noisy settings by studying their asymptotic spectral properties as dimensions grow to infinity. Any $(p \times p)$ complex Hermitian or

[2] See section B.1.1 for more details.

real symmetric matrix $\boldsymbol{M}$ has a set of p real eigenvalues $\{\ell_k\}$ (where we put several times the same eigenvalue in the set according to its multiplicity). One classically studied quantity is the empirical spectral distribution (ESD) (or empirical eigenvalue distribution, see Mingo & Speicher, 2017 and Anderson et al., 2010) of the set of all eigenvalues $\{\ell_k\}$. ESD indistinctly refers (with a slight abuse of language), to either the probability measure (also called *spectral measure* in our case),

$$\mu_{\boldsymbol{M}}(t) = \frac{1}{p}(\delta_{\ell_1}(t) + \cdots + \delta_{\ell_p}(t)),\ t \in \mathbb{R},$$

where δ_{ℓ_k} is the dirac measure with unit mass in ℓ_k, or to its associated cumulative distribution:

$$F_{\boldsymbol{M}}(t) = \int_{-\infty}^{t} d\mu_M(s).$$

Seminal works by Wigner (1955, 1958), Marchenko and Pastur (1967), and many others have established the convergence of the ESD of large random matrix ensembles (see section B.2 for the precise notions of convergence). In particular, for a sequence of matrices $\{\boldsymbol{X}_n\}_{n>0}$ of dimension $p \times n$ such that $\frac{p}{n} \underset{n\to+\infty}{\to} \alpha \leq 1$, with coefficients sampled independently and identically distributed (i.i.d.) from a (possibly complex) standard Normal distribution, the ESD of the Wishart matrix $\mathbf{S}_n = \frac{1}{n}\boldsymbol{X}_n\boldsymbol{X}_n^H$ (where $.^H$ indicates the transposed complex conjugate) converges to the Marchenko-Pastur (MP) law $\mu_{MP}(x)$ (Marchenko & Pastur, 1967) with density

$$\frac{d\mu_{MP}}{dx}(x) = \begin{cases} \frac{1}{2\pi\alpha x}\sqrt{(b-x)(x-a)}, & a \leq x \leq b, \\ 0, & \text{otherwise}, \end{cases} \tag{2.7}$$

with $a = (1-\sqrt{\alpha})^2$ and $b = (1+\sqrt{\alpha})^2$. Additionally, the smallest and largest eigenvectors converge to a and b, respectively. Importantly, these convergences also hold in the case $\alpha > 1$, but equation 2.7 is modified to account for the rank deficiency of the Wishart matrix, imposing $p - n$ zero eigenvalues in the spectrum (see section B.3.1 for details).

Notably, recent developments in the field of random matrix theory extend the classic results that were only valid for independent coefficients (uncorrelated Wishart matrices) to various forms of dependencies between coefficients. For instance, El Karoui (2007, 2008) showed that the ESD and the distribution of the largest eigenvalue for a sequence of matrices $\{\boldsymbol{X}_n\}_{n>0}$ with general covariance matrices (not necessarily with identity covariance matrix) follow similar laws and Banna, Merlevède, and Peligrad (2015) investigate the case of symmetric random matrices with correlated entries.

Furthermore, the behavior of high-dimensional autocovariance matrices in the context of discrete time stochastic processes is discussed in Liu, Aue, and Paul (2015) and Bhattacharjee and Bose (2016). Applications of this framework have also been considerably extended including global finance (Namaki et al., 2020) and various aspects of machine learning and signal processing such as shallow (Louart, Liao, & Couillet, 2018) and deep (Pennington & Bahri, 2017; Pennington & Worah, 2019) neural networks, denoising (Bun, Bouchaud, & Potters, 2017) and dimensionality reduction (Johnstone & Onatski, 2020).

In this study, we show that the martingale fluctuations of the coupling matrices also cause spectral convergence to the MP law in the absence of actual coupling between the signals. Recent results on the low-rank perturbation (Capitaine & Donati-Martin, 2016; Loubaton & Vallet, 2011; Benaych-Georges & Nadakuditi, 2012) of random matrices suggest this convergence can be exploited to further assess the significance of the largest eigenvalues of the coupling matrix with respect to the null hypothesis that they only reflect random fluctuations.

3 Assessment of Univariate Coupling

3.1 Mathematical Formulation. We consider the setting of $K \geq 1$ independent trials of measurements on $[0,\ T]$ available to estimate the coupling statistics by the trial average

$$\widehat{\mathrm{c}}_K = \frac{1}{K}\sum_{k=1}^{K}\int_0^T x(t)dN^{(k)}(t),$$

where $\{N^{(k)}\}$ are K independent copies of the process $N(t)$, associated with each trial. As this letter focuses on the statistical properties induced by the intrinsic variability of point process realizations, we assumed above that the continuous signal does not change across trials. However, including some forms of variability across trials, such as random time shifts affecting all processes in the same way, would not affect the results, barring additional technical details.

We exploit a central limit theorem (CLT) for martingales to show the residual variability (difference between the empirically estimated $\widehat{\mathrm{c}}_K$ and the expected coupling c^* of equation 2.6) is asymptotically normally distributed. We formally state it in theorem 1.

Theorem 1. *Assume* $(\mathcal{F}_t, x(t), \lambda(t))$ *satisfy assumption 1, and* $x(t)$ *real-valued. Then,*

$$\mathbb{E}[\widehat{\mathrm{c}}_K] \triangleq \mathrm{c}^* = \int_0^T x(t)\lambda(t)dt \quad \textit{and} \quad \mathrm{Var}[\widehat{\mathrm{c}}_K] = \frac{1}{K}\int_0^T x^2(t)\lambda(t)dt.$$

Moreover, as the number of trials increases, fluctuations converge in distribution:

$$\sqrt{K}\,(\widehat{\mathrm{c}}_K - \mathrm{c}^*) \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(0, \int_0^T x^2(t)\lambda(t)dt\right).$$

Sketch of the proof. We rely on the decomposition of equation 2.5. As described in section B.1.1, the martingale property is preserved by the stochastic integral term and allows us to exploit a martingale CLT to prove convergence to a gaussian distribution. □

The case of $x(t)$ complex-valued can be dealt with by distinguishing the real and imaginary parts of the signal, as is done in the proofs of the following corollaries. We can exploit theorem 1 to derive the asymptotic properties of the PLV introduced in section 2.1. For that, we adapt the empirical estimate of equation 2.1 to the K trials setting introduced above and define

$$\widehat{\mathrm{PLV}}_K = \frac{1}{\sum_{k=1}^K N_k} \sum_{k=1}^K \sum_{j=1}^{N_k} e^{i\phi(t_j^k)}, \tag{3.1}$$

where N_k is the number of events observed during trial k and $\{t_j^k\}$ is the collection of the time stamps of these events. The specificity of this multitrial estimate is to use a single normalization constant corresponding to the total number of events pooled across trials.[3] For this estimate, we get the following result.

Corollary 1. *Assume* $(\mathcal{F}_t, x(t) = e^{i\phi(t)}, \lambda(t))$ *satisfy assumption 1, where* ϕ *is real-valued and stands for the phase of the signal* x*. Then the expectation of the PLV statistics* $\widehat{\mathrm{PLV}}_K$ *estimated from* K *trials of measurements on* $[0, T]$ *tends to the limit*

$$\mathrm{PLV}^* = \int_0^T e^{\mathbf{i}\phi(t)}\lambda(t)dt/\Lambda(T), \quad \textit{with} \quad \Lambda(T) = \int_0^T \lambda(t)dt. \tag{3.2}$$

Moreover, as $K \to +\infty$*, the residual,*

$$\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right), \tag{3.3}$$

[3] This allows the normalization factor to converge to a deterministic quantity as $K \to +\infty$ equation 2.1.

converges in distribution to a zero-mean complex gaussian variable Z (i.e., the joint distribution of real and imaginary parts is gaussian), such that

$$\mathrm{Cov}\left(\begin{bmatrix} Re\{Z\} \\ Im\{Z\} \end{bmatrix}\right) = \frac{1}{\Lambda(T)^2}\int_0^T M(t)\lambda(t)dt,$$

$$where\, M(t) = \begin{bmatrix} \cos^2(\phi(t)) & \sin(2\phi(t))/2 \\ \sin(2\phi(t))/2 & \sin^2(\phi(t)) \end{bmatrix}.$$

Sketch of the proof. This relies on applying theorem 1 to the real and imaginary parts of $e^{i\phi(t)}$. In addition, the coupling between both quantities is taken into account by replacing the variance of univariate quantities $\widetilde{V}(t)$ in theorem 1 by a covariance matrix that can be assessed with martingale results given in section B.1.1. □

Remark 1. For the simple case of a T/k-periodic sinusoidal signal (k integer), such that $\phi(t) = 2\pi kt/T$, and a sinusoidal modulation of the intensity with phase shift φ_0 and modulation amplitude $\varkappa$ such that

$$\lambda(t) = \lambda_0\left(1 + \varkappa\cos\left(\phi(t) - \varphi_0\right)\right),\ \lambda_0 > 0,\ 0 \le \varkappa \le 1,$$

we get easily with trigonometric identities that $\mathrm{PLV}^* = \frac{1}{2}\varkappa e^{i\varphi_0}$ and the residual of equation 3.3 converges to an isotropic complex gaussian of total variance[4] $\frac{1}{\lambda_0 T}$ such that the coupling strength $\varkappa$ affects the mean but not the variance of the PLV estimate.

Also, it is easy to see that if $\lambda(t)$ is modulated by a sine wave at a different integer multiple $m \neq k$ of the fundamental frequency $1/T$, such that

$$\lambda(t) = \lambda_0 + \varkappa\cos\left(2\pi mt/T - \varphi_0\right),$$

the PLV^* vanishes and the residual's variance remains the same. These properties make PLV straighforward to interpret and test for sinusoidal coupling with a carefully chosen observation duration T. Assumption 3 and corollary 5, in appendix C, provide formal statements of this remark.

We can use corollary 1 to predict the statistics of PLV estimates for other models of phase-locked spike trains. A classical model uses the von Mises distribution (also known as circular normal distribution) with parameter $\kappa \ge 0$ to model the concentration of spiking probability around a specified locking phase ϕ_0 (for more details, see Ashida et al., 2010). The original model uses a purely sinusoidal time series by assuming a linearly

[4] The sum of the variances of real and imaginary parts.

increasing phase $\phi(t) = 2\pi ft$, where f is the modulating frequency, to derive the intensity of an inhomogeneous Poisson spike train,

$$\lambda(t) = \lambda_0 \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right). \tag{3.4}$$

resulting in an analytical expression for the asymptotic complex-valued PLV,

$$\mathrm{PLV}^* = e^{\mathbf{i}\varphi_0} \frac{\int_0^\pi \cos(\theta) \exp(\kappa \cos(\theta)) d\theta}{\int_0^\pi \exp(\kappa \cos(\theta)) d\theta} = e^{\mathbf{i}\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)},$$

with the I_k's denoting the modified Bessel functions of the first kind for k integer (see Abramowitz & Stegun, 1972, p. 376):

$$I_k(\kappa) = \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta.$$

Compared to the sinusoidal coupling described in remark 1, whose PLV magnitude can reach at most $1/2$, this model can achieve arbitrarily large PLV, which might explain why it is more frequently used in applications.

Corollaries 2 and 3 derive the asymptotic covariance of the variability of the PLV estimate around this theoretical value (which is novel to the best of our knowledge). Furthermore, the results are derived in a more general model setting accounting for "biases"[5] due to nonlinear phase increases $\phi(t)$ and observation intervals that are not multiples of the modulating oscillation period. It should be noted that the mentioned biases are inherent in the estimator's definition. They happen independent of additional biases originating from the phase estimation procedure (e.g., phase extraction via Hilbert transform; see Kovach, 2017).

We thus assume a coupling, parameterized by κ between a possibly nonlinearly increasing phase $\phi(t)$ and a point process with intensity

$$\lambda(t) = \lambda_0 \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) \frac{d\phi}{dt}(t). \tag{3.5}$$

Note that for linearly increasing phases, this coupling amounts to the classical von Mises model of equation 3.4. The additional factor $\frac{d\phi}{dt}(t)$ allows preserving the analytical expression of PLV statistics even for nonlinearly increasing phases, providing a novel generalization of the von Mises model (see corollary 4 in appendix C for a simplified version of corollary 2, assuming a linearly increasing phase $\phi(t) = 2\pi ft$ with frequency f).

[5] They are biases in the sense that one would expect a coupling measure to vanish if there is no coupling in the data generating procedure.

Corollary 2. *Under the assumptions of corollary 1, assume additionally that $\phi(t)$ is continuous, strictly increasing, and piece-wise differentiable on $[0,\ T]$ and the intensity of the point-process is given by equation 3.5 for a given $\kappa \geq 0$, then the expectation of the multitrial PLV estimate converges (for $K \to +\infty$) to*

$$\mathrm{PLV}^* = \frac{\int_{\phi(0)}^{\phi(T)} e^{\mathbf{i}\theta} \exp(\kappa \cos(\theta - \varphi_0))d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa \cos(\theta - \varphi_0))d\theta}. \tag{3.6}$$

If in addition $[0,\ T]$ corresponds to an integer number of periods of the oscillation,

$$\mathrm{PLV}^* = e^{\mathbf{i}\varphi_0} \frac{\int_0^{\pi} \cos(\theta) \exp(\kappa \cos(\theta))d\theta}{\int_0^{\pi} \exp(\kappa \cos(\theta))d\theta} = e^{\mathbf{i}\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}, \tag{3.7}$$

and the scaled residual $\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right)$ converges to a zero mean complex gaussian Z with the following covariance:

$$\begin{aligned} &\mathrm{Cov}\begin{bmatrix} Re\{Ze^{-i\varphi_0}\} \\ Im\{Ze^{-i\varphi_0}\} \end{bmatrix} \\ &= \frac{1}{2\lambda_0(\phi(T) - \phi(0))I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}. \end{aligned} \tag{3.8}$$

Sketch of the proof. This is based on plugging the intensity function $\lambda(t)$ of equation 3.5 in corollary 1. Using change of variable in the integrals ($\phi(t)$ to θ) and exploiting the symmetries of the functions, the integrals in the analytical expressions of the expectation and covariance turn into modified Bessel functions I_k for k integer. □

The above result has important consequences for the assessment of PLV from data. In particular, it exhibits key experimental requirements for PLV estimates to match the classical Bessel functions expression of equation 3.7: (1) evaluate PLV on an integer number of periods (this is critical for trials with short duration) and (2) take into account the fluctuations of the rate of increase of the phase $\phi(t)$ across the oscillation period. This second point is critical in applications where the phase is inferred from signals (such as LFPs) through the Hilbert transform, as nonlinearities of the underlying phenomena may lead to nonsinusoidal oscillations, with periodic fluctuations of the time derivative of the phase $\phi'(t)$. To further emphasize the consequences of this aspect, we also derive the asymptotic distribution of PLV for a homogeneous Poisson process that corresponds to the special case $\kappa = 0$ of the classical von Mises coupling of equation 3.5. Although there is no actual coupling between events and the continuous signal in such a

case,[6] the nonlinear phase increase leads asymptotically (for K large) to a nonvanishing PLV estimate and to false detection of coupling.

Corollary 3. *Under the assumptions of corollary 1, we assume additionally that the point process is homogeneous Poisson with rate λ_0 and that $\phi(t)$ is strictly increasing (almost everywhere) and differentiable on $[0, T]$. Let $\theta \mapsto \tau(\theta)$ be its inverse function (such that $\tau(\phi(t)) = t$). Then the expectation of $\widehat{PLV}_K$ converges (for $K \to +\infty$) to*

$$\mathrm{PLV}^* = \frac{\int_{\phi(0)}^{\phi(T)} e^{\mathbf{i}\theta}\tau'(\theta)d\theta}{\phi(T) - \phi(0)}, \tag{3.9}$$

and the scaled residual,

$$Z = \sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right),$$

converges to a zero mean complex gaussian:

$$\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right) \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathrm{Cov}(Z)\right),$$

with the following covariance:

$$\mathrm{Cov}(Z) = \frac{1}{\lambda_0 T^2}\int_{\phi(0)}^{\phi(T)} \begin{bmatrix} \cos^2(\theta) & \sin(2\theta)/2 \\ \sin(2\theta)/2 & \sin^2(\theta) \end{bmatrix} \tau'(\theta)d\theta.$$

Sketch of the proof. The result stems from using the intensity function λ_0 in corollary 1 and then using change of variable in the integrals and exploiting the symmetries of the functions. □

This corollary will be further illustrated in the next paragraphs.

3.2 Application to Bias Assessment. Corollary 3 predicts scenarios where in the absence of modulation of spiking activity (having a constant intensity function $\lambda(t) = \lambda_0$), the expectation of the PLV estimates remains far from zero even when the number of trials is large, that is, the coupling between a homogeneous point process and a continuous oscillatory signal would appear significant and reflect a form of bias. Corollary 3 allows computing this bias and therefore correcting it.

[6] In the sense that we can generate the homogeneous spike train and the oscillation without parametric models that do not share any information.

One such case is when the observation interval is not an integer number of oscillation periods. To demonstrate this analytically, we can start from the PLV expectation with the constant intensity λ_0,

$$\mathrm{PLV}^* = \frac{\int_0^T e^{i\phi(t)}\lambda(t)dt}{\int_0^T \lambda(t)dt} = \frac{\lambda_0 \int_0^T e^{i\phi(t)}dt}{\lambda_0 \int_0^T dt} = \frac{1}{T}\int_0^T e^{i\phi(t)}dt. \tag{3.10}$$

Furthermore, we assume $\phi(t)$ has linear phase (assumption 3): $\phi(t) = 2\pi ft$, where f is the frequency of oscillation of the continuous signal. We then get

$$\mathrm{PLV}^* = \frac{1}{T}\int_0^T e^{i2\pi ft}dt = \frac{1}{2\pi\gamma_T i}\left(e^{2\pi\gamma_T i} - 1\right), \tag{3.11}$$

where $\gamma_T = Tf$ is the ratio of the length of the time series (T) to the period of oscillation $\frac{1}{f}$. As is noticeable in equation 3.11, the coupling measure PLV* is not zero when γ_T is not an integer number. Notably, this bias affects both the magnitude and the phase of the PLV* estimate.

Furthermore, even using an observation interval covering an integer number of periods, nonlinear increases in phase may lead to a nonvanishing PLV. This can be demonstrated with a simple example. Again, we can start from the original definition of PLV expectation, equation 3.2, but now we do not assume the linearity of the phase. As introduced in corollary 3, let $\theta \mapsto \tau(\theta)$ be the inverse of $\phi(t)$, and let us use equation 3.9 to compute the PLV*. Taking a sinusoidal modulation over the oscillation period, $\tau(\theta) = \theta + \epsilon\sin(\theta)$ with $|\epsilon| < 1$,[7] we get a nonvanishing asymptotic expected PLV:

$$\mathrm{PLV}^* = \frac{1}{2\pi}\int_0^{2\pi} e^{\mathrm{i}\theta}(1 + \epsilon\cos(\theta))d\theta = \epsilon\int_0^{\pi} e^{\mathrm{i}\theta}\cos(\theta)d\theta = \epsilon/2 \neq 0, \text{ if } \epsilon \neq 0.$$

Our theoretical framework can be used for developing methods to correct such biases. In the linear phase setting, bias can be avoided simply by using an integer number of periods for coupling estimation. In the case of a nonlinear phase evolution of the continuous signal, we can use the theoretical phase (if available) or its empirical estimate to evaluate PLV* under constant spike intensity assumptions with equation 3.9 and subtract this quantity to the estimated PLV. For resolving issues that arise due to the nonlinearity of the estimated phase, specialized methods have been suggested. For instance, Hurtado, Rubchinsky, and Sigvardt (2004) dealt with phase jumps (a particular form of nonlinearity) by interpolating the signal from the available data before and after the sudden change and Cole and

[7] To guarantee the phase to be strictly increasing.

Voytek (2019) introduced a cycle-by-cycle method for analyzing oscillatory dynamics. In this method, they consider a linear phase for each detected cycle of oscillation. Therefore, with this linear choice of phase, one can avoid the spurious coupling that can appear due to phase nonlinearities. Based on our framework, theoretically motivated methods that are not relying on the linearization of the phase can be developed.

3.3 Simulations. We demonstrate the outcome of our theoretical results using simulated phase-locked spike trains (similar to what has been introduced in corollaries 2 and 4) and sinusoidal oscillations. For generating phase-locked spike trains, we adopt the method introduced in Ashida et al. (2010). As the model has already been described elsewhere, we restrict ourselves to a brief explanation.

To generate phase-locked or periodic spike trains based on the classical von Mises model with rate $\lambda(t)$ as introduced in equation 3.4, we use a purely sinusoidal continuous signal $x(t)$ with linearly increasing phase $\phi(t) = 2\pi ft$, with $f = 1$ Hz and various coupling strength (κ) (see appendix E for lists of parameters used for each figure). Based on this simulation we perform two numerical experiments to demonstrate the practical relevance of our (asymptotic) theoretical results.

3.3.1 Experiment 1. In order to demonstrate the validity of corollaries 2 and 4, in Figure 2 we show the empirical distribution of the normalized residual of the PLV estimate and compare it to its asymptotic theoretical distribution. We simulate two cases, one with homogeneous Poisson spike trains ($\kappa = 0$) and one with phase-locked spike trains ($\kappa = 0.5$) with Poisson statistics. In both cases, we observe the agreement between theory and simulation, as the joint distribution of real and imaginary part approaches an isotropic gaussian. The slightly non-gaussian shape of the real part histogram for $\kappa = 0.5$ suggests, however, a slower convergence to the normal distribution in the case of coupled signals.

3.3.2 Experiment 2. We demonstrate an application of corollary 3 for bias evaluation with a simple simulation. In section 3.2 we pointed out that using a noninteger fT (T is not a multiple of the oscillation period) can lead to spurious correlation between the point process and the oscillatory continuous signal. By using equation 3.11 we can compute this bias.

We use a simulation similar to the one used in the previous experiment with an oscillatory signal and a homogeneous Poisson spike train ($\kappa = 0$) and investigate the coupling between these two signals. If the length of the continuous signal is not an integer number of the oscillation period, the PLV estimate has a nonzero empirical mean (see Figures 3A and B) while when it is a multiple of number of the oscillation period, the estimate matches the ground truth (see Figure 3C). In Figure 3D we compare the theoretical prediction and the numerical simulation for various length of the signals,

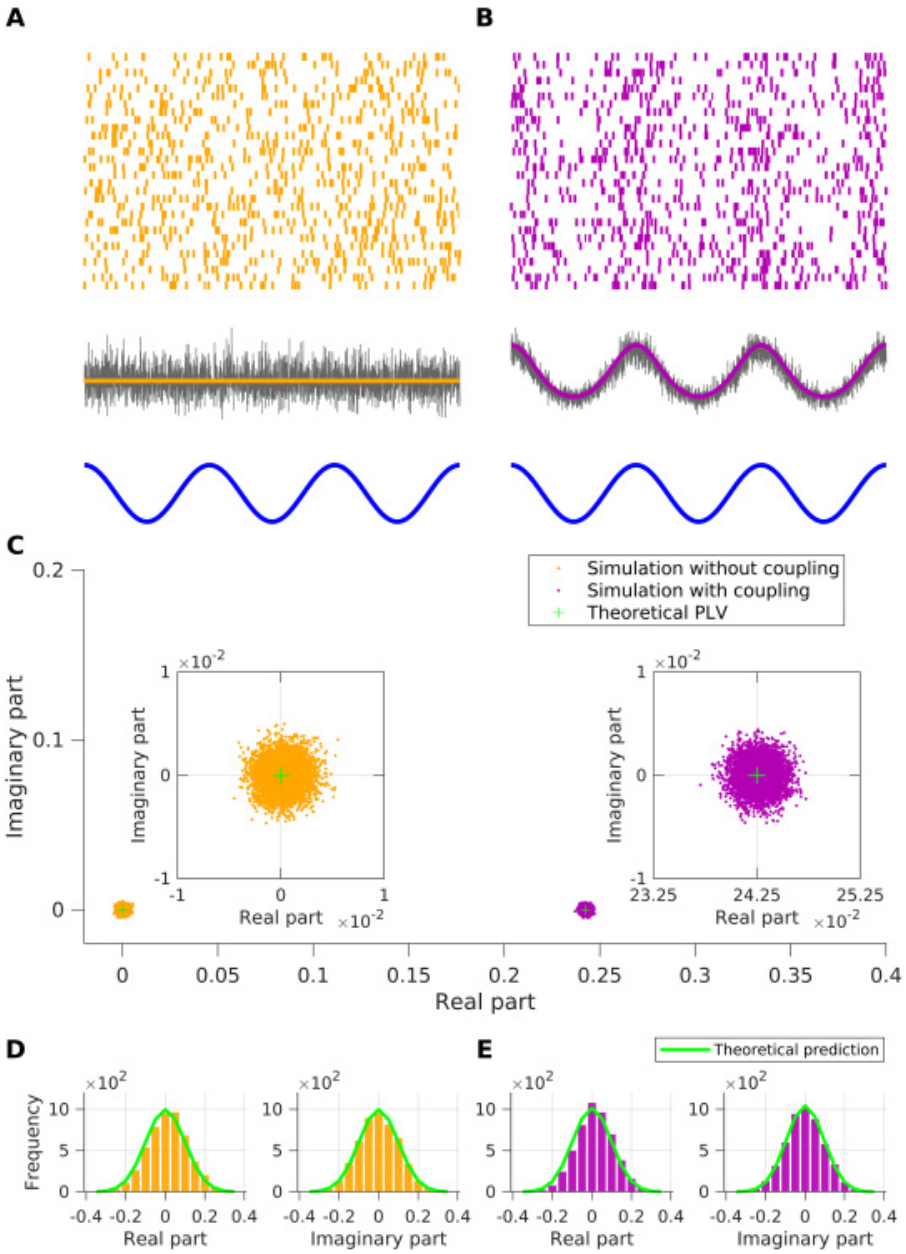


Figure 2: Simulation of (A) homogeneous Poisson spike trains and (B) phase-locked spike train with Poisson statistics (von Mises model with $\kappa = 0.5$). First row: Example raster plot of the spikes. Second row: Empirical firing rate (gray line) and ground truth firing rate (orange and purple traces). Third row: Continuous signal $x(t)$. (C) Scatter plots represent the complex-valued PLVs estimates. Each dot represents one realization of the simulation. Insets depict the zoomed version of both distributions. Green crosses indicate the theoretical complex-valued PLV. (D, E) Histograms of real and imaginary parts of scaled residuals for simulations (D) without coupling and (E) with coupling. Green lines indicate the theoretical predictions of corresponding distributions according to corollaries 2 and 4, and the bars indicate the empirical distributions. Note the subtle difference between real and imaginary parts in panels D versus E. See Table 1 for parameters used for this figure.

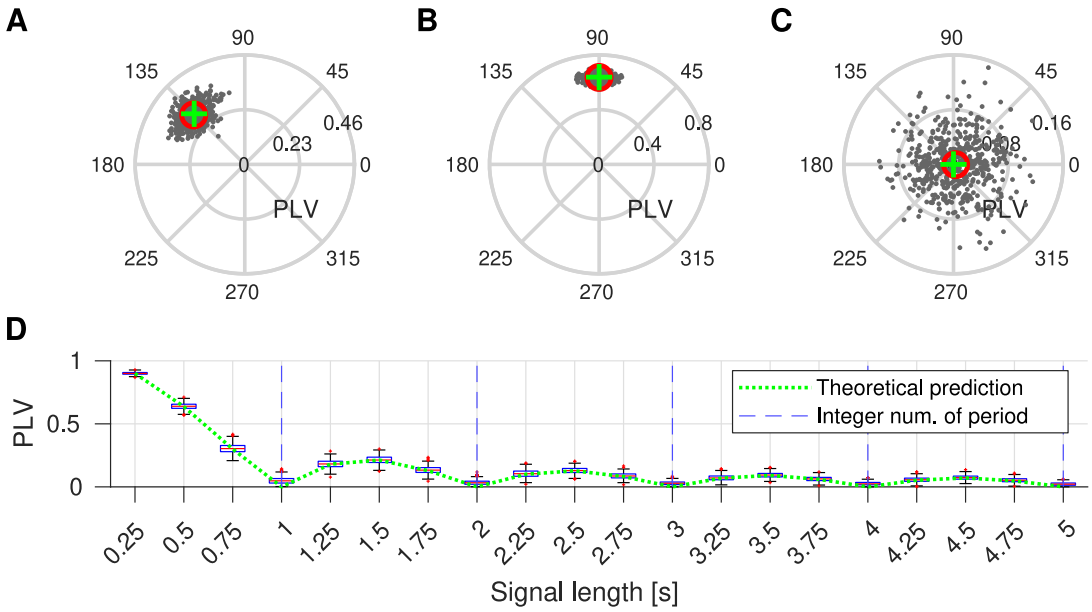


Figure 3: (A–C) Distribution of simulated complex-valued PLVs (gray dots), average of the simulated PLVs (red circle), and theoretical prediction based on equation 3.11 (green crosses) for (A) $\gamma_T = 0.75$, (B) $\gamma_T = 0.5$, and (C) $\gamma_T = 1$. All complex-valued PLVs are represented in the complex plane. Angles indicate the locking phase and the radius the PLV. (D) PLV for different interval lengths T. Box plots represent the simulated PLVs, and the dashed green trace represents theoretical prediction of the expectation based on equation 3.11. Vertical broken blue lines indicate integer number of oscillation periods. See Table 2 for parameters used for this figure.

showing that this effect disappears when the observation window covers a larger number of oscillation periods.

4 Assessment of Multivariate Coupling

High-dimensional data sets have become increasingly important in biology (Bühlmann, Kalisch, & Meier, 2014). More specifically in neuroscience, state-of-the-art multichannel electrophysiology recording systems (Dickey, Suminski, Amit, & Hatsopoulos, 2009; Jun et al., 2017; Juavinett, Bekheet, & Churchland, 2019) allow the simultaneous recording of thousands of sites (Pesaran et al., 2018; Jun et al., 2017; Buzsáki, 2004; Fukushima, Chao, & Fujii, 2015). This growth in dimensionality requires the development of appropriate tools (Stevenson & Kording, 2011; O'Leary, Sutton, & Marder, 2015; Gao & Ganguli, 2015; Williamson, Doiron, Smith, & Yu, 2019) for computing an interpretable summary of the coupling between neurophysiological quantities reflecting the collective dynamics of the underlying neural ensembles (Truccolo, 2016; Safavi et al., 2020). To achieve this aim, deriving low-rank approximations of high-dimensional matrices is supported

by empirical evidence and theoretical predictions of the existence of low-dimensional structures in neural activity (Ermentrout & Kleinfeld, 2001; Ermentrout & Pinto, 2007; Truccolo, Hochberg, & Donoghue, 2010; Gallego, Perich, Miller, & Solla, 2017; Mastrogiuseppe & Ostojic, 2018; Sohn, Narain, Meirhaeghe, & Jazayeri, 2019; Cueva et al., 2020). This section provides statistical results for such approximation in the context of the coupling between point processes and continuous signals.

As a natural extension of the scalar case discussed in the previous section, we now consider the expected coupling matrix $\boldsymbol{C}^*$ between an n-dimensional vector of counting processes $\boldsymbol{N}$ with associated intensity vector $\boldsymbol{\lambda}(t)$ and a multivariate p-dimensional signal $\boldsymbol{x}(t)$, and its estimate based on independent trials $\widehat{\boldsymbol{C}}_K$, respectively defined as

$$\boldsymbol{C}^* = \int_0^T \boldsymbol{x}(t)\boldsymbol{\lambda}(t)^\top dt \quad \text{and} \quad \widehat{\boldsymbol{C}}_K = \frac{1}{K}\sum_{k=1}^{K}\int_0^T \boldsymbol{x}(t)d\boldsymbol{N}^{(k)}(t)^\top. \tag{4.1}$$

In this multivariate setting, the coupling matrix between the point process and continuous signal can be characterized by the singular value(s) of $\boldsymbol{C}^*$,

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0,$$

and associated orthonormal singular vectors $\{(\boldsymbol{u}_k, \boldsymbol{v}_k)\}$, such that

$$\boldsymbol{C}^* = \sum_{k=1}^{p} \boldsymbol{u}_k \sigma_k \boldsymbol{v}_k^H.$$

When the dimension of the coupling matrix gets large, recovering the entire structure of $\boldsymbol{C}^*$ using its estimate $\widehat{\boldsymbol{C}}_K$ becomes unlikely due to the fluctuations of individual coupling coefficients investigated in the previous section. However, the largest singular values may remain reliably estimated because they correspond to low-rank structures of the matrix that stand out from the noise. Random matrix theory provides justifications for this approach by characterizing the spectral properties of "noisy" matrices. Up to a normalization explained later, this will involve indirectly characterizing the behavior of the empirical singular values $\{\widehat{\sigma}_k\}$ of the estimate matrix $\widehat{\boldsymbol{C}}_K$ by analyzing the eigenvalues of the hermitian matrix $\frac{1}{n}\widehat{\boldsymbol{C}}_K\widehat{\boldsymbol{C}}_K^H$ denoted

$$\ell_1 \geq \ell_2 \geq \ldots \geq \ell_p \geq 0.$$

These are related to each other by the relation $\widehat{\sigma}_k = \sqrt{n\ell_k}$ for all k.

4.1 Mathematical Formulation. We now replace assumption 1 to adapt to this multivariate setting. By restricting ourselves to homogeneous

Poisson processes, we investigate a null hypothesis of no coupling between continuous signals and point processes. Let us denote $\bar{x}$ the complex conjugate of x and δ the Kronecker delta symbol:

$$\delta_{lj} = \begin{cases} 1, & \text{if } l = j, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

Assumption 2 (Complex Multivariate Case). We consider an infinite sequence $\{x_j(t)\}_{j\geq 1}$ of complex valued left-continuous deterministic functions uniformly bounded on $[0,\ T]$ and assume

(1) For all $i,\ j \geq 1$, $\frac{1}{T}\int_0^T \bar{x}_i x_j dt = \delta_{ij}$ and $\int_0^T x_i x_j dt = 0$.
(2) For all $i \geq 1$, $\int_0^T x_i dt = 0$,
(3) There exist $0 < \lambda_{\min} < \lambda_{\max}$ and a sequence of independent homogeneous Poisson processes $\{N_i\}_{i\in\mathbb{N}^*}$'s with associated rates $\{\lambda_i\}_{i\in\mathbb{N}^*}$ in the interval $[\lambda_{\min},\ \lambda_{\max}]$.

While the assumptions on $\{x_i(t)\}$ are designed for complex signals, which is the classical case when dealing with PLV-like quantities, the results of this section also hold for real signals by using the assumption $\frac{1}{T}\int_0^T x_i x_j dt = \delta_{ij}$ instead of the above condition 1. Condition 2 is also added to ensure that there is no trivial bias leading to a nonvanishing expectation of the coupling coefficients (as illustrated in section 3.2). Indeed, when the time average of each signal vanishes, based on theorem 1, the expectation of all univariate coupling measures for a homogeneous Poisson process vanishes. We then exploit a multivariate generalization of the martingale CLT to characterize the distribution of the coupling matrix given these assumptions.

Theorem 2. *For given $n,\ p \geq 1$ and all $K \geq 1$, we use sequences of signals defined in assumption 2 to build multivariate continuous signal $\boldsymbol{x}(t) = (x_j)_{j=1\ldots p}$ and K independent copies of multivariate Poisson process $\boldsymbol{N}(t) = (N_i)_{i=1\ldots n}$ with rate vector $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_n]^\top$. Then the normalized coupling matrix $\sqrt{K}\widehat{\boldsymbol{C}}_K diag(\sqrt{T\boldsymbol{\lambda}})^{-1}$, with $\widehat{\boldsymbol{C}}_K$ given by equation 4.1, converges in distribution for $K \to +\infty$ to a matrix with i.i.d. complex standard normal coefficients.*

Sketch of the proof. This essentially uses a generalization of the CLT to multivariate point processes described in Aalen et al. (2008, appendix B). Based on the statistics of stochastic integrals presented in section B.1.1, assumptions on $\boldsymbol{x}$ entail vanishing correlations between all matrix coefficients and lead to the analytical expression of the covariance matrix. □

This result suggests that for large n and $p = p(n)$, coupling matrices $\widehat{\boldsymbol{C}}_K^n$ of increasing size can be used to build the Wishart-like matrix sequence,

$$\boldsymbol{S}_n \triangleq \frac{K}{n}\widehat{\boldsymbol{C}}_K^n \mathrm{diag}(T\boldsymbol{\lambda})^{-1}(\widehat{\boldsymbol{C}}_K^n)^H, \tag{4.3}$$

whose ESD may converge to the Marchenko-Pastur law. This is, however, not guaranteed by classical results due to the nongaussianity and dependence of the matrix coefficients of $\widehat{C}_K^n$ for fixed n and K. Convergence will thus depend on how much the departure from these assumptions plays a role as n becomes large. We show in the following theorem that increasing the number of trials as a function of the dimension guarantees convergence to the MP law.

Theorem 3. *In addition to assumption 2, assume an increasing, positive integer sequence* $\{p(n), K(n)\}_{n\in\mathbb{N}^*}$ *such that* $\frac{p(n)}{n} \underset{n\to+\infty}{\longrightarrow} \alpha \in (0, +\infty)$*, and*

$$\frac{1}{n^2K(n)^2}\sum_{\Gamma}\left(\int_0^T \bar{x}_j x_l x_{j'} \bar{x}_{l'} dt\right)^2 \to 0, \text{ uniformly in } k \leq n, \tag{4.4}$$

where $\Gamma = \{(j, l, j', l') : 1 \leq j, l, j', l' \leq p\} \setminus \{(j, l, j', l') : j = j' \neq l = l' \text{ or } j = l' \neq j' = l\}$*. Consider the sequence* $\{\widehat{C}^n_{K(n)}\}_{n\in\mathbb{N}^*}$ *built as in theorem 2 for* $p = p(n)$*; then the corresponding sequence* $\{\boldsymbol{S}_n\}$ *defined by equation 4.3 has an ESD converging weakly with probability one to the MP law of equation 2.7.*

Sketch of the proof. We use theorem 1.1 of Bai and Zhou (2008) addressing the case of matrices with dependence of coefficients within columns. We use Itô's formula (see appendix B) to check the simplified necessary conditions provided in corollary 1.1 of Bai and Zhou (2008). This implies convergence of the Stieltjes transform to the same function as the transform of the MP distribution. By classical results on the Stieltjes transform (Anderson et al., 2010, theorem 2.4.4), this implies weak convergence to the MP measure (i.e., convergence for the weak topology; see appendix B.2). □

Remark 2. Condition in equation 4.4 determines how many trials are needed at most for spectral convergence. Due to the uniform boundedness assumption on signal $\boldsymbol{x}(t)$ and given the number of terms in the sum bounded by n^4, we can already see that $\frac{n}{K(n)} \to 0$, that is, having the number of trials increasing at an even slightly faster rate than the dimension is enough for convergence for any choice of continuous signals respecting orthonormality assumption 2. However, there are cases where even fewer trials than dimensions are required. An important example is the Fourier basis of the $[0, T]$ interval, $\boldsymbol{x}_l(t) = \frac{1}{\sqrt{T}}\exp(\boldsymbol{i}2\pi lt/T)$. Then all terms in the sum of equation 4.4 vanish, except the ones satisfying $j - j' - l + l' = 0$, such that we are left with a number of bounded terms that scale with n^3. As a consequence, the condition on the number of trials to achieve spectral convergence becomes $\frac{\sqrt{n}}{K(n)} \to 0$, such that we need increasingly fewer trials than dimensions. On the contrary, due to the uniform bound that we impose on the signals, choosing a basis of signals with decreasing support, such as a wavelet basis, typically departs from our condition 1 of assumption 2, as the normalization of condition 1 of assumption 2 imposes unit

norm on each signal, requiring their amplitude to increase as their support decreases, violating the uniform bound assumption. This limitation supports the intuition that statistical regularities exploited by our asymptotic results deteriorate with highly transient signals.

This convergence of the spectral measure to the MP law guarantees eigenvalues do not accumulate in a large proportion above the upper end of the support of the MP law; however, they do not provide rigorous guarantees regarding convergence of individual eigenvalues and, in particular, the largest eigenvalue. Although such convergence is satisfied in classical settings (gaussian i.i.d. coefficients), they typically require stronger assumptions than for the (weak) spectral convergence to the MP law, and still only very few results are available in the non-i.i.d. setting. We could, however, prove such convergence by adding a constraint to our model.

Theorem 4. *In addition to assumption 2, assume all homogeneous rates λ_k are equal. Assume two increasing, positive integer sequences $\{p(n), K(n)\}_{n\in\mathbb{N}^*}$ such that*

$$\frac{p(n)}{n}\to\alpha\in(0,+\infty)\quad \textit{and}\quad \frac{1}{K(n)}\sum_{1\le i,k\le p(n)}\int_0^T |x_i x_j|^2(t)dt < B, \tag{4.5}$$

for some constant B. Then $\boldsymbol{S}_n$ defined in equation 4.3 has an ESD converging weakly with probability one to the MP law of equation 2.7. Moreover, let ℓ_1 and ℓ_p be the largest and smallest eigenvalues of $\{\boldsymbol{S}_n\}$, respectively. Then in probability

$$\ell_1(n)\to(1+\sqrt{\alpha})^2\quad \textit{and}\quad \ell_p(n)\to(1-\sqrt{\alpha})^2\mathbf{1}_{\alpha<1}.$$

Sketch of the proof. The identical intensities assumption allows us to use the result of Chafaï and Tikhomirov (2018) for matrices with i.i.d. columns. We first checked that their proof holds also for the complex case by replacing symmetric matrices by Hermitian matrices and squared scalar product by an absolute squared Hermitian product. We satisfy their strong tail projection (STP) assumption using Chebyshev's inequality. The necessary fourth-order moment conditions exploit the same stochastic integration results as theorem 3. □

Remark 3. Without additional assumptions, the moment condition of equation 4.5 is satisfied by choosing $K(n) = n^2$ (as there are p^2 bounded moments, scaling as n^2 when n grows). It is likely from the proof that taking into account more information about the moments of the continuous signal sequence $\{x_j\}$, we can achieve convergence with a lower rate of increase for the number of trials. This is left to future work.

This result thus provides the guarantees that under a null hypothesis of no coupling (due to homogeneity of the Poisson processes), the extreme eigenvalues of S will asymptotically cover exactly the full support

of the MP law. This will be used in section 4.2 to assess the significance of the eigenvalues ℓ_k by simply checking whether they are larger than $(1+\sqrt{\alpha})^2$.

This significance analysis relies as well on understanding what happens to the eigenvalues when the model departs from the null hypothesis. In a practical setting, we hypothesize that the coupling matrix has a deterministic structure superimposed on the martingale noise modeled in the above results. One qualitative justification of this assumption can be found in remark 1, showing that for sinusoidal coupling, a nonvanishing expectation proportional to the coupling is superimposed to martingale noise, whose distribution is unaffected by coupling, such that the noisy part of the matrix satisfies the conditions of the above theorems. As typically done in applications, we are mostly interested in the low-rank structure associated with the largest singular values of the coupling matrix, providing an interpretable summary of the multivariate interactions.

This naturally leads to a modeling departure from the null hypothesis with a low-rank perturbation assumption. In such a case, the eigenvalues related to significant coupling are expected to be reflected in the spectrum of the perturbed matrix, such that they can be isolated from the remaining eigenvalues associated with the martingale noise. This intuition is justified by results in the case of the Wishart ensemble (Loubaton & Vallet, 2011); see also Benaych Georges and Nadakuditi (2012) for a more general result and Capitaine and Donati Martin (2016) for an overview of matrix perturbation results), that we restate here:

Theorem 5 (From Loubaton & Vallet, 2011, Theorem 6). *Let $\boldsymbol{X}_n$ be an $n \times p$ sequence of i.i.d. complex gaussian matrices defined in section 2.3 and A_n be a finite rank perturbation of the null matrix with nonzero eigenvalues θ_i. Let $\boldsymbol{M}_n = (\frac{1}{\sqrt{n}}\boldsymbol{X}_n + A_n)(\frac{1}{\sqrt{n}}\boldsymbol{X}_n + A_n)^H$. Then as $n \to \infty$ and $\frac{p}{n} \to \alpha \in (0,1)$, almost surely,*

$$\lambda_i(M_n) \to \begin{cases} \frac{(1+\theta_i)(c+\theta_i)}{\theta_i}, & \text{if } \theta_i > \sqrt{\alpha}, \\ (1+\sqrt{\alpha})^2, & \text{otherwise}. \end{cases}$$

A demonstration that this further applies rigorously to our nongaussian, non-i.i.d. case is left to further work (but see Benaych-Georges & Nadakuditi, 2012, for a generalization in this direction). This result shows the upper end of the MP support is indeed the critical threshold for the eigenvalues of A_n to stand out from the noise. Below this threshold, the largest eigenvalue convergence to the upper end of the support of the MP distribution is not informative about θ_i. Above this threshold, the value of θ_i can be recovered and detected by comparing the largest eigenvalue to the upper end of the MP distribution.

We next illustrate the interest of these theoretical predictions in the context of neural time series for reliably quantifying the interplay between multichannel LFP signals and the spiking of multiple neurons. Nevertheless, the results are potentially applicable in other domains as well. In neuroscience, $\boldsymbol{x}$ may represent LFP measurements collected on each recording channel and $\boldsymbol{N}$ the spiking activity of different neurons, called units. The number of recording channels n_c and recorded units n_u correspond to p and n, respectively. These numbers may differ, and as a consequence, the coupling matrix is generally rectangular.

4.2 Application to Significance Assessment. In order to statistically assess the significance of the largest singular value(s) of coupling matrix $\widehat{\boldsymbol{C}}_K^n$, considered as a measure of coupling between point processes and continuous signals, we need a null hypothesis. Hypothesis testing based on the generation of surrogate data is one of the common methods for significant assessment in neuroscience and other fields. Generating appropriate surrogate data can be not only challenging (see Grün, 2009, and Elsayed & Cunningham, 2017, for examples in neuroscience), but also computationally expensive due to the increasingly large dimensions of modern data sets. Exploiting our theoretical results for this setting allows us to perform such statistical assessment in a principled way, without using surrogate data and sparing computational resources.

In order to exploit the results of the theoretical part, it is best to preprocess the $p \times q$ matrix of time-discretized signals $\mathbf{L}$ that correspond to q samples over interval $[0, T]$ with sampling interval $\Delta = T/q$. The chosen signals are driven by the application (in our case, they are preprocessed LFPs, see section 4.3 for a simulation reproducing the context of neurophysiolgy data). We assume the rows of $\mathbf{L}$ sum to zero to match condition 2 of assumption 2 (and avoid bias in the coupling measure similar to what is described in section 3.3). We then need to process further this signal such that condition 1 of assumption 2 is satisfied approximately. In order to achieve this, we perform classical whitening of the signals to generate matrix $\mathbf{X}$, the discrete time approximation of $\boldsymbol{x}(t)$, according to

$$\mathbf{X} = \mathbf{W}\mathbf{L}, \quad \text{with} \quad \mathbf{W} = \left(\frac{1}{\mathrm{q}}\mathbf{L}\mathbf{L}^{\mathrm{H}}\right)^{-\frac{1}{2}}, \tag{4.6}$$

where the power in the expression of the whitening matrix $\mathbf{W}$ describes the inversion of a matrix square root, typically achieved via eigenvalue decomposition, and which may require PCA-like dimensionality reduction in practice to minimize the numerical effects of small eigenvalues. This procedure decorrelates the martingale fluctuations of coefficients within the same column of the coupling matrix (see theorem 2), a key requirement for convergence to the MP law.

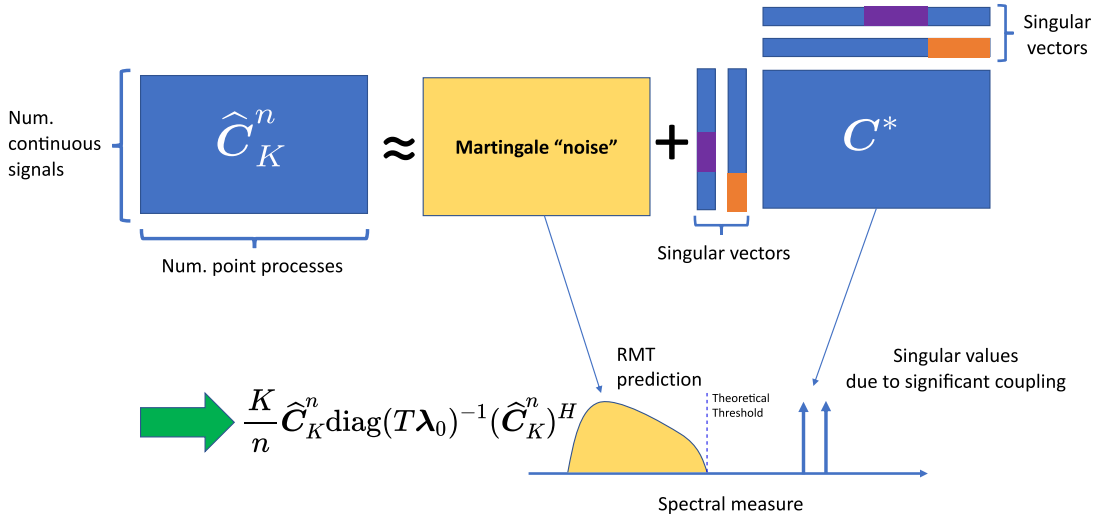


Figure 4: We assume $\widehat{C}_K^n$ is a superposition of martingale noise and a low-rank deterministic matrix C^* reflecting the actual coupling. If the singular values of a normalized version of C^* are large enough (larger than the upper end of the MP law support), theory suggests that they will correspond to the largest eigenvalues of S_n appearing beyond the support of MP distributed eigenvalues reflecting martingale noise. They can thus be detected with a simple thresholding approach (see equation 4.7).

As explained in section 4.1, theoretical results support using $\theta_{DET} = (1 + \sqrt{\alpha})^2$, the upper end of the support of the MP law, as a detection threshold for the significance of the eigenvalues of the Hermitian matrix,

$$S_n = \frac{K}{n}\widehat{C}_K^n \mathrm{diag}(T\boldsymbol{\lambda}_0)^{-1}(\widehat{C}_K^n)^H.$$

The null hypothesis of nonsignificance of the kth largest singular value $\widehat{\sigma}_k$ of the normalized coupling matrix,

$$\sqrt{K}\widehat{C}_K^n \mathrm{diag}(\sqrt{T\boldsymbol{\lambda}_0})^{-1},$$

should thus be rejected if the corresponding kth largest eigenvalue ℓ_k of S_n is superior to the significance threshold, leading to the condition

$$\widehat{\sigma}_k = \sqrt{n\ell_k} > \sqrt{n\theta_{DET}} = \sqrt{n}(1 + \sqrt{\alpha}). \tag{4.7}$$

Therefore, this last condition on the empirical singular values is used to identify those reflecting a significant coupling between the multivariate point process and continuous signal. An illustration of our overall significance assessment approach is shown in Figure 4.

4.3 Simulation. We use a simulation to demonstrate the outcome of our (asymptotic) theoretical results on mutlivariate coupling. Similar to the simulations of section 3.3 for the univariate case, we use simulated phase-locked spike trains with Poisson statistics. The main difference between this simulation and the previous one is in synthesizing the LFP. In order to simulate multichannel oscillatory signals that lead to a low-rank structure for $\boldsymbol{C}^*$, we use a combination of noisy oscillatory components.

The LFPs contain N_{osc} oscillatory groups of channels; each channel l within the same group contains the same oscillatory component with index $j(l)$, with the time course of all these components being $O_j(t) = e^{2\pi \boldsymbol{i} f_j t}$, $j \in \{1, \ldots, N_{osc}\}$, with all frequencies f_j in the range $[f_{\min},\ f_{\max}]$, and all multiples of $1/T$. Due to the necessary time axis discretization, the bracket notation $[t]$ indicates the oscillation is sampled at equispaced discrete times $t = \{k\Delta\}_{k=1,\ldots,q}$. The synthesized discrete time multichannel LFP ($\boldsymbol{\Psi}[t] = \{\psi_l[t]\}_{l=1,\ldots,n_c}$) can be written as

$$\Psi_l[t] = O_{j(l)}[t] \odot \exp\left(\boldsymbol{i}\eta_l[t]\right), \tag{4.8}$$

with $\odot$ entrywise product and $\{\eta_l[t]\}$ i.i.d. sampled (white) phase noises contaminating each channel independently (see appendix D for more details).

In this simulation, the frequencies of the oscillatory components range from 11 to 15 Hz. We used 100 LFP channels ($n_c = 100$) and different choices for the number of spiking units (10, 50, and 90). Spiking activities are simulated in different scenarios, with and without coupling to the LFP oscillations. In the latter case, we have two populations of neurons (each consisting of 1/5th of the total number of neurons) that are each coupled to one of the oscillatory groups of LFP channels. Both populations are coupled to their respective oscillation with identical strength ($\kappa = 0.15$) and phase ($\phi_0 = 0$).

To compute the coupling matrix $\widehat{\boldsymbol{C}}_K$, we first preprocess $\boldsymbol{\Psi}[t]$ by applying bandpass filtering in a range covering $[f_{\min},\ f_{\max}]$ and convert it to an analytic signal via the discrete time Hilbert transform, leading to data matrix $\mathbf{L}$, following the standards of PLV analysis in neuroscience (Chavez, Besserve, Adam, & Martinerie, 2006).

This signal matrix is then whitened according to equation 4.6 to yield matrix $\mathbf{X}$, the discrete time version of $\boldsymbol{x}(t)$. The coupling matrix $\widehat{\boldsymbol{C}}_K$ is then computed according to equation 4.1 using 10 trials (barring trivial approximation to the closest time sample in $\mathbf{X}$).

Then in order to approximate the normalization $\sqrt{K}\widehat{\boldsymbol{C}}_K^n \mathrm{diag}(\sqrt{T\boldsymbol{\lambda}_0})^{-1}$ based on empirical data, we use the total number of events for unit u occurring across all K trials $N^u_{\mathrm{tot}} = \sum_{k=1}^K N^u_k$ and multiply each column u of the coupling matrix by

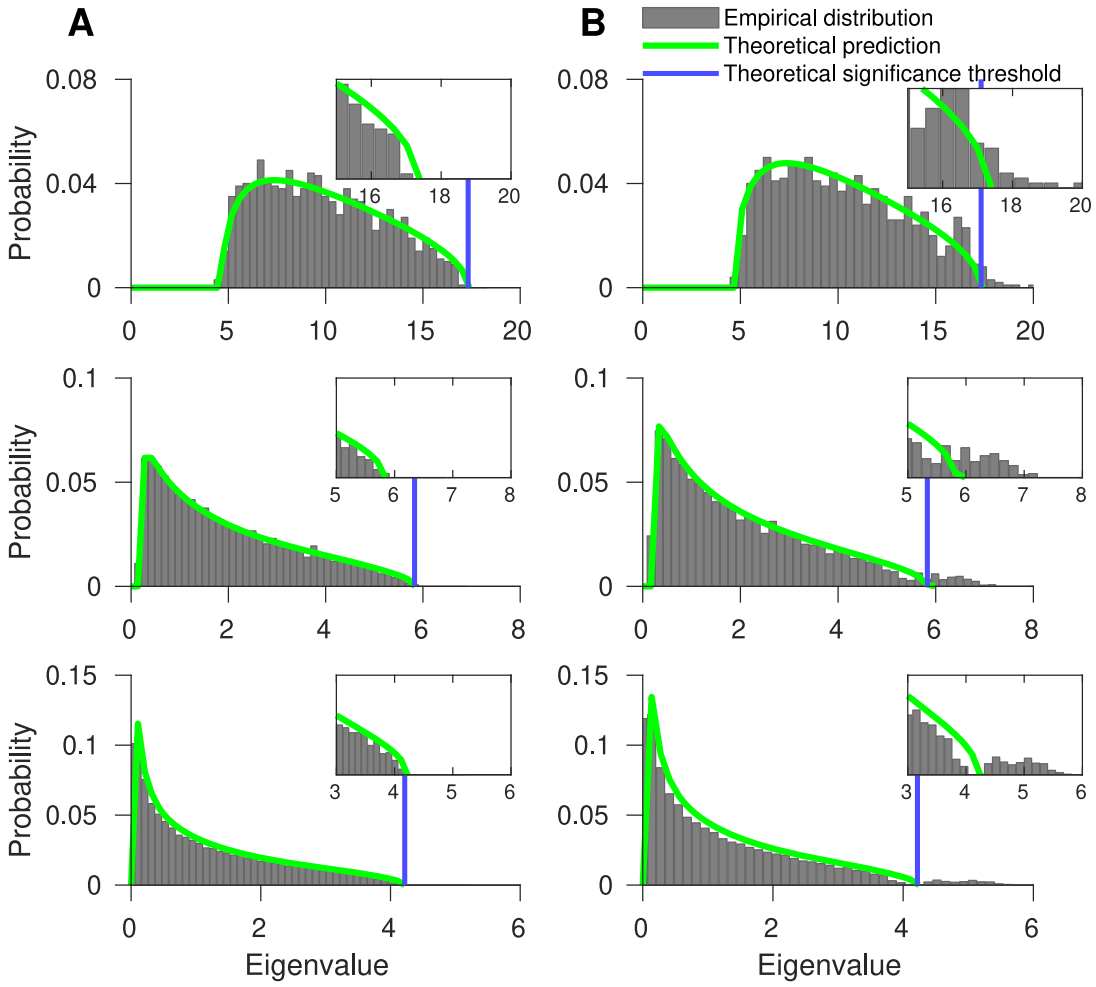


Figure 5: Theoretical Marchenko-Pastur distribution (green lines) and empirical distribution (gray bars) for (A) simulation without coupling ($\kappa = 0$) and (B) with coupling ($\kappa = 0.15$) between multivariate spikes and LFP. Rows represent the spectral distribution of simulations with different number of spiking units—rows 1, 2, and 3, respectively, 10, 50, and 90 (which leads to different α for MP law). Insets zoom the tail of the distributions. Parameters used for this figure are denoted in Table 3.

$$\frac{K}{\sqrt{N^u_{\text{tot}}}} \approx \frac{K}{\sqrt{K \int_0^T \lambda_u(t)dt}}$$

for the corresponding unit u, asymptotically matching the theoretical normalization in the homogeneous Poisson case.

We observe in Figure 5A that in the absence of coupling, the distribution of eigenvalues originating from the random matrix structure is very close to the theoretically predicted MP distribution. In Figure 5B, we have coupling between spike and eigenvalues reflecting the coupling beyond the MP support (blue line in Figure 5), and the eigenvalue bulk below the threshold is

also close to MP distribution. This suggests an easy thresholding approach for significance assessment.

5 Discussion

5.1 Insights for Data Analysis. Our theoretical results provide guarantees for specific coupling models to respect univariate and multivariate asymptotic statistics that can be easily exploited for statistical testing. The required assumptions provide guidelines for practical settings that are likely of interest beyond the strict framework that we imposed to get the rigorous results. For the univariate coupling measure, corollaries and simulations point out the importance of the choice of the observation interval $[0, T]$, which is particularly sensitive when considering short intervals covering only a few oscillation periods. This is the case when doing time-resolved analysis or dealing with experiments with short trial durations. Moreover, the univariate results also emphasize the effect of nonlinear phase increases, highly relevant in neuroscience due to the pervasive effects of nonlinear dynamics in the mesoscopic signals. Our results provide asymptotic bias correction terms that can be used for statistical testing.

In the same way, theoretical results in the multivariate setting may seem to be constrained by our assumptions, but they provide critical guidelines to interpret singular values. First, whitening the continuous signals and normalizing the coupling by the square root of the rate are key preprocessing steps for making the asymptotic behavior of the martingale noise invariant to the specifics of the data at hand. This then reduces to an analytical model, the MP law, dependent on only a single matrix shape parameter. After assessment of the significance of the singular value of the normalized coupling matrix, it is of course possible to revert these preprocessing steps to get a low-rank approximation of the original coupling matrix (nonnormalized, nonwhitened) to summarize the significant coupling structure in an interpretable way. A second insight provided by the multivariate results is the role of fourth-order moments of the continuous signals, represented by the integrals of order four monomials of components of $\boldsymbol{x}(t)$, in the MP convergence results. The magnitude of these moments determines the number of trials asymptotically needed to achieve convergence. Since these moments can be estimated empirically, we can check how they grow with the dimension of the signals in a specific application. With our minimal assumptions on the signals, the number of trials need only to grow at most sublinearly in the dimension for spectral convergence; however, we could only show that convergence of the largest eigenvalues requires at most quadratic increase in the dimension n. This last result might be improved in future work, with extra assumptions, to reach linear growth.

Our theoretical results can be extended in two directions. The first is toward exploiting point processes different from inhomogeneous Poisson

(e.g., Hawkes processes) in order to be able to apply the framework in the context of stochastic intensities. The second direction is toward exploiting recent developments in RMT, in order to develop a probabilistic significance assessment.

5.2 Extension of Signal Assumptions. Our theoretical results assume deterministic continuous signals and point process intensities (see assumption 1). This entails limitations, such as implicitly assuming the considered point processes are (homogeneous or inhomogeneous) Poisson processes. This assumption may be too restrictive in realistic scenarios (for examples in neuroscience, see Deger, Helias, Boucsein, & Rotter, 2012; Reimer, Staude, Ehm, & Rotter, 2012; Nawrot et al., 2008; Shinomoto, Shima, & Tanji, 2003; Maimon & Assad, 2009; Shinomoto et al., 2009). However, the stochastic integration methods that provide the basis of our results allow the treatment of random signals and intensities, provided they are predictable, which encompasses a wide enough class of processes to cover most applications (Protter, 2005). Our results thus have potential for generalizations to the case of random continuous signal, with the difference that the variance of the estimates would increase due to the additional variability induced by the signal fluctuations, and to the case of random intensities, leading to different asymptotic properties of the coupling measures, which may or may not have simple analytical expressions.

As a potential direction for extending this framework, Hawkes processes (Hawkes, 1971) are point processes for which the probability of occurrence of future events can also depend on the sequence of past events. Due to this history dependency, they are also called self-exciting processes. Hawkes processes are used for modeling recurrent interactions in various fields; for instance—in finance it is used to model buy or sell transaction events on stock markets (Embrechts, Liniger, & Lin, 2011) in geology to model the origin times and magnitudes of earthquakes (Ogata, 1988), in online social media to model user actions over time (Rizoiu, Lee, Mishra, & Xie, 2017), and even modeling reliability of information on the web and controlling the spread misinformation (Tabibian et al., 2017; Kim, Tabibian, Oh, Schölkopf, & Gomez-Rodriguez, 2018), and in neuroscience to model spike trains (Krumin et al., 2010). We conjecture that such history dependency can be incorporated in our analytic treatment of the coupling measure, such that our theoretical results can be extended to this model.

5.3 Extension beyond Binary Significance Assessment. We show that the Marchenko-Pastur distribution provides a good approximation of the distribution of eigenvalues in the absence of coupling, and the upper end of its support approximates the largest eigenvalue. This provides us a threshold to assess the significance of empirical singular values. Nevertheless, this hard thresholding approach does not take into account the actual fluctuations of the largest eigenvalue around this asymptotic upper end of the

support and thus does not provide meaningful *p*-value for the statistical test.

It has been shown that the appropriately rescaled and recentered[8] largest eigenvalue of Wishart matrices is asymptotically distributed as the Tracy-Widom distribution—for example, see Johnstone (2001); Tracy and Widom (2002); and El Karoui (2003, 2005, 2007). However, note that in some cases of practical relevance, the normal distribution might be more appropriate (Bai & Yao, 2008). Such asymptotic distribution of the largest eigenvalue can be exploited for reporting a theoretical *p*-value for the significance of the coupling and therefore extending the significance assessment from a binary decision to a probabilistic one. For example, Kritchman and Nadler (2009) exploit this idea (but in a simpler scenario) to determine the number of signal components in noisy data. This extension would allow a precise probabilistic assessment of the significance of weaker couplings leading to eigenvalues in the neighborhood of the asymptotic threshold introduced above.

6 Conclusion

We investigated the statistical properties of coupling measures between continuous signals and point processes. We first used martingale theory to characterize the distributions of univariate coupling measures such as the PLV. Then, based on multivariate extensions of this result and RMT, we established predictions regarding the null distribution of the singular values of coupling matrices between a large number of point processes and continuous signals and a principled way to assess significance of such multivariate coupling. These theoretical results build a solid basis for the statistical assessment of such coupling in applications dealing with high dimensional data.

Appendix: Proofs of Theorems in the Main Text

Proof of Theorem 1. For the first part of the theorem (expectation), we use the martingale $M^{(k)}$ associated with each copied process $N^{(k)}$ to rewrite

$$\widehat{c}_K = \frac{1}{K}\sum_{k=1}^{K}\int_0^T x(t)dM^{(k)}(t) + \frac{1}{K}\sum_{k=1}^{K}\int_0^T x(t)\lambda(t)dt(t). \tag{A.1}$$

Elements of the sum in the first term are then zero mean martingale, and by linearity, so is the whole term. As a consequence (using the zero mean

[8] The required recentering and rescaling of the eigenvalues is studied in the literature (Johnstone, 2001; El Karoui, 2003, 2007).

property), the expectation of the first term is zero so only the second term remains:

$$\mathbb{E}\left[\widehat{c}_K\right] = \int_0^T x(t)\lambda(t)dt(t).$$

We then exploit a central limit theorem (CLT) for martingales to prove the second part of the theorem (convergence to gaussian distribution). To satisfy the CLT in such a case, it is sufficient to find a particular martingale $\widetilde{M}^{(K)}$ sequence that will satisfy the conditions described in Aalen et al. (2008, p. 63) ($\xrightarrow{P}$ indicate convergence in probability):

1. $\mathrm{Var}(\widetilde{M}^{(K)}(t)) \xrightarrow[K\to+\infty]{P} \widetilde{V}(t)$ for all $t \in [0, T]$, with $\widetilde{V}$ increasing and $\widetilde{V}(0) = 0$.
2. Informally, the size of the jumps of $\widetilde{M}^{(K)}$ tends to zero (see Aalen et al., 2008, p. 63). Formally, for any $\epsilon > 0$, the martingale $\widetilde{M}_\epsilon^{(K)}(t)$ gathering the jumps $> \epsilon$ satisfies $\mathrm{Var}\left(\widetilde{M}_\epsilon^{(K)}(t)\right) \xrightarrow[K\to+\infty]{P} 0$.

Then $\widetilde{M}^{(K)}(t)$ converges in distribution to a gaussian martingale of variance $\widetilde{V}(t)$.

To achieve these conditions, we define $M^{(k)}$, the sequence of i.i.d. zero mean martingales defined on $[0, T]$ canonically associated with the point process of each trial $N^{(k)}$. Then we build martingales $M_x^{(k)}(t) = \int_0^t x(s)dM^{(k)}(s)ds$ and construct $\widetilde{M}^{(K)} = 1/\sqrt{K}\sum_{k=1}^K M_x^{(k)}$.

The variance of this latter martingale (also called its predictable variation processes) can be computed based on the rules provided in section B.1.1. First, due to trial independence,

$$\widetilde{V}(t) = \mathrm{Var}\left(\widetilde{M}^{(K)}(t)\right) = \mathrm{Var}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^K M_x^{(k)}(t)\right) = \sum_{k=1}^K \mathrm{Var}\left(\frac{1}{\sqrt{K}}M_x(t)\right), \tag{A.2}$$

and using equation B.4, we get

$$\widetilde{V}(t) = \frac{1}{K}\sum \int_0^t x^2(t)\lambda(t)dt = \int_0^t x^2(t)\lambda(t)dt. \tag{A.3}$$

Equation A.3 clearly fulfills CLT's condition 1.

For condition 2, due to assumption 1, $x(t)$ is bounded, such that there is a $B > 0$ satisfying $|x(t)| < B$ over $[0, T]$. As a consequence, the size of all jumps is bounded by $B/\sqrt{K}$, and for any ϵ, $\widetilde{M}_\epsilon^{(K)}(t)$ is the constant zero for $K > \frac{B^2}{\epsilon^2}$ and condition 2 is satisfied.

Fulfillment of both conditions leads to convergence in distribution to a gaussian martingale of variance $\widetilde{V}(t)$;

$$\widetilde{M}^{(K)} \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(0, \int_0^T x^2(t)\lambda(t)dt\right). \tag{A.4}$$

Finally, using equation A.1, we conclude the proof by noticing that the above martingale corresponds exactly to the quantity $\sqrt{K}\,(\widehat{c}_K - c^*)$. Therefore,

$$\sqrt{K}\,(\widehat{c}_K - c^*) \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(0, \int_0^T x^2(t)\lambda(t)dt\right). \tag{A.5}$$

□

Proof of Corollary 1. We apply theorem 1 to $e^{i\phi(t)}$ (i.e., replacing $x(t)$ with $e^{i\phi(t)}$). As $e^{i\phi(t)}$ is complex valued, we should have a covariance function for its predictable variation process $\widetilde{V}(t)$. The covariance between a martingale's real part,

$$M_{\mathrm{Re}}(t) = \int_0^t \mathrm{Re}(e^{i\phi(s)})dM(s)ds,$$

and imaginary part,

$$M_{\mathrm{Im}}(t) = \int_0^t \mathrm{Im}(e^{i\phi(s)})dM(s)ds,$$

is given by

$$\int_0^t \mathrm{Re}(e^{i\phi(s)})\mathrm{Im}(e^{i\phi(s)})\lambda(s)ds. \tag{A.6}$$

The diagonal elements of the covariance function are the predictable variation process of M_{Re} and M_{Im} that can be computed based on equation B.4, and the off-diagonal elements are the covariance between martingale's real and imaginary part that can be computed based on equation B.5. Therefore, the covariance function for its predictable variation process is

$$\begin{aligned}&\mathrm{Cov}\left(\begin{bmatrix}\mathrm{Re}\{Z\}\\ \mathrm{Im}\{Z\}\end{bmatrix}\right)\\ &= \begin{bmatrix}\int_0^t \left(\mathrm{Re}(e^{i\phi(s)})\right)^2\lambda(s)ds & \int_0^t \mathrm{Re}(e^{i\phi(s)})\mathrm{Im}(e^{i\phi(s)})\lambda(s)ds\\ \int_0^t \mathrm{Re}(e^{i\phi(s)})\mathrm{Im}(e^{i\phi(s)})\lambda(s)ds & \int_0^t \left(\mathrm{Im}(e^{i\phi(s)})\right)^2\lambda(s)ds\end{bmatrix}\end{aligned} \tag{A.7}$$

$$= \int_0^t \begin{bmatrix} \cos^2(\phi(s)) & \sin(2\phi(s))/2 \\ \sin(2\phi(s))/2 & \sin^2(\phi(s)) \end{bmatrix} \lambda(s)ds. \tag{A.8}$$

Similar to theorem 1, as $K \to +\infty$, the residuals converge in distribution to a zero-mean complex gaussian variable Z (i.e., the joint distribution of real and imaginary parts is gaussian):

$$\sqrt{K}\left(\widehat{\mathrm{c}}_K - \mathrm{c}^*\right) \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(0, \mathrm{Cov}(Z)\right).$$

Because theorem 1 guarantees that the $\sqrt{K}(\widehat{c}_K - c^*)$ tends to a gaussian with finite variance, $\widehat{c}_K$ tends to the Dirac measure at c^*.

However, given that we use $x(t) = e^{i\phi(t)}$, $\widehat{\mathrm{c}}_K$ is not exactly the multitrial PLV estimate—more precisely,

$$\widehat{\mathrm{c}}_K = \frac{1}{K}\sum_{k=1}^{K}\int_0^T e^{i\phi(t)}dN^{(k)}(t) = \frac{1}{K}\sum_{k=1}^{K}\sum_{j=1}^{N_k} e^{i\phi(t_j^k)} = \frac{\left(\sum_{k=1}^{K} N_k\right)}{K}\widehat{\mathrm{PLV}}_K.$$

Thus, we can write $\widehat{\mathrm{PLV}}_K = \nu_K \cdot \widehat{\mathrm{c}}_K$, with $\nu_K = \frac{K}{\left(\sum_{k=1}^{K} N_k\right)}$. With the same techniques (using $x(t) = 1$), we can show convergence in the distribution of ν_K to a constant:

$$\frac{1}{\nu_K} = \frac{\left(\sum_{k=1}^{K} N_k\right)}{K} = 1/K\sum_k \int_0^T 1 \cdot dN^{(k)} \underset{K\to+\infty}{\longrightarrow} \int_0^T \lambda(t)dt = \Lambda(T).$$

This leads to

$$\nu_K \underset{K\to+\infty}{\longrightarrow} \frac{1}{\Lambda(T)}.$$

Following a version of Slutsky's theorem (Mittelhammer, 1996, theorem 5.10), since ν_k and $\widehat{\mathrm{c}}_K$ tend to a limit in distribution, and one of these limits is a constant, the product tends to the product of the limits such that we get

$$\mathrm{PLV}^* = \lim_{K\to\infty} \nu_K \cdot \widehat{\mathrm{c}}_K = \frac{c^*}{\Lambda(T)}$$

and can decompose the PLV residual as follows:

$$\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right) = \sqrt{K}\nu_K\left(\widehat{\mathrm{c}}_K - \mathrm{c}^*\right) + \sqrt{K}\left(\nu_K \mathrm{c}^* - \mathrm{PLV}^*\right).$$

Taking the limit of the above equation, the second term clearly vanishes (see the above limit of ν_K), and the first term, using again the limit of products, leads to the final result:

$$\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right) \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(0, \frac{1}{\Lambda(T)^2}\mathrm{Cov}(Z)\right).$$

□

Proof of Corollary 2. We use the intensity function introduced in equation 3.5 in corollary 1. The PLV asymptotic value (PLV^*) can be derived from definition introduced in equation 3.2:

$$\mathrm{PLV}^* = \frac{\int_0^T e^{i\phi(t)}\lambda(t)dt}{\int_0^T \lambda(t)dt} \tag{A.9}$$

$$= \frac{r_o \int_0^T e^{i\phi(t)} \exp(\kappa\cos(\phi(t)-\varphi_0))\phi'(t)dt}{r_o \int_0^T \exp(\kappa\cos(\phi(t)-\varphi_0))\phi'(t)dt}. \tag{A.10}$$

We change the integration variable from $\phi(t)$ to θ:

$$\mathrm{PLV}^* = \frac{\int_{\phi(0)}^{\phi(T)} e^{i\theta} \exp(\kappa\cos(\theta-\varphi_0))d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa\cos(\theta-\varphi_0))d\theta}. \tag{A.11}$$

To simplify the integral (bring the φ_0 out of the integral), we change the integration variable again, from θ to ψ, ($\psi = \theta - \varphi_0$):

$$\mathrm{PLV}^* = \frac{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} e^{i(\psi+\varphi_0)} \exp(\kappa\cos(\psi))d\psi}{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} \exp(\kappa\cos(\psi))d\psi} \tag{A.12}$$

$$= e^{i\varphi_0} \frac{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} e^{i\psi} \exp(\kappa\cos(\psi))d\psi}{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} \exp(\kappa\cos(\psi))d\psi}. \tag{A.13}$$

Given that that integrand is a 2π-periodic functions (thus, the integral is invariant to translations of the integration interval), we get

$$\mathrm{PLV}^* = e^{i\varphi_0} \frac{\int_{-\pi}^{\pi} e^{i\psi} \exp(\kappa\cos(\psi))d\psi}{\int_{-\pi}^{\pi} \exp(\kappa\cos(\psi))d\psi}.$$

Observing that the integrand of the denominator is even, while for the numerator, the imaginary part is odd and the real part is even, we get

$$\mathrm{PLV}^* = e^{i\varphi_0} \frac{\int_0^{\pi} \cos(\psi)\exp(\kappa\cos(\psi))d\psi}{\int_0^{\pi} \exp(\kappa\cos(\psi))d\psi}.$$

This proves the first part of the corollary—equation 3.6. By using the integral form of the modified Bessel functions I_k for k integer (see, e.g., Watson, 1995, p. 181):

$$I_k(\kappa) = \frac{1}{\pi}\int_0^{\pi} \cos(k\theta)\exp(\kappa\cos(\theta))d\theta + \frac{\sin(k\pi)}{\pi}\int_0^{+\infty} e^{-\kappa\cosh t - kt}dt \quad (A.14)$$

$$= \frac{1}{\pi}\int_0^{\pi} \cos(k\theta)\exp(\kappa\cos(\theta))d\theta, \quad (A.15)$$

we can derive the compact form:

$$\mathrm{PLV}^* = e^{\mathbf{i}\varphi_0}\frac{I_1(\kappa)}{I_0(\kappa)}. \quad (A.16)$$

The covariance matrix of the asymptotic distribution can be easily derived by plugging equation 3.5 as $\lambda(t)$ in corollary 1 and integrating on $[0,\ T]$:

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2}\int_0^T \cos^2(\phi(t))\exp\left(\kappa\cos(\phi(t)-\varphi_0)\right)\phi'(t)dt. \quad (A.17)$$

Based on the above developments and noticing that the integration intervals correspond to $2\pi\gamma_T$, with γ_T the number of oscillation periods, we have

$$\Lambda(T) = \lambda_0 2\gamma_T\pi I_0(\kappa) = \lambda_0 2\frac{\phi(T)-\phi(0)}{2\pi}\pi I_0(\kappa),$$

such that

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{\lambda_0\left(\phi(T)-\phi(0)\right)^2 I_0(\kappa)^2} \times \int_0^T \cos^2(\phi(t))\exp\left(\kappa\cos(\phi(t)-\varphi_0)\right)\phi'(t)dt. \quad (A.18)$$

To simplify the rest of the derivations, we transform the complex variable coordinates by using $e^{\mathbf{i}\phi(t)}e^{-\mathbf{i}\varphi_0}$ instead of $e^{\mathbf{i}\phi(t)}$ as predictable with respect to $\{\mathcal{F}_t\}$ (i.e., replacing $x(t)$ with $e^{\mathbf{i}\phi(t)}e^{-\mathbf{i}\varphi_0}$ in theorem 1). With this change, equation A.18 becomes

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{\lambda_0\left(\phi(T)-\phi(0)\right)^2 I_0(\kappa)^2} \times \int_0^T \cos^2(\phi(t)-\varphi_0)\exp\left(\kappa\cos(\phi(t)-\varphi_0)\right)\phi'(t)dt. \quad (A.19)$$

We change the variable of the integral from $\phi(t) - \varphi_0$ to θ and use the following trigonometric identity,

$$\cos^2(\theta) = \frac{1}{2}\left(1 + \cos(2\theta)\right), \tag{A.20}$$

to obtain

$$\begin{aligned}(\mathrm{Cov}(Z))_{11} &= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \int_{\phi(0)}^{\phi(T)} \left(1 + \cos(2\theta)\right) \exp\left(\kappa \cos(\theta)\right) d\theta \\ &= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \\ &\quad \times \int_{\phi(0)}^{\phi(T)} \left(\exp\left(\kappa \cos(\theta)\right) + \cos(2\theta) \exp\left(\kappa \cos(\theta)\right)\right) d\theta.\end{aligned}$$

Using again that the integration interval is $2\pi\gamma_T$ with γ_T integer, and integrates 2π-periodic functions (thus, the integral is invariant to translations of the integration interval), we get

$$\begin{aligned}(\mathrm{Cov}(Z))_{11} &= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \left[\int_0^{2\pi\gamma_T} \exp\left(\kappa \cos(\theta)\right) d\theta \right. \\ &\quad \left. + \int_0^{2\pi\gamma_T} \cos(2\theta) \exp\left(\kappa \cos(\theta)\right) d\theta \right], \\ (\mathrm{Cov}(Z))_{11} &= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \left[2\gamma_T \pi I_0(\kappa) + 2\gamma_T \pi I_2(\kappa)\right] \qquad \text{(A.21)} \\ &= \frac{2\pi\gamma_T}{2\lambda_0 \left(\phi(T) - \phi(0)\right)^2 I_0(\kappa)^2} \left[I_0(\kappa) + I_2(\kappa)\right] \qquad \text{(A.22)} \\ &= \frac{1}{2\lambda_0 \left(\phi(T) - \phi(0)\right) I_0(\kappa)^2} \left[I_0(\kappa) + I_2(\kappa)\right], \qquad \text{(A.23)}\end{aligned}$$

where γ_T is the number of oscillation periods contained in $[0,\ T]$.

We can have a similar calculation for the imaginary part, $(\mathrm{Cov}(Z))_{22}$, as well, but using the identity $\sin^2(\theta) = \frac{1}{2}\left(1 - \cos(2\theta)\right)$ instead of equation A.20. The off-diagonal elements of the covariance matrix vanish due to symmetries of integrand.

Therefore, we showed that for a given $\kappa \geq 0$, scaled residual

$$Z' = e^{-i\varphi_0} \sqrt{K} \left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right)$$

converges to a zero mean complex gaussian with the following covariance:

$$\begin{aligned}\mathrm{Cov}\begin{bmatrix}\mathrm{Re}\{Z'\}\\ \mathrm{Im}\{Z'\}\end{bmatrix} &= \begin{bmatrix}\mathrm{Re}\{Ze^{-i\varphi_0}\}\\ \mathrm{Im}\{Ze^{-i\varphi_0}\}\end{bmatrix}\\ &= \frac{1}{2\lambda_0(\phi(T)-\phi(0))I_0(\kappa)^2}\begin{bmatrix}I_0(\kappa)+I_2(\kappa) & 0\\ 0 & I_0(\kappa)-I_2(\kappa)\end{bmatrix}.\end{aligned}$$

□

Proof of Corollary 3. Similar to corollary 2, we can derive the asymptotic PLV, equation 3.9, for this case, from the definition in equation 3.2. We apply the intensity function $\lambda = \lambda_0$ in corollary 1. The PLV asymptotic value (PLV*) can be derived simply by changing the integration variable from $\phi(t)$ to θ (and let $\theta \mapsto \tau(\theta)$ be its inverse).

The covariance matrix of the asymptotic distribution, can be derived by the procedure we used for the proof of corollary 2. We plug the rate λ_0 as $\lambda(t)$ in corollary 1 and integrate on $[0,\ T]$:

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2}\int_0^T \cos^2(\phi(t))dt. \tag{A.24}$$

By changing the variable from $\phi(t)$ to θ, we get

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2}\int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta. \tag{A.25}$$

As $\Lambda(T) = \int_0^T \lambda_0 dt = \lambda_0 T$, we have

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2}\int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta \tag{A.26}$$

$$= \frac{1}{\lambda_0 T^2}\int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta. \tag{A.27}$$

With a similar calculation for other coefficients of the covariance matrix, we get

$$\mathrm{Cov}(Z) = \frac{1}{\lambda_0 T^2}\int_{\phi(0)}^{\phi(T)} \begin{bmatrix}\cos^2(\theta) & \sin(2\theta)/2\\ \sin(2\theta)/2 & \sin^2(\theta)\end{bmatrix}\tau'(\theta)d\theta.$$

Therefore, we showed that the scaled residual,

$$Z = \sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right),$$

converges to a zero mean complex gaussian:

$$\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right) \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \mathrm{Cov}(Z)\right).$$

□

Proof of Theorem 2. Similar to the proof of theorem 1, we rely on a CLT, but this time adapted to the case of vector-valued martingales (Aalen et al., 2008, appendix B) to prove this theorem.

We start from the single trial empirical vector-valued coupling measure of equation 4.1:

$$\boldsymbol{C} = \int_0^t \boldsymbol{x}(t) d\boldsymbol{N}(t)^\top. \tag{A.28}$$

As for the univariate case, under mild assumptions, we can associate a martingale with a vector-valued counting process $\boldsymbol{N}(t)$:

$$\boldsymbol{M}(t) = \boldsymbol{N}(t) - \int_0^t \boldsymbol{\lambda}(s) ds. \tag{A.29}$$

As in this theorem, we assume $\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0, t \in [0,\ T]$, we get

$$\boldsymbol{M}(t) = \boldsymbol{N}(t) - \boldsymbol{\lambda}_0 t. \tag{A.30}$$

The $(p \times n)$ matrix-valued martingale for the empirical coupling matrix of equation 4.1, resulting from stochastic integration, is

$$\boldsymbol{M}_x(t) = \int_0^t \boldsymbol{x}(s) d\boldsymbol{M}^\top(s) ds \tag{A.31}$$

and can be decomposed similarly to equation B.3 as

$$\boldsymbol{M}_x(t) = \int_0^t \boldsymbol{x}(s) d\boldsymbol{N}(s)^\top - \int_0^t \boldsymbol{x}(s) \boldsymbol{\lambda}_0 ds. \tag{A.32}$$

By generalizing the steps of theorem 1, we introduced the $(p \times n)$-variate martingale:

$$\widetilde{\boldsymbol{M}}^{(K)}(t) = 1/\sqrt{K} \sum_{k=1}^{K} \boldsymbol{M}_x^{(k)}(t) \tag{A.33}$$

$$= 1/\sqrt{K} \sum_{k=1}^{K} \int_0^t \boldsymbol{x}(s) \left(d\boldsymbol{M}^{(k)}\right)^\top (s) ds. \tag{A.34}$$

We now state the CLT theorem for multivariate stochastic integrals.

Proposition 1 (Multivariate Martingale CLT; Aalen et al., 2008, Section B.3). *Given the (real) matrix valued predictable functions $\boldsymbol{H}^{(K)}(t)$, consider the multivariate stochastic integral of multivariate martingale $\boldsymbol{M}^{(K)}$ with intensity vector $\boldsymbol{\lambda}^{(K)}(t)$:*

$$\int_0^t \boldsymbol{H}^{(K)}(u) d\boldsymbol{M}^{(K)}(u).$$

Assume:

1. $\int_0^t \boldsymbol{H}^{(K)}(u) diag\{\boldsymbol{\lambda}^{(K)}(u)\} \boldsymbol{H}^{(K)}(u)^\top du \xrightarrow{P} \boldsymbol{V}(t)$.
2. $\sum_{j=1}^k \int_0^t (\boldsymbol{H}^{(K)}(u))^2 \mathbf{1}_{|\boldsymbol{H}^{(K)}(u)|>\epsilon} \lambda_j^{(K)}(u) du \xrightarrow{P} 0$, *for all* $t \in [0.T]$ *and* $\epsilon > 0$.

The above stochastic integral converges in distribution to a mean-zero gaussian martingale of covariance $\boldsymbol{V}(t)$.

We notice that when summing across K trials (see equation A.34), deterministic signals $\boldsymbol{x}$ remain identical and point processes are pooled across K-trials. Given that trials are independent, the counting processes derived from the trial-pooled Poisson processes $\sum_{k=1}^K \boldsymbol{N}^{(k)}(t)$ are distributed as multivariate Poisson processes with intensity vector $K\boldsymbol{\lambda}_0$, such that

$$\widetilde{\boldsymbol{M}}^{(K)}(t) = 1/\sqrt{K} \int_0^t \boldsymbol{x}(s) d\boldsymbol{P}^\top(s) ds, \tag{A.35}$$

where $\boldsymbol{P}$ is the martingale associated with the pooled process,

$$\boldsymbol{P}(t) = \left(\sum_{k=1}^K \boldsymbol{N}^{(k)}(t) \right) - \int_0^t K\boldsymbol{\lambda}(s) ds. \tag{A.36}$$

Given that the coupling matrix is matrix valued, we have to vectorize it in order to apply the above CLT. Let Vec{.} be the operator that concatenates the successive columns of a matrix into a larger column vector. $\widetilde{\boldsymbol{M}}^{(K)}(t)$ is a $(p \times n)$-variate matrix-valued process, and its vectorized version, $\text{Vec}\{\widetilde{\boldsymbol{M}}^{(K)}(t)\}$, is a $(pn \times 1)$-variate vector process. We can write equation A.35 in vectorized form as

$$\text{Vec}\{\widetilde{\boldsymbol{M}}^{(K)}(t)\} = \int_0^t \boldsymbol{H}(s) d\boldsymbol{P}^\top(s) ds,$$

with the $(pn \times n)$-variate block diagonal matrix:

$$\boldsymbol{H}(s) = \frac{1}{\sqrt{K}} \begin{bmatrix} \boldsymbol{x}(s) & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}(s) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \boldsymbol{x}(s) \end{bmatrix}. \tag{A.37}$$

The variance of $\mathrm{Vec}\{\widetilde{\boldsymbol{M}}^{(K)}(t)\}$ (a $(pn \times pn)$-variate covariance matrix which is also called predictable variation process) can be written, based on proposition 1, as

$$\widetilde{\boldsymbol{V}}(t) = \int_0^t \boldsymbol{H}(s)\mathrm{diag}\left\{\boldsymbol{\lambda}(s)\right\}\boldsymbol{H}(s)^\top ds. \tag{A.38}$$

Since we assume a constant intensity function, $\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0 = \{\lambda_k\}_k$ ($(n \times 1)$-variate matrix), we can simplify equation A.38 as follows:

$$\widetilde{\boldsymbol{V}}(t) = \int_0^t \boldsymbol{H}(s)\mathrm{diag}\left\{K\boldsymbol{\lambda}_0\right\}\boldsymbol{H}(s)^\top ds. \tag{A.39}$$

Replacing $\boldsymbol{H}(s)$ with the block diagonal matrix defined in equation A.37 leads us to

$$\widetilde{\boldsymbol{V}}(t) = \frac{1}{K} \begin{bmatrix} \int_0^t K\lambda_1 \boldsymbol{x}(s)\boldsymbol{x}(s)^H ds & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \int_0^t K\lambda_2 \boldsymbol{x}(s)\boldsymbol{x}(s)^H ds & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \int_0^t K\lambda_n \boldsymbol{x}(s)\boldsymbol{x}(s)^H ds \end{bmatrix} \tag{A.40}$$

$$= \begin{bmatrix} \lambda_1 \int_0^t \boldsymbol{x}(s)\boldsymbol{x}(s)^H ds & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \lambda_2 \int_0^t \boldsymbol{x}(s)\boldsymbol{x}(s)^H ds & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \lambda_n \int_0^t \boldsymbol{x}(s)\boldsymbol{x}(s)^H ds \end{bmatrix}. \tag{A.41}$$

This fulfills condition 1 of the CLT for all $t \in [0,\ T]$. For the second condition, it is enough to see that the coeffcients of $\boldsymbol{H}$ are bounded by a term decreasing in $\frac{1}{\sqrt{K}}$. The CLT is thus satisfied, and we get convergence in the distribution to a zero-mean complex gaussian of covariance $\widetilde{\boldsymbol{V}}(t)$ for each t. Specializing the result for $t = T$, we get, based on assumption 2, a diagonal covariance matrix with block-constant diagonal coefficients,

$$\widetilde{\boldsymbol{V}}(T) = \begin{bmatrix} T\lambda_1 \boldsymbol{I}_p & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & T\lambda_2 \boldsymbol{I}_p & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & T\lambda_n \boldsymbol{I}_p \end{bmatrix}, \tag{A.42}$$

where $\boldsymbol{I}_p$ indicates the $(p \times p)$ identity matrix, which provides the covariance matrix of the (vectorized) coefficients of matrix $\sqrt{K}\widehat{\boldsymbol{C}}_K$.

Therefore, for the normalized coupling matrix, $\widehat{\boldsymbol{C}}_K \mathrm{diag}(\sqrt{T\boldsymbol{\lambda}_0})^{-1}$, the column by-column normalization, normalizes each block of the above covariance matrix by a multiplicative term $\frac{1}{T\lambda_k}$ to lead to an identity covariance. This proves convergence of the normalized coupling matrix in distribution for $K \to +\infty$ to a random matrix with i.i.d. unit variance complex gaussian coefficients (because lack of correlations implies independence in the gaussian case):

$$\sqrt{K}\mathrm{Vec}\{\widehat{\boldsymbol{C}}_K \mathrm{diag}(\sqrt{T\boldsymbol{\lambda}_0})^{-1}\} \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(\mathbf{0}_{pn}, \mathrm{I}_{\mathrm{pn}}\right). \tag{A.43}$$

Proof of Theorem 3. Based on proposition 6 in section B.3, we need only to check the four following necessary conditions, using the Kronecker delta notation of equation 4.2:

1. $\mathbb{E}\bar{X}_{jk}X_{lk} = \delta_{lj}$, for all k.
2. $\frac{1}{n}\max_{j\neq l}\mathbb{E}\left|\bar{X}_{jk}X_{lk}\right|^2 \to 0$ uniformly in $k \leq n$.
3. $\frac{1}{n^2}\sum_{\Gamma}\left(\mathbb{E}[\left(\bar{X}_{jk}X_{lk} - \delta_{lj}\right)\left(X_{j'k}\bar{X}_{l'k} - \delta_{j'l'}\right)\right)^2 \to 0$ uniformly in $k \leq n$, where $\Gamma = \{(j, l, j', l') : 1 \leq j, l, j', l' \leq p\} \setminus \{(j, l, j', l') : j = j' \neq l = l' \text{ or } j = l' \neq j' = l\}$.
4. $p/n \to \alpha \in (0, \infty)$.

Based on the same developments as theorem 2, we use the auxiliary processes

$$X_{lk}(t) = \frac{\sqrt{K}}{\sqrt{\lambda_k T}}\frac{1}{K}\int_0^t x_l(s)dP_k(s) = \frac{1}{\sqrt{K\lambda_k T}}\int_0^t x_l(s)dP_k(s) = \int_0^t H_{lk}(s)dP_k(s)$$

with P_k zero-mean martingale associated with the Poisson process of intensity $K\lambda_k$ (see equation A.36) and

$$H_{lk}(t) = \frac{x_l(t)}{\sqrt{K\lambda_k T}},$$

and will denote $X_{lk} = X_{lk}(T)$—that is, random variables that we are concerned with are the final values (at $t = T$) of those processes.

Condition 1 is a direct application of results from equation A.42 in the proof of theorem 2 because $\mathbb{E}\left[\bar{X}_{jk}X_{lk}\right]$ is the covariance between the coefficients of the normalized coupling matrix.

For condition 2, let us first evaluate

$$\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2.$$

For that, we can use Ito's formula of equation B.10 and derive the expression of $\bar{X}_{jk}X_{lk}$ as a stochastic integral, using the function $F(\bar{X}_{jk}, X_{lk}) = \bar{X}_{jk}X_{lk}$ We obtain

$$\begin{aligned}
\bar{X}_{jk}X_{lk} &= -\int_0^T \left(X_{lk}\bar{H}_{jk}(s) + \bar{X}_{jk}H_{lk}(s)\right) K\lambda_k ds \\
&\quad + \int_0^T \left[\left(\bar{X}_{jk}(s_-) + \bar{H}_{jk}(s_-)\right)\left(X_{lk}(s_-) + H_{lk}(s_-)\right)\right. \\
&\quad \left. - \bar{X}_{jk}X_{lk}(s_-)\right]\left(dP_k(s) + K\lambda_k dt\right), \\
&= \int_0^T \left(X_{lk}\bar{H}_{jk}(s_-) + \bar{X}_{jk}H_{lk}(s_-)\right) dP_k(s) \\
&\quad + \int_0^T \left[\bar{H}_{jk}(s_-)H_{lk}(s_-)\right]\left(dP_k(s) + K\lambda_k ds\right). \qquad \text{(A.44)}
\end{aligned}$$

The first term is a stochastic integral of a zero mean martingale, while the second term is a stochastic integral of a Poisson counting process, which we can verify (due to assumption 2) that it has mean δ_{ij}. As a consequence, $\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2$ is the variance of the above expression, which is (by stochastic integral formula)

$$\begin{aligned}
\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2 &= -\int_0^T \mathbb{E}\left[\left(X_{lk}(s_-)\bar{H}_{jk}(s_-) + \bar{X}_{jk}(s_-)H_{lk}(s_-)\right)^2\right] K\lambda_k ds \\
&\quad + \int_0^T \left[\bar{H}_{jk}(s_-)H_{lk}(s_-)\right]^2 K\lambda_k ds. \qquad \text{(A.45)}
\end{aligned}$$

Applying again the formula for predictable variation process, we obtain

$$\mathbb{E}\left|\bar{X}_{jk}X_{lk}-\delta_{lj}\right|^2=-\int_0^T\left[\int_0^s\left(H_{lk}(u)\bar{H}_{jk}(s_-)+\bar{H}_{jk}(u)H_{lk}(s_-)\right)^2K\lambda_k du\right]K\lambda_k ds$$
$$+\int_0^T\left[\bar{H}_{jk}(s_-)H_{lk}(s_-)\right]^2K\lambda_k ds. \tag{A.46}$$

Due to assumption 2, this expression is bounded uniformly for any values of $i,\ j, n, k$, and condition 2 is fulfilled.

For condition 3, we use the auxiliary result presented in proposition 2 to compute the required fourth-order moments:

$$\begin{aligned}
&\frac{1}{K^2\lambda_k^2}\mathbb{E}\left[\left(\bar{X}_{jk}X_{lk}\right)\left(X_{j'k}\bar{X}_{l'k}\right)\right]=\int_0^T H_{lk}H_{j'k}ds\int_0^T\bar{H}_{jk}\bar{H}_{l'k}ds\\
&+\int_0^T H_{lk}\bar{H}_{jk}ds\int_0^T H_{j'k}\bar{H}_{l'k}ds+\int_0^T H_{lk}\bar{H}_{l'k}ds\int_0^T\bar{H}_{jk}H_{j'k}ds\\
&+\frac{1}{K\lambda_k}\int_0^T H_{lk}\bar{H}_{jk}H_{j'k}\bar{H}_{l'k}ds\\
&=\frac{1}{\lambda_k^2T^2K^2}\left[\int_0^T x_l x_{j'}ds\int_0^T\bar{x}_j\bar{x}_{l'}ds+\int_0^T x_l\bar{x}_j ds\int_0^T x_{j'}\bar{x}_{l'}ds\right.\\
&\left.+\int_0^T x_l\bar{x}_{l'}ds\int_0^T\bar{x}_{jk}x_{j'k}ds\right]+\frac{1}{K^3\lambda_k^3T^2}\int_0^T x_l\bar{x}_j x_{j'}\bar{x}_{l'}ds.
\end{aligned}$$

We first consider the term consisting in all products of two integrals, which we call *integral product term*; the last term in this expression will be dealt with independently. Given assumption 2, it is clear that for $l,\ j,\ j', l'$, all different from each other, the integral product term is vanishing. If there happen to be only two indices that are equal, the moment also vanishes (at least one term of each product vanishes). For the case $j=l=k'=l'$, the integral product term possibly does not vanish, but is uniformly bounded, and only n terms satisfy this relation, such that it will not affect the limit of the relevant expression for condition 3 (due to the $1/n^2$ factor).

It remains the case in which three indices exactly are identical. In such a case, one among δ_{jl} or $\delta_{j'l'}$ is one while the other is zero. Take $\delta_{jl}=1$ and $\delta_{j'l'}=0$ without loss of generality, assuming $j=l=j'\neq l'$. The relevant quantity of condition 3 is

$$\begin{aligned}
&\frac{1}{K^2\lambda_k^2}\mathbb{E}\left[\left(\bar{X}_{jk}X_{lk}-1\right)\left(X_{j'k}\bar{X}_{l'k}\right)\right]\\
&=\frac{1}{K^2\lambda_k^2}\mathbb{E}\left[\left(\bar{X}_{jk}X_{lk}\right)\left(X_{j'k}\bar{X}_{l'k}\right)\right]-\frac{1}{K^2\lambda_k^2}\mathbb{E}\left[X_{j'k}\bar{X}_{l'k}\right]
\end{aligned}$$

$$= \frac{1}{\lambda_k^2 T^2 K^2}\left[\int_0^T x_l x_{j'} ds \int_0^T \bar{x}_j \bar{x}_{l'} ds + \int_0^T \left(x_l \bar{x}_j - T\right) ds \int_0^T x_{j'} \bar{x}_{l'} ds \right.$$
$$\left. + \int_0^T x_l \bar{x}_{l'} ds \int_0^T \bar{x}_{jk} x_{j'k} ds\right] + \frac{1}{K^3 \lambda_k^3 T^2} \int_0^T x_l \bar{x}_j x_{j'} \bar{x}_{l'} ds,$$

in which, due to assumption 2, the integral product term still vanishes. As a consequence, the asymptotic behavior we are interested in is given by the behavior of the remaining single integral term of the moment: $\frac{1}{K\lambda_k}\int_0^T x_l \bar{x}_j x_{j'} \bar{x}_{l'} ds$ (the only remaining nonvanishing terms are bounded and intervene only in n terms of the sum), such that

$$\lim \frac{1}{n^2} \sum_{\Gamma} \left(\mathbb{E}\left[\left(\bar{X}_{jk} X_{lk} - \delta_{lj}\right)\left(X_{j'k} \bar{X}_{l'k} - \delta_{j'l'}\right)\right]\right)^2$$
$$= \lim \frac{1}{n^2 K^2 \lambda_k^2} \sum_{\Gamma} \left(\mathbb{E}\left[\left(\bar{X}_{jk} X_{lk} - \delta_{lj}\right)\left(X_{j'k} \bar{X}_{l'k} - \delta_{j'l'}\right)\right]\right)^2. \tag{A.47}$$

Thus condition 3 is satisfied due to the theorem's assumption.

To sum up, all four necessary conditions for the application of proposition 6 are fulfilled (condition 4 is part of the assumptions), and the convergence to the MP law follows immediately. □

Proof of Theorem 4. Let us use the result of Chafaï and Tikhomirov (2018) adapted to our complex case and adapt the dimension notation ($n \to p(n)$, $m_n \to n$, but we keep the notation X_n). We additionally checked in all proofs and lemmas that the result still holds when we replace symmetric matrices by Hermitian ones and the scalar product of real vectors by Hermitian products of complex vectors, putting an absolute value on the Hermitian product when the original scalar product was squared. We consider $\{X_n\}$, a sequence of isotropic (i.e. identity covariance) zero mean random vectors and consider the empirical covariance matrix estimated from observing n independent copies of X_n,

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{X}_n^{(k)} \boldsymbol{X}_n^{(k)H}.$$

We rely on the strong tail projection property (STP) that guarantees convergence of the spectral measure of the empirical covariance to the MP law, and convergence of the extreme eigenvalues to the ends of the MP support.

Definition 1 (Strong Tail Projection Property (STP)). *STP holds when there exist* $f : \mathbb{N} \to [0, 1]$, $g : \mathbb{N} \to \mathbb{R}^+$ *such that* $f(r) \to 0$ *and* $g(r) \to 0$ *as* $r \to \infty$, *and for every* $p \in \mathbb{N}$, *for any orthogonal projection* $P : \mathbb{C}^p \to \mathbb{C}^p$ *of rank* $r > 0$, *for any real* $t > f(r).r$ *we have*

$$\mathbb{P}\left(\|P\boldsymbol{X}_n\|^2 - r \geq t\right) \leq \frac{g(r)r}{t^2}.$$

By noting that $\mathbb{E}\,\|PX_n\|^2 = r$, we can use Chebyshev's inequality to satisfy such property. Let σ^2 be the variance of $\|P\boldsymbol{X}_n\|^2$. The inequality leads to, for any t,

$$\mathbb{P}\left(\|P\boldsymbol{X}_n\|^2 - r \geq \sigma t\right) \leq \mathbb{P}\left(\left|\|P\boldsymbol{X}_n\|^2 - r\right| \geq \sigma t\right) \leq \frac{1}{t^2},$$

so we get $\mathbb{P}\left(\|P\boldsymbol{X}_n\|^2 - r \geq t\right) \leq \sigma^2/t^2$ and just need to find an upper bound of σ^2 of the form $g(r)r$. To limit the complexity of the rank-dependent analysis, we will look for $g(r) = C/r$ for a fixed positive constant C, such that we just need to bound the above variance by a constant. Finer bounds are likely possible but left to future work.

In our specific case, in line with the proof of theorem 3, we use

$$\boldsymbol{X}_n = \int_0^T \frac{\boldsymbol{x}(t)}{\sqrt{K\lambda T}} dP(t),$$

with P the compensated Poisson process martingale of rate $K\lambda$. In an orthonormal basis adapted to the othogonal projection P with rank r, we can rewrite

$$\|PX_n\|^2 = \sum_{k=1}^{r} |\langle \boldsymbol{w}_k,\, \boldsymbol{X}_n\rangle|^2\,,$$

where $\{\boldsymbol{w}_k\}$ are r orthonormal vectors in $\mathbb{C}^p$. Then we have

$$\sigma^2 = \sum_{k,l\leq r} \mathbb{E}\left[|\langle \boldsymbol{w}_k,\, \boldsymbol{X}_n\rangle|^2\, |\langle \boldsymbol{w}_l,\, \boldsymbol{X}_n\rangle|^2 - 1\right].$$

Using similar fourth-order moment results as in theorem 3 (based on proposition 2) leads to an expansion for which all terms vanish but one per expectation, leading to

$$\sigma^2 = \frac{1}{K\lambda T^2} \sum_{k,l\leq r} \int_0^T \left\langle \boldsymbol{w}_k,\, \boldsymbol{x}(t)\right\rangle \left\langle \boldsymbol{x}(t),\, \boldsymbol{w}_k\right\rangle \left\langle \boldsymbol{w}_l,\, \boldsymbol{x}(t)\right\rangle \left\langle \boldsymbol{x}(t),\, \boldsymbol{w}_l\right\rangle dt,$$

which can be rewritten using the Hermitian operator $\mathcal{X}$ acting on the space of $p \times p$ matrices as a positive definite bilinear form,

$$\mathcal{X}(U, V) = \int_0^T \left\langle V,\, \boldsymbol{x}\boldsymbol{x}^H(t)\right\rangle \left\langle \boldsymbol{x}\boldsymbol{x}^H(t),\, U\right\rangle dt,$$

with associated eigenvalues $\xi_1 \geq ... \geq \xi_{p^2} \geq 0$ such that

$$\sigma^2 = \frac{1}{K\lambda T^2} \sum_{k,l \leq r} \mathcal{X}\left(\boldsymbol{w}_k \boldsymbol{w}_l^H, \boldsymbol{w}_k \boldsymbol{w}_l^H\right).$$

This sum is maximized when the r^2 unitary tensor matrices of the sum $\boldsymbol{w}_k \boldsymbol{w}_l$ are eigenvectors associated with the largest eigenvalues of the operator, such that we get

$$\sigma^2 \leq \frac{1}{K\lambda T^2} \sum_{k=1 \leq r^2} \xi_k,$$

which is itself upper bounded by the trace of the operator, leading to

$$\sigma^2 \leq \frac{1}{K\lambda T^2} \sum_{k,l \leq p(n)} \int_0^T |x_k x_l|^2 dt,$$

which is bounded according to the theorem's assumptions, completing the proof.

Appendix B: Additional Background and Useful Results

B.1 Jump Processes. Jump processes exhibit discontinuities related to the occurrence of random events, which are distributed according to the given point process models. In this letter, we are concerned with jump times distributed according to (possibly inhomogeneous) Poisson processes.

B.1.1 Martingales Related to Counting Processes. As introduced in section 2.2 (see equation 2.3), under mild assumptions, we can associate a zero-mean martingale with a counting process $N(t)$:

$$M(t) = N(t) - \int_0^t \lambda(s) ds. \tag{B.1}$$

In addition, in our case (deterministic intensity), the variance of $M(t)$ is given by

$$V(t) = \mathbb{E}\left[M(t)^2\right] = \int_0^t \lambda(s) ds.$$

B.1.2 Stochastic Integrals. Now, if we consider for a deterministic predictable process H (with regard to the same filtration $\mathcal{F}_t$), the stochastic integration

$$M_H(t) = \int_0^t H(s)dM(s)ds. \tag{B.2}$$

Using equation B.1, we can write

$$M_H(t) = \int_0^t H(s)dN(s) - \int_0^t H(s)\lambda(s)ds, \tag{B.3}$$

which is equivalent to equation 2.5, which introduced the separation of the deterministic component of empirical coupling measure from the (zero-mean) random fluctuations of the measure. $M_H(t)$ is also a zero-mean martingale with respect to history $\{\mathcal{F}_t\}$. This trivially entails that $\mathbb{E}[M_H(t)] = 0$ at all times.

A.1.3 Second Order Statistics. In addition, the second-order statistics of such stochastic integrals can be explicitly derived from the original intensities. In particular, for $M_H(t) = \int_0^t H(s)dM(s)ds$, we have the variance

$$V_H(t) = \mathbb{E}\left[M_H(t)^2\right] = \int_0^t H(s)^2\lambda(s)ds, \tag{B.4}$$

which corresponds to its predictable variation process (see Aalen et al., 2008, sec. 2.2.6). A similar result applies to covariance as well. Let G and H be deterministic predictable; then

$$V_{H,G}(t) = \mathbb{E}[M_H(t)M_G(t)] = \int_0^t H(s)G(s)\lambda(s)ds. \tag{B.5}$$

Importantly, we note that this nonvanishing covariance reflects the fact that both stochastic integrals are computed from the same realization of $M(t)$. If two stochastic integrals are derived from independent point processes, the resulting covariance between them is zero.

B.1.4 General Jump Stochastic Processes. For the proofs of our results, it is convenient to state some general results for jump processes that combine deterministic and a jump stochastic integral, decomposable as

$$X(t) = X(0) + \int_0^t f(X(s), s)ds + \int_0^t h(X(s), s)dN(s), \tag{B.6}$$

with $N(t)$ a Poisson process with intensity $\lambda(t)$, f and h square integrable. This clearly includes the martingales defined above.

B.1.5 Mean Stochastic Jump Integrals. According to Hanson (2007, theorem 3.20), we can compute the expectation of $X(t)$ defined in equation B.6:

$$\mathbb{E}[X(t)] = \mathbb{E}[X(0)] + \int_0^t f(X(s), s)ds + \int_0^t \mathbb{E}\left[h(X(s), s)\right] \lambda(s)ds. \quad \text{(B.7)}$$

This allows retrieval of the zero-mean property of the stochastic integral of martingales.

B.1.6 Itô's Formula. Itô's formula or Itô's lemma is an identity to find the differential of a function of a stochastic process. It is a counterpart of the chain rule used to compute the differential of composed functions. We restrict ourselves to the case of a time-independent scalar function of a jump process, while different formulas exist for other cases.

A generalized chain rule for the time derivative of such processes allows deriving an integral formula for scalar process $Y(t) = F(X(t))$ with F continuously differentiable (see Hanson, 2007, lemma 4.22, rule 4.23):

$$\begin{aligned} Y(t) = Y(0) &+ \int_0^t \frac{dF}{dx}(X(s)) f(X(s), s)ds \\ &+ \int_0^t \left[F\left(X(s_-) + h(X(s_-), s)\right) - F(X(s_-))\right] dN(s), \end{aligned} \quad \text{(B.8)}$$

where $X(s_-) = \lim_{t \to s_-} X(t)$ indicates the left limit.

For a scalar function of a multivariate process $Y(t) = F(\boldsymbol{X}(t))$ with

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) + \int_0^t \boldsymbol{f}(\boldsymbol{X}(s), s)ds + \int_0^t \boldsymbol{h}(\boldsymbol{X}(s), s)dN(s), \quad \text{(B.9)}$$

the generalization is straightforward:

$$\begin{aligned} Y(t) = Y(0) &+ \int_0^t \sum_k \frac{dF}{dx_k}(\boldsymbol{X}(s)) f_k(\boldsymbol{X}(s), s)ds \\ &+ \int_0^t \left[F\left(\boldsymbol{X}(s_-) + \boldsymbol{h}(\boldsymbol{X}(s_-), s)\right) - F(\boldsymbol{X}(s_-))\right] dN(s). \end{aligned} \quad \text{(B.10)}$$

This allows retrieving the expression of martingale second-order statistics presented above, as well as computing higher-order moments required in the proof of theorem 3.

An application of this formula that we will use follows:

Proposition 2. *Assume that* $W(t) = \int_0^t A(s)dM(s)$, $X(t) = \int_0^t B(s)dM(s)$, $Y(t) = \int_0^t C(s)dM(s)$, *and* $Z(t) = \int_0^t D(s)dM(s)$ *are stochastic integrals with respect to the same (possibly inhomogeneous) Poisson process martingale* $M(t) = N(t) - \int_0^t \lambda(s)ds$ *with intensity* $\lambda(t)$. *Then*

$$\begin{aligned}\mathbb{E}\left[WXYZ\right](t) = &\int_0^t ABCD(s_-)\lambda(s)ds \\ &+ \left(\int_0^t AB(s)\lambda(s)ds\right)\left(\int_0^t CD(s)\lambda(s)ds\right) \\ &+ \left(\int_0^t AC(s)\lambda(s)ds\right)\left(\int_0^t BD(s)\lambda(s)ds\right) \\ &+ \left(\int_0^t AD\lambda(s)(s)ds\right)\left(\int_0^t BC(s)\lambda(s)ds\right). \qquad \text{(B.11)}\end{aligned}$$

Proof. We apply the above formula to $F(W, X, Y, Z) = WXYZ$, yielding

$$\begin{aligned}WXYZ(t) = &-\int_0^t (AXYZ(s) + WBYZ(s) + WXCZ(s) + WXYD(s))\lambda ds \\ &+ \int_0^t [(W(s_-) + A)(X(s_-) + B)(Y(s_-) + C)(Z(s_-) + D) \\ &- WXYZ(s_-)]dN(s).\end{aligned}$$

Expanding the second term, we obtain the formula

$$\begin{aligned}WXYZ(t) = &\int_0^t (AXYZ(s) + WBYZ(s) + WXCZ(s) + WXYD(s))\, dM(s) \\ &+ \int_0^t (ABYZ(s_-) + AXCZ(s_-) + AXYD(s_-) + WBCZ(s_-) \\ &+ WBYD(s_-) + WXCD(s_-))dN(s) + \int_0^t ABCD(s_-)dN(s) \\ &+ \int_0^t (ABCZ(s_-) + AXCD(s_-) + ABYD(s_-) + WBCD(s_-))dN(s).\end{aligned}$$

The first and last integral terms in this formula have vanishing expectation, the first because it is a stochastic integral of zero mean martingale M, the last because each term inside the integral contains only one random

variable, which is itself a stochastic integral of the martingale M (and thus zero mean). Thus, for the expectation, we get

$$\begin{aligned}\mathbb{E}[WXYZ](t) = \int_0^t ABCD(s_-)d\lambda(s) + \int_0^t (AB\mathbb{E}YZ(s_-) + AC\mathbb{E}XZ(s_-) \\ + AD\mathbb{E}XY(s_-) + BC\mathbb{E}WZ(s_-) + BD\mathbb{E}WY(s_-) \\ + CD\mathbb{E}WX(s_-))\lambda(s)ds. \end{aligned} \tag{B.12}$$

Based on the Itô integral formula, one can easily derive an expression for the expectation of each product of two variables (see equation A.44), leading to, after reordering the terms,

$$\begin{aligned}\mathbb{E}[WXYZ](t) = \int_0^t ABCD(s_-)d\lambda(s) + \int_0^t \Bigg(AB(s_-)\int_0^s CD(u_-)\lambda(u)du \\ + CD(s_-)\int_0^s AB(u_-)\lambda(u)du + AC(s_-)\int_0^s BD(u_-)\lambda(u)du \\ + BD(s_-)\int_0^s AC(u_-)\lambda(u)du + AD(s_-)\int_0^s BC(u_-)\lambda(u)du \\ + BC(s_-)\int_0^s AD(u_-)\lambda(u)du\Bigg)\lambda(s)ds. \end{aligned} \tag{B.13}$$

We then observe that the terms inside the integral can be paired such that the integral form of the product derivative formula ($\int f \int g = \int \left(g \int f + f \int f\right)$) can be applied, leading directly to equation B.11. □

B.2 Notions of Convergence. In contrast to finite-dimensional vectors, there are different and nonequivalent notions of convergence for functions and random variables. We explain the two types of convergence encountered in this letter. For a random variable X, we consider its probability measure μ_X such that

$$\mu_X(A) = P(X \in A),$$

and its associated cumulative distribution function (CDF),

$$F_X(x) = \mu_X((-\infty, x]) = P(X \leq x).$$

B.2.1 Convergence in Distribution. The classical definition is based on the CDF.

Definition 2 (Convergence in Distribution). *We say that the sequence of random variables $\{X_n\}$ converges in distribution (or in law) to X whenever*

$$F_{X_n}(x) \underset{n\to+\infty}{\longrightarrow} F_X,$$

at all continuity points of F_X. This is then denoted $X_n \xrightarrow{D} X$.

An equivalent definition can be formulated in terms of weak convergence:

Proposition 3. *$X_n \xrightarrow{D} X$ if and only if, for any bounded continuous function f,*

$$\mathbb{E}\big[f(X_n)\big] = \int f d\mu_{X_n} \to \mathbb{E}\big[f(X)\big] = \int f d\mu_X,$$

that is, in classical topological terms, the measure μ_{X_n} converges weakly to μ_X.

The generalization to multidimensional variables encountered in theorem 2 consists simply in replacing the cumulative distribution by its multivariate version, $F_{\boldsymbol{X}}(x) = P(X_1 < x_1, \ldots, X_n < x_n)$, in definitions. A simple necessary and sufficient condition for $\boldsymbol{X} \longrightarrow \boldsymbol{Y}$ is that for all vectors $\boldsymbol{t}$, $\boldsymbol{t}^\top \boldsymbol{X} \longrightarrow \boldsymbol{t}^\top \boldsymbol{Y}$ (this is the Cramér-Wold theorem, see Billingsley, 1995).

B.2.2 Convergence in Probability. This stronger notion of convergence denotes $X_n \xrightarrow{P} X$, stating that for any $\epsilon > 0$,

$$P\left(|X_n - X| > \epsilon\right) \underset{n\to+\infty}{\longrightarrow} 0. \tag{B.14}$$

It can be shown that convergence in probability implies convergence in distribution. The converse is true only in special cases:

Proposition 4. *If X converges in distribution to a (deterministic) constant c, then it also converges to it in probability.*

An extension to the multivariate case is obtained in finite vector spaces by replacing the absolute value in equation B.14 by any norm, or simply by requiring the convergence of all components individually.

B.2.3 Convergence of Random Measures. The ESDs are random measures, and as such, random variables, leaving in an infinite-dimensional space of measures. This means that for a fixed realization ω, the random measure $\boldsymbol{\mu}$ takes the deterministic value $\boldsymbol{\mu}(\omega)$.

Several types of convergence can be defined. First, the notion of *convergence weakly in probability* can be seen as a combination of the above definitions. It is known that the weak convergence of deterministic measures (see proposition 3) can be associated with a (nonunique) metric (the topological

space of weak convergence is metrizable). Let us pick such a metric $\rho(\mu, \nu)$ between two deterministic measures; then:

Definition 3 (Convergence Weakly in Probability). *The sequence of random measures $\boldsymbol{\mu}_n$ converges weakly in probability to the deterministic measure ν for any $\epsilon > 0$:*

$$P\left(\rho(\boldsymbol{\mu}_n, \nu) > \epsilon\right) \underset{n\to+\infty}{\longrightarrow} 0. \tag{B.15}$$

Next, we can also define convergence with probability 1 (also called *almost sure convergence*).

Definition 4 (Convergence (Weakly) with Probability One). *The sequence of random measures $\boldsymbol{\mu}_n$ converges weakly with probability one to the deterministic measure ν for any $\epsilon > 0$:*

$$P\left(\rho(\boldsymbol{\mu}_n(\omega), \nu) \underset{n\to+\infty}{\longrightarrow} 0\right) = 1. \tag{B.16}$$

As for the case of scalar random variables, convergence with probability one implies convergence in probability.

B.3 Random Matrix Theory. Random matrix theory resulted from fairly recent developments in high-dimensional statistics. It has various application in physics (Guhr, Müller-Groeling, & Weidenmüller, 1998; Dousal, Majumdar, & Schehr, 2016), machine learning (Pennington & Bahri, 2017; Pennington & Worah, 2017; Louart et al., 2018), and neuroscience (Timme, Geisel, & Wolf, 2006; Veraart et al., 2016; Almog et al., 2019).

B.3.1 Wishart Ensemble. Let $\boldsymbol{X}$ be a $p \times n$ data matrix. Assume that the coefficients of $\boldsymbol{X}$, x_{ij} are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$. $\mathcal{N}_{\mathbb{C}}$ specifies a standard complex normal distribution. By definition, this means that $x_{ij} = x_{ij}^{real} + \boldsymbol{i}x_{ij}^{imag}$, where $x_{ij} = x_{ij}^{real}$ and x_{ij}^{imag} are independent (real) $\mathcal{N}(0, \frac{1}{2})$. This implies that columns of $\boldsymbol{X}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_p, \mathrm{I_p}\right)$ and, similarly, the real and imaginary parts are $\mathcal{N}\left(\mathbf{0}_p, \mathrm{I_p}/2\right)$.

As n grows and $\frac{p}{n} \underset{n\to+\infty}{\to} \alpha \in (0, +\infty)$, the ESD of the so-called Wishart ensemble, $\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^H$, converges to the Marchenko-Pastur law $\mu_{MP}(x)$ (Marchenko & Pastur, 1967) with density

$$\frac{d\mu_{MP}}{dx}(x) = \frac{1-\alpha}{\alpha}\mathbf{1}_{\alpha>1}\delta_0 + \frac{1}{2\pi\alpha x}\sqrt{(b-x)(x-a)}\mathbf{1}_{[a,b]}, \tag{B.17}$$

with $a = (1-\sqrt{\alpha})^2$ and $b = (1+\sqrt{\alpha})^2$ (see examples for Marchenko-Pastur law for different values of α in Figure 6).

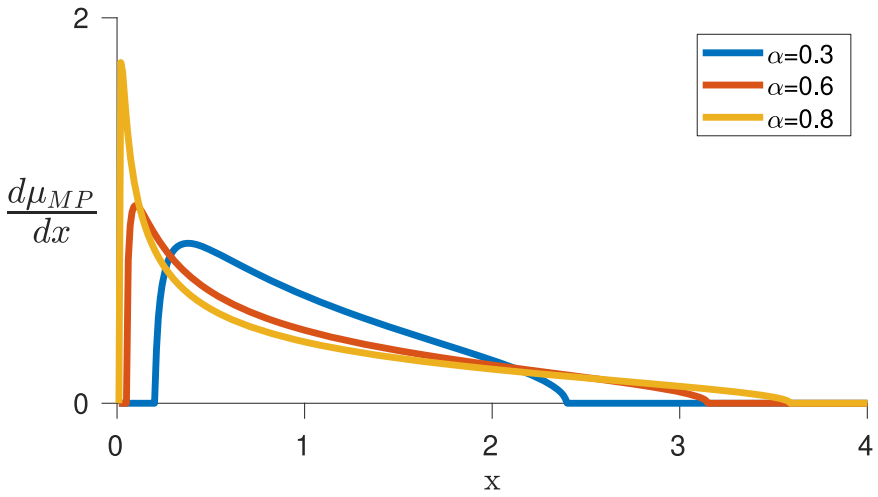


Figure 6: Density of the Marchenko-Pastur law for different values of the aspect ratio of the matrices, α, in equation 2.7.

We wrote here the general formula that holds for all $\alpha > 0$, accounting for zero eigenvalues with a Dirac mass in zero in the rank-deficient case $\alpha > 1$.

B.3.2 Stieltjes Transform of ESD. The Stieltjes transform is a very useful tool to establish the convergence of ESD and determine its limit. The Stieltjes transform of a measure μ is defined as

$$m_\mu(z) = \int \frac{1}{x - z} d\mu(x), z \in \mathbb{C} \setminus \mathbb{R}.$$

A key example for us is the Stieltjes transform of the MP law:

$$m(z) = \frac{1 - c - z + \sqrt{(1 + c - z)^2 - 4c}}{2cz}.$$

Many important results relate measures to their Stieltjes transform. We only need the property that the Stieltjes transform identifies the limit of a sequences of measures, with the following proposition that immediately derives from Anderson et al. (2010, theorem 2.4.4).

Proposition 5. *If two sequences of random measures $\{\boldsymbol{\mu}_k\}$ and $\{\boldsymbol{\nu}_k\}$ converge weakly in probability to a deterministic with identical Stieltjes transform, they converge to the same measure.*

B.3.3 Convergence to MP for Matrices with Dependent Coefficients. Based on the above, we can now write a result that is a combination of results found in Bai and Zhou (2008—mainly theorem 1.1 and corollary 1.1) adapted to

our specific case. We consider a sequence of random matrices $\{\boldsymbol{X}_n\}$ with independent columns and study the ESD of

$$\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{X}_n\boldsymbol{X}_n^H.$$

In the following proposition, we use the Kronecker delta symbold delta (see equation 4.2) and denote by $\bar{X}$ the complex conjugate of X.

Proposition 6. *Let As $n \to \infty$, and assume the following. Let*

1. $\mathbb{E}\bar{X}_{jk}X_{lk} = \delta_{lj}$, *for all k.*
2. $\frac{1}{n}\max_{j\neq l}\mathbb{E}\left|\bar{X}_{jk}X_{lk} - \delta_{lj}\right|^2 \to 0$ *uniformly in* $k \leq n$.
3. $\frac{1}{n^2}\sum_{\Gamma}\left(\mathbb{E}\left(\bar{X}_{jk}X_{lk} - \delta_{lj}\right)\left(X_{j'k}\bar{X}_{l'k} - \delta_{j'l'}\right)\right)^2 \to 0$ *uniformly in* $k \leq n$, *where* $\Gamma = \{(j, l, j', l') : 1 \leq j, l, j', l' \leq p\} \setminus \{(j, l, j', l') : j = j' \neq l = l' \text{ or } j = l' \neq j' = l\}$.
4. $p/n \to \alpha \in (0, \infty)$.

Then, with probability 1, the ESD of $\boldsymbol{S}_n$ tends (weakly) to the MP law.

Sketch of the proof. We use theorem 1.1 from Bai and Zhou (2008) combined with the sufficient condition of corollary 1.1, assuming the identity matrix $\boldsymbol{T}_n$. These conditions are compatible with the case of the Wishart ensemble, such that the ESD converges to a distribution with the same Stieltjes transform as the MP law.[9] As a consequence of proposition 5, we get that the limit ESD is the MP law. □

Appendix C: Additional Corollaries

The additional results in this appendix are corollaries based on simplifying assumption 3, where a linear phase is considered instead of the general assumption on phase that was used in corollaries 2 and 3.

Assumption 3. Assume that $\phi(t)$ is a linear function of t on $[0,\ T]$,

$$\phi(t) = mt, \quad m = 2\pi f = 2\pi/\tau, \tag{C.1}$$

where $f > 0$ (interpretable as the frequency of an oscillation for the continuous signal) and γ_T is the ratio of length (T) of signal to period of oscillation τ:

$$\gamma_T = \frac{T}{\tau} = \frac{\phi(T) - \phi(0)}{2\pi}.$$

[9] This requires checking that the self-consistency equation 1.1 in Bai and Zhou (2008) has a unique solution, which they establish by equation 1.2.

Corollary 4. *Under the assumptions of corollary 2, assume additionally assumption 3 is also satisfied, and the intensity of the point-process is given by*

$$\lambda(t) = \lambda_0 \exp(\kappa \cos(\phi(t) - \varphi_0)), \tag{C.2}$$

for a given $\kappa \geq 0$. Then the expectation of the multitrial PLV estimate converges (for $K \to +\infty$) to

$$\mathrm{PLV}^* = \frac{\int_0^T e^{i2\pi ft} \exp(\kappa \cos(2\pi ft - \varphi_0))dt}{\int_0^T \exp(\kappa \cos(2\pi ft - \varphi_0))dt}. \tag{C.3}$$

If, in addition, $[0, T]$ corresponds to an integer number $\gamma_T > 0$ of periods of the oscillation,

$$\mathrm{PLV}^* = e^{i\varphi_0} \frac{\int_{\phi(0)}^{\phi(T)} \cos(\theta) \exp(\kappa \cos(\theta))d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa \cos(\theta))d\theta} = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}, \tag{C.4}$$

and the scaled residual $\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right)$ converges to a zero mean complex gaussian Z with the following covariance:

$$\mathrm{Cov}\begin{bmatrix} Re\{Ze^{-i\varphi_0}\} \\ Im\{Ze^{-i\varphi_0}\} \end{bmatrix} = \frac{1}{2\lambda_0 T I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}. \tag{C.5}$$

Proof. We use the intensity function introduced in equation C.2. The PLV asymptotic value (PLV^*) can be derived from definition introduced in equation 3.2 by using assumption 3:

$$\mathrm{PLV}^* = \frac{\int_0^T e^{i\phi(t)} \lambda(t)dt}{\int_0^T \lambda(t)dt} \tag{C.6}$$

$$= \frac{\lambda_0 \int_0^T e^{i\phi(t)} \exp(\kappa \cos(\phi(t) - \varphi_0))dt}{\lambda_0 \int_0^T \exp(\kappa \cos(\phi(t) - \varphi_0))dt} \tag{C.7}$$

$$= \frac{\lambda_0 \int_0^T e^{imt} \exp(\kappa \cos(mt - \varphi_0))dt}{\lambda_0 \int_0^T \exp(\kappa \cos(mt - \varphi_0))dt}. \tag{C.8}$$

We change the integration variable from mt to θ:

$$\mathrm{PLV}^* = \frac{\int_{\theta(0)}^{\theta(T)} e^{i\theta} \exp(\kappa \cos(\theta - \varphi_0))d\theta}{\int_{\theta(0)}^{\theta(T)} \exp(\kappa \cos(\theta - \varphi_0))d\theta}. \tag{C.9}$$

To simplify the integral (bring the φ_0 out of the integral), we change the integration variable again, from θ to ψ, ($\psi = \theta - \varphi_0$),

$$\mathrm{PLV}^* = \frac{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} e^{i(\psi+\varphi_0)} \exp(\kappa \cos(\psi))d\psi}{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} \exp(\kappa \cos(\psi))d\psi} \tag{C.10}$$

$$= e^{i\varphi_0} \frac{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} e^{i\psi} \exp(\kappa \cos(\psi))d\psi}{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} \exp(\kappa \cos(\psi))d\psi}. \tag{C.11}$$

When $[0,\ T]$ corresponds to an integer number of periods of the oscillation (i.e., is an integer number), and given that the integration interval is $2\pi\gamma_T$, and integrates 2π-periodic functions (thus the integral is invariant to translations of the integration interval), we have

$$\mathrm{PLV}^* = e^{i\varphi_0} \frac{\int_{-\pi}^{\pi} e^{i\psi} \exp(\kappa \cos(\psi))d\psi}{\int_{-\pi}^{\pi} \exp(\kappa \cos(\psi))d\psi}.$$

Observing that the integrand of the denominator is even, while for the numerator the imaginary part is odd and the real part is even, we get

$$\mathrm{PLV}^* = e^{i\varphi_0} \frac{\int_{0}^{\pi} \cos(\psi) \exp(\kappa \cos(\psi))d\psi}{\int_{0}^{\pi} \exp(\kappa \cos(\psi))d\psi}.$$

We prove the first part of the corollary, equation C.3. By using the integral form of the modified Bessel functions I_k for k integer (see Watson, 1995, p. 181)

$$I_k(\kappa) = \frac{1}{\pi} \int_0^{\pi} \cos(k\theta) \exp(\kappa \cos(\theta))d\theta + \frac{\sin(k\pi)}{\pi} \int_0^{+\infty} e^{-\kappa \cosh t - kt} dt \tag{C.12}$$

$$= \frac{1}{\pi} \int_0^{\pi} \cos(k\theta) \exp(\kappa \cos(\theta))d\theta, \tag{C.13}$$

we can derive the compact form:

$$\mathrm{PLV}^* = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}. \tag{C.14}$$

The covariance matrix of the asymptotic distribution can be easily derived by plugging equation C.2 as $\lambda(t)$ in corollary 1 and integrating on $[0,\ T]$

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(\phi(t)) \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) dt. \tag{C.15}$$

As we have

$$\Lambda(T) = \lambda_0 T I_0(\kappa),$$

we can continue with equation C.15 as,

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{\lambda_0 T^2 I_0(\kappa)^2} \int_0^T \cos^2(\phi(t)) \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) dt. \tag{C.16}$$

To simplify the rest of the derivations, we transform the complex variable coordinates by using $e^{i\phi(t)}e^{-i\varphi_0}$ instead of $e^{i\phi(t)}$ as predictable with respect to $\{\mathcal{F}_t\}$ (i.e., replacing $x(t)$ with $e^{i\phi(t)}e^{-i\varphi_0}$ in theorem 1). With this change, equation C.16 becomes

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{\lambda_0 T^2 I_0(\kappa)^2} \int_0^T \cos^2(\phi(t) - \varphi_0) \exp\left(\kappa \cos(\phi(t) - \varphi_0)\right) dt. \tag{C.17}$$

Then we change the variable of the integral from $mt - \varphi_0$ to θ (and consequently dt to $\frac{1}{m}d\theta$) and use the following trigonometric identity,

$$\cos^2(\theta) = \frac{1}{2}\left(1 + \cos(2\theta)\right), \tag{C.18}$$

to obtain

$$\begin{aligned}(\mathrm{Cov}(Z))_{11} &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \int_{\theta(0)}^{\theta(T)} \left(1 + \cos(2\theta)\right) \exp\left(\kappa \cos(\theta)\right) d\theta \\ &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \int_{\theta(0)}^{\theta(T)} \left(\exp\left(\kappa \cos(\theta)\right) + \cos(2\theta) \exp\left(\kappa \cos(\theta)\right)\right) d\theta.\end{aligned}$$

Given that the integral is invariant to translations of the integration, we get

$$\begin{aligned}(\mathrm{Cov}(Z))_{11} &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \left[\int_0^{2\pi\gamma_T} \exp\left(\kappa \cos(\theta)\right) d\theta \right. \\ &\qquad \left. + \int_0^{2\pi\gamma_T} \cos(2\theta) \exp\left(\kappa \cos(\theta)\right) d\theta\right] \\ (\mathrm{Cov}(Z))_{11} &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \left[2\gamma_T \pi I_0(\kappa) + 2\gamma_T \pi I_2(\kappa)\right]\end{aligned} \tag{C.19}$$

$$= \frac{2\pi\gamma_T}{2m\lambda_0 T^2 I_0(\kappa)^2}\left[I_0(\kappa)+I_2(\kappa)\right] \tag{C.20}$$

$$= \frac{mT}{2m\lambda_0 T^2 I_0(\kappa)^2}\left[I_0(\kappa)+I_2(\kappa)\right] \tag{C.21}$$

$$= \frac{1}{2\lambda_0 T I_0(\kappa)^2}\left[I_0(\kappa)+I_2(\kappa)\right]. \tag{C.22}$$

We can have a similar calculation for the imaginary part, $(\mathrm{Cov}(Z))_{22}$, as well, but using the identity $\sin^2(\theta) = \frac{1}{2}(1-\cos(2\theta))$ instead of equation A.20. The off-diagonal elements of the covariance matrix vanish due to symmetry of integrand.

Therefore, we showed that for a given $\kappa \geq 0$, scaled residual

$$Z' = e^{-i\varphi_0}\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right)$$

converges to a zero mean complex gaussian with the following covariance:

$$\mathrm{Cov}\begin{bmatrix}\mathrm{Re}\{Z'\}\\ \mathrm{Im}\{Z'\}\end{bmatrix} = \begin{bmatrix}\mathrm{Re}\{Ze^{-i\varphi_0}\}\\ \mathrm{Im}\{Ze^{-i\varphi_0}\}\end{bmatrix} = \frac{1}{2\lambda_0 T I_0(\kappa)^2}\begin{bmatrix}I_0(\kappa)+I_2(\kappa) & 0\\ 0 & I_0(\kappa)-I_2(\kappa)\end{bmatrix}.$$

□

Corollary 5. *Assume $\phi(t) = 2\pi kt/T$, with $k > 0$ integer, and a sinusoidal modulation of the intensity at frequency m/T, with $m > 0$ integer possibly different from k, phase shift φ_0, and modulation amplitude $\varkappa$ such that*

$$\lambda(t) = \lambda_0\left(1+\varkappa\cos\left(2\pi mt/T-\varphi_0\right)\right),\ \lambda_0 > 0,\ 0 \leq \varkappa \leq 1, \tag{C.23}$$

and the point process is homogeneous Poisson with rate λ_0. Then the expectation of the PLV estimate converges (for $K \mapsto +\infty$) to

$$\mathrm{PLV}^* = \frac{1}{2}\varkappa e^{i\varphi_0}\delta_{km}, \tag{C.24}$$

where δ_{km} denotes the Kronecker symbol. Moreover, the asymptotic covariance of $Z = \sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right)$ is

$$\mathrm{Cov}\begin{bmatrix}Re\{Z\}\\ Im\{Z\}\end{bmatrix} = \frac{1}{2\lambda_0 T}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}. \tag{C.25}$$

Proof. Similar to corollary 2, we can derive the asymptotic PLV (see equation C.24) for this case, from the definition in equation 3.2. We use the

assumed phase $\phi(t) = 2\pi kt/T$ and apply the intensity function defined in equation C.23 in corollary 1:

$$\mathrm{PLV}^* = \frac{\int_0^T e^{i\phi(t)}\lambda(t)dt}{\int_0^T \lambda(t)dt} \tag{C.26}$$

$$= \frac{\int_0^T e^{i2\pi kt/T}\left(1 + \varkappa\cos\left(2\pi mt/T - \varphi_0\right)\right)dt}{\int_0^T \left(1 + \varkappa\cos\left(2\pi mt/T - \varphi_0\right)\right)dt}. \tag{C.27}$$

By using Euler's formula, we can write the second term in the numerator as weighted sum of exponentials ($\cos(x) = \frac{1}{2}(e^{ix} + e^{-ix})$),

$$\mathrm{PLV}^* = \frac{1}{2}\frac{\int_0^T e^{i2\pi kt/T} + \varkappa\int_0^T e^{i2\pi kt/T}\left(e^{i(2\pi mt/T - \varphi_0)} + e^{-i(2\pi mt/T - \varphi_0)}\right)dt}{\int_0^T dt + \int_0^T \varkappa\cos\left(2\pi mt/T - \varphi_0\right)dt} \tag{C.28}$$

$$= \frac{1}{2}\frac{\int_0^T e^{i2\pi kt/T} + \varkappa\int_0^T e^{i2\pi(k+m)t/T}e^{i\varphi_0} + \varkappa\int_0^T e^{-i2\pi(k-m)t/T}e^{i\varphi_0}dt}{\int_0^T dt + \int_0^T \varkappa\cos\left(2\pi mt/T - \varphi_0\right)dt} \tag{C.29}$$

$$= \frac{1}{2}\frac{\int_0^T e^{i2\pi kt/T} + \varkappa e^{i\varphi_0}\int_0^T e^{i2\pi(k+m)t/T} + \varkappa e^{i\varphi_0}\int_0^T e^{-i2\pi(k-m)t/T}dt}{\int_0^T dt + \varkappa\int_0^T \cos\left(2\pi mt/T - \varphi_0\right)dt}. \tag{C.30}$$

Given that $k, m > 0$ and we are integrating over full periods, all terms vanish except the last term in the numerator (if and only if $k = m$) and the first term in the denominator. Therefore we have,

$$\mathrm{PLV}^* = \frac{1}{2}\frac{\varkappa e^{i\varphi_0}\int_0^T e^{-i2\pi(k-m)t/T}dt}{\int_0^T dt} \tag{C.31}$$

$$= \frac{1}{2}\varkappa e^{i\varphi_0}\delta_{km}. \tag{C.32}$$

We prove the first part of the corollary.

The covariance matrix of the asymptotic distribution can be derived by the procedure we used for the proof of corollary 2. We plug the rate $\lambda(t)$ assumed in the corollary (see equation C.23) and integrate on $[0,\ T]$,

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2}\int_0^T \cos^2(2\pi kt/T)\left(1 + \varkappa\cos\left(2\pi mt/T - \varphi_0\right)\right)dt, \tag{C.33}$$

and use the trigonometric identity, equation A.20, to get

$$(\mathrm{Cov}(Z))_{11} = \frac{\lambda_0}{2\Lambda(T)^2}\int_0^T (1+\cos(4\pi kt/T))\,(1+\varkappa\cos(2\pi mt/T-\varphi_0))\,dt. \tag{C.34}$$

In the resulting equation,

$$\begin{aligned}&(\mathrm{Cov}(Z))_{11}\\ &= \frac{\lambda_0}{2\Lambda(T)^2}\left[\int_0^T dt + \varkappa\int_0^T \cos(2\pi mt/T-\varphi_0)\,dt + \int_0^T \cos(4\pi kt/T)dt\right.\\ &\quad\left.+\varkappa\int_0^T \cos(4\pi kt/T)\cos(2\pi mt/T-\varphi_0)\,dt\right],\end{aligned} \tag{C.35}$$

all terms except the first one vanish. The second and third vanish as we integrate in the full period, and the last term vanishes given that

$$\begin{aligned}&\int_0^T \cos(4\pi kt/T)\cos(2\pi mt/T-\varphi_0)\,dt\\ &= \cos(\varphi_0)\int_0^T \cos(4\pi kt/T)\cos(2\pi mt/T)\,dt\\ &\quad+\sin(\varphi_0)\int_0^T \cos(4\pi kt/T)\sin(2\pi mt/T)\,dt,\end{aligned} \tag{C.36}$$

and k and m are integers.

Finally, given that $\Lambda(T) = \int_0^T \lambda(t)dt = \lambda_0 T$, we have

$$(\mathrm{Cov}(Z))_{11} = \frac{1}{2\lambda_0 T}. \tag{C.37}$$

We have a similar calculation for the imaginary part, $(\mathrm{Cov}(Z))_{22}$. The off-diagonal elements of the covariance matrix vanish due to the symmetry of he integrand.

Therefore, we showed that for the scaled residual,

$$Z = \sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right)$$

converges to a zero mean isotropic complex gaussian:

$$\sqrt{K}\left(\widehat{\mathrm{PLV}}_K - \mathrm{PLV}^*\right) \underset{K\to+\infty}{\longrightarrow} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \frac{1}{2\lambda_0 T}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right).$$

□

Appendix D: Circular Noise

We use random numbers drawn from the von Mises distribution to generate noise for the phase of an oscillation. Consider the oscillation $O^{orig}[t] = e^{2\pi \boldsymbol{i} f t}$, where the bracket indicates the oscillation is sampled at equispaced discrete times $t = \{k\Delta\}_{k=1,\dots,q}$. Then $O[t]$ is a noisy version of this oscillation, which is perturbed in the phase

$$O[t] = e^{2\pi \boldsymbol{i} f t} \exp\left(\boldsymbol{i}\eta[t]\right), \tag{D.1}$$

where $\eta[t]$ is sampled i.i.d. from the zero-mean von Mises distribution $\mathcal{M}(0, \kappa)$ at each time t. Notably, κ is the dispersion parameter; therefore, larger κ correspond to smaller variance of the noise. In the simulation used in section 4.3, we use $\kappa = 10$.

In the simulation for the multivariate case, we use N_{osc}-dimensional vector of oscillations, $\boldsymbol{O}^{orig}[t] = \{O_j^{orig}[t]\}_{j=1,\dots,N_{osc}}$, and sample i.i.d. the noise for each oscillation, leading to the vector time series $\boldsymbol{\eta}[t]$. In this case, the noisy oscillations can be written as

$$\boldsymbol{O}[t] = \boldsymbol{O}^{orig}[t] \odot \exp\left(\boldsymbol{i}\boldsymbol{\eta}[t]\right), \tag{D.2}$$

where $\odot$ is (entrywise) Hadamard product.

The advantage of such phase noise is to preserve the spectral content of the original oscillation better than conventional normal noise. Nevertheless, using conventional normal (white) noise (on both the real and imaginary parts of the oscillation) did not change the results significantly.

Appendix E: Tables of Parameters

The choice of parameters used in the figures in the main text. In all simulations, $\phi_0 = 0$.

Table 1: Parameters Used for Simulations in Figure 2.

Parameter	Description	A	B
f	Frequency	1 Hz	
K	Number of trials	5000	
T	Simulation length	5 s	
λ_0	Average firing rate	20 Hz	
N_S	Number of simulations	5000	
κ	Modulation strength	0	0.5

Table 2: Parameters Used for Simulations in Figure 3.

Parameter	Description	A	B	C	D
f	Frequency	1 Hz			
K	Number of trials	10			
T	Simulation length	0.75 s	0.5 s	1 s	x-axis
λ_0	Average firing rate	30 Hz			
N_S	Number of simulations	500			
κ	Modulation strength	0			

Table 3: Parameters Used for Simulations in Figure 5.

Parameter	Description	A1	A2	A3	B1	B2	B3
f	Frequency	5 oscillatory components, 11–15 Hz					
K	Number of trials	10					
T	Simulation length	11 s					
λ_0	Average firing rate	20 Hz					
N_S	Number of simulations	100					
κ	Modulation strength	0			0.15		
n_c	Number of LFP channels	100					
n_s	Number of spiking units	10	50	90	10	50	90
κ_{noise}	Dispersion parameter of phase noise	10					

Code Availability

The code to reproduce our simulation results is at https://github.com/shervinsafavi/safavi_neuralComp2021.

Acknowledgments

We are very grateful to Afonso Bandeira and Asad Lodhia for fruitful discussions at the beginning of the project. We thank Edgar Dobriban for pointing us to Bai and Yao (2008) and Joachim Werner and Michael Schnabel for their excellent IT support. This work was supported by the Max Planck Society.

References

Aalen, O. O., Borgan, Ø., & Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view*. New York: Springer.

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover.

Almog, A., Buijink, M. R., Roethler, O., Michel, S., Meijer, J. H., Rohling, J. H. T., & Garlaschelli, D. (2019). Uncovering functional signature in neural systems via random matrix theory. *PLOS Computational Biology*, *15*(5), e1006934.

Anderson, G. W., Guionnet, A., & Zeitouni, O. (2010). *An introduction to random matrices*. Cambridge: Cambridge University Press.

Ashida, G., Wagner, H., & Carr, C. E. (2010). Processing of phase-locked spikes and periodic signals. In S. Rotter& S. Grn (Eds.), *Analysis of parallel spike trains* (pp. 59–74). New York: Springer.

Aydore, S., Pantazis, D., & Leahy, R. M. (2013). A note on the phase locking value and its properties. *NeuroImage*, *74*, 231–244.

Bai, Z., & Yao, J.-f. (2008). Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, *44* (3), 447–474.

Bai, Z., & Zhou, W. (2008). Large sample covariance matrices without independence structures in columns. *Statistica Sinica*, *18*, 425–442.

Banna, M., Merlevède, F., & Peligrad, M. (2015). On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries. *Stochastic Processes and Their Applications*, *125*(7), 2700–2726.

Benaych-Georges, F., & Nadakuditi, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, *111*, 120–135.

Bhattacharjee, M., & Bose, A. (2016). Large sample behaviour of high dimensional autocovariance matrices. *Annals of Statistics*, *44*(2), 598–628.

Billingsley, P. (1995). *Probability and measure*. New York: Wiley.

Brillinger, D. R. (1981). *Time series: Data analysis and theory*. Philadelphia: SIAM.

Bühlmann, P., Kalisch, M., & Meier, L. (2014). High-dimensional statistics with a view toward applications in biology. *Annual Review of Statistics and Its Application*, *1*, 255–278.

Bun, J., Bouchaud, J.-P., & Potters, M. (2017). Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, *666*, 1–109.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, *7* (5), 446–451.

Buzsaki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.

Buzsaki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents–EEG, ECOG, LFP and spikes. *Nat. Rev. Neurosci.*, *13* (6), 407–20.

Buzsaki, G., Logothetis, N., & Singer, W. (2013). Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms. *Neuron*, *80*(3), 751–764.

Buzsaki, G., & Schomburg, E. W. (2015). What does gamma coherence tell us about inter-regional neural communication? *Nat. Neurosci.*, *18*, 484–489.

Capitaine, M., & Donati-Martin, C. (2016). *Spectrum of deformed random matrices and free probability.* arXiv:1607.05560.

Chafaï, D., & Tikhomirov, K. (2018). On the convergence of the extremal eigenvalues of empirical covariance matrices with dependence. *Probability Theory and Related Fields*, *170*(34), 847–889.

Chavez, M., Besserve, M., Adam, C., & Martinerie, J. (2006). Towards a proper estimation of phase synchronization from time series. *J. Neurosci. Methods*, *154*(1–2), 149–160.

Cole, S., & Voytek, B. (2019). Cycle-by-cycle analysis of neural oscillations. *Journal of Neurophysiology*, *122*(2), 849–861.

Cueva, C. J., Saez, A., Marcos, E., Genovesio, A., Jazayeri, M., Romo, R., . . . Fusi, S. (2020). Low-dimensional dynamics for working memory and time encoding. *PNAS, 117*, 23021–23032.

Dai, H., Wang, Y., Trivedi, R., & Song, L. (2016). Recurrent coevolutionary latent feature processes for continuous-time recommendation. In *Proceedings of the First Workshop on Deep Learning for Recommender Systems* (pp. 29–34). New York: ACM.

De, A., Valera, I., Ganguly, N., Bhattacharya, S., & Rodriguez, M. G. (2016). Learning and forecasting opinion dynamics in social networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems*, *29* (pp. 397–405). Red Hook, NY: Curran.

Deger, M., Helias, M., Boucsein, C., & Rotter, S. (2012). Statistical properties of superimposed stationary spike trains. *Journal of Computational Neuroscience*, *32*(3), 443–463.

Dickey, A. S., Suminski, A., Amit, Y., & Hatsopoulos, N. G. (2009). Single-unit stability using chronically implanted multielectrode arrays. *J. Neurophysiol.*, *102*(2), 1331–1339.

Doussal, P. L., Majumdar, S. N., & Schehr, G. (2016). Large deviations for the height in 1D Kardar-Parisi-Zhang growth at late times. *Europhysics Letters*, *113*(6), 60004.

Einevoll, G. T., Kayser, C., Logothetis, N. K., & Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat. Rev. Neurosci.*, *14*(11), 770–785.

El Karoui, N. (2003). *On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity*. arXiv:0309355(math).

El Karoui, N. (2005). Recent results about the largest eigenvalue of random covariance matrices and statistical application. *Acta Physica Polonica. Series B*, *B35*(9), 2681–2697.

El Karoui, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Annals of Probability*, *35*(2), 663–714.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Annals of Statistics*, *36*(6), 2757–2790.

Elsayed, G. F., & Cunningham, J. P. (2017). Structure in neural population recordings: An expected byproduct of simpler phenomena? *Nature Neuroscience*, *20*(9), 1310.

Embrechts, P., Liniger, T., & Lin, L. (2011). Multivariate Hawkes processes: An application to financial data. *Journal of Applied Probability*, *48*(A), 367–378.

Ermentrout, B., & Pinto, D. (2007). Neurophysiology and waves. *SIAM News*, *40*(2).

Ermentrout, G. B., & Kleinfeld, D. (2001). Traveling electrical waves in cortex: Insights from phase dynamics and speculation on a computational role. *Neuron*, *29*(1), 33–44.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn. Sci.*, *9*(10), 474–80.

Fries, P. (2015). Rhythms for cognition: Communication through coherence. *Neuron*, *88*, 220–35.

Fukushima, M., Chao, Z. C., & Fujii, N. (2015). Studying brain functions with mesoscopic measurements: Advances in electrocorticography for non-human primates. *Current Opinion in Neurobiology*, *32*, 124–131.

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural manifolds for the control of movement. *Neuron*, *94*, 978–984.

Gao, P., & Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, *32*, 148–155.

Grosmark, A. D., & Buzsáki, G. (2016). Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science*, *351*(6280), 1440–1443.

Grosmark, A. D., Mizuseki, K., Pastalkova, E., Diba, K., & Buzsáki, G. (2012). REM sleep reorganizes hippocampal excitability. *Neuron*, *75*(6), 1001–1007.

Grün, S. (2009). Data-driven significance estimation for precise spike correlation. *Journal of Neurophysiology*, *101*(3), 1126–1140.

Guhr, T., Müller-Groeling, A., & Weidenmüller, H. A. (1998). Random-matrix theories in quantum physics: Common concepts. *Physics Reports*, *299*(4–6), 189–425.

Hanson, F. B. (2007). *Applied stochastic processes and control for jump-diffusions: Modeling, analysis and computation*. Philadelphia: SIAM.

Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, *33*(3), 438–443.

Herreras, O. (2016). Local field potentials: Myths and misunderstandings. *Front. Neural Circuits*, *10*, 101.

Hurtado, J. M., Rubchinsky, L. L., & Sigvardt, K. A. (2004). Statistical method for detection of phase-locking episodes in neural oscillations. *Journal of Neurophysiology*, *91*(4), 1883–1898.

Jiang, H., Bahramisharif, A., van Gerven, M. A. J., & Jensen, O. (2015). Measuring directionality between neuronal oscillations of different frequencies. *NeuroImage*, *118*, 359–367.

Johnson, D. H. (1996). Point process models of single-neuron discharges. *Journal of Computational Neuroscience*, *3*(4), 275–299.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, *29*, 295–327.

Johnstone, I. M., & Onatski, A. (2020). Testing in high-dimensional spiked models. *Annals of Statistics*, *48*(3), 1231–1254.

Juavinett, A. L., Bekheet, G., & Churchland, A. K. (2019). Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *eLife*, *8*, e47188.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., . . . Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), 232–236.

Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 324–332). New York: ACM.

Kovach, C. K. (2017). A biased look at phase locking: Brief critical review and proposed remedy. *IEEE Transactions on Signal Processing*, *65*(17), 4468–4480.

Kritchman, S., & Nadler, B. (2009). Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, *57*, 3930–3941.

Krumin, M., Reutsky, I., & Shoham, S. (2010). Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Front. Comput. Neurosci.*, *4*, 147.

Lepage, K. Q., Kramer, M. A., & Eden, U. T. (2011). The dependence of spike field coherence on expected intensity. *Neural Computation*, *23*(9), 2209–2241.

Li, Z., Cui, D., & Li, X. (2016). Unbiased and robust quantification of synchronization between spikes and local field potential. *J. Neurosci. Methods*, *269*, 33–8.

Liljenstroem, H. (2012). Mesoscopic brain dynamics. *Scholarpedia*, *7*(9), 4601.

Liptser, R. S., & Shiryaev, A. N. (2013a). *Statistics of random processes: I. General theory*. New York: Springer Science & Business Media.

Liptser, R. S., & Shiryaev, A. N. (2013b). *Statistics of random processes II: Applications*. New York: Springer Science & Business Media.

Liu, H., Aue, A., & Paul, D. (2015). On the Marčenko–Pastur law for linear time series. *Annals of Statistics*, *43*(2), 675–712.

Louart, C., Liao, Z., & Couillet, R. (2018). A random matrix approach to neural networks. *Ann. Appl. Probab.*, *28*(2), 1190–1248.

Loubaton, P., & Vallet, P. (2011). Almost sure localization of the eigenvalues in a gaussian information plus noise model: Application to the spiked models. *Electronic Journal of Probability*, *16*, 1934–1959.

Maimon, G., & Assad, J. A. (2009). Beyond Poisson: Increased spike-time regularity across primate parietal cortex. *Neuron*, *62*(3), 426–440.

Marchenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, *114*(4), 507–536.

Mastrogiuseppe, F., & Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, *99*(3), 609–623.

Mingo, J. A., & Speicher, R. (2017). *Free probability and random matrices*. New York: Springer.

Mitra, P. (2007). *Observed brain dynamics*. New York: Oxford University Press.

Mittelhammer, R. C. (1996). *Mathematical statistics for economics and business*. New York: Springer.

Namaki, A., Ardalankia, J., Raei, R., Hedayatifar, L., Hosseiny, A., Haven, E., & Jafari, G. R. (2020). *Analysis of the global banking network by random matrix theory*. arXiv:200714447.

Nawrot, M. P., Boucsein, C., Molina, V. R., Riehle, A., Aertsen, A., & Rotter, S. (2008). Measurement of variability dynamics in cortical spike trains. *Journal of Neuroscience Methods*, *169*(2), 374–390.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27.

O'Leary, T., Sutton, A. C., & Marder, E. (2015). Computational models in the age of large datasets. *Current Opinion in Neurobiology*, *32*, 87–94.

Pennington, J., & Bahri, Y. (2017). Geometry of neural network loss surfaces via random matrix theory. In *Proceedings of the International Conference on Machine Learning* (pp. 2798–2806).

Pennington, J., & Worah, P. (2017). Nonlinear random matrix theory for deep learning. In I. Guyon, Y. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 2637–2646). Red Hook, NY: Curran.

Pennington, J., & Worah, P. (2019). Nonlinear random matrix theory for deep learning. *J. Stat. Mech., 2019*(12), 124005.

Pereda, E., Quiroga, R. Q., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, *77*(1–2), 1–37.

Pesaran, B., Vinck, M., Einevoll, G. T., Sirota, A., Fries, P., Siegel, M., . . . Srinivasan, R. (2018). Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation. *Nature Neuroscience*, *21*, 903–919.

Peterson, E. J., & Voytek, B. (2018). *Healthy oscillatory coordination is bounded by single-unit computation*. bioRxiv:309427.

Protter, P. E. (2005). *Stochastic integration and differential equations*. New York: Springer.

Reimer, I. C., Staude, B., Ehm, W., & Rotter, S. (2012). Modeling and analyzing higher-order correlations in non-Poissonian spike trains. *Journal of Neuroscience Methods*, *208*(1), 18–33.

Rizoiu, M.-A., Lee, Y., Mishra, S., & Xie, L. (2017). *A tutorial on Hawkes processes for events in social media.* arXiv:1708.06401.

Rosenblum, M., Pikovsky, A., Kurths, J., Schäfer, C., & Tass, P. A. (2001). Phase synchronization: From theory to data analysis. In F. Moss & S. Gielen (Eds.), *Handbook of biological physics* (vol. 4, pp. 279–321). Amsterdam: Elsevier.

Safavi, S., Panagiotaropoulos, T., Kapoor, V., Ramirez-Villegas, J. F., Logothetis, N. K., & Besserve, M. (2020). *Uncovering the organization of neural circuits with generalized phase locking analysis*. bioRxiv. https://doi.org/10.1101/2020.12.09.413401

Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., . . . Toyama, K. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLOS Computational Biology*, *5*(7).

Shinomoto, S., Shima, K., & Tanji, J. (2003). Differences in spiking patterns among cortical neurons. *Neural Computation*, *15*(12), 2823–2842.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, *13*(1), 334–350.

Sohn, H., Narain, D., Meirhaeghe, N., & Jazayeri, M. (2019). Bayesian computation through cortical latent dynamics. *Neuron*, *103*, 934–907.

Stevenson, I. H., & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, *14*(2), 139–142.

Tabibian, B., Valera, I., Farajtabar, M., Song, L., Schölkopf, B., & Gomez-Rodriguez, M. (2017). Distilling information reliability and source trustworthiness from digital traces. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 847–855). New York: ACM.

Timme, M., Geisel, T., & Wolf, F. (2006). Speed of synchronization in complex networks of neural oscillators: Analytic results based on random matrix theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *16*(1), 015108.

Tracy, C. A., & Widom, H. (2002). *Distribution functions for largest eigenvalues and their applications*. arXiv:0210034(math-ph).

Truccolo, W. (2016). From point process observations to collective neural dynamics: Nonlinear Hawkes process GLMs, low-dimensional dynamics and coarse graining. *Journal of Physiology–Paris*, *110*(4, Part A), 336–347.

Truccolo, W., Hochberg, L. R., & Donoghue, J. P. (2010). Collective dynamics in human and monkey sensorimotor cortex: Predicting single neuron spikes. *Nat. Neurosci.*, *13*, 105–111.

Veraart, J., Novikov, D. S., Christiaens, D., Ades-Aron, B., Sijbers, J., & Fieremans, E. (2016). Denoising of diffusion MRI using random matrix theory. *NeuroImage*, *142*, 394–406.

Vinck, M., Battaglia, F. P., Womelsdorf, T., & Pennartz, C. (2012). Improved measures of phase-coupling between spikes and the local field potential. *J. Comput. Neurosci.*, *33*(1), 53–75.

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., & Pennartz, C. M. (2010). The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. *NeuroImage*, *51*(1), 112–22.

Watson, G. N. (1995). *A treatise on the theory of Bessel functions*. Cambridge: Cambridge University Press.

Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math*, *62*, 548.

Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, *67*, 325–327.

Williamson, R. C., Doiron, B., Smith, M. A., & Yu, B. M. (2019). Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Current Opinion in Neurobiology*, *55*, 40–47.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, *316*, 1609–1612.

Zarei, M., Jahed, M., & Daliri, M. R. (2018). Introducing a comprehensive framework to measure spike-LFP coupling. *Frontiers in Computational Neuroscience*, *12*.

Received July 18, 2020; accepted January 19, 2021.

Uncovering the organization of neural circuits with generalized phase locking analysis

Shervin Safavi[1,2], **Theofanis Panagiotaropoulos**[1,3], **Vishal Kapoor**[1], **Juan F. Ramirez-Villegas**[1,4], **Nikos K. Logothetis**[1,6,7], **Michel Besserve**[1,5*]

***For correspondence:**
michel.besserve@tuebingen.mpg.de

[1]Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen, Germany; [2]IMPRS for Cognitive and Systems Neurosurgeon, University of Tübingen, Tübingen, Germany; [3]Cognitive Neuroimaging Unit, INSERM, CEA, Université Paris-Saclay, NeuroSpin center, Gif/Yvette, France; [4]Institute of Science and Technology Austria (IST Austria), Klosterneuburg, Austria; [5]Department of Empirical Inference, Max Planck Institute for Intelligent Systems and MPI-ETH center for Learning Systems, Tübingen, Germany; [6]International Center for Primate Brain Research, Songjiang, Shanghai, China; [7]Centre for Imaging Sciences, Biomedical Imaging Institute, The University of Manchester, Manchester, UK

Abstract Spike-field coupling characterizes the relationship between neurophysiological activities observed at two different scales: on the one hand, the action potential produced by a neuron, on the other hand a mesoscopic "field" signal, reflecting subthreshold activities. This provides insights about the role of a specific unit in network dynamics. However, assessing the overall organization of neural circuits based on multivariate data requires going beyond pairwise approaches, and remains largely unaddressed. We develop *Generalized Phase Locking Analysis* (GPLA) as an multichannel extension of univariate spike-field coupling. GPLA estimates the dominant spatio-temporal distributions of field activity and neural ensembles, and the strength of the coupling between them. We demonstrate the statistical benefits and interpretability of this approach in various biophysical neuronal network models and Utah array recordings. In particular, we show that GPLA, combined with neural field modeling, help untangle the contribution of recurrent interactions to the spatio-temporal dynamics observed in multi-channel recordings.

Keywords
Spike-field coupling, microcircuits, prefrontal cortex, neural field models, neural mass models, hippocampus, phase-locking.

Introduction

Understanding brain function requires uncovering the relationships between neural mechanisms at different scales (Einevoll et al., 2019): from single neurons to microcircuits (Rasch et al., 2008, 2009), from microcircuits to a single brain area (Li et al., 2009) and from a single area to the whole brain (Schwalm et al., 2017; Zerbi et al., 2019). Therefore, it is crucial to develop data analysis tools to investigate the cooperative mechanisms that connect one level of organization to the next.

A well-studied example of such a collective organization phenomenon is oscillatory neuronal activity. Neural oscillations are hypothesized to support computations in the brain (Peterson and

Voytek, 2018) and various cognitive functions, such as perceptual binding (Engel et al., 1999), visual awareness (Dwarakanath et al., 2020), attention (Niebur et al., 1993) and memory (Buzsaki, 2006). These oscillations manifest themselves in Local Field Potentials (LFP), a mesoscopic extracellular signal (Liljenstroem, 2012) resulting from ionic currents flowing across the cellular membranes surrounding the electrode. LFP oscillatory activity partly reflects a number of subthreshold processes shared by units belonging to underlying neuronal ensembles and responsible for the coordination of their activity (Buzsaki et al., 2012; Einevoll et al., 2013; Herreras, 2016). As a consequence, a broad range of empirical investigations support the importance of analyzing oscillatory dynamics observed in LFPs (for reviews see Buzsaki et al. (2012, 2013); Einevoll et al. (2013); Herreras (2016); Pesaran et al. (2018)).

In particular, the relationship between spiking activity and LFP has broad implications for mesoscale mechanisms of network coordination. For instance, spike-field coupling relates to synaptic plasticity, triggering changes in the spike sequences generated by neural ensembles (Grosmark et al., 2012; Grosmark and Buzsáki, 2016). Moreover, cognitive functions, such as attention, are hypothesized to rely on interactions between various neural populations coordinated by network oscillations, which modulate the excitability of target populations so that they spike during time windows facilitating their communication (Fries, 2005, 2015; Womelsdorf et al., 2007).

Common approaches for investigating the spike-LFP oscillatory coupling are typically restricted to pairwise spike-LFP interactions (Zeitler et al., 2006; Ashida et al., 2010; Vinck et al., 2010, 2012; Jiang et al., 2015; Li et al., 2016; Zarei et al., 2018) which are suboptimal for modern datasets. Indeed, state-of-the-art multichannel electrophysiology systems (Dickey et al., 2009; Jun et al., 2017; Juavinett et al., 2019) allow simultaneous recording of hundreds or even thousands of sites (Pesaran et al., 2018; Jun et al., 2017; Buzsáki, 2004; Fukushima et al., 2015). This growth in dimensionality and complexity requires the parallel development of appropriate methodologies and models (Stevenson and Kording, 2011; O'Leary et al., 2015; Gao and Ganguli, 2015; Williamson et al., 2019). In particular, recent technological advances offer an unprecedented opportunity to study the relationship of large scale collective organization binding the activity of individual units (e. g. as spiking activity) with spatio-temporal field potential dynamics (e. g. wave patterns (Ermentrout and Pinto, 2007; Ermentrout and Kleinfeld, 2001)). Alongside this experimental progress, there is a growing need for conceptual and methodological frameworks to investigate this relationship. In particular, computing an interpretable summary of the coupling between neurophysiological quantities is of paramount importance in this high dimensional setting.

We develop a "Generalized Phase Locking Analysis" (GPLA) to address this growing need. GPLA characterizes and assesses statistically at once the coupling between the spiking activity of large populations of units and large-scale spatio-temporal patterns of LFP. After demonstrating GPLA's interpretability on toy simulated datasets and statistical benefits with respect to pairwise coupling approaches, we use neural field models to illustrate how this method characterizes key aspects of the underlying neural circuits. In particular, we show, on simulated data with a larger degree of realism, how GPLA can untangle the contribution of recurrent interactions to the observed spatio-temporal dynamics. Finally, GPLA's interpretability is exploited in the analysis of Utah array recordings from the macaque prefrontal cortex.

Results

Generalizing spike-oscillation coupling analysis to the multivariate setting

Our analysis relies first on characterizing the coupling between signals originating from a pair of recording channels. On the one hand, we consider the time-varying spike rate $\lambda(t)$ of a given unit; and on the other hand oscillatory activity $L_f(t)$ is derived from the LFP by band-pass filtering in a band of center frequency f. We assume $L_f(t)$ is the complex analytic signal representation of this oscillation, computed using the Hilbert transform (Chavez et al., 2006), such that $L_f(t) = a_f(t)e^{i\phi_f(t)}$, where $a_f(t)$ and $\phi_f(t)$ are the instantaneous amplitude and phase of the oscillation, respectively.

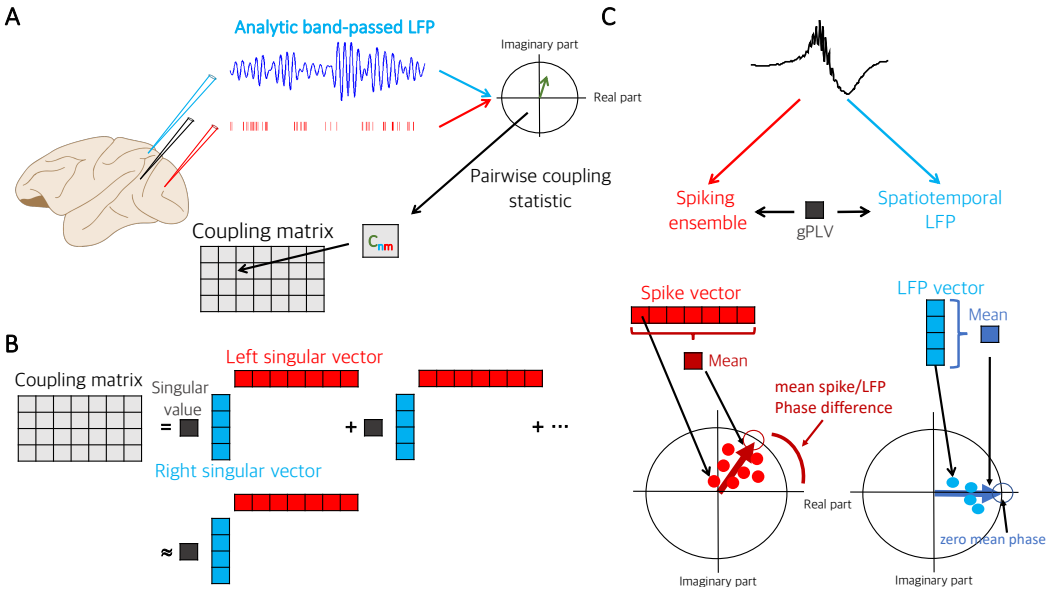


Figure 1. Schematic illustration of Generalized Phase Locking Analysis (GPLA).
(A) The coupling matrix is built from electrophysiology data by gathering pairwise complex phase locking estimates (based on Equation 2) of all spike-LFP pairs in a rectangular matrix. Coefficients (C_{nm}) contain information similar to complex-valued PLV up to a scaling factor: the magnitude indicates the strength of coupling and the angle reflects the average timing of the spike occurrence within the period of the corresponding LFP oscillation. **(B)** The coupling matrix can be approximated using its largest singular value and the corresponding singular vectors. **(C)** Singular vectors represent the dominant LFP (blue array) and spiking patterns (red array) and the singular value (d_1), called generalized Phase Locking Value (gPLV), characterizes the spike-field coupling strength for the phenomenon under study and the chosen frequency. The magnitude of each vector entry indicates a relative coupling of the corresponding unit/channel and the phase indicates the relative timing with respect to other units/channels. By convention, the phase of the LFP vector coefficients' average is set to zero, such that the phase of the spike vector average reflects the overall phase shift of the spike pattern with respect ot the LFP pattern.

The coupling between these signals can be characterized by the covariance

$$c(f) = \left\langle \lambda(t) L_f(t) \right\rangle = \left\langle \lambda(t) a_f(t) e^{i\phi_f(t)} \right\rangle = |c| e^{i\Phi_c} = |c| \left(\cos(\Phi_c) + i \sin(\Phi_c) \right) , \quad (1)$$

where the $\langle \cdot \rangle$ indicates averaging across time and experimental trials. The complex number c then reflects the strength of coupling through its modulus $|c|$, and the dominant LFP phase of spiking through its argument Φ_c (see Figure 1A). This coupling measure is a modification of the Phase-Locking Value (PLV) (Ashida et al., 2010), and differs from the latter mainly through the incorporation of the amplitude of the oscillation in the averaging, and the absence of normalization by the spike rate. Although $c(f)$ is straightforward to estimate based on observed spike times, leading to the quantity noted $\hat{c}(f)$ (see Safavi et al. (2020)), interpreting its value in terms of the underlying neural mechanisms is challenging, due to the largely unknown specific properties of the considered cell and recording site. Alternatively, synthesizing the information provided by a large number of couplings may provide a more reliable picture of the functioning of the underlying circuits.

Generalized Phase Locking Analysis (GPLA) is introduced as a multivariate statistical analysis technique to estimate the key properties of a matrix $\boldsymbol{C}(f)$ consisting of the pairwise coupling between a large number of units and LFP channels. Given the spike times $\{t_k^m\}$ for unit m and the analytic signal $L_f^n(t)$ that is filtered around frequency f for LFP channel n, the (n, m) coordinate of the coupling matrix $\boldsymbol{C}(f)$ is estimated by summing the values taken by the analytic signal at all spike

times (see Figure 1A),

$$\hat{\boldsymbol{C}}(f)_{n,m} = \sum_k L_f^n(t_k^m)\,. \tag{2}$$

As schematized in Figure 1B, to derive a compact and interpretable representation from this typically large matrix, we compute the Singular Value Decomposition (SVD) of the coupling matrix estimate and approximate it with the term of largest singular value d_1 leading to the approximation

$$\hat{\boldsymbol{C}} \quad = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^H \quad = \sum_{k=1}^{p} d_k \boldsymbol{u}_k \boldsymbol{v}_k^H \approx d_1 \boldsymbol{u}_1 \boldsymbol{v}_1^H\,, \tag{3}$$

where $\boldsymbol{v}^H$ indicate the transpose conjugate of vector $\boldsymbol{v}$. In this expression, the singular value d_1 is a positive scalar, that we will call generalized Phase Locking Value ($gPLV$), and quantifies the magnitude of the coupling. The associated complex valued singular vectors in this factorization will be respectively called the *LFP vector* $\boldsymbol{u} = \boldsymbol{u}_1$ and the *spike vector* $\boldsymbol{v} = \boldsymbol{v}_1$. As illustrated in Figure 1C, the spike vector indicates the pattern of coordinated spiking activity most coupled to LFP oscillations, while the LFP vector reflects the dominant spatio-temporal pattern of LFP involved in this coupling. Importantly, the relationship between $\boldsymbol{u}$ and $\boldsymbol{v}$ reflects the phase lag between spiking and LFP activities. Multiplication of both singular vectors by the same unit complex number leads to the exact same approximation as Equation 3, reflecting that GPLA only measures the relative phase between LFP and spikes. To resolve this ambiguity in our analyses, we adopt the convention of setting the phase of the average of LFP vector coefficients $\langle \boldsymbol{u} \rangle = \frac{1}{n_c} \sum_k u_k$ to zero, as illustrated in Figure 1C. As a consequence, the phase of the mean of the spike vector coefficients $\langle \boldsymbol{v} \rangle = \frac{1}{n_u} \sum_k v_k$ reflects the difference of mean phases between spiking and LFP activities. See section Detailed GPLA methodology for electrophysiology data in STAR Methods for more details.

Illustration of GPLA on toy examples

To illustrate how GPLA can provide an intuitive summary of the coupling between the population of spiking units and LFPs, we use toy simulations in which transient LFP oscillations (Figure 2A) modulate the firing probability in 18 spike trains (attributed to neuron-like units). The firing patterns of these neurons depend on the coupling strength and the delay with respect to the global oscillation. We introduce four simple models in Figure 2C-F, and demonstrate how the resulting spike vector summarizes the internal structure of the model. For each model, we schematically represent the coupling strength by the line thickness (absence of a line indicates no coupling) and the delay with respect to the global oscillation by the line color (Figure 2C-F, first column). In model 1-3 (Figure 2C-E) the spikes are coupled to the LFP (see STAR Methods, section Simulation of phase-locked spike trains for details) and in model 4 (Figure 2F) there is no coupling and the spikes were generated with a homogeneous Poisson process (presence or absence of coupling is also reflected in the coupling strength assessed by gPLV, as demonstrated in Figure 2B). Exemplary spike trains for each model are displayed in the second column of Figure 2C-F overlaid on the magnified version of the oscillation. Model 1 instantiates a global oscillation driving a synchronous population of neurons and this structure is also reflected in the resulting spike vector of the model as all the coefficients collapse to a single point (Figure 2C, third column). Model 2 corresponds to wave-like discharges of neurons (similar to the case of "delayed excitations from single oscillator" described by Ermentrout and Kleinfeld (2001)) which is also reflected in the evenly distributed phases of the spike vector coefficients over a 180 degrees interval (Figure 2D). Model 3 accounts for the activity of different cell populations that fire together predominately at three distinct phase value of the LFP and this synchrony is also reflected in the spike vector as the coefficients of the spike vector are clustered together in a complex plane (Figure 2E, third column) i. e. the group to which each unit belongs to is reflected in the phase of the associated spike vector coefficient. In model 4, there is no coupling between spikes and LFP, and the coefficients of the spike vector of this model are isotropically distributed in the complex plane (Figure 2F, third column). In all the above cases, the spike vector resulting from GPLA summarizes the coupling structure in an intuitive and compact way. Because

this setting has a single LFP channel, this analysis straightforwardly combines univariate coupling measures of each unit. However, statistical analysis of GPLA is different from the univariate case, as we show next.

Advantages of GPLA over univariate spike-field coupling

Beyond the above benefits for interpreting multivariate data, we investigated whether GPLA can be advantageous over its univariate counterpart from a statistical perspective by simulating transient oscillations (Figure 3A) and phase-locked spikes. The neural populations consisted of 3 groups, each of them locked to a different phase (0, 120, and 240 degrees) of the same oscillation, corrupted with additive noise. An illustrative simulation in the case of low noise and large amount of observed spike is shown in Figure 3B, together with the corresponding spike vector in Figure 3C, providing results similar to Fig. 2E.

For quantitative analysis, we first consider the setting of a single LFP channel and a handful of neurons are the focus of the analysis. Such recordings are still common and valuable in human electrophysiology experiments for understanding cognition (Mukamel and Fried, 2011; Fried et al., 2014). The main difference between this simulation and the one presented in Figure 2E is the weaker coupling of individual neurons leading to values at the edge of significance (assessed with the surrogate-based test, see section Significance assessment of gPLV in STAR Methods). While pooling the spikes from all units into a single spike train to get a *pooled Phase-Locking-Value* (pPLV) would result in a higher statistical power, it requires the distribution of the locking phase to be homogeneous across units (e.g. in the case of Figure 2C, but not for Figure 2D and E). In contrast, GPLA may be able to exploit the spike times from multiple neurons to achieve better statistical power in assessing the global coupling between spikes and LFPs. We ran 5000 simulations with only 3 units and compared the coupling assessment based on PLV, pPLV, and gPLV. Figure 3D represents the estimated PLVs, with averages matching the couplings obtain with a larger number of spikes Figure 3C. Performance of each measure is assessed based on its detection rate, which is defined as the percentage of simulations for which significant coupling is detected, as assessed using spike-jittered surrogate data (see STAR Methods section Significance assessment of gPLV) and with a significance threshold of 5%. As it is demonstrated in Figure 3E, gPLV detection outperforms the competing approaches (PLV and pPLV).

Beyond better detection of significant coupling, GPLA-based estimation of pairwise couplings based on the approximation of Equation 3 may be more accurate than individual estimates when the data is very noisy and multivariate, benefiting from the SVD procedure to disentangle noise from the ground truth coupling. To demonstrate this, we performed simulations similar to the above, but using 50 LFP channels containing oscillations driving spike-LFP coupling, contaminated by different levels of noise (i. e. adding Gaussian noise with different variances to the transient oscillation, see STAR Methods for details), and modulating the firing rates of the units, lower firing rates leading to a larger amount of martingale noise for the PLV estimates (Safavi et al., 2020). An example LFP trace with (black) and without (blue) noise is exemplified in Figure 3F and an example coupling matrix in the presence of noise is also illustrated in Figure 3G. In this case, the coupling matrix has rank one, as all the units are locked to a single frequency. We ran the simulations with different amounts of LFP noise (indicated on the x-axis of Figure 3H-I) and computed the coupling coefficients (similar to Figure 3G) and compare it to ground truth (based on Equation 32). Signal-to-Noise Ratio (SNR) was defined as the ratio of estimation error (the difference between estimated PLV and the ground truth) to coupling strength (PLV) and was used to compare the quality of GPLA-based and univariate estimation (indicated in the y-axis of Figure 3H-I). As this simulation demonstrates, the estimation error of the coupling coefficients is larger for the univariate estimation than for the GPLA-based approach for a broad range of noise levels (Figure 3H-I).

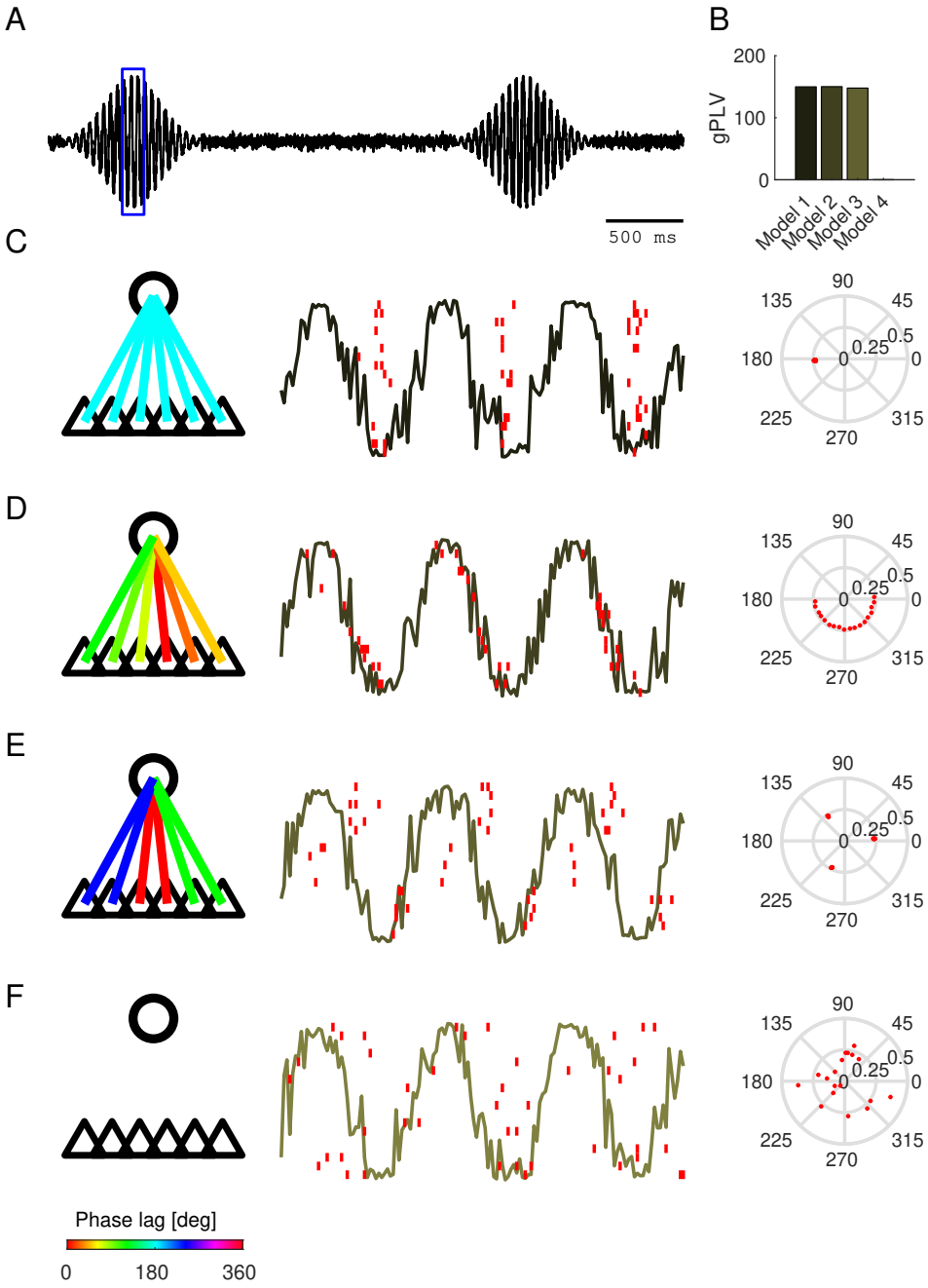


Figure 2. Illustration of GPLA on simple simulations.
(A) Normalized amplitude of LFP-like oscillatory signals. **(B)** gPLVs for different models demonstrated in C-F **(C-F)** Various scenarios of spike-LFP coupling. Left: schematic representation of the modulating LFP oscillation (circle), and 6 representative neuron-like-units (indicated by the triangles). The color of each connecting line indicates the locking phase (see bottom colorbar for color code). Center: LFP-like signals within the window specified by the blue box in A and spikes are represented by overlaid red vertical lines. Right: resulting spike vector is represented in the third column. **(C)** Spiking activity globally synchronized to the trough of the LFP oscillation. **(D)** Sequential discharge of spikes coupled to the LFP. **(E)** Three clusters of neurons discharge at different phases of the LFP oscillation (a similar model was also used in Figure 3). **(F)** Spiking activity uncoupled to LFP oscillation (independent homogeneous spike trains).

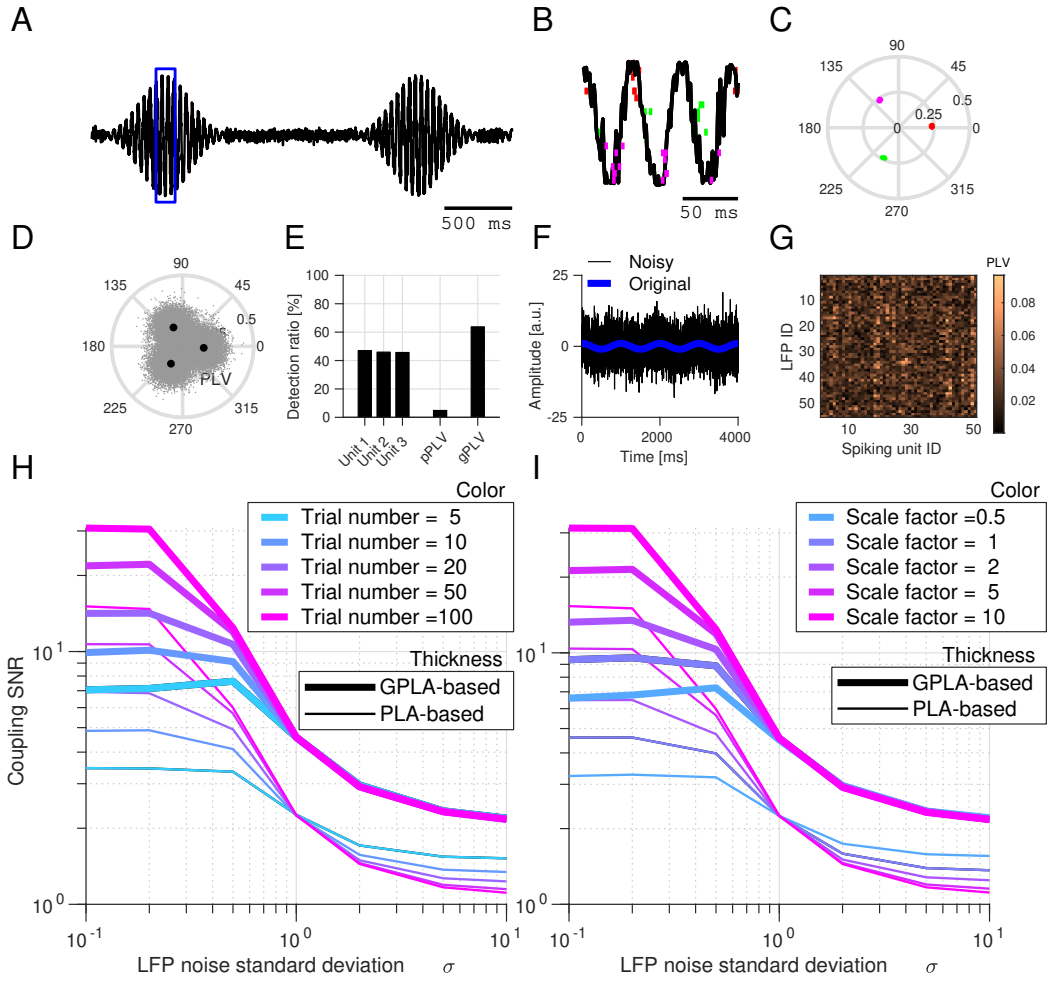


Figure 3. Comparison of GPLA and uni-variate spike-field coupling
(A) Normalized amplitude of LFP-like transient oscillatory signal with additive Gaussian white noise. **(B)** LFP-like signal and overlaid spike raster (colored vertical lines – colors indicate populations with the same locking phase) within the window specified by the blue box in (A). **(C)** Spike vector coefficients in the complex plane (colors correspond to B). Each dot represents one coefficient of the spike vector corresponding to a single neuron (note that within each cluster dots are overlapping as they are similarly coupled). **(D)** Complex PLVs represented in the complex plane. Angles indicate the locking phase and the radius the PLV. The gray point clouds indicate the PLV of multiple simulations and larger black dots indicate the average values. **(E)** Performance comparison (in percentage of simulations with significant coupling) of PLV, pooled PLV (pPLV) and gPLV, for three individual neurons. **(F)** Example oscillation, original (blue trace) and noisy (black trace). **(G)** Example coupling matrix related to simulation with a large amount of noise ($\sigma = 5$) **(H-I)** Comparison of GPLA-based and PLA-based estimation of PLVs for (H) different number of trials and (I) different levels of firing rate. Signal-to-Noise Ratio (SNR) is defined as the estimation error (the difference between estimated PLV and the ground truth) to coupling strength (PLV).

Statistical properties of GPLA

While in the previous section, GPLA's significance was assessed using surrogate data, this approach is computationally expensive and provides limited insights into the statistical properties of GPLA estimates. We now investigate this question using theory, and exploit it to assess more efficiently the significance of multivariate coupling. Singular values and vectors estimated by GPLA have an intrinsic variability due to the stochasticity of spiking activity, which can be investigated through stochastic integration and random matrix theory (Anderson et al., 2010; Capitaine and Donati-Martin, 2016). In the absence of coupling between spikes and LFP, appropriate preprocessing allows deriving analytically the asymptotic distribution of univariate and multivariate coupling measures (Safavi et al., 2020), including the convergence of the squared singular values to the classical Marchenko-Pastur (MP) law (Marchenko and Pastur, 1967). Based on the MP law, we can define an upper bound on the largest singular values of the coupling matrix, such that exceeding this bound indicates the significance of the coupling (for more details see STAR Methods section Analytical test and Safavi et al. (2020)), leading to a *analytical test*.

We assessed the performance of this test on simulated spikes and LFPs with or without coupling as follows. Briefly, we synthesized multivariate LFP activity by linearly superimposing several oscillations ($O_k(t)$ demonstrated in Figure 4A) with different multiplicative weights applied for each LFP channel and generated the spike trains of each unit with Poisson statistics. As for the coupling between spikes and LFPs, $2/5^{\text{th}}$ of the units were coupled to the LFP oscillations (exemplified in Figure 4B), while the remaining units had homogeneous Poisson spike trains (for details see the STAR Methods, section Simulation of phase-locked spike trains). The estimated coupling matrix computed based on Equation 2 for a simulation with 100 spike trains and 100 LFPs is exemplified in Figure 4C, where we have two coupled populations, one coupled to the lowest-frequency and one coupled to the highest-frequency oscillatory component of LFP (respectively top-right and bottom-left relatively bright blocks in Figure 4C and sample spike trains and LFP in Figure 4B).

Computing the SVD of the coupling matrix after application of the preprocessing explained in STAR Methods, section LFP pre-processing, results in a spectral distribution for the squared singular values, which matches the prediction of the theory (Figure 4D-E). In the absence of coupling between spikes and LFP signals (Figure 4D), the distribution of the eigenvalues closely follows the MP law and in the presence of coupling, the largest eigenvalue exceeds the significance bound predicted by Random Matrix Theory (RMT) (see STAR Methods section Analytical test for more details).

We further quantified the type I and II error of this analytical test. For type I error, we ran the simulations with non-zero coupling between spikes and LFP signals. GPLA was able to indicate a significant coupling between spike and LFP even when the coupling strength is as small as 0.05 (no coupling case corresponds to 0 strength and the perfect coupling corresponds to 1). One benefit of our analytical test is that the statistical power does not degrade with the increasing dimension of the data through the number of recording channels (Figure 4F). Notably, this is in contrast with univariate methods, assessing independently the significance of pairwise couplings resulting from each pair of spike train and LFP. In this case, it is necessary to correct for multiple comparisons before defining the significance threshold, Therefore, the statistical power will typically degrade with the increase in the number of units/LFPs. This is particularly relevant for weaker couplings as they may lose significance after correction for multiple comparisons . Lastly, in order to quantify the type II error of the significance test, we ran the simulation with no coupling between spikes and LFP and quantified the number of false positives. Our results show that our analytical test has a negligible ($< 5\%$) false positive error (Figure 4G).

We also quantified the performance of the method for estimating the number of populations coupled to different rhythms. Similar to the simulation explained earlier (Figure 4A-C), we simulate multiple (1-10) non-overlapping cell assemblies synchronized to different LFP rhythms (with different frequencies within a narrow range of 11-15.5 Hz). When the coupling is larger than a minimal

strength (coupling strength of 0.5), the method was able to capture the number of populations with very low error, MSE < 0.015 (Figure 4H).

A neural mass model of spike-LFP dynamics

While the above results have addressed the meaning of GPLA's outcomes in toy models, their neuroscientific interpretation requires modeling of the underlying neural network. We study this question in the context of a two-population *neural field model*: a spatially distributed rate model of the activity of two interacting homogeneous populations: excitatory pyramidal cells (E population) and inhibitory interneurons ("I" population) (Nunez and Srinivasan, 2006; Wilson and Cowan, 1973). The model is governed by three basic input-output relations (see STAR Methods, section Analytical neural field modeling of spike-field coupling) and depicted in Figure 5A: (1) the dynamics of the average somatic membrane potentials V_E and V_I of each population is governed by afferent post-synaptic currents η originating from other cortical or subcortical structures as well as recurrent excitatory and inhibitory post-synaptic currents (EPSC and IPSC) s_E and s_I; (2) the population spike rates λ_E and λ_I are a function of their respective membrane potentials; and (3) EPSC and IPSC are each controlled by the spike rate of their afferent population (E and I respectively). As a consequence, the dynamics of a two population neural mass model can be described based on six state variables: the excitatory variables (V_E, λ_E, s_E) and the inhibitory variables (V_I, λ_I, s_I), together with the additional PSC η reflecting exogenous input to the structure. In the context of large-scale recordings, the neural population can be distributed across one or several spatial directions, possibly following the spatial spread of neurons' soma, axons, and dendrites. Following a classical approximation depicted in Figure 5B, inhibitory connections are considered to be local (Sik et al., 1995; Schomburg et al., 2012; Taxidis et al., 2012), such that global coupling between cells surrounding distinct recording sites is assumed to happen exclusively through excitatory axons ($s_E(x)$ may depend on λ_E at other spatial locations than x) and the possibly space-dependent exogenous input current $\eta(x)$.

For the LFP $L(t)$, resulting from the conduction of trans-membrane currents in the extracellular space, we assume the contribution of currents flowing through the membrane of interneurons is negligible, based on the weakness of the anisotropy induced by their dendritic geometry across the population (Nó, 1947; Lindén et al., 2011; Mazzoni et al., 2015). The LFP thus results exclusively from pyramidal cell's membrane currents. Which currents (IPSC, EPSC, leak current, exogenous current) affect the most the recorded LFP at a given spatial location depends on multiple factors: the geometry of the cells, the distribution of synapses (inhibitory, excitatory, exogenous) onto them, and the geometry of the electrodes (Buzsaki et al., 2012, 2013; Einevoll et al., 2013). Figure 5C provides a schematic of how the differentiated location of synaptic boutons over the dendritic tree may result in variable algebraic contributions of each type of current to each recording channel.

Low rank linear response theory and frequency analysis

Neural mass models as described above are typically assumed non-linear, and use a static sigmoidal transformation to convert membrane potentials into spike rates (see section Analytical neural field modeling of spike-field coupling in STAR Methods and Jirsa and Haken (1996)). However, assuming small-amplitude perturbations in the neighborhood of an operating point, important aspects of their structure can be captured by linearization of such a model (see Methods and e.g. Moran et al. (2007); Ledoux and Brunel (2011); Pinotsis et al. (2012)) This leads to a linear time-invariant model, whose behavior is fully characterized by its amplitude and phase response to oscillatory inputs at each frequency. This allows studying how the underlying microcircuit parameters influence the GPLA properties. We provide qualitative illustrations of this idea in simplistic scenarios. First, as illustrated in Figure 5D, consider the microcircuit receives exogenous input exclusively onto the pyramidal cells' dendrites (no feedforward inhibition) and I cells receive local excitatory input, but do not synapse back onto E cells (no feedback inhibition). We assume additionally that subthreshold activity is dominated by the exogenous input currents and proportional to the measured LFP. The

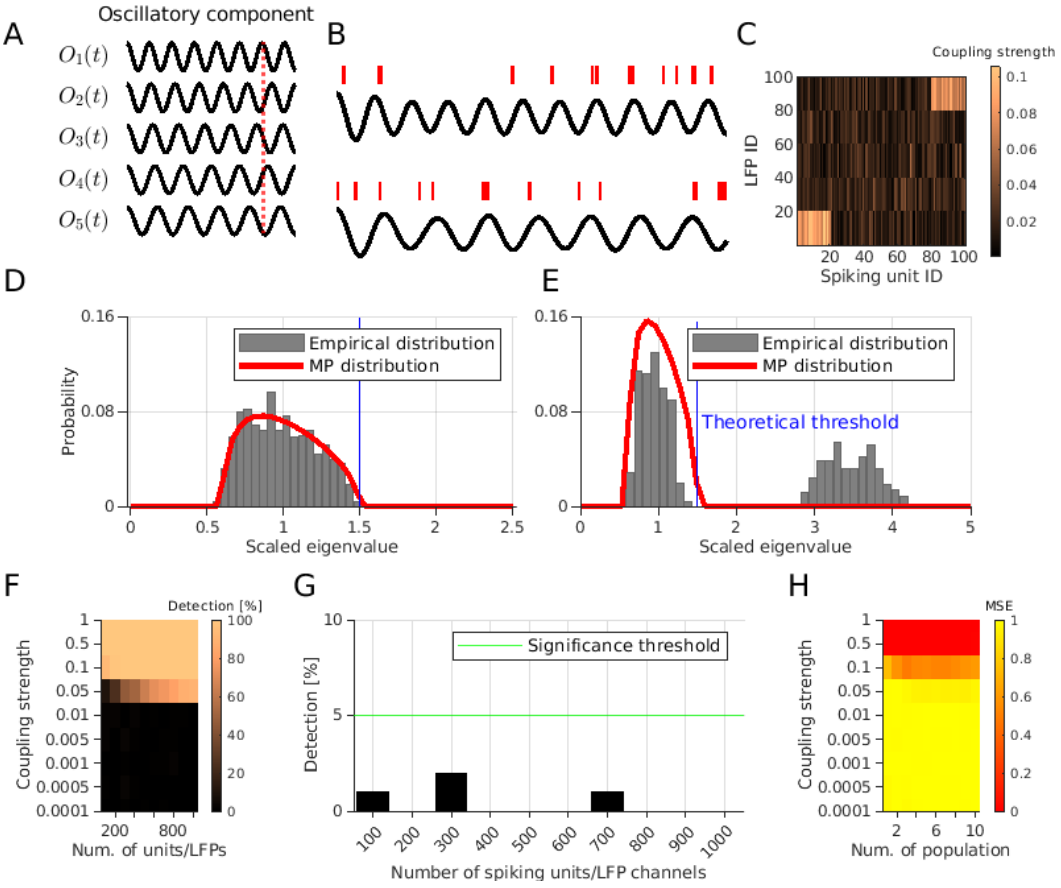


Figure 4. Quantifying statistical properties of GPLA with theoretical significance test.
(A) LFPs are synthesized by mixing several oscillatory components ($O_k(t)$). The vertical red line evidences the phase shift between them. **(B)** Two exemplary spike trains (each from one of the coupled populations) and the corresponding LFPs. In the LFP trace on the top, the oscillatory component with the highest frequency is dominated and for the bottom one, the lowest frequency component is the dominated oscillation. **(C)** An exemplary coupling matrix for a simulation with two coupled populations. **(D-E)** Theoretical Marchenko-Pastur distribution (red lines) and empirical distribution (gray bars) for (D) simulation without coupling and (E) with coupling between multivariate spikes and LFP **(F)** Performance of GPLA for the detection of coupling between spike trains and LFPs for different strength of coupling (y-axis) and different number of spiking units/LFP channels. **(G)** Type I error for different numbers of spiking units/LFP channels (x-axis), quantified as the percentage of simulations wherein a significant coupling between spike trains and LFPs is detected in absence of ground truth coupling. The horizontal green line indicates the %5 threshold. **(H)** Mean-squared-error of GPLA-based estimation of the number of populations coupled to LFP for varying coupling strengths (y-axis) and numbers of coupled populations (x-axis).

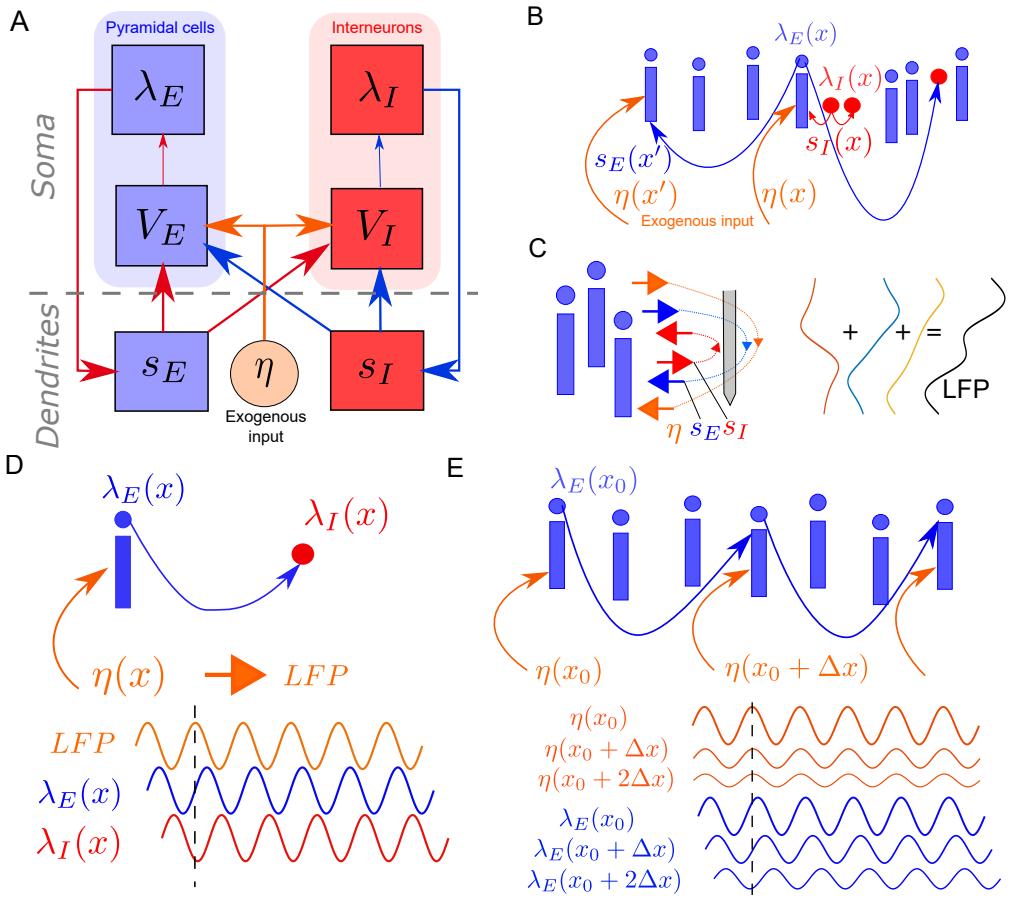


Figure 5. Generative model of spike-LFP coupling.
(A) A two-population neural field model of neural dynamics. V_k, λ_k and s_k indicate respectively somatic membrane potential, firing rate and post-synaptic current for Excitatory (k=E) and Inhibitory (k=I) populations. η indicates the exogenous input to the circuit. Arrows indicates the causal dependence between variables of the model. **(B)** Schematic representation of the model's connectivity: local inhibition and long range excitation, together with the driving by exogenous synaptic currents. **(C)** Schematic representation of the contribution of postsynaptic currents to the electric field, affected by the spatial distribution of synapses over the dendritic tree and the geometry of pyramidal cells. From left to right: Schematic representation of pyramidal neurons, electric field, electrode (gray bar), contribution of each current (EPSC, IPSC and exogenous current, leak current is also contributing to LFP but is not shown) to the LFP profile along the electrode's axis **(D)** Simple microcircuit structure leading to a temporal ordering of the local activities of different kinds $LFP \rightarrow excitation \rightarrow inhibition$ **(E)** Simple microcircuit structure leading to a temporal ordering of activities of the same kind across space: the location receiving stronger exogenous input leads other locations, such that amplitude gradient leads to phase gradients.

lag induced by the membrane potential dynamics results in a positive (frequency-dependent) delay of excitatory activity with respect to the LFP (reflecting the input), while inhibitory activity is itself delayed with respect to excitation. For an exogenous input oscillating at a given frequency, this implies a phase lag configuration between the (oscillatory) response of these variables, irrespective of the considered frequency.

Going beyond the relationship between different variables observed at the same location, we study how the activity of the same variable varies across space. For such a scenario, we extend spatially the previous model (no feedforward and feedback inhibition), by adding horizontal E-E connectivity with strength monotonously decreasing with distance, and study the activity resulting from a spatially inhomogeneous oscillatory input, with larger input amplitude at a given side (on the left in Figure 5E). Again due to the delay induced by membrane dynamics, this results in the propagation of the activity from one side to the opposite. This leads to an interesting relationship between the phase and amplitude of oscillatory activity: the location of largest amplitude is ahead of time with respect to the neighboring locations with smaller amplitudes. Note in particular that these propagation-like patterns are induced by the network's horizontal connectivity, while the input to the structure does not have phase lags at different locations (Ermentrout and Kleinfeld, 2001). These two examples indicate that phase and amplitude of oscillatory activities are informative about the underlying microcircuit structure and dynamics. More realistic scenarios should however take into account recurrent interactions between cell populations, and these above phases-amplitude relationships become frequency-dependent, as we will see in the next sections.

In order to analyze these more complex scenarios, a systematic way to evaluate the coupling between network variables at a given frequency is required. Consider the coupling between firing rate ($\lambda_E(x_1,t)$) and LFP ($L(x_2,t)$) at two different locations x_1 and x_2.

The linearity of the system implies the existence of transfer functions (denoted H_{λ_E} and H_L respectively), linking their activity to the time domain Fourier transform of the exogenous input $\hat{\eta}(x,f)$. Crucially, the first singular value approximation of GPLA can be obtained by assuming the space and time dependence of the input are separable ($\hat{\eta}(x',t) = n(x')\epsilon(t)$ implying $\hat{\eta}(x',f) = n(x')\hat{\epsilon}(f)$). This leads to both rate and LFP being proportional to the exogenous input at a given frequency, with the following expressions based on the transfer functions:

$$\lambda_E(x,f) = \int H_{\lambda_E}(x,x',f)n(x')\hat{\epsilon}(f)dx' = \psi_E(x,f)\hat{\epsilon}(f) \quad \text{and} \quad L(x,f) = \psi_L(x,f)\hat{\epsilon}(f)\,. \tag{4}$$

such that the coupling between spikes and LFP at locations x_1 and x_2 writes

$$C_{x_1,x_2}(f) = \psi_L(x_2,f)\psi_E(x_1,f)^* E(f)\,, \tag{5}$$

where $E(f)$ is the input power spectral density at frequency f and z^* denotes the complex conjugate of z. This shows that the coupling between L at x_2 and λ_E at x_1 is separable in the spatial variables (x_1, x_2), characterized by two functions of space: one for the field, ψ_L, and one for the excitatory spiking, ψ_E. Three quantities, ψ_L, ψ_E, and $E(f)$, are thus enough to reconstruct any pairwise coupling up to a scaling factor. In particular, as $E(f)$ is a positive number, the phase of C reflects a property of the circuit, independent of the exogenous input. Importantly, this analysis also applies to the coupling between the same variables at different locations, phase gradient thus informing us about the underlying spatial connectivity, and can account for excitatory and inhibitory activity at the same time.

In practice, we have access to a finite number of spatial measurement points of $C_{x_1,x_2}(f)$ corresponding to electrode channels where L and λ are recorded. This corresponds to the rectangular matrix $\boldsymbol{C}(f)$ estimated by GPLA, combining excitatory and inhibitory units. The above separability then implies that $\boldsymbol{C}(f)$ is a rank-one matrix, such that we can in principle apply Singular Value Decomposition (SVD) to estimate the coupling properties using the largest singular value and the associated left and right singular vectors, which are spatially discretized approximations of the above functions ψ_L and ψ_E (up to a multiplicative constant). Overall, the spatial distribution of the

phase and magnitude of each singular vector is influenced by the underlying recurrent dynamics and its propagation across horizontal connections (shaping the transfer functions such as H_{λ_E} and H_L), as well as the type of currents that dominate the LFP. As we will demonstrate in the next sections, the analysis of these features across frequencies is a rich source of information to validate assumptions about local network organization based on experimental multivariate data.

Application to spike-field dynamics during sharp wave-ripples

The phenomenon of hippocampal Sharp Wave-Ripples (SWR) is one of the most striking examples of neural activity entangling spike and LFP dynamics in multiple frequency bands, attributable to specific mechanisms of the underlying microcircuit (Buzsáki et al., 1992). Specifically, SWRs are brief episodes observed in hippocampal LFP traces combining a high-frequency oscillation (the ripple) to a low-frequency biphasic deflection (the sharp-wave). Moreover, these LFP activities are well known to be synchronized with spiking activity, with different cell-types firing at specific phases of the ripple oscillation (Buzsáki et al., 1992), although spike-field coupling at lower frequencies has also been demonstrated (Ramirez-Villegas et al., 2015).

We use simulations of in-vivo SWR described in Ramirez-Villegas et al. (2018) in order to demonstrate what insights GPLA can provide about the underlying hippocampal network mechanisms. The model generates realistic spiking and LFP activity in subfields CA1 and CA3, based on populations of two-compartment Hodgkin-Huxley neurons distributed along two one dimensional grids representing the strata of each subfield. In this model, the connectivity of CA3 is characterized by strong recurrent excitatory auto-associational $E-E$ connections, together with $E \rightarrow I$ connections and short-range $I \rightarrow E$ and $I \rightarrow E$ synapses (see Figure 6A for a schematic representation). In contrast, local $E-E$ connections are absent in CA1, but both E and I cells receive feedforward excitation from CA3. LFPs were generated from the total trans-membrane currents using line current density approximation, and measured by two laminar multi-shank electrodes (see STAR Methods, section Simulation of hippocampal sharp wave-ripples for more details).

We first assay the insights provided by GPLA on SWRs by analyzing a single hippocampal subfield, CA1, as various studies suggest SWRs emerge from it in response to afferent CA2- and CA3-ensemble synchronous discharges (Csicsvari et al., 2000; Oliva et al., 2016). In this simulation, LFP and unit recordings are distributed along two orthogonal spatial directions (laminar for LFPs and horizontal for units). We use a total of 157 peri-ripple traces of simulated LFPs and spikes of both populations (inhibitory and excitatory) of duration approximately 1 sec. Exemplary traces of simulated LFP and population firing rate of the CA1 population (pyramidal cells and inhibitory interneurons belonging to CA1) are shown in Figure 6B.

GPLA results for representative frequency bands are provided in Figure 6C-E and for all bands covering the 1-180Hz interval in supplemental Figure 10. The overall coupling magnitude (gPLV) was significant is all frequencies (Figure 6C), according to both surrogate (based on spike jittering, $p<0.05$) and analytical (based on random matrix theory) tests. In particular, the strongest coupling is detected in the ripple band (80-180 Hz), which is compatible with results based on classical univariate techniques (Buzsáki et al., 1992).

The LFP vectors exhibit a biphasic electric potential profile along the spatial axis along which the electrode contacts are distributed, typical of laminar recordings (Figure 6D), and corresponding to the field generated by the dipolar geometric arrangement of sources and sinks in the parallel two-compartment models of pyramidal neurons used for this simulation. Notably, the two-compartment approximation has been shown to lead to results comparable to simulations based on passive membrane currents of realistic morphologies (see Figure 6D (right) and Ramirez-Villegas et al. (2018) for more details). To check the quantitative agreement between the LFP vector and the original model of LFP generation in this simulation, we computed analytically the total LFP generated passively by all pyramidal cells using the original LFP simulation code of Ramirez-Villegas et al. (2018), and assuming all cells have identical trans-membrane currents flowing through their somatic and dendritic compartments (see STAR Methods section Computation of the laminar LFP profile).

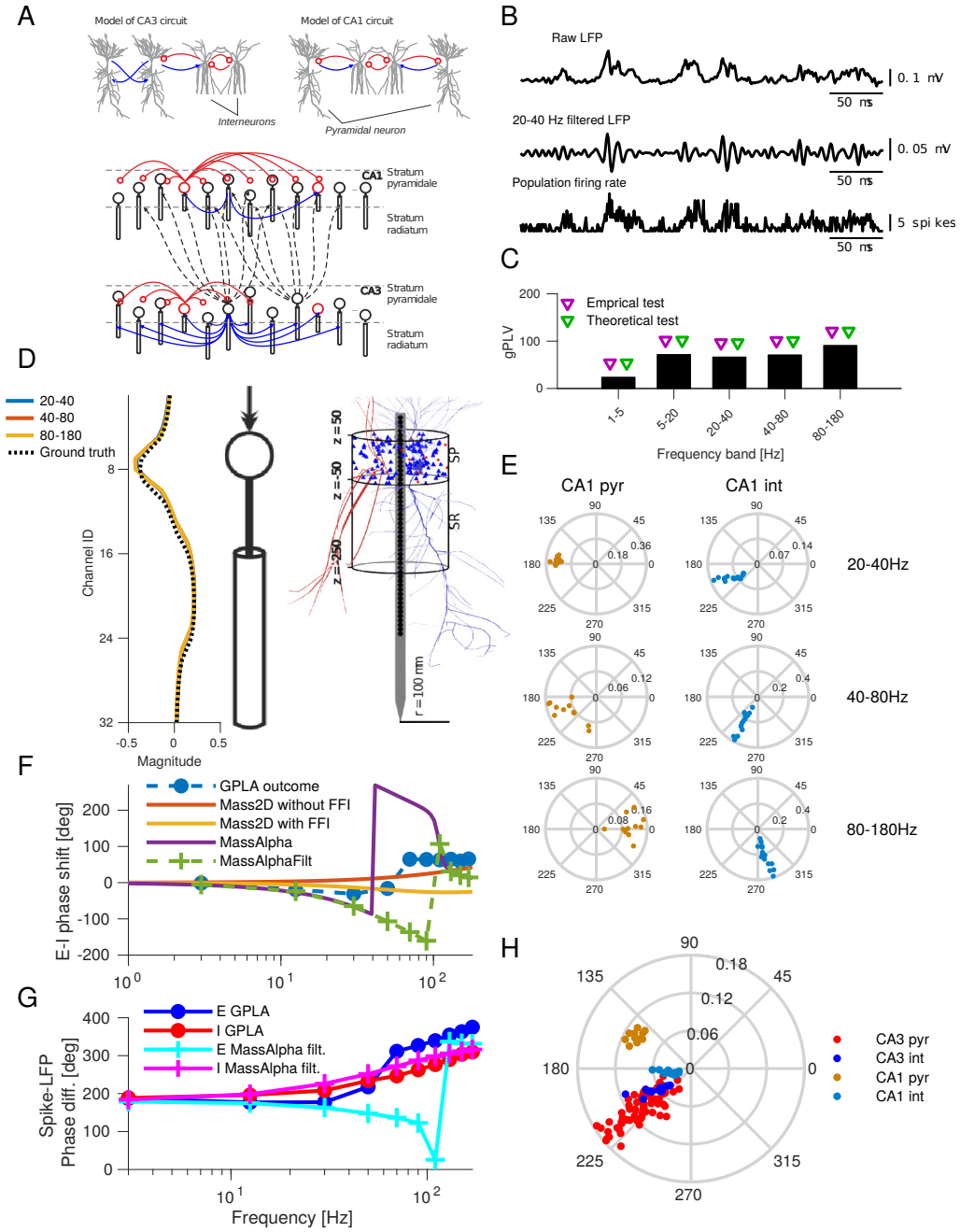


Figure 6. Analysis of hippocampal SWRs generated by a biophysical model
(A) Hippocampal multi-compartment model. Top: Canonical circuits of CA1 and CA3. Bottom: Schematic of the whole model (blue, excitatory connections; red, inhibitory, modified from Ramirez-Villegas et al. (2018). **(B)** From top to bottom: Example broad band CA1 LFP trace, band-pass filtered trace of the CA1 LFP in ripple band (80-180 Hz), and population firing rate of CA1 neurons. **(C)** CA1 gPLVs. Triangles indicate the significance assessed based on empirical (blue triangles, $p<0.05$) and theoretical (red triangles) tests. **(D)** Left: LFP vectors for GPLA of CA1, with a schematic of the pyramidal two-compartment model. Right: Location of the somata and expample morphologies of each cell type (blue, pyramidal cells; red, interneurons) in an equivalent model with many compartments (right hand-side schematic modified from Ramirez-Villegas et al. (2018)). **(E)** Spike vector coefficients for CA1 in several frequency bands (left: pyramidal cells, right: interneurons). **(F)** Average phase lag between LFP and spike vectors across frequencies for: outcome GPLA on hippocampal SWRs, theoretical analysis of *Mass2D* (without and with feedforward inhibition) and *MassAlpha* neural mass models. Dashed green line indicate *MassAlpha* filtered over the frequency bands used for GPLA. **(G)** Difference between phases of E and I populations based on GPLA the *MassAlpha* neural mass model filtered in the same bands (IPSP was used as LFP proxy). **(H)** Spike vector resulting GPLA jointly applied to CA1 and CA3 in the gamma band (20-40 Hz).

While the dendritic current reflects the post-synaptic input of the cell, somatic currents are taken opposite to preserve the charge neutrality of each cell. The resulting theoretical LFP profile of the pyramidal population are highly similar to the LFP vector (cosine similarity > 0.97 for LFP vector of all three frequencies in Figure 6D). Note that the sign of the LFP vector, only reflects our convention of setting the phase of its mean to zero: as the LFP vector coefficients are divided into two groups of opposite sign, a positive sign is attributed to the set of coefficients that weight the most in the overall sum. In the context of laminar recordings, one could as well adopt a different convention ascribing a fixed sign to peri-somatic LFP, reflecting the sign obtained through classical analyses, e. g. triggered averaging based on spikes or oscillatory peak (Sullivan et al., 2011). This suggest that the LFP vector can be exploited to further study the current sources and sinks causing the LFP, e. g. through current source density analysis (Wójcik, 2013).

The spike vectors components distribution in the complex plane (Figure 6E) support that both E and I cells are synchronized to CA1 LFP in the ripple band (80-180 Hz), but at a different phase, in line with experimental and simulation results (Buzsáki et al., 1992; Ramirez-Villegas et al., 2018). This extends our previous observation of Figure 5D to a more realistic case of recurrent E-I interactions. Notably, we observe that while GPLA does not exploit neuron type information, the spike vector can clearly differentiate pyramidal cells from interneurons based on their respective phase in the spike vector, showing that interneurons lead pyramidal cells in lower frequency bands, barring a drastic change in the distribution of phases in high frequencies (see also Figure 6F). This can be used not only for inferring cell types from experimental data Varga et al. (2014), but also to infer the organization and dynamics of neural circuits based on analytical or computational models.

In order to illustrate this second approach, we use the linear response framework (see above section Low rank linear response theory and frequency analysis) to analyze the phase of the populations in neural mass models of the microcircuit activity with different levels of complexity. In line with (Ledoux and Brunel, 2011), we first took into account only membrane time constants in the *Mass2D* model (resulting from membrane capacitance and leak currents), and neglected completely synaptic dynamics (see STAR Methods, section Analysis and simulation of two population neural mass models). As a result, *Mass2D* has a 2 dimensional state space, allowing for resonance due to interactions between pyramidal cells and interneurons (a mechanism exploited in the *PING* model of gamma oscillations (Traub et al., 1999)). However, as shown in Figure 6F, the predicted phase shift across frequencies could neither account for the driving by interneurons in CA1, nor for phase changes in high frequencies (>30Hz). Notably, incorporating strong feedforward inhibition (FFI) did not improve the qualitative match between the analytical predictions and GPLA's outcome. The inappropriateness of *Mass2D* is in line with the current understanding of SWR emergence in CA1 through the pacing of pyramidal activity by delayed I-I interactions (Stark et al., 2014), as the model does not account for such interactions.

The emergence of oscillations through I-I interactions is supported by mathematical modeling showing that sufficiently strong delayed recurrent inhibition gives rise to resonance or sustained oscillations (Brunel and Wang, 2003). The presence of these oscillations can be accounted for in an extended version of our neural mass model by including an additional synaptic delay and/or a synaptic time constant for I-I synapses (Ledoux and Brunel, 2011). We build the *MassAlpha* model that introduces a lag in the post-synaptic dynamics through the use of the so-called alpha synapses (see Methods for details). Interestingly, the resulting E-I phase shift of this model is in qualitative agreement with GPLA estimation (Figure 6F), exhibiting a reversal in the lead-lag relation between populations as frequency grows. This supports the ability of GPLA, combined with appropriate simulations, to capture key characteristics of the underlying circuitry (in this simulation having oscillation dynamics and synaptic delays).

Another interesting property of the network is the phase shift between each individual population and the LFP, which is simply reflected in the phases of the spike vector coefficients averaged across each population (E and I). Given that the LFP is a linear combination of all post-synaptic currents of the network, we can use modeling to evaluate which of these currents is the most repre-

sentative of the observed spike-LFP phase relation. As shown in Figure 6G (also see Supplementary Figure 9 for EPSP proxy), the choice of the IPSP as an LFP proxy in the MassAlpha model accounts qualitatively, as frequency increases, for (1) monotonous phase increase of the I population, (2) the phase slope reversal of the E population. Overall, the simulation results are also in line with the experimentally observed synchronizations of these populations. In contrast, using EPSP as an LFP proxy still fails to reproduce these two aspects (see Supplementary Figure 9), illustrating how GPLA, beyond microcircuit dynamics, may also help address the cellular underpinnings of experimentally observed LFP (Teleńczuk et al., 2017). This overall suggests that GPLA combined with neural mass modeling of a structure can provide insights into the microcircuit dynamics underlying phenomena as complex as sharp-wave ripples, despite neglecting many biophysical details.

Moreover, GPLA can also provide further insights when recordings from multiple regions are available. As an illustration, Figure 6H depicts coefficients of the spike vector for the joint analysis of CA1 and CA3 (LFP and spikes of both areas is used for GPLA) showing CA1 and CA3 neurons are all coupled to the field activity with cell-type-specific phases in the gamma band (20-40 Hz) (see supplemental Figure 11) that are consistent with the GPLAs obtained from individual structures (see supplemental Figure 10). This suggests that the gamma rhythm has a dominant coherent component spanning the two structures, and thus may support communication between subfields during SWR and consequently memory trace replay, consistently with current hypotheses (Carr et al., 2012).

Application to spatio-temporal patterns of neural field models

In order to assess the interpretability of GPLA in the context of electrode arrays covering a piece of the cortical surface, we simulated a 2D neural field model as described in Figure 5. We used an exponentially decaying horizontal excitatory connectivity with a spatial scale constant $r_0 = 440\mu m$, in line with recent analyses of cortical recordings Teleńczuk et al. (2017) (see STAR Methods section Analysis and simulations of neural field models for details). The spatio-temporal dynamics was down-sampled spatially on a grid with a step size $\Delta x = 800\mu m$, representing the inter-electrode distance of a putative electrode array of $1.2cm$ size. We compared the spatio-temporal dynamics for two choices of connectivity for which the input-free network has a stable equilibrium. First, we consider the *weak inhibition* case (Figure 7A), for which inhibitory (I) cells have weak feedback inhibition ($I \rightarrow E$), relative to the self-excitation caused by $E - E$ horizontal connections. The resulting activity is akin to stochastic fluctuations, due to the exogenous input, around a *stable node* equilibrium. Second, in the *strong inhibition* case (Figure 7B), the larger excitability of inhibitory neurons strengthens their influence on excitation and leads to activity fluctuating around a *stable spiral equilibrium*, reflecting a tendency of perturbations to oscillate around this point (Figure 7B) (Onslow et al., 2014). In both cases, the computed excitatory population rate is used to simulate the spike train of one excitatory unit per spatial electrode on this grid, in line with the observation that excitatory units are more easily detected experimentally due to their open field configuration (Logothetis and Panzeri, 2014). GPLA is then computed between this excitatory spiking activity and different LFP proxies. The results in Figure 7C-H are computed using the total EPSP resulting from horizontal E-E connections as LFP proxy (i. e. excluding exogenous excitation). We observe key differences between the GPLA of the two systems, predicted by linear response theory (see STAR Methods section Analysis and simulations of neural field models)

First, as reflected in the gPLV values (Figure 7C), spike-field coupling appears stronger in the lower frequency bands in the case of weak recurrent inhibition, while in the case of strong recurrent inhibition we observe a stronger coupling at intermediate frequencies. Notably, the peak of spike-field coupling in intermediate frequencies for strong inhibition is in line with models of the prefrontal cortex with the same enhanced feedback inhibition (Sherfey et al., 2018), exhibiting a resonance in the beta range (25Hz).

Second, as demonstrated in the previous neural mass model simulation, the global spike-LFP phase shift may also be informative about the underlying neural circuits. We can compute the

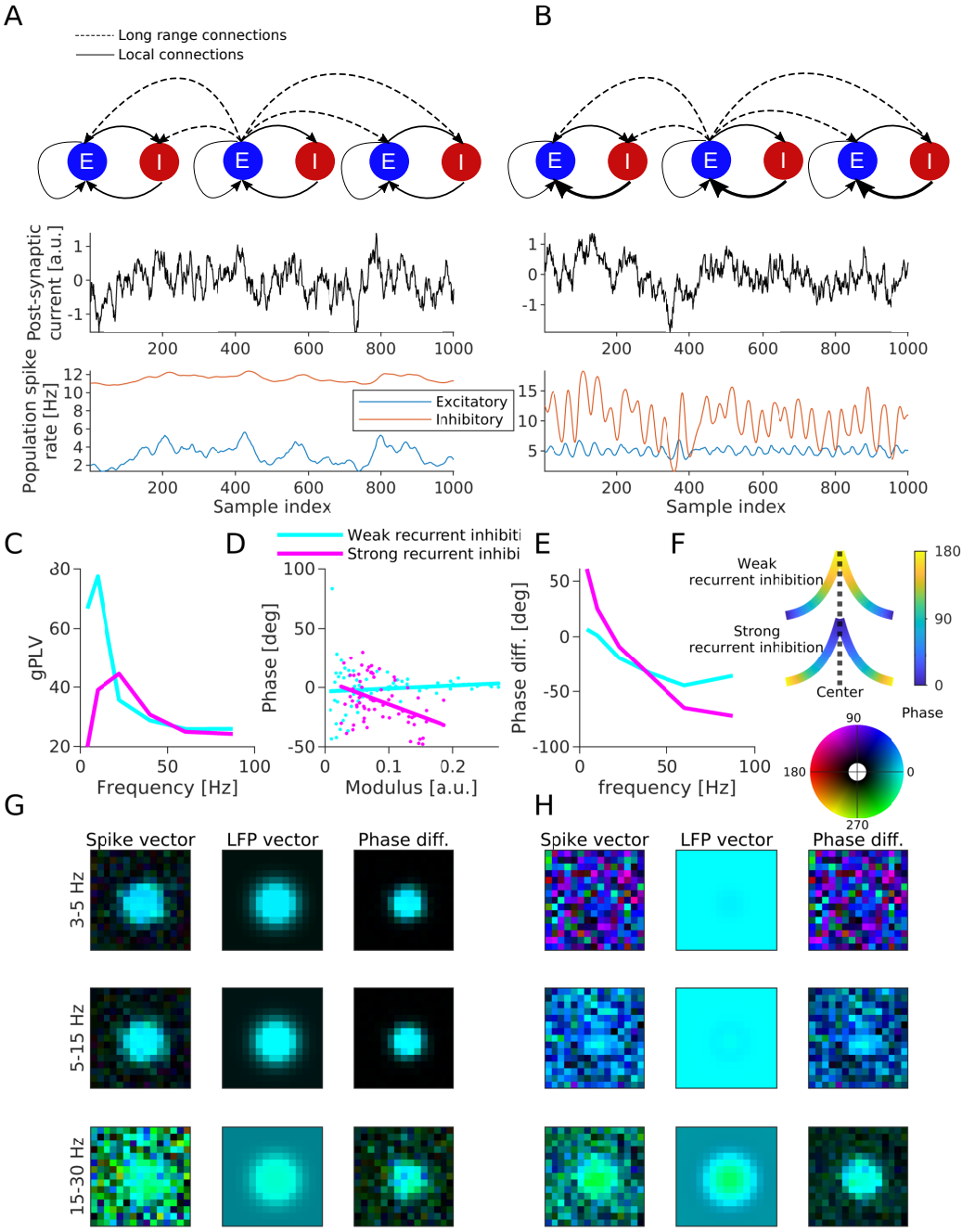


Figure 7. Neural field simulation. (A) Simulation of a neural field with weak recurrent inhibition. Example time course of (from top to bottom) exogenous input, E- and I- populations rates at the center of the neural field (in green, blue and red). **(B)** Same as A for a neural field strong recurrent inhibition, which shows oscillatory dynamics due to increase in recurrent activity. **(C)** gPLV as a function of frequency for both models. **(D)** Phase of spike vector coefficients as a function of its modulus for the frequency band associated with maximum gPLV for both models (each dot one coefficient, and the continuous lines are plotted based on linear regression). **(E)** Shift between the averaged phase of spike vector and averaged phase of LFP vector, as a function of frequency. **(F)** Schematic of the spike vector's phase gradient in the two models according to Equation 6. X-axis is the distance from center and y-axis is the connectivity strength. Line color indicates the phase according to the colorbar on the right. **(G)** Resulting GPLA in 3 frequency bands (indicated on the left) for neural field with weak recurrent inhibition (model schematized in A). **(H)** Same as G for neural field with strong recurrent inhibition (model schematized in B). In both G and H, color of pixel code the values of spike/LFP vector coefficients, with colorbar on top of H. Colors are represented in HSV mode, in which a complex number ($re^{i\phi}$) is represented by hue and brightness of a pixel. Hue of a pixel indicates the phase (ϕ) and the brightness of a pixel indicates the magnitude (r).

average phase shift between the spike and LFP vectors as a function of the frequency band to see a clear difference between the two models. Strong recurrent inhibition leads to phase advance of the spiking activity in the low frequency, in contrast with the weak recurrent inhibition case showing a consistent lag of excitatory spiking across frequencies (Figure 7E).

Third, the relationship between the spatial variations of modulus and phase of the spike vector is different across these two networks. In the simulation with strong recurrent inhibition, the phase of spike vector coefficients as a function of their modulus for the frequency band associated with maximum gPLV for each model indicates that the phase of the spike vector coefficients decreases (i. e. the oscillation lags further relative to the LFP) for larger modulus ($p < 10^{-4}$, linear regression), whereas, in the simulation with weak recurrent inhibition, phase is not correlated with modulus ($p > 0.3$, linear regression) (Figure 7D).

This last difference between the two cases can be directly interpreted based on the spatial maps of spike vector coefficients across the array. Indeed, models exhibit a different radial phase map in both situations (Figure 7E-H), reflecting how phase changes as magnitude decreases when going away from the center (the location with largest input). This gradient can be predicted by theoretical analysis of a one dimensional neural field as we show in detail in STAR Methods section Spatio-temporal phase analysis in 1D. Briefly, to obtain the spike vector, the input spatial pattern at a given temporal frequency f is approximately convolved by a spatial convolution kernel of the form

$$k(x) = e^{-|x|a(f)} = e^{-|x|\mathrm{Re}[a(f)]} e^{-i|x|\mathrm{Im}[a(f)]} . \tag{6}$$

The first term of this kernel has a negative real number multiplied by distance in the exponential that makes the activity decrease away from the locations where exogenous input is the highest, as intuitively expected from the horizontal connectivity of the circuit. In contrast, the imaginary number in the argument of the second term enforces a spatial phase gradient in response to the input, which depends on the sign of the imaginary part of a. If this sign is positive, responses at the location of highest input will be ahead of time with respect to their surrounding in the considered band, as reflected by their larger spike vector phase in the top illustration of Figure 7F. In case $\mathrm{Im}[a]$ is negative, locations with highest input are lagging behind (bottom illustration of Figure 7F). The frequency-dependent complex number $a(f)$ that controls this behavior reveals the key microcircuit properties that determine our observation through the approximation relation (valid a low frequency)

$$a^2 \approx \frac{1}{r_0^2}\left(1 + \nu_{E\leftarrow I}\nu_{I\leftarrow E} - \nu_{E\leftarrow E} - i2\pi\tau f(2\nu_{E\leftarrow I}\nu_{I\leftarrow E} - \nu_{E\leftarrow E})\right) . \tag{7}$$

It can be deduced from this expression (taking the square root of this number) that the imaginary part of a will depend on the relative strength of recurrent inhibition onto pyramidal cells, controlled by $\nu_{E\leftarrow I}\nu_{I\leftarrow E}$, with respect to recurrent excitation controlled by $\nu_{E\leftarrow E}$. Intuitively, $\nu_{E\leftarrow I}\nu_{I\leftarrow E} >> \nu_{E\leftarrow E}$ will tend to suppress the input that created the response, generating a wave converging to the points were the input was highest. The theory also predicts that large values of $\nu_{E\leftarrow I}\nu_{I\leftarrow E}$, as used in the strongly recurrent simulation, can generate strong phase gradients. In contrast, linear stability constrains the values of $\nu_{E\leftarrow E}$ to remain small, reflecting our choice for the simulations, and resulting in a comparatively moderate slope for the weakly recurrent case.

Overall, contrasting these two cases of neural field connectivity shows that changes in strength of feedback inhibition are reflected not only in the dominant frequency of spike-LFP synchronization (Figure 7C), but also in the spike-LFP shifts of the GPLA results (Figure 7E), and in the relationship between modulus and phase of spike vector coefficients (Figure 7D). Notably, these observations are being made in absence of specific oscillatory activity nor spatial phase gradient of the exogenous input (which influences the activity synchronously across the array). Therefore, it supports that the observation of complex coordinated activity involving phase spatial gradients, such as oscillations and traveling waves-like phase gradients, in array recordings may emerge from local

recurrent interactions in the recorded regions, instead of resulting from the passive driving by spatio-temporally coordinated activity originating from other brain regions.

Analysis of Utah array data in the prefrontal cortex

After testing the capabilities of GPLA on simulations, we investigated its performance on a large-scale electrophysiological dataset. We applied the method on Utah array (10×10 electrodes, inter-electrode distance $400\mu m$) recordings from the ventrolateral prefrontal cortex of one anaesthetized rhesus monkey (see Figure 8A). LFP signals were preprocessed as described in STAR Methods, section Animal preparation and intracortical recordings, and multi-unit activity with a minimum of 5 Hz firing rate was used. Recorded signals are exemplified in Figure 8B-F. Exemplary LFP traces are illustrated in Figure 8B. Each trace is recorded from the location specified in Figure 8C. Spike trains are also displayed in Figure 8D (for the same epoch used in Figure 8B). As the analysis is performed in band-limited frequency ranges we also exemplified band-passed LFP signals (together with spikes) in Figure 8E and F. The dataset consisted of 200 trials of visual stimulation (10 sec) and intertrials (10 sec) each 20 sec.

Computing GPLA in different frequency bands revealed that the strongest coupling was in the alpha range (5-15Hz) (Figure 8G). Furthermore, we assessed the significance of coupling with both surrogate and analytical tests (see STAR Methods, section Significance assessment of gPLV). GPLA above 50 or 60 Hz should be considered with caution, as in high frequencies the spike-LFP relationships can be largely effected by the local coupling between spikes and LFP recorded from the same electrode. Furthemore, contamination of high frequency LFP bands by spike waveforms (Zanos et al., 2011; Ray and Maunsell, 2011b) may bias spike-LFP coupling towards interactions within the same electrode contact instead of capturing coordination at the mesoscopic scale.

Figure 8I-K further shows the spike and LFP vectors for the three frequencies with largest coupling according to their gPLVs (for other frequencies see Supplementary Figure 13). Representing spike and LFP vectors in the complex plane (Figure 8I-J first column), suggests that the relative phases of spike and LFP vectors are different across these three frequencies. To demonstrate the difference more clearly, we fit von Mises distributions to the pooled phase of all coefficients of the vectors (Figure 8I-J second column). The sign of the spike-LFP phase differences changes across frequencies, with spikes ahead of time with respect to LFP in low frequency, while lagging at higher frequencies. This behavior is similar to the above analysis of strongly recurrent neural field model (Figure 7G), when EPSP is taken as an LFP proxy.

The spatial mappings of the LFP and spike vectors on the Utah array (Figure 8I-J, third and fourth column) also demonstrate a spatial structure in the modulus and phase of the LFP and spike vectors, revealing localized regions with stronger participation in the locking, in particular in the beta range 15-30 Hz (green pixels at the middle-top and -bottom in Figure 8K, fourth column). We hypothesize this is due to a higher activation of spatially localized populations, as supported by anatomical studies of the PFC (Elston, 2000, 2003) and electrophysiological (Safavi et al., 2018) studies.

Furthermore, in the lower frequencies, in particular the alpha band (5-15 Hz) exhibiting the strongest coupling between spike and LFP, the spike vector coefficients' moduli are significantly negatively correlated with their phase (Figure 8H ($p < 10^{-6}$, linear regression). Interestingly, we observe again a similar behavior in the above neural field simulation with strong recurrent inhibition, but not in the simulation with weak recurrent inhibition (Figure 7F). Overall, these results suggest a neural field with excitatory horizontal connections and strong local recurrent inhibition as a plausible model for the recorded prefrontal circuits, in line with what has been suggested by previous modeling work (Sherfey et al., 2018, 2020).

Discussion

In spite of the strong relation between LFP activity and the neural mechanisms involved in the coordination of the underlying networks (Buzsaki et al., 2012, 2013; Einevoll et al., 2013; Herreras, 2016; Pesaran et al., 2018), spike-LFP relationship is still not systematically exploited in extensive

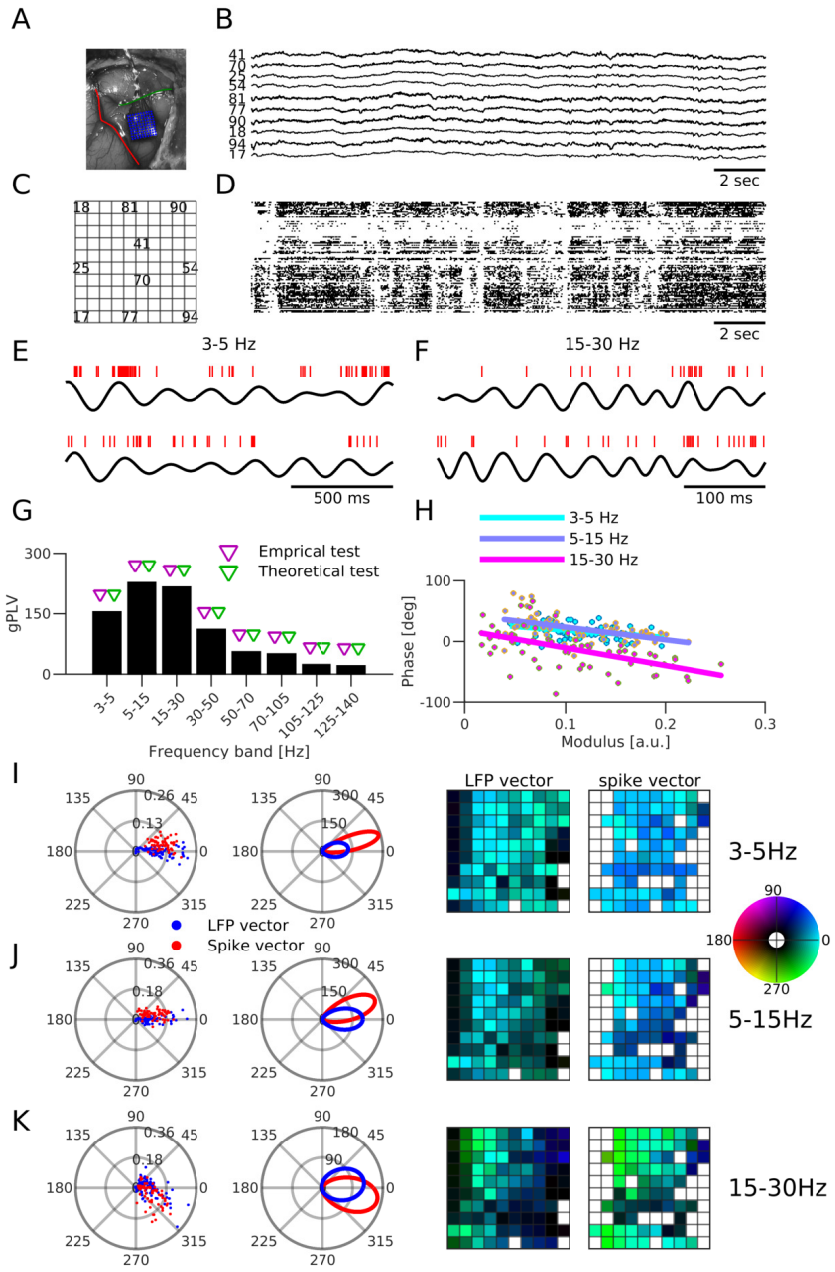


Figure 8. Application to electrophysiological recordings in non-human primate PFC.
(A) Location of the Utah array, anterior to the arcuate sulcus (red line) and inferior to the principal sulcus (green line). **(B)** Broadband trace of the recorded LFP (from the recording channels indicated in C). **(C)** Utah array spatial map identifying channel IDs shown in B. **(D)** Spike rasters for all recorded neurons. **(E-F)** Example spike trains (red bars) and filtered LFP (black traces) in the frequency ranges (E) 3-5 Hz and (F) 15-30 Hz. **(G)** gPLV values. Triangles indicate the significance assessed based on surrogate (blue triangles) and analytical test (red triangles) tests. **(H)** Phase of spike vector coefficients as a function of its modulus for the frequencies indicated in the legend (one dot per coefficient, continuous lines indicate linear regression). **(I-K)** LFP and spike vectors for frequency (I) 3-5 Hz and (F) 15-30 Hz. First column depicts the LFP (blue dots) and spike (red dots) in the complex plane. Second column depicts the fitted von Mises distribution to phase of LFP and spike vectors. Third and forth columns respectively represent the spatial distribution of phase of LFP and spike vectors values on the array (see C). White pixels in the third column (LFP vector) indicate the recording channels that were not used in the recording and in the forth column (spike vector), white pixels indicate the recording channels with insufficient number of spikes (multiunit activity with a minimum of 5 Hz firing). In the last two columns, colors are represented in HSV mode, in which a complex number ($re^{i\phi}$) is represented by hue and brightness of a pixel. Hue of a pixel indicates the phase (ϕ) and the brightness of a pixel indicates the modulus (r). The colorbar is depicted on the right.

quantitative analysis of highly multivariate neural recordings. One potential reason could be the lack of multivariate methodologies for investigating spike-LFP coupling beyond a single pair of spiking unit and a LFP channel.

In this study, we develop the Generalized Phase Locking Analysis (GPLA), to the best of our knowledge, as the first biophysically interpretable *multivariate* method for investigating the coupling between spatially distributed spiking and LFP activity. GPLA summarizes the coupling between multiple LFP spatio-temporal patterns and multiple spiking units in a compact way. At a given frequency, the spike and LFP vectors represent the dominant LFP and spiking spatio-temporal distribution and the generalized Phase Locking Value (gPLV), characterizes the strength of the coupling between LFP and spike patterns.

We demonstrate that by using outcomes of GPLA such as the overall spike-LFP phase shift, the phase shift between different cell types (excitatory and inhibitory), and the spatial phase gradients, we can extract information about the overall organization of the recorded structure. In particular, we demonstrated that GPLA features can reflect properties pertaining to the organization of recurrent interactions in the recorded region that are not directly accessible by simple measurements. First, application to realistic simulations of hippocampal SWR revealed various characteristics of hippocampal circuitry with minimal prior knowledge. Second, in order to better understand the interpretation of spike and LFP vectors' spatial structure, we also simulated spatially extended neural field models and demonstrate that phase gradients of spike and LFP vectors in these neural field models reflect the underlying microcircuit connectivity (such as the strength of recurrent interactions). Finally, the application of GPLA to experimental recordings suggests a global coupling between spiking activity and LFP traveling wave in vlPFC in line with our simulations of a neural field endowed with strong recurrent inhibition.

Statistical properties of the gPLV were investigated to develop an empirical and theoretical framework for the assessment of the significance of the coupling. The theoretical statistical test makes the method applicable to high dimensional electrophysiology data with low run-time complexity, which is important for modern probes such as Neuropixel with 960 recording sites (Jun et al., 2017). Notably, the conventional statistical testing procedures based on the generation of surrogate data are computationally very expensive as their computation time is scaled by the number of recorded neurons and number of surrogate datasets generated for construction of a null distribution.

In summary, GPLA is a powerful tool to quantify, statistically assess and interpret the interactions between spiking activity and large-scale dynamics of current recording techniques.

Comparison to existing approaches

To the best of our knowledge, there are very few studies that include the information of multiple LFP channels *and* multiple spiking units for investigating spike-LFP interactions. Among those few studies which employ a multi-variate approach, none fully exploited the multivariate nature of spiking activity recorded from multiple sites, but instead, restrict the multivariate aspect only to LFP channels.

Spike-Triggered Average (STA) of LFP is one of the common multivariate technique for characterizing spike-LFP relationship (Jin et al., 2008; Nauhaus et al., 2009). Spike-triggered average of LFP has been also interpreted as a measure of functional connectivity (Nauhaus et al., 2009), therefore it is also interpretable (but also see Ray and Maunsell (2011a) and Nauhaus et al. (2012)). Although in LFP spike-triggered averaging the multivariate aspect of the LFP signals is exploited, ultimately only individual spiking units are being used as each spike-triggered average is only computed based on a single spiking unit. However, considering the information of all spiking units simultaneously can be potentially informative. For instance, the phase difference between different types of neurons (e. g. excitatory and inhibitory neurons) (Klausberger et al., 2003) can inform us about the organization of the neural circuit (as we illustrate with one of our simulations) Varga et al. (2014). Notably, even more sophisticated extensions of spike-triggered averaging of LFP (Teleńczuk et al.,

2017) are still using information of the individual spiking units in their methodology. In a similar vein, Canolty et al. (2010) by exploiting maximum entropy models, showed that the probability of spiking can be modulated not only by the phase of the local LFP, but also by the LFP phase in multiple distant regions. This study was also limited to individual spiking units.

Limitations and potential extensions

One limitation of GPLA is that it considers the underlying networks dynamics to be fixed for the analyzed data. Identifying transient activities that potentially show differentiated signatures of cell assemblies activation is of paramount importance (Harris, 2005; Buzsaki, 2010). As mentioned earlier, LFPs result from the superposition of electric potentials from multiple sources and can capture various *coordinated or cooperative* phenomena. LFP decomposition techniques can temporally isolate these epochs of coordinated activity and application of GPLA to these epochs can characterize how each neuron is participating in the collective activity and/or to what degree, it is coupled to the larger-scale dynamics.

Furthermore, GPLA can also be improved by exploiting a better univariate estimation method. Various novel methodologies for assessing the coupling between a pair of spike train and LFP time course have been developed in recent years (Grasse and Moxon, 2010; Vinck et al., 2010; Lepage et al., 2011; Vinck et al., 2012; Jiang et al., 2015; Zarei et al., 2018; Li et al., 2016) each providing some improvements over classical measures such as PLV and SFC. Replacing coefficients of coupling matrix (Eq 15) with better estimators of the coupling between spike and LFPs can bring those benefits to GPLA as well. For instance, Zarei et al. (2018) proposed a bias-free estimation of spike-LFP coupling in low firing regime which can also improve GPLA if we replace the coefficients of the coupling matrix (Equation 15) with the new estimator. Nevertheless, the current settings i. e. using a normalized version of the conventional PLV (Equation 13) is suitable for the theoretical statistical test that we introduce in Analytical test. Alternative univariate coupling estimates should be adapted to the requirement of the Random Matrix Theory to preserve the statistical benefits of our approach. This typically requires knowing the asymptotic distribution of the coupling statistics and devising and appropriate normalization thereafter. In case the new coupling measures are not adaptable to theoretical statistical testing, the Surrogate-based test that we discussed earlier remains applicable at the expense of heavier computational costs.

Neuroscientific interpretation of GPLA

Due to the complexity of the structure and dynamics of spatially extended neural networks, even a statistically sound approach such as GPLA (as demonstrated in simulations), might be challenging to interpret in terms of biological mechanisms. In order to ease interpretability, we related it to linear response analysis of neural field models. Thanks to this approach, we could link several features of GPLA to a mechanistic interpretation. First, we chose a biophysically realistic model of hippocampal ripples in order to use a system for which the underlying mechanism are (1) well understood, (2) more complex than the neural mass models used to interpret GPLA results. Despite the discrepancy between models, this showed that increasing the complexity of neural mass models using properties that are qualitatively in line with the key ground truth underlying mechanisms (e. g. inhibitory synaptic delays), allowed reproducing qualitatively GPLA results of these simulations, making the approach more interpretable. This allowed in particular (1) to show the LFP vector reflects the laminar distribution of field potential generated by current dipoles, (2) to link the phases of the spike vector to cell types and recurrent I-I dynamics,

Next, we used neural field simulations in order to find interpretations of GPLA characteristics that can be exploited in the context of electrode array recording in the cortex. This is an important step as the mechanisms underlying spatio-temporal phenomena observed in *in vivo* recordings remain largely elusive. While keeping the complexity of these models minimal (using exponentially decaying horizontal excitation and local inhibition), we could already observe the microcircuit structure produced non-trivial qualitative changes in the GPLA outcome, in particular regarding

the phase gradients of spike and LFP vectors across the array. Finally, our analysis of Utah array recordings suggests the key GPLA features exhibited in simulation can also be estimated in real data and provide insights in the underlying organization of the circuit.

Overall, the simple rate models we investigated have the benefit of lending themselves to approximate analytical treatment, providing direct insights into the role played by network parameters in GPLA characteristics. However, refinements of these first analyses are possible by improving the biological realism of population models and will allow checking the robustness of our simplifying assumptions. In particular, neural data recorded with multi-electrode arrays may be investigated in light of the knowledge about the horizontal connectivity of the structure, which may not be monotonous (for example see recent findings on non-monotonous correlation structure in V1 (Rosenbaum et al., 2017) and PFC (Safavi et al., 2018)).

More generally, a mechanistic interpretation of GPLA results in a given structure strongly relies on the accuracy of the assumptions made to perform analytical and/or computational modeling. We have shown neural mass models can approximate the population behavior of rather complex dynamics, such as sharp-wave ripples, and further sophistication of mass models will certainly help generalize the interpretability of GPLA analysis to include more biophysically realistic results (Jirsa and Haken, 1996; Qubbaj and Jirsa, 2007; Costa et al., 2016).

Ultimately, we believe GPLA provides insights for understanding the distributed information processing in higher-tier cortical areas such as PFC and hippocampus where investigating distributed spike-LFP interactions has been shown to be insightful for elucidating the mechanisms of cognitive functions such as working memory, memory consolidation and spatial navigation. Indeed, previous studies exploited spike-LFP relationship to shed light on the coordination mechanisms involved in working memory (Liebe et al., 2012; Markowitz et al., 2015), memory consolidation and spatial navigation (Taxidis et al., 2015; Fernández-Ruiz et al., 2017). Certainly, exploiting multivariate methods for investigating spike-LFP relationships can provide further insights about the coordination mechanisms involved in such cognitive functions.

Acknowledgments

We thank Britni Crocker for help with preprocessing of the data and spike sorting; Joachim Werner and Michael Schnabel for their excellent IT support. Andreas Tolias for help with the initial implantation's of the Utah arrays; scidraw.io for providing a free repository of high-quality scientific drawings (in particular Macauley Smith Breault for providing her brain drawing in this repository). This work was supported by the Max Planck Society.

Author contribution

Conceptualization, S.S., T.I.P., M.B.; Methodology, S.S., J.F.R.-V. and M.B.; Software, S.S. and M.B.; Formal Analysis, S.S. and M.B.; Investigation, S.S., T.I.P., V.K. and M.B.; Resources, N.K.L.; Data Curation, S.S., T.I.P., V.K., and M.B.; Writing - Original Draft, S.S. and M.B.; Writing - Review & Editing: S.S., T.I.P., V.K., J.F.R.-V., N.K.L. and M.B.; Visualization, S.S. and M.B.; Supervision and Project administration, T.I.P. and M.B.; Funding acquisition, N.K.L.

References

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and event history analysis: a process point of view*. Statistics for Biology and Health. Springer, New York, NY. OCLC: 254319944.

Abramowitz, M., Stegun, I. A., et al. (1972). Handbook of mathematical functions with formulas, graphs, and mathematical tables.

Anderson, G. W., Guionnet, A., and Zeitouni, O. (2010). *An Introduction to Random Matrices*. Cambridge University Press, Cambridge; New York.

Ashida, G., Wagner, H., and Carr, C. E. (2010). Processing of Phase-Locked Spikes and Periodic Signals. In *Analysis of Parallel Spike Trains*, Springer Series in Computational Neuroscience, pages 59–74. Springer, Boston, MA.

Belitski, A., Gretton, A., Magri, C., Murayama, Y., Montemurro, M. A., Logothetis, N. K., and Panzeri, S. (2008). Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. *J Neurosci*, 28(22):5696–709.

Brunel, N. and Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? i. synaptic dynamics and excitation-inhibition balance. *Journal of Neurophysiology*, 90(1):415–430.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7(5):446–451.

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford University Press.

Buzsaki, G. (2010). Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362–85.

Buzsaki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents–eeg, ecog, lfp and spikes. *Nat Rev Neurosci*, 13(6):407–20.

Buzsáki, G., Horváth, Z., Urioste, R., Hetke, J., and Wise, K. (1992). High-frequency network oscillation in the hippocampus. *Science*, 256(5059):1025–1027.

Buzsaki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron*, 80(3):751–64.

Canolty, R. T., Ganguly, K., Kennerley, S. W., Cadieu, C. F., Koepsell, K., Wallis, J. D., and Carmena, J. M. (2010). Oscillatory phase coupling coordinates anatomically dispersed functional cell assemblies. *Proc Natl Acad Sci U S A*, 107:17356–61.

Capitaine, M. and Donati-Martin, C. (2016). Spectrum of deformed random matrices and free probability. *arXiv preprint arXiv:1607.05560*.

Carnevale, N. T. and Hines, M. L. (2006). *The NEURON Book*. Cambridge University Press, Cambridge, UK ; New York, illustrated edition edition.

Carr, M. F., Karlsson, M. P., and Frank, L. M. (2012). Transient slow gamma synchrony underlies hippocampal memory replay. *Neuron*, 75(4):700–713.

Chavez, M., Besserve, M., Adam, C., and Martinerie, J. (2006). Towards a proper estimation of phase synchronization from time series. *J Neurosci Methods*, 154(1-2):149–60.

Costa, M. S., Weigenand, A., Ngo, H.-V. V., Marshall, L., Born, J., Martinetz, T., and Claussen, J. C. (2016). A Thalamocortical Neural Mass Model of the EEG during NREM Sleep and Its Response to Auditory Stimulation. *PLOS Computational Biology*, 12(9):e1005022.

Csicsvari, J., Hirase, H., Mamiya, A., and Buzsáki, G. (2000). Ensemble Patterns of Hippocampal CA3-CA1 Neurons during Sharp Wave–Associated Population Events. *Neuron*, 28(2):585–594.

Datta, B. N. (2010). *Numerical Linear Algebra and Applications*.

Denker, M., Roux, S., Timme, M., Riehle, A., and Grun, S. (2007). Phase synchronization between LFP and spiking activity in motor cortex during movement preparation. *Neurocomputing*, 70(10-12):2096–2101.

Dickey, A. S., Suminski, A., Amit, Y., and Hatsopoulos, N. G. (2009). Single-Unit Stability Using Chronically Implanted Multielectrode Arrays. *J. Neurophysiol.*, 102(2):1331–1339.

Dwarakanath, A., Kapoor, V., Werner, J., Safavi, S., Fedorov, L. A., Logothetis, N. K., and Panagiotaropoulos, T. I. (2020). Prefrontal state fluctuations control access to consciousness. *bioRxiv*, page 2020.01.29.924928.

Einevoll, G. T., Destexhe, A., Diesmann, M., Grün, S., Jirsa, V., de Kamps, M., Migliore, M., Ness, T. V., Plesser, H. E., and Schürmann, F. (2019). The Scientific Case for Brain Simulations. *Neuron*, 102(4):735–744.

Einevoll, G. T., Kayser, C., Logothetis, N. K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat Rev Neurosci*, 14(11):770–85.

Elston, G. N. (2000). Pyramidal cells of the frontal lobe: All the more spinous to think with. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20:RC95.

Elston, G. N. (2003). Cortex, cognition and the cell: new insights into the pyramidal neuron and prefrontal function. *Cerebral Cortex*, 13(11):1124–1138.

Engel, A. K., Fries, P., Konig, P., Brecht, M., and Singer, W. (1999). Temporal binding, binocular rivalry, and consciousness. *Conscious Cogn*, 8(2):128–51.

Ermentrout, B. and Pinto, D. (2007). Neurophysiology and waves.

Ermentrout, G. B. and Kleinfeld, D. (2001). Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron*, 29(1):33–44.

Fernández-Ruiz, A., Oliva, A., Nagy, G. A., Maurer, A. P., Berényi, A., and Buzsáki, G. (2017). Entorhinal-CA3 Dual-Input Control of Spike Timing in the Hippocampus by Theta-Gamma Coupling. *Neuron*, 93(5):1213–1226.e5.

Fisher, N. I. (1995). *Statistical Analysis of Circular Data*. Univ. Press, Cambridge, repr., 1. paperback ed edition.

Fletcher, C. (1991). Computational techniques for fluid dynamics; vol 1.

Fried, I., Rutishauser, U., Cerf, M., and Kreiman, G., editors (2014). *Single Neuron Studies of the Human Brain: Probing Cognition*. The MIT Press, Cambridge, Massachusetts.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends Cogn Sci*, 9(10):474–80.

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, 88:220–35.

Fukushima, M., Chao, Z. C., and Fujii, N. (2015). Studying brain functions with mesoscopic measurements: Advances in electrocorticography for non-human primates. *Current Opinion in Neurobiology*, 32:124–131.

Gao, P. and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, 32:148–155.

Grasse, D. W. and Moxon, K. A. (2010). Correcting the bias of spike field coherence estimators due to a finite number of spikes. *Journal of neurophysiology*, 104(1):548–58.

Grosmark, A. D. and Buzsáki, G. (2016). Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. *Science*, 351(6280):1440–1443.

Grosmark, A. D., Mizuseki, K., Pastalkova, E., Diba, K., and Buzsáki, G. (2012). REM Sleep Reorganizes Hippocampal Excitability. *Neuron*, 75(6):1001–1007.

Grün, S. (2009). Data-Driven Significance Estimation for Precise Spike Correlation. *Journal of Neurophysiology*, 101(3):1126–1140.

Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nature reviews. Neuroscience*, 6(5):399–407.

Herreras, O. (2016). Local Field Potentials: Myths and Misunderstandings. *Front Neural Circuit*, 10:101.

Jiang, H., Bahramisharif, A., van Gerven, M. A. J., and Jensen, O. (2015). Measuring directionality between neuronal oscillations of different frequencies. *NeuroImage*, 118:359–367.

Jin, J. Z., Weng, C., Yeh, C.-I., Gordon, J. A., Ruthazer, E. S., Stryker, M. P., Swadlow, H. A., and Alonso, J.-M. (2008). On and off domains of geniculate afferents in cat primary visual cortex. *Nat. Neurosci.*, 11(1):88–94.

Jirsa, V. K. and Haken, H. (1996). Field theory of electromagnetic brain activity. *Physical Review Letters*, 77(5):960–963.

Juavinett, A. L., Bekheet, G., and Churchland, A. K. (2019). Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *eLife*, 8:e47188.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydin, C., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Hausser, M., Karsh, B., Ledochowitsch, P., Lopez, C. M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P. D., Rossant, C., Sun, W. L., Svoboda, K., Carandini, M., Harris, K. D., Koch, C., O'Keefe, J., and Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551:232–236.

Klausberger, T., Magill, P. J., Marton, L. F., Roberts, J. D., Cobden, P. M., Buzsaki, G., and Somogyi, P. (2003). Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature*, 421(6925):844–8.

Ledoux, E. and Brunel, N. (2011). Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. *Frontiers in Computational Neuroscience*, 5:25.

Lepage, K. Q., Kramer, M. A., and Eden, U. T. (2011). The dependence of spike field coherence on expected intensity. *Neural computation*, 23(9):2209–41.

Li, C. Y., Poo, M. M., and Dan, Y. (2009). Burst spiking of a single cortical neuron modifies global brain state. *Science*, 324(5927):643–6.

Li, Z., Cui, D., and Li, X. (2016). Unbiased and robust quantification of synchronization between spikes and local field potential. *Journal of neuroscience methods*, 269:33–8.

Liebe, S., Hoerzer, G. M., Logothetis, N. K., and Rainer, G. (2012). Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat. Neurosci.*, 15(3):456–462, S1–2.

Liljenstroem, H. (2012). Mesoscopic brain dynamics. *Scholarpedia*, 7(9):4601.

Lindén, H., Tetzlaff, T., Potjans, T. C., Pettersen, K. H., Grün, S., Diesmann, M., and Einevoll, G. T. (2011). Modeling the Spatial Reach of the LFP. *Neuron*, 72(5):859–872.

Logothetis, N., Merkle, H., Augath, M., Trinath, T., and Ugurbil, K. (2002). Ultra high-resolution fmri in monkeys with implanted rf coils. *Neuron*, 35(2):227–42.

Logothetis, N. K., Guggenberger, H., Peled, S., and Pauls, J. (1999). Functional imaging of the monkey brain. *Nat Neurosci*, 2(6):555–62.

Logothetis, N. K. and Panzeri, S. (2014). Local Field Potential, Relationship to BOLD Signal. In Dieter Jaeger, R. J., editor, *Encyclopedia of Computational Neuroscience*, pages 1–11. Springer New York.

Marchenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536.

Markowitz, D. A., Curtis, C. E., and Pesaran, B. (2015). Multiple component networks support working memory in prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*.

Maynard, E. M., Nordhausen, C. T., and Normann, R. A. (1997). The utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalography and clinical neurophysiology*, 102(3):228–239.

Mazzoni, A., Linden, H., Cuntz, H., Lansner, A., Panzeri, S., and Einevoll, G. T. (2015). Computing the Local Field Potential (LFP) from Integrate-and-Fire Network Models. *PLoS computational biology*, 11(12):e1004584.

Moran, R. J., Kiebel, S. J., Stephan, K. E., Reilly, R., Daunizeau, J., and Friston, K. J. (2007). A neural mass model of spectral responses in electrophysiology. *NeuroImage*, 37(3):706–720.

Mukamel, R. and Fried, I. (2011). Human Intracranial Recordings and Cognitive Neuroscience. *Annu. Rev. Psychol.*, 63(1):511–537.

Nauhaus, I., Busse, L., Carandini, M., and Ringach, D. L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. *Nature neuroscience*, 12(1):70–6.

Nauhaus, I., Busse, L., Ringach, D. L., and Carandini, M. (2012). Robustness of traveling waves in ongoing activity of visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(9):3088–94.

Niebur, E., Koch, C., and Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Res*, 33(18):2789–802.

Nó, R. L. D. (1947). Action potential of the motoneurons of the hypoglossus nucleus. *J. Cell. Comp. Physiol.*, 29(3):207–287.

Nunez, P. L. and Srinivasan, R. (2006). *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA.

O'Leary, T., Sutton, A. C., and Marder, E. (2015). Computational models in the age of large datasets. *Current Opinion in Neurobiology*, 32:87–94.

Oliva, A., Fernandez-Ruiz, A., Buzsaki, G., and Berenyi, A. (2016). Role of Hippocampal CA2 Region in Triggering Sharp-Wave Ripples. *Neuron*, 91:1342–55.

Onslow, A. C., Jones, M. W., and Bogacz, R. (2014). A canonical circuit for generating phase-amplitude coupling. *PloS one*, 9(8):e102591.

Pesaran, B., Vinck, M., Einevoll, G. T., Sirota, A., Fries, P., Siegel, M., Truccolo, W., Schroeder, C. E., and Srinivasan, R. (2018). Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation. *Nat. Neurosci.*, page 1.

Peterson, E. J. and Voytek, B. (2018). Healthy oscillatory coordination is bounded by single-unit computation. *bioRxiv*, page 309427.

Pinotsis, D. A., Moran, R. J., and Friston, K. J. (2012). Dynamic causal modeling with neural fields. *Neuroimage*, 59(2):1261–1274.

Pinsky, P. F. and Rinzel, J. (1994). Intrinsic and network rhythmogenesis in a reduced traub model for CA3 neurons. *J Comput Neurosci*, 1(1):39–60.

Platkiewicz, J., Stark, E., and Amarasingham, A. (2017). Spike-Centered Jitter Can Mistake Temporal Structure. *Neural Computation*, 29(3):783–803.

Qubbaj, M. R. and Jirsa, V. K. (2007). Neural field dynamics with heterogeneous connection topology. *Physical review letters*, 98:238102.

Quiroga, R. (2007). Spike sorting. *Scholarpedia*, 2(12):3583.

Rabiner, L. R., Mcclellan, J. H., and Parks, T. W. (1975). Fir digital-filter design techniques using weighted chebyshev approximation. *Proceedings of the IEEE*, 63(4):595–610.

Ramirez-Villegas, J. F., Logothetis, N. K., and Besserve, M. (2015). Diversity of sharp-wave-ripple LFP signatures reveals differentiated brain-wide dynamical events. *Proceedings of the National Academy of Sciences of the United States of America*, 112:E6379–87.

Ramirez-Villegas, J. F., Willeke, K. F., Logothetis, N. K., and Besserve, M. (2018). Dissecting the synapse- and frequency-dependent network mechanisms of in vivo hippocampal sharp wave-ripples. *Neuron*, 100(5):1224–1240.e13.

Rasch, M., Logothetis, N. K., and Kreiman, G. (2009). From neurons to circuits: Linear estimation of local field potentials. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29:13785–96.

Rasch, M. J., Gretton, A., Murayama, Y., Maass, W., and Logothetis, N. K. (2008). Inferring spike trains from local field potentials. *J Neurophysiol*, 99(3):1461–76.

Ray, S. and Maunsell, J. H. (2011a). Network rhythms influence the relationship between spike-triggered local field potential and functional connectivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(35):12674–82.

Ray, S. and Maunsell, J. H. R. (2011b). Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biology*, 9(4):e1000610.

Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. (2017). The spatial structure of correlated neuronal variability. *Nature neuroscience*, 20:107–114.

Safavi, S., Dwarakanath, A., Kapoor, V., Werner, J., Hatsopoulos, N. G., Logothetis, N. K., and Panagiotaropoulos, T. I. (2018). Nonmonotonic spatial structure of interneuronal correlations in prefrontal microcircuits. *PNAS*, page 201802356.

Safavi, S., Logothetis, N. K., and Besserve, M. (2020). From univariate to multivariate coupling between continuous signals and point processes: A mathematical framework. *ArXiv200504034 Q-Bio Stat*.

Schomburg, E. W., Anastassiou, C. A., Buzsáki, G., and Koch, C. (2012). The Spiking Component of Oscillatory Extracellular Potentials in the Rat Hippocampus. *J. Neurosci.*, 32(34):11798–11811.

Schwalm, M., Schmid, F., Wachsmuth, L., Backhaus, H., Kronfeld, A., Aedo Jury, F., Prouvot, P. H., Fois, C., Albers, F., van Alst, T., Faber, C., and Stroh, A. (2017). Cortex-wide BOLD fMRI activity reflects locally-recorded slow oscillation-associated calcium waves. In *eLife*, volume 6.

Sherfey, J., Ardid, S., Miller, E. K., Hasselmo, M. E., and Kopell, N. J. (2020). Prefrontal oscillations modulate the propagation of neuronal activity required for working memory. *Neurobiology of Learning and Memory*, 173:107228.

Sherfey, J. S., Ardid, S., Hass, J., Hasselmo, M. E., and Kopell, N. J. (2018). Flexible resonance in prefrontal networks with strong feedback inhibition. *PLOS Computational Biology*, 14(8):e1006357.

Sik, A., Penttonen, M., Ylinen, A., and Buzsaki, G. (1995). Hippocampal CA1 interneurons: An in vivo intracellular labeling study. *J. Neurosci.*, 15(10):6651–6665.

Somers, D. C., Nelson, S. B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.*, 15(8):5448–5465.

Stark, E., Roux, L., Eichler, R., Senzai, Y., Royer, S., and Buzsáki, G. (2014). Pyramidal cell-interneuron interactions underlie hippocampal ripple oscillations. *Neuron*, 83(2):467–480.

Stevenson, I. H. and Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, 14(2):139–142.

Sullivan, D., Csicsvari, J., Mizuseki, K., Montgomery, S., Diba, K., and Buzsáki, G. (2011). Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: influence of dentate and entorhinal cortical activity. *Journal of Neuroscience*, 31(23):8605–8616.

Taxidis, J., Anastassiou, C. A., Diba, K., and Koch, C. (2015). Local Field Potentials Encode Place Cell Ensemble Activation during Hippocampal Sharp Wave Ripples. *Neuron*, 87(3):590–604.

Taxidis, J., Coombes, S., Mason, R., and Owen, M. R. (2012). Modeling sharp wave-ripple complexes through a CA3-CA1 network model with chemical synapses. *Hippocampus*, 22(5):995–1017.

Teleńczuk, B., Dehghani, N., Le Van Quyen, M., Cash, S. S., Halgren, E., Hatsopoulos, N. G., and Destexhe, A. (2017). Local field potentials primarily reflect inhibitory neuron activity in human and monkey cortex. *Sci. Rep.*, 7:40211.

Traub, R. D. and Miles, R. (1995). Pyramidal cell-to-inhibitory cell spike transduction explicable by active dendritic conductances in inhibitory cell. *Journal of computational neuroscience*, 2(4):291–298.

Traub, R. D., Whittington, M. A., Buhl, E. H., Jefferys, J. G., and Faulkner, H. J. (1999). On the mechanism of the gamma –> beta frequency shift in neuronal oscillations induced in rat hippocampal slices by tetanic stimulation. *The Journal of Neuroscience*, 19(3):1088–1105.

Varga, C., Oijala, M., Lish, J., Szabo, G. G., Bezaire, M., Marchionni, I., Golshani, P., and Soltesz, I. (2014). Functional fission of parvalbumin interneuron classes during fast network events. *Elife*, 3.

Vinck, M., Battaglia, F. P., Womelsdorf, T., and Pennartz, C. (2012). Improved measures of phase-coupling between spikes and the Local Field Potential. *J Comput Neurosci*, 33(1):53–75.

Vinck, M., van Wingerden, M., Womelsdorf, T., Fries, P., and Pennartz, C. M. (2010). The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. *Neuroimage*, 51(1):112–22.

Williamson, R. C., Doiron, B., Smith, M. A., and Yu, B. M. (2019). Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Current Opinion in Neurobiology*, 55:40–47.

Wilson, H. R. and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80.

Wójcik, D. K. (2013). Current Source Density (CSD) Analysis. In Jaeger, D. and Jung, R., editors, *Encyclopedia of Computational Neuroscience*, pages 1–10. Springer, New York, NY.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., and Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science*, 316:1609–12.

Zanos, T. P., Mineault, P. J., and Pack, C. C. (2011). Removal of spurious correlations between spikes and local field potentials. *Journal of neurophysiology*, 105(1):474–86.

Zarei, M., Jahed, M., and Daliri, M. R. (2018). Introducing a comprehensive framework to measure spike-lfp coupling. *Frontiers in Computational Neuroscience*, 12.

Zeitler, M., Fries, P., and Gielen, S. (2006). Assessing neuronal coherence with single-unit, multi-unit, and local field potentials. *Neural computation*, 18:2256–81.

Zerbi, V., Floriou-Servou, A., Markicevic, M., Vermeiren, Y., Sturman, O., Privitera, M., von Ziegler, L., Ferrari, K. D., Weber, B., De Deyn, P. P., Wenderoth, N., and Bohacek, J. (2019). Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation. *Neuron*, 103(4):702–718.e5.

STAR Methods

Contact for reagent and resource sharing

Further information and requests for reagents and resources may be directed to and will be fulfilled by the Lead Contact, Dr. Michel Besserve (michel.besserve@tuebingen.mpg.de).

Experimental model and subject details

The neural data used in this study were recorded from the ventrolateral prefrontal cortex (vlPFC) of one anaesthetised adult, male rhesus monkey (macaca mulatta) by using Utah microelectrode arrays [Blackrock Microsystems (Maynard et al., 1997)] (more details on these experiments are provided in a previous study exploiting this data by Safavi et al. (2018)). All experiments were approved by the local authorities (Regierungspräsidium) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

Method details

Detailed GPLA methodology for electrophysiology data

GPLA proceeds in several steps: preprocessing of multi-channel LFP signals, construction of the coupling matrix, and its low-rank approximation. Finally, parameters of this low-rank approximation are standardized following specific normalization conventions allowing their easy interpretation and comparison. These steps are described in the following subsections.

LFP pre-processing

Prior to computing couplings, the LFP signal is pre-processed, first by filtering in the frequency band of interest. The choice of the filter bandwidth for the purpose of extracting the instantaneous phase or analytic signal in a particular band is subjected to a trade-off. On one hand, the signal requires a narrow enough band-pass filtering to provide us a proper estimate of the phases (Chavez et al., 2006). On the other hand, the filtered signal should preserve the temporal dynamics in the frequency of interest. The second step is extracting the analytical signal using the Hilbert transform, resulting in a complex-valued signal containing both the amplitude and phase of LFP. In the optional third step (see section Necessity of whitening and post-processing), we whiten the LFPs. We need to decorrelate LFP signal recorded in the different channels by applying a whitening operator. It is necessary to be able to use tools from Random Matrix Theory (Anderson et al., 2010) (the rationale for the inclusion of the whitening step is elaborated in section Analytical test and Safavi et al. (2020)).

We consider LFPs and spiking units are recorded repeatedly over K trials, and each trial has length T (number of time-points). We represent LFPs of trial k by $L^{(k)}$, which is a $(n_c \times T)$ matrix, where n_c is the number of LFP recording channels. To simplify the notations, by $L^{(k)}$ we refer to analytical signals, i. e. band-passed in a particular frequency range and Hilbert transformed signals. We denote the collection of $N_m^{(k)}$ spike times of unit m at trial k by $\{t_j^{m,(k)}\}_{j=1\dots N_m^{(k)}}$ ($\{t_j^{m,(k)}\}$ contains the time-point indices of the LFP data for which spikes occur).

We introduce a *reduced-ranked* whitening operator which is a modified version of the conventional whitening that decorrelate the data, in this case, LFP signals. We customized this procedure in order to accommodate GPLA's needs, i. e. (1) avoid over-amplification of noise components of LFP (which are reflected in smaller eigenvalues of LFP covariance matrix) in the whitening operator, and (2) eliminate factors of variability that are not consistent across trials.

In our *reduced-ranked* whitening, we first reduce the rank of the LFP covariance matrix, by truncating the eigenvalue decomposition of LFP covariance matrix. We choose the number of components such that 99% of variance is explained with the reduced rank covariance matrix. In order to find the number of components that 99% of variance of LFP covariance matrix, we

concatenate LFPs of all trials into a larger $n_c \times KT$ matrix, denoted by L and compute the eigenvalue decomposition of the covariance matrix,

$$\mathrm{Cov}\,(L) = \frac{1}{T} L L^H\,, \tag{8}$$

where $.^H$ indicates the transpose complex conjugate (should be noted that, analytical signal L, is a complex-valued matrix). We denote the number of components needed to explain 99% of variance of LFP covariance matrix by n_c^{eff}. We find the reduced number of components, n_c^{eff}, based on all trials, and we use n_c^{eff} to define the whitening operator of individual trials. The reduced rank single-trial LFP covariance matrix is denoted by $\mathrm{Cov}^{\mathrm{red}}\left(\mathrm{L}^{(\mathrm{k})}\right)$, and computed as follows

$$\mathrm{Cov}^{red}\left(L^{(k)}\right) = \sum_{p=1}^{n_c^{eff}} \lambda_p^{(k)} x_p^{(k)} (x_p^{(k)})^H\,, \tag{9}$$

where $\lambda_k^{(k)}$ and $x_k^{(k)}$ respectively denote the eigenvalue and eigenvectors of the LFP covariance matrix of trial k. We denote the whitened LFP of trial k by $L_w^{(k)}$, and compute it as follows,

$$L_w^{(k)} = (\Lambda^{(k)})^{\frac{-1}{2}} (X^{(k)})^H L^{(k)}\,, \tag{10}$$

where $\Lambda^{(k)}$ is a $n_c^{eff} \times n_c^{eff}$ diagonal matrix containing the eigenvalues of the above single-trial reduced rank LFP covariance matrix, and $X^{(k)}$ is a is a $n_c \times n_c^{eff}$ matrix containing the eigenvectors $x_k^{(k)}$.

Coupling matrix

Given the spike times of a single spike train $\{t_j^{(k)}\}_{j=1\ldots N^{(k)}}$ and $L_w^{(k)}$ a single channel pre-processed LFP analytic signal (as explained in section LFP pre-processing) and its phase ϕ (= $\angle L$), the conventional measure of spike-LFP coupling, Phase Locking Value (PLV), defined as follows:

$$PLV = \frac{1}{N^{tot}} \sum_{k=1}^{K} \sum_{j=1}^{N^{(k)}} \exp\left(i\phi_{t_j^{(k)}}^{(k)}\right)\,, \tag{11}$$

where, i is the imaginary unit ($i^2 = -1$), and $N^{(k)}$ is the number of spikes occurring during the trial k, N_{tot} is the total number of spikes occurred across all trials, i. e.

$$N^{tot} = \sum_{k=1}^{K} N^{(k)}\,. \tag{12}$$

In addition to PLV, we introduce a similar coupling statistics, denoted by c,

$$c = \frac{1}{\sqrt{N^{tot}}} \sum_{k=1}^{K} \sum_{j=1}^{N^{(k)}} L_{t_j^{(k)}}^{(k)}\,, \tag{13}$$

to be used when the theoretical significance test is intended to be used (see section Analytical test). The coupling statistics c is different from PLV in two ways, First, in PLV only the phase information from the continuous signal is used, while for c, we use both the phase and amplitude of the LFP signal. This is motivated by evidence that inclusion of the amplitude can improve the coupling measure (Denker et al., 2007) by weighting the contribution of spikes in the coupling measure by the LFP amplitude at the correspond spike time, as well as by theoretical considerations (see STAR Methods section Analytical test for more details). The second difference is, for c we have normalization by square root of the number of spikes rather the number of spikes (division by $\sqrt{N^{tot}}$ in Equation 13 versus N^{tot} in Equation 11). Basically, a scaling by $\sqrt{N^{tot}}$ is needed to normalize the variance of entires of the coupling matrix to 1, in order to be able to use tools from Random Matrix Theory (Anderson et al., 2010) (see Safavi et al. (2020) for more details).

A multivariate generalization of the coupling statistics, could be achieved by collecting the coupling statistics between all spiking units and LFP signals. Given spike times $\{t_j^{m,(k)}\}_{j=1\ldots N_m^{(k)}}$, $\phi_w^{(k)}$ LFP

phase, and $L_w^{(k)}$ the analytical LFP, we can define the coupling matrix $\boldsymbol{C}$, based on PLV (Equation 11) as follows,

$$(\boldsymbol{C})_{n,m} = \frac{1}{N_m^{tot}} \sum_{k=1}^{K} \sum_{j=1}^{N_m^{(k)}} \exp\left(i(\phi^{(k)})_{n,t_j^{(k)}}\right), \tag{14}$$

or based on c (Equation 13),

$$(\boldsymbol{C})_{n,m} = \frac{1}{\sqrt{N_m^{tot}}} \sum_{k=1}^{K} \sum_{j=1}^{N_m^{(k)}} (L^{(k)})_{n,t_j^{(k)}}, \tag{15}$$

where m, j and n respectively indicate the index of spiking unit, index of spike time and index of LFP channel and N_m refers to number of spikes recorded in spiking unit m [1].

Let n_c and n_s be the number of LFP channels and number of spiking units, respectively, $\boldsymbol{C}$ is thus a $n_c \times n_s$ complex-valued matrix (or $n_c^{eff} \times n_s$ if whitening is applied). As n_c (or n_c^{eff}) and n_s are not necessarily equal in electrophysiological datasets, the coupling matrix is not square in general.

Our coupling matrix is thus designed as a multivariate generalization of univariate coupling measures in order to capture the overall synchronization between the spiking activity and the phase of a global oscillatory dynamics in a given frequency band.

Low rank decomposition

Each column of the coupling matrix $\boldsymbol{C}$ has a common spiking unit whose locking is computed with respect to different LFP channels (called LFP vectors). Conversely, each row collects the phase locking values of all spiking channels to a common LFP reference channel. In order to achieve a compact and interpretable representation of this high dimensional object, we compute the Singular Value Decomposition (SVD) of the coupling matrix of the form

$$\boldsymbol{C} \quad = U D V^H \quad = \sum_{k=1}^{p} d_k u_k v_k^H, \tag{16}$$

where (d_k) is a tuple of positive scalars, the singular values (SV), listed in decreasing order. The complex valued vectors u_k and v_k are, respectively, the n_c/n_c^{eff}- and n_s-dimensional singular vectors associated to a given SV d_k. One important property of SVD is that keeping only the first term in Equation (16), with SV d_1, achieves the best rank-one approximation of the matrix, $C \approx d_1 u_1 v_1^H$, in the least square sense (Datta, 2010, Theorem 7.29).

Post-processing

In order to make the outputs of GPLA interpretable, we introduce a few post-processing steps. An unwhitening and rescaling procedure is introduced to reverse some normalization discussed in previous sections LFP pre-processing, Coupling matrix, and Low rank decomposition, and a rotational transformation is introduced in order to represent the singular vectors in a more interpretable fashion.

Representation of singular vectors: Following the conventional mathematical representation of SVD in Equation 3, U and V are unitary matrices i. e. $U^H U = I$ and $V^H V = I$.[2] This implies that all singular vectors are unit norm, and all the information regarding the strength of coupling is absorbed in the singular values on the diagonal matrix D. As explained in main text (see sections Low rank linear response theory and frequency analysis and Generalizing spike-oscillation coupling analysis to the multivariate setting), the relative magnitude and phase of singular vectors coefficients can be used to interpret the *relative* contribution of individual LFP channel and individual spiking unit to the coordinated pattern captured by the largest singular value.

We can summarize the coupling matrix with three quantities:

$$\boldsymbol{C} \sim (g\boldsymbol{PLV}).v_{LFP} v_{spike})^H. \tag{17}$$

[1] Reader can also refer to Safavi et al. (2020, Section 4) for a different formulation of computing the coupling matrix.

[2] For spike vector, V, should be noted that, V is a unitary after appropriate normalization that was discussed in the previous paragraph

However the coefficient of both singular vectors can be rotated of the same arbitrary angle in the complex plane, as the rotation transformation in the complex plane does not change the SVD factorization, i. e.

$$udv^H = udv^H e^{-i\theta_0} e^{i\theta_0} = e^{-i\theta_0} ud \left(e^{-i\theta_0} v\right)^H . \tag{18}$$

We exploit this free parameter to make the GPLA more neuroscientifically interpretable by rotating both spike and LFP vectors with $\overline{-\phi_{LFP}}$, where $\overline{\phi_{LFP}}$ and $\overline{\phi_{spike}}$ are the average spike and LFP phases, defined as,

$$\overline{\phi_{LFP}} = \angle \sum_{i=1}^{n_c} (v_{LFP})_i , \tag{19}$$

$$\overline{\phi_{spike}} = \angle \sum_{i=1}^{n_u} (v_{spike})_i . \tag{20}$$

The rationale behind it is to center the coefficient of the rotated LFP vector ($\widetilde{v_{LFP}} = v_{LFP} e^{-i\overline{\phi_{LFP}}}$) around zero phase in the complex plane and the rotated spike vector, $\widetilde{v_{spike}} = v_{spike} e^{-i\overline{\phi_{LFP}}}$, preserves the angular difference of Φ_d of the spikes with respect to the LFP, defined as

$$\Phi_d = \overline{\phi_{LFP}} - \overline{\phi_{spike}} . \tag{21}$$

With this chosen convention, we obtain the final GPLA factorization

$$C \sim (gPLV).\widetilde{v_{LFP}}\widetilde{v_{spike}}^H . \tag{22}$$

We can also apply the phase difference between average LFP and spike vectors (Φ_d) to gPLV as it can summarize the overall phase shift between LFP and spikes. Given that gPLV is always a real positive value, by this convention, we add an extra information to gPLV.

We thus define a *complex gPLV* ($\widetilde{gPLV} = gPLV e^{-i\Phi_d}$) which its magnitude indicate the coupling strength between spikes and LFPs as in phase locking value (PLV) and its angle indicates the overall phase difference between spiking activity and LFP which is similar to locking phase in classical univariate phase locking analysis. This is an arbitrary choice to some degree, neverthelss it allows to interpret the GPLA output very similar to classical univariate phase locking analysis.

Needless to mention, when the magnitude of gPLV is small, this overall phase difference is not meaningful (similar to the case where PLV is small, the locking phase is not meaningful).

Unwhitening: As discussed in section LFP pre-processing, due to theoretical considerations, and in particular for applicability of our analytical significance test (see Significance assessment of gPLV), we whiten the LFPs prior to any other processing. In order to retrieve the original structure of the LFP i. e. retrieve all the correlations that were present in the original LFP signals but was diminished by the whitening, we need to revert the whitening i. e. unwhitening the LFP vector resulting from GPLA. This can be achieved by computing the unwhiten operator and apply it to the LFP vector,

$$v_{LFP}^{unwhiten} = W^{-1} v_{LFP} , \tag{23}$$

where W^{-1} is the unwhitening operator. In order to find this operator, we first we concatenate whitened LFPs of all trials (resulting from Equation 10) into a larger matrix L_w ($n_c^{eff} \times KT$). Then we estimate W^{-1} by using a linear regression with unwhitned and whitened LFPs (W^{-1} is the $n_c \times n_c^{eff}$ matrix of coefficient for regression).

Rescaling: As introduced in Equation 13, coefficient of coupling matrix are normalized by the square root of number of spikes. This choice of normalization, is different from the one use in conventional PLV (Equation 11). This will lead to inhomogeneous weighting of spiking units according to their variability of their firing rate. To avoid this potentially misleading weighting, we divide the spike vector by the square root of number of spikes,

$$v_{Spike}^{rescaled} = v_{Spike} \oslash \vec{N} , \tag{24}$$

where $\oslash$ is (entrywise) Hadamard division and $\vec{N} = \{N_m^{tot}\}_{m=1,\dots,n_s}$, which is a vector consist of total spike counts (similar to Equation 12) of all the neurons (indexted by m) used in GPLA. Furthermore,

to preserve the original norm of the spike vector (unit magnitude), we also need to normalize the spike vector by its norm,

$$v_{Spike}^{final} = \frac{v_{Spike} \oslash \vec{N}}{\left\| v_{Spike} \oslash \vec{N} \right\|}. \tag{25}$$

Necessity of whitening and post-processing

The whitening (and the subsequent post-processing) is necessary to have the advantage of applicability of the analytical significance test. LFPs are typically very correlated signals, and such correlation will be reflected in the magnitude of the singular values (and consequently gPLV), therefore, by whitening such correlations will be removed prior to investigating spike-LFP coupling. Nevertheless, multivariate analysis of spike-LFP coupling with GPLA even without whitening can still be beneficial for investigating spike-LFP relation sip (see Figure 3). If statistical testing based on surrogate data is intended, it is possible to skip the whitening step and proceed directly with the constructing the coupling matrix and low rank estimation. In that case, entries of the coupling can be filled by conventional PLVs (see Equation 14), or other choices of spike-LFP coupling measures (Grasse and Moxon, 2010; Vinck et al., 2010; Lepage et al., 2011; Vinck et al., 2012; Jiang et al., 2015; Zarei et al., 2018; Li et al., 2016) (also see the section Limitations and potential extensions for further elaboration). In this case, whitening of the LFP can be skipped and subsequent "Unwhitening and rescaling" discussed in section Post-processing is not necessary anymore.

Optional normalization for gPLV

As gPLV is a singular value of a matrix, it grows with the dimensions of the coupling matrix. This makes the comparison of gPLV resulting from different datasets difficult. For instance, assume the hypothetical situation of having two datasets recorded from two homogeneous populations of neurons, if the strength of coupling is the same in two populations, the populations with a larger amount of recorded neurons (therefore larger dimension of the coupling matrix) will have larger gPLV. Certainly, this can be misleading for investigating the spike-LFP coupling with GPLA when datasets with variable number of spiking units and/or LFP channels. To overcome this issue we suggest normalizing the gPLV to become independent of the size of the neural population (dimension of the coupling matrix) and the number of channels. When we consider the the entries of coupling matrix, C, to be PLV (LFPs are not whitened and Equation 14 is used for constructing the coupling matrix), pairwise coupling static is bounded ($|PLV| \leq 1$). When all the entities of the coupling matrix C attain their maximum value, gPLV will also gain the maximum possible value. Therefore, we can exploit it to normalize the gPLV. For a coupling matrix having maximum coupling for all pairs ($(C)_{n,m} = 1$ and C, a $n_c \times n_s$ matrix), then $gPLV_{max} = \sqrt{n_c n_s}$. Therefore, if we normalize the original gPLV by the maximum value it can achieve ($gPLV_{max} = \sqrt{n_c n_s}$, calculated is based on the dimensionality of matrix C), then the gPLV will be bounded by 1 as well. Moreover, with this normalization, gPLV is also comparable to PLV (if we have a homogeneous population of neurons, otherwise these quantities are not comparable).

Significance assessment of gPLV

In order to statistically assess the significance of coupling between spikes and LFP based on gPLV, we develop a surrogate- and a Random Matrix Theory (RMT)-based statistical testing framework exposed in Safavi et al. (2020). Hypothesis testing based on the generation of surrogate data is a common method for significant assessment in neuroscience. Nevertheless, not only generating appropriate surrogate data can be challenging (for a review see Grün (2009)), but also computationally expensive. This motivates the development of an analytical test exploiting minimal computational resources.

Surrogate-based test

In contrast to uni-variate methods for which the distribution under a null hypothesis is more likely to be (possibly approximately) derived based on theoretical analysis (e.g. Rayleigh test for PLV (Fisher, 1995, Chapter 4)), such approaches are usually unavailable in multi-variate settings (nevertheless, we have developed one for gPLV, see section Analytical test). Following a common alternative approach, we build the null distribution by generating many surrogate datasets Grün (2009). The resulting gPLVs values forms an empirical H_0 distribution that can be used to compute the p-value for statistical assessment of the significance gPLV in the data. Importantly, the choice of the appropriate surrogate data according to characteristics of neural data is critical. For instance, generating surrogate data by shuffling inter-spike-intervals (ISI) is not an appropriate method when we have non-stationarity in firing rates (Grün, 2009).

In this work, we used an *interval*-jittering rather than a *spike-centered*-jittering[3], as the former was reported to be more reliable for detecting temporal structures in spike data (Platkiewicz et al., 2017). We devised the two following spike-jittering-based methods for GPLA. We also verify the appropriateness of our jittering approaches with various simulations (see the Results).

Simple interval jitter: Each surrogate dataset is generated by jittering all the spikes (from all neurons) with a particular jittering window (or dither width). In the interval jittering, per each spike, a new spike time is drawn within the jittering window around the spike. The timing of jittered spikes should be drawn from a uniform distribution. The size of the jittering window can be specified by the frequency wherein the spike-LFP coupling is being investigated. The smallest jittering window (or dither width) that can be used in order to destroy the temporal structure potentially exists in the range of frequency-of-interest. In the phase-locking analysis of electrophysiological data we usually extract the analytic signal or instantaneous phase of LFP by applying Hilbert transform on band-limited LFP signals (Chavez et al., 2006). The central frequency of the band-limited filter can be used for specifying the jittering window (or dither width), i. e. jittering window is the inverse of this central frequency.

Group preserved jitter: Similar to "simple interval jitter" we generate each surrogate dataset by relocating all the spikes within a window. For each surrogate data, we first divide the spike trains into equally-sized windows. Then we circularly shift the spike sequence within each window for all neurons together using a uniformly distributed time shift. Notably, we use a single random value for circular shifting of all neuron's spiking within the window. This size of this window should be chosen similar to the previous method ("simple interval jitter") i. e. based on the central frequency of the band-limited filter. The rationale behind this method of generation surrogate data is *relative* timing of the spikes could be associated to a large degree to the ansamble activity irrespective of the coupling to the LFP. Therefore, the relative timing of the spikes might not be impaired in the absence of coupling to global dynamics of the LFP. With "group preserved jittering" the relative timing is preserved and the coupling to the LFP is destroyed.

Analytical test

Challenges in generation of surrogate data (Grün, 2009) and considerable increase in the dimensionality of datasets (Pesaran et al., 2018; Jun et al., 2017; Buzsáki, 2004; Fukushima et al., 2015), suggest that deriving mathematically (asymptotic) properties of GPLA under the null hypotheses, as is done for univariate testing (e. g. Rayleigh test for PLV (Fisher, 1995, Chapter 4)) is an interesting alternative.

In a companion work (Safavi et al., 2020), by using martingale theory (Aalen et al., 2008) we derive an asymptotic distribution for the entries of the coupling matrix in fairly general settings. Furthermore, by exploiting RMT (Anderson et al., 2010) we can find a good approximation of the distribution of eigenvalues (or singular values) of the coupling matrix in absence of coupling between spikes and LFPs. This provides a null hypothesis for the statistical testing of the largest eigenvalues (or singular values) of the coupling matrix, which corresponds to gPLV in our setting.

[3] Interval- and spike-centered-jittering are also known as hard and soft dithering respectively.

Table 1

Figure num.	Osc. type	Num. of oscillatory component	Equations
Figure 2	Transient	1	14,31
Figure 3 A-C	Transient	1	14,31
Figure 3 D-I	Sustained	1	31,32
Figure 4 A-G	Sustained	5	15,31
Figure 4 H	Sustained	1-10	15, 31

As mathematical details are described in Safavi et al. (2020), we restrict ourselves to a brief explanation. When the LFP signal is whitened, and under a null hypothesis reflecting an *absence of coupling*, the coupling matrix which is constructed based on Eq 13, asymptotically converges to a matrix with i.i.d. complex standard normal coefficients (Safavi et al., 2020, Theorem 2), and the Marchenko-Pastur (MP) law then provides an approximation of the distribution of its squared singular values (Safavi et al., 2020, Theorem 3).

This law (Marchenko and Pastur, 1967) has the density

$$\frac{d\mu_{MP}}{dx}(x) = \begin{cases} \frac{1}{2\pi\alpha x}\sqrt{(b-x)(x-a)}, & a \le x \le b, \\ 0, & \text{otherwise}, \end{cases} \tag{26}$$

with $a = (1-\sqrt{\alpha})^2$ and $b = (1+\sqrt{\alpha})^2$ which are the upper and low bounds of the support of the distribution. Based on on the these bounds we can define a significance threshold, θ_{DET}, for the largest eigenvalue of hermitian matrix, $\boldsymbol{S} = \frac{K}{n_u}\boldsymbol{C}\boldsymbol{C}^H$:

$$\theta_{DET} = (1+\sqrt{\alpha})^2 . \tag{27}$$

The null hypothesis can be rejected if, the largest eigenvalue of $\boldsymbol{S}$ (denoted by ℓ_1) is superior to the significance threshold:

$$\ell_1\left(\boldsymbol{S}_n\right) > \theta_{DET} . \tag{28}$$

Therefore, there is a significant coupling between the multi-channel spikes and LFPs, if

$$gPLV > \sqrt{n_u \theta_{DET}} . \tag{29}$$

As mentioned above to be able to use the result of Safavi et al. (2020, Theorem 3), we need to whiten the LFP signal first, as described in LFP pre-processing. Furthermore, satisfying this theorem requires the coupling matrix to be normalized appropriately based on the spike rate of each unit (as defined in Equation 15).

For computing α on neural data, the *reduced ranked* $n_c^{eff} < n_c$ entailed by the whitening procedure (see LFP pre-processing more details), the *effective* dimensionality of the coupling matrix changes from $n_c \times n_u$ to $n_c^{eff} \times n_u$ (which depends on the spectral content of the LFP). This leads to a modification of Equation 27 as follows:

$$\theta_{DET} = (1+\sqrt{\alpha_{eff}})^2 , \tag{30}$$

where $\alpha_{eff} = n_c^{eff}/n_u$.

Simulation of phase-locked spike trains

We use simulated phase-locked spike trains and noisy oscillations as a toy model to demonstrate the potential applications of GPLA. The core principles of simulations used in both Figure 2, 4 and 3 are explained in the following paragraphs and the specializations used for individual figures are provided at the end (also summerized in table 1).

For generating phase-locked spike trains, we adopt the method introduced in (Ashida et al., 2010). As the model has already been described elsewhere we restrict ourselves to a brief explanation. We sample the spike times from an inhomogeneous Poisson process with rate $\lambda(t)$,

$$\lambda(t) = \lambda_0 \exp\left(\kappa \cos(2\pi f t - \varphi_0)\right) , \tag{31}$$

where φ_0 is the locking phase of the spikes with respect to the oscillation, κ is the concentration parameter of the spikes around the locking phase, that specify the strength of coupling between spikes and the oscillation, λ_0 is propositional to the average firing rate over time ($\lambda_0 I_0(\kappa)$ is the average firing rate), and f is the frequency of oscillatory modulation of the spike trains.

Furthermore, we can also derive an analytical expression for the complex-valued PLV to be used as ground truth PLV (used in Figure 3),

$$PLV^* = e^{\mathbf{i}\varphi_0} \frac{\int_0^\pi \cos(\theta)\exp(\kappa\cos(\theta))d\theta}{\int_0^\pi \exp(\kappa\cos(\theta))d\theta} = e^{\mathbf{i}\varphi_0}\frac{I_1(\kappa)}{I_0(\kappa)} , \tag{32}$$

where PLV^* indicate the ground truth value, and the I_k's denoting the modified Bessel functions of the first kind for k integer (see e. g. Abramowitz et al. (1972, p. 376)):

$$I_k(\kappa) = \frac{1}{\pi}\int_0^\pi \cos(k\theta)\exp(\kappa\cos(\theta))d\theta . \tag{33}$$

For the simulation used in Figure 4, we construct the LFP by superimposing $N_{osc} \leq 10$ oscillatory components $O_j(t) = e^{2\pi i f_j t}$, $j \in \{1, \dots, N_{osc}\}$ that the frequency of oscillations are limited in range of $[f_{min}, f_{max}]$. Each LFP signal is a weighted sum of these oscillatory components. We can represent these weights in a $(n_c \times N_{osc})$-variate matrix (we call it *mixing matrix*, and denot it by W), where each row of mixing matrix indicate the weights for the corresponding LFP channel. Thus, the synthetized multichannel LFP ($\mathbf{\Psi}(t) = \{\psi_l(t)\}_{l=1,\dots,n_c}$) can be written as the product of the mixing matrix (W) and the oscillatory basis ($\boldsymbol{O}(t) = \{O_j(t)\}_{j=1,\dots,N_{osc}}$),

$$\mathbf{\Psi}(t) = W\boldsymbol{O}(t) + \boldsymbol{\eta}(t) , \tag{34}$$

where $\boldsymbol{\eta}(t)$ is additive white noise on both real and imaginary parts.

In this simulation, the frequency of the oscillatory components range from 11Hz to 15 Hz, and the mixing is the following,

$$W = \begin{bmatrix} \boldsymbol{w}_d & \boldsymbol{w}_0 & \cdots & \cdots & \boldsymbol{w}_0 \\ \boldsymbol{w}_0 & \boldsymbol{w}_d & \boldsymbol{w}_0 & \cdots & \boldsymbol{w}_0 \\ \boldsymbol{w}_0 & \boldsymbol{w}_0 & \ddots & \ddots & \boldsymbol{w}_0 \\ \boldsymbol{w}_0 & \ddots & \ddots & \ddots & \boldsymbol{w}_0 \\ \boldsymbol{w}_0 & \cdots & \cdots & \boldsymbol{w}_0 & \boldsymbol{w}_d \end{bmatrix} . \tag{35}$$

where $\boldsymbol{w}_d = w_d \mathbf{1}_{N_g}$ and $\boldsymbol{w}_0 = w_d \mathbf{0}_{N_g}$, $w_d = 1$ and $w_0 = 0.1$ ($\mathbf{1}_{N_g}$, is a $(N_g \times 1)$-variate all-one column vector). This simple structure of the mixing matrix (which is close to a block diagonal matrix) implies that each LFP channel contains one dominant oscillatory component with a specific frequency (as in each row, there is only one oscillatory component with large coefficient w_d and a specific frequency).

For the simulation used in Figure 3A-C, oscillations originate from a single oscillatory source, but in order to make transient rather sustained oscillations, they were multiplied by a Gaussian window (with the size of 20 cycles of oscillation) around random events. The timing of transitory events was governed by a homogeneous Poisson process. Moreover, the spiking activities are phase-locked to the phase of the oscillations as the spike times were drawn from an inhomogeneous Poisson process with the rate specified in Equation 31. For the rest of the simulations in Figure 3 a single sustained oscillation has been used.

Simulation of hippocampal sharp wave-ripples

The model was introduced and described in Ramirez-Villegas et al. (2018). We thus restrict ourselves to a brief explanation of the characteristics that are the most relevant to GPLA analysis.

Network architecture

We use a model of part of the hippocampal formation, accounting for the dynamics of CA1 and CA3 subfields during non-Rapid Eye Movement (non-REM) sleep. Cells of each subfield consist of 150 units (135 pyramidal neurons and 15 interneurons), arranged on a one dimensional array, along the x-axis. The connectivity of CA3 is characterized by strong recurrent excitatory auto-associational pyramidal-pyramidal connections, together with pyramidal-interneuron connections and short-range and interneuron-pyramidal synapses. In contrast, CA1 connectivity is implemented as a "feedback and reciprocal inhibition" circuit, including only pyramidal-interneuron, interneuron-pyramidal, and interneuron-interneuron synapses, all located in their peri-somatic region (see Figure 6A for the schema of the model)

Cell dynamics

Each neuron is modeled with two compartments: dendritic and axosomatic, the dynamics of each follows a Hodgkin-Huxley type (conductance-based) equation (Pinsky and Rinzel, 1994; Traub and Miles, 1995). Notably, they include a non-linear slow dendritic calcium channel responsible for the bursting activity.

Computation of the laminar LFP profile

The procedure for computing laminar LFP profiles was also described in Ramirez-Villegas et al. (2018). Briefly, the trans-membrane current of each compartment of each cell is modeled as a line source (Schomburg et al., 2012) that were placed with the equal distance across the horizontally in a Stratum Pyramidale (SP) of 100μm thickness, with an axosomatic compartments height of 80μm for both pyramidal neurons and interneurons (Traub and Miles, 1995). Total dendritic arbor height of pyramidal cells was 200μm, corresponding to the CA1 Stratum Radiatum (SR). LFPs were captured through two multi-channel electrodes (mimicking laminar probes), each with 16 recording sites disposed along the vertical axis (denoted by z), 20 μm apart covering the simulated axosomatic and apical dendritic fields of CA1 and CA3. Each electrode crosses the corresponding linear cell arrangement (perpendicularly) in its middle.

The extracellular medium is modeled as a uniform and isotropic ohmic conductor with resistivity $\rho = 333\Omega\text{cm}$. The potential in the extracellular medium is governed by the Poisson equation $\nabla^2\phi = \frac{1}{\sigma}\frac{d\xi}{dt} = -\frac{I_t}{\sigma}$, where $\sigma = \frac{1}{\rho}$ is the conductivity of the extracellular space $\left[\frac{\text{S}}{\text{m}}\right]$. With these assumptions, the extracellular potential $\phi\left(y_0, r, t\right)$ at the algebraic depth z_0 and a radial distance r, measured over the compartment's length limits (z_1 and z_2, respectively indicate the algebraic depth and the bottom of the top of the cylindrical compartment with length $L = z_2 - z_1$) can be computed by

$$\phi\left(z_0, z_1, z_2, r, t\right) = \frac{1}{4\pi\sigma L}\int_{z_1}^{z_2}\frac{I\left(t\right)}{\sqrt{\left(z - z_0\right)^2 + r^2}}\text{dz} = \frac{1}{4\pi\sigma}\frac{I\left(t\right)}{L}\ln\left[\frac{\sqrt{\left(z_1 - z_0\right)^2 + r^2} - \left(z_1 - z_0\right)}{\sqrt{\left(z_2 - z_0\right)^2 + r^2} - \left(z_2 - z_0\right)}\right]. \quad (36)$$

after solving the integral with standard procedures. Accounting for the contribution of all compartments and cells, the total extracellular potential $\phi_{\text{tot}}\left(z_0, t\right)$ at a given depth z_0 is

$$\phi_{\text{tot}}\left(z_0, t\right) = \sum_i\sum_j \phi_{i,j}\left(z_0, z_{1,i,j}, z_{2,i,j}, r_{i,j}, t\right), \quad (37)$$

where $\phi_{i,j}$ is the potential generated by the total transmembrane current of the j^{th} compartment of the i^{th} cell, located at radial distance $r_{i,j}$ from the electrode.

Note that since the neuron models considered in this work are two-compartmental, charge conservation within the cell implies that the total absolute somatic transmembrane currents equal the absolute of the total dendritic transmembrane currents (which also follows the charge conservation principle), leading to a dipolar distribution of the LFP contribution for each cell.

Equations 36-37 describe the way LFPs are simulated (as the low-frequency parts of the extracellular potential) in the original biologically realistic model that generated the LFP data we use for

GPLA analysis. We also exploit it to provide an approximative LFP laminar profile for the population of pyramidal cells based on this equation, by injecting the same constant current to all cells and compartment of the linear arrangement, but having opposite signs for the axosomatic and dendritic compartment to respect charge neutrality and thus the dipolar structure (Figure 6D (broken line)).

Neuron exclusion criterion for GPLA

To reduce the small sample bias caused by a low number of spike events, we only use neurons that had a minimum average firing rate of 3 Hz firing. Nevertheless, using all neurons did not change the results significantly. For instance, in Figure 6G (in contrast to Figure 6A-F and H where excluded neurons based on their firing), to be compatible with the neural mass model we did not exclude any neuron.

Analytical neural field modeling of spike-field coupling

In order to justify and interpret our approach, we use a rate-based neural field model. Units are grouped in populations according to their cell-type on spatial localization. Spiking activity of a specific population p at possibly multidimensional location x is represented by its average spike rate $\lambda_p(x,t)$. Simultaneously, the LFP $L(X,t)$ is recorded at locations reflected by possibly different coordinates X.

Rate model of circuit dynamics

We follow classical neural field models, stating that the rate of each population evolves as a monotonous function of the membrane potential, itself controlled by post-synaptic currents (PSCs). Dynamics of the membrane potential V_p of each population p, is assumed to be governed by the following differential equation:

$$\frac{dV_p}{dt}(x,t) + \tau_p V_p(x,t) = \alpha_p \eta(x,t) + \sum_k \nu_{p\leftarrow k} s_k(x,t)\,, \tag{38}$$

where η represents the post-synaptic current generated by the external input (for which no spiking activity is available), s_k the normalized[4] post-synaptic current from the afferent population k, whose effect on target population p is scaled by synaptic strength $\nu_{p\leftarrow k}$.

The relationship between the normalized post-synaptic current at location x and spiking activity of the afferent population activity is modeled by spatio-temporal integration (Wilson and Cowan, 1973; Somers et al., 1995; Jirsa and Haken, 1996), possibly taking into account the propagation speed v_0 along the axons

$$s_k(x) = \int c_k(x,X)\lambda_k(x, t - |x - X|/v_0)dX, \tag{39}$$

where the connectivity kernel $c_p(x,X)$ models the density of synapses of neurons whose soma is located at target location x, with afferent neurons having their somas at location X. The integral covers the spatial domain where units' somas can be found, and may thus be 1-, 2- or 3-dimensional depending on the model. The kernel c_k reflects the spatial spread of axonal arborizations and as such can be approximated based on anatomical studies.

The elements finalizing the description of the state of the system are the relations between each population's membrane potential and rate, modeled by

$$\lambda_p = a_p(V_p)\,, \tag{40}$$

where a_p is a typically sigmoidal activation function (these are the only non-linearities considered in our neural mass equations), leading to the overall dynamical system represented in Figure 5A.

[4] By "normalized", we mean that s_k is a numerical quantity independent of the target population, possible target-specific differences in the PSCs being taken into account in the connectivity parameters $\nu_{p\leftarrow k}$, without loss of generality.

From synaptic currents to LFPs

The local field potential is the lower frequency (<150Hz) part of the electrical potential recorded in the extracellular space, generated by the transmembrane currents (Einevoll et al., 2013). Considering that active currents mostly reflect spiking activity, whose dynamics lies mostly above the typical LFP frequency range, we approximate the LFP as resulting from the linear superposition of passive membrane currents triggered by post-synaptic input currents (Mazzoni et al., 2015), leading to the equation

$$L(y,t) = \sum_p \int f_{p,e}(y,x)\eta(x,t)dx + \sum_{p,k} \int f_{p,k}(y,x)s_k(x,t)dx\,,$$

where $f_{p,k}(.,x)$ represents the electrical field spatial distribution generated by trans-membrane currents of the p cells with soma located at x, resulting from exciting them with post-synaptic unit currents of population k. Note due to charge neutrality of the cells, trans-membrane currents across the membrane of individual cells sum to zero, such that input currents at the levels of synapses are compensated by opposite trans-membrane currents away from them, typically leading to dipolar current distributions. Along the same lines, $f_{p,e}$ is the distribution associated with post-synaptic current resulting from exogenous inputs to p cells. Differences between these fields according to the afferent populations are due to the respective distribution of their synaptic button over the efferent cell, preferentially targeting either the peri-somatic or distant dendritic sites, as illustrated in Figure 5C. These field distributions are assumed dominated by currents originating from pyramidal cells, due to their individual and collective geometric arrangement (Nó, 1947; Lindén et al., 2011; Mazzoni et al., 2015), such that we can simplify the above equation to obtain

$$L(y,t) = \int f_{E,e}(y,x)\eta(x,t)dx + \sum_k \int f_{E,k}(y,x)s_k(x,t)dx\,. \tag{41}$$

Spike-LFP relation

Analysis of the frequency response of neural network models is a useful approach to understand their characteristics and underlying mechanisms (Ledoux and Brunel, 2011; Sherfey et al., 2018). In our case, this can be performed analytically by linearizing Equation 40 around an operating point, the neural field model becomes a linear time-invariant system controlled by the exogenous input $\eta(t)$. We can thus compute transfer functions for each variable of the system that will determine their response to a sinusoidal exogenous input at frequency f, based on the computation of temporal Fourier transforms of the signals. For a given signal, $s(t)$ the temporal Fourier transform at frequency f is given by

$$S(f) = \mathcal{F}_t[s](f) = \int_{\mathbb{R}} s(t)e^{-i2\pi ft}dt\,.$$

By applying the Fourier transforms on the left- and right-hand-side of dynamical equations, we can derive transfer functions linking the responses of each network variable to the exogenous input at a given frequency. We provide in Table 2 a list of time-domain variables and their corresponding notations for their time-domain Fourier transform, used in analytical developments.

Specifically, for a general exogenous input signal with time-domain Fourier transform $\mathcal{E}(X,f)$, where f denotes the temporal frequency, we obtain the input-output relation for population rates and LFP activity L

$$\lambda_p(x,f) = \int H_{\lambda_p}(x,X,f)\mathcal{E}(X,f)dX \quad \text{and} \quad L(y,f) = \int H_L(y,X,f)\mathcal{E}(X,f)dX\,.$$

A major simplification of this expression occurs when the exogenous input is separable in time and space,

$$\eta(X,t) = n(X)e(t))\,, \tag{42}$$

leading to $\mathcal{E}(X,f) = n(X)E(f)$ after temporal Fourier transform. This simplifying assumption models a number of typical inputs to the structure, including sinusoidal standing waves ($\eta(x,t) = n(x)e^{i2\pi ft}$) and traveling plane waves ($\eta(x,t) = e^{i(2\pi ft - kx)}$). This results in a simple expression for the covariance

Table 2. List of neural field and mass model variables (k indicates neuron population (afferent in case of synapse property))

Description	Symbol	Temporal Fourier transform	Equation
Exogenous input (spatio-temporal)	$\eta(x,t)$	$\mathcal{E}(x,f)$	38
Exogenous input (temporal)	$e(t)$	$\mathcal{E}(f)$	42
Spike rate	$\lambda_k(x,t)$	$\Lambda_k(x,f)$	43
Membrane potential	$v_k(x,t)$	$V_k(x,f)$	43
Synaptic activity	$s_k(x,t)$	$S_k(x,f)$	38, 39,41
Activation function	$a_k(x,t)$	$A_k(x,f)$	40

Table 3. List of neural mass model parameters

Parameter name	Symbol	Value (*Mass2D*)	Value (*MassAlpha*)
E membrane time constant	τ	$20ms$	$25ms$
I membrane time constant	δ	$5ms$	$25ms$
$E \leftarrow I$ synaptic strength	$\nu_{E\leftarrow I}$	0.01	0.85
$I \leftarrow E$ synaptic strength	$\nu_{I\leftarrow E}$	0.5	0.3
Alpha synapse time constant	σ	n.a.	$15ms$
$I \leftarrow I$ synaptic strength	$\nu_{I\leftarrow I}$	n.a. (accounted for in δ)	12.0

estimated across experimental trials between $rate$ and LFP at two possibly different spatial points (x, y)

$$\left\langle L(y,f)\lambda_p(x,f)\right\rangle = \left(\int H_L(y,X,f)n(X)dX\right)\left(\int H_{\lambda_p}(x,X,f)n(X)dX\right)|e(f)|^2 .$$

in which the input intervenes only as a multiplicative positive constant, and which is separable in both space variables x and y. As a consequence, the rank one approximation of the covariance between spiking units and LFP channels activity estimated by GPLA is informative about the microcircuit properties, as we explained in the STAR methods sections describing the analysis of the neural mass and neural field models.

Analysis and simulation of two population neural mass models

General description

The generic dynamic model of Equation 38 is exploited to describe network activity at a single location (i.e. we neglect the spatial extent of the considered structure) containing two cell types: pyramidal (E) and inhibitory (I), leading to the linear equations:

$$V_E + \tau_E \frac{dV_E}{dt} = \nu_{E\leftarrow E}s_E - \nu_{E\leftarrow I}s_I + \eta \quad (43)$$

$$V_I + \tau_I \frac{dV_I}{dt} = \nu_{I\leftarrow E}s_E - \nu_{I\leftarrow I}s_I + \alpha\eta \quad (44)$$

where $\boldsymbol{\nu}$ is a matrix gathering the non-negative synaptic strengths between populations, η the exogenous input to the network, with $\alpha \geq 0$ controlling the ratio between feed-forward excitation and inhibition. The term $\nu_{kj}s_j$ is the population averaged post-synaptic potential from population j to population k. In order to study quantitatively the effect of connectivity changes in the microcircuit, in this expression of the post-synaptic current, we isolate the synaptic strength coefficient $\nu_{k\leftarrow j}$, from a perisynaptic activity, that summarizes the dynamical processes occurring pre- and post-synaptically (synaptic delay, time constant induced by the post-synaptic channel conductance, ...). In the simplest case we assume peri-synaptic activity s_j can be approximated by the spike rate of population j, λ_j (up to a multiplicative constant that is absorbed by ν_{kj}). Alternatively, we model synaptic dynamics by a linear differential equation controlled by this rate (see model *MassAlpha* below).

The neural mass models will be analyzed with linear response theory, such that the a_k's of Equation 40 will be linearized around an equilibrium point of the dynamical system (that can be computed for vanishing input $\eta = 0$), and the resulting multiplicative constants will be themselves absorbed in the connectivity matrix $\boldsymbol{\nu}$, leading to replacing Equation 40 by

$$\lambda_k = V_k \,. \tag{45}$$

Next we describe the two linearized neural mass models exploited to interpret GPLA results of hippocampal simulations (see Figure 6). Parameter values for both models are reported in Table 3

Mass2D: E-I interactions without synaptic dynamics

Starting from Equation 43 and using Equation 45, the linearized system can be trivially reduced to the two-dimensional dynamical system (up to rescaling of the connectivity matrix coefficients)

$$\lambda_E + \tau \frac{d\lambda_E}{dt} = -\nu_{E\leftarrow I}\lambda_I + \eta \,, \tag{46}$$

$$\lambda_I + \delta \frac{d\lambda_I}{dt} = \nu_{I\leftarrow E}\lambda_E + \alpha\eta \,. \tag{47}$$

where τ and δ are time constants derived from membrane time constant τ_k and recurrent synaptic connection ν_{kk}. Linear response analysis then relies on the Laplace transform (with Laplace variable p) of these equation

$$\Lambda_E(p) + \tau p \Lambda_E = -\nu_{E\leftarrow I}\Lambda_I + N(p) \,, \tag{48}$$

$$\Lambda_I(p) + \delta p \Lambda_I = \nu_{I\leftarrow E}\Lambda_E + \alpha N(p) \,. \tag{49}$$

For the case of *no feedforward inhibition* ($\alpha = 0$), this leads to the ratio of excitatory to inhibitory activity in the Laplace domain

$$\frac{\Lambda_E}{\Lambda_I}(p) = \frac{\delta p + 1}{\nu_{I\leftarrow E}} \tag{50}$$

resulting in excitatory activity being in advance of $\tan^{-1} 2\pi f \tau$ with respect to inhibitory activity at frequency f.

For the case of strong feedforward inhibition ($\alpha = 1$), this leads to

$$\frac{\Lambda_E}{\Lambda_I}(p) = \frac{\delta p + 1 - \nu_{E\leftarrow I}}{\tau p + 1 + \nu_{I\leftarrow E}} \tag{51}$$

such that the phase shift between population is of constant sign across frequencies, but may be positive or negative depending on the exact parameters' values governing the E-I dynamics. Plots summarizing these situations are provided in Figure 6D.

MassAlpha: E-I interactions with alpha type synaptic impulse response

Together with Equation 43, we include in addition a non-trivial synaptic dynamics in the form of the differential equation (assuming same dynamic for both AMPA and GABA synapses),

$$\sigma^2 \frac{d^2 s_k}{dt^2} + 2\sigma \frac{ds_k}{dt} + s_k = \lambda_k \,, \tag{52}$$

This corresponds the classical *alpha synapse* used in computational models (e.g. implemented in the *NEURON* software (Carnevale and Hines, 2006)), modeling the response to a single spike with the alpha function

$$s(t) = \frac{1}{\sigma^2} t e^{-t/\sigma} \,. \tag{53}$$

By combining these equations with linear activations (Equation 45), the dynamics of the circuit is summarized by a 6-dimensional state-space model that can be studied analytically with linear response theory.

Analysis and simulations of neural field models

When taking into account the spatial extension of the network, neural mass models can be extended to neural field models, where the variables described above possibly depend on space. Consider one or two spatial dimensions tangential to the layers of the network (assuming a layered organization like the hippocampus or cortex), the key phenomenon that should be additionally modeled is then the coupling between activity in different locations of the network entailed by horizontal connections. In line with the literature and to simplify the analysis, we will consider only the excitatory connections are spatially extended. With respect to the above generic neural field model Equations 38-40, equations pertaining to synaptic activity and rate need only to be specified. We use the model introduced by Jirsa and Haken (1996, Equation 15) with a spatial diffusion term with characteristic distance r_0 and axonal propagation speed $v_0 = r_0\gamma$, which takes the form of a damped wave equation:

$$\frac{1}{\gamma^2}\frac{\partial^2 s_E}{\partial t^2} + \frac{2}{\gamma}\frac{\partial s_E}{d\partial t} + s_E - r_0^2\Delta s_E = \lambda_E + \frac{1}{\gamma}\frac{\partial \lambda_E}{\partial t}\,, \tag{54}$$

where Δ is the Laplacian operator, while $s_I = V_I$ to encode purely local inhibition, eliminating redundant multiplicative factors.

To specify Equation 40, we use sigmoid activation functions for both AMPA and GABA synapses,

$$\lambda_E = \frac{Q_E}{1+\exp(-\chi_E.(V_E - V_{th,E}))}\,, \tag{55}$$

$$\lambda_I = \frac{Q_I}{1+\exp(-\chi_I.(V_I - V_{th,I}))}\,, \tag{56}$$

whose parameters (maximum rate Q_k, spiking threshold $V_{th,k}$ and excitability χ_k for population k) are adjusted in order to obtain different types of dynamics, either evolving around a stable equilibrium point (model *FieldStable*) or with a clear oscillatory activity (model *FieldOsc*).

Spatio-temporal phase analysis in 1D

Before simulating 2D neural field models with an explicit method, we investigate analytically properties of a simplified 1D model. In this case, the partial differential Equation 54 corresponds to an exponentially decaying connectivity, with axonal propagation speed $v_0 = r_0\gamma$, such that the resulting post-synaptic current takes the integral form (see Jirsa and Haken (1996, Equation (14)))

$$s_E(x,t) = \frac{1}{2r_0}\int \exp(-|x-X|/r_0)\lambda_E(X, t-|x-X|/v_0)dX\,. \tag{57}$$

If we take the neural field equation in the context of horizontal connection along non-myelinated axons ($v_0 \sim 1m/s$, $r_0 < 1mm$), the typical value of γ is beyond 1000 such that if we focus on frequencies below $200Hz$, we may neglect for the temporal derivatives of the partial differential equation. This leads to the following approximation for the dynamics of excitatory post-synaptic current.

$$s_E - r_0^2\frac{\partial^2 s_E}{dx^2} = \lambda_E\,. \tag{58}$$

In an unbounded 1D medium, assuming that activities vanish at large distances, we can use the spatial Fourier transform $\widehat{f}(t,z) = \mathcal{F}_x[f(t,x)](z) = \int_{\mathbb{R}} f(t,x)e^{-2i\pi zx}dx$ to derive the expression of s_E as a function of λ_E

$$\widehat{s_E}(t,z) = \frac{1}{1+(2\pi z)^2 r_0^2}\widehat{\lambda_E}(t,z)\,. \tag{59}$$

In order to get back to the original spatial position domain, we use a general Fourier transform relation for an arbitrary complex parameter a such that $\mathrm{Re}[a] \geq 0$:

$$\mathcal{F}_x\left[\frac{1}{2a}e^{-|x|a/r_0}\right](z) = \frac{r_0}{(2\pi z r_0)^2 + a^2}\,, \tag{60}$$

This formula can be inverted, considering an arbitrary complex number b, and defining $\sqrt{b}$ to be the unique complex number such that $\sqrt{b}^2 = b$ and $\mathrm{Re}\sqrt{b} \geq 0$, we get

$$\mathcal{F}_z^{-1}\left[\frac{r_0}{(2\pi z)^2 r_0^2 + b}\right](x) = \frac{1}{2\sqrt{b}} e^{-|x|\sqrt{b}/r_0}\,, \tag{61}$$

For the particular case $b = 1$, we get $\sqrt{b} = 1$, which, in the spatial position domain, leads to

$$s_E(t,x) = \int_{\mathbb{R}} \lambda_E(t,y) h_{r_0}(x-y)dy = h_{r_0} * \lambda_E(t,x)\,, \tag{62}$$

where $*$ denotes spatial convolution and $h_{r_0}(x) = \frac{1}{2r_0} e^{-|x|/r_0}$. This reflects that horizontal connectivity generates EPSCs corresponding to a spatial smoothing of the excitation rate spatial distribution.

After linearizing around the operating point of the network (absorbing again the resulting multiplicative constant in the connectivity matrix), we obtain the equation of the dynamics by modifying Equation 38 (assuming neither long range nor feedforward inhibition)

$$\lambda_E(t,x) + \tau \frac{d\lambda_E}{dt} = \nu_{E\leftarrow E} s_E(t,x) - \nu_{E\leftarrow I}\lambda_I(t,x) + \eta(t,x)\,, \tag{63}$$

$$\lambda_I(t,x) + \delta \frac{d\lambda_I}{dt} = \nu_{I\leftarrow E} s_E(t,x)\,, \tag{64}$$

where the synaptic strength values incorporate multiplicative constants resulting from the linearization of Equations 55-56. By computing the temporal (with frequency variable f) and spatial Fourier transform of each equation, we get (using $p = i2\pi f$)

$$\widehat{\Lambda_E} + \tau p \widehat{\Lambda_E} = \nu_{E\leftarrow E}\widehat{S_E} - \nu_{E\leftarrow I}\widehat{\Lambda_I} + \hat{H}\,, \tag{65}$$

$$\Lambda_I(t,x) + \delta p \widehat{\Lambda_I} = \nu_{I\leftarrow E}\widehat{S_E}\,. \tag{66}$$

Eliminating $\widehat{\Lambda_I}$ we get

$$(1+\tau p)\widehat{\Lambda_E} = \left(\nu_{E\leftarrow E} - \nu_{E\leftarrow I}\frac{\nu_{I\leftarrow E}}{1+\delta p}\right)\widehat{S_E} + \hat{H}\,. \tag{67}$$

Combined with the spatially Fourier transformed horizontal connectivity Equation 58 (using $k = i2\pi z$)

$$(1 - r_0^2 k^2)\hat{S}_E = \widehat{\Lambda_E}\,, \tag{68}$$

this leads to

$$\widehat{S_E} = \frac{1}{1+\tau p}\frac{\hat{E}}{-r_0^2 k^2 + 1 + \frac{1}{1+\tau p}\left(\frac{\nu_f}{1+\delta p} - \nu_{E\leftarrow E}\right)}\,, \tag{69}$$

where we define the *feedback inhibition gain* $\nu_f = \nu_{E\leftarrow I}\nu_{I\leftarrow E}$.

Introducing our time-space separability assumption on the exogenous input

$$\eta(x,\,t) = n(x)\epsilon(t) \tag{70}$$

leads to

$$\widehat{S_E}(z,\,f) = \frac{1}{1+i2\pi\tau f}\frac{E(f)\hat{n}(z)}{-r_0^2(2\pi z)^2 + 1 + \frac{1}{1+i2\pi\tau f}\left(\frac{\nu_f}{1+i2\pi\delta f} - \nu_{E\leftarrow E}\right)}\,, \tag{71}$$

By defining

$$b = 1 + \frac{1}{1+i2\pi\tau f}\left(\frac{\nu_f}{1+i2\pi\delta f} - \nu_{E\leftarrow E}\right)\,, \tag{72}$$

and using the inverse spatial Fourier transform of Equation 61, we get

$$S_E(x,\,f) = \frac{1}{2r_0\sqrt{b}}\frac{E(f)}{1+i2\pi\tau f} n(x) * e^{-|x|\sqrt{b}/r_0}\,, \tag{73}$$

Assuming the exogenous input does not impose a spatial phase gradient to the structure (i.e. $n(x)$ is positive real for all locations up to a multiplicative constant), the phase gradient at a given frequency

will be controlled by the imaginary part of $\sqrt{b}$. Specifically, to investigate qualitatively the phase gradient around a peak of activity of the exogenous input, we assume that $n(x)$ is a dirac at $x = 0$. Then the spatial variation of the phase around $x = 0$ take the form

$$\phi(x) = -\frac{|x|}{r_0}\mathrm{Re}\left[\sqrt{b}\right] . \tag{74}$$

This dirac approximation, does not match well our simulations (using a Gaussian shape spatial input distribution). However, computing spiking activity form S_E based on Equation 58, to obtain the spatial distribution of the spike vector, will have a deblurring effect compensating the convolution by $n(x)$ in Equation 75, making in closer to a Dirac. As a consequence, we will interpret the data based on the following approximation

$$\lambda_E(x,\, f) \approx C\frac{1}{2r_0\sqrt{b}}\frac{E(f)}{1 + i2\pi\tau f}e^{-|x|\sqrt{b}/r_0} , \tag{75}$$

up to a multiplicative constant C.

In order to investigate the qualitative effect of the microcircuit connectivity on this spatial gradient, we assume $\tau = \delta$ and use a low (temporal) frequency assumption of the form $f << 1/\tau$, such that we can exploit a first order expansion for the fractions containing the term $\tau p << 1$. This leads to the approximation

$$b \approx 1 + (1 - i2\pi\tau f)\left(\nu_f(1 - i2\pi\tau f) - \nu_{E\leftarrow E}\right) \approx 1 + \nu_f - \nu_{E\leftarrow E} - i2\pi\tau f(2\nu_f - \nu_{E\leftarrow E}) . \tag{76}$$

Simple geometric considerations show that the sign of the imaginary part of b is the same as the sign of its square root, such that under our simplifying assumption, the sign of the gradient taken algebraically from center ($x = 0$) to surround ($|x| > 0$) is the sign of

$$2\nu_f - \nu_{E\leftarrow E} , \tag{77}$$

showing that strong feedback inhibition will tend to put the populations surrounding $x = 0$ in advance with respect to this center point, while weak feedback inhibition (with respect to feedback excitation), will to generate a phase lag of the surround with respect to the center.

Neural field simulation in 2D

While the above analysis is much easier to perform in 1D, in most structures (and in particular cortex), the domain spanned by horizontal connectivity is better approximated by a 2D domain, which can also be sampled by modern electrode arrays. We thus simulate the dynamics of such 2D system to get insight into the characteristics revealed by GPLA analysis in this context. We use simplified notations for the 2D (in space) time-varying scalar fields $V(t, x, y) = s_E((x, y), t)$ and $I(t, x, y) = \lambda_E((x, y), t)$. Let Δx and Δt be the spatial and temporal grid spacings, and $V^n_{j,l} = V(n\Delta t, j\Delta x, l\Delta x)$ the discretized field. We use a Forward Time Centered Space (FTCS) finite difference scheme to simulate the above neural field model (Fletcher, 1991). FTCS relies on making the approximations

$$\frac{\partial V}{\partial t}(t, x, y) \approx \frac{1}{\Delta t}\left(V^{n+1}_{j,l} - V^n_{j,l}\right) , \quad \frac{\partial^2 V}{\partial t^2}(t, x, y) \approx \frac{1}{(\Delta t)^2}\left(V^{n+1}_{j,l} + V^{n-1}_{j,l} - 2V^n_{j,l}\right) , \tag{78}$$

$$\text{and} \quad \frac{\partial^2 V}{\partial x^2}(t, x, y) \approx \frac{1}{(\Delta x)^2}\left(V^n_{j+1,l} + V^n_{j-1,l} - 2V^n_{j,l}\right) . \tag{79}$$

Applying these approximations to Equation 54, leads to an explicit scheme for the field values at time $(n + 1)\Delta t$ based on all values at time $n\Delta t$ and $(n - 1)\Delta t$.

$$V^{n+1}_{j,l} = \frac{1}{G + 1}\left((G - 1)V^{n-1}_{j,l} + 2V^n_{j,l} + G^2\left(R^2\left(K_\ell *_d \boldsymbol{V}^n\right)_{j,l} + I^n_{k,l} + \frac{1}{G}\left(I^n_{k,l} - I^{n-1}_{k,l}\right) - V^n_{j,l}\right)\right) , \tag{80}$$

where $G = \gamma\Delta t$, $R = r_0/\Delta x$ and $K_\ell *_d$ denotes the discrete 2D spatial convolution with the discrete Laplace operator

$$K_\ell = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} . \tag{81}$$

The parameters chosen for both models presented in main text are reported in Table 4.

Table 4. List of neural field model parameters

Parameter name	Symbol	Value (*weak Rec.*)	Value (*strong Rec.*)
E membrane time constant	τ_E	$20ms$	$20ms$
I membrane time constant	τ_I	$20ms$	$20ms$
E-E synaptic strength	$\nu_{E\leftarrow E}$	0.2	0.2
I-I synaptic strength	$\nu_{I\leftarrow I}$	0	0
$E \rightarrow I$ synaptic strength	$\nu_{E\leftarrow I}$	0.2	0.2
$I \rightarrow E$ synaptic strength	$\nu_{I\leftarrow E}$	1	1
E excitability	χ_E	1	1
I excitability	χ_I	0.1	3.33
E sigmoid threshold	$V_{th,E}$	0	0
I sigmoid threshold	$V_{th,I}$	0	0
E maximum rate	Q_E	$20Hz$	$20Hz$
I maximum rate	Q_E	$20Hz$	$20Hz$

Quantification and statistical analysis

Parameter estimation of von Mises distribution

The von Mises distribution (VM), which is also known as "circular normal" distribution is the counterpart of the Gaussian distribution for circular data (Fisher, 1995, Chapter 3). We used it for various purposes in this work (e. g. to model the spiking probability to synthesize phase-locked spike trains).

The VM distribution takes the form,

$$p(\phi|\varphi_0,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left(\kappa\cos(\phi-\varphi_0)\right), \tag{82}$$

where $I_0(\kappa)$ is the modified Bessel function of order zero (Equation 33). In Figure 8 and 13, we fit a VM distribution to the pooled phases of spike and LFP vectors coefficients. We use a maximum likelihood (ML) method for estimating the two parameters of the VM distribution, φ_0 and κ (Fisher, 1995). The ML estimation of φ_0 is simply the sample mean direction, denoted by $\bar{R}$ (for spike-LFP data is the locking phase). Maximum likelihood estimation of, $\hat{\kappa}$, is the solution of following equation:

$$A_1(\hat{\kappa}) = \bar{R}, \tag{83}$$

and A_1 is a ratio of two modified Bessel functions:

$$A_1(x) = \frac{I_1(x)}{I_0(x)}. \tag{84}$$

Approximate solutions are available for $\hat{\kappa}$ (Fisher, 1995, sec. 4.5.5)

$$\hat{\kappa} = \begin{cases} 2\overline{R} + \overline{R}^3 + 5\overline{R}^5/6 & \overline{R} < 0.53 \\ -0.4 + 1.39\overline{R} + 0.43/(1-\overline{R}) & 0.53 \le \overline{R} < 0.85 \\ 1/\left(\overline{R}^3 - 4\overline{R}^2 + 3\vec{R}\right) & \overline{R} \ge 0.85 \end{cases} \tag{85}$$

where $\overline{R}$ is the resultant length of the phases.

Animal preparation and intracortical recordings

The methods for surgical preparation, anesthesia, and presentation of visual stimuli for the Utah array recordings have been described in previous studies (see Logothetis et al. (1999, 2002); Belitski et al. (2008); Safavi et al. (2018)).

Data collection

Neural signals were recorded with a NeuroPort Cortical Microelectrode Array (Blackrock Microsystems, Salt Lake City, Utah USA). An array was implanted in the inferior convexity of the prefrontal cortex (see Safavi et al. (2018) for more details). The arrays are 4mm $\times$ 4mm with a 10 by 10 electrode configuration. Neural signals recorded from 96 of the available 100 electrodes. Neural activity was recorded in 200 trials. Each trial consisted of a 10s period of movie presentation, followed by 10s of a blank screen (inter-trial).

LFP extraction

The raw signals were low-pass filtered using an 8th order Chebyshev Type 1 filter with a cut-off frequency of 200Hz and a pass-band ripple less than 0.05dB. Forward and backward filtering was used to minimize phase distortions caused by the filtering. Next, the filtered signal was decimated to a sampling frequency of approximately 500Hz.

Spike detection

For detecting multi-unit spikes, the raw signal was band-pass filtered using a minimum–order finite impulse response (FIR) filter (Rabiner et al., 1975) with pass-band cut-off frequencies of 600Hz to 5800Hz and stop-band cut-off frequencies of 400Hz and 6000Hz, with at least 65dB attenuation in the stop-bands and less than 0.002dB ripple within the pass-band. The amplitude threshold for spike detection was set to 5 standard deviations above the average of the filtered signal (Quiroga, 2007). To spare computational costs, the standard deviation of the signal for each channel was estimated using a smaller, randomly chosen section of the filtered signal. Spike times with inter-spike intervals less than the refractory period of 0.5ms were eliminated.

The complex spectral structure of transient LFPs reveals subtle aspects of network coordination across scales and structures.

Michel Besserve[1, 2], Shervin Safavi[1], Bernhard Schölkopf[2] and Nikos K. Logothetis[1, 3]

[1]Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany.
[2]Department of Empirical Inference, Max Planck Institute for Intelligent Systems, Spemannstrasse 38, 72076 Tübingen, Germany.
[3]Centre for Imaging Sciences, Biomedical Imaging Institute, The University of Manchester, Manchester M13 9PT, United Kingdom.

Corresponding Author: Michel Besserve
Spemanstrasse 38,
72076 Tuebingen
Germany
Email: michel.besserve@tuebingen.mpg.de

Keywords: Local field potentials, oscillations, neural events, Neural Event Triggered-fMRI, Hippocampus, Thalamus.

Abstract

Brain Local Field Potentials (LFP) exhibit various dynamical patterns reflecting underlying cooperative network mechanisms at multiple scales that are still largely unexplored in subcortical structures. Such patterns have long been characterized by their activity in various frequency bands. However such decomposition fails to capture fundamental features electrical signals. Using an band free decomposition approach applied to single channel recordings from anesthetized monkey in Hippocampus, we identifies dynamical patterns, such as Sharp Wave-Ripple complexes, and characterizes their brain-wide properties using concurrent fMRI recordings. We also link LFP events to the underlying local activity changes in the recorded structure and show this relationship is highly structure specific. Overall, this approach helps elucidate the relationship between LFP transient activity, the local circuit dynamics and brain-wide interactions.

Introduction

The ability of mammals to react quickly to environmental changes requires their brain circuits, highly interconnected at multiple scales (Abeles, 1982, Honey et al., 2007), to coordinate their activity efficiently. Local Field Potential (LFP) activity, generated by transmembrane currents originating from the cells near to the intracortical electrode tip (Mitzdorf, 1985, Logothetis and Wandell, 2004), is an important marker of neural cooperation as it reflects several local perisynaptic integrative processes (Logothetis, 2008). In particular, transient phenomena, such as rhythmic bursts (Roopun et al., 2008, Wang, 2010, Buzsáki, 2006), or more complex patterns, such as Sharp Wave-Ripples (SPW-R) (Buzsaki et al., 1992), reflect various network mechanisms instrumental to brain function (Buzsáki, 2006, Wang, 2010). Fully characterizing these neural events across time and brain structures would allow decomposing brain activity into sequences of *atoms* of neural computation and greatly help understand the dynamics of neural information processing at a system level (Marcus et al., 2014). To achieve this characterization, classical LFP frequency bands have been established in the clinical electroencephalography and electrophysiology literature, dominated by the study of neocortex and a few other structures. These bands are scrutinized in numerous studies investigating functions as diverse as sleep, perception or motor control.

However, the LFP dynamic patterns and associated mechanisms are highly dependent on the detailed local properties of the network, including the recurrent local organization of canonical microcircuits, synaptic inputs from other brain structures, as well as diffuse neuromodulatory inputs from brainstem and basal forebrain. Not only the above circuit structures affect LFPs by shaping the sequences of excitations and inhibitions of various neuronal types, but also the detailed geometrical arrangement of cells will determine which part of the network activity is be captured by the extracellular electrical field. While these circuit properties show a relative stability across the mammalian neocortex, subcortical regions exhibit a large diversity. As a consequence, studying a subcortical region by scrutinizing its LFP activity in predefined frequency bands might result in a suboptimal, if not misleading, description of relevant activity in this particular structure. In addition, several aspects reflecting the often non-linear properties of the mechanisms generating LFP neural events are not well captured by a decomposition of the LFP signal in fixed frequency bands. In particular, informative aspects of transient neural activity also lies in the more complex features of neural activity, such as harmonics and cross-frequency coupling, showing that LFP activity can involve several frequencies at the same time (Thiagarajan et al., 2010, Abeysuriya et al., 2014a, Canolty et al., 2006, Contreras et al., 1996). These observations call for a general *data-driven* methodology to identify relevant events in LFP data without relying on an *a priori* choice of frequency bands to scrutinize. To overcome the issues of fixed-band LFP analysis, stimulus-informed methods have been introduced for partitioning of the frequency domain (Belitski et al., 2008, Montemurro et al., 2008, Magri et al., 2012), but cannot be used when overt stimuli or behavioral information are unavailable, as it is the case in most sleep and anesthesia studies. One the other hand, independent component analysis techniques or realistic biophysical generative models of LFPs (Makarov et al., 2010,

Einevoll et al., 2007, Schomburg et al., 2014) can extract relevant LFP components with less prior information, but are not appropriate when only one or few recording channels are available.

In the present study, we propose to take advantage of the rich dynamical information available in single channel LFP data to detect all types of dynamical patterns using a two-step methodology. First, LFPs are modeled as a sum of dynamical components, identified by their Power Spectral Density (PSD) profile. Second, we detect frequently occurring temporal patterns in these components using a dictionary learning approach. We apply this methodology to signals recorded with extracellular electrodes in the CA1 area of the hippocampus (denoted Hp along this manuscript) of anesthetized monkeys. This subcortical structure have as privileged relationship to neocortex by reactivating cortical assemblies encoding past sensory experiences (Ji and Wilson, 2007). We use concurrent whole brain functional Magnetic Resonance Imaging (fMRI) recordings to characterize the LFP events according to their associated Blood Oxygen Level Dependent (BOLD) responses in multiple structures, such that both their local and brain-wide properties can be assessed simultaneously. A clustering of the events detected by our approach identifies 6 different neural events in this structure. In addition, the temporal couplings of hippocampal events suggest that the events' occurrences are temporally organized at a supra-second time scale, possibly reflecting a global brain state dynamics. These results illustrate how the specificity of transient LFP neural activity in various brain structures can be quantified using an appropriate identification of relevant neural events. The repertoire of LFP events built in this way will help elucidate how a given brain structure participates in the overall brain activity at multiple temporal and spatial scales.

Results

Dynamical properties of LFP signals

Extracellular potentials from the pyramidal layer of the CA1 region of hippocampus (Hp), or from LGN were recorded simultaneously with whole-brain fMRI in anesthetized monkeys. The detailed experimental setup has been previously described in (Logothetis et al., 2012) (see Materials and Methods). As illustrated in Figure 1a, the LFP time-course undergoes fast changes in its spectral content over time as witnessed by the occurrence of oscillatory bursts at different time scales. Such changes can be captured with low run time complexity by computing the Short Term Fourier Transform (STFT) spectrogram, also represented on Figure 1a. Bursts of high and low frequency oscillations appear at different locations of the spectrogram (see Figure 1a, orange and green rectangles respectively). The occurrence of such transient brain rhythms has been reported in a large number of brain structures and species (Buzsáki and Draguhn, 2004) and associated to various underlying mechanisms (Brunel and Wang, 2003, Contreras et al., 1996). Interestingly, Figure 1a also shows increases in the magnitude of the spectrogram involving broader frequency ranges (purple rectangle). The variety of transient spectral profiles observed in this example LFP time series likely reflects different underlying cooperative mechanisms as

supported by modeling and experimental studies. For example, recurrent interactions within and between populations of excitatory and inhibitory neurons in local cortical microcircuits generates high frequency oscillations in the LFP (Brunel and Wang, 2003) (Figure 1b, orange arrows). Alternatively, post-synaptic currents generated by a volley of synchronous action potentials in a remote afferent brain region typically result in slower LFP waveforms (Buzsaki, 1986) (Figure 1b, green arrows). More generally, combinations of such local and long-range interactions can generate more complex LFP patterns involving a broad range of frequencies. As a consequence, our approach to discover cooperative mechanisms in a recorded brain structure relies on detecting transient LFP activities characterized by specific spectral profiles and further study their properties.

Time varying spectral decomposition of LFPs

To identify these profiles efficiently, the STFT spectrogram matrix shown in Figure 1a is modeled as a linear superposition of a few spectral profiles associated to the different underlying events. Across time, each spectral profile is assumed to keep the same shape, but its contribution to the spectrogram can vary in magnitude. Stated in this form, retrieving these spectral profiles from data can be cast as a Non-negative Matrix Factorization problem (NMF) (Seung and Lee, 2001). Let **S** be the (frequency x time) spectrogram matrix, NMF seeks optimal matrices **W** and **H** having only positive coefficients and such that:

$$\mathbf{S} \approx \mathbf{WH} \tag{1}$$

The number of columns of **W** is the same as the number of lines of **H** and sets the assumed number of spectral profiles K appearing across time in the data. As illustrated in Figure 1c, while the columns of **W** correspond to the spectral profile of each of these components, the lines of **H** are weights quantifying the time-varying contribution of the spectral profiles to each time window. The NMF problem is solved by minimizing a dissimilarity measure called *divergence*, quantifying how close the factorization on the right-hand side of Equation 1 is to the original spectrogram matrix (Seung and Lee, 2001, Févotte et al., 2009, Sra and Dhillon, 2005). While the Euclidean norm is the simplest and most widely used divergence, we argue that NMF based on the Itakura-Saito divergence (IS-NMF), previously used in music analysis (Févotte et al., 2009) is particularly suited to LFP analysis due to its scale invariance properties (see Supplementary Methods). This allows the method to automatically adapt to the typical "1/f" distribution of the PSDs observed in neural time series (Novikov et al., 1997, He et al., 2010) and illustrated by the decrease of spectrogram values with increasing frequency in Figure 1a. As a consequence, IS-NMF can better detect low power high frequency spectral profiles frequently observed in empirical LFP data, as illustrated in simulations in Supplementary Results and Supplementary Fig. 1. In addition, the IS-NMF approach allows for a probabilistic modelling of the original signal, which associates NMF results to a decomposition of the original time series into dynamical components (Smaragdis et al., 2014) that we will exploit in later sections.

The outcome of IS-NMF for the recordings shown in Figure 1a using 3 components is also represented in Figure 1c. The estimated spectral profiles (normalized by their maximum value) show that IS-NMF captures various aspects of neural activity altogether covering a wide range of frequencies (from 2-150 Hz). The time-varying contribution of the spectral profiles exhibit isolated peaks corresponding to the occurrence of transient events in the LFP.

Classification of spectral components in hippocampus and LGN

To compare the components detected by IS-NMF in Hp and LGN, we ran the algorithm on several sessions (21 for Hp recordings, 11 for LGN recordings), each of them consisting in 10-20 experiments of 10 min recordings of spontaneous activity. The number of spectral profiles in the NMF decomposition is a parameter subject to trade-of: a large number of components leading to a finer description of neural activity, but at the expense of a larger number of samples necessary to robustly estimate the solution. We assessed the robustness of the decomposition with cross-validation and chose the largest number of components which across sessions led to an average cosine similarity between cross-validated spectral profiles above 80% (see Materials and Methods). The optimal number of 4 components was found for Hp, while 3 components were optimal for LGN (see Supplementary Fig. 2a-b). To analyze the properties of the resulting profiles across sessions, they were pooled together and then clustered (using the above mentioned optimal number of profiles as the number of clusters) based on their pairwise cosine similarity using the normalized cut graph clustering algorithm (Shi and Malik, 2000). The NMF outcome of a few sessions showing components reflecting artifact contamination (recognizable in the spectral profiles by the presence of sharp peaks distributed over a wide range of frequencies) were excluded from this clustering analysis (4 out of 21 for Hp, 1 out of 11 for LGN). The characteristics of the resulting clusters of spectral profiles are summarized on Figure 1d-e. The normalized spectra of Hp clusters are shown in Figure 1d, exhibiting a low frequency component (in blue), as well as several spectra having most of their energy in a specific frequency band: 15-40 Hz, 40-90 Hz, and finally 90-140 Hz. In comparison, the three LGN clusters, computed with the same clustering technique, are shown on Figure 1e and consist in one low frequency component (<15 Hz), one component with most of its energy in a narrow band (15-25 Hz), as well as a broad high frequency component (25-130 Hz). In comparison to Hp components, LGN spectral profiles thus extend to a lower frequency range, suggesting differences in the underlying network dynamics of these structures. The observed transient spectral profiles describing neural activity can differ in many respects, which are not only limited to their peak frequency. Previous modeling studies have shown that the non-linear properties of the underlying neural network can affect the shape of the PSD of the observed neural time series, for example by generating harmonics at multiples of the peak frequency (Abeysuriya et al., 2014a, Abeysuriya et al., 2014b, Breakspear et al., 2006, Robinson et al., 2002). While a more detailed study of these spectral properties is left to further studies, we characterized our spectral profiles using two simple parameters: their peak frequency and their spectral centroid (see Supplementary Methods). Spectral centroid measures the center of mass of the spectral profile distribution across frequencies. As a consequence, the closer it gets to the

peak frequency; the closer is the spectral profile to a single narrow-band peak deprived from harmonics, and the closer is the associated time course to a sinusoid (see illustration Figure 1f, more details are provided in Supplementary Methods and Supplementary Fig. 3). Hence, we call the ratio of the spectral centroid to the peak frequency the *spectral purity ratio* and use it to quantify putative non-linear network interactions associated to each profile (see results Fig. 1g-h). Interestingly, the spectral purity ratio was higher for LGN spectral profiles than for Hp profiles ($p<10^{-4}$, one-sided Wilcoxon rank sum test, n=81), suggesting more prominent non-linear network dynamics in LGN.

Time course of dynamical components

While IS-NMF results provide an overview of the spectral profiles present in LFP data, it is important to understand how such profiles contribute to the time course of neural activity and whether they can be interpreted in terms of known phenomena. One important feature of the IS-NMF decomposition is to associate a time course to each spectral component according to the Gaussian Composite Model (Févotte et al., 2009, Smaragdis et al., 2014), leading to a linear decomposition of the LFP signal into a sum of dynamical components in the time domain (see Supplementary Methods). After applying an invertible decomposition of the signal into blocks using overlapping tapering windows (Allen, 1977), the time course of each component can be estimated efficiently for each block based on NMF results by using a time varying filter bank. In practice, as illustrated in Figure 2a-b, normalized NMF profiles for a given time window define the filter bank applied to the LFP in the Fourier domain to generate the time courses of each dynamical component. This approach is performed on successive overlapping time windows and the full time course of each component is reconstructed by simply summing the contribution of all windows as illustrated in Figure 2c.

In Figure 3a-b, we represent examples of the estimated time course of dynamical components for Hp and LGN recordings respectively. It can be seen that while events are difficult to identify in the LFP traces, each component exhibits occurrences of stereotypical patterns, some of which appearing sinusoidal (Figure 3a green component), while others exhibit more complex shapes (Figure 3a cyan component). This decomposition thus helps us detecting potentially interesting neural events and their characteristics. To detect such events automatically, we apply to each component a shift-invariant dictionary learning technique inspired by work in music analysis (Mailhé et al., 2008). This approach approximates each time series by a superposition of shifted and rescaled typical patterns appearing in the time course. The recursive estimation of these patterns is schematized in Figure 3c and explained in Materials and Methods (see also the simulation study in Supplementary Results). The method parameters were chosen in cross-validation (see Supplementary Methods). Figure 3a-b shows the resulting detected events surrounded by colored rectangles and the estimated dictionary patterns of each dynamical component are shown on Figure 3d-e for Hp and LGN examples shown in Figure 3a-b respectively. Interestingly, it is possible to identify among detected patterns important events reported in the literature as hippocampal SPW-R (Figure 3a top component). Being able to detect

these events with our approach is particularly interesting because it is a compound of a low frequency waveform, the sharp wave and a high frequency oscillation, the ripple. It thus shows the capability of the reconstruction technique to detect patterns with complex (non-sinusoidal) time course corresponding to non-linear dynamical interactions.

Local and brain-wide neural event properties

The LFP neural events, although detected at a single recording site, likely reflect mechanisms engaging a broader set of structures. We take advantage of the simultaneously recorded fMRI signal to compute the Neural Event Triggered(NET)-fMRI activity in cortical and sub-cortical Regions of Interest (RoI) to assess how the occurrence of a neural event relates to metabolic changes in various brain structures (see illustration Figure 4a and Materials and Methods). We first computed the magnitude of the NET-fMRI responses following event onset in two broad brain regions that were relevant in previous studies : neocortex and thalamus (Logothetis et al., 2012) (see Materials and Methods). The raster plot of thalamic response against cortical response are represented Figure 4b respectively. These plots show a clear and significant negative correlation between thalamic and cortical response (Spearman: $\rho = -.855$, $p < 10^{-7}$). Detected Hp events thus tend to have opposite metabolic correlates in neocortex and thalamus, extending previous observations regarding hippocampal SPW-R (Logothetis et al., 2012). We checked that this effect could not be explained by an overall negative correlation between the two structures. As show in Figure 4f, the distribution of correlation between fMRI recordings from thalamus and cortex has a positive median (Wilcoxon signed rank test, $p<.001$), and thus cannot account for the opposite signs of thalamic and cortical responses during hippocampal events. These results support that events detected in Hp reflect competitive interactions between neocortex and thalamus, and we quantify this property for each individual event by defining a *Thalamocortical Competition Index* (TCI) as the algebraic coordinate of each point projected in Figure 4b along the regression line relating thalamic to cortical response. Large positive TCI values (towards bottom right) represent events with large positive cortical response and large negative thalamic response, while negative values reflect the opposite effects. According to the sign of TCI, we split all our observed events in two categories: the cortex-activating (CA) events, with positive TCI, and the thalamus-activating (TA) events, with negative TCI. We then analyzed the properties of these categories separately by applying a clustering procedure on each subset based on the peak frequency of the Fourier transform of their dictionary pattern (see Materials and Methods). The optimal number of clusters was chosen according to the silhouette index (Desgraupes, 2013) (see Materials and Methods). To overview together the brain-wide and local properties of the clustered events, we show in Figure 4c their corresponding dictionary patterns, mapped according to the magnitude of their TCI as well as the peak frequency of their LFP dictionary pattern. CA patterns were clustered in 3 subtypes, associated to LFP activity in lower frequencies (median peak frequency of 4.5 Hz), the classical EEG alpha band (10.0 Hz median frequency peak) and the classical EEG beta band (median peak frequency 21.9 Hz). The CA clusters are also in the number of 3. Two of these clusters (Clusters 5 and 6) correspond to low gamma and ripple events according to their median peak frequencies at 37.5 Hz and 94.3 Hz

respectively, and are associated to strong cortical activation and thalamic deactivation, as previously reported (Logothetis et al., 2012). The additional CA event subtype (Cluster 4), has a frequency range similar to Cluster 1 (4.5 Hz median peak frequency). However, one can notice in the dictionary patterns belonging to this cluster an additional high frequency oscillation, as illustrated by the magnified example pattern of Figure 4c.

To better understand this exception, we assessed in more detail the neurophysiological differences between low frequency cortex-activating Hp events from their thalamus-activating counterpart. We thus computed the neural event triggered average of the LFP spectrogram (see Materials and Methods) for all Hp events. The average of these spectrograms across clusters, showing the frequency content of LFPs specific to each cluster in a peri-event time window, are shown on Figure 5a. While all spectrograms exhibit a peak in power at event onset around previously measured median peak frequency, we observe that Hp Cluster 4 has an additional high frequency peak (157 Hz peak frequency), corresponding to the previously observed oscillation apparent in the dictionary patterns (see example Figure 5a). This event cluster thus clearly corresponds to hippocampal SPW-R events (Buzsaki et al., 1992, Ylinen et al., 1995), paradigmatic of neural events involving multiple frequencies.

While events have been found to fall in two categories, CA or TA, their brain-wide properties might differ in other respects, reflecting differences in brain-wide network interactions. To check in more detail the differences between large scale metabolic changes associated to each Hp event cluster, we plot the average time course of NET-fMRI responses in two representative brain structures (Figure 5b). While the three TA event subtypes, have in common a deactivation in cortical structures, some NET-fMRI responses deviate above 0 before the stimulus onset (see green arrow in Figure 5b). We quantified the significance of this pre-activation phenomenon in cortex by averaging the responses across all cortical RoIs over a period of 6s preceding stimulus onset. The distribution of average NET-fMRI pre-activation magnitudes for TA clusters is represented on Figure 6a. Interestingly, only the events with lower frequency oscillation patterns (Cluster 1) show significant cortical pre-activation according to a Wilcoxon signed rank test ($p<.01$, Bonferoni corrected), which suggests this low frequency event might not emerge spontaneously from hippocampus, but would rather be caused by earlier changes in brain activity.

Event temporal dynamics

While the fMRI correlates of neural event clusters were analyzed independently in the above section, their occurrence might be linked dynamically; making two types of events more probable to occur in close temporal proximity. Such dynamical coupling may in turn relate to the NET-fMRI results observed above in the case of Hp events. In particular, the similarity between the NET-fMRI responses associated to each CA events might reflect the occurrence of these different events in close temporal proximity. In addition, the cortical pre-activation phenomenon observed for Hp cluster 1 could be due to the occurrence of a CA event several seconds before event onset. To test these conjectures, we computed the time resolved conditional intensities (or

second order intensity) of one type of event (the number of events occurring per time unit) given a "conditioning" event of a given type occurs at time zero (Brillinger, 1976). While the overall conditional intensity plots for all possible event pairs are provided in Supplementary Fig. 8, Figure 5 c provides key example of dynamic coupling between ripples and other events. First, as previously observed (Logothetis et al., 2012), the detected ripples have a significant probability to occur in sequences. This is validated by the increased ripple intensity in the neighborhood of a ripple event onset. In addition, ripples have also more chances to occur in close neighborhood to Sharp-waves. This confirms the classical result that Sharp-waves and ripples frequently occur together, although not systematically (Ramirez-Villegas et al., 2015). Overall, there is a considerable coupling between the occurrence of all CA events in Hp, possibly explaining the similarity between the NET-fMRI signatures of sharp-waves, ripples and gamma oscillations. In addition, among CA events, ripples have specifically more probability to occur before lower frequency oscillation (cluster 1). This again provides a putative explanation for the cortical pre-activation associated to low frequency oscillations. Overall, the NET-fMRI responses may not only reflect the activity related to a given isolated event, but also the metabolic changes due more broadly to the sequences of events occurring in multiple structures, that are dynamically linked to the detected event.

Local neural activity during neural events

We observed a similarity of the relationship between brain wide fMRI activity and LFP patterns in two subcortical structures: high frequencies are related to an increase in the neocortical BOLD signal. Previous results from LFP recordings in primary visual cortex (Logothetis et al., 2001, Murayama et al., 2010), showed that high frequency LFP activity was in good correspondence with the BOLD signal in the tissue surrounding the recording site, and a better predictor of the fMRI signal than spiking activity estimated using the MUA signal. To investigate whether the relationship of LFP patterns to the underlying local neural activity is similar in subcortical structures recorded in the present study, we use two quantifications of neural activity: on the one hand the amount of multi-unit spiking in the underlying populations measured electrophysiologically, and on the other hand the metabolic activity measured by the BOLD signal. To quantify the massed firing rate in the neighborhood of the electrode, we extracted the Multiple Unit Activity (MUA) signal by filtering the extracellular signal in the (800-3000 Hz) frequency range and rectifying it. The distribution of average MUA activity changes in a 400ms peri-event time window (excluding a 10ms peri-event window to avoid artifacts due to event detection), Z-scored with respect to randomized events, are shown for each event clusters in both structures on Figure 6c and 6d respectively. Local metabolic activity changes are quantified as in the previous section, using the magnitude of the NET-fMRI responses (see Material and Methods) in the manually labeled ROI associated to each structure where the electrode is located, namely LGN and Hippocampus. Distributions of local metabolic changes for each event clusters are shown for each structure on Figure 6e and 6f. Interestingly comparison of MUA and local metabolic activity show similar trends, significantly positive metabolic changes corresponding to significant MUA increases. On the other hand, other events do not appear to have significantly

negative metabolic changes (although median values are negative), and are related to no significant changes in MUA activity. Most importantly, there is a clear relationship between local activity changes and the TA/CA property of events in both structures, but with a major difference. While CA events lead to the largest increase in local activity in Hp, TA events are the ones reflect increase in local activity in LGN. This implies that LFP events in similar frequency bands have a largely structure dependent relationship to the underlying level of local activity. In particular, these results suggest that low frequency (below 25Hz) events relate to local metabolic increases in LGN, but not in Hp. Overall, the results emphasize that the local network properties associated to neural events are largely structure dependent.

Neural events in the model of thalamocortical system

In order to investigate, to what degree the detected neural events are informative about the cellular processes (e.g. membrane potentials and ionic currents), we exploit a simulation of thalamocortical system developed by (Costa et al., 2016). The putative relationship between neural events and cellular dynamics is import from two aspects. First, presence of a relationship between neural events and cellular dynamics signify the fact that neural events are not just statistically important pattern in the LFPs, but also they can be mechanically meaningful.
Second, presence of a relationship adds another piece of evidence that neural events, not only provide us a time window that meso-scale dynamics is closely related to macro-scale dynamics (Logothetis et al., 2012), but also the micro-scale dynamics of the brain.

We identify neural events in the membrane potential of excitatory population only in the thalamus module (Figure 6a), as a crude proxy of LFP signal (see the method section for the justification). Applying our method on approximated LFP (will be called briefly LFP rather approximated LFP in the remaining of this section for the simplicity) led to 3 types of neural events (Figure 6b) depict LFP surrounding the exemplary of each neural event). Interestingly, two of the identified events are the well-known type I and type II thalamic spindles.

Further analysis of the cellular dynamics in the vicinity of the neural events with high amplitude (determined based on dictionary learning), suggests that each neural event has a distinct profile cellular dynamic. We build a large feature vector from the time course of all membrane potentials and the calcium current surrounding the neural events and reduce its dimensional with t-distributed stochastic neighbor embedding (tSNE) (van der Maaten et al. 2008). Interestingly, a 2-dimensional representation the cellular dynamics underlying neural event demonstrate clear clusters (Figure 6c).

Discussion

Neural information processing relies on cooperative phenomena which manifest themselves in the complexity of recorded brain signals. We introduced a principled approach to exploit the dynamical properties of single channel LFPs to detect these phenomena with minimal prior

assumptions. Two key features of this approach are that it does not rely at all on band-pass filtering of the signals, and that it associates to each detected event a time resolved pattern of activity. Compared to analysis in predefined frequency bands, these features are beneficial for electrophysiology data analysis and modeling purposes, as they avoid mixing the effects of different types of events, such as CA and TA events, which would result from choosing a frequency band that is not adapted to the data at hand. In addition, this approach facilitates the interpretability of the results, as it can automatically capture non-linear properties of events, such as the involvement of several frequencies in Sharp-Wave Ripples, and computes typical temporal patterns of the events that are easily identifiable in the LFP time course. Such results can be further exploited in modelling studies in order to guide the design of the network mechanisms that generate such LFP properties (see also Supplementary Discussion for additional discussion regarding the chosen approach).

We applied this methodology to ongoing macaque LFP recordings in two structures: hippocampus and LGN and were able to detect and characterize previously reported phenomena without any prior knowledge. Since whole-brain fMRI activity was recorded concurrently with electrophysiology, we studied the large scale brain activity during the detected events. We used the NET-fMRI methodology (Logothetis et al., 2012) to quantify the level of activation or deactivation related to each event in a large variety of cortical and subcortical RoIs. In line with earlier results focused on SPW-R events (Logothetis et al., 2012), the detected events in both structures reflect a competition between cortical and subcortical regions. While the mechanisms and functional role of this thalamo-cortical competition are yet unknown, we speculate that the large variety of neuromodulatory inputs to thalamus may be involved (Varela, 2014), and that such mechanism can avoid different information processing pathways to interfere. In particular, the replay of memory traces in cortex triggered by hippocampal SPW-R (Ji and Wilson, 2007) should not be altered by sensory information reaching cortex through thalamic relays to ensure correct encoding of those memories. Importantly, the observed thalamo-cortical competition is specifically related to the events observed in the recorded subcortical structures, as the fMRI signals in cortex and thalamus are overall positively correlated when considering the whole time course. This suggests that for each type of event occurring in a brain structure, it is possible to associate information routing pathways that links the activity of this structure to the overall brain activity. Interestingly, the relationship of LFP rhythms (excluding SPW-R complexes) to brain-wide metabolic changes share common features in Hp and LGN structures, higher frequencies being related to cortical activation and lower ones to thalamic activation. In addition, lower frequency events in both recorded structures were associated to metabolic changes in cortex prior to the onset of the event, suggesting that these events may be triggered by other events happening earlier in the neocortex.

In contrast to their effect on cortex, the relationship between the frequency of LFP events and the underlying activity in the structure in which LFPs are recorded is more structure specific. While previous work in the visual cortex has shown that high frequency oscillation was reliably

associated to spiking as well as BOLD signal increases (Murayama et al., 2010, Logothetis et al., 2001), we observe a similar relationship in Hippocampus but not in LGN. In contrast, low frequency events (below 25Hz) are the one associated with an increased activity. This finding is supported by experimental studies reporting a decrease in LGN beta band power (20-40Hz) during visual stimulation while lower frequencies (7-15Hz) were increasing (Bastos et al., 2014), suggesting information processing in LGN is associated with increased lower frequency activity. In principle, possible explanations for such discrepancies of the relationship between LFP, fMRI and population spiking activity pertaining to the microcircuit anatomical and functional organization have been reviewed in Logothetis (2008).

The nature of the higher frequency events (above 25Hz) that we detect in LGN, as well as thalamus activating lower frequency events in CA1, remains elusive and requires further experimental and modeling studies. All these events have in common to be associated with weak local BOLD and spiking activity, suggesting that they correspond to mainly subthreshold mechanisms. This subthreshold activity might be strongly influenced by synaptic and neuromodulatory inputs for other brain structures. For example low frequency activity in hippocampus might reflect the synaptic filtering of neocortical delta and spindle oscillations. In support of this view, delta and spindles are associated to an electroencephalographic pattern, the K-complex, associated to neocortical down-states (Cash et al., 2009) and thereby justifying the neocortical down-regulation associated to these events. Connecting the occurrence of these events to brain state fluctuations may also explain the cortical pre-activation observed in our NET-fMRI results and the increase probability of ripples to occur several seconds prior to low frequency events. Indeed, ripples reportedly occur more frequently during down- to up-state transitions during slow-wave sleep (Battaglia et al., 2004) while lower frequencies would then be associated to the following up- to down-state transitions. Overall, these results illustrate how our approach to extract neural events allows a detailed study of large scale network interactions and their largely unexplored underlying mechanisms concurrently happening in multiple brain structures.

Materials and methods

Surgical procedures, electrophysiology and fMRI recordings

Experimental and surgical procedures have been detailed in a previous study (Logothetis et al., 2012). In summary, a total of 24 recording sessions were carried out in 4 anesthetized male rhesus monkeys (Macaca mulatta). Head holders and recording chambers were located stereotaxically based on high-resolution anatomical MRI scans. Hippocampal recordings were conducted in the anterior part of the hippocampus in the right hemisphere of each animal. Additionally, thalamic recording were recorded in the Lateral Geniculate Nucleus (LGN) in some animals in the same hemisphere. All recording hardware, including the electrodes and amplifiers for simultaneous fMRI and multi-site electrophysiology recordings, was developed at the Max Planck Institute for Biological Cybernetics. Multi-contact recordings were performed

around the pyramidal layer of the hippocampal CA1 subfield (8 to 14 mm anterior of the interaural line). Fine adjustment of the recording electrode was achieved by intermediate MRI anatomical scans. Functional imaging was carried out in a vertical 4.7 Tesla scanner, in which each animal was positioned in a custom-made chair. Typically, 22 axial slices were acquired, covering the entire brain. BOLD activity was acquired at a resolution of 2 seconds. During all experiments, anesthesia was maintained with remifentanil (0.5-2 µg/kg/min) in combination with a fast-acting paralytic mivacurium chloride (5-7 mg/kg/h), only mildly affecting the magnitude and time course of neural and vascular responses (Logothetis et al., 2012, Goense and Logothetis, 2008). All experimental and surgical procedures were approved by the local authorities (Regierungspraesidium, Tübingen Referat 35, Veteriärwesen) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

Processing and Analysis of Neural Data

Analyses of electrophysiology and fMRI data were performed using MATLAB (The MathWorks). LFP data was cleaned from electromagnetic gradient artifacts, down-sampled and low pass filtered at 660 Hz (Logothetis et al., 2012). In addition to anatomical criteria, we selected in each recording session one hippocampal recording tip belonging to stratum pyramidale (PL) based on the relative power in the ripple band (100-200 Hz). In each session containing thalamic recordings, we chose the recording tip in LGN having the maximum multi-unit response to a polar checkerboard stimulus.

Overlap-add decomposition and short term Fourier analysis

Spectrogram analysis was performed on the LFP time series by tapering the signal using sliding 600ms Hanning windows with 50% overlap. Tapering windows were normalized such that at any time point, the sum of overlapping tapering windows is one, which ensures an exact reconstruction of the time course of the original signal based on short-term Fourier transform values using the overlap-add technique. The tapered signal was then Fourier transformed using the FFT algorithm and squared to get the spectrogram values for this window.

Non-negative matrix factorization

The IS-NMF algorithm was applied to the spectrogram matrix obtained according to the previous paragraph. The factorization was initialized by drawing spectral profiles at random using uniformly distributed coefficient on the unit interval, and the corresponding time contributions were initialized using least square regression (see Supplementary information). The NMF decomposition was then optimized using the multiplicative update algorithm (Sra and Dhillon, 2005) and the stability of the solution was enforced using an iterative bootstrap approach to find a good initialization of the components (see Supplementary Methods). To determine an optimal number of components, the robustness of the result was assessed by cross-validation using two subsets of experimental recordings (with 50% overlap): the spectral profiles obtained by the approach using these two subsets were matched and compared using the cosine

similarity. The average cosine similarity between the matched spectral profiles resulting from each subsets was used as an indicator of the robustness of the result.

Dictionary learning

The principle of this approach is illustrated in Figure 3c. After initialization by random patterns, the algorithm alternates between a Matching Pursuit step, which detects iteratively the event locations by finding the time points where the component time course has maximum similarity with one of the dictionary patterns, and an Singular Value Decomposition (SVD) step in which the dictionary patterns are updated by computing the singular vector of largest singular value of the matrix gathering all peri-event time windows of the component's time series at locations where a given pattern was detected. Good convergence of the algorithm was reached after 30 iterations. The shift-invariant dictionary learning procedure has two free parameters: the number of dictionary patterns and the total number of events to detect. These parameters were selected using a cross-validation procedure described in Supplementary Methods. The performance of the approach was also evaluated in Supplementary Results.

Clustering of Neural Events

Neural events where clustered according to the peak frequency of the Fourier transform of their associated dictionary pattern. We excluded from this analysis events with residual MRI artifact contamination, that we detected using their peri-event LFP spectrogram: the Fourier transform of the spectrogram time course averaged over higher frequencies (133-300 Hz) and the corresponding event was excluded whenever the magnitude absolute Fourier transform at its peak frequency was larger than three times the average magnitude of the Fourier transform at the frequencies surrounding the peak frequency. Events with their peak frequency within [49-51 Hz], corresponding to line noise were also excluded from this analysis. We defined a distance between two events as the absolute difference between the two frequencies, normalized by their maximum. We subtract this distance to 1 to define a similarity measure which is used to cluster events using a graph clustering procedure (Shi and Malik, 2000). To avoid outliers to bias the clustering procedure, events having less than 2 neighbors within a distance of less than .2, were also excluded from the procedure. This approach was repeated for a number of clusters ranging from 2 to 12, the quality of the clustering for each number of clusters was assessed using the silhouette index based again on the same distance metric. The silhouette index quantifies how grouped the clusters are within each cluster with respect to the whole dataset (Rousseeuw, 1987). We evaluated the graph of the silhouette index against the number of clusters, and selected the smallest number of clusters achieving a local maximum of this graph.

Neural Event triggered measures

To characterize local and large scale properties of neural events, we computed two quantities: NET-fMRI and spectrogram, based on the peri-event time course of the fMRI and LFP signal respectively. NET-fMRI maps were computed by averaging the peri-event time course ranging from 20 second before event onset until 20 second after event onset. Peri-event signals were preliminarily detrended. Spectrograms were estimated by first computing the continuous wavelet

transform with a complex Morlet wavelet on a 2 second interval centered on each event onset. The spectrogram values were obtained by averaging the squared modulus of the wavelet transform across events. Both quantities were normalized by computing a z-score with respect to randomized events in the following way. Randomized event were generated on each 10 min experiment by taking the original event time stamps, compute the empirical distribution of inter-event intervals, and randomly drawing with replacement from this distribution the same number of intervals as the original. Time stamps of randomized events were generated by picking at random and initial point in the first 20s of the recording and iteratively adding the randomized inter-event intervals.

Descriptive statistics and tests

In figures describing statistics with boxplots, on each box, the top and bottom are the 25th and 75th percentiles of the samples, respectively; the line in the middle of each box indicates the sample median; the dashed lines extending below and above each box are drawn from the ends of the interquartile ranges to the furthest observation (extreme points not considered as outliers); crosses (if any) in the diagrams indicate outliers of the samples. A data point is considered as an outlier whenever it is larger than Q3+1.5*(Q3-Q1) or smaller than Q1-1.5*(Q3-Q1), Q1 and Q3 indicating the 25th and 75th percentiles, respectively. If not specified, statistical tests presented in text and figures are two-sided.

References

ABELES, M. 1982. *Local cortical circuits: An Electrophysiological study*, Springer.

ABEYSURIYA, R. G., RENNIE, C. J. & ROBINSON, P. A. 2014a. Prediction and verification of nonlinear sleep spindle harmonic oscillations. *J Theor Biol,* 344**,** 70-7.

ABEYSURIYA, R. G., RENNIE, C. J., ROBINSON, P. A. & KIM, J. W. 2014b. Experimental observation of a theoretically predicted nonlinear sleep spindle harmonic in human EEG. *Clin Neurophysiol,* 125**,** 2016-23.

ALLEN, J. B. 1977. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on,* 25**,** 235-238.

BASTOS, A. M., BRIGGS, F., ALITTO, H. J., MANGUN, G. R. & USREY, W. M. 2014. Simultaneous recordings from the primary visual cortex and lateral geniculate nucleus reveal rhythmic interactions and a cortical source for gamma-band oscillations. *J Neurosci,* 34**,** 7639-44.

BATTAGLIA, F. P., SUTHERLAND, G. R. & MCNAUGHTON, B. L. 2004. Hippocampal sharp wave bursts coincide with neocortical up-state transitions. *Learning & Memory,* 11**,** 697-704.

BELITSKI, A., GRETTON, A., MAGRI, C., MURAYAMA, Y., MONTEMURRO, M. A., LOGOTHETIS, N. K. & PANZERI, S. 2008. Low-Frequency Local Field Potentials and Spikes in Primary Visual Cortex Convey Independent Visual Information. *J Neurosci,* 28**,** 5696-5709.

BREAKSPEAR, M., ROBERTS, J. A., TERRY, J. R., RODRIGUES, S., MAHANT, N. & ROBINSON, P. A. 2006. A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. *Cereb Cortex,* 16**,** 1296-313.

BRILLINGER, D. R. 1976. Estimation of 2nd-Order Intensities of a Bivariate Stationary Point Process. *Journal of the Royal Statistical Society Series B-Methodological,* 38**,** 60-66.

BRUNEL, N. & WANG, X.-J. 2003. What Determines the Frequency of Fast Network Oscillations With Irregular Neural Discharges? I. Synaptic Dynamics and Excitation-Inhibition Balance. *Journal of Neurophysiology,* 90**,** 415-430.

BUZSAKI, G. 1986. Hippocampal sharp waves: their origin and significance. *Brain Res,* 398**,** 242-52.

BUZSÁKI, G. 2006. *Rhythms of the Brain*, Oxford University Press.

BUZSÁKI, G. & DRAGUHN, A. 2004. Neuronal oscillations in cortical networks. *Science,* 304**,** 1926-1929.

BUZSAKI, G., HORVATH, Z., URIOSTE, R., HETKE, J. & WISE, K. 1992. High-frequency network oscillation in the hippocampus. *Science,* 256**,** 1025-7.

CANOLTY, R. T., EDWARDS, E., DALAL, S. S., SOLTANI, M., NAGARAJAN, S. S., KIRSCH, H. E., BERGER, M. S., BARBARO, N. M. & KNIGHT, R. T. 2006. High gamma power is phase-locked to theta oscillations in human neocortex. *Science,* 313**,** 1626-1628.

CASH, S. S., HALGREN, E., DEHGHANI, N., ROSSETTI, A. O., THESEN, T., WANG, C., DEVINSKY, O., KUZNIECKY, R., DOYLE, W., MADSEN, J. R., BROMFIELD, E., EROSS, L., HALASZ, P., KARMOS, G., CSERCSA, R., WITTNER, L. & ULBERT, I. 2009. The human K-complex represents an isolated cortical down-state. *Science,* 324**,** 1084-7.

CONTRERAS, D., TIMOFEEV, I. & STERIADE, M. 1996. Mechanisms of long-lasting hyperpolarizations underlying slow sleep oscillations in cat corticothalamic networks. *The Journal of Physiology,* 494**,** 251-264.

DESGRAUPES, B. 2013. Clustering Indices.

EINEVOLL, G. T., PETTERSEN, K. H., DEVOR, A., ULBERT, I., HALGREN, E. & DALE, A. M. 2007. Laminar population analysis: estimating firing rates and evoked synaptic activity from multielectrode recordings in rat barrel cortex. *Journal of neurophysiology,* 97**,** 2174-2190.

FÉVOTTE, C., BERTIN, N. & DURRIEU, J.-L. 2009. Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis. *Neural Computation,* 21**,** 793-830.

GOENSE, J. & LOGOTHETIS, N. K. 2008. Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology,* 18**,** 631-640.

HE, B. J., ZEMPEL, J. M., SNYDER, A. Z. & RAICHLE, M. E. 2010. The temporal structures and functional significance of scale-free brain activity. *Neuron,* 66**,** 353-369.

HONEY, C. J., KÖTTER, R., BREAKSPEAR, M. & SPORNS, O. 2007. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Science USA,* 104**,** 10240-10245.

JI, D. Y. & WILSON, M. A. 2007. Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience,* 10**,** 100-107.

LOGOTHETIS, N. K. 2008. What we can do and what we can not do with fMRI. *Nature,* 453**,** 869-878.

LOGOTHETIS, N. K., ESCHENKO, O., MURAYAMA, Y., AUGATH, M., STEUDEL, T., EVRARD, H. C., BESSERVE, M. & OELTERMANN, A. 2012. Hippocampal-cortical interaction during periods of subcortical silence. *Nature,* 491**,** 547-53.

LOGOTHETIS, N. K., PAULS, J., AUGATH, M., TRINATH, T. & OELTERMANN, A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature,* 412**,** 150-157.

LOGOTHETIS, N. K. & WANDELL, B. A. 2004. Interpreting the BOLD signal. *Annu Rev Physiol,* 66**,** 735-69.

MAGRI, C., MAZZONI, A., LOGOTHETIS, N. K. & PANZERI, S. 2012. Optimal band separation of extracellular field potentials. *Journal of neuroscience methods,* 210**,** 66-78.

MAILHÉ, B., LESAGE, S., GRIBONVAL, R., BIMBOT, F., VANDERGHEYNST, P. & OTHERS. 2008. Shift-invariant dictionary learning for sparse representations: extending K-SVD. 16th EUropean SIgnal Processing COnference (EUSIPCO'08), 2008 2008.

MAKAROV, V. A., MAKAROVA, J. & HERRERAS, O. 2010. Disentanglement of local field potential sources by independent component analysis. *Journal of computational neuroscience,* 29**,** 445-457.

MARCUS, G., MARBLESTONE, A. & DEAN, T. 2014. Neuroscience. The atoms of neural computation. *Science,* 346**,** 551-2.

MITZDORF, U. 1985. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol Rev,* 65**,** 37-100.

MONTEMURRO, M. A., RASCH, M. J., MURAYAMA, Y., LOGOTHETIS, N. K. & PANZERI, S. 2008. Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Current Biology,* 18**,** 375-380.

MURAYAMA, Y., BIE$MANN, F., MEINECKE, F. C., MÜLLER, K.-R., AUGATH, M., OELTERMANN, A. & LOGOTHETIS, N. K. 2010. Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA. *Magnetic resonance imaging,* 28**,** 1095-1103.

NOVIKOV, E., NOVIKOV, A., SHANNAHOFF-KHALSA, D., SCHWARTZ, B. & WRIGHT, J. 1997. Scale-similar activity in the brain. *Phys. Rev. E,* 56**,** R2387-R2389.

RAMIREZ-VILLEGAS, J. F., LOGOTHETIS, N. K. & BESSERVE, M. 2015. Diversity of sharp-wave-ripple LFP signatures reveals differentiated brain-wide dynamical events. *Proc Natl Acad Sci U S A,* 112**,** E6379-87.

ROBINSON, P. A., RENNIE, C. J. & ROWE, D. L. 2002. Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. *Phys Rev E Stat Nonlin Soft Matter Phys,* 65**,** 041924.

ROOPUN, A. K., KRAMER, M. A., CARRACEDO, L. M., KAISER, M., DAVIES, C. H., TRAUB, R. D., KOPELL, N. J. & WHITTINGTON, M. A. 2008. Temporal Interactions between Cortical Rhythms. *Frontiers in Neuroscience,* 2**,** 145-154.

ROUSSEEUW, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics,* 20**,** 53-65.

SCHOMBURG, E. W., FERNÁNDEZ-RUIZ, A., MIZUSEKI, K., BERÉNYI, A., ANASTASSIOU, C. A., KOCH, C. & BUZSÁKI, G. 2014. Theta Phase Segregation of Input-Specific Gamma Patterns in Entorhinal-Hippocampal Networks. *Neuron*.

SEUNG, D. & LEE, L. 2001. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems,* 13**,** 556-562.

SHI, J. & MALIK, J. 2000. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 22**,** 888-905.

SMARAGDIS, P., FEVOTTE, C., MYSORE, G., MOHAMMADIHA, N. & HOFFMAN, M. 2014. Static and Dynamic Source Separation Using Nonnegative Factorizations: A unified view. *Signal Processing Magazine, IEEE,* 31**,** 66-75.

SRA, S. & DHILLON, I. S. Generalized nonnegative matrix approximations with Bregman divergences. Advances in neural information processing systems, 2005 2005. 283-290.

THIAGARAJAN, T. C., LEBEDEV, M. A., NICOLELIS, M. A. & PLENZ, D. 2010. Coherence potentials: loss-less, all-or-none network events in the cortex. *PLoS biology,* 8**,** e1000278-e1000278.

VARELA, C. 2014. Thalamic neuromodulation and its implications for executive networks. *Front Neural Circuits,* 8**,** 69.

WANG, X. J. 2010. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological reviews,* 90**,** 1195-268.

YLINEN, A., BRAGIN, A., NÁDASDY, Z., JANDÓ, G., SZABO, I., SIK, A. & BUZSÁKI, G. 1995. Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *The Journal of neuroscience,* 15**,** 30-46.

Figures

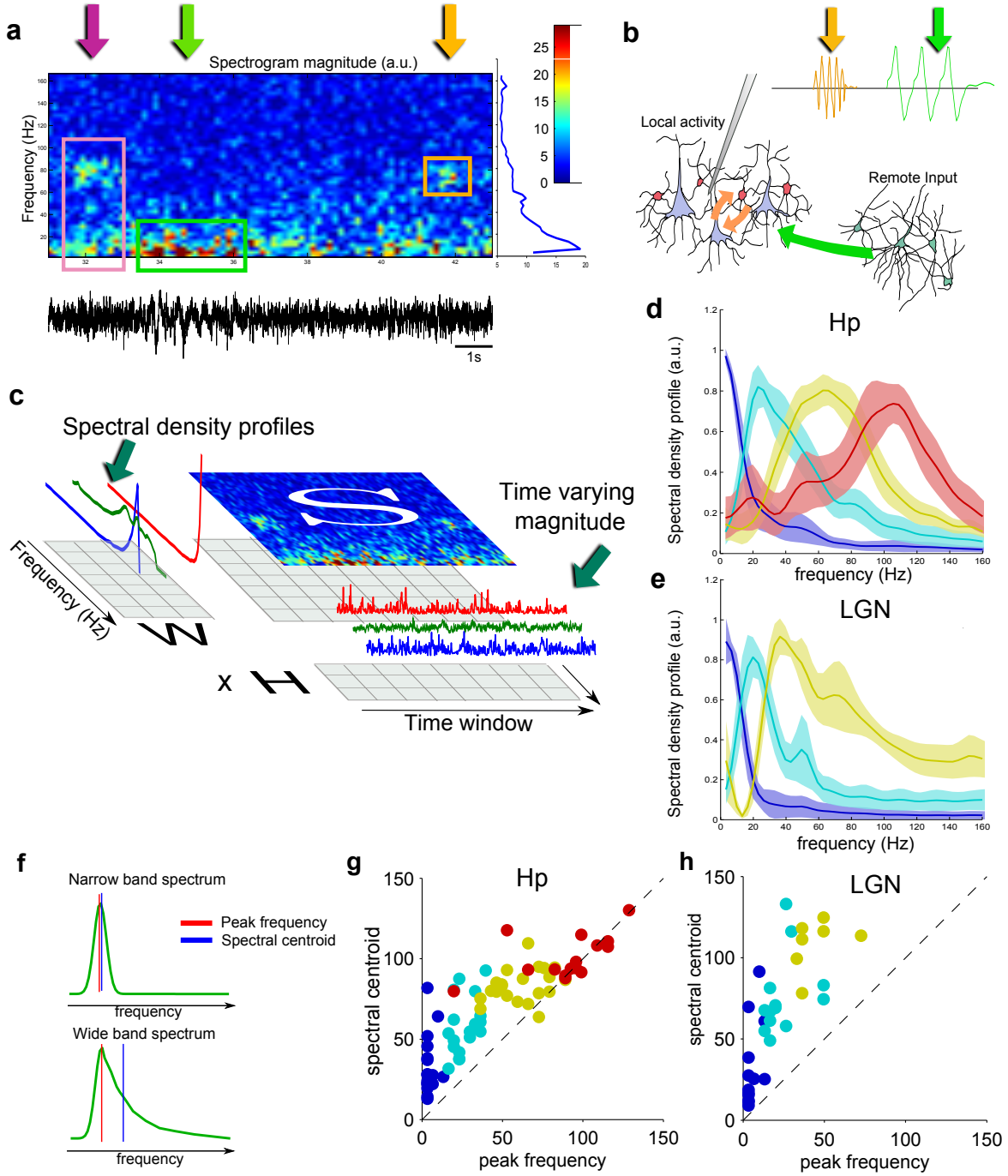


Figure 1: **Principle of neural event analysis.** (a) Example LFP from the pyramidal layer from monkey CA1 (bottom black trace) and corresponding spectrogram (top) computed with STFT. Rectangles indicate transient spectral power increases, see text. (b) Schematic representation of LFP events (top) and putative underlying neural populations (bottom). Two local groups of neurons, in gray and red interact through recurrent interactions (orange arrows) and receive input (green arrow) from a third population (in green) in a remote brain structure. (c) Principle of NMF applied to the spectrogram matrix **S** shown in panel a. **S** is approximated by the product of two matrices with non-negative coefficients: **W**, gathering the spectral profiles of 3 components, and **H**, gathering the time dependent contribution of each component to the LFP spectrogram. (d) Normalized average spectral profiles resulting from the clustering of NMF results for LFP recordings in Hp, area indicates the standard error within each cluster. (e) Same as panel d for LFP recordings in LGN. Additional analysis can be found in Supplementary Fig. 2-3. (f) Schematic representation of two types of power spectral density profiles together with their the location of their peak frequency (in red) and spectral centroid (in blue). (g) Spectral centroid against peak frequency for all hippocampal spectral profiles resulting from NMF analysis. Color indicates cluster membership as in d. (h) Spectral centroid against peak frequency for all LGN spectral profiles cluster. Color indicates cluster membership as in e.

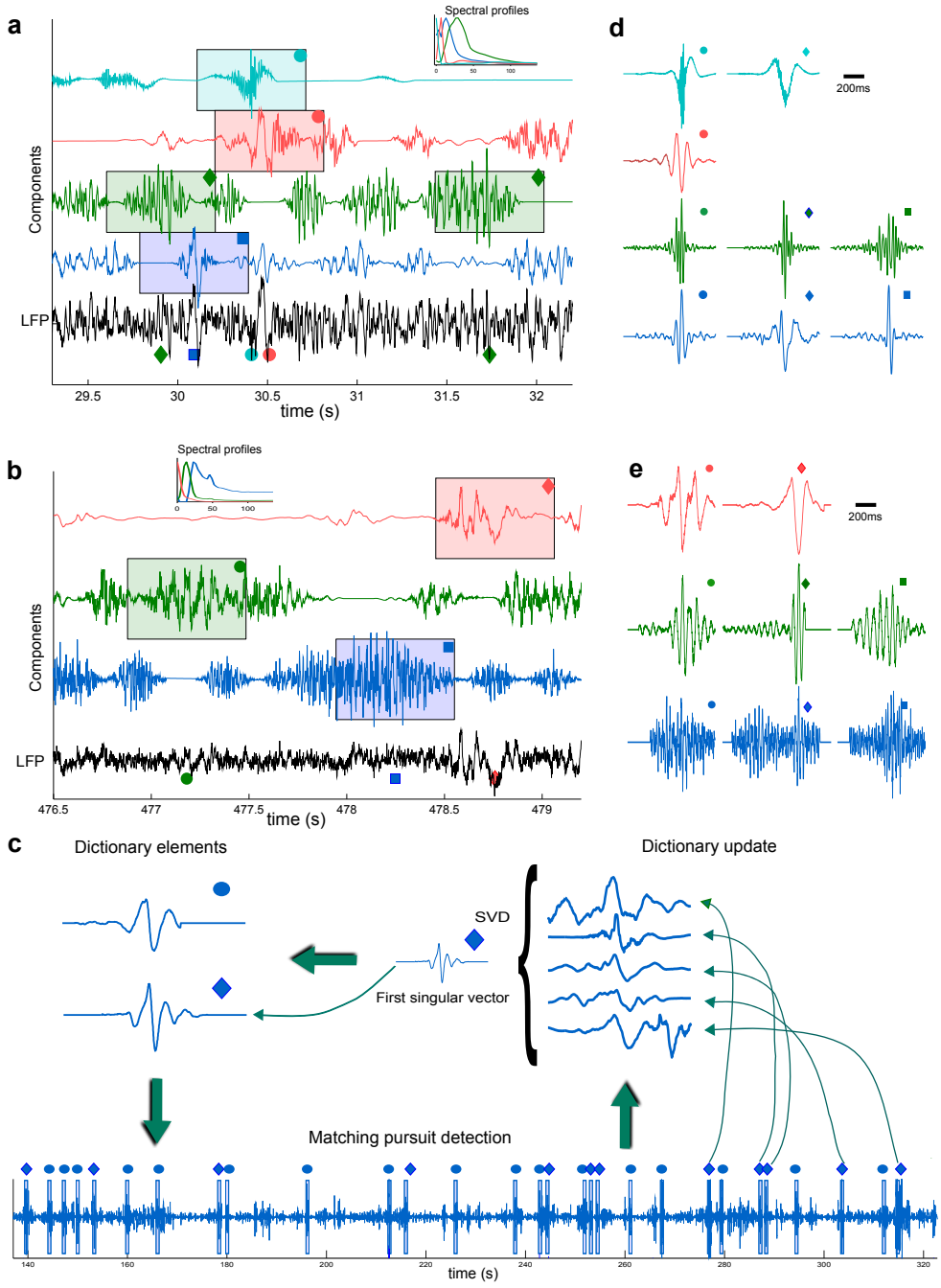


Figure 2: **Time course of neural events.** (a) Example of dynamical components extracted from CA1 LFP (bottom black trace). Rectangles indicate the location of detected events, the marker in the upper right-hand corner indicates the corresponding dictionary element in panel d. (b) Same as panel a for LFP in LGN. Rectangles indicate the location of detected events, markers indicate dictionary elements in panel e. (c) Principle of shift-invariant dictionary learning. The algorithm alternates between a Matching Pursuit step, which detects the time points where the component time course have maximum similarity with one of the dictionary elements, and an SVD step in which the dictionary patterns are updated using the SVD of the matrix gathering perievent time windows detected during the previous step. (d) Dictionary patterns learned from the dynamical components in panel b. (e) Same as (d) for components in panel c.

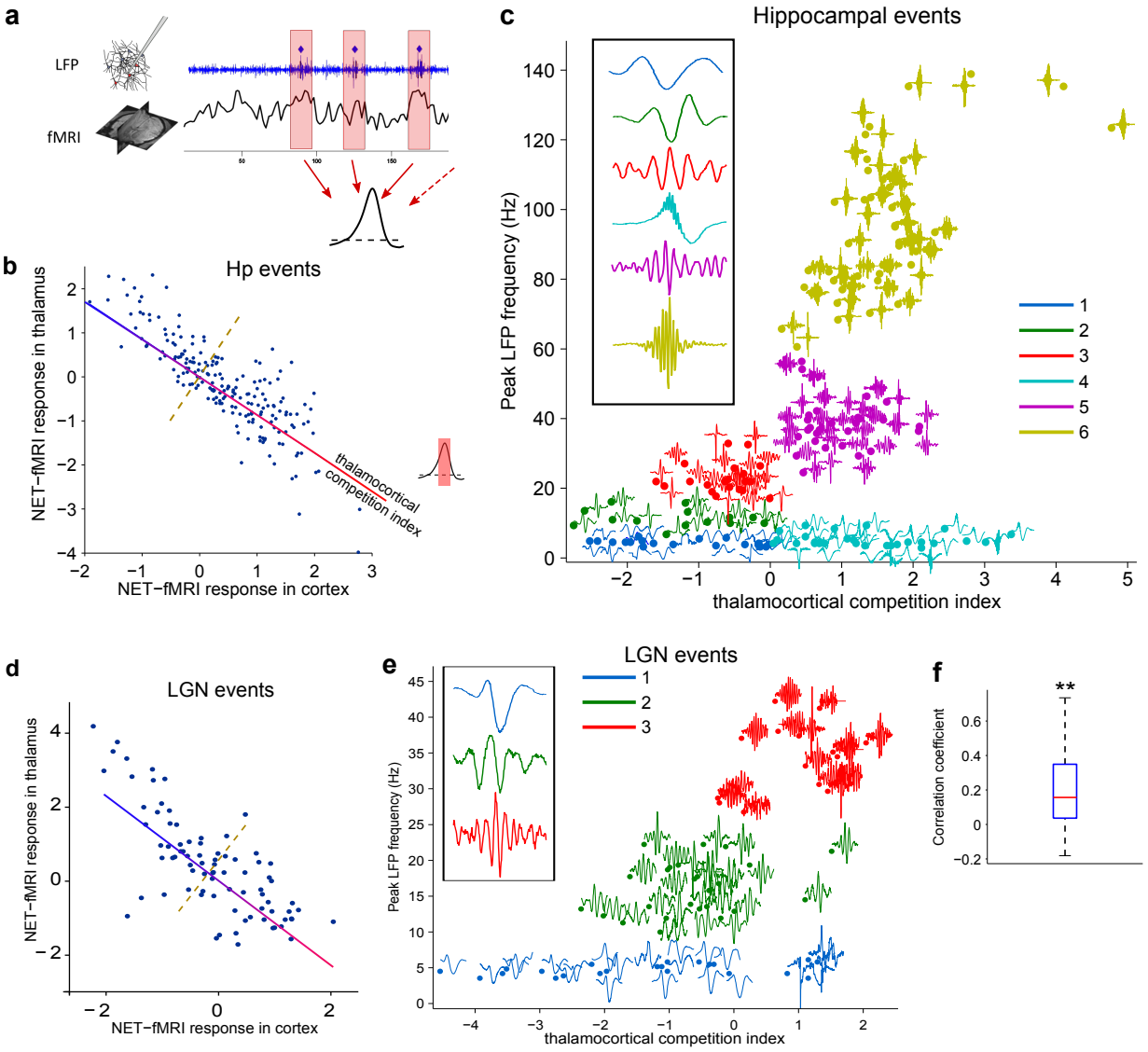


Figure 3: **Local versus brain-wide properties of neural events.** (a) Schematic representation of the principle of NET-fMRI analysis: the peri-event fMRI traces are averaged across events associated to a given dictionary pattern. (b) Raster plot of the average NET-fMRI response in thalamus and cortex associated to each detected Hp event. As illustrated by the inset, NET-fMRI response is averaged over the interval were it reaches half the maximum response. Correlation between the two variables is significant ($\rho = -.839$, Spearman, $p < 10^{-7}$, n=171). The solid line indicates the corresponding fitted linear regression function. (c) Raster plot of the peak frequency of dictionary patterns against thalamocortical competition index (TCI) for Hp events, insets represent magnified examples of dictionary patterns for each cluster. Colors indicates cluster indentity. (d) Same as panel b for LGN events.Correlation between the two variables is significant (($\rho = -.727$, Spearman, $p < 10^{-7}$, n=80)). (e) Same as panel c for LGN events. (f) Distribution of correlation coefficient between thalamic and cortical fMRI activity across sessions. (f) Distribution across sessions of average correlation between Thalamic and Cortical fMRI activity. Star indicate significantly positive median (Wilcoxon signed rank test, p<.01, n=20)

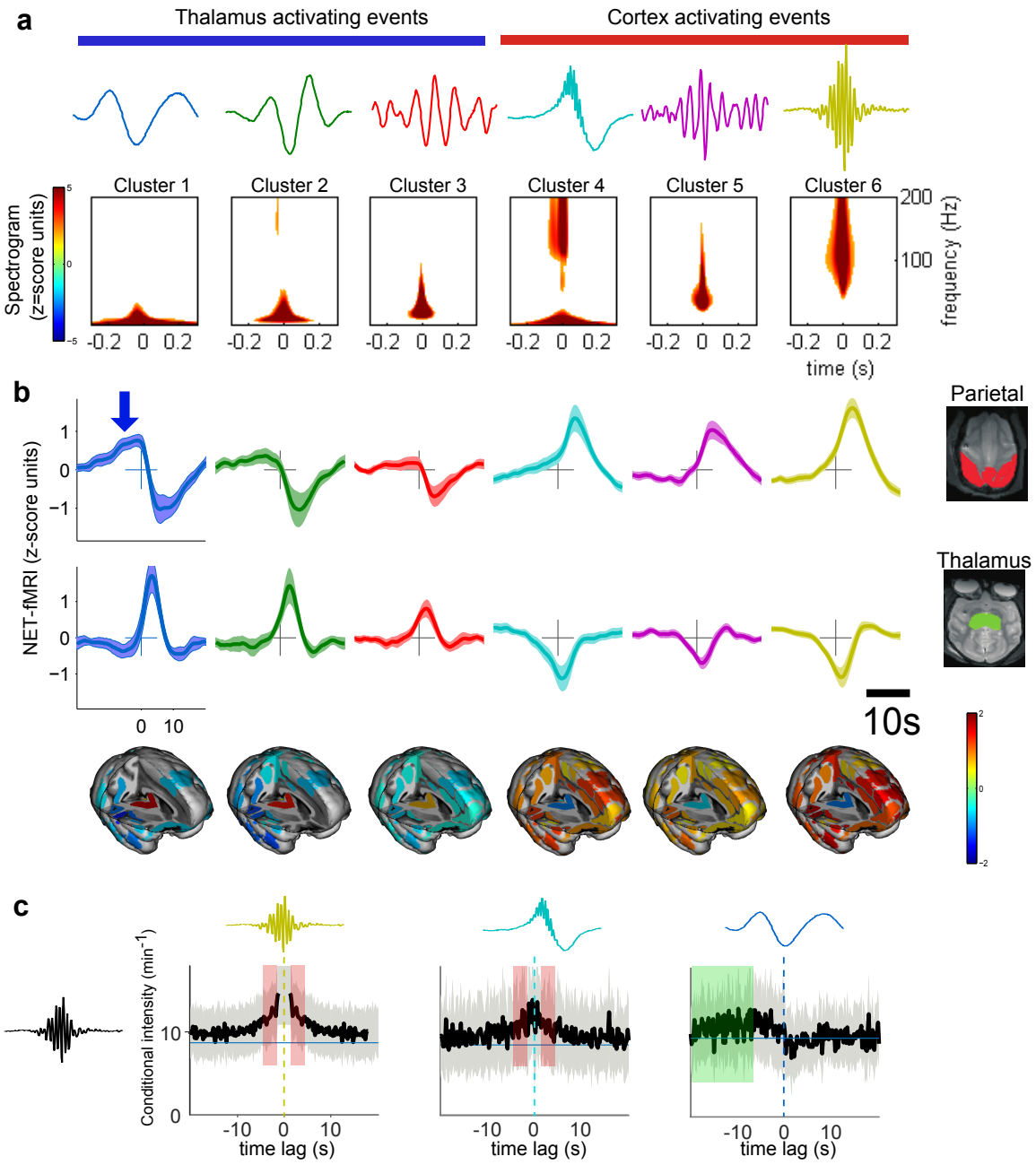


Figure 4: **Local and brain-wide physiological properties of hippocampal neural events.** (a) Average event triggered time-frequency maps of the 6 clusters of hippocampal events (Z-score with respect to randomized event onsets). Traces at the top represent example dictionary patterns of each cluster. (b) Average NET-fMRI response of the 6 clusters of hippocampal events for selected structures (Z-score with respect to randomized event onsets). Bottom insets represent the trimensional mapping of the amplitude of this response at t=3s overlaid on a template macaque brain.

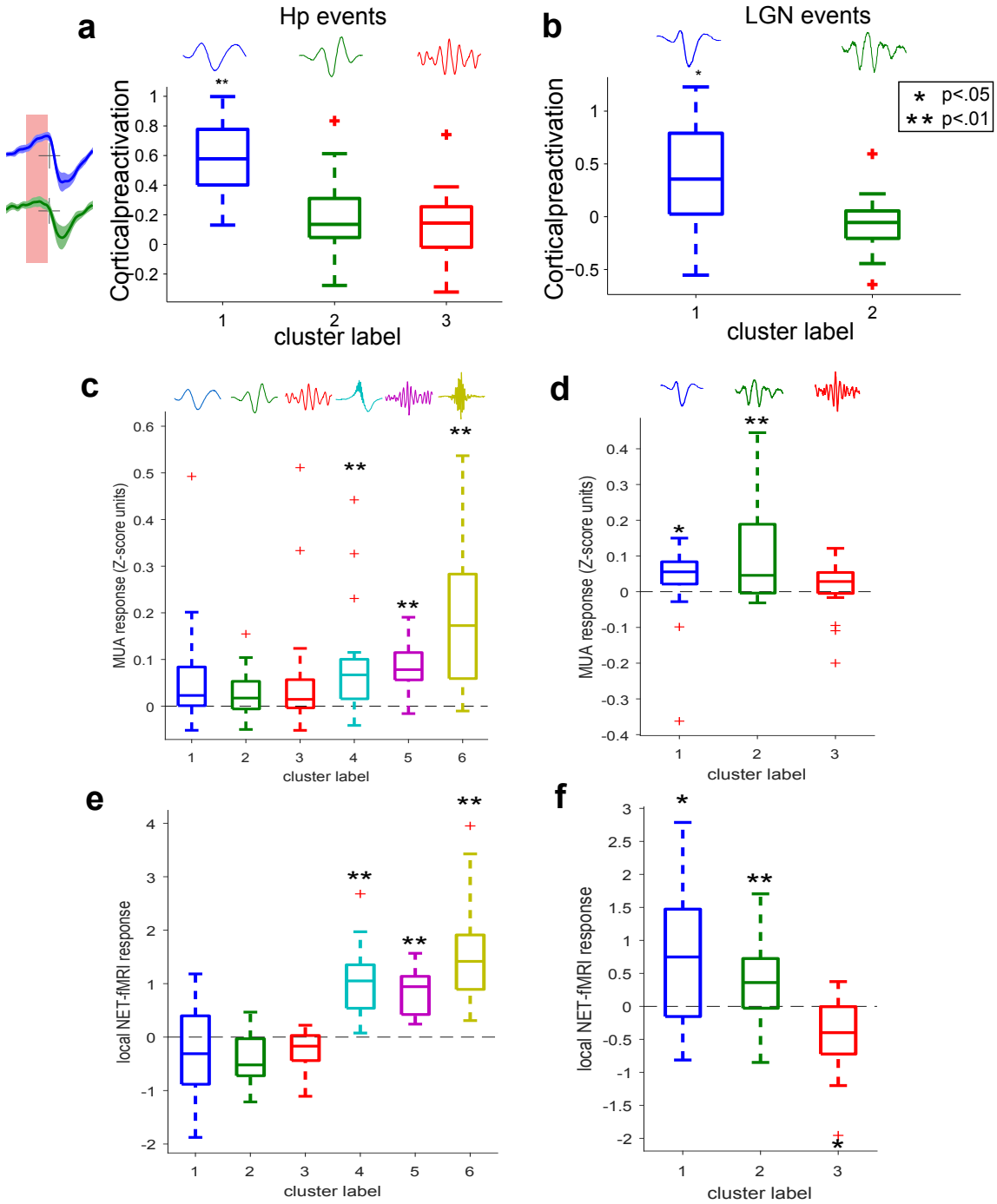


Figure 5: **Comparison of LGN and Hp event properties.** (a) Distribution of average NET-fMRI response in cortex prior to the event onset (during the pre-activation period, see text) for thalamus activating clusters. Stars indicate significantly positive values (Wilcoxon signed rank test; p<.01, Bonferoni corrected, from left to right: n=18,14, 21). Traces at the top represent example dictionary patterns of each cluster. (b) Same as panel (a) for LGN events. Stars indicate significantly non-zero median values (Wilcoxon signed rank test; p<.01, Bonferoni corrected, from left to right: n=22, 25). (c) Distribution of average MUA changes in a 400ms peri-event time window (Z-scored with respect to randomized events) for each Hp event cluster (outliers at 1.23 and .75 for cluster label 4 and .89 for cluster label 6 are out of the axis range). Stars indicate significantly non-zero median values (Wilcoxon signed rank test; p<.01, Bonferoni corrected, from left to right: n=18, 14, 21, 27, 26, 38). (d) Same as (d) for LGN events. Stars indicate significantly non-zero median values (Wilcoxon signed rank test; p<.01, Bonferoni corrected, from left to right: n=22, 25, 16). (e) Distribution of local (Hippocampal) NET-fMRI changes for each Hp event clusters. (f) Same as (e) for LGN events (NET-fMRI changes in LGN).

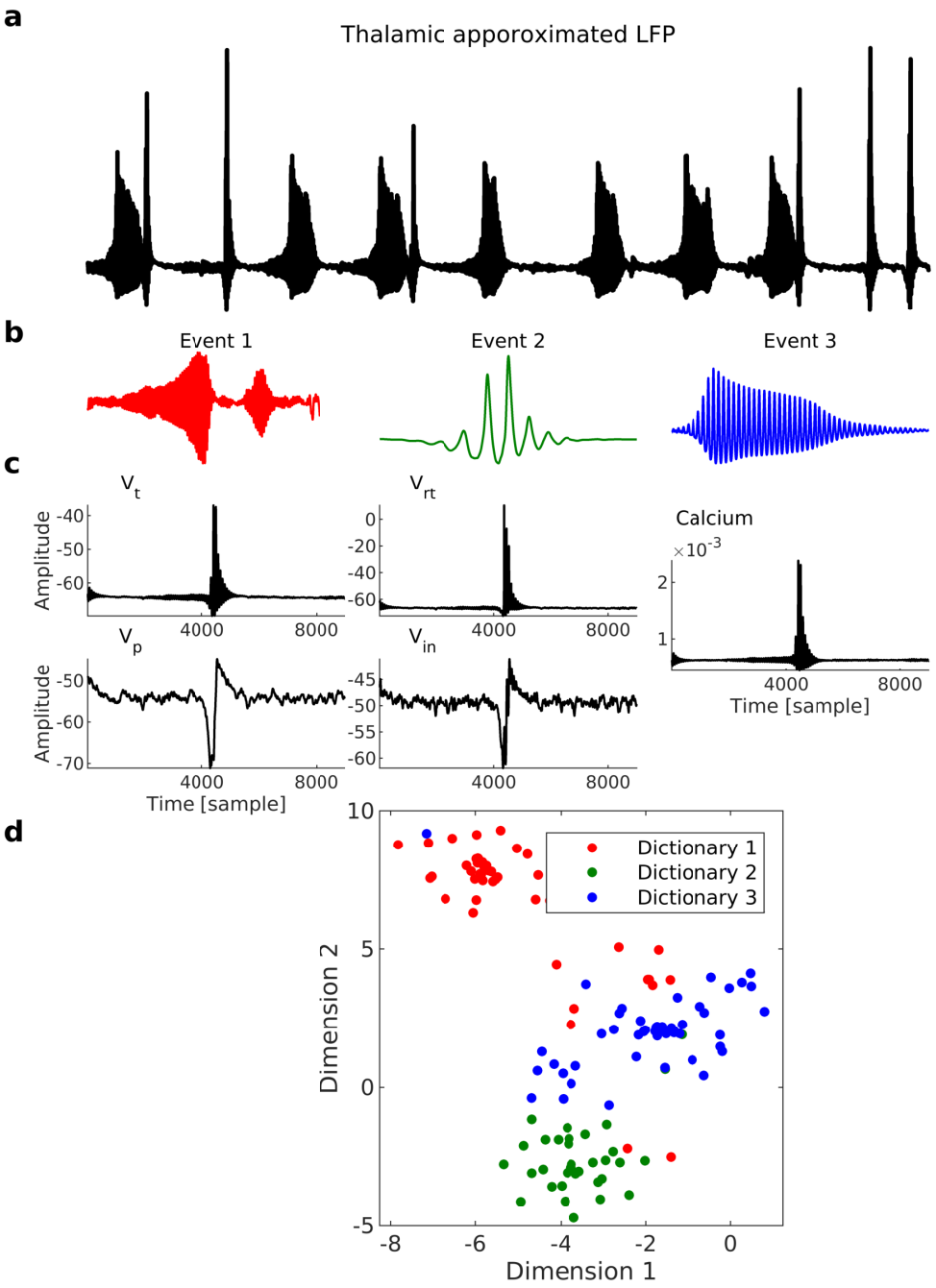


Figure 6: **Neural events in a thalamocoritical simulation.** (a) An exemplary trace for approximated LFP from thalamic module of the thalamcortical simulation. (b) Three Identified events in the approximated LFP. (c) Exemplary trace of cellular variables of the thalamcortical (membrane potential of all 4 populations and the calcium current, see the main text for the details). (d) Representation of cellular dynamics in the vicinity of the identified events (in (b)), in a 2-D subspace (based on tSNE dimensionality reduction). Colors are matching the colors of identified neural events in (b).

The complex spectral structure of transient LFPs reveals subtle aspects of network coordination across scales and structures.

Supplementary information

Michel Besserve[1,2], Shervin Safavi[1], Bernhard Schölkopf[2] and Nikos K. Logothetis[1,3]

[1]Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany.

[2]Department of Empirical Inference, Max Planck Institute for Intelligent Systems, Spemannstrasse 38, 72076 Tübingen, Germany.

[3]Centre for Imaging Sciences, Biomedical Imaging Institute, The University of Manchester, Manchester M13 9PT, United Kingdom.

1 Supplementary methods

1.1 Dynamical component decomposition of LFP signals

Using a latent variable model and an associated non-negative matrix factorization (NMF) procedure, it is possible to decompose the single channel LFP time course into a set of dynamical components. We describe in this section the model, the NMF algorithm and the dynamical component estimation.

1.1.1 Gaussian Composite Model

The Gaussian Composite Model (GCM) [Smaragdis et al., 2014] models the input signal $x(t)$ as a mixture of K latent components:

$$X(t) = \sum_{k=1}^{K} C_k(t).$$

The vector of coefficients of the Short Term Fourier Transform (SFT) of X on a given time window n of length $2T+1$ centered at time t_n is denoted by:

$$\mathbf{x}_n = (x_{f,n}) = SFT(\{X(t_n+m)\}_{m=-T,\dots,T}).$$

By linearity of the Fourier transform, the mixture can be written in the Fourier domain as:

$$\boldsymbol{x}_n = \sum_{k=1}^{K} \boldsymbol{c}_{n,k}, \quad \boldsymbol{c}_{n,k} = SFT(\{C_k(t_n+m)\}_{m=-T,\dots,T}).$$

The Fourier coefficients of each component at each non-negative frequency are assumed to be independent normally distributed random variables such that their joint distribution is a multivariate complex Gaussian with a diagonal covariance matrix:

$$\boldsymbol{c}_{k,n} \sim \mathcal{N}_c(0, h_{k,n} diag(\boldsymbol{w}_k)).$$

In this expression, the vector $\boldsymbol{\omega}_k$ represents the frequency profile of the power spectral density of component k, while $h_{k,n}$ is the time varying magnitude of component k for window n. In this way, the input signal is modeled as a mixture of several frequency components with a fixed frequency profile and a time varying contribution to each time window.

As a consequence of this Gaussian model, the distribution of Fourier coefficient of the input signal can be expressed by marginalizing the latent components, such that for each frequency f and time window n:

$$x_{f,n} \sim \mathcal{N}_c(0, \sum_k h_{k,n} w_{f,k}).$$

As a consequence, the power spectrogram is exponentially distributed with mean:

$$\overline{s_{f,n}} = \mathrm{var}(x_{f,n}) = \sum_k h_{k,n} w_{f,k} = (\mathbf{WH})_{f,n},$$

where $\mathbf{W}$ and $\mathbf{H}$ are matrices built from the coefficients $w_{f,k}$ and $h_{k,n}$ respectively.

Given the observed samples of $\boldsymbol{x}$, the minus log-likelihood of the two matrices of parameters writes

$$C_{ML}(\mathbf{W}, \mathbf{H}) = -\sum_{f,n} \log \mathcal{N}_c(x_{f,n} \,|\, 0, \sum_k w_{f,k} h_{n,k}),$$

which can be rewritten up to an additive constant using the Itakura-Saito divergence $d_{IS}(x|y) \hat{=} \frac{x}{y} - \log \frac{x}{y} - 1$:

$$C_{ML}(\mathbf{W}, \mathbf{H}) =_C \sum_{f,n} d_{IS}(|x_{f,n}|^2 \,|\, \sum_k w_{fk} h_{nk}) \hat{=} d_{IS}(\mathbf{S}|\mathbf{WH}).$$

In the last expression, all matrices being non-negative, it appears that finding the maximum likelihood estimates of GCM parameters amounts to solve an approximate non-negative matrix factorization problem with a specific loss: the Itakura-Saito divergence (the divergence for a matrix is defined as the sum of entrywise divergences).

1.1.2 Bregman divergences

The Itakura-Saito divergence is a member of a parametric family of divergences, β-divergences, which are defined for positive scalars x and y as:

$$d_\beta(x|y) = \begin{cases} \frac{1}{\beta(\beta-1)}(x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}) & \beta \in \mathbb{R}\backslash\{0,1\} \\ x\log(\frac{x}{y}) + (y-x) & \beta = 1 \\ \frac{x}{y} - \log\frac{x}{y} - 1 & \beta = 0 \end{cases}$$

Interestingly, this family includes the squared Euclidean norm ($\beta = 2$), the Kullback–Leibler divergence (1), and the Itakura-Saito divergence ($\beta = 0$) as special cases. The non-negative factorization problem can be addressed for any of these divergences in a common framework explained in the next section. One property specific to the Itakura-Saito divergence is scale invariance: assume variables x and y are rescaled by a multiplicative positive factor s, it immediately follows from the definition that $d_{IS}(sx|sy) = d_{IS}(x|y)$. As a consequence, Itakura-Saito NMF is invariant to rescaling of the lines of the spectrogram matrix $\mathbf{S}$ associated to each frequency.

1.1.3 Non-negative matrix factorization

The non-negative factorization problem for a given Bregman divergence d_β can be written as:

$$\min_{\mathbf{W}\geq 0, \mathbf{H}\geq 0} d_\beta(\mathbf{S}|\mathbf{WH}).$$

The minimum of this objective is reached at a stationary point, which satisfies the Karush-Kuhn-Tucker (KKT) optimality conditions:

$$\mathbf{W} \odot \nabla_{\mathbf{W}} d_\beta(\mathbf{S}|\mathbf{WH}) = 0 \text{ and } \mathbf{H} \odot \nabla_{\mathbf{H}} d_\beta(\mathbf{S}|\mathbf{WH}) = 0.$$

Computing the expression of the gradient $\nabla_W d_\beta(\mathbf{V}|\mathbf{WH})$, we get for the first KKT condition:

$$\mathbf{W} \odot (\mathbf{W}^T(\mathbf{S} \odot (\mathbf{WH})^{\beta-2})) = \mathbf{W} \odot (\mathbf{W}^T(\mathbf{WH})^{\beta-1}).$$

and an equivalent expression in $\mathbf{H}$ for the second KKT condition.

The multiplicative update algorithm proposes to search for a stationary point satisfying this condition by using the fixed point iteration:

$$\mathbf{W}_{k+1} = \mathbf{W}_k \odot \frac{(\mathbf{S}\odot(\mathbf{WH})_k^{\beta-2})\mathbf{H}_k^T}{(\mathbf{WH})_k^{\beta-1}\mathbf{H}_k^T}.$$

We can easily verify from this last equation that if the algorithm reaches a stationary point, the above KKT conditions are satisfied.

1.1.4 Bootstrapping and stabilization of NMF solution

While NMF is an appropriate technique for our problem, we observed that the convergence of the algorithm can be sensitive to initialization. There is no guaranty of convergence to the global optimum of the objective since it is not convex. To ensure robust convergence of the algorithm to a stable solution, we design a bootstrap technique that iteratively stabilizes the solution. The procedure follows three steps: initialization, bootstrap and final optimization described in the next paragraphs.

Initialization At the beginning, all elements of the matrix $\mathbf{W}$ are drawn independently from a uniform distribution on the interval [0 1]:

$$(\mathbf{W}_{init})_{ij} \sim \mathcal{U}([0\,,\,1]).$$

The matrix $\mathbf{H}$ is then initialized accordingly using least-square linear regression of the matrix $\mathbf{S}$ by solving:

$$\tilde{\mathbf{H}} = \arg\min_{\mathbf{H}} \|\mathbf{S} - \mathbf{W}_{init}\mathbf{H}\|^2 .$$

Since the solution of this minimization problem is not guarantied to be non-negative, we take the positive part of each entry to build the initialization matrix (negative entries are zeroed), we write this operation as:

$$\mathbf{H}_{init} = (\tilde{\mathbf{H}})^{+}.$$

Bootstrap We run 50 bootstrap iterations as follows. Starting from the initialized matrices, we partition the N columns of the matrix $\mathbf{S}$ in two subsets of equal length by random permutation of the columns, the N/2 first columns being assigned to the first subset and the last to the second.

The NMF optimization is then run separately on the two matrices built from the columns of the respective subsets resulting in two solutions: $(\mathbf{W}_1, \mathbf{H}_1)$ and $(\mathbf{W}_2, \mathbf{H}_2)$. While the matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are not related, the columns of $\mathbf{W}_1$ and $\mathbf{W}_2$ should in principle reach similar values assuming the number of samples is large enough and the solution of the NMF being stable for this number of components. However, the similarity between columns is up to a permutation of the components. We thus assess similarity by first reordering the columns so that they match together. The ordering of the components is done by computing the kernel PCA of all columns of both matrices pooled together. Then the components of each sub-matrix are reordered according to the value of their projection on the first PCA component.

Once the components are reordered, pairs of columns from each matrix are matched together according to this order and a similarity measure between them is computed according to their mean absolute log ratio. For pairs exceeding the similarity threshold of .1, the corresponding columns are kept for the initialization of the next bootstrap iteration, while the remaining columns will simply be initialized again as described in the previous subsection.

Final optimization After 50 bootstrap iterations, the spectral components found in the last 5 iterations are averaged together to build the final initialization matrix $\mathbf{W}$, and the NMF algorithm is run on the full matrix $\mathbf{S}$ to provide the final solution.

1.1.5 Wiener filtering

Under the GCM assumption, once the matrix parameters $\mathbf{H}$ and $\mathbf{W}$ have been estimated using IS-NMF, it is possible to estimate the time course of the latent components. Taking the example of the first component, estimating $C_1(t)$ from data $X(t)$ amounts to denoising the input signal corrupted by an additive noise $\eta(t)$ which is the sum of all other latent components:

$$X(t) = C_1(t) + \eta(t), \quad \eta(t) = C_2(t) + \ldots + C_K(t).$$

On each time window n, components are stationary Gaussian time series fully specified by the matrix $\mathbf{W}$ and the n^{th} column of $\mathbf{H}$, $\mathbf{H}_{.n}$ and the optimal reconstruction of source signal associated to component k in the Fourier domain is given by the generalized Wiener filter [Smaragdis et al., 2014]

$$\hat{\mathbf{c}}_{nk} = (\mathbf{W}_{.k}\mathbf{H}_{kn}) \oslash (\mathbf{W}\mathbf{H}_{.n}) \odot \mathbf{x}_n,$$

where $\odot$ denotes the entrywise product and $\oslash$ the entrywise division.The resulting estimated time course $\hat{C}_k(t)$ on time frame n can be calculated by inverse Fourier transform of $\hat{\mathbf{c}}_{nk}$. An intuitive interpretation of this procedure in the time domain is that the components correspond to the output of a time varying filter bank , the transfer function $\mathbf{f}_{kn}$ of each filter k and time frame n being given by

$$\mathbf{f}_{kn} = (\mathbf{W}_{.k}\mathbf{H}_{kn}) \oslash (\mathbf{W}\mathbf{H}_{.n}),$$

and thus correspond to a normalized version of the original spectral density $\mathbf{W}_{.k}$ of each component.

1.2 Event detection

1.2.1 Shift-invariant dictionary learning

The objective of dictionary learning is to find a sparse representation of a signal using a set of basis functions. While several sparse representations use predefined basis functions, such as wavelets, dictionary learning also learns these functions from data, in order to achieve a better representation. Among classical dictionary learning techniques, KSVD has become a reference, both for its simplicity and efficiency on real data [Aharon et al., 2006]. KSVD relies on applying two steps iteratively: orthogonal matching pursuit, which learns a sparse representation for a fixed set of basis functions, and a dictionary improvement step, which is implemented efficiently using Singular Value Decomposition (SVD). The analysis of long time series such as ongoing brain activity poses and additional challenge, since the interesting patterns in these time series can be present at multiple times that are unknown a priori. Classical dictionary learning techniques such as KSVD reveal inefficient since they need to learn a large dictionary containing similar instances with different time lags. In contrast, shift invariant dictionary leaning approaches address this issue by learning fixed dictionary patterns (also named *atoms* in the literature) and adjust an additional time lag parameter to fit each possible occurrence of these patterns. An efficient shift invariant generalization of KSVD to long time series was proposed by [Mailhé et al., 2008] in the context of music analysis.

Here we implemented a modified version adapted to our specific application. To prevent overlap between neural events, we impose to the dictionary to capture the time course of the ongoing activity with at most one pattern at a time. In this way, the detected patterns correspond directly to segment of the dynamical component under analysis. In addition to this modification, we developed a cross-validation methodology to estimate the parameters of the dictionary on empirical data, described in the next section.

1.2.2 Selection of the number of events and patterns

For a fixed number of dictionary patterns, we use a cross-validation procedure to find the optimal number of events. To focus on events frequent enough to assess their properties statistically, but rare enough such that their NET-fMRI response can be isolated, we assume each type of event to have a rate of occurrence ranging from 2 per minute to 12 per minute and tested 6 possible rates between these values (2, 4, 6, 8, 10 and 12 events per minute). For a given rate, we fixed the number of events of the shift invariant dictionary learning algorithm according to the length of the recording. After running the algorithm, the detected events were removed from the original time series, and added again to it at random times. We then run the algorithm a second time and evaluate how many of these new events were detected during this new run. We then chose the number of events achieving the best average performance in the retrieval of the randomized events. The number of patterns was initialized to 3, and then decreased as long as the proportion of events associated to each pattern stays above a minimum value (to enforce the method to focus on frequent patterns). The minimum proportion was set to be 50% of the proportion achieved by a equipopulated repartition of the events among the different patterns. This minimum proportion thus corresponds to 15% and 25% for 2 and 3 patterns respectively. If this minimum proportion is not achieved, the number of patterns is ultimately reduced to 1. The performance of this selection procedure is studied in the supplementary results section.

1.3 Characterization of spectral profiles

1.3.1 Spectral centroid and spectral purity ratio

Beyond comparing peak frequencies of spectral profiles, assessing whether the detected spectral components reflect pure narrow-band oscillations (i.e. with a power spectral density having a single narrow peak) or more complex dynamical patterns is important to understand the underlying neural mechanisms. In particular, the non-linear properties of the underlying network interactions can affect the shape of the Power Spectral Density (PSD) of the observed neural time series, for example by generating harmonics at multiples of the peak frequency [Abeysuriya et al., 2014]. To quantify this, we evaluated the spectral centroid of each spectral component. Let $\mathbf{S}(f)$ be the PSD of a given discrete signal, the spectral centroid is the center of mass of this distribution in the frequency domain, defined as:

$$c(\mathbf{S}) = \frac{\int_0^{1/2} f\mathbf{S}(f)df}{\int_0^{1/2} \mathbf{S}(f)df}.$$

As schematized on Figure 1f in main text, for a narrow-band oscillation, the spectral centroid matches very closely the peak frequency. In contrast, discrepancies between the spectral centroid and the peak frequency reveals that the energy is not well concentrated around the peak of the power spectrum, and thus indicates that the signal is not well approximated by a sinusoidal rhythm. As a consequence, we define the *spectral purity ratio* as the ratio of the spectral centroid to the peak frequency. A spectral purity ratio differing from one (or a non-zero log spectral purity ratio) thus indicates that the observed signal differs from a sinusoid and possibly reflects non-linear network interactions.

1.3.2 Illustration with wavelets

To further illustrate how the discrepancy between spectral centroid and peak frequency mark a non-sinusoidal time course, we show how these peak frequency and spectral centroid are related for example to transient patterns originating from wavelet theory [Mallat, 1999]. Since the invention of wavelet analysis, multiple types of wavelets have been designed with various properties for modeling transient signals. Among them, the Morlet wavelet exhibits the closest similarity to short lived sinusoidal oscillation. Alternatively, coiflets show sharp changes in their time course that distinguish them from a sinusoidal pattern. Supplementary Fig. 3b illustrates the time course of these wavelets and show the corresponding absolute value of their Fourier transform. While the Morlet wavelet exhibits a single peak in the Fourier domain, the coiflet shows multiple peaks with decreasing amplitude as the frequency increases. When combining multiple coiflets with different time scales, as illustrated in Supplementary Fig. 3 with two coiflets, the frequency content becomes closer to the monotonously decreasing profiles as observed in our empirical results in Figure 1d-e. From the squared Fourier transform of the wavelet patterns, we derive the same spectral peaks and centroid parameters as for our empirical results. The spectral purity ratio is .98 for the Morlet wavelet and 1.16 for the coiflet, showing non-sinusoidal wavelets exhibit larger discrepancy between these two parameters. When combining two coiflets with different scales, this ratio reached 2.6. In sum, these illustrative examples show the spectral profiles provided by the IS-NMF approach can be characterized beyond the classical peak frequency property to reflect more subtle properties of the time course of neural activity.

1.4 Analysis of thalamocoritical model

1.4.1 Biophysical simulation of thalamocoritical system

In order to investigate to what degree the detected neural events are informative about the cellular processes such as dynamics of the membrane potentials and ionic currents we exploit a simulation of thalamocortical system developed by Costa et al. [2016]. As the details of the model are described earlier [Costa et al., 2016], we restrict ourselves to a brief explanation of the model.

The model of Costa et al. [2016] is a conductance-based neural mass model [Robinson et al., 1997, Liley et al., 1999, 2002, Wilson et al., 2006]. In this class of neural models, population activity can be approximated by the mean membrane potentials, based on an empirical firing rate function [Marreiros et al., 2008]. Populations interact with each other through the synapses. The spike rate of a sender pre-synaptic population elicits a post-synaptic response in a receiving population and the dynamics of this post-synaptic response is determined by a convolution involving the conventional alpha function for synapses.

This thalamocortical neural mass model is consist of 2 modules, a thalamic and a cortical module. The architecture of each module is adopted from Weigenand et al. [2014]. Briefly each of the two modules consist of two sub-module, one excitatory and one inhibitory population. Both excitatory sub-modules of the model are reciprocally connected to both excitatory and inhibitory population of the other module. Furthermore, all sub-modules receive independent background noise in addition to what they recieve through synaptic interaction with other populations.

The thalamic module is consist of a excitatory sub-module, thalamocortical population (t), and an inhibitory sub-module, reticular (r) population. These sub-modules are connected

via AMPA and GABA synapses, but with different synaptic time constant and only the inhibitory population (reticular) has self-connection. Furthermore, in the thalamic module, various ionic currents has been incorporated for the realistic genesis of spindle oscillations. These currents consist of potassium leak current, T-type calcium currents and an rectifier current.

The cortex module similarly consist of an excitatory sub-module, population of pyramidal neurons (p), and an inhibitory sub-module, population of interneurons (i) and similar to thalamic module, they are connected via AMPA and GABA synapses. In contrast to thalamic module, in the cortex module, both sub-modules have self-connections. Furthermore, in the cortex module, some adaptation mechanisms for firing rates has been incorporated which is necessary for transitioning to down (silent) state.

1.4.2 Event detection in the thalamocortical model

Neural events in the thalamocortical model was identfied with the similar procedure used for neural data. As the extracellular field potential stems mainly from activity of pyramidal neurons [Buzsaki et al., 2012], we used the membrane potential of pyramidal neurons in thalamus module as a crude proxy of thalamus LFP and identify the neural events in the time course of the signal. First the short-term Fourier transform (STFT) of the signals over overlapping time windows was computed, then by applying non-negative matrix factorization (NMF) on the spectrograms results from STFT of the LFP, we identify the spectral profile of the characteristic transients of LFP. Lastly, by applying shift-invariant dictionary learning we temporally localize the neural events and identify their sub-types. The number of component chosen for NMF factorization was 3, based on the procedure explain in the method section of the main text.

1.4.3 Low dimensional representation of cellular dynamics

We represent the cellular dynamics during the occurrence of the events in a 2-dimensional (after dimensionality reduction) sub-space. We consider the full space span by concatenated time course of all membrane potentials and the calcium current. Membrane potentials were consist of cortex pyramidal and inhibitory population, and thalamic reticular and thamamocoritical (excitatory) populations. Around each event, a window of length 1000 sample has been used for the time course of each cellular variable. To represent cellular dynamics in low dimentional sub-space we used t-distributed stochastic neighbor embedding (tSNE) [van der Maaten and Hinton, 2008].

2 Supplementary results

2.1 Comparison of NMF techniques

We run a comparison of NMF techniques to address the problem of identifying events with different spectral profiles occurring in a time series. Events were generated by bandpass filtering three homogenous Poisson processes in different frequency bands (5Hz, 20Hz and 50Hz respectively) with impulse responses of differing amplitude (1, .5 and .1 respectively). As a consequence (see example time course in Supplementary Fig. 1a), the high frequency events have less energy than the low frequency events, as frequently observed in empirical

LFP time series. The spectral profiles corresponding to each processes are represented in Supplementary Fig. 1b. The three filtered process are then summed and a small Gaussian white noise (with standard deviation .001) added to generate the input time series. To retrieve the three spectral profiles from this mixture signal, we used two NMF approches: Euclidean NMF and IS-NMF using 4 components (one extra components is used to capture the noise). The similarity between original and retrieved spectral profiles was quantified using cosine similarity. The average performance on each approach is reported in Supplementary Fig. 1c. While both approaches perform well for low and middle frequency profiles (reaching the maximal cosine similarity value 1), only IS-NMF reliably estimates the high frequency profile.

2.2 Dictionary learning

To check the validity of our dictionary learning approach, we simulated noisy time series with different neural events occurring across time. To measure time related quantities in seconds, we assumed a sampling rate of 500Hz. Events were generated by bandpass filtering unit amplitude impulses from an homogenous Poisson process from which we excluded overlapping events by eliminating any event occurring less than 1s after a previous one. To replicate the setting reported in the main text, in which each dynamical component is analyzed separately with this approach, the time course of the generated neural events are chosen in order to have similar frequency properties but possibly different time courses that indicate different subtypes of events. This was done by choosing Butterworth band pass filters with a comment center frequency of 5Hz but different bandwidth (.89 and 4.47Hz respectively). Twenty simulations where performed for two cases: either only one of both of these subtypes of events are generated in the time series (which means the dictionary had either one or two patterns), allowing us to assess whether our approach retrieves correctly the number of event subtypes present in the data. An example simulated time series is shown in Supplementary Fig. 5a, with a magnified portion showing the two types of simulated events present in this data (Supplementary Fig. 5b), the most oscillatory events being generated by the band-pass filters with the smallest band-width. We applied the dictionary learning approach as described in Supplementary Methods in order to learn the different types of events and detect their occurrence. The approach was implemented assuming a maximum number of dictionary patterns of 3, and the number of events was selected from an equispaced grid of 8 values ranging from 66 to 400. The dictionary patterns learned from the data of Supplementary Fig. 5a is shown on panel c of the same Figure, witnessing that the shape and number of event subtypes is well recovered. We assessed quantitatively the performance of our approach by first quantifying whether the total number of generated events was correctly estimated. Supplementary Fig. 5d shows the original distribution of the number of events in the simulation, while the corresponding estimated number of events is shown in Supplementary Fig. 5e. This shows the estimated number of events is close to the original value in most cases. Moreover, the correct number of event subtypes is correctly estimated in 80% of the cases (Supplementary Fig. 5(f)).

2.3 Comparison of Hp and LGN events

According to the similarities shown in to the two dimensional maps of event properties in Figure 4c and 4e, we compared Hp Cluster 1 to LGN Cluster1, Hp Cluster 2 and 3 to LGN

Cluster 2, and Hp Cluster 4 to LGN Cluster 3. By first comparing their peak frequency (Supplementary Fig. 7a), we found the that lower frequency Hp and LGN events (labeled Cluster 1 in both cases) were comparable in this respect, so as well as gamma band events (Hp Cluster 5 and LGN Cluster 3. On the contrary, neither HP Cluster 2 nor 3 were matching the higher frequency TA LGN Cluster 2 as far as their peak frequency is concerned. In addition to differences in peak frequencies, LFP dynamics of neural events can be compared according to finer properties of the Fourier transform of their associated dictionary patterns. In particular, we can extend the analysis of spectral purity ratio performed on the NMF spectral profiles in a previous section to assess how close to a sinusoid are each event by computing the log of the spectral purity ratio, which should be close to zero for nearly sinusoidal events. Comparison of the log spectral purity ratio between Hp and LGN events (Supplementary Fig. 7b) shows that the two lower frequency TA events from each structure (Clusters 1 and 2 in each structure) are comparable. On the contrary the TA Hp cluster of highest frequency (Cluster 3) has significantly lower spectral purity ratio. In the same way, in the low gamma band, Hp events have lower spectral purity ratio than their LGN counterpart. These results are in accordance with the our analysis of spectral profiles, suggesting a larger contribution of non-linear network interactions in LGN high frequency events than in Hp events of comparable frequency.

3 Supplementary discussion

Time frequency decomposition

Our methodology is based on short-term Fourier transform (SFFT) of the LFP signals over overlapping time windows. Other types of time-frequency decompositions [Martin and Flandrin, 1985, Rioul and Vetterli, 1991] (wavelet transform, Wigner-Ville distribution) could possibly be used and possibly generalize the present approach. However, SFFT provides a much faster time frequency representation of the signal, together with a sound statistical modeling framework [Smaragdis et al., 2014]. As a consequence, while the SFFT representation shown at the beginning of the paper might look noisier than other time frequency representation, it actually carries all necessary statistical information to lead to a decomposition of the LFP time course into a sum of dynamical components. As opposed to a classical time-frequency representation, our approach thus offers a direct visualization of the time courses of different dynamical aspects of the LFP, instead of an abstract representation trough wavelet coefficients or spectrograms with a lower time resolution.

Dictionary learning

To analyze quantitatively the interesting patterns that could, to some extent, be readily visually identified in this decomposition, we performed a detailed quantitative analysis of dynamical events by learning a dictionary of recurrent dynamical patterns for each component. This approach enables to isolate different dynamical events that might not be easily identifiable from their power spectral density, as illustrated by sharp-waves (cortex-activating) and low-frequency thalamus-activating events in the hippocampus. Dictionary learning can also be related to wavelet analysis techniques used for example to provide a sparse description of a signal in classical matching pursuit [Mallat and Zhang, 1993]. Our approach indeed provides

a fully automated optimization of a sparse representation with a few dictionary patterns, while in wavelet analysis the shape of those patterns is strongly constrained by a fixed set of basis functions, which has to be selected first. While in principle, dictionary learning could be applied directly to the original LFP time course, the NMF based decomposition is critical for reducing the complexity of the time series such that even dynamical patterns with low energy can be robustly detected.

Generality of the approach

Our approach is a fully unsupervised technique: no additional information (such as behavior or sensory input) is required to assess relevant frequency bands, in contrast to other methods [Magri et al., 2012]. As a consequence our approach is particularly well suited to study ongoing activity in various passive or active states and enables a precise study of the evolution of the dynamical content of neural activity from sleep and anesthesia to cognition and complex behavior. Although our approach does not use any prior knowledge on the structure under study, it was able to retrieve the known characteristic events of both structures during passive states, namely delta oscillation, and spindles for thalamus, and shape-wave ripples and gamma oscillations for hippocampus. This supports that the assumptions underlying our method are general enough to capture a wide range of dynamical events without any tuning from the experimenter, and shows for the first time a description of these activity that is not biased by human expertise. Along this line, we noticed that many of our detected waveforms were more complex than quasi-sinusoidal oscillations. Analysis of the power law behavior of broad band LFP power spectrum has previously been suggested to quantify scale free properties and self-similarity in these time series and was proposed as a model of arrhythmic brain activity that can be taken apart from quasi-sinusoidal components [He et al., 2010]. While our results suggest that many interesting neural events are not quasi-sinusoidal, whether they are related to self-similar processes is an interesting future direction. In addition, we stress that this approach can be applied to single channel data (while generalization to multiple channels will be addressed in future work) because it does not rely on assumptions on the spatial spread of the activity due to multiple underlying current generators. The activity captured in each component may or may not reflect spatially segregated generators; in particular, a same generator can contribute to different components at different times if it is involved in different dynamical network events. The spatial characteristics of each components or events can however be further studied in the context of multiple channel recordings and/or concurrent recording of brain wide activity with neuroimaging techniques.

References

R. G. Abeysuriya, C. J. Rennie, and P. A. Robinson. Prediction and verification of nonlinear sleep spindle harmonic oscillations. *Journal of Theoretical Biology*, 344:70–77, Mar. 2014. ISSN 0022-5193. doi: 10.1016/j.jtbi.2013.11.013.

M. Aharon, M. Elad, and A. Bruckstein. -svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11): 4311–4322, 2006.

G. Buzsaki, C. A. Anastassiou, and C. Koch. The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. *Nature reviews. Neuroscience*, 13(6):407–20, May 2012. ISSN 1471-0048 (Electronic) 1471-003X (Linking). doi: 10.1038/nrn3241.

M. S. Costa, A. Weigenand, H.-V. V. Ngo, L. Marshall, J. Born, T. Martinetz, and J. C. Claussen. A Thalamocortical Neural Mass Model of the EEG during NREM Sleep and Its Response to Auditory Stimulation. *PLOS Computational Biology*, 12(9):e1005022, Sept. 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005022.

B. J. He, J. M. Zempel, A. Z. Snyder, and M. E. Raichle. The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66(3):353–369, 2010.

D. T. J. Liley, P. J. Cadusch, and J. J. Wright. A continuum theory of electro-cortical activity. *Neurocomputing*, 26–27:795–800, June 1999. ISSN 0925-2312. doi: 10.1016/S0925-2312(98)00149-0.

D. T. J. Liley, P. J. Cadusch, and M. P. Dafilis. A spatially continuous mean field theory of electrocortical activity. *Network*, 13(1):67–113, Jan. 2002. ISSN 0954-898X. doi: 10.1088/0954-898X/13/1/303.

C. Magri, A. Mazzoni, N. K. Logothetis, and S. Panzeri. Optimal band separation of extracellular field potentials. *Journal of neuroscience methods*, 210(1):66–78, 2012.

B. Mailhé, S. Lesage, R. Gribonval, F. Bimbot, P. Vandergheynst, et al. Shift-invariant dictionary learning for sparse representations: extending k-svd. In *16th EUropean SIgnal Processing COnference (EUSIPCO'08)*, 2008.

S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 2nd ed edition, 1999. ISBN 978-0-12-466606-1.

S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

A. C. Marreiros, J. Daunizeau, S. J. Kiebel, and K. J. Friston. Population dynamics: Variance and the sigmoid activation function. *NeuroImage*, 42(1):147–157, Aug. 2008. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2008.04.239.

W. Martin and P. Flandrin. Wigner-ville spectral analysis of nonstationary processes. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(6):1461–1470, 1985.

O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE signal processing magazine*, 8(LCAV-ARTICLE-1991-005):14–38, 1991.

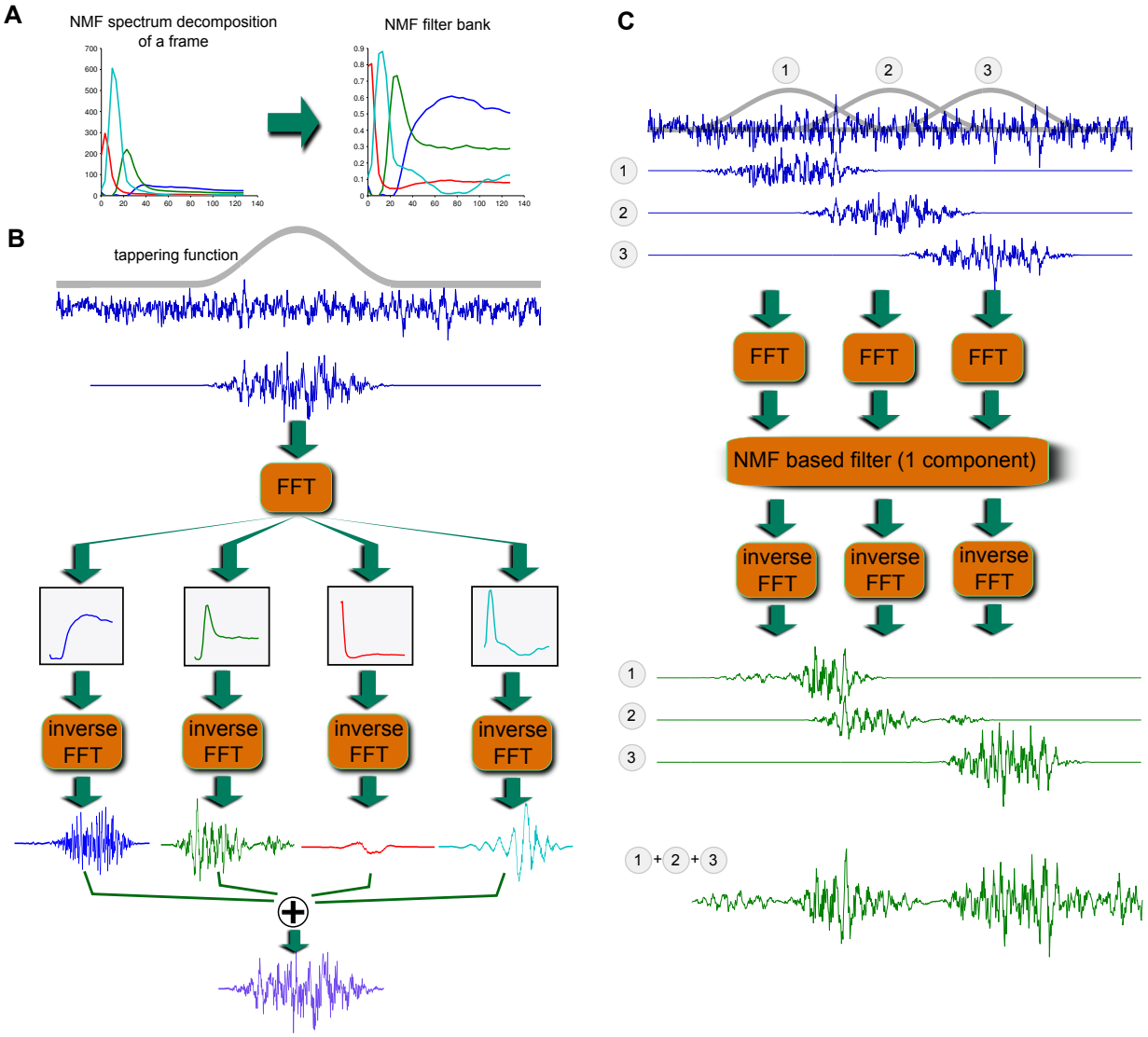
P. A. Robinson, C. J. Rennie, and J. J. Wright. Propagation and stability of waves of electrical activity in the cerebral cortex. *Phys. Rev. E*, 56(1):826–840, July 1997. doi: 10.1103/PhysRevE.56.826.

P. Smaragdis, C. Fevotte, G. Mysore, N. Mohammadiha, and M. Hoffman. Static and dynamic source separation using nonnegative factorizations: A unified view. *Signal Processing Magazine, IEEE*, 31(3):66–75, 2014.

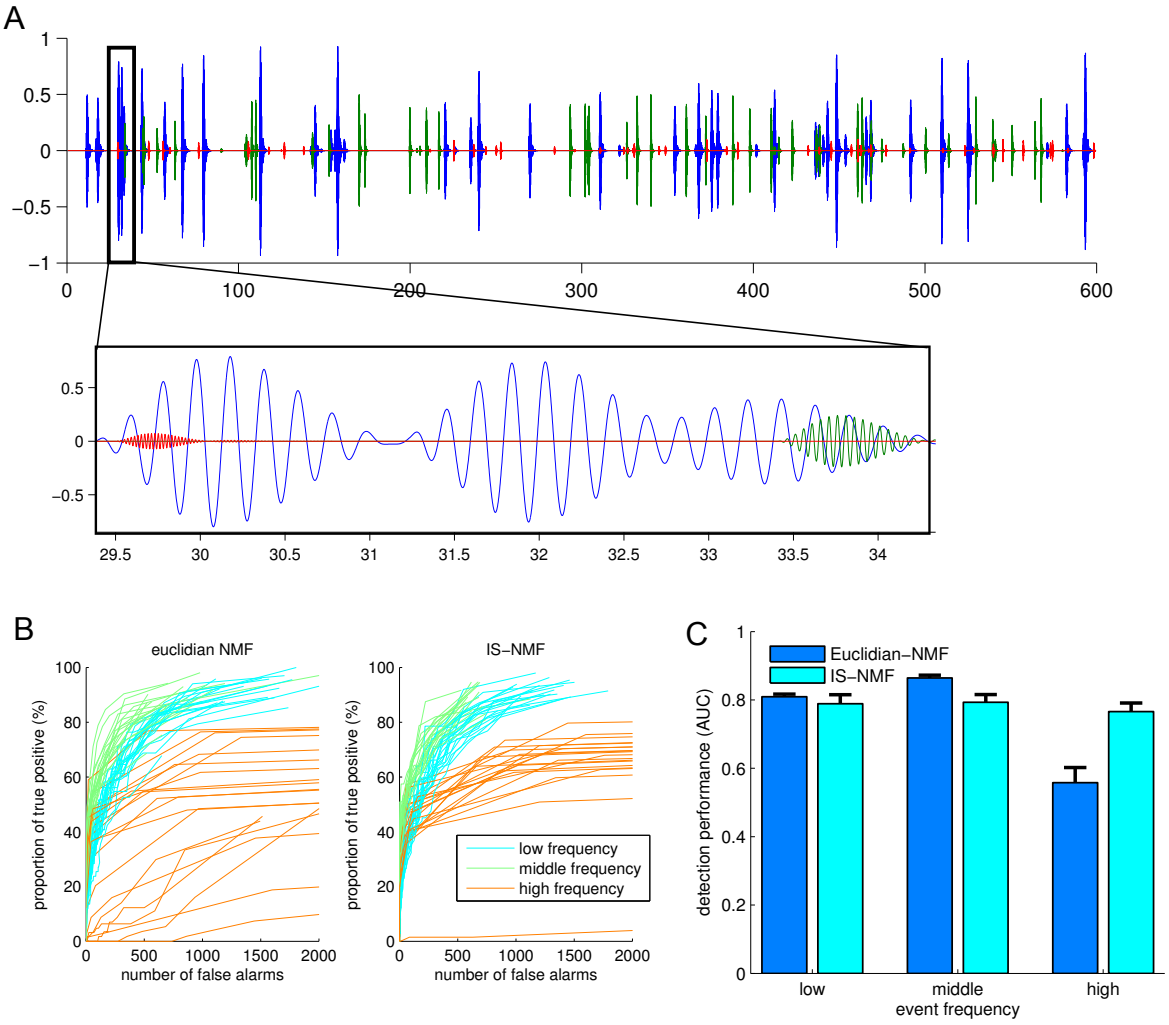
L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. *J. Mach. Learn. Res.*, 9 (86):2579–2605, 2008. ISSN 1533-7928.

A. Weigenand, M. S. Costa, H.-V. V. Ngo, J. C. Claussen, and T. Martinetz. Characterization of K-Complexes and Slow Wave Activity in a Neural Mass Model. *PLOS Computational Biology*, 10(11):e1003923, Nov. 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003923.

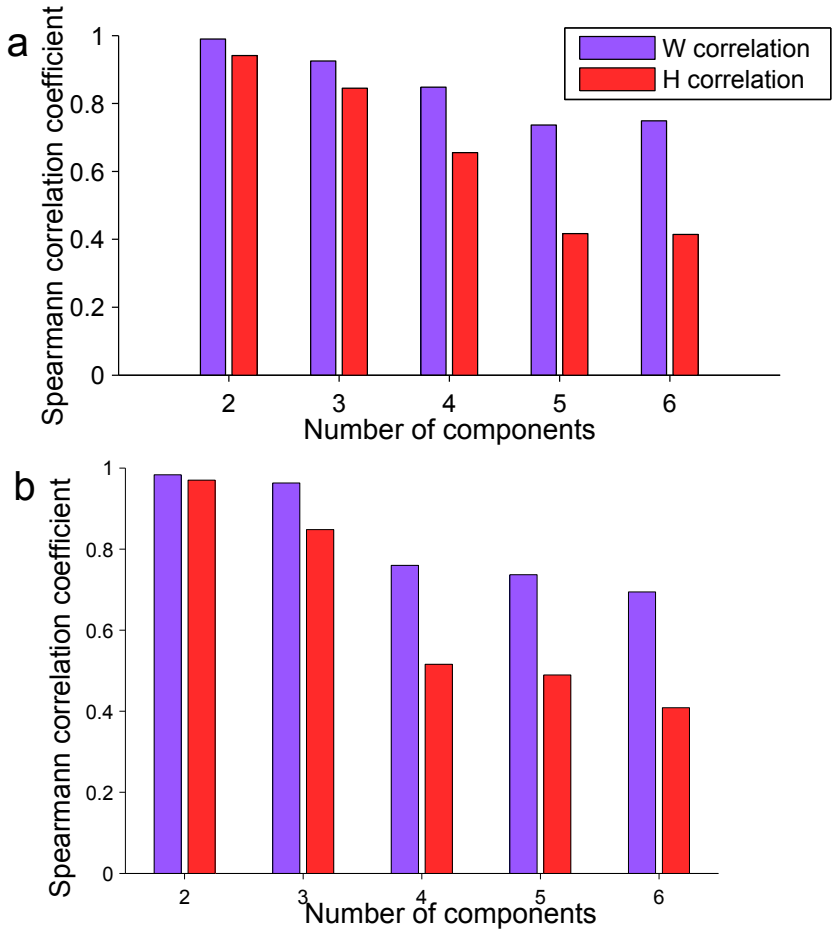
M. T. Wilson, J. W. Sleigh, D. A. Steyn-Ross, and M. L. Steyn-Ross. General Anesthetic-induced Seizures Can Be Explained by a Mean-field Model of Cortical Dynamics. *Anesthesiology*, 104(3):588–593, Mar. 2006. ISSN 0003-3022. doi: 10.1097/00000542-200603000-00026.



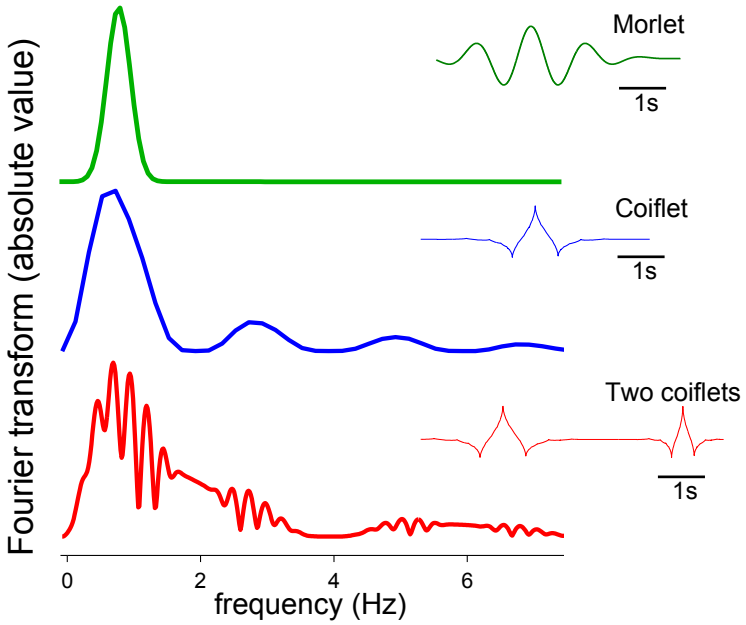
Supplementary Figure 1: **Principle of the extraction of the time course of dynamical components based on NMF results.** (a) For a given time window, the weighted spectral profiles of all NMF components are normalized for each frequency in order to sum to one. (b) These resulting functions are used as transfer functions in a filter bank of K filters, which are efficiently implemented as a simple multiplication in the frequency domain using the FFT algorithm. Activity of a given component is estimated by applying the corresponding filter to the LFP signal multiplied by a tapering function selecting the block of signal of current time window. Due to the normalization, the resulting sum of the time courses across all components reconstructs the original LFP signal. (c) The decomposition of panel b is implemented efficiently across the full time course of the signal using overlapping time windows, exemplified here with three time windows. The signal of each window is multiplied by a smooth tapering function such that the tapering coefficients of all overlapping windows at a given time point sum to one. This allows reconstructing without discontinuity the full time course of a component after filtering each frame independently by simply summing the filtered time courses contributed by each frame.



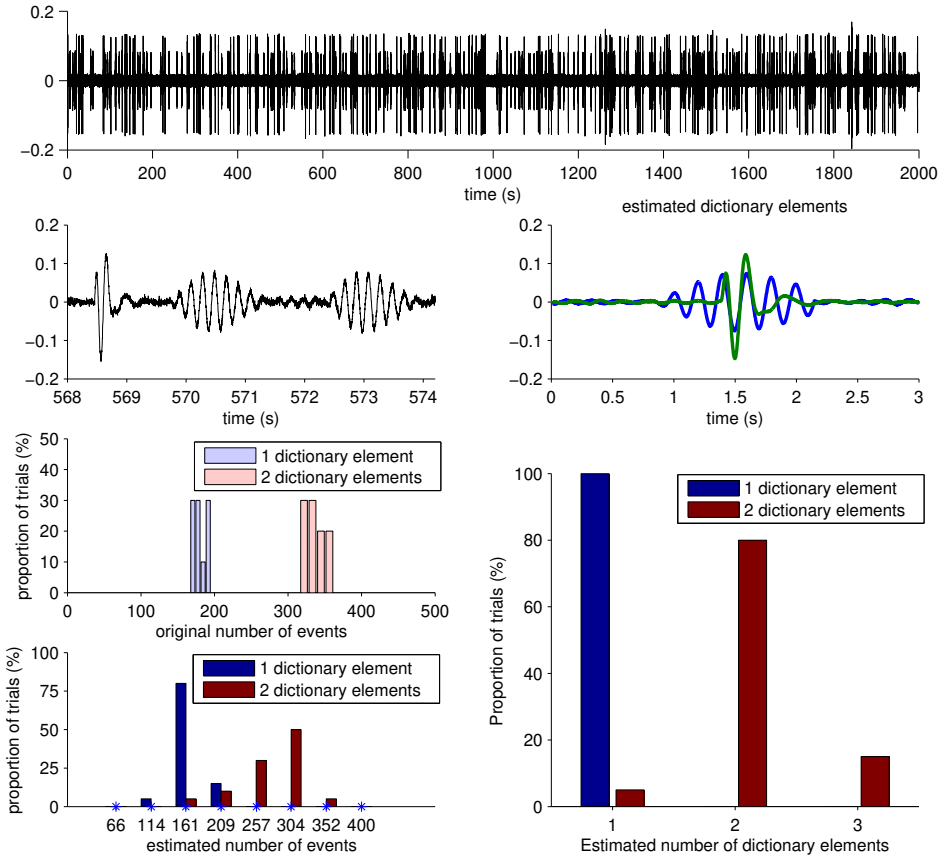
Supplementary Figure 2: **Comparison of Euclidean and Itakura-Saito NMF in simulation.** (a) Example time course of simulated dynamical components, representing low frequency (in blue), middle frequency (in green) and high frequency oscillations (in red). The inset represents a magnified version of the signals. (b) Spectral profiles associated to each components (estimated using the Welch periodogram). (c) Average cosine similarity between the original spectral profiles and those retrieved using Euclidean or Itakura-Saito NMF (error bars indicate standard deviation).



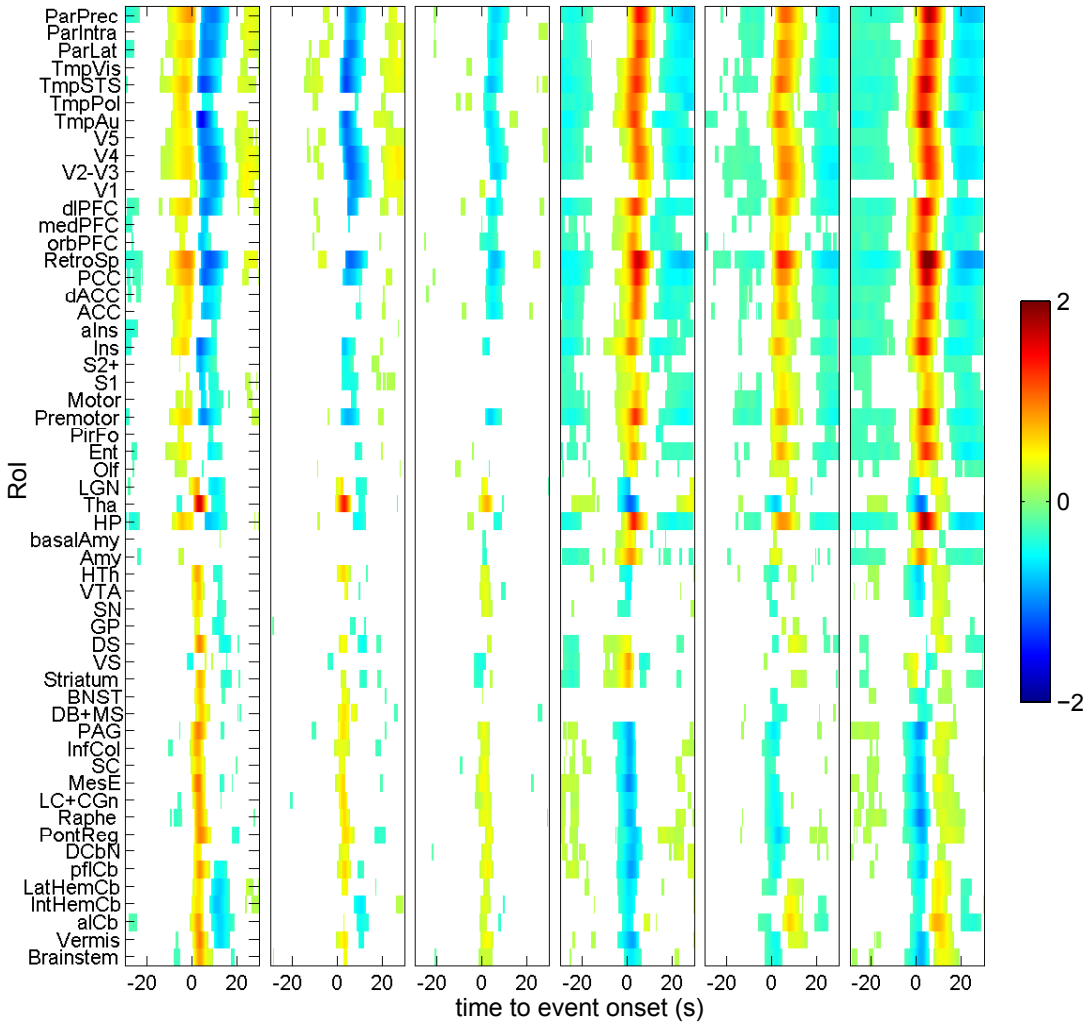
Supplementary Figure 3: **Selection of the number of NMF components.** (a) Average cosine similarity between bootstrapped NMF components averaged across all Hp recording sessions (blue: similarity between spectral profiles stored in the columns of the matrix **W**, red similarity between time-varying contribution of profiles to each time window stored in the lines of the matrix **H**) (b) Average cosine similarity between bootstrapped NMF components averaged across all LGN recording sessions.



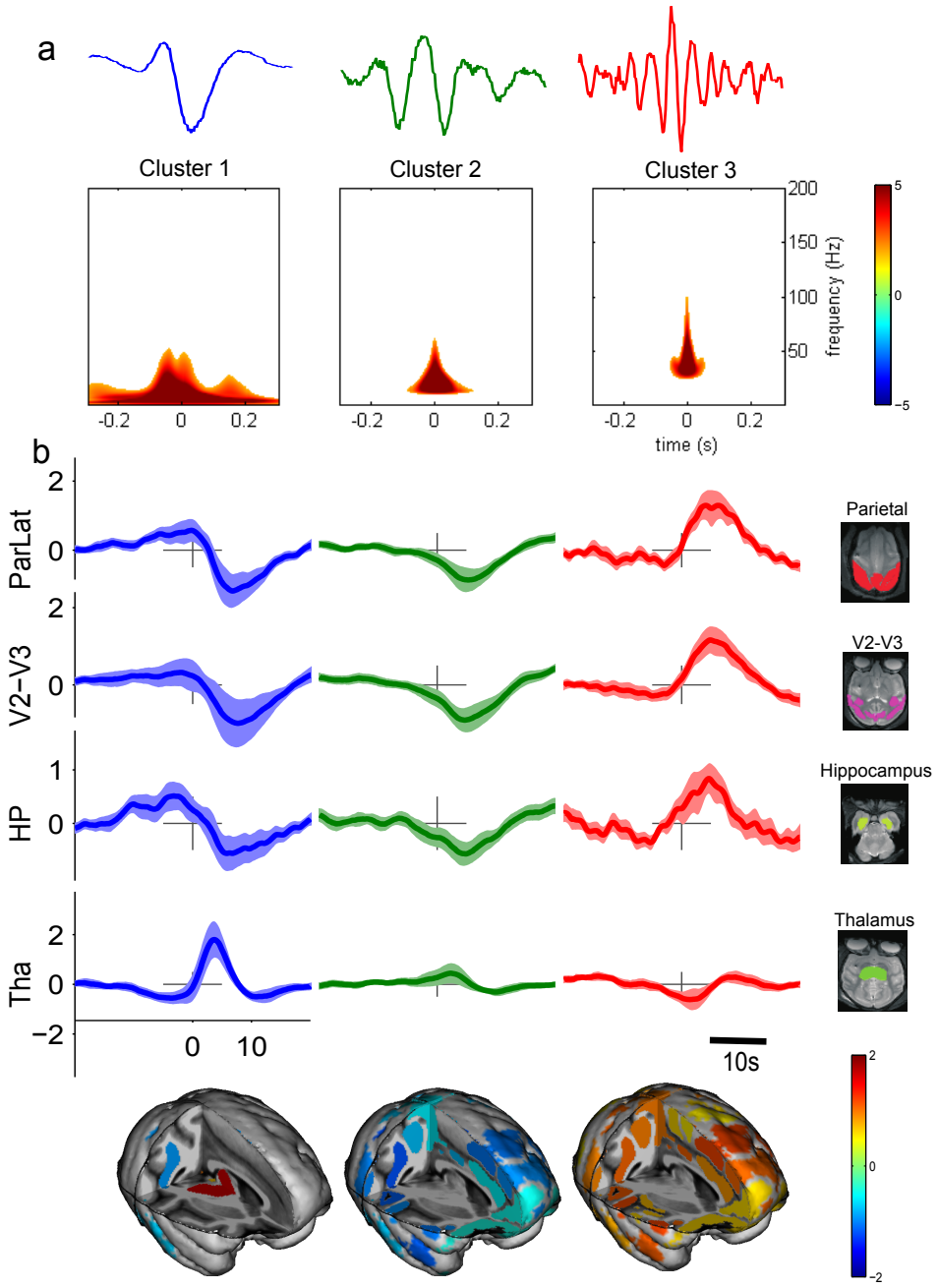
Supplementary Figure 4: **Properties of spectral profiles.** Absolute Fourier transform for different types of temporal patterns (see Supplementary Methods). Insets on the right hand side represent the time course of each pattern.



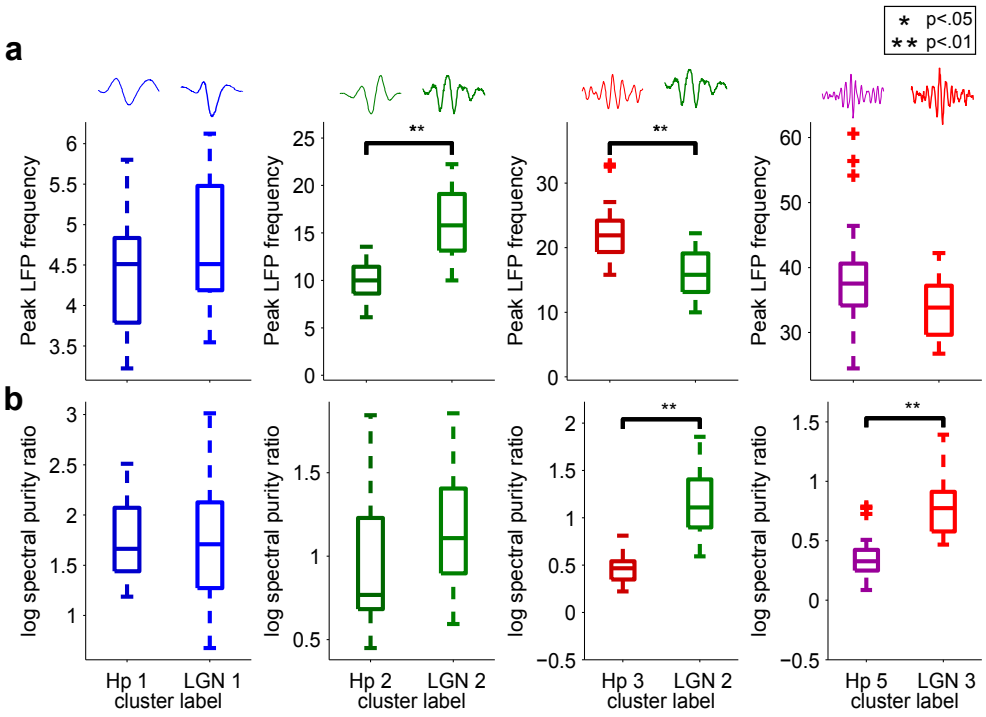
Supplementary Figure 5: **Dictionary learning.** (a) Example time course of a simulated time series with events taken from a dictionary with two patterns. (b) Magnified portion of time-course in (a) showing the occurrence of the first (blue arrow) and the second (green arrow) dictionary pattern. (c) Example of dictionary patterns estimated using our dictionary learning approach. (d) Histogram of the original number of events generated in simulations for two cases: events generated from one or two-patterns dictionary. (e) Histogram of the estimated number of events for both cases. (f) Estimated number of dictionary patterns estimated for both cases.



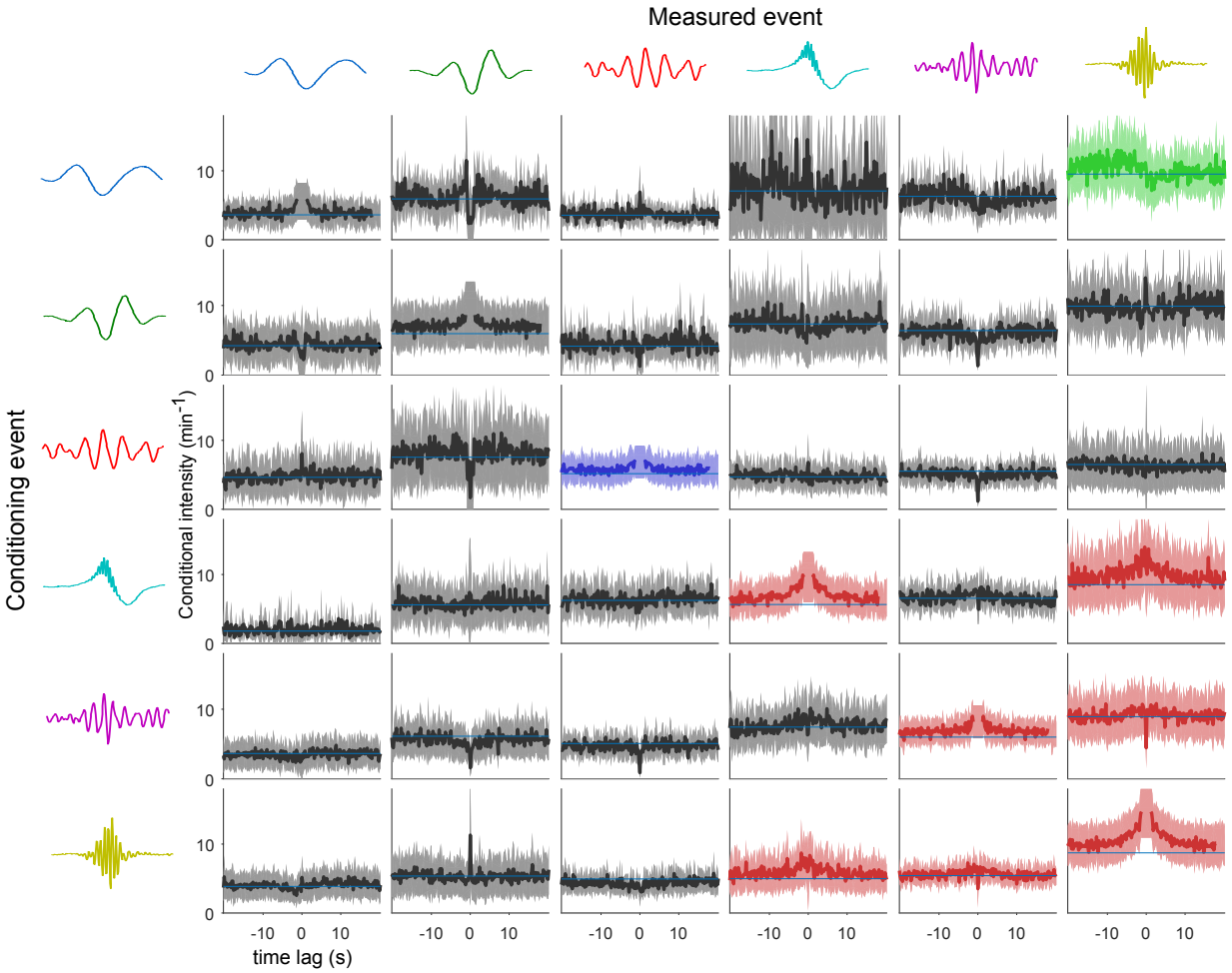
Supplementary Figure 6: **Full NET-fmri response for hippocampal events.** NET-fmri response averaged across voxels of each given ROI for each cluster of events. Rows indicate the considered ROI, columns indicate the cluster of events (same order as Figure 4). Values are z-scored with respect to randomized evens. The meaning of Roi labels on the left hand side are detailed in Supplementary Table 1.



Supplementary Figure 7: **Clustering of LGN events.** (a-b) Average time-frequency map (a) and NET-fMRI (b) of the 3 clusters of LGN events (Z-score with respect to randomized event onsets). Waveforms in panel a are examples of learned dictionary patterns for each cluster. Bottom insets represent the trimensional mapping of the amplitude of this response at t=3s overlaid on a template macaque brain.



Supplementary Figure 8: **Comparison of LGN and hippocampal event LFP properties.** (a) Comparison of the distribution of peak frequency of dictionary patterns originating from Hp and LGN structures. Horizontal lines indicate significant differences (Wilcoxon rank sum test; p<.05, Bonferoni corrected, from left to right: n=59,58,44,43). Traces at the top represent example dictionary patterns of each cluster. (b) Comparison of the distribution of log spectral purity ratio of dictionary patterns originating for Hp and LGN structures. Horizontal lines indicate significant differences (Wilcoxon rank sum test; p<.05, Bonferoni corrected, from left to right: n=59,58,44,43).



Supplementary Figure 9: **Conditional intensity of Hp events.** Estimated intensity of each type of event, conditioned on the occurence of another type of event at time 0. Columns correspond to each event of which the rate is estimated, and lines indicate the event used for conditioning. Colored curves indicate significant average deviation of the intensity from baseline for different time intervals (Wilcoxon signed rand test; $p<.05$; FDR adjusted): red plots indicate change in the [-3s,3s] time window, excluding a [-1s,1s] interval and tested across all CA event pairs, blue plots indicate the same tested for all TA pairs, green curves indicate change in the [-20s,-7s] interval tested across pairs of CA intensity, conditioned on TA events.

Group-ROI	Included Structures/Areas
V5	MT/MST
V4	V4 Complex
V2-V3	Areas V2/V3
V1	Primary Visual Cortex
ParPrec	Parietal Precuneus
ParIntra	Intraparietal Cortex
ParLat	Lateral Parietal
TmpVis	Inferotemporal Cortex
TmpSTS	Superior Temporal Sulcus
TmpAu	Temporal Auditory
TmpPol	Temporal Pole
dlPFC	Dorsolateral Prefrontal
medPFC	Medial Prefrontal
orbPFC	Orbitofrontal Cortex
S2+	Somatosensory Association
S1	Somatosensory Primate
Motor	Motor Cortex
Premotor	Premotor Cortex
RetroSp	Retrosplenial Cortex
PCC	Posterior Cingulate
ACC	Anterior Cingulate
Ins	Insular Cortex
Olf	Ofactory Cortex
PirFo	Piriform Cortex

Group-ROI	Included Structures/Areas
Ent	Entohrinal Cortex
HP	Hippocampus
Amy	Amygdala
HTh	Hypothalamus
Septum	Septum
DB	Diagonal Band
GP	Glovus Pallidus
Striatum	Striatum
LGN	Lateral Geniculate Nucleus
Tha	Thalamus
VTA	Ventral Tegmental Area
SN	Substantia Nigra
LC+CGn	Locus Coeruleus – Cent Gray
PAG	Periaqueductal Gray
InfCol	Inferior Colliculus
SC	Superior Colliculus
Raphe	Raphe Nuclei
PontReg	Pontine Region
alCb	Anterior Cerebellar Lobe
DCbN	Deep Cerebellar Nuclei
pflCb	Paraflculonodular Cerebellum
LatHemCb	Lateral cereb. hemisphere
IntHemCb	Intermediate cer. hemisphere
Vermis	Vermis

Supplementary Table 1: **ROI labels.** Nomenclature of the ROIs used in the present study. The two columns show respectively the utilized group ROI labels and the cortical or subcortical regions that they include. More information can be found in (Logothetis et al. 2012).

CSB, 2020-10-11, No. 1, 2020
Work-in-progress
CSB—SoT-20201011–v1.0

Signatures of criticality in efficient coding networks

Shervin Safavi[1,2], Matthew Chalk[3], Nikos Logothetis[1], Anna Levina[1,4,5]

Neural computation (computations brain need to perfom) and neural dynamics (dynamics of the brain activity) are two important facets for understanding the brain. Ideally, we need a single framework that can accommodate both of these aspects. An intermediate step toward developing such a framework is exploiting the models that are either centered around neural computation or neural dynamics, *but* with implications for the other aspect. Indeed, there are normative models that have implications for neural dynamics and also models of neural dynamics with implications for neural computation. Neural coding, as one of the important computations brain need to perform, is of particular interest for building bridges between neural computation and dynamics. Notably, there has been various studies that suggest potential connections between neural coding and neural dynamics. Despite the appeal of this connection, networks implementing efficient coding has not been extensively investigated for important aspects of neural dynamics (e.g. scale-freeness). In this study, we investigate for signatures of criticality in networks that can be optimized to perform efficient coding. We consider a network of leaky-integrate and Fire neurons with synaptic transmission delays whose connectivity and dynamics can be optimized for efficient coding. Previously, it was shown that the performance of such networks varies non-monotonically with the noise amplitude. We consider networks with different noise amplitudes and investigate signatures of criticality. Interestingly, in the vicinity of the optimal noise level for efficient coding, the distribution of avalanche sizes follows a power-law. When the noise amplitude is too low or too high for efficient coding, the network appears either super-critical or sub-critical, respectively. This result has important implications, as it shows how two influential, and previously disparate fields – efficient coding, and criticality – might be intimately related.

Keywords
Criticality — Efficient coding — Neural dynamics — Neural computation

[1] *Department of Physiology of Cognitive Processes, MPI for Biological Cybernetics, Tübingen, Germany*
[2] *School of Neural Information Processing, IMPRS for Cognitive and Systems Neuroscience, Tübingen, Germany*
[3] *Institut de la Vision, Sorbonne Universite, Paris, France*
[4] *Department of Computer Science, University of Tübingen, Tübingen, Germany*
[5] *Bernstein Center for Computational Neuroscience Tübingen, Tübingen, Germany*

Introduction

Understanding the computation brain needs to perform (neural computation) and the dynamics of the brain activity (neural dynamics) are two important goal of computational neuroscience [20, Chapter 1]. Indeed, within each branch, an extensive research has been done, and they led to important insight about the brain [1, 20, 26, 39]. Certainly, understanding each of the mentioned aspects (neural computation and dynamics) is important on their own, but perhaps more important than that is having a framework that can accommodate both [20, 35]. Having the ideal framework that can incorporate sophisticated computational objectives of the brain and also the rich dynamics of the brain is an appealing and ambitious goal. Nevertheless, there are already various frameworks and models that is either centered around neural computation or neural dynamics *with implications for the other aspect*. One of the first step toward developing the ideal framework could be investigating the potential connections between normative models that have implications for neural dynamics and models of neural dynamics with implications for neural computation.

On one side, various normative models of neural computation has been developed that can explain some aspects of the observed brain dynamics [10, 11, 16, 18, 31, 32, 34, 58], such as irregular spiking [11] and neural oscillations [5, 18]. Furthermore, there has been efforts to relate state of the machinery implementing a given neural computation to putative dynamical regime of the neural model. For instance Echeveste et al. [32] and Lengyel et al. [51] have developed neuronal network which implement Bayesian inference, and the network is also an attractor networks. In the most of such normative models, we optimize or train the network based on a specific computational objective (such as minimizing reconstruction error) and the features of the neural dynamic appear in the resulting network activity as a byproduct.

One the other side, one of the frameworks for explaining the neural dynamics with connection to neural computation is the "criticality hypothesis of the brain" (for reviews see [63, 65, 85]). This hypothesis states that, brain operates close to a critical state, and pertaining to neural computation, it has been shown that general information processing capabilities such as sensitivity to input [15, 48], dynamic range [48, 50, 67], and information transmission and storage [53, 55, 79, 88], and various other computational characteristics are optimized in this state. To evaluate how closeness to criticality can be beneficial for the information processing, the common approach is using a neural model (e. g. a branching network, a recurrent neural network) that can attain various states (in-

cluding critical and non-critical state) depending on control parameters (e. g. branching ratio, connection strength) and quantify how the general information processing capabilities (mentioned earlier e. g. sensitivity to input) depend on the control parameters. Typically, these capabilities appear to be optimal by tuning the network close to the critical state, but without any special treatment pertaining to the computational objective.

Overall, on the side of normative models, some aspect of neural dynamics has been explained by the previous normative models (like neural oscillations [5, 18]), nevertheless, the brain dynamics has been shown to be more complex than the reach of normative models [14, 27]. Not only in terms complexity of the observed dynamics, but also in terms of the scale [2, 38]. Normative models mentioned earlier mainly explain a single scale dynamics and that is the circuit-level dynamics, but brain has a rich multi-scale dynamics. On the criticality side, being in a state with such an optimized capabilities should be relevant for the computations in the brain, but mere adjusting for the closeness to criticality cannot provide an algorithm for a given neural computation. Therefore, we believe it is necessary to have a *complementary approach* wherein we seek for connections between models and frameworks of neural dynamics and computation. Indeed, this is complementary to the approach focusing on each of these aspects in isolation.

Neural coding is of particular interest from the perspective of a complementary approach introduced earlier, as it is one of the well established and functionally relevant computations that brain needs to perfom [73, 74] and there has been various studies that suggest potential connections between neural coding and neural dynamics [5, 11, 18, 32, 34, 45, 75, 80]. In particular, multiple recent studies provide qualitative or quantitative evidence on the usefulness of operating close to a phase transition for coding [18, 45, 75, 80]. Interestingly, phase transition is also one of the pillars of the criticality hypothesis of the brain [63, 65, 85]. Despite of the attractiveness of this connection, to best of our knowledge, networks implementing neural coding has never been investigated for signatures of criticality.

We investigate a network of Leaky-Integrate and Fire (LIF) neurons whose connectivity and dynamics are optimized for coding one-dimensional input [18]. The performance of the network varies non-monotonically with the noise amplitude. Therefore, the network can be optimized using the noise amplitude as a control or tuning parameter. We consider networks with different noise amplitudes and investigate the presence of the signatures of criticality in the networks operating with the noise levels in the vicinity of optimal noise level for efficient coding and the absence of signatures when the network operate with noise levels far from optimal level for efficient coding. Interestingly, in the vicinity of optimal noise level we observe the signatures of criticality and when the noise amplitude is too low or too high for efficient coding, the network appears either super-critical or sub-critical, respectively.

Our result suggests that coding-based optimality might co-occur with closeness to criticality or at least in the vicinity of criticality. In fact, this co-occurrence can have important implications, as it shows two influential, and previously disparate fields — efficient coding, and criticality — might be intimately related. This work suggests several promising avenues for future research on the computation and dynamics of neural system.

Results

Optimization of coding efficacy

In this study, we investigate a network of LIF neurons, consist of two populations, whose connectivity and dynamics can be optimized for coding one-dimensional input [18] (see Figure 1A for a schematic illustration and Method section for more details). This network can be optimized to encode the input efficiently by adjusting the level of noise in the membrane potentials. Input is reconstructed by performing a linear readout of the spike trains (see Boerlin et al. [11] for more details). Given an idealized network with instantaneous synapses [11], neurons that receive a common input avoid communicating redundant information via instantaneous recurrent inhibition. However, adding realistic synaptic delays leads to network synchronization that impairs coding efficiency. Chalk et al. [18] demonstrated that, in the presence of synaptic delays, this network can nonetheless be optimized for efficient coding by adding noise to the network. Network's performance depends non-monotonically on the noise amplitude, with the optimal performance achieved for an intermediate noise level (Figure 1B, center; Figure 1C, the minimum MSE). When the noise level is considerably smaller than the optimal noise, the reconstructed input tend to overshoot (Figure 1B, left), and when the noise level is considerably larger than the optimal noise, the reconstructed input is not precise (Figure 1B, right; Figure 1C, MSE for large noise levels).

Signatures of criticality

As it was illustrated in the previous section, noise can be used to adjust the network to operate in an optimized regime for efficient coding. Interestingly, the network with noise level in the vicinity of optimal noise (minimum reconstruction error) qualitatively operate between a very synchronized (Figure 1B, left) and a chaotic regime (Figure 1B, right). Indeed, this resembles the optimal operating regime which is suggested based on the criticality hypothesis of the brain. Motivated by this qualitative observation, we perform quantitative analysis on networks with different noise amplitudes to examine them for signatures of criticality.

We investigate the scale-freeness of neural avalanches in the spiking activity of the networks as a potential signature of operating close to criticality. A neuronal avalanche is defined as an uninterrupted cascade of spikes propagating through the network [8]. In a system operating close to criticality, the distribution of avalanche sizes (number of spikes in the cascade)

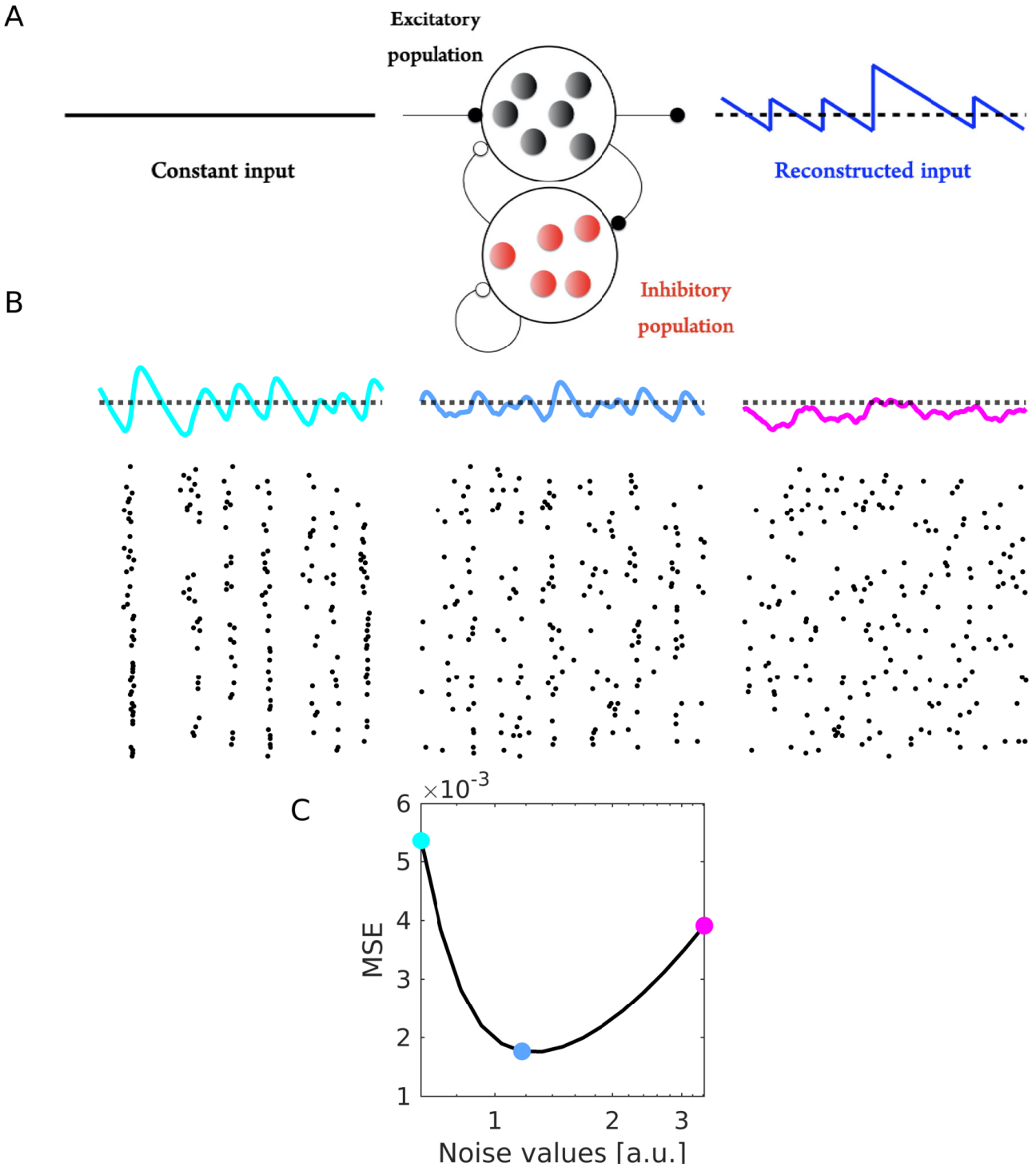


Figure 1. Illustration of efficient coding network optimization by adjusting the level of noise.
(A) One the left constant input to the network is illustrated; In the center a schematic representation of the efficient coding network is provided. Excitatory and inhibitory neurons are shown in black and red, respectively. Connections between different populations are represented by lines terminating with solid (for excitatory connections) and open (for inhibitory connections) circles. On the right, network's reconstruction of the input based on a linear readout of spikes is schematized with the blue line and the original input by a broken black line. **(B)** Each column demonstrates the output of a network with an exemplary choice of noise level. The leftmost column, an exemplary for the network with an insufficient amount of noise; The middle column, an exemplary for the network in vicinity of the optimal noise level; The rightmost column, an exemplary for the network with excessive amount of noise. Top row: Broken lines indicate the constant stimulus that network encodes. Solid lines indicate the network reconstruction of the stimulus (colors are matching with the colors of dots in C). Bottom row: Raster plots of spike trains. **(C)** Mean-squared-error(MSE) of the stimulus reconstruction for different level of noise. Spike trains and stimulus reconstruction corresponding to noise levels indicated with colorful dots exemplified in B.

follows a power-law distribution. To define the avalanches, we threshold the Inter-Spike Intervals (ISI) using an objective choice of a threshold i. e. average Inter-Event-Intervals (IEI) similar to [8] (for more details see Method section).

By analyzing the neural avalanches we observe that, in the vicinity of noise level optimized for efficient coding, neural avalanches detected in the spiking output of the network appear to be scale-free and when the noise amplitude is considerably lower or higher for efficient coding, the network appears either super-critical or sub-critical, respectively.

Furthermore, statistical analysis of neural avalanches confirm it to a large degree. In the vicinity of optimal regime for efficient coding (Figure 1B-center and C, Figure 2A-center and B-center), a power-law (PL) is the winner model, compare to log-normal (PL) (Figure 2 C-center) and exponential (exp) (Figure 2 D-center) distributions, according to their log-likelihood ratio (LLR). When the noise is considerably lower, the distribution of neural avalanches qualitatively resemble a super-critical regime, given the large bump at the tail of the distribution (Figure 2A-left and B-left) [1]. When the noise is considerably higher than the optimal noise, the distribution of neural avalanches qualitatively resemble a sub-critical regime, given it exponential-like distribution. Indeed, statistical analysis of neural avalanches also confirms that, in this regime an exponential distribution is a better model compare to other two (Figure 2C-right and D-right).

Gradual transition

We observed that by adjusting the noise level considerably different from the optimal level of the noise, coding efficacy of network degrades and the signatures of criticality disappear or become less pronounced. To investigate whether the nature of this transition is abrupt or continuous/gradual (*resembling* a first-order phase-transition or a second-order phase transition), we investigate the change of various quantities by incremental changes of noise. We investigate distribution of avalanche sizes, spike count ratio (branching ratio) and synchronization.

Similar to the procedure we used for Figure 1, we detect the neural avalanches in spiking activity of efficient coding networks with a wide ranges of noise levels, and we observe that by increasing the noise, large bump at the tail of avalanche distribution (which is typical in super-critical systems) gradually disappears and the distribution gradually follow a power-law. Interestingly, after crossing the the optimal value of noise, the avalanche distributions gradually change to an exponential distribution (Figure 3A). We also observe a similar behavior in the spike count ratio (branching ratio), which is another relevant measure of criticality [8]. Branching ratio gradually change from values less than 1 to larger than 1 (Figure 3C). Nevertheless, it should be noted that, in the critical value of noise the spike count ratio is still slightly lesser than 1 (which require further investigation).

Lastly we investigate how the synchronization in network varies by the level of noise. We quantify the degree of synchronization for each neuron by its phase-locking value (PLV) to the population firing rate (for more details see Method section). We observe that, synchronization in the network varies smoothly with noise level and the synchronization corresponding to the vicinity of the optimal value for efficient coding and criticality has intermediate values (Figure 3D) which qualitatively matches the result of Botcharova et al. [12]. Certainly, a more quantitative approach is needed to establish a better connection to result of Botcharova et al. [12].

Overall, distribution of neural avalanches, spike count ratio, and the synchronization suggest that there is a smooth transition. Nevertheless, to understand precisely the nature of this putative phase transition further investigation is necessary.

Materials and methods

Efficient coding network

The model was introduced and described extensively in previous studies by [11, 18], we thus restrict ourselves to a brief explanation of the key aspect of the model.

This network is optimized to encode a sensory input efficiently (i. e. with a minimal number of spikes) and accurately (i. e. with minimal reconstruction error). Network optimization objective is incorporated in the loss function $E(t)$,

$$E(t) = \left(x(t) - \widehat{x}(t)\right)^2 + \alpha \sum_i r_i(t) + \beta \sum_i r_i(t)^2 \,. \quad (1)$$

where $x(t)$ is a given sensory input, $\widehat{x}$ is the reconstructed sensory input, $r_i(t)$ is the firing rate of neuron i, and α and β are the weights of the $L1$ and $L2$ penalties on the firing rate.

It is assumed that the input can be reconstructed by performing a linear readout of the spike trains, more precisely, it is assumed that by a weighted leaky integration of output spike trains,

$$\tau \frac{d\hat{x}(t)}{dt} = -\hat{x}(t) + \sum_i w_i o_i(t) \,, \quad (2)$$

where o_i indicate the output spike trains for neuron i,

$$o_i(t) = \sum_k \delta\left(t - t_i^k\right) \,, \quad (3)$$

and τ is the read-out time constant [2], and w_i is a constant read-out weight associated to neuron i.

Given an idealized network with instantaneous synapses, the optimal network could be derived from first principles [11]. They showed that the dynamics of each Leaky-Integrate

[1] It should also be mentioned that, for some choices of insufficient noise level (considerably far from optimal value) power-law was the winner model as well, but only for a few values of x_{max}. Certainly this observation also need to be further investigated.

[2] In the efficient coding network used in this study (as in Chalk et al. (ref. 18)), for simplicity, the read-out time constant of the input (i. e. time-scale of $x(t)$) is the same as the time-constant of the membrane potential of the neurons. Nevertheless, in Boerlin et al. (ref. 11) they are not necessarily the same for more general computations.

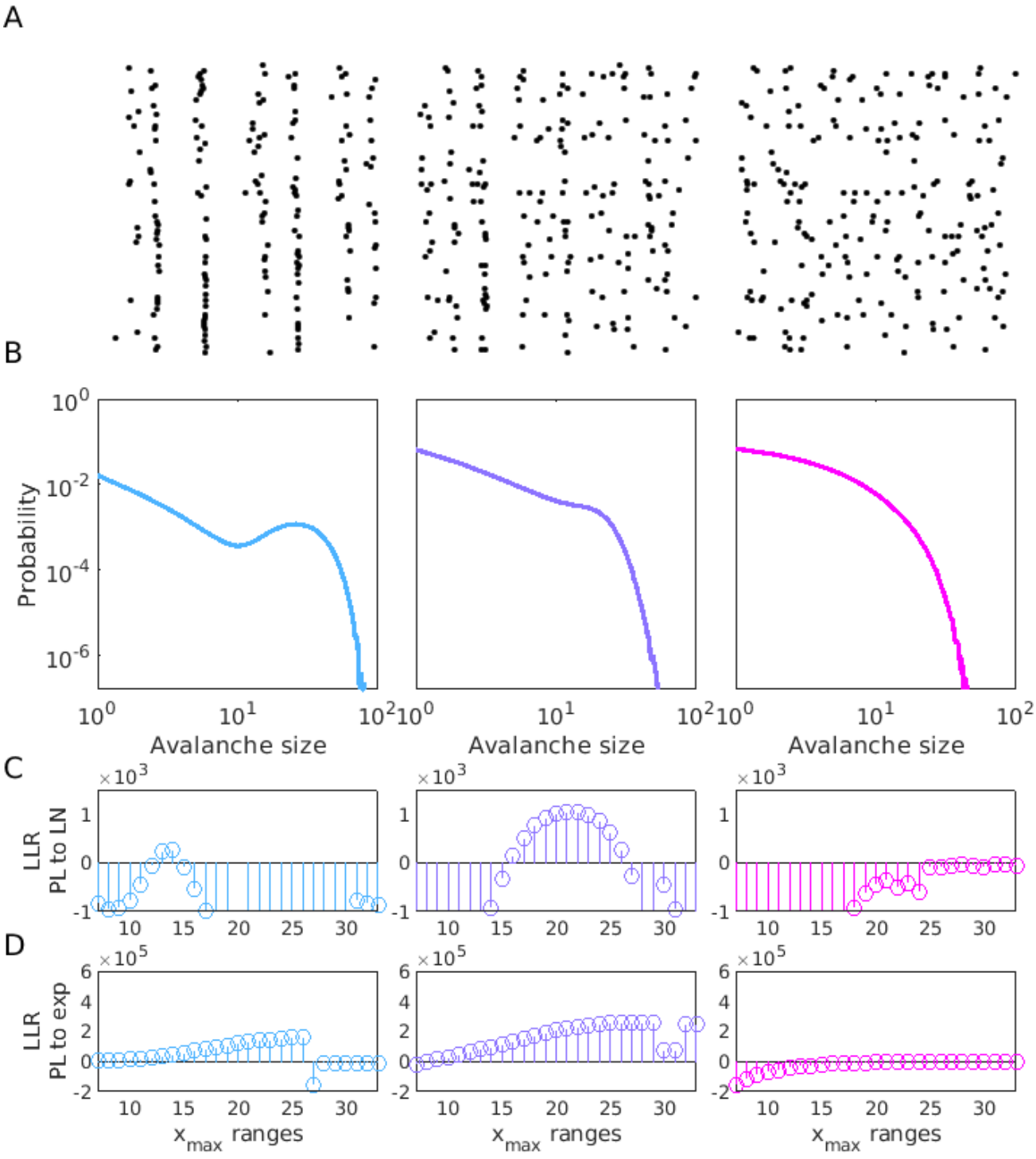


Figure 2. Illustration of neural avalanches in spiking activity of the efficient coding network for representative noise levels and their statistical analysis.
Each column pertains to a network with an exemplary choice of noise level. The leftmost column, an exemplary for the network with an insufficient amount of noise; The middle column, an exemplary for the network in vicinity of the optimal noise level; The rightmost column, an exemplary for the network with excessive amount of noise. **(A)** Raster plots of the spiking activity. **(B)** Distribution of the avalanche sizes. **(C-D)** Log-likelihood ratio (LLR) of fitted power-law (PL) distribution to log-normal (LN) distribution (C) and power-law (PL) distribution to exponential (exp) distribution (D) for different choices of x_{max} cut-off (x_{min} has been picked objectively based on procedure introduced by Clauset et al. [21]). In (C) and (D) the limits of the y-axis is chosen such that, the positive LLRs are better visible (to illustrate if/where the power-law distribution is the better model of the distribution compare to competitors).

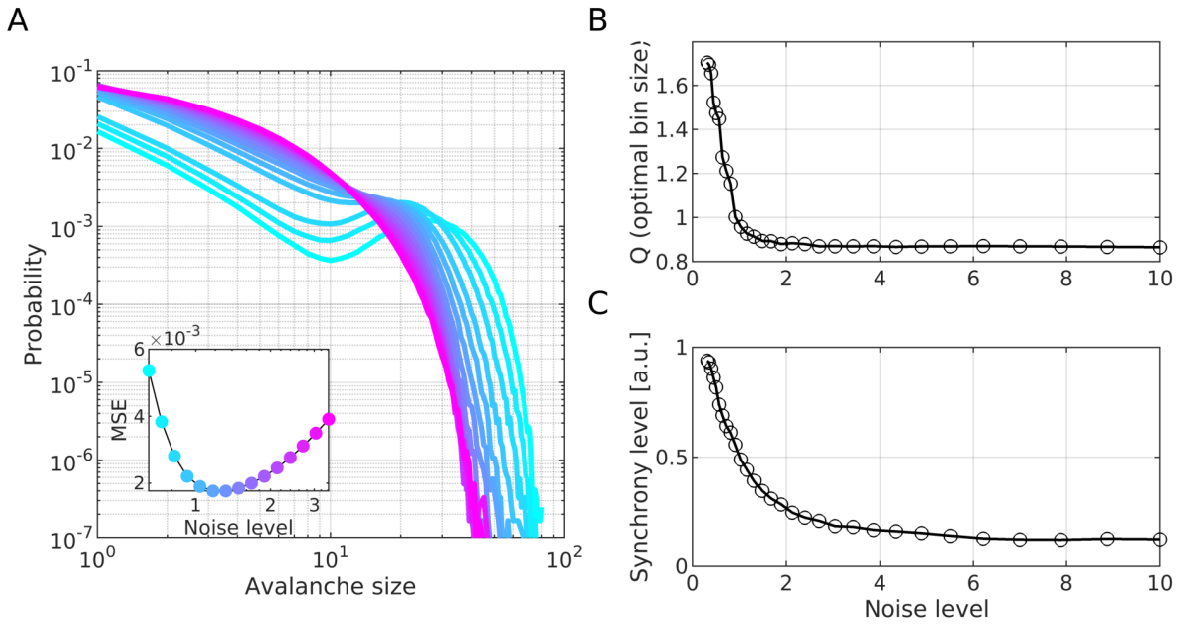


Figure 3. Illustration of gradual nature of the putative phase transition.
(A) Avalanche-size distributions in efficient coding networks with different noise levels. Inset: Mean-square-error (MSE) of stimulus reconstruction for different injected noise amplitudes. Colors of dots match the lines in the main figure. **(B)** The spike count ratio (Q) or branching ratio with optimal bin size (see the Method seciton) estimated from the spiking activity of the efficient coding network with different noise level indicated in the x-axis. **(C)** The synchrony level of the efficient coding network with different noise level.

and Fire (LIF) neuron can be expressed by conventional differential equation governing the dynamics of the membrane potentials,

$$\tau \frac{dV_i(t)}{dt} = -V_i(t) + w_i c(t) - w_i \sum_k w_k o_k(t) - \beta o_i(t) + \sigma \nu_i(t)\,, \tag{4}$$

where V_i is the membrane potential of neuron i, w_i is the constant readout which was introduced in Equation 2, $c(t)$ is the input to the network, $o_i(t)$ is the spike train of neuron i, β is the regularizer that was introduced in Equation 1, and $\nu(t)$ is a white noise with unit variance that was manually added in the original derivation of Boerlin et al. (ref. 11) for biological realism. Notably, in this network we have two types of input, a feed-forward input, $w_i c(t)$ and a recurrent input $-w_i \sum_k w_k o_k(t)$. The recurrent is result of a fully connected network. In this network, neurons that receive a common input, decorrelate their activity to avoid communicating redundant information via instantaneous recurrent inhibition.

Chalk et al. (ref. 18) introduce a more biologically plausible variants of Boerlin et al. [11] network by incoporating synaptic delays and introducing a balance network of inhibitory and exitatory population of neurons (50 neurons per each population). They incorporate realistic synaptic delays by assuming that each spike generates a continuous current input to other neurons, with dynamics described by the conventional alpha function,

$$h(t) = \begin{cases} \frac{1}{\tau_d - \tau_r}\left[e^{\frac{-(t-\tau_{tr})}{\tau_d}} - e^{\frac{-(t-\tau_{tr})}{\tau_r}}\right] & \text{if } t > \tau_{tr} \\ 0 & \text{if } t > \tau_{tr} \end{cases} \tag{5}$$

where τ_r and τ_d are respectively synaptic rise and decay times. Adding realistic synaptic delays, leads to network synchronization, which impairs coding efficiency. Chalk et al. (ref. 18) demonstrated that, in the presence of synaptic delays, this network of LIF neurons can nonetheless be optimized for efficient coding by adding noise to the network. In this study, we consider the additional noise, as white noise added to the membrane potentials. However, Chalk et al. (ref. 18) also demonstrate similar results by using other ways of incorporating noise (for instance, inducing unreliability in spike elicitation).

The original network introduced by Boerlin et al. [11] was a pure inhibitory network. Chalk et al. (ref. 18) introduced a variant of this network that respects the Dale's law. In their network, they introduce a population of inhibitory neurons that tracks the estimate encoded by the excitatory neurons, and provides recurrent feedback to the excitatory population (for further detail see Chalk et al. [18] and Boerlin et al. [11, Supernumerary material]).

Avalanche detection

In order to investigate the scale-free characteristic of the spiking activity (as a potential signature of networks operating close to criticality), similar to previous studies [8], we probe into neural avalanches. A neuronal avalanche is considered as an uninterrupted cascade of spiking activity propagating through the network [8]. In a system operating close to criticality, the distribution of avalanche sizes (number of spikes in a cascade) and avalanche life time follows a power-law (nevertheless, in this letter we have only investigated the distribution of avalanche sizes).

To define the avalanches, we first define a "population spike train". We have a binary spike train per each neuron in the network. In each time step of the simulation (which is chosen to be 0.5 msec) we can either have a spike (binary assignment of 1) or not (binary assignment of 0). Therefore, we have such binary sequences for every neurons in the network. We construct the population spike train (denoted by $o_p(t)$) by collapsing the spike trains of all neurons into a single spike train. For each time bin, if at least one spike was elicited across the entire population, the given bin was assigned to 1 otherwise 0 in the population spike train. Should be noted that, the mentioned binarization was an inherent part of the simulation, rather an additional binning of the spike trains.

The average firing rate of the population spike train ($\bar{r}_p$) defined as,

$$\bar{r}_p = \frac{1}{T}\int_0^T o_p(t)\,, \tag{6}$$

where T is the length of the population spike train (number of samples in the simulation). The inverse of $\bar{r}_p$ determines the Δ that was used to separate the avalanches.

Statistical analysis of avalanches

For assessing the goodness of power-law fit to neural avalanches, similar to recent studies [37, 91], we use the likelihood ratio (LLR) of power-law to other distributions. We use the `powerlaw` package developed by Alstott et al. [6] that was established based on methodology introduced by Clauset et al. [21].

We computed the LLR between a power-law and a log-normal (that is the strongest competitor model) and a exponential distribution for a given collection of avalanche sizes. If the power-law was the winner model in both model compressions: power-law vs. exponetitnal distribution and power-law vs. log-normal distribution, we considered power-law a better model, otherwise not. We use the significance threshold of 0.05 for both of the comparisons.

The cut-offs (lower and upper bounds, x_{min} and x_{max} respectively) for fitting the power-laws was chosen according to the size of each network. The lower bound, x_{min}, was chosen objectively based on the procedure described in Clauset et al. [21], and for the upper bound, x_{max}, we initially chose a value based visual inspection and sweep the range of values around the subjective choice to assure that the results is not sensitive to this choice.

Quantifying the synchronization

For quantifying the synchronization in the network, for each neuron we measure the phase-locking value (PLV) of spiking activity of each neuron to the population firing rate. In order to compute the PLV, first we high-pass filter the population firing rate with cut-off 3 Hz (similar to Chalk et al. [18]), then band-pass it with width of 10 Hz around the peak frequency of the spectrum of population firing rate, i. e. $f_{max} \pm 5$ Hz (peak frequency, varies with the noise). By applying the Hilbert transform, we extract the phase of population firing rate $\phi(t)$ and then PLV can be calculated as follows:

$$PLV = \frac{1}{N^{tot}}\sum_{k=1}^{K}\sum_{j=1}^{N^{(k)}} \exp\left(\boldsymbol{i}\phi^{(k)}_{t_j^{(k)}}\right)\,, \tag{7}$$

where, $\{t_j^{(k)}\}_{j=1\ldots N^{(k)}}$ indicate the spike times of the spike train, $\boldsymbol{i}$ is the imaginary unit ($\boldsymbol{i}^2 = -1$), and $N^{(k)}$ is the number of spikes occurring during the trial k (if multiple trials is used), N_{tot} is the total number of spikes occurred across all trials, i. e.

$$N^{tot} = \sum_{k=1}^{K} N^{(k)}\,. \tag{8}$$

Average PLV across all neurons has been used in Figure 3C as a measure of synchronization in the network.

Quantifying the spike count ratio

We compute the spike count ratio (Q) or branching ratio as it was used in the previous papers (e. g. Priesemann and Shriki [71]),

$$Q(\Delta t) = \left\langle \frac{A(i|\Delta t)}{A(i-1|\Delta t)} \right\rangle \tag{9}$$

A is the population activity i. e. spike trains or sequence of spike counts (if Δt is larger than 1) summed across all units of the network. Note that the average is over all bins with $A(t = i-1|\Delta t) > 0$.

Discussion

In this study we introduce a new approach to better connect neural dynamics and neural computation. Here we search for potential connection between models of neural dynamics with implication about neural computation, and normative models of neural computation with implication on neural dynamics. More specifically in this study, we search for signatures of criticality in a neuronal network that can be optimized based on the functionally relevant computational objectives of efficient coding. We investigate a network of LIF neurons whose connectivity and dynamics can be optimized for coding a one-dimensional input [18]. As the performance of the network varies non-monotonically with the noise amplitude, we investigate the presenc of the signatures of criticality in the network. Interestingly, in the vicinity of parameters lead to optimized network for coding, we observe the signatures of criticality (scale-free neural avalanches).

Criticality and efficient coding co-occur?

Our study pose an important question pertaining neural computation and dynamics, which is whether coding-based optimally *mechanistically* co-occurs with closeness to criticality? Our observation suggests, they might, but need to be further investigated. Nevertheless, we believe both negative and positive answers to this question, lead to interesting and important research directions.

A negative answer

If it turns out that, all signatures of criticality that we can observe in this model, are simply a byproduct of efficient coding, raise a new set of questions that, how and why this particular implementation of efficient coding lead to scale-free dynamics? Is it specific to this particular implementation or it is rather a general property of efficient coding networks? Does scale-free dynamics is also functionally relevant for coding efficiently? One of the directions that particularly deserve further investigation, is the functional role of neural avalanches in this network (regardless of their relationship to criticality). This is an interesting question, because of the relationship between neural avalanches and the reconstruction error. In the network in our study, size of the avalanches should be related to degree of the corrected errors (given that the network elicit spikes only when it leads to lesser error - baring the spikes elicited due to noise). Therefore, it is important to ask, does any attributes of reconstruction error should follow a scale-free distribution, and if yes, how it can be interpreted in terms of optimality? Furthermore, two lines of research could be related to this question.

First, it has been suggested that neural avalanches can reflect transient formation of cell assemblies [69] and cell assemblies has been suggested as an important neural substrate of neural computation [17, 41, 68]. This guide us to the question that, does cell assemblies represent some form reconstruction error? Answer to this question, might guide us toward a functional interpretation of neural avalanches. Second, it has been suggested that "Statistical criticality, i. e. the occurrence of power law frequency distributions, arises in samples that are maximally informative about the underlying generating process" [22] (also see [23, 40, 56, 82]). Interestingly, the same principle has been suggested for representation in deep neural network [81]. These studies show, efficient representation co-occur with some form of scale-freeness. Nevertheless, it should be noted that, the scale-freeness pointed out by Cubero et al. [22] is on a quantity different from what is typically used as neural avalanche [8]. Cubero et al. [22] consider the spiking patterns (within a fixed length) which occur k times and find the multiplicity of such patterns m_k, and they analytically showed that m_k as a function of k should follow a characteristic power-law. Perhaps, the most important difference is, neural avalanches are variable spatio-temporal patterns, whereas Cubero et al. [22] focus on a fixed temporal window (or fixed temporal windows if one consider multiple lengths [24])

A positive answer

Positive answer to this question establish a first connection between two of the important and principled frameworks suggested for optimized information processing in the brain (efficient coding and criticality). Furthermore, it also suggests a complementary approach to criticality. The common approach for evaluating and demonstrating how closeness to criticality can be beneficial for information processing in the brain, is using neural models (e. g. a branching networks, a recurrent neural networks) that can attain various states (including critical and non-critical state), and investigate how general information processing capabilities (e. g. sensitivity to input), depends on the control parameters (e. g. branching ratio, connection strength). Certainly, this approach has been insightful for criticality hypothesis of the brain, Nevertheless, the *optimized setting* implied by criticality hypothesis, does not imply a specific computation that brain needs to execute, but rather general capabilities for computation or computational primitives [3]. For instance, in this state we have the maximum sensitivity to input [15, 48], and maximum dynamic range [48, 50, 67]. Certainly such capabilities are relevant for coding sensory information and therefore it is important to be optimized, but mere adjusting for the closeness to criticality cannot provide a neural coding algorithm and its implementation for coding the sensory information given the resource constraints. Therefore, the approach we suggest in this study can complement criticality hypothesis, by connecting better the criticality hypothesis of the brain to specific neural computations.

Furthermore, in this study, beside the criticality signatures that we quantified in our efficient coding network, some aspect of our efficient coding network also match what has been associated to brain criticality. (1) The network seems to be operating between an oscillatory and a non-oscillatory regime which was also attributed to brain criticality [76]. This already suggests that bifurcation analysis might shed light on the mechanisms of [persumed] criticality in this network (also see [45] for a similar analysis in a mean-field model). (2) It has been suggested the E-I balance can be a neuronal control parameter to tune the neural circuits close to criticality [79], and indeed in this network E-I balance is one of the key factors for efficiency of the coding [30]. Nevertheless, whether having a loose or tight balance [3] matters for the criticality need to be investigated in the future studies.

Lastly, it should also be mentioned that, in spite of solid recent experimental evidence supporting the criticality hypothesis of the brain [37, 91], some of the phenomenology which has been suggested as a potential signature of criticality in the brain could have more mundane explanations; for scale-freeness of rank-frequency distributions [59, 77], for scale-freeness of neural avalanches [4, 86, 87], for thermodynamics approach to criticality [66], and for the approach of

[3] See also Lizier [52] (in particular chapter 6) which argue that closeness to criticality is a sate where [some] computing primitives (such as information storage, transfer and modification) are optimized.

renormalization group [64] [4] . Certainly, one needs to thoroughly check if any of the mentioned mechanisms has been led to signature of criticality that has been observed in our study.

Future directions

Regardless of positive or negative answer to the question we raised, our study report an important observation, and therefore motivates broad explorations that cross the conventional boundaries in computational neuroscience. Our observations encourage normative models to engage with more realistic dynamics (which the endeavor has been already initiated [10, 11, 16, 18, 28, 29, 32, 51, 84, 92]); Also, it encourages models of neural dynamics to engage with explicit neural computation objectives (where the endeavor has been already started in this domain as well [9, 19, 33, 36, 42, 47, 54, 62, 80, 83]). In the following paragraphs a few directions are briefly discussed.

Regarding neural dynamics, scale-freeness that has been observed in this efficient coding network has been an intriguing aspect of networks and their dynamics in various natural phenomena [7] [5], not only due to its association to criticality that might be controversial, but it is also important for brain pathology. Indeed, recent studies suggest that pharmacological interventions effect such dynamics [60, 61, 78] (also see for a broad review Zimmern [93]). Therefore, it might be crucial to understand the underlying mechanism of brain scale-freeness even irrespective of criticality. Whether the brain is operating close to a critical state or not, and criticality-based optimality co-occurs with the optimality in the sense of efficient coding, in this study, we show that a richer dynamics of the brain –scale-freeness– is a signatures of efficient coding as well. Overall, we believe it is justified to suggest, efficient coding can also be a promising framework for deeper investigation of neural dynamics. Therefore, efficient coding is a suitable candidate of a normative model to explore for more realistic dynamics of the brain.

Furthermore, we believe our observations opens up several promising avenues for future research toward the integration of neural computation and neural dynamics in a single framework. In particular, new directions should intend to fill the gaps that this framework did not explicitly answer. For instance our approach show evidence on *potential* connections between brain criticality and efficient coding, but is not necessarily informative on a mechanistic level. For instance, in this particular case, one interesting direction is using information-theoretic measures like Fisher information. Fisher information can be a candidate quantity that both framwork — efficient coding [89] and criticality [25, 46, 49, 72] — use to assess to closeness to their optimal point. This can be a starting point for more formal approaches.

[4] Should be also noted that, for the concerns raised by Aitchison et al. [4], Johnson et al. [44] and Yu et al. [90] proposed solution and Bowen et al. [13] suggest that presence of correlation between neurons rule out the models like the one proposed by Martinello et al. [57] and Touboul and Destexhe [87].

[5] But also see Holme [43] and references therein for a different perspective). In particular scale-free dynamics is an important aspect of brain dynamics [38, 70]

References

[1] Abbott, L. F. “Theoretical Neuroscience Rising”. English. In: *Neuron* 60.3 (2008), pp. 489–495 (cit. on p. 1).

[2] Agrawal, Vidit, Chakraborty, Srimoy, Knöpfel, Thomas, and Shew, Woodrow L. “Scale-Change Symmetry in the Rules Governing Neural Systems”. In: *iScience* 12 (2019), pp. 121–131 (cit. on p. 2).

[3] Ahmadian, Yashar and Miller, Kenneth D. “What Is the Dynamical Regime of Cerebral Cortex?” In: *ArXiv190810101 Q-Bio* (2019) (cit. on p. 8).

[4] Aitchison, Laurence, Corradi, Nicola, and Latham, Peter E. “Zipf’s Law Arises Naturally When There Are Underlying, Unobserved Variables”. en. In: *PLOS Computational Biology* 12.12 (2016), e1005110 (cit. on pp. 8, 9).

[5] Alamia, Andrea and VanRullen, Rufin. “Alpha Oscillations and Traveling Waves: Signatures of Predictive Coding?” en. In: *PLOS Biology* 17.10 (2019), e3000487 (cit. on pp. 1, 2).

[6] Alstott, Jeff, Bullmore, Ed, and Plenz, Dietmar. “Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions”. In: *PLoS ONE* 9.1 (2014), e85777 (cit. on p. 7).

[7] Barabási, Albert-László and Albert, Réka. “Emergence of Scaling in Random Networks”. en. In: *Science* 286.5439 (1999), pp. 509–512 (cit. on p. 9).

[8] Beggs, J. M. and Plenz, D. “Neuronal Avalanches in Neocortical Circuits”. In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23 (2003), pp. 11167–77 (cit. on pp. 2, 4, 7, 8).

[9] Bertschinger, N. and Natschlager, T. “Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks”. In: *Neural computation* 16 (2004), pp. 1413–1436 (cit. on p. 9).

[10] Bill, Johannes, Buesing, Lars, Habenschuss, Stefan, Nessler, Bernhard, Maass, Wolfgang, and Legenstein, Robert. “Distributed Bayesian Computation and Self-Organized Learning in Sheets of Spiking Neurons with Local Lateral Inhibition”. en. In: *PLOS ONE* 10.8 (2015), e0134356 (cit. on pp. 1, 9).

[11] Boerlin, M., Machens, C. K., and Deneve, S. “Predictive Coding of Dynamical Variables in Balanced Spiking Networks”. In: *PLoS computational biology* 9 (2013), e1003258 (cit. on pp. 1, 2, 4, 6, 9).

[12] Botcharova, M., Farmer, S. F., and Berthouze, L. “Markers of Criticality in Phase Synchronization”. English. In: *Frontiers in systems neuroscience* 8 (2014), p. 176 (cit. on p. 4).

[13] Bowen, Zac, Winkowski, Daniel E., Seshadri, Saurav, Plenz, Dietmar, and Kanold, Patrick O. "Neuronal Avalanches in Input and Associative Layers of Auditory Cortex". en. In: *bioRxiv* (2019), p. 620781 (cit. on p. 9).

[14] Breakspear, M. "Dynamic Models of Large-Scale Brain Activity". en. In: *Nature neuroscience* 20.3 (2017), pp. 340–352 (cit. on p. 2).

[15] Brochini, Ludmila, de Andrade Costa, Ariadne, Abadi, Miguel, Roque, Antônio C., Stolfi, Jorge, and Kinouchi, Osame. "Phase Transitions and Self-Organized Criticality in Networks of Stochastic Spiking Neurons". en. In: *Sci. Rep.* 6 (2016), p. 35831 (cit. on pp. 1, 8).

[16] Buesing, L., Bill, J., Nessler, B., and Maass, W. "Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons". In: *PLoS computational biology* 7 (2011), e1002211 (cit. on pp. 1, 9).

[17] Buzsaki, G. "Neural Syntax: Cell Assemblies, Synapsembles, and Readers". In: *Neuron* 68.3 (2010), pp. 362–85 (cit. on p. 8).

[18] Chalk, M., Gutkin, B., and Deneve, S. "Neural Oscillations as a Signature of Efficient Coding in the Presence of Synaptic Delays". In: *eLife* 5 (2016) (cit. on pp. 1, 2, 4, 6, 7, 9).

[19] Chen, Guozhang and Gong, Pulin. "Computing by Modulating Spontaneous Cortical Activity Patterns as a Mechanism of Active Visual Processing". en. In: *Nat Commun* 10.1 (2019), pp. 1–15 (cit. on p. 9).

[20] Churchland, Patricia Smith and Sejnowski, Terrence J. *The Computational Brain*. Computational Neuroscience. Cambridge, Mass: MIT Press, 1992 (cit. on p. 1).

[21] Clauset, Aaron., Shalizi, Cosma Rohilla., and Newman, M. E. J. "Power-Law Distributions in Empirical Data". In: *SIAM Rev.* 51.4 (2009), pp. 661–703 (cit. on pp. 5, 7).

[22] Cubero, Ryan John, Jo, Junghyo, Marsili, Matteo, Roudi, Yasser, and Song, Juyong. "Statistical Criticality Arises in Most Informative Representations". en. In: *J. Stat. Mech.* 2019.6 (2019), p. 063402 (cit. on p. 8).

[23] Cubero, Ryan John, Marsili, Matteo, and Roudi, Yasser. "Minimum Description Length Codes Are Critical". en. In: *Entropy* 20.10 (2018), p. 755 (cit. on p. 8).

[24] Cubero, Ryan John, Marsili, Matteo, and Roudi, Yasser. "Multiscale Relevance and Informative Encoding in Neuronal Spike Trains". en. In: *J Comput Neurosci* 48.1 (2020), pp. 85–102 (cit. on p. 8).

[25] Daniels, B. C., Ellison, C. J., Krakauer, D. C., and Flack, J. C. "Quantifying Collectivity". In: *Curr Opin Neurobiol* 37 (2016), pp. 106–113 (cit. on p. 9).

[26] Dayan, Peter and Abbott, L. F. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Computational Neuroscience. Cambridge, Mass: Massachusetts Institute of Technology Press, 2001 (cit. on p. 1).

[27] Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. "The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields". In: *PLoS computational biology* 4.8 (2008), e1000092 (cit. on p. 2).

[28] Deneve, S. "Bayesian Spiking Neurons I: Inference". en. In: *Neural computation* 20 (2008), pp. 91–117 (cit. on p. 9).

[29] Deneve, S. "Bayesian Spiking Neurons II: Learning". en. In: *Neural computation* 20 (2008), pp. 118–45 (cit. on p. 9).

[30] Deneve, S. and Machens, C. K. "Efficient Codes and Balanced Networks". In: *Nature neuroscience* 19 (2016), pp. 375–82 (cit. on p. 8).

[31] Dubreuil, Alexis M., Valente, Adrian, Beiran, Manuel, Mastrogiuseppe, Francesca, and Ostojic, Srdjan. "Complementary Roles of Dimensionality and Population Structure in Neural Computations". en. In: *bioRxiv* (2020), p. 2020.07.03.185942 (cit. on p. 1).

[32] Echeveste, Rodrigo, Aitchison, Laurence, Hennequin, Guillaume, and Lengyel, Máté. "Cortical-like Dynamics in Recurrent Circuits Optimized for Sampling-Based Probabilistic Inference". en. In: *Nat. Neurosci.* 23.9 (2020), pp. 1138–1149 (cit. on pp. 1, 2, 9).

[33] Eliasmith, Chris. "A Unified Approach to Building and Controlling Spiking Attractor Networks". In: *Neural Computation* 17.6 (2005), pp. 1276–1314 (cit. on p. 9).

[34] Ermentrout, G. Bard, Galán, Roberto F., and Urban, Nathaniel N. "Relating Neural Dynamics to Neural Coding". In: *Phys. Rev. Lett.* 99.24 (2007), p. 248103 (cit. on pp. 1, 2).

[35] Eurich, C. W. *Neural Dynamics and Neural Coding Two Complementary Approaches*. Tech. rep. 2003 (cit. on p. 1).

[36] Finlinson, Kathleen, Shew, Woodrow L., Larremore, Daniel B., and Restrepo, Juan G. "Optimal Control of Excitable Systems near Criticality". In: *Phys. Rev. Research* 2.3 (2020), p. 033450 (cit. on p. 9).

[37] Fontenele, Antonio J., de Vasconcelos, Nivaldo A. P., Feliciano, Thaís, Aguiar, Leandro A. A., Soares-Cunha, Carina, Coimbra, Bárbara, Dalla Porta, Leonardo, Ribeiro, Sidarta, Rodrigues, Ana João, Sousa, Nuno, Carelli, Pedro V., and Copelli, Mauro. "Criticality between Cortical States". In: *Phys. Rev. Lett.* 122.20 (2019), p. 208101 (cit. on pp. 7, 8).

[38] Freeman, Walter J. and Breakspear, Michael. "Scale-Free Neocortical Dynamics". en. In: *Scholarpedia* 2.2 (2007), p. 1357 (cit. on pp. 2, 9).

[39] Gerstner, Wulfram, Kistler, Werner M., Naud, Richard, and Paninsk, Liam. *Neuronal Dynamics, From Single Neurons to Networks and Models of Cognition*. University Printing House, Cambridge CB2 8BS, United Kingdom: Cambridge University Press, 2014 (cit. on p. 1).

[40] Haimovici, Ariel and Marsili, Matteo. "Criticality of Mostly Informative Samples: A Bayesian Model Selection Approach". en. In: *J. Stat. Mech.* 2015.10 (2015), P10013 (cit. on p. 8).

[41] Harris, K. D. "Neural Signatures of Cell Assembly Organization". en. In: *Nature reviews. Neuroscience* 6.5 (2005), pp. 399–407 (cit. on p. 8).

[42] Hidalgo, J., Grilli, J., Suweis, S., Munoz, M. A., Banavar, J. R., and Maritan, A. "Information-Based Fitness and the Emergence of Criticality in Living Systems". In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (2014), pp. 10095–100 (cit. on p. 9).

[43] Holme, Petter. "Rare and Everywhere: Perspectives on Scale-Free Networks". en. In: *Nat. Commun.* 10.1 (2019), p. 1016 (cit. on p. 9).

[44] Johnson, James Kenneth, Wright, Nathaniel C., Xia, Ji, and Wessel, Ralf. "Single-Cell Membrane Potential Fluctuations Evince Network Scale-Freeness and Quasicriticality". en. In: *bioRxiv* (2018), p. 498477 (cit. on p. 9).

[45] Kadmon, Jonathan, Timcheck, Jonathan, and Ganguli, Surya. "Predictive Coding in Balanced Neural Networks with Noise, Chaos and Delays". In: *ArXiv200614178 Cond-Mat Q-Bio Stat* (2020) (cit. on pp. 2, 8).

[46] Kalloniatis, Alexander C., Zuparic, Mathew L., and Prokopenko, Mikhail. "Fisher Information and Criticality in the Kuramoto Model of Nonidentical Oscillators". In: *Phys. Rev. E* 98.2 (2018), p. 022302 (cit. on p. 9).

[47] Kim, Christopher M and Chow, Carson C. "Learning Recurrent Dynamics in Spiking Networks". In: *eLife* 7 (2018), e37124 (cit. on p. 9).

[48] Kinouchi, O. and Copelli, M. "Optimal Dynamical Range of Excitable Networks at Criticality". English. In: *Nat Phys* 2 (2006), pp. 348–352 (cit. on pp. 1, 8).

[49] Kuebler, Eric S., Calderini, Matias, Lambert, Philippe, and Thivierge, Jean-Philippe. "Optimal Fisher Decoding of Neural Activity Near Criticality". en. In: *The Functional Role of Critical Dynamics in Neural Systems*. Ed. by Nergis Tomen, J. Michael Herrmann, and Udo Ernst. Springer Series on Bio- and Neurosystems. Cham: Springer International Publishing, 2019, pp. 159–177 (cit. on p. 9).

[50] Larremore, Daniel B., Shew, Woodrow L., and Restrepo, Juan G. "Predicting Criticality and Dynamic Range in Complex Networks: Effects of Topology". In: *Phys. Rev. Lett.* 106.5 (2011), p. 058101 (cit. on pp. 1, 8).

[51] Lengyel, Máté, Kwag, Jeehyun, Paulsen, Ole, and Dayan, Peter. "Matching Storage and Recall: Hippocampal Spike Timing–Dependent Plasticity and Phase Response Curves". en. In: *Nat. Neurosci.* 8.12 (2005), pp. 1677–1683 (cit. on pp. 1, 9).

[52] Lizier, Joseph T. *The Local Information Dynamics of Distributed Computation in Complex Systems*. eng. Springer Theses. Berlin: Springer, 2013 (cit. on p. 8).

[53] Lukovic, M., Vanni, F., Svenkeson, A., and Grigolini, P. "Transmission of Information at Criticality". English. In: *Physica A* 416 (2014), pp. 430–438 (cit. on p. 1).

[54] Maass, W. "Searching for Principles of Brain Computation". English. In: *Curr. Opin. Behav. Sci.* 11 (2016), pp. 81–92 (cit. on p. 9).

[55] Marinazzo, D., Pellicoro, M., Wu, G., Angelini, L., Cortes, J. M., and Stramaglia, S. "Information Transfer and Criticality in the Ising Model on the Human Connectome". In: *PloS one* 9 (2014), e93616 (cit. on p. 1).

[56] Marsili, Matteo, Mastromatteo, Iacopo, and Roudi, Yasser. "On Sampling and Modeling Complex Systems". English. In: *J Stat Mech-Theory E* 2013 (2013), P09003 (cit. on p. 8).

[57] Martinello, Matteo, Hidalgo, Jorge, di Santo, Serena, Maritan, Amos, Plenz, Dietmar, and Muñoz, Miguel A. "Neutral Theory and Scale-Free Neural Dynamics". In: *arXiv* (2017) (cit. on p. 9).

[58] Mastrogiuseppe, Francesca and Ostojic, Srdjan. "Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks". English. In: *Neuron* 99.3 (2018), 609–623.e29 (cit. on p. 1).

[59] Mastromatteo, Iacopo and Marsili, Matteo. "On the Criticality of Inferred Models". en. In: *J. Stat. Mech.* 2011.10 (2011), P10012 (cit. on p. 8).

[60] Meisel, Christian. "Antiepileptic Drugs Induce Subcritical Dynamics in Human Cortical Networks". en. In: *PNAS* 117.20 (2020), pp. 11118–11125 (cit. on p. 9).

[61] Meisel, Christian, Schulze-Bonhage, Andreas, Freestone, Dean, Cook, Mark James, Achermann, Peter, and Plenz, Dietmar. "Intrinsic Excitability Measures Track Antiepileptic Drug Action and Uncover Increasing/Decreasing Excitability over the Wake/Sleep Cycle". en. In: *PNAS* 112.47 (2015), pp. 14694–14699 (cit. on p. 9).

[62] Michiels van Kessenich, L., Berger, D., de Arcangelis, L., and Herrmann, H. J. "Pattern Recognition with Neuronal Avalanche Dynamics". In: *Phys. Rev. E* 99.1 (2019), p. 010302 (cit. on p. 9).

[63] Mora, Thierry and Bialek, William. "Are Biological Systems Poised at Criticality?" In: *J Stat Phys* 144 (2011), pp. 268–302 (cit. on pp. 1, 2).

[64] Morrell, Mia C., Sederberg, Audrey J., and Nemenman, Ilya. "Latent Dynamical Variables Produce Signatures of Spatiotemporal Criticality in Large Biological Systems". In: *ArXiv200804435 Q-Bio* (2020) (cit. on p. 9).

[65] Muñoz, Miguel A. "Colloquium: Criticality and Dynamical Scaling in Living Systems". In: *Rev. Mod. Phys.* 90.3 (2018), p. 031001 (cit. on pp. 1, 2).

[66] Nonnenmacher, Marcel, Behrens, Christian, Berens, Philipp, Bethge, Matthias, and Macke, Jakob H. "Signatures of Criticality Arise from Random Subsampling in Simple Population Models". en. In: *PLOS Computational Biology* 13.10 (2017), e1005718 (cit. on p. 8).

[67] Nur, Tazima, Gautam, Shree Hari, Stenken, Julie A., and Shew, Woodrow L. "Probing Spatial Inhomogeneity of Cholinergic Changes in Cortical State in Rat". En. In: *Sci. Rep.* 9.1 (2019), p. 9387 (cit. on pp. 1, 8).

[68] Papadimitriou, Christos H., Vempala, Santosh S., Mitropolsky, Daniel, Collins, Michael, and Maass, Wolfgang. "Brain Computation by Assemblies of Neurons". en. In: *PNAS* (2020) (cit. on p. 8).

[69] Plenz, D. and Thiagarajan, T. C. "The Organizing Principles of Neuronal Avalanches: Cell Assemblies in the Cortex?" English. In: *Trends in neurosciences* 30 (2007), pp. 101–10 (cit. on p. 8).

[70] Dietmar Plenz and Ernst Niebur, eds. *Criticality in Neural Systems*. eng. 1. edition. Reviews of Nonlinear Dynamics and Complexity. Weinheim: Wiley-VCH verlag GmbH & Co. KGaA, 2014 (cit. on p. 9).

[71] Priesemann, Viola and Shriki, Oren. "Can a Time Varying External Drive Give Rise to Apparent Criticality in Neural Systems?" en. In: *PLOS Computational Biology* 14.5 (2018), e1006081 (cit. on p. 7).

[72] Prokopenko, Mikhail, Lizier, Joseph T., Obst, Oliver, and Wang, X. Rosalind. "Relating Fisher Information to Order Parameters". In: *Phys. Rev. E* 84.4 (2011), p. 041116 (cit. on p. 9).

[73] Rodrigo Quian Quiroga and Stefano Panzeri, eds. *Principles of Neural Coding*. Boca Raton: CRC Press, 2013 (cit. on p. 2).

[74] Rieke, Fred, Warland, David, Rob de de Ruyter van Steveninck, and Bialek, William. *Spikes: Exploring the Neural Code*. A Bradford Book, 1999 (cit. on p. 2).

[75] Roeth, Kai, Shao, Shuai, and Gjorgjieva, Julijana. "Efficient Population Coding Depends on Stimulus Convergence and Source of Noise". en. In: *bioRxiv* (2020), p. 2020.06.15.151795 (cit. on p. 2).

[76] Santo, Serena di, Villegas, Pablo, Burioni, Raffaella, and Muñoz, Miguel A. "Landau–Ginzburg Theory of Cortex Dynamics: Scale-Free Avalanches Emerge at the Edge of Synchronization". en. In: *PNAS* (2018), p. 201712989 (cit. on p. 8).

[77] Schwab, D. J., Nemenman, I., and Mehta, P. "Zipf's Law and Criticality in Multivariate Data without Fine-Tuning". English. In: *Physical review letters* 113.6 (2014) (cit. on p. 8).

[78] Seshadri, Saurav, Klaus, Andreas, Winkowski, Daniel E., Kanold, Patrick O., and Plenz, Dietmar. "Altered Avalanche Dynamics in a Developmental NMDAR Hypofunction Model of Cognitive Impairment". En. In: *Transl. Psychiatry* 8.1 (2018), p. 3 (cit. on p. 9).

[79] Shew, W. L., Yang, H., Yu, S., Roy, R., and Plenz, D. "Information Capacity and Transmission Are Maximized in Balanced Cortical Networks with Neuronal Avalanches". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.1 (2011), pp. 55–63 (cit. on pp. 1, 8).

[80] Shriki, Oren and Yellin, Dovi. "Optimal Information Representation and Criticality in an Adaptive Sensory Recurrent Neuronal Network". en. In: *PLOS Computational Biology* 12.2 (2016), e1004698 (cit. on pp. 2, 9).

[81] Song, Juyong, Marsili, Matteo, and Jo, Junghyo. "Resolution and Relevance Trade-Offs in Deep Learning". en. In: *J. Stat. Mech.* 2018.12 (2018), p. 123406 (cit. on p. 8).

[82] Sorbaro, Martino, Herrmann, J. Michael, and Hennig, Matthias. "Statistical Models of Neural Activity, Criticality, and Zipf's Law". en. In: *The Functional Role of Critical Dynamics in Neural Systems*. Ed. by Nergis Tomen, J. Michael Herrmann, and Udo Ernst. Springer Series on Bio- and Neurosystems. Cham: Springer International Publishing, 2019, pp. 265–287 (cit. on p. 8).

[83] Sussillo, D. "Neural Circuits as Computational Dynamical Systems". English. In: *Curr Opin Neurobiol* 25 (2014), pp. 156–63 (cit. on p. 9).

[84] Tanaka, Takuma, Kaneko, Takeshi, and Aoyagi, Toshio. "Recurrent Infomax Generates Cell Assemblies, Neuronal Avalanches, and Simple Cell-Like Selectivity". In: *Neural Computation* 21.4 (2008), pp. 1038–1067 (cit. on p. 9).

[85] Tkacik, G. and Bialek, W. "Information Processing in Living Systems". English. In: *Annu Rev Conden Ma P* 7 (2016), pp. 89–117 (cit. on pp. 1, 2).

[86] Touboul, J. and Destexhe, A. “Can Power-Law Scaling and Neuronal Avalanches Arise from Stochastic Dynamics?” English. In: *PloS one* 5 (2010), e8982 (cit. on p. 8).

[87] Touboul, Jonathan and Destexhe, Alain. “Power-Law Statistics and Universal Scaling in the Absence of Criticality”. In: *Phys. Rev. E* 95.1 (2017), p. 012413 (cit. on pp. 8, 9).

[88] Vanni, F., Lukovic, M., and Grigolini, P. “Criticality and Transmission of Information in a Swarm of Cooperative Units”. English. In: *Physical review letters* 107 (2011), p. 078103 (cit. on p. 1).

[89] Wei, Xue-Xin and Stocker, Alan A. “Mutual Information, Fisher Information, and Efficient Coding”. In: *Neural Computation* 28.2 (2015), pp. 305–326 (cit. on p. 9).

[90] Yu, S., Ribeiro, T. L., Meisel, C., Chou, S., Mitz, A., Saunders, R., and Plenz, D. “Maintained Avalanche Dynamics during Task-Induced Changes of Neuronal Activity in Nonhuman Primates”. In: *eLife* 6 (2017) (cit. on p. 9).

[91] Zanoci, Cristian, Dehghani, Nima, and Tegmark, Max. “Ensemble Inhibition and Excitation in the Human Cortex: An Ising-Model Analysis with Uncertainties”. In: *Phys. Rev. E* 99.3 (2019), p. 032408 (cit. on pp. 7, 8).

[92] Zeldenrust, Fleur, Gutkin, Boris, and Denéve, Sophie. “Efficient and Robust Coding in Heterogeneous Recurrent Networks”. en. In: *bioRxiv* (2019), p. 804864 (cit. on p. 9).

[93] Zimmern, Vincent. “Why Brain Criticality Is Clinically Relevant: A Scoping Review”. English. In: *Front. Neural Circuits* 14 (2020) (cit. on p. 9).

frontiers in PSYCHOLOGY

OPINION ARTICLE
published: 19 September 2014
doi: 10.3389/fpsyg.2014.01063

Is the frontal lobe involved in conscious perception?

Shervin Safavi[1,2†], Vishal Kapoor[1,2†], Nikos K. Logothetis[1,3] and Theofanis I. Panagiotaropoulos[1]*

[1] Department Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2] International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, Tübingen, Germany
[3] Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, UK
*Correspondence: theofanis.panagiotaropoulos@tuebingen.mpg.de

† These authors have contributed equally to this work.

Edited by:
Jaan Aru, University of Tartu, Estonia

Reviewed by:
Wolfgang Einhauser, Philipps-Universität Marburg, Germany
Jaan Aru, University of Tartu, Estonia

Keywords: prefrontal cortex, conscious visual perception, frontal lobe, binocular rivalry, perceptual suppression

When studying the neural mechanisms underlying conscious perception we should be careful not to misinterpret evidence, and delineate these mechanisms from activity which could reflect the prerequisites or consequences of conscious experiences (Aru et al., 2012; De Graaf et al., 2012). However, at the same time, we need to be careful not to exclude any relevant evidence about the phenomenon.

Recently, novel paradigms have attempted to dissociate activity related to conscious perception from activity reflecting its prerequisites and consequences. In particular, one of these studies focused on resolving the role of frontal lobe in conscious perception (Frässle et al., 2014). Through a clever experimental design that contrasted blood-oxygen-level-dependent (BOLD) activity elicited during binocular rivalry with and without behavioral reports, Frässle et al. (2014) suggested that frontal lobe, or a large part of it, may not be necessary for conscious perception *per se*. Rather frontal areas are involved in processing the consequences of conscious perception like monitoring the perceptual content in order to elicit the appropriate report of the subjective experience. In particular, Frässle et al. showed that behavioral reports of conscious experiences resulted in increased and more widespread activity of the frontal lobe compared to a condition without behavioral reports, where spontaneous transitions in the content of consciousness were estimated through the objective measures like optokinetic nystagmus (OKN) and pupil dilation. The authors of this study concluded that "*frontal areas are associated with active report and introspection rather than with rivalry per se.*" Therefore, activity in prefrontal regions could be considered as a consequence rather than a direct neural correlate of conscious experience.

However, a previous study (Panagiotaropoulos et al., 2012) that measured directly neural activity in the macaque lateral prefrontal cortex (LPFC) using extracellular electrophysiological recordings could help to narrow down the role of frontal activity in conscious perception and exclude the contribution of cognitive or motor consequences in prefrontal neural activity during visual awareness. Specifically, the activity of feature selective neurons in the macaque LFPC was shown to be modulated in accordance with the content of subjective perception, without any confound from motor action (i.e., behavioral reports). Using binocular flash suppression (BFS), a paradigm of robust, externally induced perceptual suppression and without any requirement of behavioral reports, neurons in the LPFC were found to increase or decrease their discharge activity when their preferred stimulus was perceptually dominant or suppressed, respectively. Therefore, since neuronal discharges in the LPFC follow the content of conscious perception even without any motor action, the conclusion of Frässle et al. (2014) about the role of frontal lobe activity in rivalrous perception needs to be refined. Prefrontal activity can indeed reflect the content of conscious perception under conditions of rivalrous stimulation and this activity should not be necessarily considered as the result of a motor action or self-monitoring required for active report. Moreover, the results obtained by Frässle et al. (2014) do not anatomically preclude the entire prefrontal cortex from having a role in conscious perception. Specifically, the BOLD activity related to rivalry in their experiment is still present in the right inferior frontal lobe and right superior frontal lobe (Zaretskaya and Narinyan, 2014). Further, activation of dorso- LPFC in conscious perception of Mooney images was also reported in a study that explicitly controlled for activity elicited by motor action (Imamoglu et al., 2012).

It is true that the BFS-related prefrontal activity cannot conclude on a mechanistic, causal involvement of prefrontal activity in driving spontaneous transitions in conscious perception. This is because BFS is a paradigm of externally induced perceptual suppression and is therefore not directly informative about the role of recorded activity in spontaneous transitions. Therefore, the possibility remains open that the kind of prefrontal activity observed in the macaque LPFC during BFS is not a causal factor for conscious perception but rather reflects some other aspects of monitoring that are not directly related to motor action. For example, prefrontal activity could just reflect a read-out from other areas like the inferior temporal cortex (Sheinberg and Logothetis, 1997) that also reliably reflects the content of conscious perception. However, if this is the case, it triggers the question why this activity that closely follows the content of subjective perception is observed in the LPFC even in the absence of any behavioral report. Overall, it motivates further investigation to understand whether prefrontal activity

has a mechanistic role in conscious perception or it might underlie some monitoring functions that are not necessarily bound to motor action.

Similar to this debate on the role of LPFC in visual awareness, the last decade witnessed disagreement on whether activity in primary visual cortex reflects subjective perception as monitored with electrophysiology and fMRI (Leopold and Logothetis, 1996; Tong, 2003; Maier et al., 2008; Keliris et al., 2010; Leopold, 2012). Measuring both electrophysiological activity and the BOLD signal in the same macaques engaged in an identical task of perceptual suppression finally provided the solution (Maier et al., 2008; Leopold, 2012). Therefore, in order to investigate and resolve the role of PFC in visual perception, one must take a similar approach that utilizes multiple measurement techniques simultaneously or in the same animal along with a careful experimental design. The experimental tasks should not only segregate the effect of various cognitive processes such as attention or introspection in comparison to awareness (Watanabe et al., 2011; Frässle et al., 2014), but also use an objective criterion to decode the content of conscious experience (Frässle et al., 2014), therefore separating perception-related activities from the subsequent behavioral report. Such an approach could therefore robustly delineate the prerequisites and consequences of conscious experience and reveal the true correlates of conscious perception.

Lastly, although such a multimodal approach could provide us substantial insights into the activity underlying the representation of conscious content, whether or not this activity has a causal role in mediating perception remains to be understood. Although a number of studies indeed point to a causal involvement of prefrontal cortex in conscious perception (reviewed in Dehaene and Changeux, 2011), a systematic study which directly interferes with prefrontal activity during a task of subjective perception is currently, to the best of our knowledge, missing. While utilizing objective criteria as indicators of perceptual transitions, systematic perturbation of the PFC (such as cooling, transcranial magnetic stimulation, microstimulation, or optogenetics) and observing concomitant changes in the temporal dynamics of perceptual transitions could reveal its causal contribution. Indeed, patients with frontal lesions are impaired in their ability to switch from one subjective view of an ambiguous figure to the other (for example see Ricci and Blundo, 1990, but also see a different case study from Valle-Inclán and Gallego, 2006).

We would like to conclude that in formulating our conclusions related to prerequisites, consequences and true correlates of conscious experiences, we need to have an *integrative view* on the available evidence. Our investigations and conclusions about the neural correlates of consciousness must not only entail better-designed experiments but also diverse experimental techniques (e.g., BOLD fMRI, electrophysiology) that could measure brain activity on different spatial and temporal scales (Panagiotaropoulos et al., 2014). Such a multi-modal approach holds great promise in refining our current understanding of conscious processing.

ACKNOWLEDGMENT

This work was supported by the Max Planck Society and the International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, Germany.

REFERENCES

Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. doi: 10.1016/j.neubiorev.2011.12.003

De Graaf, T. A., Hsieh, P. J., and Sack, A. T. (2012). The "correlates" in neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 191–197. doi: 10.1016/j.neubiorev.2011.05.012

Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018

Frässle, S., Sommer, J., Jansen, A., Naber, M., and Einhauser, W. (2014). Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* 34, 1738–1747. doi: 10.1523/JNEUROSCI.4403-13.2014

Imamoglu, F., Kahnt, T., Koch, C., and Haynes, J. D. (2012). Changes in functional connectivity support conscious object recognition. *Neuroimage* 63, 1909–1917. doi: 10.1016/j.neuroimage.2012.07.056

Keliris, G. A., Logothetis, N. K., and Tolias, A. S. (2010). The role of the primary visual cortex in perceptual suppression of salient visual stimuli. *J. Neurosci.* 30, 12353–12365. doi: 10.1523/JNEUROSCI.0677-10.2010

Leopold, D. A. (2012). Primary visual cortex: awareness and blindsight. *Annu. Rev. Neurosci.* 35, 91–109. doi: 10.1146/annurev-neuro-062111-150356

Leopold, D. A., and Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553. doi: 10.1038/379549a0

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., and Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat. Neurosci.* 11, 1193–1200. doi: 10.1038/nn.2173

Panagiotaropoulos, T. I., Deco, G., Kapoor, V., and Logothetis, N. K. (2012). Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron* 74, 924–935. doi: 10.1016/j.neuron.2012.04.013

Panagiotaropoulos, T. I., Kapoor, V., and Logothetis, N. K. (2014). Subjective visual perception: from local processing to emergent phenomena of brain activity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369:20130534. doi: 10.1098/rstb.2013.0534

Ricci, C., and Blundo, C. (1990). Perception of ambiguous figures after focal brain lesions. *Neuropsychologia* 28, 1163–1173. doi: 10.1016/0028-3932(90)90052-P

Sheinberg, D. L., and Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc. Natl. Acad. Sci. U.S.A.* 94, 3408–3413. doi: 10.1073/pnas.94.7.3408

Tong, F. (2003). Primary visual cortex and visual awareness. *Nat. Rev. Neurosci.* 4, 219–229. doi: 10.1038/nrn1055

Valle-Inclán, F., and Gallego, E. (2006). Chapter 13 Bilateral frontal leucotomy does not alter perceptual alternation during binocular rivalry. *Prog. Brain Res.* 155, 235–239. doi: 10.1016/S0079-6123(06)55013-7

Watanabe, M., Cheng, K., Murayama, Y., Ueno, K., Asamizuya, T., Tanaka, K., et al. (2011). Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science* 334, 829–831. doi: 10.1126/science.1203161

Zaretskaya, N., and Narinyan, M. (2014). Introspection, attention or awareness? The role of the frontal lobe in binocular rivalry. *Front. Hum. Neurosci.* 8:527. doi: 10.3389/fnhum.2014.00527

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 July 2014; accepted: 04 September 2014; published online: 19 September 2014.

*Citation: Safavi S, Kapoor V, Logothetis NK and Panagiotaropoulos TI (2014) Is the frontal lobe involved in conscious perception? Front. Psychol. **5**:1063. doi: 10.3389/fpsyg.2014.01063*

This article was submitted to Consciousness Research, a section of the journal Frontiers in Psychology.

Nonmonotonic spatial structure of interneuronal correlations in prefrontal microcircuits

Shervin Safavi[a,b,1], Abhilash Dwarakanath[a,1], Vishal Kapoor[a,b], Joachim Werner[a], Nicholas G. Hatsopoulos[c], Nikos K. Logothetis[a,d,2], and Theofanis I. Panagiotaropoulos[a,e,2]

[a]Department Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany; [b]International Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72074 Tübingen, Germany; [c]Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637; [d]Division of Imaging Science and Biomedical Engineering, University of Manchester, 72074 Manchester, United Kingdom; and [e]Cognitive Neuroimaging Unit, Commissariat à l'Énergie Atomique, Division Sciences de la Vie (DSV), Institut d'imagerie Biomédicale (I2BM), INSERM, Université Paris-Sud, Université Paris-Saclay, Neurospin Center, 91191 Gif/Yvette, France

Contributed by Nikos K. Logothetis, February 20, 2018 (sent for review June 21, 2017; reviewed by Mayank R. Mehta and Bijan Pesaran)

Correlated fluctuations of single neuron discharges, on a mesoscopic scale, decrease as a function of lateral distance in early sensory cortices, reflecting a rapid spatial decay of lateral connection probability and excitation. However, spatial periodicities in horizontal connectivity and associational input as well as an enhanced probability of lateral excitatory connections in the association cortex could theoretically result in nonmonotonic correlation structures. Here, we show such a spatially nonmonotonic correlation structure, characterized by significantly positive long-range correlations, in the inferior convexity of the macaque prefrontal cortex. This functional connectivity kernel was more pronounced during wakefulness than anesthesia and could be largely attributed to the spatial pattern of correlated variability between functionally similar neurons during structured visual stimulation. These results suggest that the spatial decay of lateral functional connectivity is not a common organizational principle of neocortical microcircuits. A nonmonotonic correlation structure could reflect a critical topological feature of prefrontal microcircuits, facilitating their role in integrative processes.

functional connectivity | prefrontal cortex | network structure | long-range interactions | noise correlations

The intraareal connectivity patterns of neural populations in the mammalian neocortex frequently repeat across cortical areas (1–4). Such canonical rules with general validity are important in understanding basic organizational principles and ensemble computations in cortical networks (1–3, 5). Nevertheless, identifying deviations from these rules between sensory and higher-order, association cortical areas could reveal properties leading to cortical network specialization and higher cognitive functions (1–3, 5, 6).

The spatial structure of intraareal functional connectivity is frequently inferred by measuring the trial-by-trial correlated variability of neuronal discharges (spike count correlations) (7). One of the most well-established properties (a canonical feature) of intraareal, mesoscopic, functional connectivity is a so-called limited-range correlation structure, reflecting a monotonic decrease of spike count correlations as a function of spatial distance and tuning similarity (7–17). However, this distance-dependent decrease of correlations has been almost exclusively derived from recordings in primary sensory cortical areas or inferred from recordings in the prefrontal cortex (PFC) with various constraints like a rather limited scale (18) (see also *Discussion*). As a result, it is currently unclear whether known differences in the structure of anatomical connectivity across the cortical hierarchy could also give rise to different spatial patterns of functional connectivity (19–22).

Specifically, the rapid spatial decay of correlations in sensory cortex is widely assumed to reflect a similar rapid decay in lateral anatomical connectivity and excitation (23). In early visual cortical areas, correlations rapidly decrease as a function of distance (refs. 12, 14, and 17; but also see refs. 24 and 25) in a manner that closely reflects anatomical findings about the limited spread and density of intrinsic lateral connections (19, 26–29). However, lateral connections are significantly expanded in later stages of the cortical hierarchy, like the PFC (19, 21, 28–31). In this higher-order association area, lateral connections commonly extend to distances up to 7–8 mm (28, 29, 31), while patches of connected populations are both larger and more distant from each other compared with sensory cortex (29, 32). Although horizontal axons in macaque V1 can extend up to 4 mm, they do not form clear patches, and for distances of 2–3 mm laterally to the injection patch border, only a small number of cells are labeled in comparison with higher-order areas (19, 27, 29, 33, 34). In addition to the more extended intrinsic lateral connectivity, associational input from other cortical areas to the PFC also forms stripes with an average distance of 1.5 mm and contributes to the spatial periodicities in lateral organization (35). Finally, the proportion of lateral excitatory connections is higher in the PFC (95%) compared with V1 (75%) (36).

Whether these significant differences in the structural architecture of the PFC compared with early sensory areas also result in a distinct spatial pattern of functional connectivity is currently unknown. Intuitively, higher probability of long-range lateral excitatory connections and stripe-like associational input patterns could give rise to strong spike count correlations across local and spatially remote populations, with weaker correlations for populations

Significance

The spatial structure of correlated activity of neurons in lower-order visual areas has been shown to linearly decrease as a measure of distance. The shape of correlated variability is a defining feature of cortical microcircuits, as it constrains the computational power and diversity of a region. We show here a nonmonotonic spatial structure of functional connectivity in the prefrontal cortex (PFC) where distal interactions are just as strong as proximal interactions during visual engagement of functionally similar PFC neurons. Such a nonmonotonic structure of functional connectivity could have far-reaching consequences in rethinking the nature and role of prefrontal microcircuits in various cognitive states.

Author contributions: N.K.L. and T.I.P. designed research; A.D., V.K., J.W., and T.I.P. performed research; N.G.H. and N.K.L. contributed new reagents/analytic tools; S.S., A.D., V.K., and T.I.P. analyzed data; and S.S., A.D., V.K., N.G.H., N.K.L., and T.I.P. wrote the paper.

Reviewers: M.R.M., University of California, Los Angeles; and B.P., New York University.

The authors declare no conflict of interest.

[1]S.S. and A.D. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: nikos.logothetis@tuebingen.mpg.de or theofanis.panagiotaropoulos@tuebingen.mpg.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1802356115/-/DCSupplemental.

Published online March 27, 2018.

in intermediate distances. To address this question, we recorded simultaneously the activity of large neural populations in the inferior convexity of the macaque PFC during both anesthetized and awake states using multielectrode Utah arrays (37). In both anesthetized and awake states, the spatial pattern of pairwise correlated variability was nonmonotonic with significantly positive long-range correlations. A major source of nonmonotonicity could be attributed to the spatial pattern of correlated variability between functionally similar neurons.

Results

We used multielectrode Utah arrays (4 × 4 mm, 10 × 10 electrodes, interelectrode distance 400 μm, electrode length 1 mm; Fig. 1*A*) to record spiking activity from the inferior convexity of the ventrolateral PFC (vlPFC) during repeated visual stimulation with movie clips in two anesthetized macaque monkeys (Fig. 1*B*) and with sinusoidal gratings, drifting in eight different directions, in two awake behaving macaques (Fig. 1*C*). To evaluate the effect of structured visual input on correlated variability, we contrasted periods of visual stimulation to intertrial as well as spontaneous activity (long periods of neural activity without any task demands). Both anesthetized- and awake-state recordings resulted in the simultaneous monitoring of multiple, well-isolated single units that remained stable for several hours of recording (Fig. 1*D*). On average, in each dataset, we recorded from 103 ± 16 (mean ± SEM) single units and 5,305 ± 1,681 pairs during anesthesia (Fig. S1*A*) and 107 ± 14 single units and 5,758 ± 1,675 pairs during wakefulness (Fig. S1*B*).

Spatial Structure of Correlated Variability During Anesthesia and Wakefulness. It has been repeatedly shown that correlated variability of spike counts in early sensory, especially visual, areas in different species decreases as a function of lateral distance, with strong interactions for proximal and progressively weaker interactions for distal (up to 4 mm) neurons (14, 15, 17). We investigated the same relationship between spike count correlations (r_{sc}) and lateral distance up to 4 mm in the vlPFC.

Visual stimulation with movie clips during anesthesia or with drifting gratings during wakefulness gave rise to a spatial pattern in the structure of correlated variability that was fundamentally different compared with early sensory areas: strong and positive long-range (>2.5 mm) correlations that were comparable to the average magnitude of local (up to 1 mm) correlations and significantly weaker correlations for intermediate distances (red curves in Fig. 2 and Fig. S1 *A* and *B*).

We evaluated the statistical significance of nonmonotonicity and long-range correlations by comparing the distributions of pairwise correlations in populations recorded from nearby (0.5 mm for anesthetized and 1 mm for awake), intermediate (2.5 mm), and distant (3.5–4 mm) sites during visual stimulation. The choice of these particular distance bins was based on the local extrema of correlated variability as a function of distance

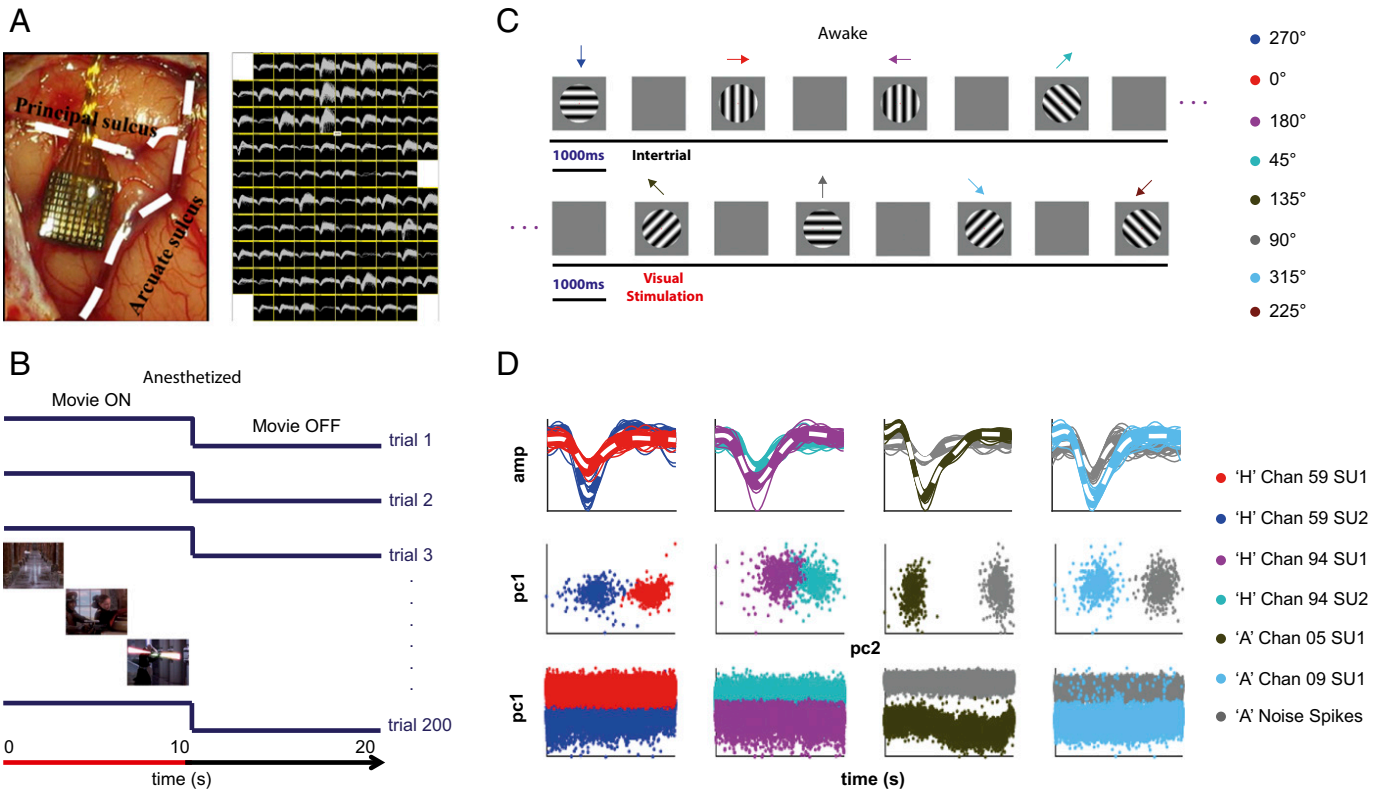


Fig. 1. Implantation, visual stimulation, and quality of single unit isolation. (*A*) Location of the implanted array with respect to arcuate and principal sulci and an example of typical waveforms acquired across the implanted cortical patch during a typical recording session in an awake animal. (*B*) Anesthetized visual stimulation protocol: 10 s of movie clip presentation was interleaved with 10-s-long intertrial (stimulus off) periods for 200 repetitions. (*C*) Awake visual stimulation protocol: The macaques initiated each trial by fixating on a red dot for 300 ms, following which a drifting sinusoidal grating was presented monocular for 1,000 ms. After 1,000 ms of visual stimulation and a 300-ms stimulus-off period, liquid reward was delivered for successful fixation throughout the trial period. An intertrial period of 1,000 ms preceded the next trial. Each block of trials comprised eight different motion directions (exemplified by differently colored arrows) presented in a random order. (*D*) Single unit isolation quality: Each column shows the activity recorded from four channels recorded in two different datasets, one from each of the two monkeys (monkeys H and A). The 500 example waveforms for single units (shown as colored clusters) and noise spikes (multiunit activity shown as gray clusters) along with the mean waveform in dashed white, and their corresponding first and second principal components (pc1 and pc2) are shown in the first and second row, respectively. In the last row, the first principal component of all of the waveforms in a cluster is plotted over time, demonstrating stability of recordings and single unit isolation for periods lasting ~3.5–4 h. amp, amplitude.

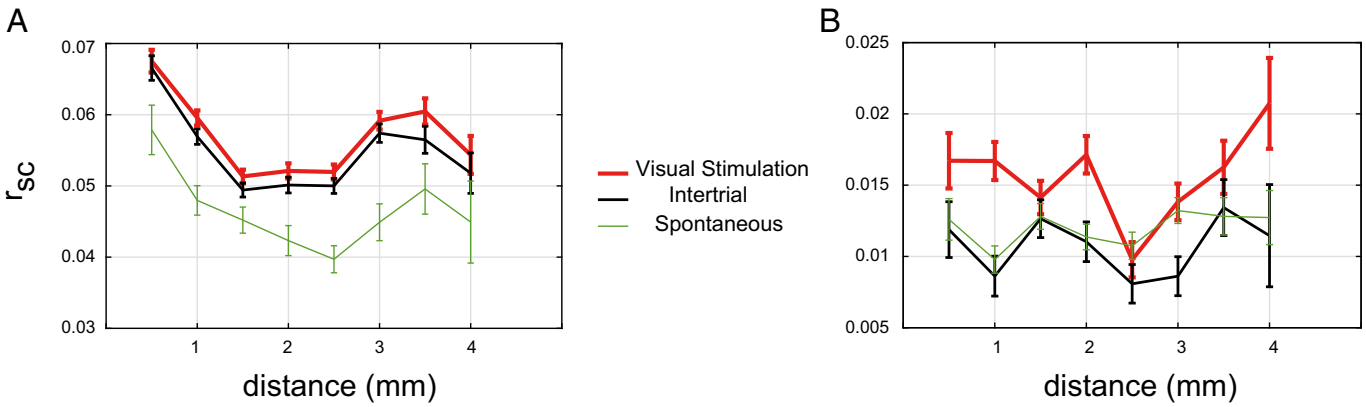


Fig. 2. Spatial structure of correlated variability. (*A*) Spike count correlations (r_{sc}) during visual stimulation (red), intertrial (black), and spontaneous activity (green) as a function of lateral spatial distance (millimeters) between cell pairs for anesthetized-state recordings (error bars represent mean ± SEM). (*B*) Same as *A* for awake-state recordings.

during visual stimulation. For all of the comparisons made across these distance bins, to assess the significance of the differences in correlated variability, we used the Wilcoxon rank-sum test (unless otherwise mentioned explicitly). Moreover, as we made the comparisons across the three key distance bins, we assessed the significance after a Bonferroni correction for multiple comparisons (corrected $P = 0.0167$). Summary statistics of Bonferroni-corrected P values are available in Tables S1–S3 (for anesthetized and awake data).

Average correlated variability between neurons located in intermediate distances was significantly lower compared with very proximal neurons in the anesthetized ($\overline{r_{sc}}^{0.5\ mm} = 0.0675 \pm 0.0016$ vs. $\overline{r_{sc}}^{2.5\ mm} = 0.0520 \pm 0.0010$; $P < 10^{-10}$; Fig. 2*A*, red curve, and Fig. S2*A*) state. In the awake recordings, correlations among nearby neurons, i.e., at a pairwise distance of 1 mm, showed a significant difference from those at the local minimum of the spatial correlation structure ($\overline{r_{sc}}^{1\ mm} = 0.0167 \pm 0.0013$ vs. $\overline{r_{sc}}^{2.5\ mm} = 0.0098 \pm 0.0012$; $P = 0.0038$; Fig. 2*B*, red curve, and Fig. S2*D*). Following this minimum, correlations during anesthesia significantly increased from 2.5 to 3 mm ($\overline{r_{sc}}^{2.5\ mm} = 0.0520 \pm 0.0010$ vs. $\overline{r_{sc}}^{3\ mm} = 0.0592 \pm 0.0012$; $P = 0.012$) and 3.5 mm ($\overline{r_{sc}}^{3.5\ mm} = 0.0605 \pm 0.0018$; $P = 0.0040$). A similar increase in correlated variability for progressively more distant populations was also observed in the awake state, where correlations significantly increased from 2.5 mm to both 3.5 mm ($\overline{r_{sc}}^{2.5\ mm} = 0.0098 \pm 0.0012$ vs. $\overline{r_{sc}}^{3.5\ mm} = 0.0163 \pm 0.0019$; $P = 0.0076$) and 4 mm ($\overline{r_{sc}}^{2.5\ mm} = 0.0098 \pm 0.0012$ vs. $\overline{r_{sc}}^{4\ mm} = 0.0207 \pm 0.0032$; $P = 0.0079$).

In the awake-state recordings, the average magnitude of correlations for distant populations, located 3.5–4 mm apart, was not different from the respective magnitude for nearby pairs ($\overline{r_{sc}}^{1\ mm} = 0.0167 \pm 0.0013$ vs. $\overline{r_{sc}}^{3.5\ mm} = 0.0163 \pm 0.0019$, $P = 0.7$; and $\overline{r_{sc}}^{4\ mm} = 0.0207 \pm 0.0032$, $P = 0.3$; Fig. 2*B*, red curve, and Fig. S2*D*). In addition, both local and distant average correlations were significantly positive ($P < 0.005$, t test). However, in the anesthetized recordings, despite the significant increase of correlations for distant neurons compared with intermediate distances, long-range correlations remained significantly lower compared with nearby neurons ($\overline{r_{sc}}^{0.5\ mm} = 0.075 \pm 0.0016$ vs. $\overline{r_{sc}}^{3.5\ mm} = 0.0605 \pm 0.0018$, $P = 0.0056$; and $\overline{r_{sc}}^{4\ mm} = 0.0544 \pm 0.0027$, $P = 0.0008$; Fig. 2*A*, red curve, and Fig. S2*A*), suggesting that anesthesia-induced fluctuations had a nonhomogeneous impact on the spatial structure of correlations. Such nonhomogeneous, state-dependent weighting on the spatial structure of correlations has been reported in previous studies of primary visual cortex as well (12).

The nonmonotonic structure in correlated variability could not be ascribed to random spatial variability in firing rates, since it could be observed even when correlated variability was estimated for populations with matched geometric mean firing rates across lateral distances (Fig. S3). Furthermore, to confirm that the intrinsic nonuniformity of spatial sampling with Utah arrays did not lead to the nonmonotonic structure of correlated variability, we used a bootstrapping analysis of our spatial sampling (Fig. S4). This analysis showed that equalized resampling of pairs across distance bins also resulted in a nonmonotonic correlated variability structure (Fig. 2).

The decrease in correlations from nearby neuronal pairs (0.5 mm in the anesthetized state and 1 mm in the awake state) to 2.5 mm and the increase from 2.5 to 3.5 or 4 mm was observed in both the anesthetized and awake states. However, in the awake-state recordings, we also observed an additional pronounced peak at 2 mm (Fig. 2*B* and Fig. S2*D*). Lack of this peak at intermediate distances in our anesthetized recordings is compatible with other studies performed during anesthesia and provides further evidence for a nonhomogeneous, state-dependent weighting on the spatial structure of correlated variability (12, 14, 17, 25). These common features in the spatial structure of correlated activity across anesthetized ($\overline{r_{sc}}^{anesth}$) and awake ($\overline{r_{sc}}^{awake}$) states were observed despite significant differences in the average magnitude of correlations [$\overline{r_{sc}}^{anesth} = 0.0574 \pm 3 \times 10^{-4}$ (mean ± SEM) vs. $\overline{r_{sc}}^{awake} = 0.0153 \pm 3 \times 10^{-4}$; $P = 0$; Fig. 3*A*]. Despite being very close to zero, average correlations during visual stimulation were significantly positive during the awake state ($P < 10^{-104}$; t test).

Visual Stimulation Shapes the Spatial Structure of Correlated Variability. We evaluated the impact of structured visual stimulation on the spatial pattern of correlated variability by comparing correlations during visual stimulation with movie clips (during anesthesia) or drifting sinusoidal gratings (during wakefulness) to the respective pattern during intertrial and spontaneous activity periods. Compared with periods of intertrial activity, visual stimulation resulted in a significant increase of correlated variability in both anesthetized recordings ($\overline{r_{sc}}$ visual $= 0.0574 \pm 3 \times 10^{-4}$ vs. $\overline{r_{sc}}$ intertrial $= 0.0554 \pm 3 \times 10^{-4}$; $P < 10^{-3}$; Fig. 3*B*) and awake recordings ($\overline{r_{sc}}$ visual $= 0.0153 \pm 3 \times 10^{-4}$ vs. $\overline{r_{sc}}$ intertrial $= 0.011 \pm 4 \times 10^{-4}$; $P = 6.7 \times 10^{-5}$; Fig. 3*C*).

Visual stimulation in the awake state significantly shaped a spatially inhomogeneous, nonmonotonic structure of correlated variability. In striking contrast to the significant differences observed for the same lateral distances during visual stimulation, we found that correlations during the intertrial period were not different between local and intermediate populations ($\overline{r_{sc}}^{0.5\ mm} = 0.0119 \pm 0.0020$ vs. $\overline{r_{sc}}^{2.5\ mm} = 0.0081 \pm 0.0014$, $P > 0.3$; Fig. 2*B*, black curve, and Fig. S2*B*) or intermediate and distant populations ($\overline{r_{sc}}^{2.5\ mm} = 0.0081 \pm 0.0014$ vs. $\overline{r_{sc}}^{3.5\ mm} = 0.0134 \pm 0.0020$, $P = 0.08$; and vs. $\overline{r_{sc}}^{4\ mm} = 0.0115 \pm 0.0036$, $P > 0.25$; Fig. 2*B*, black curve, and Fig. S2*B*). Spatially homogeneous correlations were also observed during periods without any structured visual input or task engagement, in data collected during spontaneous activity (Fig. 2*B*, green curve). In these epochs, we also found no difference between local and intermediate correlations ($\overline{r_{sc}}^{0.5\ mm} = 0.0126 \pm 0.0015$ vs. $\overline{r_{sc}}^{2.5\ mm} = 0.0107 \pm 9.5 \times 10^{-4}$;

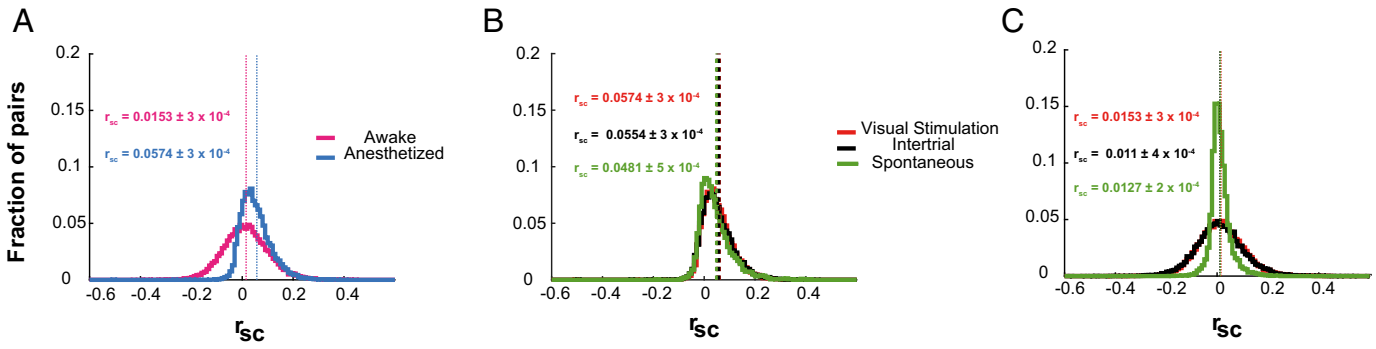


Fig. 3. Distributions of correlated variability across different states and conditions. (*A*) Distribution of pairwise correlated variability (fraction of pairs) and mean values (dotted lines) during visual stimulation for anesthetized (blue) and awake (pink) recordings. Correlated variability was significantly stronger during anesthesia as a result of a shift in the distribution of pairwise correlations toward positive values. (*B*) Same as *A* for anesthetized-state recordings during visual stimulation (red), intertrial (black), and spontaneous activity (green) periods. (*C*) Same as *B* for awake-state recordings.

$P = 0.6$) and very similar correlations between intermediate and distant populations ($\overline{r_{sc}}^{2.5\ mm} = 0.0107 \pm 9.5 \times 10^{-4}$ vs. $\overline{r_{sc}}^{3.5\ mm} = 0.0128 \pm 0.0013$, $P > 0.9$; vs. $\overline{r_{sc}}^{4\ mm} = 0.0127 \pm 0.0019$, $P > 0.17$; Fig. 2*B*, green curve).

We quantified the magnitude of spatial inhomogeneity in the structure of correlations across different conditions and states (*Experimental Procedures*). A clear difference in the rate of changes in correlated variability was observed in awake-state recordings (Fig. 4*A*), where visual stimulation resulted in the strongest spatial variability and intertrial activity in the weakest (almost constant average correlation as a function of lateral distance). A similar spatial variability was also observed under anesthesia (Fig. 4*B*); however, the average rate of change was comparable across the two conditions of visual stimulation and intertrial, but different during spontaneous activity. The difference in the structure of functional connectivity between visual stimulation and intertrial periods across anesthesia and awake states could be attributed to the lack of saccadic eye movements in intertrial periods during anesthesia. Saccadic eye movements reset visual perception (38), and their absence could create a persistent network state, showing no reset, resulting in very similar patterns of correlations during visual stimulation and intertrial periods.

These results suggest that the spatial structure of correlated variability in the PFC is inhomogeneous. The magnitude of inhomogeneity depended not only on the variation of global states such as wakefulness or anesthesia, but most importantly on behavioral demands, i.e., visual stimulation, intertrial (anticipation of the succeeding trial), or spontaneous activity (no behavioral load). Although traces of inhomogeneity in the spatial structure of correlations were observed during spontaneous activity or intertrial periods, structured visual stimulation during the awake state appeared to result in the strongest spatial inhomogeneity in the correlation structure.

Prevalence of Nonmonotonic Spatial Structure in Functionally Similar Populations. Lateral connectivity in PFC has been hypothesized to preferentially target neurons with functional similarities (e.g., similar spatial tuning), similar to iso-orientation columns in the visual cortex [Goldman-Rakic (39)]. Therefore, we next examined whether the source of the nonmonotonic correlated variability could be traced to populations of neurons that were modulated similarly by visual input. First, tuning functions for each recorded unit were obtained based on the discharge response to sinusoidal gratings drifting in eight different directions (Fig. 5). The correlation between tuning functions (signal correlation; r_{signal}) provided a measure of functional similarity among the recorded pairs (see *Experimental Procedures* for more details). We analyzed the relationship between the spatial structure of functional connectivity and functional similarity of pairwise responses (i.e., signal correlations). The relationship between signal correlations, noise correlations, and interneuronal distance (Fig. 6*A*) points to a stronger nonmonotonic trend for pairs with positive signal correlations.

Specifically, we computed the noise correlation across distance bins for pairs with positive signal correlations ($0.1 < r_{signal} < 0.9$) during visual stimulation (Fig. 6 *A* and *B*). A nonmonotonic trend could be observed; however, the differences between the first local maximum to the local minimum, and the local minimum to the next local maximum were marginally significant ($\overline{r_{sc}}^{0.5\ mm} = 0.0266 \pm 0.0032$ vs. $\overline{r_{sc}}^{2.5\ mm} = 0.0165 \pm 0.0021$, $P = 0.07$; and $\overline{r_{sc}}^{2.5\ mm} = 0.0165 \pm 0.0021$ vs. $\overline{r_{sc}}^{4\ mm} = 0.0266 \pm 0.0054$, $P = 0.1$; green curve in Fig. 6*B*). The nonmonotonic trend was also confirmed from fitting first- and second-degree polynomials to these data. The adjusted-R^2 goodness-of-fit measure for a line (first-degree polynomial, monotonic) was −0.15, whereas the same measure for a quadratic function (second-degree polynomial, the simplest nonmonotonic function) (40) yielded a value of 0.3, pointing to the quadratic curve being a much better fit to the data (Fig. 7*A*). Progressively higher thresholds for signal correlation, resulting in sampling populations with stronger functional similarity, did not qualitatively change these effects that were characterized by a significant decrease in intermediate distance (~2.5 mm) correlations (Fig. S5 *B*, *E*, and *H*).

During intertrial periods, correlated variability of the same population of functionally similar neurons (functional similarity estimated during the visual stimulation period) was homogenous (Fig. 6 *C* and *D*). For positive signal correlations ($0.1 < r_{signal} < 0.9$) during visual stimulation, the strength of correlated variability between nearby and intermediate neurons was not significantly different ($\overline{r_{sc}}^{0.5\ mm} = 0.015 \pm 0.0033$ vs. $\overline{r_{sc}}^{2.5\ mm} = 0.0111 \pm 0.0022$, $P = 0.3$; and $\overline{r_{sc}}^{1\ mm} = 0.008 \pm 0.0023$ vs. $\overline{r_{sc}}^{2.5\ mm} = 0.0111 \pm 0.0022$, $P = 0.9$;

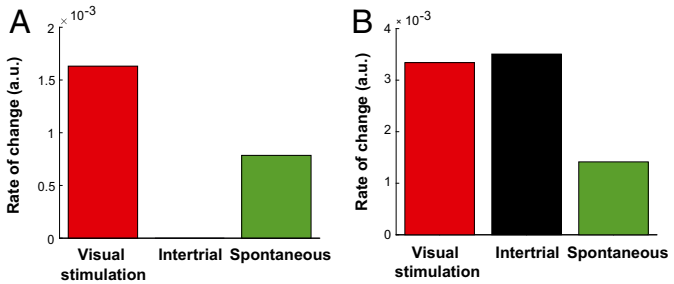


Fig. 4. Quantification of spatial inhomogeneity in the structure of correlated variability. (*A*) Spatial inhomogeneity in the structure of correlations across different conditions during awake-state recordings. Spatial inhomogeneity was quantified by computing the average of the absolute rate of change in the correlation structure across successive distance bins (only those rates significantly different in successive distance bins; see also *Experimental Procedures*). (*B*) Same as *A* for anesthetized-state recordings.

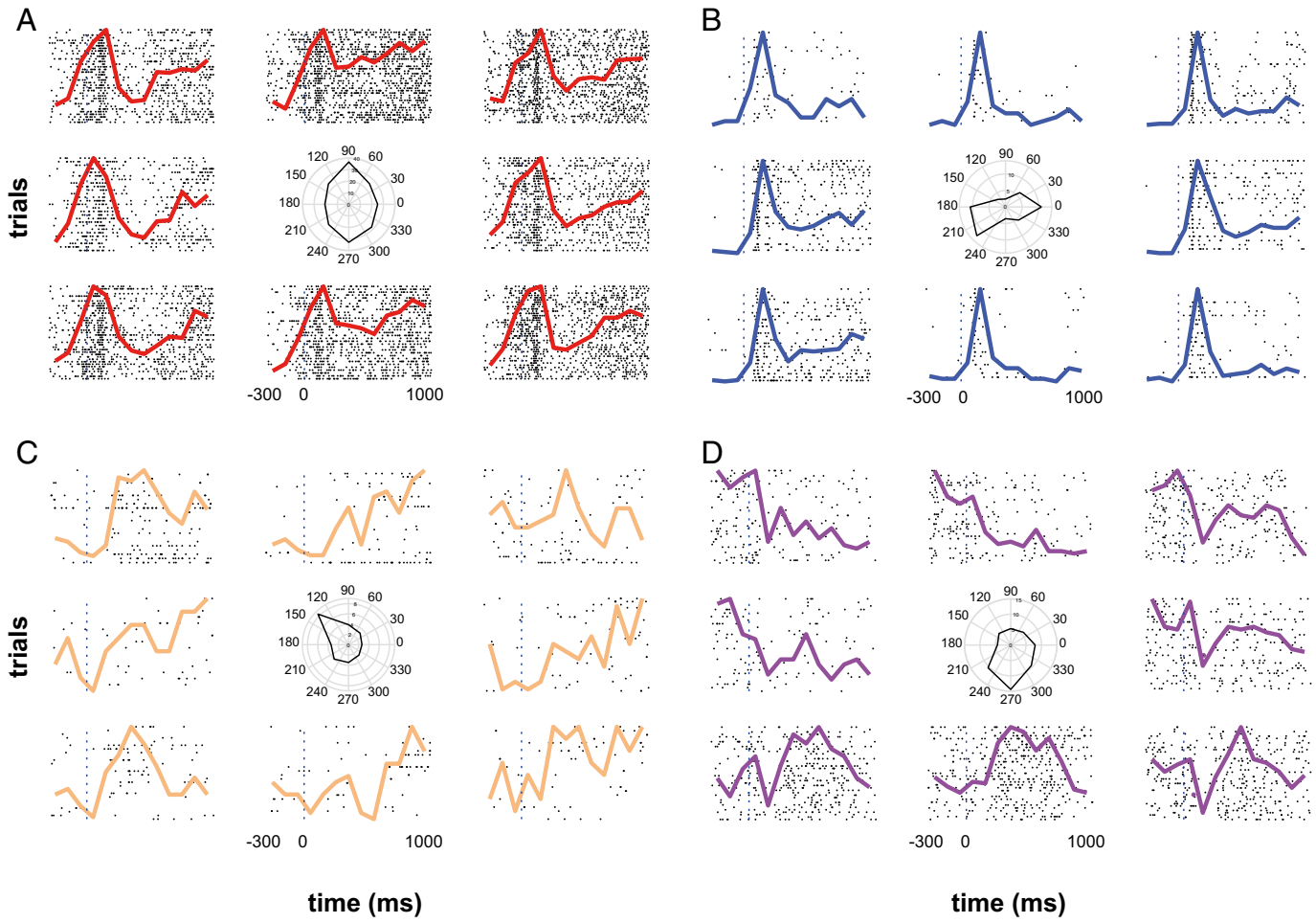


Fig. 5. Visual modulation of single unit activity during wakefulness. Spike raster plots and overlaid peristimulus time histograms (PSTHs) for four single units across all eight orientations (0–315°) are shown. Polar plots for each unit show the preferred direction(s) of motion. The green lines in the center indicate the resultant length and direction (see *SI Experimental Procedures* for more details). *A* and *B* are typical examples of bimodally tuned prefrontal units with significant responses for opposite directions of motion (orientation-selective responses). *C* and *D* are examples of sharper, unimodal responses for a particular direction of motion.

green curve in Fig. 6*D*), and correlations between neurons in intermediate and distant locations were also very similar ($\overline{r_{sc}}^{2.5\ mm} = 0.0111 \pm 0.0022$ vs. $\overline{r_{sc}}^{3.5\ mm} = 0.0077 \pm 0.0034$, $P = 0.8$; and vs. $\overline{r_{sc}}^{4\ mm} = 0.0057 \pm 0.0068$, $P = 0.9$). A similar fitting procedure as used for the data in the visual-stimulation period was also used to test for the observed trends in the intertrial period. Fitting a line yielded an adjusted R^2 value of 0.48, whereas fitting a quadratic function yielded an adjusted R^2 value of 0.5, pointing to both fits being quantitatively similar (Fig. 7*B*). However, whereas the quadratic fit in the visual-stimulation period yielded a U-shaped curve that clearly displayed the equivalence between local and distant populations, this equivalence was not seen during the intertrial period, where distant populations were weakly correlated compared with local populations.

Furthermore, when local (0.5–1 mm) and distant (3.5–4 mm) populations were pooled at a spatial resolution of 1 mm, a clear and specific strengthening of correlated variability at the flanks was observed during structured visual stimulation epochs ($P^{0.5–1\ mm}{}_{VisStim\ \text{vs.}\ Intertrial} = 0.00024$; $P^{3.5–4\ mm}{}_{VisStim\ \text{vs.}\ Intertrial} = 0.017$; $P^{2.5}{}_{VisStim\ \text{vs.}\ Intertrial} = 0.14$). Together, the nonmonotonic spatial structure of functionally similar populations during visual stimulation was stronger compared with the more homogeneous structure (Fig. 7*B*) of the same population during the intertrial period, pointing to a role of structured visual input in shaping the nonmonotonic structure of correlated variability in functionally similar populations.

When the same fitting procedure as described above was performed on pairs of functionally dissimilar neurons (i.e., $-0.9 < r_{signal} < -0.1$), a linearly increasing trend provided a slightly better fit compared with a quadratic fit (adjusted R^2 linear = 0.33; adjusted R^2 quadratic = 0.2). However, when local and distant populations were binned as for the high-signal correlation pairs, correlated variability in the flanks during visual stimulation and intertrial was identical ($P^{0.5–1\ mm}{}_{VisStim\ \text{vs.}\ Intertrial} = 0.64$; $P^{3.5–4\ mm}{}_{VisStim\ \text{vs.}\ Intertrial} = 0.7$). Examples of pairwise neuronal responses from neurons with similar signal correlations and sampled from short, intermediate, and long lateral distances are presented in Fig. 6 *E*–*G*.

Several factors changed between the anesthetized vs. awake animal experiments. For instance, recordings were performed in different monkeys, using different stimuli (movie clips vs. moving grating) and different data-acquisition systems and spike extraction methods. Despite these differences, interneuronal correlations showed a similar spatial structure in both anesthetized and awake recordings with visual stimulation. More specifically, these results suggest that spatial inhomogeneities in the functional architecture of the PFC arise from strong local and long-range lateral functional interactions between functionally similar neurons, which are particularly pronounced during structured visual stimulation in the awake state.

Discussion

Spatial Structure of Prefrontal Correlated Variability and Relationship to Anatomical Structure. Spatial decay in the strength of spike count correlations on a mesoscopic scale, up to 4 mm of lateral

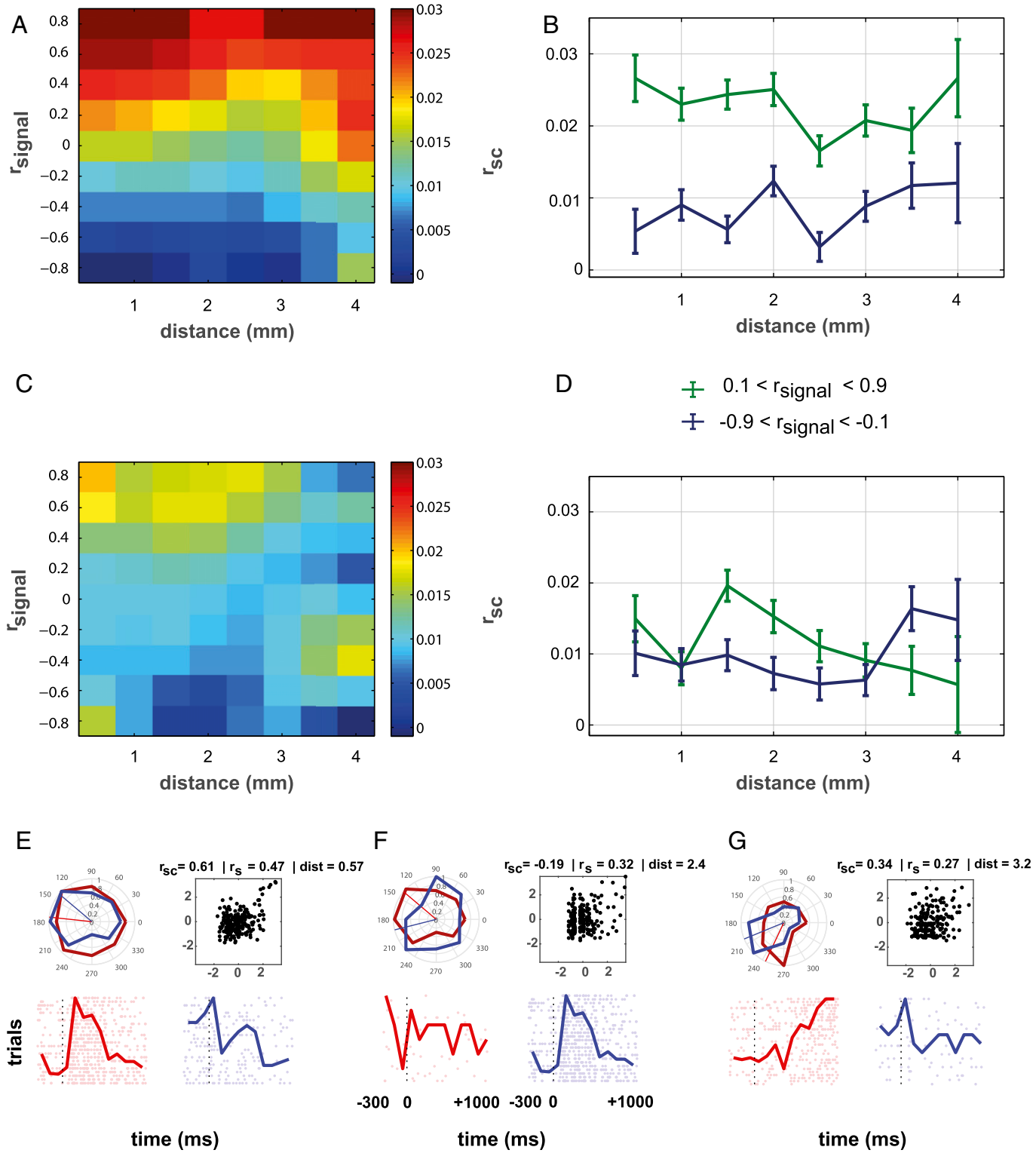


Fig. 6. Effect of functional similarity on the spatial structure of correlated variability. (*A*) Correlated variability as a function of distance and signal correlation for the pooled data recorded during visual stimulation in the awake state. The color of each pixel indicates the average correlated variability for pairs that their signal correlation and distance landed in the specific bin. Pixels containing <10 pairs were removed (white pixels). Correlated variability values are indicated by the color bar at the right of the image. Data were smoothed with a 2D Gaussian (SD of 1 bin) for display purposes. (*B*) Correlated variability as a function of distance (similar to Fig. 2) among neuronal pairs with signal correlation >0.1 and <0.9, i.e., the nonzero upper part of matrix represented in *A* with green line; and among pairs with signal correlation higher than −0.9 and less than −0.1, i.e., the nonzero lower part of matrix represented in *A* with blue line (mean ± SEM as error bars). (*C* and *D*) Same as *A* and *B* for the intertrial period. The signal correlation matrix is computed from the visual stimulation period, and the correlated variability of these populations in the intertrial period is mapped onto the pixels in *C*. (*E*–*G*) Three example pairs with high signal correlations and high, low, and high correlated variability from the nearest, the intermediate, and the farthest distance bins, respectively. The polar plot shows the vector sum of the tuning for each neuron in a given pair, while the scatter plots depict their z-score normalized responses. Example raster plots are overlaid with the PSTHs for the preferred direction of motion. Despite the sparseness in firing for some of the neurons, sharp tunings can be observed (compare raster plots with polar plots).

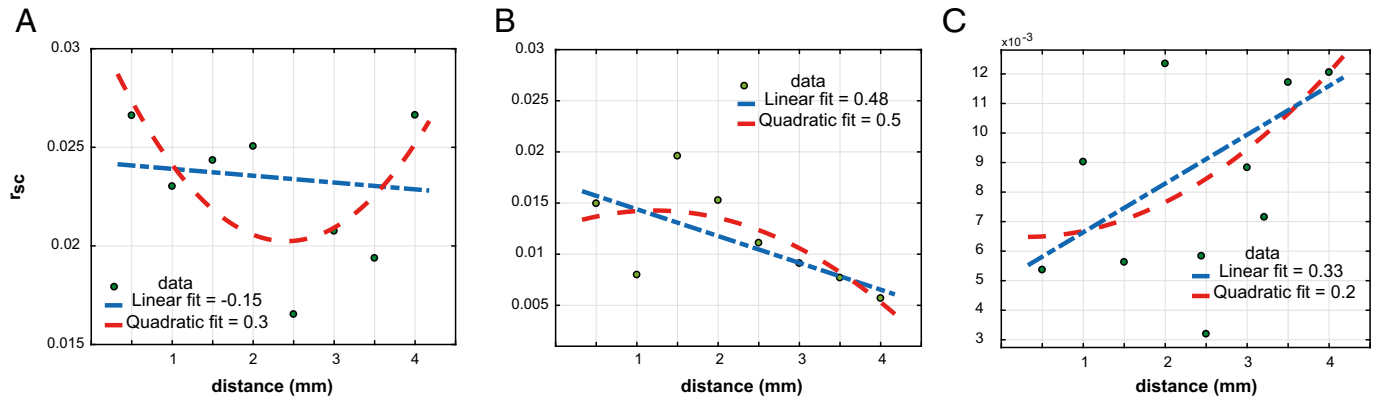


Fig. 7. Fitting trends to the relationship between noise correlations and distance for functionally similar populations. (*A*) Linear ($y = ax + b$) and quadratic ($y = ax^2 + bx + c$) trends fit to noise correlations as a function of distance for functionally similar neurons ($0.1 < r_{signal} < 0.9$) during visual stimulation periods. A negative adjusted R^2 value for the linear fit and a positive R^2 value of 0.3 for the quadratic fit clearly demonstrate a nonmonotonic trend being shaped by visual stimulation. A positive symmetric convexity also points toward a strengthening of local and long-range connectivity during visual stimulation. (*B*) Linear ($y = ax + b$) and quadratic ($y = ax^2 + bx + c$) trends fit to noise correlations as a function of distance for functionally similar neurons ($0.1 < r_{signal} < 0.9$) during intertrial periods. Very similar R^2 values for both the fits (linear = 0.48 and quadratic = 0.5) demonstrate that the quadratic trend is not much different from a linearly decreasing trend. Moreover, the asymmetric negative convexity of the quadratic curve points to a lack of strengthening of local and long-range correlations when no visual stimulation is present. (*C*) Linear ($y = ax + b$) and quadratic ($y = ax^2 + bx + c$) trends fit to noise correlations as a function of distance for functionally dissimilar neurons ($-0.9 < r_{signal} < -0.1$) during visual stimulation periods. The fitting results display a monotonically increasing trend of noise correlations as a function of distance. These pairs of neurons do not display the characteristic positive convexity shown by functionally similar neurons where local and distant populations have equivalent correlated variability, pointing to a different mechanism of functional connectivity driving this relationship.

distance, is largely considered a canonical feature of functional connectivity. Our results suggest that this spatial decay is not observed in the PFC, since nearby and distant neurons are correlated to the same degree, thus reflecting a fundamentally different lateral functional connectivity structure compared with primary sensory areas like V1 (14, 17, 24, 25). Such a functional connectivity pattern is likely to directly reflect the underlying anatomical organization of prefrontal neural populations into spatially distributed clusters connected through local and long-range excitatory collaterals (28, 32, 36). Indeed, in awake-state recordings, the spatially inhomogeneous correlation pattern reflected bumps of ∼1.5- to 2-mm width (Fig. 2*B*), which closely matches the spatial distribution (∼1.5 mm maximum width) of laterally labeled stripes of neuronal assemblies in supragranular prefrontal layers [Kritzer and Goldman-Rakic (28) and Pucak et al. (32)].

Although purely anatomical methods cannot identify functional similarities across connected populations (and vice versa), an influential hypothesis of structural connectivity in the PFC assumes that long-range excitatory collaterals target clusters of neurons with similar functional preference, like spatial tuning (39). Our results provide experimental evidence supporting this hypothesis, since correlated variability of functionally similar neurons was a major source of spatial inhomogeneities, on a spatial scale that closely matches the anatomical estimates of periodicities in lateral connections and associational input. In contrast, functionally dissimilar neurons showed a strengthening of correlated variability across distance, but did not display any clear periodicity. Interestingly, despite a columnar structure of orientation preference in V1 [Hubel et al. (41)], correlated variability was significantly lower for distant populations, potentially reflecting a much weaker influence of lateral connections. Although a definite answer to the exact relationship between structural and functional connectivity in the PFC could be provided in the future from functional anatomy techniques, the spatial scale consistency across anatomical and functional connectivity measures seems to suggest that, indeed, structural connectivity is likely to cluster functionally similar prefrontal populations into local and distant functionally connected ensembles.

The spatial pattern of horizontal connections could be one likely source of the nonmonotonic correlation structure in PFC. Another source could be ascribed to spatially distributed input from associational or thalamic areas to the PFC (35, 42, 43). Regardless of the underlying mechanism, the impact of spatially clustered, similarly tuned, correlated prefrontal neurons to distant cortical and subcortical targets may facilitate the role of PFC in large-scale transmission and integration of information. Specifically, such prefrontal clusters could be thought of as separate channels of information that project to distant cortical and subcortical areas (42, 43). Correlated prefrontal output could coordinate these distant targets and therefore contribute to large-scale information processing.

Spatial Structure of Prefrontal Correlated Variability and Integrative Processing. PFC is a central subnetwork playing a crucial role in cognitive computations due to an increase in the integrative aspect of information processing in higher-order cortical areas (44, 45). This progressive increase in integrative functions across the cortical hierarchy was recently suggested to be mediated by a similar hierarchy in the timescales of intrinsic fluctuations that arise due to systematic changes in the anatomical structure, like heterogeneous connectivity of local circuitry (6, 46, 47). A nonmonotonic spatial structure of correlated variability differentiates prefrontal functional connectivity from primary sensory areas and could therefore be relevant to the emergence of prefrontal-specific timescales (6, 46–48). Network topology was recently suggested to affect timescales since physical distance between connected nodes was shown to increase as timescale lengthened (49).

From a graph-theoretical perspective, a spatially inhomogeneous connectivity profile, combining strong local and long-range functional connectivity, similar to what we observed in the PFC for functionally similar populations, could reflect a network with shorter average path length and higher average clustering coefficient compared with a network with monotonically decreasing correlations and/or uncorrelated long-range functional connectivity (like V1) (50). These topological features are known to facilitate efficient integrative processing (51, 52) and could reflect a fundamental characteristic of laterally organized prefrontal microcircuits compared with primary visual cortex, where, despite positive local correlations, long-range activity on the same spatial scale is uncorrelated (25).

Some recent findings shed light on the spatial functional organization of prefrontal populations that could be critical for integrative processing (53–55). Kiani et al. (54) revealed a natural grouping of prefrontal neurons into isolated clusters that remained stable across various conditions (e.g., different epochs

of task, spontaneous activity), therefore suggesting that intrinsic lateral connections play a prominent role in shaping functional parcellation in PFC. In another study, Markowitz et al. (55) found that different working memory stages are implemented in the PFC by spatially and functionally segregated subnetworks. More importantly, the spiking activity of these subnetworks during working memory is coordinated, indicating a distributed network that integrates different aspects of working memory through long-range interactions. Our findings, revealing spatially distributed clusters of correlated neurons with similar feature selectivity, provide further evidence for the existence and function of long-range functional interactions within the PFC, which seems to be instrumental for higher-order integrative processing.

Comparison with Previous Studies of Correlated Variability in the PFC. Experimental constraints prevented previous studies in dorsolateral areas of the PFC, around the principal sulcus, from capturing a nonmonotonic correlation structure (9, 56, 57). These studies were constrained by a maximum interelectrode distance of 1 mm, and our findings up to this distance were indeed consistent, showing a decrease in correlations up to 1 mm.

A number of other factors might also have prevented previous studies that used Utah arrays in other areas of the PFC to capture the nonmonotonic spatial structure of correlations that we report here. First, it is likely that the nonmonotonic structure is specific for this particular region of PFC, i.e., vlPFC, since none of these studies involved recordings in the vlPFC, but rather in area 8A (i.e., the frontal eye fields) (54, 58). The source of this region-specific discrepancy between our results and previous studies (54, 58) could be potentially traced to differences in the involvement of various prefrontal regions in visual processing. For example, the probability of finding feature-selective neurons (e.g., direction selective neurons) may be higher in the vlPFC compared with area 8A [Hussar and Pasternak (59)]. Since our data validated the spatially nonmonotonic correlation structure during visual stimulation with movie clips and direction of motion, the lack of a similar spatial structure in the frontal eye fields could be due to its differential functional role.

Leavitt et al. (58) recorded using 4×4-mm Utah arrays in area 8A and found a monotonically decreasing correlation structure. However, hardware limitations allowed them to record simultaneously from blocks of only 32 electrodes each time, limiting the spatial coverage that would prevent an extensive examination of the potential spatial anisotropy in area 8A. Kiani et al. (54), using the same electrode arrays, recorded simultaneously from all 96 electrodes and also reported a monotonic decrease of correlations for multiunit activity for distances up to 4 mm. However, the length of electrodes was 1.5 mm, compared with the 1-mm length used in our recordings. Therefore, the monotonically decreasing correlations might be due to layer-specific effects as reported in primary visual cortex (60, 61).

Comparison with Previous Studies of Correlated Variability in Primary Visual Cortex. Rosenbaum et al. (25) recently provided evidence for a nonmonotonic correlation structure in superficial layers (L2/3) of primary visual cortex without strong long-range correlations. In particular, they reanalyzed data collected with Utah arrays during anesthesia and, only after removing the effect of latent shared variability, found that nearby neurons were weakly but significantly correlated, neurons at intermediate distances were negatively correlated, and distant neurons were uncorrelated (r_{sc} not different from 0).

There are some major differences between this study and our results from prefrontal recordings on the same spatial scale. First, the average correlation coefficient for distant (3–4 mm apart) neurons in these V1 recordings was not different from zero, which implies an absence of correlation rather than weak correlation between distant populations. In striking contrast, the average magnitude of long-range correlations for the same distance in the awake PFC recordings was (*i*) significantly positive and (*ii*) comparable to the magnitude of correlations for nearby neurons. This suggests that long-range (3–4 mm) functional connectivity in PFC is stronger and in fact results in significant long-range correlations compared with the primary visual cortex, where, despite a weak nonmonotonicity, the average correlated variability between distant neurons is not different from zero. The second, and more important, difference pertains to the conditions under which the nonmonotonic structure was detectable. The Rosenbaum et al. (25) results in V1 suggested an underlying nonmonotonic functional connectivity that was washed out by the strong modulatory effects of global state fluctuations (e.g., up and down states) observed during anesthesia in macaques (12) and in rodents during anesthesia and quiet wakefulness (62, 63). Specifically, the nonmonotonic correlation structure was revealed only after removing the effect of global latent fluctuations via Gaussian process factor analysis (GPFA). This suggests that a nonmonotonic structure in the primary visual cortex should be directly detectable in data recorded from awake animals where the anesthesia-induced global fluctuations are absent. However, to the best of our knowledge, until now there has been no direct experimental evidence in awake V1 recordings. In contrast, Ecker et al. (24) found a flat correlation structure in awake V1 recordings using tetrode arrays, which was also revealed after removing latent fluctuations from anesthetized recordings using the same technique (12). Regardless of the underlying reason for this discrepancy between the two above-mentioned studies in the V1 (e.g., layer specificity or the effect or number of samples), our recordings in the PFC provide direct evidence for a nonmonotonic, long-range correlation structure during wakefulness, without the need for removing latent sources of covariance, i.e., without application of GPFA or any other similar tool involving theoretical assumptions like stationarity of responses or the number of latent factors that contribute in driving correlated variability.

Conclusion

Overall, our results suggest that the mesoscopic functional connectivity architecture of vlPFC is fundamentally different compared with early sensory cortices such as V1 or V4. Correlated variability in the vlPFC is spatially nonmonotonic, and a major source of nonmonotonicity is the spatial pattern of correlations between neurons with similar functional properties. A nonmonotonic functional connectivity profile with strong and equivalent local and long-range interactions might reflect the underlying machinery for large-scale coordination of distributed information processing in the PFC.

Experimental Procedures

Electrophysiological Recordings. Extracellular electrophysiological recordings were performed in the inferior convexity of the lateral PFC of two anesthetized and two awake adult, male rhesus macaques (*Macaca mulatta*) by using Utah microelectrode arrays [Blackrock Microsystems (37)]. The array (4×4 mm with a 10×10 electrode configuration and interelectrode distance of 400 μm) was placed 1–2 mm anterior to the bank of the arcuate sulcus and below the ventral bank of the principal sulcus, thus covering a large part of the inferior convexity in the vlPFC (Fig. 1*A*). For the awake experiments, monkeys were implanted with form-specific titanium head posts on the cranium after modeling the skull based on an anatomical MRI scan acquired in a vertical 7T scanner with a 60-cm-diameter bore (Biospec 47/40c; Bruker Medical). The methods for surgical preparation and anesthesia have been described (64–66). All experiments were approved by the local authorities (Regierungspräsidium) and were in full compliance with the guidelines of the European Community (European Union Vendor Declaration 86/609/EEC) for the care and use of laboratory animals.

Data Acquisition and Spike Sorting. Broadband neural signals (0.1–32 kHz in the anesthetized recordings and 0.1–30 kHz in the awake recordings) were recorded by using a Neuralynx (Digital Lynx) data-acquisition system for the anesthetized recordings and Neural Signal Processors (Blackrock Microsystems) for the awake recordings.

In the anesthetized data, to detect spiking activity, we first bandpass-filtered (0.6–5.8 kHz) the broadband raw signal using a minimum-order finite impulse response filter (67) with 65-dB attenuation in the stop bands and <0.002-dB ripple within the pass band. A Gaussian distribution was fit to

randomly selected chunks of the filtered signal to compute the noise variance, and the amplitude threshold for spike detection was set to five times the computed variance. Spike events with interspike intervals less than a refractory period of 0.5 ms were eliminated. Those events that satisfied the threshold and refractory period criteria were kept for spike sorting.

In the awake experiments, broadband data were filtered between 0.3 and 3 kHz by using a second-order Butterworth filter. The amplitude for spike detection was set to five times the median absolute deviation (68). The criterion for rejection of spikes was the same as described above. All of the collected spikes were aligned to the minimum. For spike sorting, 1.5 ms around the peak, i.e., 45 samples, were extracted.

Automatic clustering to detect putative single neurons in both the awake and anesthetized data were achieved by a split and merge expectation–maximization (SMEM) algorithm that fits a mixture of Gaussians to the spike feature data which consisted of the first three principal components. For the anesthetized data, the SMEM algorithm by Ueda et al. (69) was used. Details of the spike-sorting method used in this study have been described using tetrodes (24, 70). For the awake data, the KlustaKwik algorithm (71, 72) was used. The spike-sorting procedure was finalized in both cases through visual inspection by using the program Klusters (73).

Visual Stimulation. In anesthetized recordings, full-field visual stimulation of 640×480 resolution with 24-bit true color at 60 Hz for each eye was presented by using a Windows machine equipped with an OpenGL graphics card (Wildcat series; 3DLABS). We used 10-s epochs from a commercially available movie [*Star Wars Episode I*, the Battle of Naboo (74)]. Hardware double buffering was used to provide smooth animation. The experimenter's monitor and the video interface of a fiber-optic stimulus presentation system (Silent Vision; Avotec) were driven by the VGA outputs. The field of view was 30 (horizontal) $\times$ 23 (vertical) degrees of visual angle, and the focus was fixed at two diopters. Binocular presentation was possible through two independently positioned plastic, fiber-optic glasses; however, in this study, we used monocular stimulation (either left or right eye). The contact lenses for the eyes had matched diopter with an Avotec projector to focus images on the retina. Positioning was aided by a modified fundus camera (RC250; Carl Zeiss) with an attachment to hold the projector on the same axis of the camera lens. After observing the foveal region, the projector was fixed relative to the animal.

In the awake recordings, the visual stimuli were generated by in-house software written in C/Tcl and used OpenGL implementation. Stimuli were displayed by using a dedicated graphics workstation (TDZ 2000; Intergraph Systems) with a resolution of 1,280 $\times$ 1,024 and a 60-Hz refresh rate. An industrial PC with one Pentium CPU (Advantech) running the QNX real-time operating system (QNX Software Systems) controlled the timing of stimulus presentation, digital pulses to the Neuralynx system (anesthetized) or the Blackrock system (awake), and acquisition of images. Eye movements were captured by using an IR camera at 1-kHz sampling rate using the software iView (SensoriMotoric Instruments GmBH). They were monitored online and stored for offline analysis by using both the QNX-based acquisition system and the Blackrock data-acquisition system.

In the anesthetized recordings, neural activity was recorded in 200 trials of repeated stimulus presentation. Each trial consisted of the same 10-s-long movie clip, followed by 10 s of a blank screen (intertrial). In the awake experiments, two monkeys were trained to fixate on a red square of size 0.2° of visual angle subtended on the eye ~45 cm from the monitors and maintain fixation within a window of 1.5–2° of visual angle. The location of the red fixation square was adjusted to the single eye vergence of each individual monkey. After 300 ms of fixation, a moving grating of size 8°, moving at a speed of 12° (monkey H) and 13° (monkey A) per second with a spatial frequency of 0.5 cycles per degree of visual angle and at 100% contrast was presented for 1,000 ms. The gratings encompassed eight different directions of motion, viz. 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° (Fig. 1*B*), pseudorandomized within a block of eight trials. After 1,000 ms, a 300-ms stimulus-off period preceded the completion of the trial. The monkeys were given a liquid reward (either water or juice) at the end of the trial, if they maintained fixation within the specified fixation window during the entire duration of the trial. Every successful trial was followed by a 1,000-ms intertrial period. On average, we found $32 \pm 5\%$ of all recorded neurons to be visually modulated. The stimuli, although presented through a stereoscope (due to the data being collected on the same day with other experiments requiring dichoptic viewing conditions), were always presented monocularly in the left eye to remain consistent with the monocular stimulation protocol used in the anesthetized recordings. In both anesthetized and awake recordings, to ensure accurate control of stimulus presentation, a photodiode was attached to the experimenter's monitor, permitting the recording of the exact presentation time of every single frame.

In the awake recordings, spontaneous activity datasets were collected on days different from those of the task recordings. The monkeys were allowed to move their eyes freely or have their eyes closed. The recording chamber was sound-resistant and dark. In the anesthetized recordings, spontaneous activity datasets were recorded between periods of visual stimulation. In both the awake and anesthetized recordings of spontaneous activity, the monitors were turned off. The duration of each spontaneous activity dataset was between 40 and 80 min.

Tuning Functions and Signal Correlations. Tuning curves for each detected single unit were computed by averaging the firing rate across trials for each of the eight presented directions of motion. Signal correlations, defined as the correlation coefficient between the tuning curves of a neuronal pair, were also computed (7). In addition to classical tuning curves (direction and orientation selectivity), other types of tunings, such as inverted tunings, for example, have also been reported in the electrophysiological studies of the macaque PFC (75). Because of this variability in the observed tuning properties of detected single units, signal correlation provides a more general measure of response similarity, and therefore it was used to investigate the correlation structure that arises from this functional similarity.

Spike Count Correlations. To compute spike count correlation (r_{sc}) during the anesthetized state, we divided the period of visual stimulation into 10 periods, each being 1,000 ms long, and considered these periods as different successive stimuli. The intertrial period was also binned in the same way. In the awake data, visual stimulation and intertrial periods were 1,000 ms long each, thus being consistent with the anesthetized experiments. We estimated spike counts over 1,000 ms due to the stimulus length used in previous studies of correlated variability. In spontaneous datasets (both anesthetized and awake), the entire length of the recording epoch was split into periods of 1,000 ms that were treated as a trial.

The spike count correlation coefficients were computed similarly to previous studies in primary visual areas (10, 24, 64). First, for each condition (either presentation of each moving grating in awake experiment or a single bin of movie clip in the anesthetized experiment), we normalized the spike counts across all trials by converting them into *z* scores (10). For each pair, we computed the Pearson's correlation coefficient for the two vectors z_i and z_j as follows:

$$c(r_i, r_j) = E[z_i z_j]. \quad [1]$$

After computing $c(r_i, r_j)$ for each condition, we averaged across conditions to obtain the correlation value. Equivalently, one can concatenate z scores for all of the conditions in long vectors and find the expectation of their product. To account for possible nonphysiological correlations between detected neurons, which could happen, for example, due to shorts between recording electrodes, a threshold of 5 SDs above the mean correlation value was set, and the outliers were discarded.

Quantification of Spatial Inhomogeneities in Correlated Variability. We quantified the inhomogeneity in the spatial structure of correlated variability across different conditions and states by computing the mean of the absolute rate (i.e., first differential) of correlation changes across lateral distance. To estimate the first differential with respect to distance, we subtracted the mean correlation values of consecutive bins that were significantly different (Wilcoxon rank-sum test, alpha level 0.05). If no significant change between two consecutive bins was observed, the derivative at that point was set to zero.

Curve Fitting Procedures. A two-parameter line ($y = ax + b$) and a three-parameter quadratic function ($y = ax^2 + bx + c$) were fit via a minimization of the least-squared error to the results in Fig. 6 *B* and *D* by using the in-built Curve Fitting Toolbox in MATLAB (Version 2016b). The chosen functions (40) were fit to the mean noise correlation functions, which were weighted by the SEM of each data point as the individual data points spanned varying number of observations.

ACKNOWLEDGMENTS. We thank Britni Crocker and Zeynab Razzaghpanah for help with preprocessing of the data and spike sorting; Yusuke Murayama and the other technical and animal care staff for excellent technical assistance; Prof. Andreas Tolias for help with the initial implantations of the Utah arrays; and Dr. Michel Besserve and Christos Constantinidis and Rodrigo Quian Quiroga for their comments on a previous version of this manuscript. This work was supported by the Max Planck Society.

1. Douglas RJ, Martin KAC (2004) Neuronal circuits of the neocortex. *Annu Rev Neurosci* 27:419–451.
2. Harris KD, Mrsic-Flogel TD (2013) Cortical connectivity and sensory coding. *Nature* 503:51–58.
3. Douglas RJ, Martin KAC, Whitteridge D (1989) A canonical microcircuit for neocortex. *Neural Comput* 1:480–488.
4. Douglas RJ, Martin KAC (2007) Mapping the matrix: The ways of neocortex. *Neuron* 56:226–238.
5. Miller KD (2016) Canonical computations of cerebral cortex. *Curr Opin Neurobiol* 37: 75–84.
6. Murray JD, et al. (2014) A hierarchy of intrinsic timescales across primate cortex. *Nat Neurosci* 17:1661–1663.
7. Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. *Nat Neurosci* 14:811–819.
8. Denman DJ, Contreras D (2014) The structure of pairwise correlation in mouse primary visual cortex reveals functional organization in the absence of an orientation map. *Cereb Cortex* 24:2707–2720.
9. Constantinidis C, Goldman-Rakic PS (2002) Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J Neurophysiol* 88:3487–3497.
10. Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *J Neurosci* 21:1676–1697.
11. Rothschild G, Nelken I, Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex. *Nat Neurosci* 13:353–360.
12. Ecker AS, et al. (2014) State dependence of noise correlations in macaque primary visual cortex. *Neuron* 82:235–248.
13. Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. *Phys Rev E Stat Nonlin Soft Matter Phys* 64:051904.
14. Smith MA, Kohn A (2008) Spatial and temporal scales of neuronal correlation in primary visual cortex. *J Neurosci* 28:12591–12603.
15. Schulz DPA, Sahani M, Carandini M (2015) Five key factors determining pairwise correlations in visual cortex. *J Neurophysiol* 114:1022–1033.
16. Chelaru MI, Dragoi V (2016) Negative correlations in visual cortical networks. *Cereb Cortex* 26:246–256.
17. Smith MA, Sommer MA (2013) Spatial and temporal scales of neuronal correlation in visual area V4. *J Neurosci* 33:5422–5432.
18. Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001) Coding specificity in cortical microcircuits: A multiple-electrode analysis of primate prefrontal cortex. *J Neurosci* 21:3646–3655.
19. Amir Y, Harel M, Malach R (1993) Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J Comp Neurol* 334:19–46.
20. Elston GN (2003) Cortex, cognition and the cell: New insights into the pyramidal neuron and prefrontal function. *Cereb Cortex* 13:1124–1138.
21. Lund JS, Yoshioka T, Levitt JB (1993) Comparison of intrinsic connectivity in different areas of macaque monkey cerebral cortex. *Cereb Cortex* 3:148–162.
22. Gochin PM, Miller EK, Gross CG, Gerstein GL (1991) Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Exp Brain Res* 84:505–516.
23. Martin KAC, Roth S, Rusch ES (2014) Superficial layer pyramidal cells communicate heterogeneously between multiple functional domains of cat primary visual cortex. *Nat Commun* 5:5252.
24. Ecker AS, et al. (2010) Decorrelated neuronal firing in cortical microcircuits. *Science* 327:584–587.
25. Rosenbaum R, Smith MA, Kohn A, Rubin JE, Doiron B (2017) The spatial structure of correlated neuronal variability. *Nat Neurosci* 20:107–114.
26. Voges N, Schüz A, Aertsen A, Rotter S (2010) A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Prog Neurobiol* 92:277–292.
27. Angelucci A, et al. (2002) Circuits for local and global signal integration in primary visual cortex. *J Neurosci* 22:8633–8646.
28. Kritzer MF, Goldman-Rakic PS (1995) Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J Comp Neurol* 359:131–143.
29. Tanigawa H, Wang Q, Fujita I (2005) Organization of horizontal axons in the inferior temporal cortex and primary visual cortex of the macaque monkey. *Cereb Cortex* 15: 1887–1899.
30. Fujita I, Fujita T (1996) Intrinsic connections in the macaque inferior temporal cortex. *J Comp Neurol* 368:467–486.
31. Levitt JB, Lewis DA, Yoshioka T, Lund JS (1993) Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *J Comp Neurol* 338:360–376.
32. Pucak ML, Levitt JB, Lund JS, Lewis DA (1996) Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *J Comp Neurol* 376:614–630.
33. Schwartz ML, Goldman-Rakic PS (1984) Callosal and intrahemispheric connectivity of the prefrontal association cortex in rhesus monkey: Relation between intraparietal and principal sulcal cortex. *J Comp Neurol* 226:403–420.
34. Yoshioka T, Blasdel GG, Levitt JB, Lund JS (1996) Relation between patterns of intrinsic lateral connectivity, ocular dominance, and cytochrome oxidase-reactive regions in macaque monkey striate cortex. *Cereb Cortex* 6:297–310.
35. Goldman-Rakic PS, Schwartz ML (1982) Interdigitation of contralateral and ipsilateral columnar projections to frontal association cortex in primates. *Science* 216:755–757.
36. Melchitzky DS, González-Burgos G, Barrionuevo G, Lewis DA (2001) Synaptic targets of the intrinsic axon collaterals of supragranular pyramidal neurons in monkey prefrontal cortex. *J Comp Neurol* 430:209–221.
37. Maynard EM, Nordhausen CT, Normann RA (1997) The Utah intracortical electrode array: A recording structure for potential brain-computer interfaces. *Electroencephalogr Clin Neurophysiol* 102:228–239.
38. Paradiso MA, Meshi D, Pisarcik J, Levine S (2012) Eye movements reset visual perception. *J Vis* 12:11.
39. Goldman-Rakic PS (1995) Cellular basis of working memory. *Neuron* 14:477–485.
40. Mandel J (1981) Fitting curves and surfaces with monotonic and non-monotonic four parameter equations. *J Res Natl Bur Stand* 86:1.
41. Hubel DH, Wiesel TN, Stryker MP (1978) Anatomical demonstration of orientation columns in macaque monkey. *J Comp Neurol* 177:361–380.
42. Giguere M, Goldman-Rakic PS (1988) Mediodorsal nucleus: Areal, laminar, and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. *J Comp Neurol* 277:195–213.
43. Eblen F, Graybiel AM (1995) Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *J Neurosci* 15:5999–6013.
44. Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
45. Modha DS, Singh R (2010) Network architecture of the long-distance pathways in the macaque brain. *Proc Natl Acad Sci USA* 107:13485–13490.
46. Chen J, Hasson U, Honey CJ (2015) Processing timescales as an organizing principle for primate cortex. *Neuron* 88:244–246.
47. Chaudhuri R, Knoblauch K, Gariel M-A, Kennedy H, Wang X-J (2015) A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* 88:419–431.
48. Chaudhuri R, Bernacchia A, Wang X-J (2014) A diversity of localized timescales in network activity. *elife* 3:e01239.
49. Timme N, et al. (2014) Multiplex networks of cortical and hippocampal neurons revealed at different timescales. *PLoS One* 9:e115764.
50. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:186–198.
51. Latora V, Marchiori M (2003) Economic small-world behavior in weighted networks. *Eur Phys J B* 32:249–263.
52. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. *Phys Rev Lett* 87:198701.
53. Masse NY, Hodnefield JM, Freedman DJ (2017) Mnemonic encoding and cortical organization in parietal and prefrontal cortices. *J Neurosci* 37:6098–6112.
54. Kiani R, et al. (2015) Natural grouping of neural responses reveals spatially segregated clusters in prearcuate cortex. *Neuron* 85:1359–1373.
55. Markowitz DA, Curtis CE, Pesaran B (2015) Multiple component networks support working memory in prefrontal cortex. *Proc Natl Acad Sci USA* 112:11084–11089.
56. Tsujimoto S, Genovesio A, Wise SP (2008) Transient neuronal correlations underlying goal selection and maintenance in prefrontal cortex. *Cereb Cortex* 18:2748–2761.
57. Sakurai Y, Takahashi S (2006) Dynamic synchrony of firing in the monkey prefrontal cortex during working-memory tasks. *J Neurosci* 26:10141–10153.
58. Leavitt ML, Pieper F, Sachs A, Joober R, Martinez-Trujillo JC (2013) Structure of spike count correlations reveals functional interactions between neurons in dorsolateral prefrontal cortex area 8a of behaving primates. *PLoS One* 8:e61503.
59. Hussar CR, Pasternak T (2009) Flexibility of sensory representations in prefrontal cortex depends on cell type. *Neuron* 64:730–743.
60. Hansen BJ, Chelaru MI, Dragoi V (2012) Correlated variability in laminar cortical circuits. *Neuron* 76:590–602.
61. Smith MA, Jia X, Zandvakili A, Kohn A (2013) Laminar dependence of neuronal correlations in visual cortex. *J Neurophysiol* 109:940–947.
62. Hahn TTG, Sakmann B, Mehta MR (2007) Differential responses of hippocampal subfields to cortical up-down states. *Proc Natl Acad Sci USA* 104:5169–5174.
63. Petersen CCH, Hahn TTG, Mehta M, Grinvald A, Sakmann B (2003) Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. *Proc Natl Acad Sci USA* 100:13638–13643.
64. Belitski A, et al. (2008) Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. *J Neurosci* 28:5696–5709.
65. Logothetis NK, Guggenberger H, Peled S, Pauls J (1999) Functional imaging of the monkey brain. *Nat Neurosci* 2:555–562.
66. Logothetis N, Merkle H, Augath M, Trinath T, Ugurbil K (2002) Ultra high-resolution fMRI in monkeys with implanted RF coils. *Neuron* 35:227–242.
67. Rabiner LR, McClellan JH, Parks TW (1975) FIR digital filter design techniques using weighted Chebyshev approximation. *Proc IEEE* 63:595–610.
68. Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* 16:1661–1687.
69. Ueda N, Nakano R, Ghahramani Z, Hinton GE (2000) SMEM algorithm for mixture models. *Neural Comput* 12:2109–2128.
70. Tolias AS, et al. (2007) Recording chronically from the same neurons in awake, behaving primates. *J Neurophysiol* 98:3780–3790.
71. Kadir SN, Goodman DFM, Harris KD (2014) High-dimensional cluster analysis with the masked EM algorithm. *Neural Comput* 26:2379–2394.
72. Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsáki G (2000) Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol* 84:401–414.
73. Hazan L, Zugaro M, Buzsáki G (2006) Klusters, NeuroScope, NDManager: A free software suite for neurophysiological data processing and visualization. *J Neurosci Methods* 155:207–216.
74. Lucas G (1999) *Star Wars: Episode I—The Phantom Menace* [motion picture], director Lucas G (Lucasfilm).
75. Zhou X, Katsuki F, Qi X-L, Constantinidis C (2012) Neurons with inverted tuning during the delay periods of working memory tasks in the dorsal prefrontal and posterior parietal cortex. *J Neurophysiol* 108:31–38.

Decoding the contents of consciousness from prefrontal ensembles

Vishal Kapoor[1,†,*], Abhilash Dwarakanath[1,†], Shervin Safavi[1,2], Joachim Werner[1], Michel Besserve[1,3], Theofanis I. Panagiotaropoulos [1,4,*,#], Nikos K. Logothetis[1,5,#]

1. Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany
2. International Max Planck Research School, Tübingen 72076, Germany
3. Department of Empirical Inference, Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany
4. Cognitive Neuroimaging Unit, CEA, DSV/I2BM, INSERM, Universite Paris-Sud, Universite Paris-Saclay, Neurospin Center, 91191 Gif/Yvette, France
5. Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK

[†]*These authors contributed equally to this work*

 # Equal Senior-authorship

**Correspondence: vishal.kapoor@tuebingen.mpg.de,*

theofanis.panagiotaropoulos@tuebingen.mpg.de

ABSTRACT

Multiple theories attribute to the primate prefrontal cortex a critical role in conscious perception. However, opposing views caution that prefrontal activity could reflect other cognitive variables during paradigms investigating consciousness, such as decision-making, monitoring and motor reports. To resolve this ongoing debate, we recorded from prefrontal ensembles of macaque monkeys during a no-report paradigm of binocular rivalry that instigates internally driven transitions in conscious perception. We could decode the contents of consciousness from prefrontal ensemble activity during binocular rivalry with an accuracy similar to when these stimuli were presented without competition. Oculomotor signals, used to infer conscious content, were not the only source of these representations since visual input could be significantly decoded when eye movements were suppressed. Our results suggest that the collective dynamics of prefrontal cortex populations reflect internally generated changes in the content of consciousness during multistable perception.

One sentence summary

Neural correlates of conscious perception can be detected and perceptual contents can be reliably decoded from the spiking activity of prefrontal populations.

INTRODUCTION

One of the most elusive problems in science is to understand the biological basis of consciousness (*1–3*). A seminal paper almost 30 years ago, incited researchers that *"the time is ripe for an attack on the neural basis of consciousness"* and proposed conscious visual perception as a form of consciousness that is within the reach of neuroscience (*4*). Since then, several theoretical treatises (*5–8*) including the frontal lobe hypothesis (*6*), the higher order (HOT) (*8*) and global neuronal workspace (GNW) theories (*5*, *9*) have suggested a critical role for the brain's prefrontal cortex (PFC) in mediating conscious perception. Evidence supporting its involvement comes from functional magnetic resonance imaging (fMRI) (*9*, *10*), experience of visual hallucinations upon electrical stimulation of the region (*11*, *12*), impaired conscious processing following PFC lesions in patients (*13–18*) as well as intracortical recordings of neural activity (*19–22*). In contrast, alternative theories like the IIT (integrated information theory) identifies the neural elements mediating consciousness as the system having maximal internal cause-effect power (*23*). Its proponents among others have suggested that the neural substrates underlying conscious perception can be traced to a "posterior cortical hot zone", with PFC being primarily critical for processing the behavioral and cognitive consequences of conscious perception (*24–26*) like task demands and monitoring, introspection and motor reports, rather than consciousness per se (*26–29*). For example, frontal cortex was found to be dramatically more active during motor reports, when blood-oxygen-level-dependent (BOLD) fMRI signal modulation was compared between reported and unreported spontaneous changes in the content of consciousness (*27*, *30*, *31*). Additionally, reduced frontal activation accompanied inconspicuous and unreportable switches in perception, when contrasted against a condition, where perceptual changes were easily discernible (*29*). Together, these reports could suggest that frontal activity is related to consequences of perception, thus casting doubt on its role in conscious content representations (*13*, *25*, *26*, *32*). However, the univariate fMRI

analysis comparing report vs no-report conditions (*27*) as well as the indirect nature and limited spatial resolution of BOLD fMRI signal compared to neuronal recordings leaves open the possibility that prefrontal ensembles do reflect the content of consciousness even without report requirements.

We examined this hypothesis by simultaneously probing the discharge activity of large neural populations in the inferior convexity of the macaque PFC with multielectrode arrays during a no-report binocular rivalry (BR) paradigm. BR belongs to the family of multistable perceptual phenomena (*7*, *33*, *34*), which allow a dissociation of conscious perception from sensory input, by inducing in an observer spontaneous fluctuations in the content of consciousness without a change in sensory stimulation. BR instigates these perceptual fluctuations through presentation of incongruent, dichoptic visual input to corresponding retinal locations, resulting in stochastic, endogenously driven alternations in subjective perception. For a certain duration, only one of the two images is consciously experienced, while the other is perceptually suppressed (*33*, *34*). Similar to other paradigms utilized in investigating conscious perception, the standard practice in BR requires humans and macaques to manually report their percepts (e.g. by pressing levers). Therefore neural activity related to consciousness could be conflated with signals related to its consequences such as voluntary motor reports, decision making and introspection (*35–37*). Objective indicators of perception provide a solution to this problem. For example during rivalry between opposing directions of motion, the polarity of optokinetic nystagmus (OKN) reflex, a combination of smooth pursuit and fast saccadic eye movements, is known to be tightly coupled to the reported direction of motion (*38–41*). Thus, the reflexive nature of OKN can be exploited as an objective measure of changes in the content of consciousness (*41*), removing any confounds in the neural activity originating from voluntary motor reports.

We therefore combined neural ensemble recordings with no-report BR between opposing directions of motion and found that prefrontal neural activity reflects the OKN-inferred content of visual consciousness during spontaneous perceptual switches. Monitoring simultaneously large neural populations allowed us to observe for each spontaneous perceptual transition, the collective dynamics of neuronal ensembles representing two percepts that compete for access to consciousness. Conscious content could be successfully decoded from feature specific ensemble vectors in single instances of internally generated perceptual transitions, thus reinforcing the role of prefrontal populations in mediating conscious perception.

RESULTS

We exposed two rhesus macaques to a no report BR paradigm, which consisted of two trial types, physical alternation (PA) and binocular rivalry (BR) (see Figure 1A and methods). Each trial started with a fixation spot cueing the animal to initiate fixation, which lasted ~300 milliseconds, followed by an upward or downward drifting stimulus, presented monocularly for 1 or 2 seconds. After this initial phase, during BR trials, a second stimulus drifting in the opposite direction was presented to the contralateral eye, typically inducing perceptual suppression of the first stimulus, a phenomenon known as binocular flash suppression (BFS) (*42*, *43*). Following this period, visual competition ensued, resulting in spontaneous perceptual switches between the two opposing directions of motion. In contrast, PA trials consisted of exogenously driven changes in perception by alternating monocular presentations of the same stimuli used for instigating BR.

The rivaling, oppositely drifting stimuli elicited optokinetic nystagmus (OKN), (Figure 1B), which served as a reliable indicator for the contents of perception (*38–41*). BFS resulted in a switch in the polarity of the OKN, indicating perceptual dominance of the newly presented

direction of motion (marked in grey). Following this initial period of externally induced perceptual suppression, OKN polarity could be observed switching again, occasionally more than once until the end of the trial, indicating spontaneous, internally driven, changes in conscious contents. We evaluated the onset and offset of perceptual dominance periods in every trial based on the stability of the OKN pattern (see methods). These well-defined epochs of stable perceptual dominance during both BFS and BR displayed a gamma distribution (Figure 1C), typical of multistable perception dynamics (*44*) with an average dominance duration of 3.22 ± 0.102 and 1.92 ± 0.037 seconds respectively.

We targeted the inferior convexity of the PFC (Figure 2A and methods), where neurons display selective responses to complex visual stimuli as well as direction of motion (*45–47*). Neural representations of conscious content are directly related to such feature selectivity. If neurons during BR reliably increase their firing rate each time their preferred stimulus is perceived and suppress their activity, when it is perceptually suppressed, then they explicitly represent conscious content. Figure 2B displays discharges of four such, simultaneously recorded prefrontal units and OKN during a single BR trial. Two units (33 and 119) fired more when downward drifting stimulus was presented while the other two (44 and 167) displayed stronger modulation for the opposite direction of motion during a PA trial (Supplementary Figure 2.1). Spiking activity of these units was also correlated with subjective changes in conscious perception in a BR trial, for both externally induced perceptual suppression (BFS) and a subsequent spontaneous switch in conscious content (BR) (Figure 2B and D).

We analyzed separately the spiking activity during perceptual dominance and suppression periods either (i) externally induced during BFS, or (ii) brought about by an endogenous spontaneous switch, in BR. Selectivity of neural activity was analyzed both before and after such perceptual switches and compared to selectivity in corresponding temporal phases from PA trials (see methods). Figure 2D displays the average spike density functions of

units 33 (preferring downward motion) and 167 (preferring upward motion). The two units were recorded simultaneously from distant electrodes on the array (Figure 2C) and displayed robust modulation during both PA and BR trials with spiking activity switching reliably for both externally induced (Wilcoxon rank sum test, during PA trials (temporally analogous phase to flash suppression dominance) for unit 33, $p_{PA\text{-}33} = 2.82*10^{-14}$ and for binocular flash suppression phase during BR trials, $p_{BFS\text{-}33} = 2.39*10^{-6}$; for unit 167, $p_{PA\text{-}167} = 1.13*10^{-17}$ and $p_{BFS\text{-}167} = 1.20*10^{-9}$) and internally driven perceptual switches (Wilcoxon rank sum test, during PA trials (temporally equivalent phase to rivalry dominance) for unit 33, $p_{PA\text{-}33} = 8.72*10^{-15}$ and perceptual dominance during BR trials, $p_{BFS\text{-}33} = 7.18*10^{-8}$; for unit 167, $p_{PA\text{-}167} = 1.49*10^{-4}$ and $p_{BFS\text{-}167} = 7.8*10^{-3}$). The recorded sites displaying similar stimulus preference formed clusters throughout the $16mm^2$ recorded area of the inferior prefrontal convexity (Figure 2C and Supplementary Figure 2.1).

We compared the stimulus selectivity of all recorded units (n = 987 and see methods) during subjective perception in BFS and BR with their selectivity during purely sensory, monocular stimulus presentations in PA trials, using a d' index (see methods) (*20*, *48*). A large majority of units exhibiting significant stimulus selectivity in PA (see Methods), fired more when their preferred stimulus was perceived compared to its perceptual suppression during BR trials (BFS - 85.38 % (292/342) and BR - 76.09 % (277/364)), with 53.8 % (184/342) and 40.38% (147/364) of them being significantly modulated, respectively (Wilcoxon rank sum test, $p<0.05$, also see Table 1) (Figure 3A), suggesting that ongoing perceptual content is robustly encoded in PFC. Moreover, compared to earlier visual regions, where many units display significant modulation during perceptual suppression of their preferred stimulus, which were proposed as part of an inhibitory mechanism independent of the mechanisms of perception, suggestive of non conscious processing (*49*); such units were a small minority in PFC (BFS - 2.92 % (10/342) and BR - 4.12 % (15/364)). Furthermore, many units displayed significant preference only during BFS (26.51 %, 70/264) and BR (34.4 %, 85/247), suggesting

that individual prefrontal units contribute more reliably to conscious perception during visual competition.

Over all, the recorded units displayed considerable heterogeneity in stimulus preference strength (d') during BFS (PA - 0.3985 ± 0.0131, BFS - 0.4471 ± 0.0133) and BR (PA - 0.3075 ± 0.0098, BFS - 0.2719 ± 0.0082) (Figure 3A). Importantly, selectivity strength was a critical factor determining significant perceptual modulation in BFS and BR. For units with strong sensory selectivity (d' > 1), around 90% were significantly perceptually modulated (Wilcoxon rank sum test, $p < 0.05$) (90 % (72/80) for BFS and 86.96 % (40/46) for BR). This indicates that the percentage of perceptually modulated units in the PFC is remarkably similar to the temporal lobe (*50*), thus suggesting that there are at least two cortical regions, where neuronal activity explicitly represents conscious contents. Furthermore, these results show that the activity of prefrontal units correlates with internally driven switches in the subjective perception of more simple visual features like direction of motion, in addition to the externally induced perceptual suppression of faces and more complex stimuli (*20*).

Plotting the population spiking activity averaged across all units which displayed significant modulation and similar preference across the first monocular switch phase of PA and the temporally corresponding flash suppression phase of the BR trials, revealed a strong, early peak response followed by a long sustained response when a preferred stimulus was presented and a dramatic suppression in activity during presentation of the non-preferred stimulus (Figure 3B, upper row). The average population activity during the BFS phase displayed robust perceptual modulation firing more when a preferred stimulus was perceived, and less when a preferred stimulus was suppressed by a non-preferred stimulus stimulating the contralateral eye (Figure 3B, lower row). Similarly, reliable perceptual modulation was observed, when stimuli were perceived following spontaneous changes in perception during BR (Figure 3B middle column) as well as around spontaneous perceptual switches (Figure 3B,

last column, also see Supplementary Figures 3.1 and 3.2). Similar results were obtained when neural activity in PA was aligned to OKN changes (see methods), as in BR trials (Supplementary Figures 3.3, 3.4, 3.5, and Table 2).

Probing the PFC with multi-electrode arrays allowed us to monitor simultaneously, feature specific ensembles, displaying preferential responses to stimuli drifting in opposite directions. We therefore examined the population code for single instances of different types (i.e., upward to downward and downward to upward motion) of exogenous stimulus and endogenous, spontaneous perceptual transitions. Prefrontal ensemble activity correlated with both exogenous stimulus changes in PA and subjective changes in perceptual content during BR trials (Figure 4A). We utilized a multivariate decoding approach (*51*) to assess the reliability with which we could predict conscious perceptual contents from ensemble activity on single cases of perceptual transitions (see methods). During PA switches, the classifier discriminated between the two stimuli strongly above chance (50%), and generalized across the total duration of a given stimulus presentation (Figure 4B, upper row) suggesting a static population code (*52*). Similarly, a classifier trained on BR activity also discriminated between periods of perceptual dominance and suppression for the two competing stimuli and generalized around perceptual switches (Figure 4B, lower row), similar to that observed during stimulus switches. Importantly, strong temporal generalization of the classifier trained and tested across PA and BR before and after a switch, indicates an invariance in the population code representing sensory input and its subjective experience (*51*). This cross trial type generalization was highly significant (permutation test, $p<0.002$, see methods) when it was carried out during two temporal windows (400 ms), before (-200 ms to - 600) and after (200ms to 600 ms) a switch (Figure 4C). Similarly strong decoding of perceptual content was possible in individual datasets (Supplementary Figure 4.1). Together, these results suggest that the prefrontal population code underlying sensory input and subjective perception is not only

similar, but also reliable and robust. Similar results were obtained when PA trials were aligned to the OKN change instead of the digital pulses for stimulus presentation, (Supplementary Figure 4.2).

Finally, given that in our experiments the OKN is tightly linked to perceptual content, we dissociated neural activity related to oculomotor processes from activity related to visual input. For a majority of the recorded units (n = 747), we estimated their preference to direction of motion in a control experiment during two paradigms, namely fixation Off and fixation On. During fixation Off, the presentation of visual motion elicited OKN similar to BR, while during fixation On, the eye movements were suppressed since macaques were required to fixate a centrally presented spot (Supplementary Figure 5.1 and Supplementary Figure 5.3). We focused our analysis on the upward and downward motion directions, used for instigating rivalry, to make a direct comparison with BR. We found that a majority of the units displaying significant stimulus selectivity across the two control paradigms retained their stimulus preference (fix On - 69.56 %, 48/69; fix Off - 56.25 %, 81/144). Only a small percentage of units (fix On - 14.49 %, 10/69 fix Off - 6.94 %, 10/144; Wilcoxon rank sum test, $p < 0.05$) exhibited a significant preference to stimuli with opposing motion content across the two paradigms (Supplementary figure 5.2 for typical tuning curves). Ensemble population PSTHs (see methods) of significantly modulated units during fixation Off or fixation On preferring the same motion direction in both paradigms are displayed in Figure 5A. In both paradigms, average firing rate increased when a preferred motion direction was presented and decreased in response to the non preferred visual input. We investigated if a classifier trained on neural responses to stimuli which elicited OKN could reliably predict the stimuli, when they were viewed with the eye movements suppressed, and vice versa. We observed strongly above chance (50%) decoding accuracy of the classifier during both conditions (Figure 5B). Importantly, a classifier trained on individual paradigms could generalize across them (Figure

5C) and decode with significant accuracy (permutation test, $p<0.002$, see methods) thus suggesting that prefrontal ensemble activity contains stimulus information, and is not just driven by the eye movements. Similar results were obtained when decoding analysis was performed using the trials from the fixation On paradigm, where any eye movements within the fixation window were further controlled (see methods)(Supplementary Figure 5.4). These findings are in line with previous work suggesting that frontal cortex responds to visual motion both in the presence and absence of OKN (*47*, *53*) and suggest that motion content signals contribute to the activity of the tested population. Neurons in this prefrontal region reflect a mixture of perceptual and oculomotor signals (*54*) and are selective for motion stimuli even when the monkeys fixate (*47*). Such comodulation was recently reported in V4 and inferotemporal cortex where microsaccades were found to contribute in attention related neuronal responses (*55*). Future investigations could ascertain, if a similar mechanism is relevant for prefrontal responses in BR.

DISCUSSION

These results suggest that feature selective units in the primate PFC reliably reflect the dynamics of internally generated changes in the content of subjective perception even without voluntary perceptual reports. While addressing an ongoing debate between GNW and IIT about the neural correlates of conscious perception in the PFC (*13*, *25*, *26*, *32*, *56*), we demonstrate that the contents of subjective experience can be reliably decoded from the activity of prefrontal ensembles during single instances of internally driven transitions in conscious perception.

BR offers a distinct advantage over other paradigms of visual consciousness such as BFS or visual masking due to the stochastic, internally driven changes in the subjective perceptual content in the absence of any concomitant changes in visual stimulation (*34*, *49*, *57*). Hence, it confers a unique opportunity to observe neural dynamics contemporaneous with

spontaneous changes in the contents of subjective experience. When paired with electrophysiological investigations of the non human primate visual system (*57*, *58*), BR has revealed that the proportion of feature selective neurons reliably reflecting conscious content increases as one progresses in the visual cortical hierarchy from early visual areas (*48*, *59*, *60*) to later temporal regions (*20*, *21*, *50*). Recent single unit recordings in human medial frontal and anterior cingulate cortex areas during BR found non-selective modulation of neural activity before spontaneous perceptual transitions, suggesting that some frontal cortical areas might reflect the prerequisites of conscious perception than conscious content per se (*22*). In contrast, our results demonstrate conscious content representations in a subregion of the macaque lateral PFC, where cells are selective for faces, complex visual objects and direction of motion (*45*–*47*, *61*, *62*) and reciprocally connected with the inferotemporal cortex (*63*). Importantly, previous electrophysiological studies in the PFC during conscious perception either utilized a motor report (*19*, *21*, *22*, *64*), and were therefore conflated by consequences of conscious perception, or investigated perceptual modulation among neurons selective to faces and complex objects with a no-report BFS paradigm. In BFS, perceptual dominance and suppression are externally induced due to an abrupt and strong change in the feedforward input and not endogenously driven as in BR, wherein neural activity modulations could contribute causally towards changes in conscious perception (*20*). Hence, our results collected during unreported spontaneous transitions in conscious perception unequivocally demonstrate the existence of prefrontal representations of conscious content.

Our findings are in sharp contrast to the conclusions of recent imaging studies showing reduced involvement of the PFC in conscious perception (*27*–*29*). However, constraints in the spatiotemporal resolution of the BOLD signal and its complex relationship with neural activity limit the interpretations from imaging data, especially so, when null findings are reported (*65*, *66*). Such limits in spatiotemporal resolution are particularly relevant to the frontal cortex,

where individual neurons often display a high degree of mixed selectivity (*67, 68*) or distinct temporal patterns of activity during perceptual paradigms (*69*). For example, we find that units displaying preferential responses to opposite directions of motion are frequently distributed in close proximity (~0.4mm) throughout the electrode array (Figure 2C and supplementary figure 2.3). Such spatial variability of stimulus selectivity remains difficult to capture with fMRI. A recent attempt with high-field fMRI offering better spatial resolution could identify clusters activated by competing perceptual representations and reported a relatively low correlation between sensory and perceptual representations in early visual areas, confirming earlier electrophysiological studies (*70*). In contrast, recent work utilizing fMRI in conjunction with multivariate pattern analysis revealed neural correlates of consciously perceived location in anterior brain regions such as the frontal cortex, beyond early visual areas (*71*). Such approaches hold great promise in providing whole brain representations of conscious content.

Utilizing multivariate pattern analysis for decoding the contents of conscious perception from neuronal ensembles in the PFC lays the foundation for a comparative approach using direct neuronal recordings, aimed at investigating the population code subserving conscious contents across cortical regions. It may further help elucidate the details of the mechanism responsible for inducing spontaneous changes in the content of consciousness. In summary, our results demonstrate that prefrontal ensemble activity explicitly reflects internally generated changes in conscious contents since only a miniscule percentage of units fired significantly more when their preferred stimulus is perceptually suppressed. They, therefore lend support to theoretical approaches such as the GNW and HOT, which attribute an essential role for the PFC in mediating consciousness in general and conscious perception in particular (*8, 72, 73*). Interestingly, apart from conscious contents, PFC was recently shown to control also the level of consciousness in rodents (*74*), suggesting a more general role of this area in awareness. While GNW and HOT have recently received criticism because of this region's

functional relevance to cognitive consequences of perception and motor processes, we address with this study, one such confound, namely the volitional motor report (*35*). Future work aimed at elucidating the causal mechanisms of conscious perception not just in the PFC, but the primate brain in general could greatly benefit from employing direct activation of such perceptually modulated ensembles (*75*). In combination with carefully designed experimental approaches, it could help us both understand the relationship and disentangling the mechanisms underlying consciousness from other cognitive processes (*76*) such as introspection (*27*, *77*), attention (*78*), decision making (*79*, *80*) or cognitive control (*37*, *81*).

METHODS

Binocular rivalry paradigm, control paradigms and stimulus presentation

The task consisted of two trial types, namely, the physical alternation (PA) trials and binocular rivalry (BR) trials. Both trial types started with the presentation of a red fixation spot (subtending 0.2 degree of visual angle), cueing the animal to initiate fixation. Upon successful fixation for 300 milliseconds within a fixation window (±8°), a drifting sinusoidal grating (size: 8 degrees (radius), speed: 12-13 degrees/sec, spatial frequency: 0.5 cycles per degree, gratings were drifting vertically up or down) was monocularly presented. During one recording session, we used random dot motion stimulus (field of view 8 degrees (radius), speed 13 degrees/sec, 200 limited lifetime dots and 100% coherence). After one or two seconds, the first stimulus was removed and a second stimulus drifting in the opposite direction was presented in the contralateral eye in PA trials. In BR trials, the second stimulus was added to the contralateral eye without removing the first stimulus. This typically results in perceptual suppression of the first stimulus and is denoted by flash suppression (*20*, *42*, *43*, *48*) in Figure 1A. After this period, visual input alternated physically between oppositely drifting gratings in the PA

condition. In the BR condition, the percept of the animal switched endogenously between the discordant visual stimuli, whose temporal histogram could be approximated with a gamma distribution (Figure 1C). The total duration of a single trial/observation period was between 8-10 seconds. Note that the perception of the animal displayed in Figure 1A is identical in the two conditions, even though the underlying visual input is monocular in PA trials, while it is dichoptic in the case of BR. The eye (where the first stimulus was presented), motion direction (which was presented first) and trial types (PA or BR) were pseudorandomized and balanced in a single dataset. During the entire period of a trial, animals maintained their gaze within a fixation window, which was the same size (±8°) as the stimulus. A liquid reward was given to the animal upon successful maintenance of gaze within the window for the entire trial duration.

The eye movement control experiments using the fixation Off and fixation On paradigms were carried out on a subset of recording sessions (4/6, 2 - H'07, 2 - A'11). Both paradigms consisted of trials, where the macaques were presented with a visual stimulus drifting in one of eight randomly chosen directions for one second (Supplementary Figure 5.1). Each trial started with the presentation of a fixation spot for ~300 milliseconds, following which a drifting visual stimulus was presented for one second. However, there was one key difference across the two paradigms. During fixation Off, the fixation spot disappeared as soon as the visual stimulus was presented, eliciting OKN and the fixation window (the window within which the animal was required to maintain its gaze) was the entire stimulus (±8°). In contrast, during the fixation On paradigm, a fixation spot overlaid on the stimulus and a smaller fixation window (~±1 to ±2 degrees) indicated that the monkeys must fixate to complete the trial and receive reward, thus suppressing eye movements. Stimulus parameters were identical to the ones used during the BR paradigm.

Dichoptic visual stimulation was carried out with the aid of a stereoscope and displayed at a resolution of 1280X1024 on the monitors (running at a 60 Hz refresh rate) using a dedicated

graphics workstation. Prior to the presentation of the BR paradigm, we carried out a previously described calibration procedure (*48*) which ensured that the stimuli presented on the two monitors through the stereoscope were appropriately aligned and could be fused binocularly. It started with the animal participating in a fixation-saccade task, wherein visual input was at first presented monocularly to the left eye. The task required brief period of fixation on a centrally presented red fixation spot (its location was adjusted according to single eye vergence for each individual monkey), following which a peripheral fixation target was presented in one of eight different directions. Animal was trained to make a saccade to the presented target for obtaining a liquid reward. During this period, the eye position was centered within a fixation window, using a custom designed linear offset amplifier. After this, a second procedure was carried out, wherein the fixation target was first presented to the left eye for a brief duration, after which it was switched off and immediately presented to the right eye. The animal typically responded with a saccade, whose amplitude, provided an estimate of the offset between the fixation spot displayed on the two monitors. This offset was confirmed with several repetitions of this procedure and it served as a correction factor. The visual stimuli were aligned taking into account this correction factor, thus enabling their fusion.

The visual stimuli and the task were designed with an in-house software written in C/Tcl. A QNX real-time operating system (QNX Software Systems) managed the precise temporal presentation of the visual stimuli, and sent digital pulses to the Blackrock recording system. An infrared camera captured eye movements (1kHz sampling rate) with the software iView (SensoriMotoric Instruments GmBH, Germany). Besides monitoring eye movements online, they were also stored for offline analysis in both QNX-based acquisition system as well as Blackrock neural data acquisition system. We used the latter to align the neural data.

Surgical procedures

Two healthy rhesus monkeys (*Macaca mulatta*), H'07 and A'11 participated in behavioral and electrophysiological recordings. All experiments were approved by the local authorities (Regierungspräsidium, protocol KY6/12 granted to TIP as the principal investigator) and were in full compliance with the guidelines of the European community (EUVD 86/609/EEC) for the care and use of laboratory animals . Each animal was implanted with a cranial headpost (material: titanium) custom designed to fit the skull based upon a high resolution MR scan collected using a 4.7 tesla scanner (Biospec 47/70c; Bruker Medical, Ettlingen, Germany). The headpost implantation was carried out while the animal was under general anesthesia and prior to the beginning of behavioral training in the BR paradigm. Details of the surgical procedures have been previously described (*82*). The MR scan also aided in localizing the inferior convexity of the LPFC. Post behavioral training in the task, the animals underwent another surgery, where a Utah microelectrode array (Blackrock Microsystems, Salt Lake City, Utah USA; (*83*)) was implanted in the PFC. The array had a 10 by 10 electrode configuration and was 4mm by 4 mm in size, with an inter-electrode distance of 400μm and electrode length of 1 mm. We implanted the array ventral to the principal and anterior to the arcuate sulcus, thus aiming to cover a large part of the inferior convexity in the ventrolateral PFC (Figure 2A).

Electrophysiology data acquisition

All behavioral training and electrophysiological recordings were carried out with the animals seated in a custom designed chair. Data presented here was collected across six recording sessions in two macaques (4 - H'07 and 2 - A'11). Broadband neural signals (0.1 - 30 kHz) were recorded with the Neural Signal Processors (Blackrock Microsystems) and band-pass filtered offline between 0.6 - 3 kHz using a 2nd order Butterworth filter. Spikes were detected with an amplitude threshold set at five times the median absolute deviation (*84*). Any

spike events larger than 50 times the mean absolute deviation were discarded. Further, spike events with an inter-spike interval of less than the refractory period of 0.5 ms were also discarded. Events satisfying the aforementioned criterion of threshold and the refractory period were kept for further analysis. Collected spike events were aligned to their minima and 45 samples (1.5 milliseconds) around the peak were extracted for spike sorting. An automatic clustering procedure identified putative single neurons via a Split and Merge Expectation-Maximisation algorithm which fits a mixture of Gaussians on the spike feature data consisting of the first three principal components of the spike waveforms. Inspection and manual cluster cutting was carried out in Klusters (Lynn Hazan, Buzsáki lab, Rutgers, Newark NJ). The details of the spike sorting algorithms have been described elsewhere (*85*). The spiking waveforms, recorded under a given channel, which could not be sorted to a given single unit were collected and denoted as a multi-unit. For the analysis presented in this study, we combined individual single units and multi-units recorded and they are referred to as units.

Selectivity of unit activity

Each BR trial was visually inspected with the aid of a custom written GUI in MATLAB and the onset and end of a perceptual dominance (during the rivalry phase) was manually marked using the onset of a change in the slow phase of the OKN as a criterion. Two authors VK and AD marked the datasets.

Selectivity of a given unit was assessed separately for PA and BR trials by comparing the spike counts elicited during the presentation (PA) or perception (BR) of downward vs. upward drifting stimuli, using a Wilcoxon rank sum test. For unit selectivity during BR trials, spiking response was aligned to the onset of two events, invoking a perceptual change, namely the (i) onset of flash suppression phase and (ii) onset of a perceptual dominance during spontaneous switches in rivalry. Unit selectivity was similarly assessed during analogous

temporal phases of PA trials. The presentation of the second stimulus during PA is temporally corresponding with the presentation of the second stimulus during the BR trial, which constitutes the flash suppression phase. All subsequent stimulus presentations during a PA trial, can be considered equivalent to the perceptual dominance phases during BR. Therefore selectivity of the spiking responses during these periods was computed for assessing unit selectivity during PA trials. Further, among the periods described above, we considered only those epochs during PA and BR trials for computing selectivity, which consisted of perceptual dominance (BR) or monocular presentation (PA) of a given stimulus lasting at least 1000 milliseconds. With respect to perceptual switches, we analyzed transitions, which consisted of at least 1000 milliseconds of clear dominance (judged by a stable OKN pattern), before and after an OKN switch. To compare with PA as close as possible, we analyzed only those transitions with an interval up to 250 milliseconds between the end of the preceding dominance, and the onset of the next. Data was aligned to the onset of the forward dominance. Corresponding temporal phases of stimulus switches during PA trials, included at least one second monocular presentation of a given stimulus followed by the presentation of an oppositely drifting stimulus in the contralateral eye (compared to the preceding visual presentation) for a minimum duration of 1000 milliseconds. Selectivity was assessed both before (-1000 to 0) and after (0 to 1000) the perceptual (BR) and stimulus switches (PA) by collecting all spikes elicited in a 1000 millisecond period. Any relevant figures presented in the main body of the paper were obtained by analyzing PA trials which were aligned to the TTL pulse signaling a stimulus change. In addition, we visually inspected and marked the onset and offset of the visual stimulus during PA trials similarly to the way these episodes were marked for BR trials, based upon the change in the OKN direction. The selectivity analysis (Figure 3) were repeated with PA trials aligned according to this new criterion and the results are presented in Supplementary Figure 3.3, 3.4, 3.5.

D-prime calculation

For every unit, we computed a preference index denoted as d', by quantifying the strength of its selectivity during PA and BR trials. It was calculated as follows:

$$\mathbf{d' = (\mu_p - \mu_{np})/sqrt((var_p + var_{np})/2)}$$

where, $\boldsymbol{\mu_p}$ and $\boldsymbol{\mu_{np}}$ is the average spiking response of a given unit during the presentation of its preferred and non-preferred stimulus, calculated over a duration of 1000 milliseconds after a physical stimulus change during PA or a perceptual change during BR trials. The difference between these two quantities is normalized by the square root (**sqrt**) of the average pooled variance ($\mathbf{var_p}$ and $\mathbf{var_{np}}$) of the response distributions.

Conventional population PSTHs and ensemble PSTHs

Population PSTHs (Figure 3) were computed by averaging the average neural activity of selective units in response to their preferred and non preferred stimuli. The activity of each unit was calculated as the mean response of the unit during specific temporal phases (flash suppression, perceptual dominance and switches) in 50 ms bins during PA or BR trials. For the flash suppression and perceptual dominance phases, we identified all units which displayed significant modulation either during PA or BR trials. With respect to switches, all units which displayed significant modulation (and maintained stimulus preference) before and after a switch during both trial types were identified. In all three cases, the population PSTH was computed by averaging the activity of all units which displayed preference to the same motion direction across PA and BR. In addition, population PSTHs with units significantly selective in the PA or BR conditions were also computed (Supplementary Figures 3.1, 3.4 and 3.2, 3.5, respectively). Population activity related to switches included units which were significantly modulated both before and after the switch for the same motion direction in PA (Supplementary

Figure 3.1, 3.4) or BR (Supplementary Figure 3.2, 3.5). Additionally, we generated average ensemble population PSTHs. Here we refer to a population of units displaying preference for the same stimulus as a neuronal ensemble. Population of units which contributed to ensemble PSTHs were identified similarly to conventional population PSTHs. Therefore, the population of units contributing to Figure 3B,C (Switches) and Figure 4A is identical. However, PSTHs were computed differently. First, the activity elicited by all units preferring the downward and upward motion directions were separately averaged for each transition in 50 ms bins, providing a population vector of each neural ensemble for every switch. Next, each of these traces was normalized by subtracting the minimum and dividing it by the maximum activity. Finally, traces were collected across all transitions across datasets and were averaged to generate the average ensemble population PSTHs, presented in Figure 4A. Such an ensemble population PSTH complements the decoding approach, which utilizes the population response on single trials aimed at ascertaining the ongoing sensory input (PA) or perceptual experience (BR). Ensemble population PSTHs for the control paradigms presented in Figure 5A were generated similarly as described above, but without the normalization step. For the ensemble activity related to presentation of the preferred stimulus, all trials where the preferred stimulus of the units comprising the two different ensembles (upward and downward motion) were presented were pooled together and an average was computed. Similarly, all trials where ensemble's non preferred stimulus was presented were pooled together and averaged for ensemble activity pertaining to the non-preferred stimulus.

Decoding Analysis

Multivariate pattern analysis was utilized to assess if the spiking activity of neuronal ensembles in the prefrontal cortex contained information about the stimulus on a single transition basis. In this regard, we used a maximum correlation coefficient classifier (*52*) implemented as a part

of the neural decoding toolbox (*86*). All the recorded units (N = 987) across the two monkeys were pooled as a pseudopopulation for the decoding analysis pertaining to the BR paradigm (Figure 4). This is similar to previous studies (*52*, *87*) where units recorded during independent sessions are treated as simultaneously recorded. Firing rates of each of these units during 15 randomly selected stimulus (PA trials) and perceptual switches (BR trials) were utilized. A z-score normalization (subtracting the mean activity and dividing by the standard deviation) of each unit's response was done before it participated in the classification procedure in order to assure that units with high spike rates do not influence the decoding procedure disproportionately. We used 15 cross-validation splits, implying that for 14 switches used for training the pattern classifier, one was leaved out to put in the test. This procedure was repeated 50 times (resample runs) to estimate the classification accuracy with a different randomly chosen cross-validation split in each run. All decoding accuracy estimates are zero-one-loss results. Each pixel in cross temporal generalization plots (Figure 4B, 5B and Supplementary Figures 4.1B, 4.2B, 5.4B) depicts the classification accuracy computed with firing rates in 150 ms bins, sampled every 50 ms. This bin duration was chosen, because it has been previously used successfully for decoding visual input from neural activity in the frontal and temporal cortex (*52*, *87*). Similar steps as described above were used for decoding analysis in the control paradigms (Figure 5 and Supplementary figure 5.4), with one difference. Units which were significantly modulated in either of the two tasks and preferred the same stimulus (N = 104) participated in the decoding procedure.

Statistical significance of the classification accuracy was estimated using a permutation test, which involved running the decoding analysis on the data with labels shuffled (*51*, *86*). This procedure was repeated 500 times with parameters related to binning, cross validation splits as well as resample runs identical to those used for standard decoding of correctly labeled data. The resultant classification accuracies obtained served as a null distribution. If the decoding

results obtained without shuffling the labels were greater than all values within the null distribution, they were considered as significant ($p<1/500 = 0.002$). Significance of decoding accuracy was computed using this procedure for the results presented in figure 4C (also supplementary figure 4.1 and 4.2) and 5C (also supplementary figure 5.4).

Selection of trials with reduced eye position variance

To create a robust dataset for decoding the visual stimulus during passive observation of monocular stimuli, we only picked those trials corresponding to upward and downward moving gratings where the OKN during passive fixation was relatively flat, i.e. there were no strong drifts in the direction of motion during suppression of eye movements. Firstly, the Y coordinate of the eye movement signal was detrended to remove any DC offset. It was then filtered below 20Hz to remove high-frequency noise and blinks. Next, the double differential was computed and compared to a flat line (i.e. with a slope of 0 and an intercept corresponding to the baseline of the OKN signal) using a least-squared-error minimization method. The sum of the squared error for each trial was collected across all sessions. All those trials whose sum of least squared error was less than the median of the distribution of these errors were selected for further analysis. This method resulted in a selection of trials with reduced variance of the eye movements signal (Supplementary Figure 5.3).

TABLES

Table 1

Number (N) and proportion (%) of significantly modulated units during the BR paradigm

Trial Type	Physical Alternation		Binocular Rivalry	
Temporal Phase	**Physical Alternation**	**Sensory Dominance**	**BFS Dominance**	**BR dominance**
N of significantly modulated units	**342**	**364**	**264**	**247**
N and % of significantly modulated units* in PA, with similar average stimulus preference in BR	**292/342 85.38%**	**277/364 76.09%**		
N and % of significantly modulated units* in BR, with similar stimulus preference in PA			**229/264 86.74%**	**199/247 80.56%**
N of significantly modulated units in PA with a d'>1	**80**	**46**		
N and % of significantly modulated units* in PA (d'>1), with similar preference and significantly modulated in BR (ALL)	**72/80 90%**	**40/46 86.96%**		

*significantly modulated here in a particular condition refers to units displaying significantly stronger activity to one of the two stimuli using a Wilcoxon rank sum test at an alpha value of 0.05.

Table 2

Number and proportion of significantly modulated units during the BR paradigm

(Physical alternation trials aligned to the change in the OKN direction)

Trial Type	Physical Alternation	Binocular Rivalry
Temporal Phase	Sensory Dominance	BR dominance
N of significantly modulated units	408	247
N and % of significantly modulated units* in PA, with similar average stimulus preference in BR	324/408 79.41%	
N and % of significantly modulated units*in BR, with similar stimulus preference in PA		199/247 80.56%
N of significantly modulated units in PA with a d'>1	68	
N and % of significantly modulated units* in PA (d'>1), with similar preference and significantly modulated in BR (ALL)	49/68 72.05%	

FIGURES

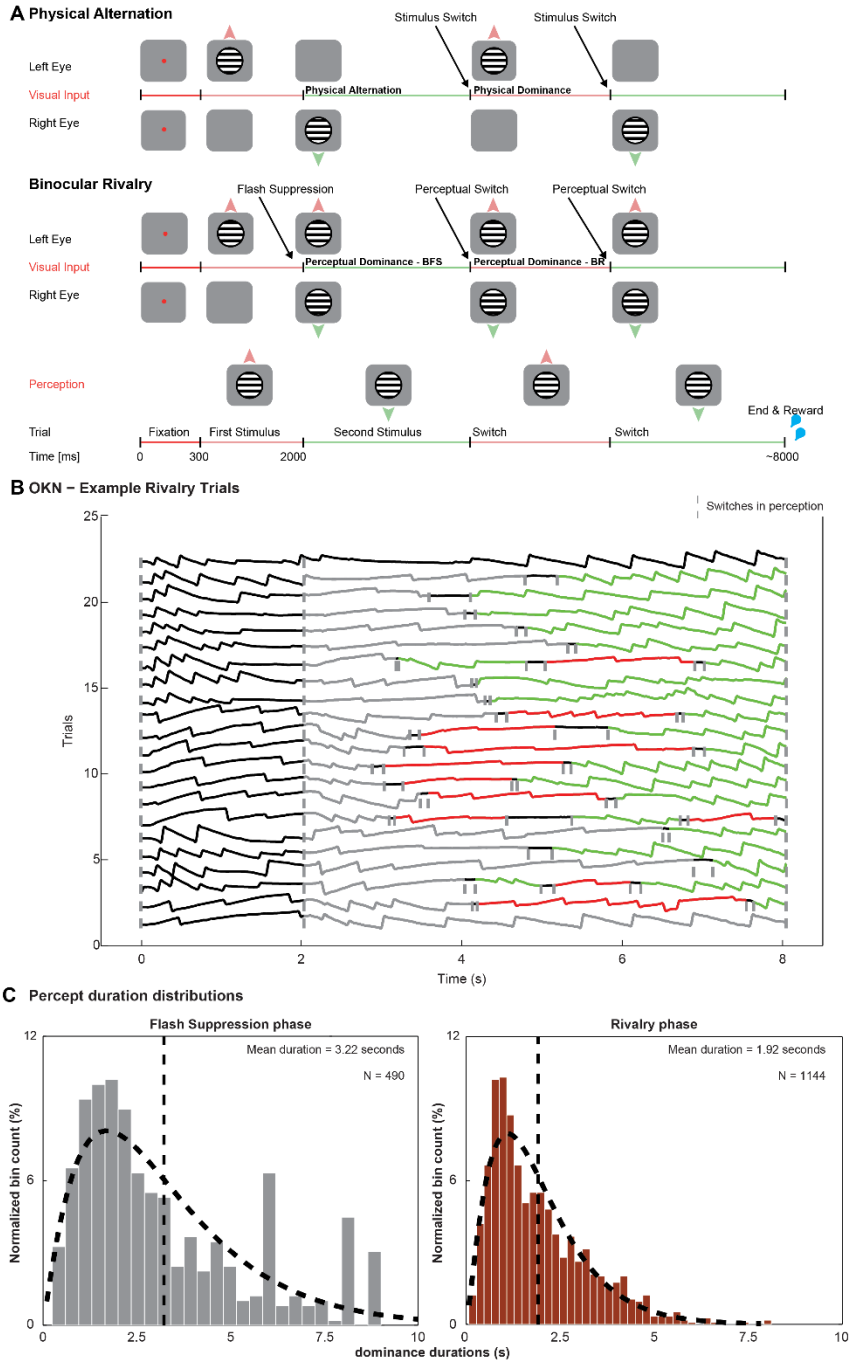


FIGURE 1

Binocular rivalry paradigm and behavior

(A). The task consisted of two trial types, namely, the physical alternation (PA) trials and binocular rivalry (BR) trials. Both trial types started with the presentation of a fixation spot,

cueing the animal to initiate fixation. Upon successful fixation for 300 milliseconds, a drifting sinusoidal grating was monocularly presented. After 1 or 2 seconds, the first stimulus was removed and a second grating drifting in the opposite direction was presented in the contralateral eye during PA trials. During BR trials, the second stimulus was added to the contralateral eye without removing the first stimulus, inducing perceptual suppression of the first stimulus (Flash Suppression). After this period, visual input alternated between upward and downward moving gratings during PA trials (Stimulus Switch). During BR trials, the percept of the animal could randomly switch between the discordant visual stimuli (Perceptual Switch). Note that perceived direction displayed in the bottom row schematic is identical, even though the underlying visual input is monocular in PA and dichoptic during BR. (B) OKN elicited during example BR trials from one recording session. The gray vertical dashed line denotes the beginning of the flash suppression phase. Subsequent dominance phases are color coded and their beginning and end are marked with shorter grey dashed lines. Note that on the last example trial the flash suppression resulted in a prolonged continuous suppression of the previously presented direction of motion, while on the first trial, flash suppression was not effective and the initially presented direction of motion remained dominant. (C) Perceptual dominance distributions during flash suppression and rivalry phases could be approximated well with a gamma distribution.

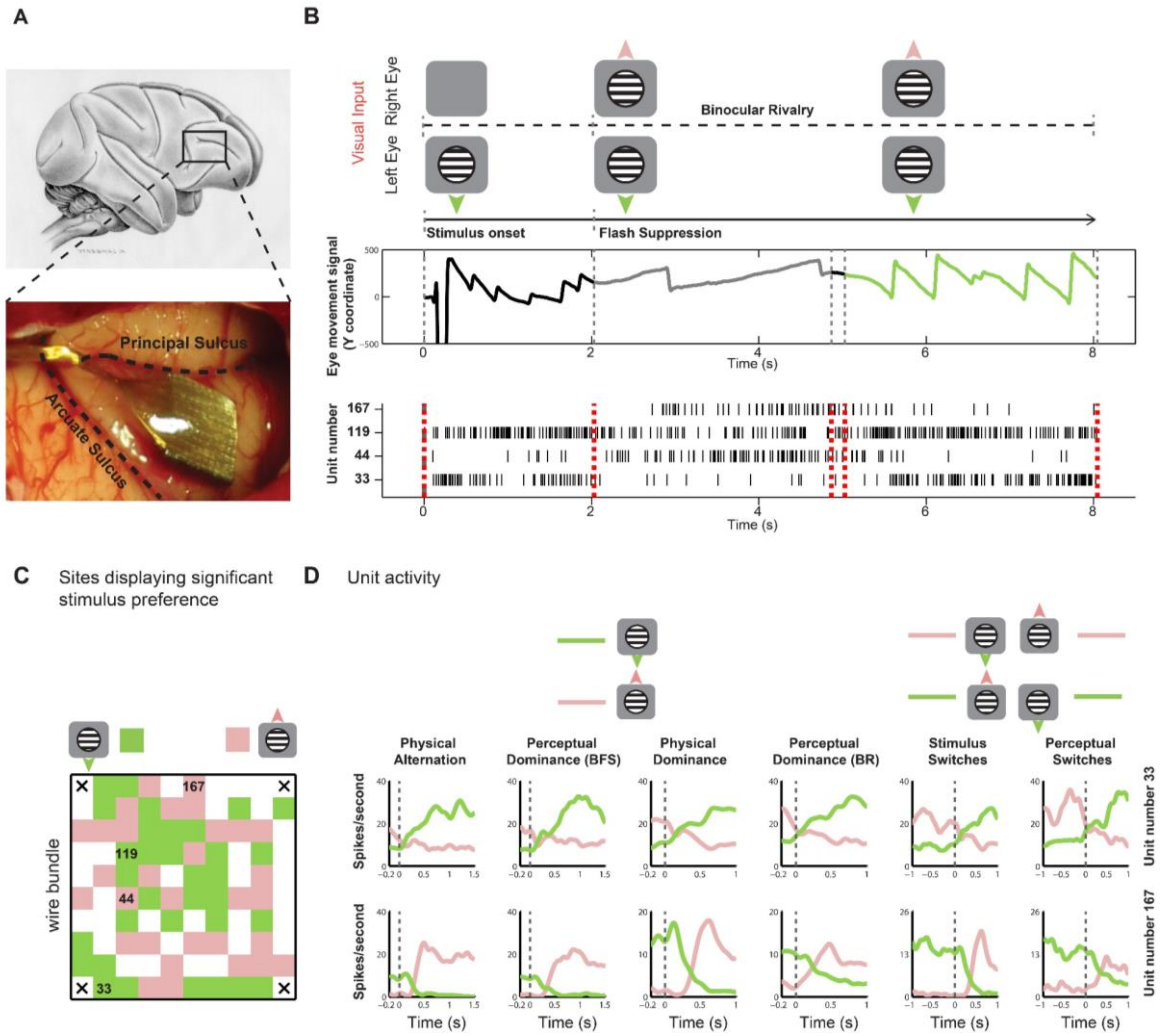


FIGURE 2

Example unit responses

(A) The location of the implanted Utah array in the inferior convexity of the PFC on a schematic macaque brain and in one of the animals. (B) Visual input, OKN and corresponding spiking activity during an example BR trial. The trial started with monocular presentation of downward drifting grating. An upward drifting grating was added to the contralateral eye 2000 ms later, which resulted in perceptual suppression (flash suppression) of downward motion, inferred from a change in the OKN direction. Externally induced perceptual suppression lasted for ~3000 ms after which a spontaneous switch reinstated the perception of the downward motion.

In the spike raster plot, while unit numbers 33 and 119 display strong spiking activity when downward drifting grating is perceptually dominant, unit numbers 44 and 167 respond stronger when upward drifting grating is perceived. (C) Projection of all sites with significant stimulus preference during the flash suppression phase of the PA trials on the array for one recording session. The location of the units presented in (B) are marked. Green and pink pixels reflect sites, where spiking activity (unsorted spikes recorded from a given electrode), responded more to upward or downward drifting gratings respectively. (D) Average spike density functions of two units recorded (unit 33, selective for downward motion in the upper row, unit 167, selective for upward motion in the lower row) simultaneously in the PFC during PA and BR trials. Pink and green colors in the first four columns correspond to the response elicited by presentation or perception of downward and upward drifting grating respectively. In the last two columns, we plot the activity elicited during a stimulus or perceptual switch from downward to an upward drifting grating (pink) and vice versa (green). The activity of both units is very similar during PA and BR, thus displaying robust perceptual modulation.

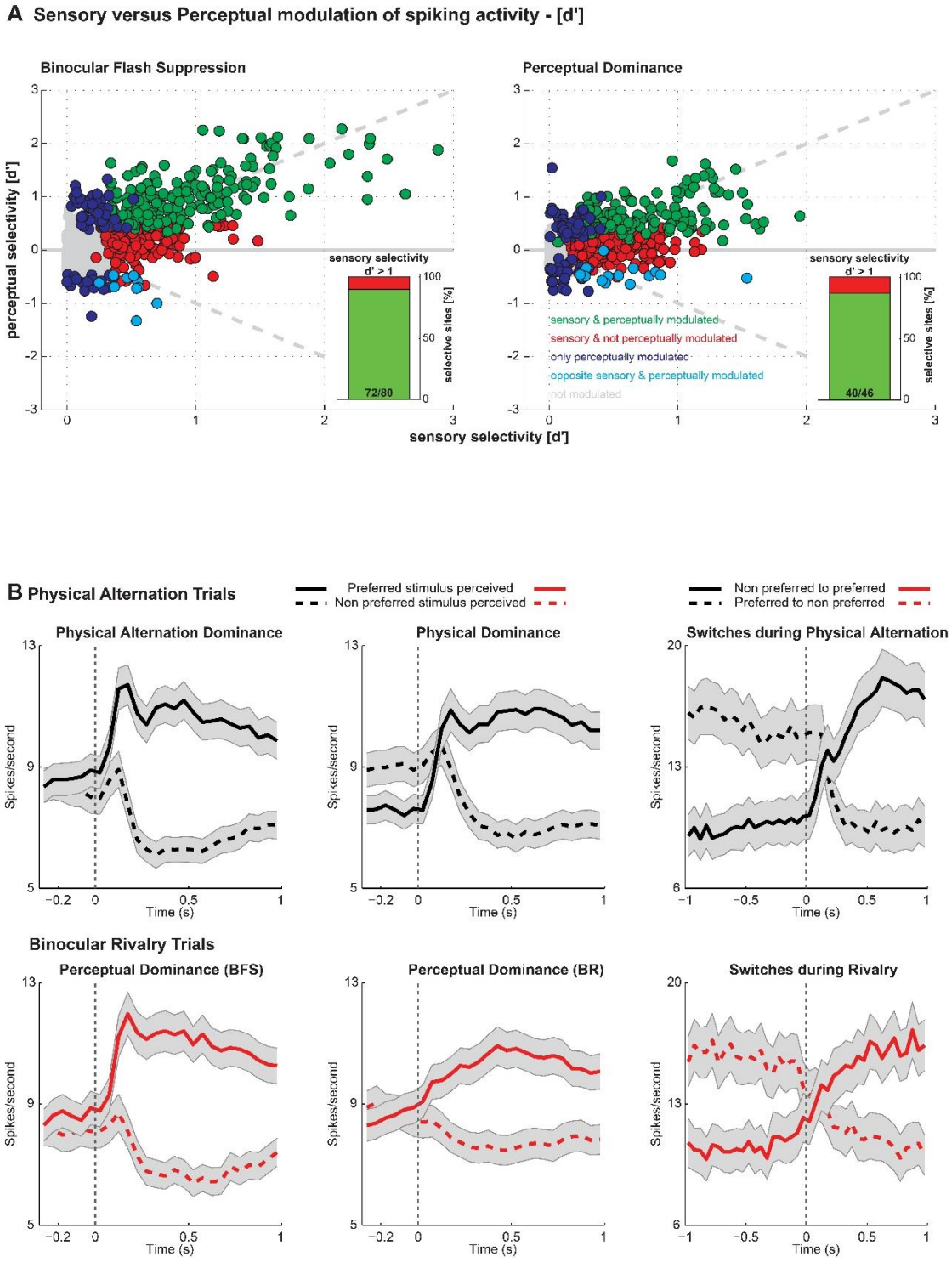


FIGURE 3

Sensory (PA) versus perceptual (BR) modulation of spiking activity. (A) Scatter plot of sensory vs. perceptual selectivity (d′) for all units (dots) across all datasets for BFS and BR. Units showing no significant modulation in PA or BR trials are displayed in grey, those with significant modulation during both conditions in green, units which display significant preference only during PA trials in red and units displaying significant modulation only during BR trials are displayed in blue. In cyan the small percentage of units which fired more when their preferred stimulus was perceptually suppressed across the two conditions. The proportion of perceptually modulated units for both BFS (90%) and BR (86%) increased as a function of sensory selectivity strength (insets showing perceptual modulation for d'>1). (B). Average population spiking activity in PA and BR. The population activity averaged across all units which were significantly modulated during PA (upper row) or BR (lower row) trials and preferred the same stimulus is plotted for the flash suppression (left), the perceptual dominance (middle) phase and switches (right) during BR and temporally matched phases in PA. Displayed are two traces of population activity, one, calculated when the unit's preferred stimulus was dominant (thick lines) and the second, when the unit's preferred stimulus was suppressed because of the dominance of its non-preferred stimulus (dashed lines). Population activity reliably followed phenomenal perception during perceptual transitions brought about exogenously with flash suppression as well as endogenously driven during BR. A remarkable similarity in population activity across the two trial types indicates strong and robust perceptual modulation.

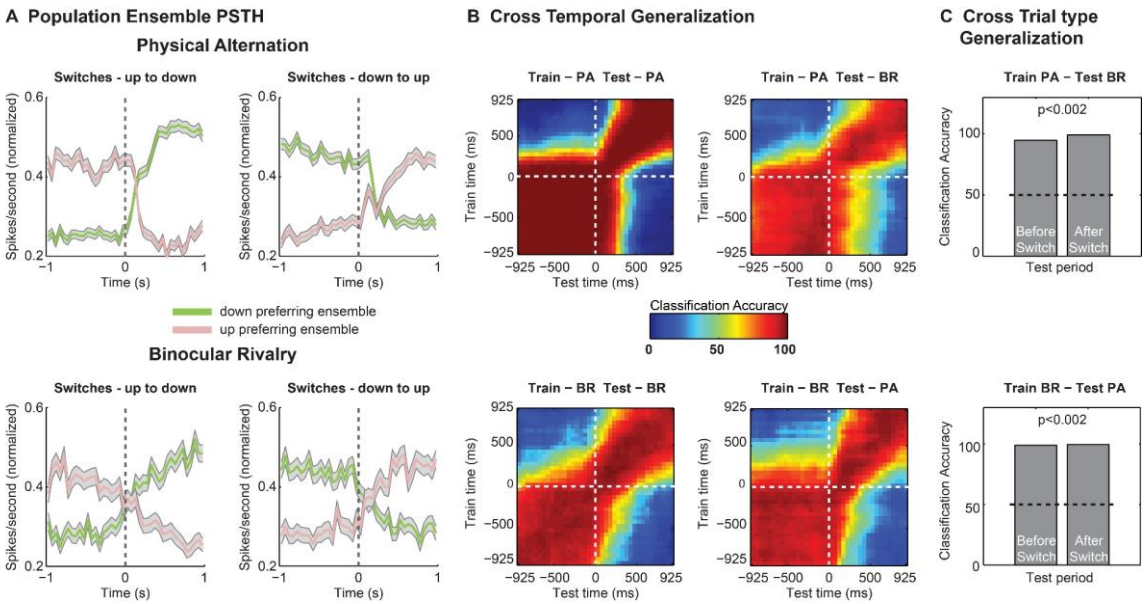


FIGURE 4

Decoding the contents of conscious perception from simultaneously recorded prefrontal ensembles. (A) Normalized spiking activity of down (green) and up (pink) preferring ensembles of units during up to down or down to up, PA (upper row) and BR (lower row) switches showing reliable modulation of neuronal ensembles during both external stimulus changes and internally generated switches in conscious perception. (B) Cross-temporal decoding of stimulus contents around switches in perception during PA and BR trials and generalization across the two. Classification accuracy was computed for each pair of train and test time windows around a switch (see methods) in steps of 50 ms, using 150 ms bins. (C) Cross trial type generalization was highly significant (permutation test, $p < 0.002$), suggesting that the underlying population code is invariant to the trial type, and therefore encodes perceptual contents.

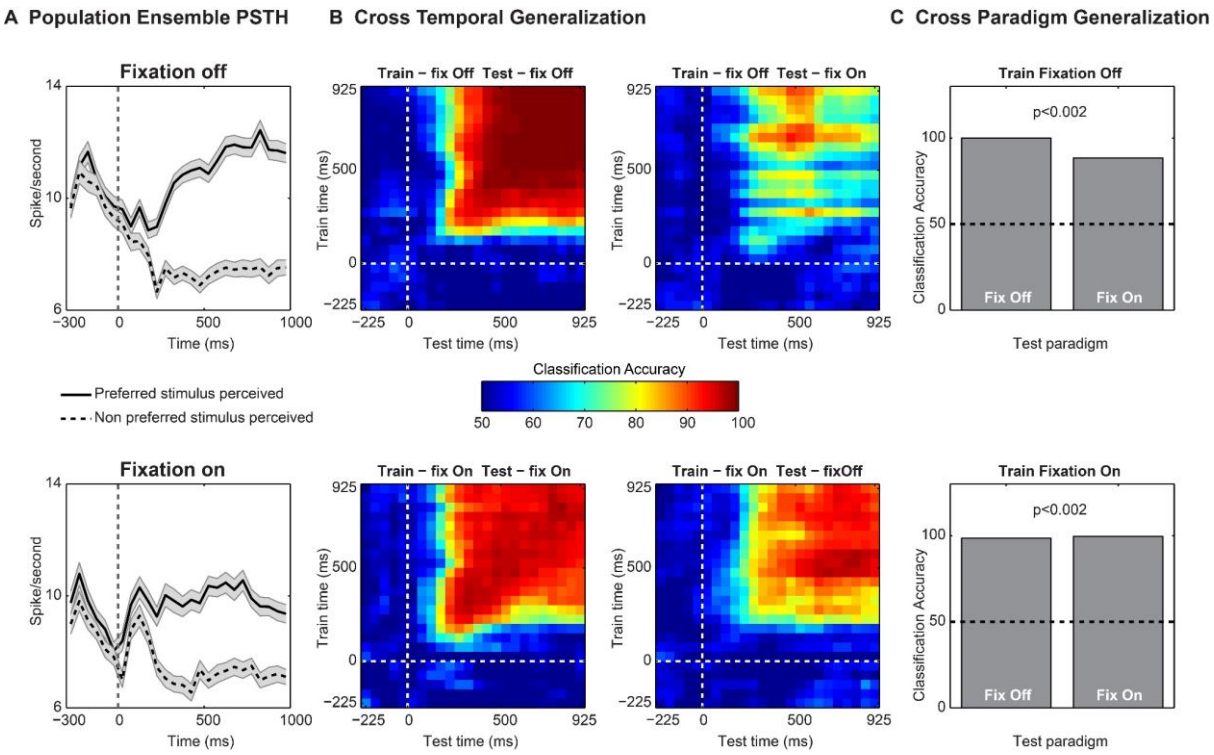
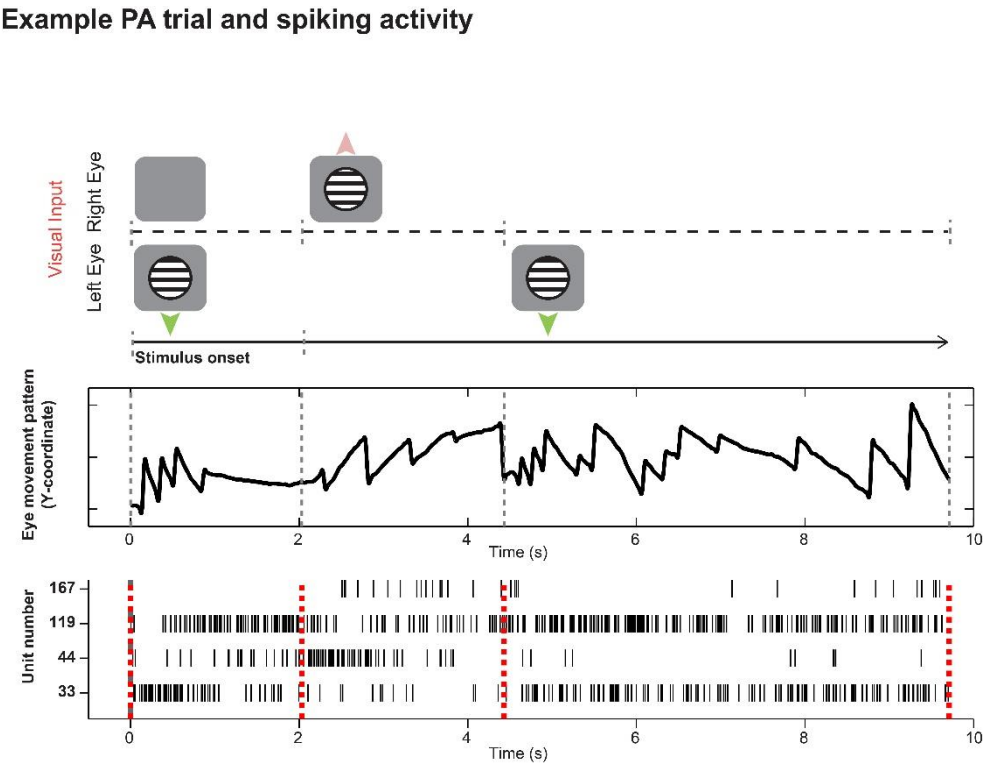


FIGURE 5

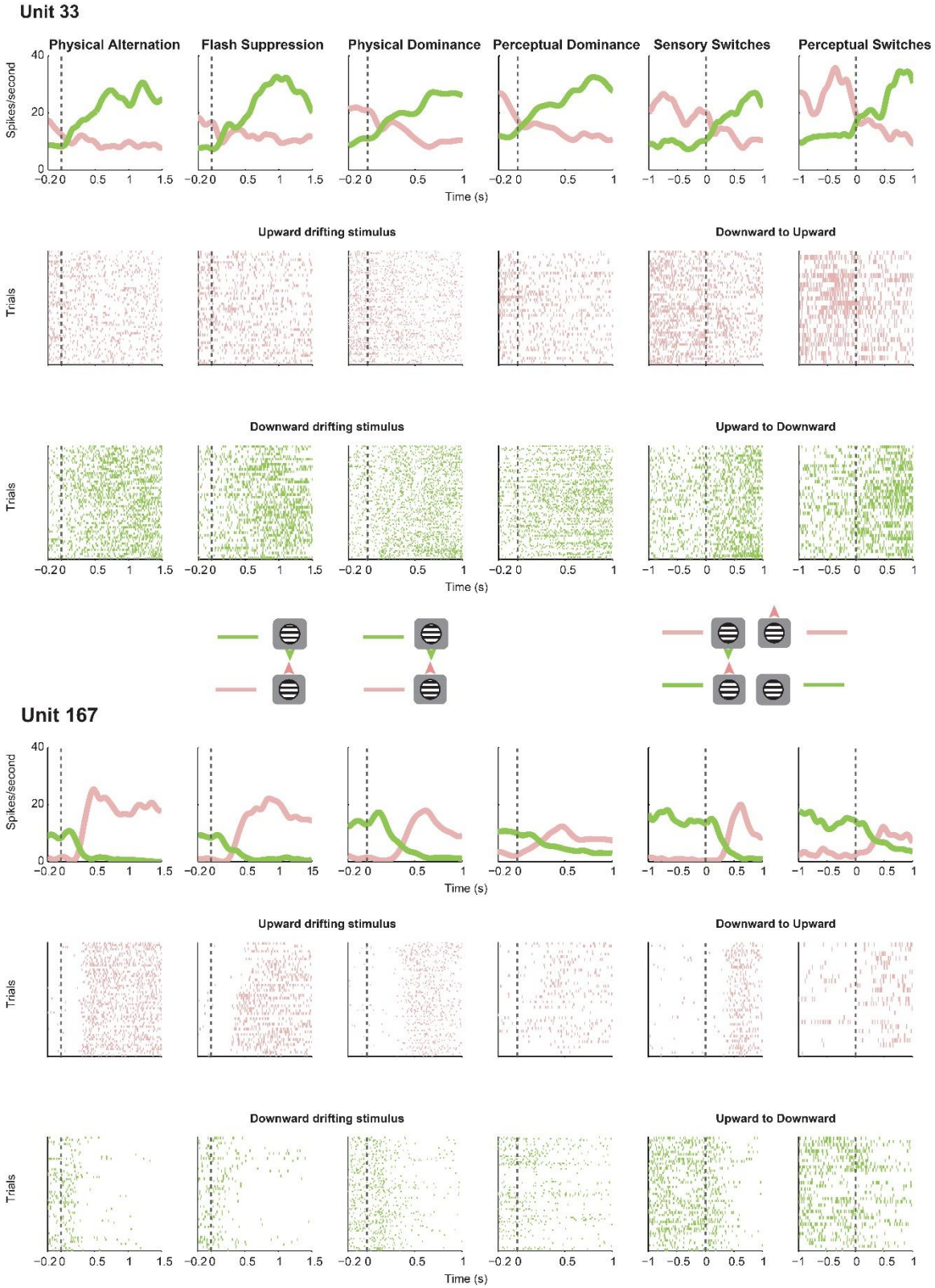
Invariance of the population code to motion content in the presence and during the suppression of OKN eye movements assessed with multivariate pattern analysis. (A) Ensemble population spiking activity (see methods) during the fixation Off and fixation On paradigm for units which were significantly modulated in either of the two paradigms, and preferred the same motion direction (B) Cross-temporal decoding of stimulus contents during the two paradigms. Decoding accuracy was tested for each pair of train and test time windows during the two paradigms as well as across them, with binning parameters similar to figure 4 (C) The cross paradigm invariance of the population code was tested by training a classifier on activity in one paradigm and testing on the other, for a single bin of 400 ms (starting 400 ms post stimulus onset) during the presentation of the visual stimulus. We observed significant (permutation test, $p < 0.05$) cross-task generalization accuracy, thus suggesting that the underlying code is largely invariant to the presence of large OKN, and encodes stimulus motion contents.

SUPPLEMENTARY FIGURES



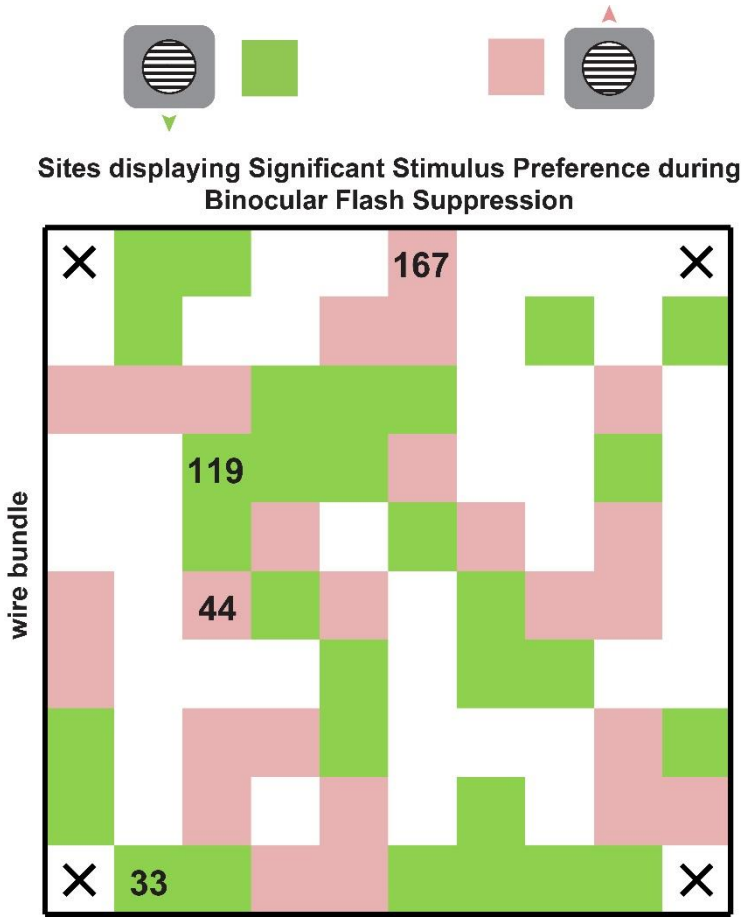
Supplementary Figure 2.1

Y-coordinate of the eye movement signal displaying the OKN, and concomitantly recorded spiking activity during an example PA trial for the same units presented in Figure 2B. The trial started with monocular presentation of downward drifting grating. 2000 ms later, the first grating was removed and simultaneously, an upward drifting grating was presented to the contralateral eye. A stimulus switch was externally induced at ~4500 ms, and a change in the OKN polarity was observed right after. While unit number 33 and 119 display strong spiking activity when downward drifting grating is presented, unit numbers 44 and 167 respond strongly to the presentation of upward drifting grating, thus modulated in a similar way as for perceptual switches during BR in Figure 2B.



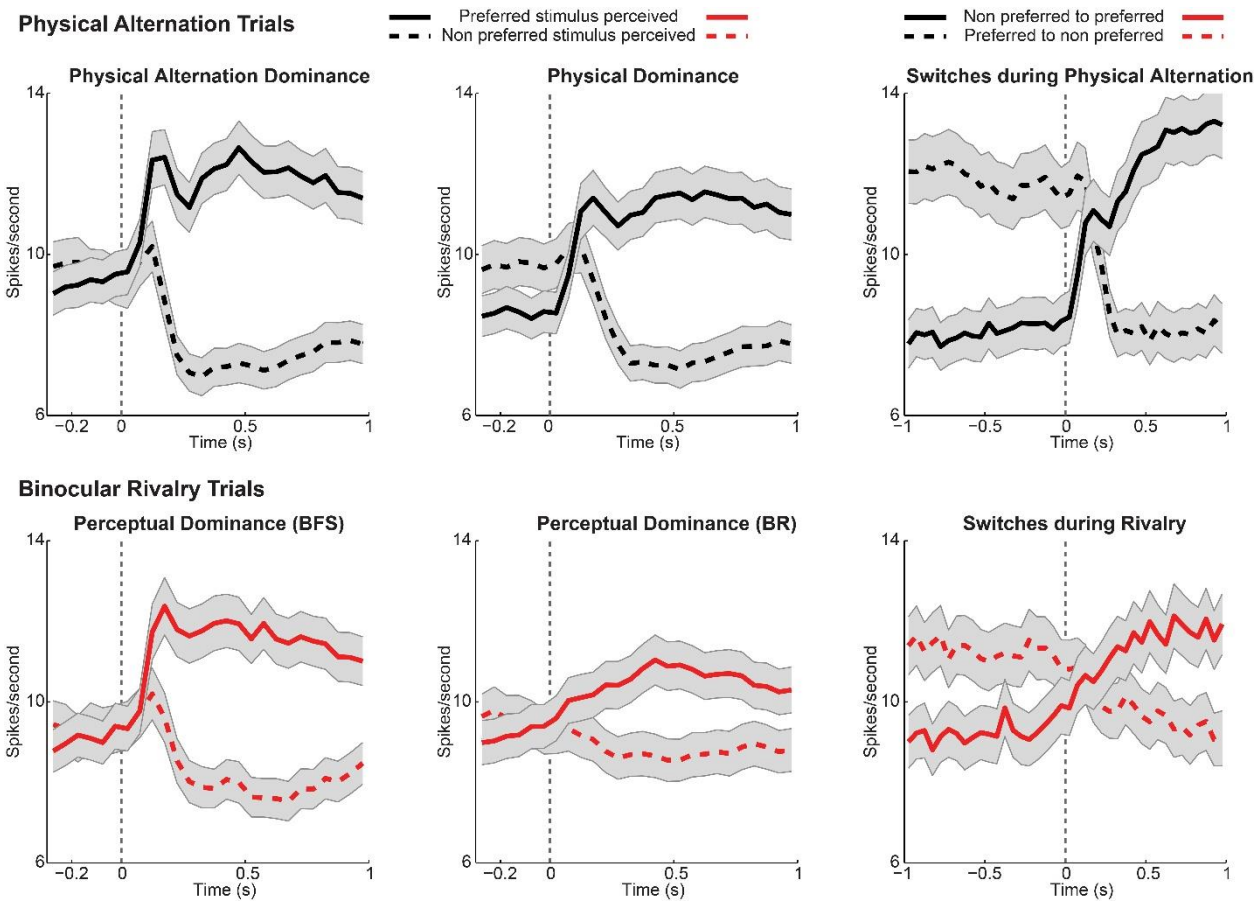
Supplementary Figure 2.2

Spike density functions and raster plots, for the units displayed in Figure 2D. Unit 33 displayed stronger activity to grating drifting down, while Unit 167 fired more, when a grating drifting up was presented in PA or perceived in BR. Stronger activity of units is evident also in the spike rasters. With respect to the first four columns of spike rasters: displayed in pink are responses related to grating drifting upwards, while in green is spiking related to the stimulus drifting down. The last two columns display spiking activity as pink rasters for a down to up switch, while in green for an up to down switch.



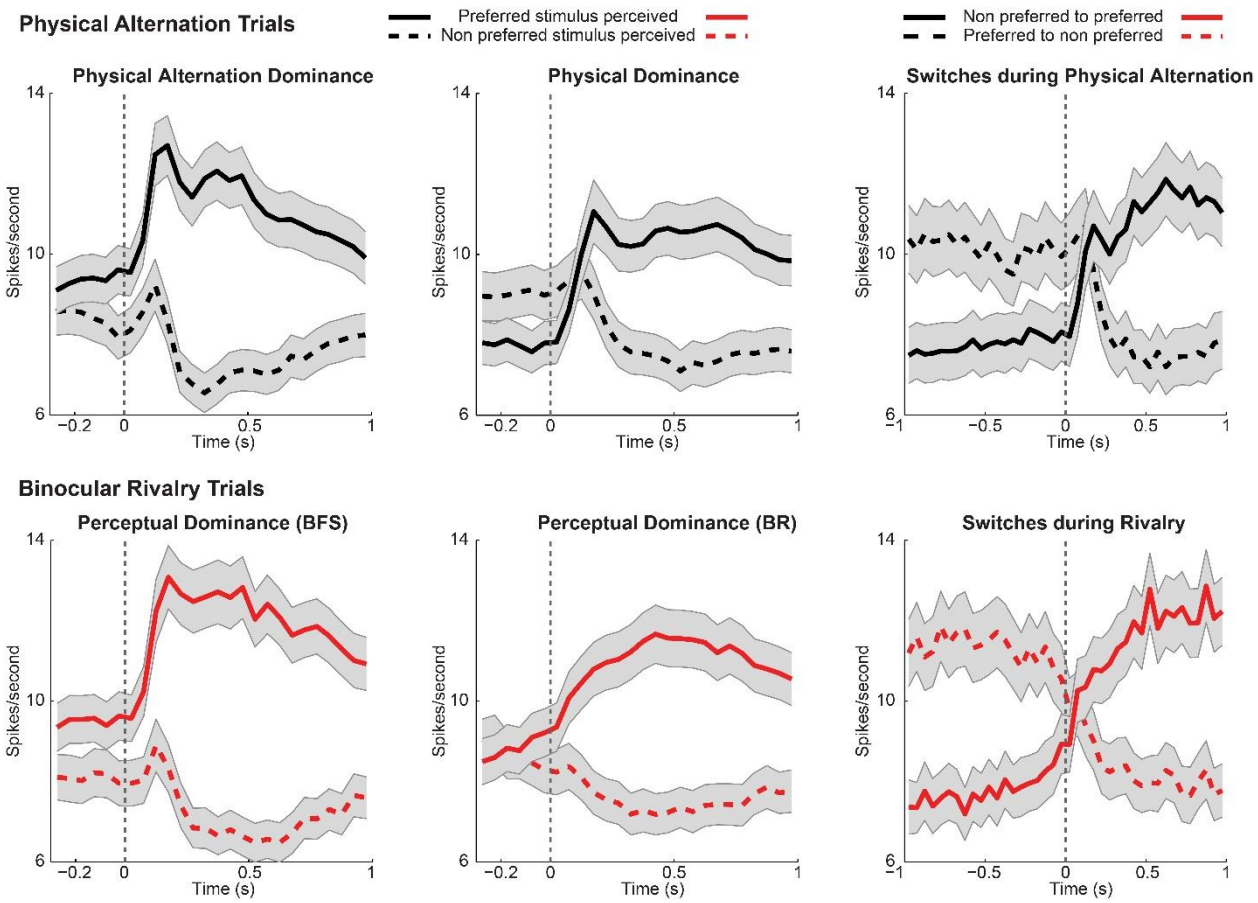
Supplementary Figure 2.3

Sites which displayed significant stimulus preference during the flash suppression phase of the BR trials during one recording session is projected back on the array. The numbers denote the location of units displayed in Figure 2B. Green and pink pixels reflect sites, where the spiking activity (unsorted spiking activity recorded from a given electrode), responded more to upward or downward drifting gratings respectively.



Supplementary Figure 3.1

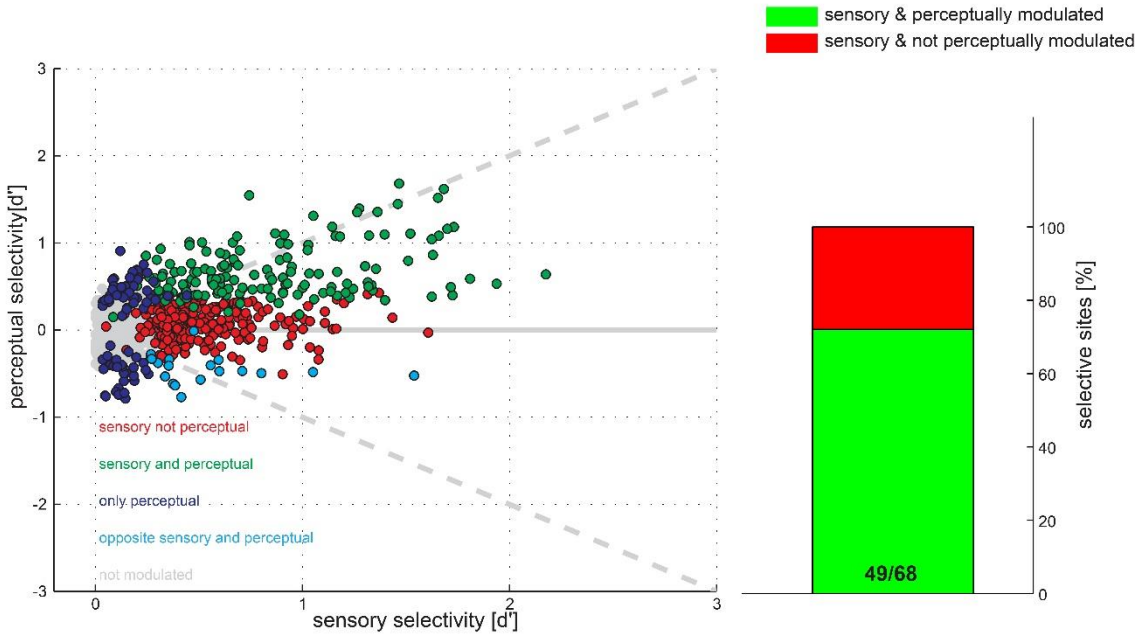
Similar to Figure 3B, the average population spiking activity during PA and BR trials is presented across the various temporal phases of the paradigm (flash suppression, perceptual dominance and switches) for units significantly modulated during PA trials. For switches, selectivity was estimated both before and after the stimulus change. Units significantly modulated more for the same visual stimulus both before and after the stimulus switch were used.



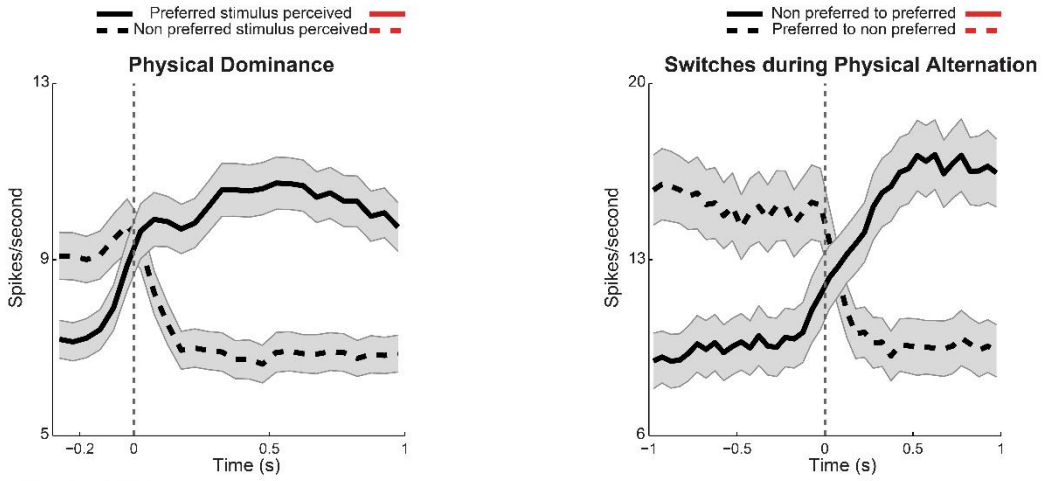
Supplementary Figure 3.2

Similar to Figure 3B, presented here is the average population activity during PA and BR trials across the various temporal phases of the paradigm. The population activity was computed using units which were significantly modulated during BR trials. For switches, selectivity was estimated both before and after the perceptual change. Units which were significantly modulated more for the same perceived motion direction both before and after the perceptual transition were used.

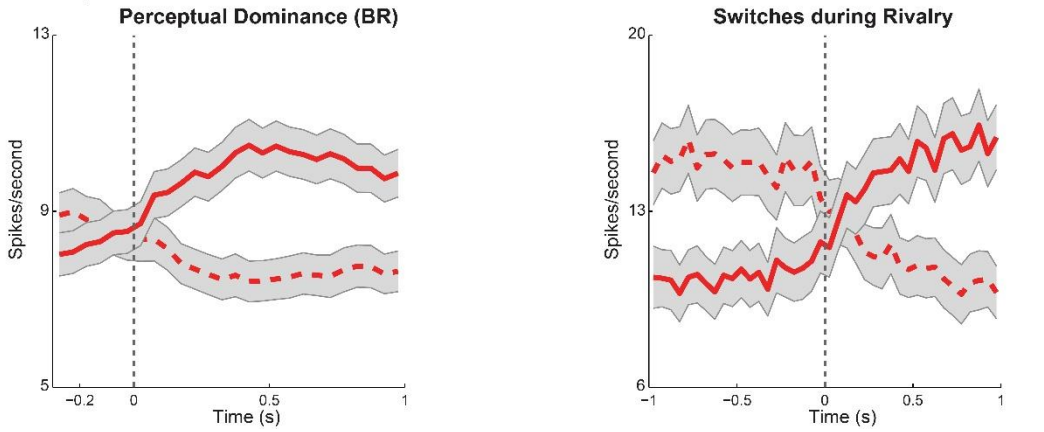
A **Sensory versus Perceptual modulation of spiking activity - [d']**



B **Physical Alternation Trials**

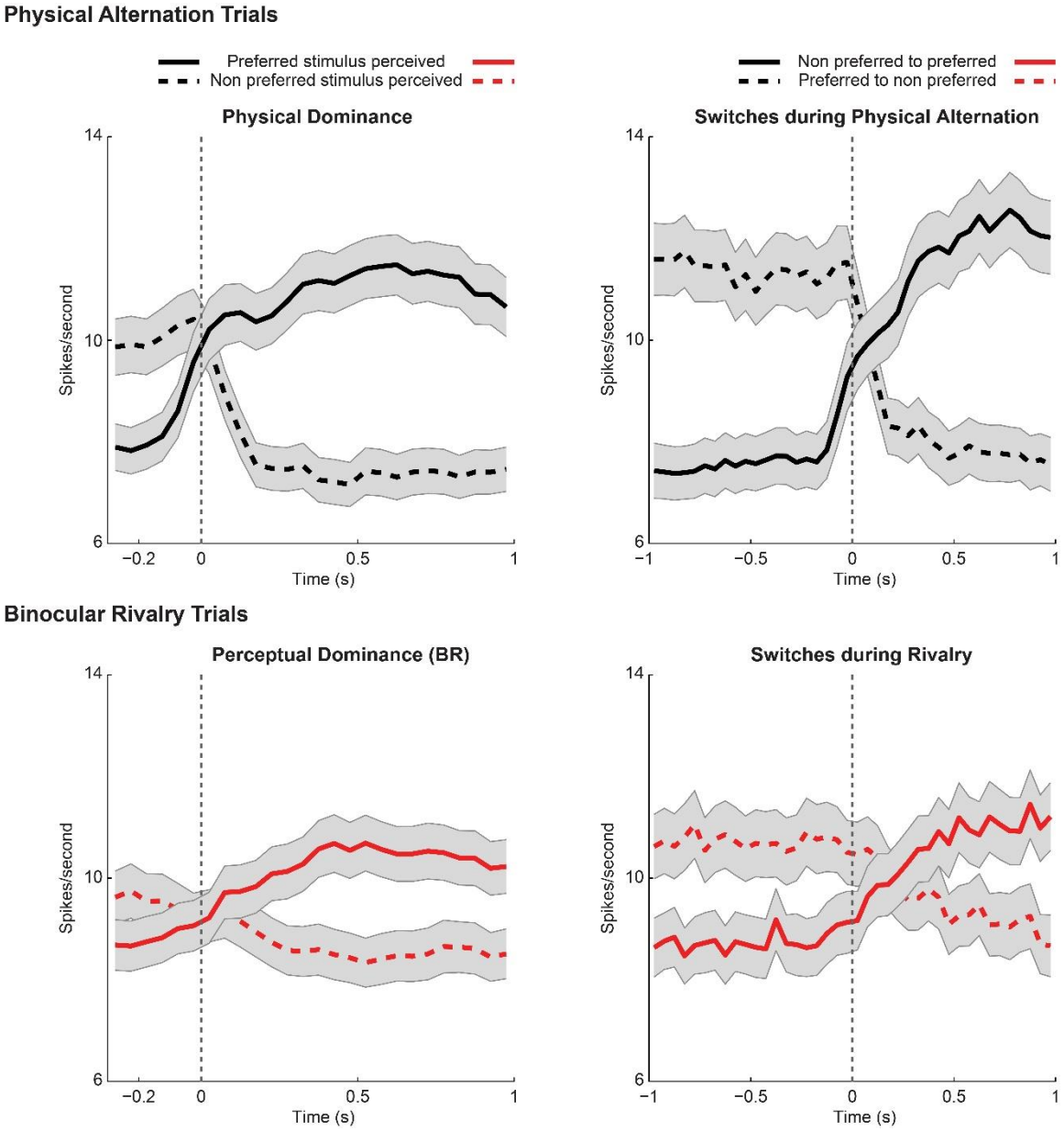


Binocular Rivalry Trials



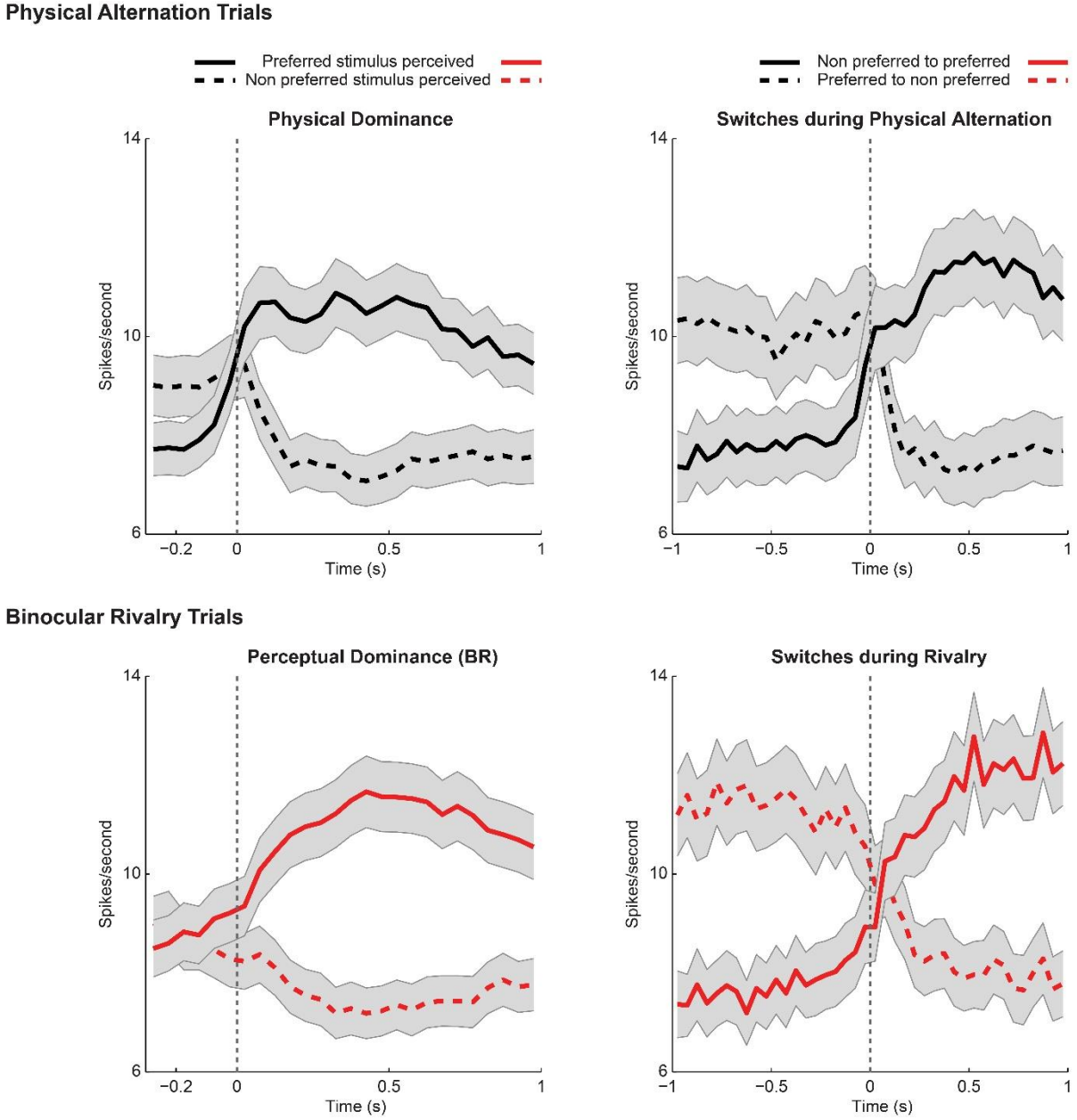
Supplementary Figure 3.3

Similar to Figure 3, these plots display the results obtained when the onset and offset of the visual stimulus in physical alternation trials was aligned to the change in OKN (see methods). (A) Scatter plot of sensory versus perceptual preference (d′) for all recorded units is displayed. Each dot denotes a unit. Units showing no significant modulation in PA or BR trials are displayed in grey while those with significant modulation during both conditions are colored green. In red are units which display significant preference only during PA trials. Units displaying significant modulation only during BR trials are displayed in blue, while in cyan are units which fired more when their preferred stimulus was perceptually suppressed. As evident from the scatter plot, the proportion of PA modulated units which are also significantly modulated during BR increase as a function of the strength of sensory selectivity (d′). The right column displays the proportion of PA modulated units with a d′ greater than 1, which were also significantly modulated during perceptual dominance phase in BR trials (green). The right column plots a similar scatter for perceptual dominance phase of the task. (B) Displayed below is the average population spiking activity during PA and BR trials. Similar to Figure 3, population activity was computed by averaging across all units which were significantly modulated during PA or BR trials and preferred the same stimulus. Although the PA trials which participated in this analysis were aligned to change in OKN, the population activity observed across the two trial types was remarkably similar indicating clear and robust perceptual modulation in the units recorded in the vlPFC. Population activity reliably followed phenomenal perception during perceptual transitions brought about both exogenously with flash suppression as well as endogenously driven during BR.



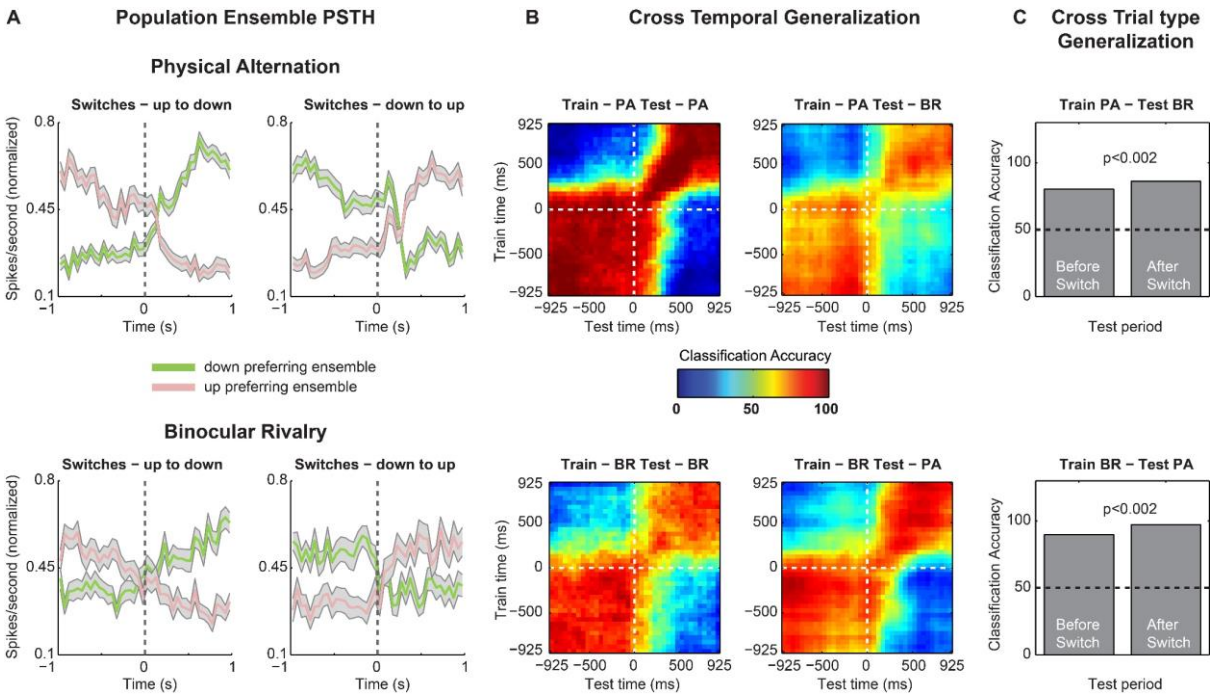
Supplementary Figure 3.4

Similar to Figure 3.1 B, results obtained with PA trials aligned to the change in OKN. Presented across three columns is the average population spiking activity during PA and BR trials. The population activity averaged across all units which were significantly modulated during PA trials is plotted here during three temporal phases, namely, the flash suppression phase, the perceptual dominance phase and switches.



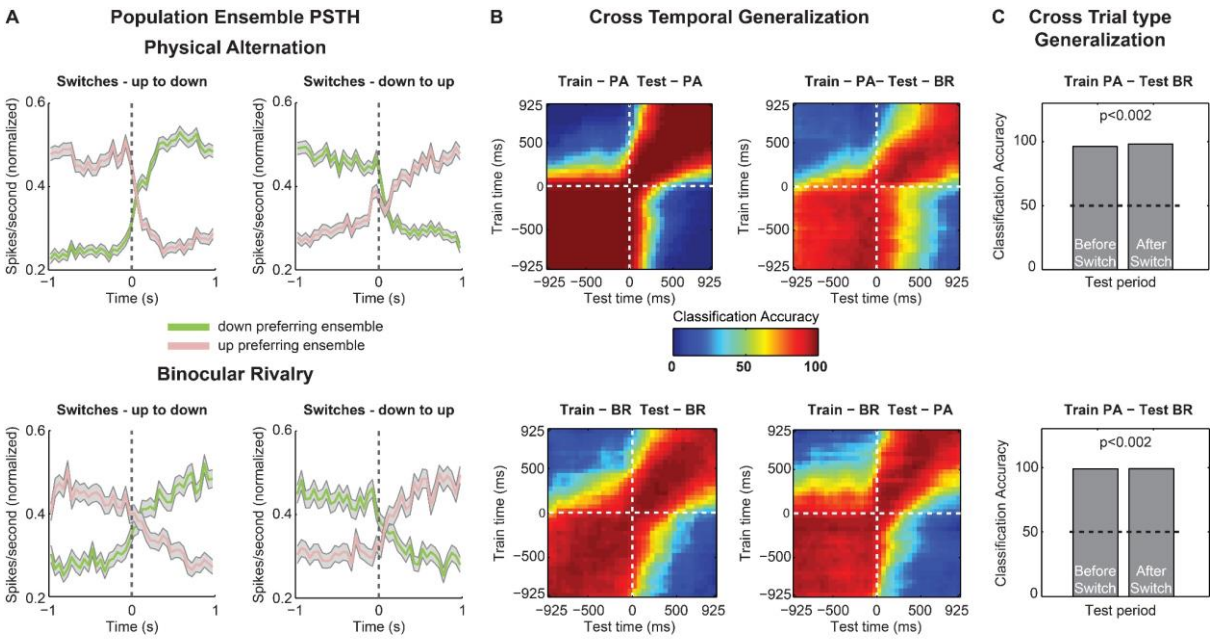
Supplementary Figure 3.5

Similar to Figure 3.2, results obtained with manually marked PA trials. Presented across three columns is the average population spiking activity during PA and BR trials. The population activity averaged across all units which were significantly modulated during BR trials is plotted here during three temporal phases, namely, the flash suppression phase, the perceptual dominance phase and switches.



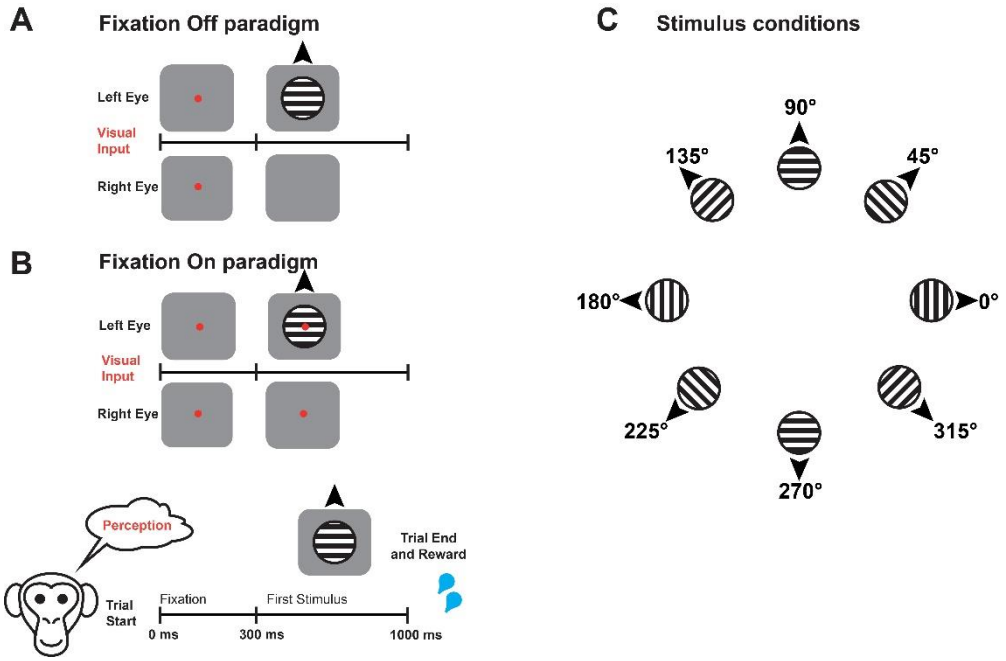
Supplementary Figure 4.1

Similar to Figure 4, the results obtained with the multivariate pattern analysis from an individual dataset are displayed. Both stimulus and perceptual contents could be successfully decoded from simultaneously recorded units in an individual dataset.



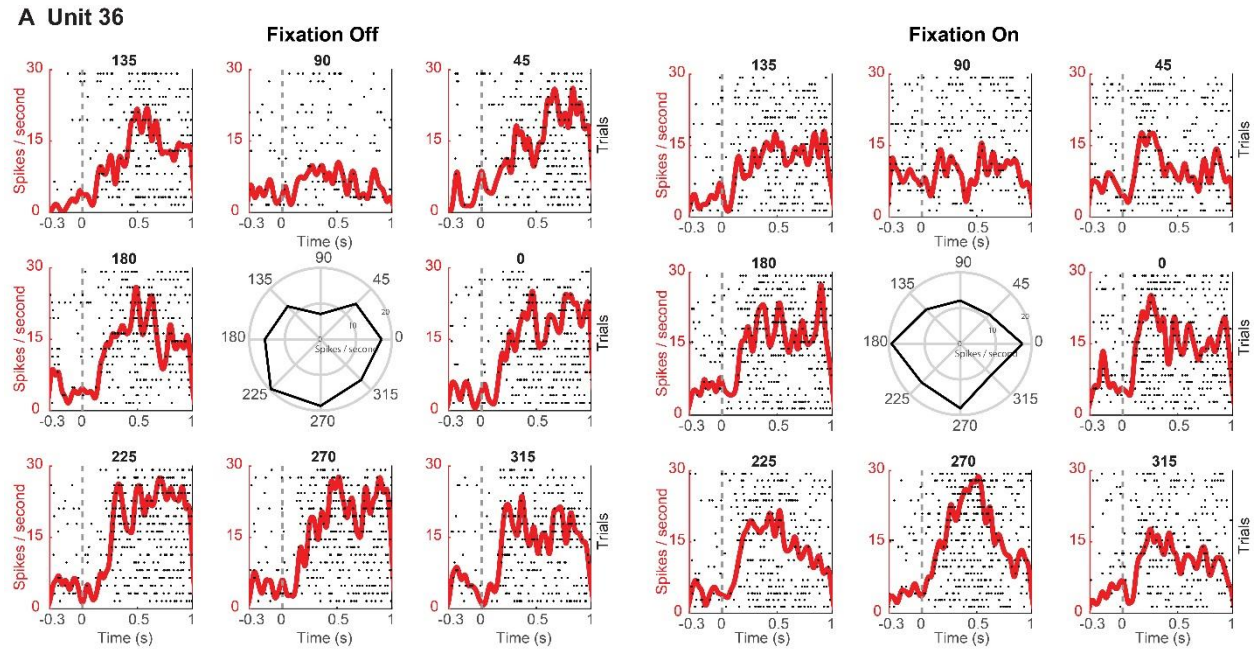
Supplementary Figure 4.2

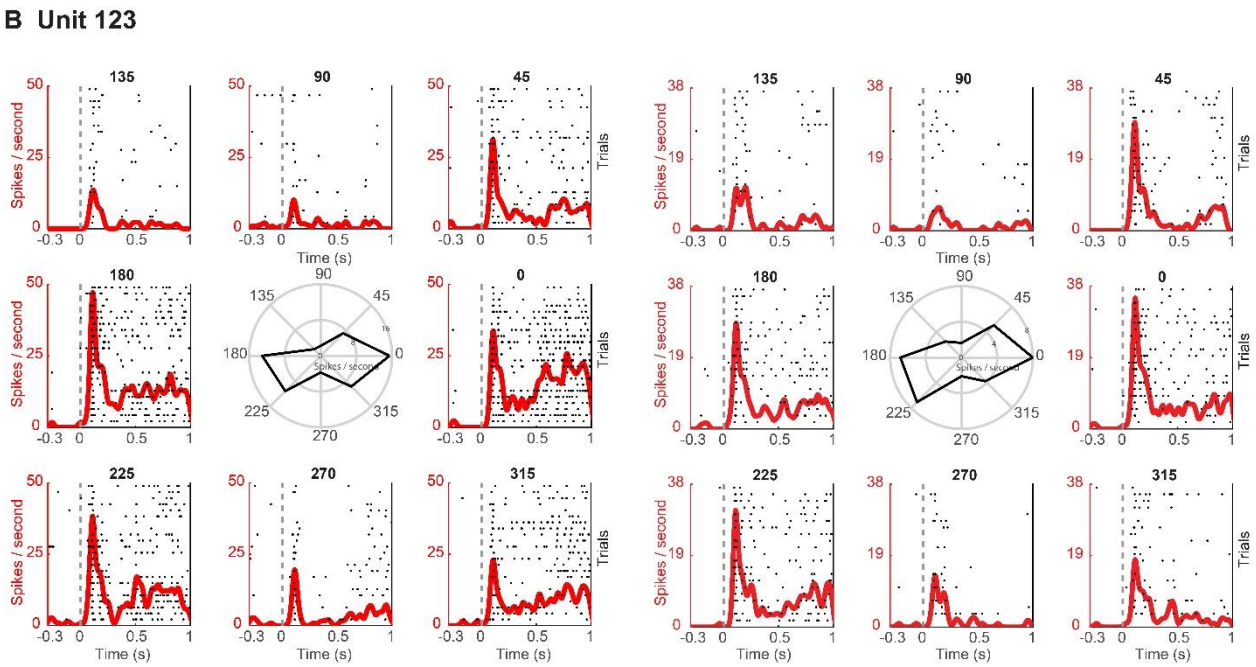
Similar to Figure 4, plotted here are the results obtained with the multivariate pattern analysis, when PA trials were aligned to the change in OKN instead of the TTL pulse (see methods). In (A) is the normalized spiking activity of neuronal ensembles (see methods) during the two different switch types. (B) Cross temporal decoding within and generalization across the two trial types. (C) A cross trial type generalization was carried out over a single temporal window of 400 ms before and after a switch.



Supplementary Figure 5.1

To test the contribution of eye movements on neural activity, animals participated in two control paradigms, namely fixation Off and fixation On, in two different blocks. Both of them started with cueing the animal to fixate for 300 ms, after which a stimulus drifting in a particular direction was presented monocularly. (A) During fixation Off, the fixation spot was removed at the onset of the stimulus, thus inducing optokinetic nystagmus eye movements. (B) During fixation On, the stimulus was presented without removal of the fixation spot, and the animal was required to maintain its gaze within a window (±1 or ±2 degrees) until the end of the trial, in order to receive a juice reward. (C) During both paradigms, on each trial, a stimulus drifting in one of eight different directions was presented.

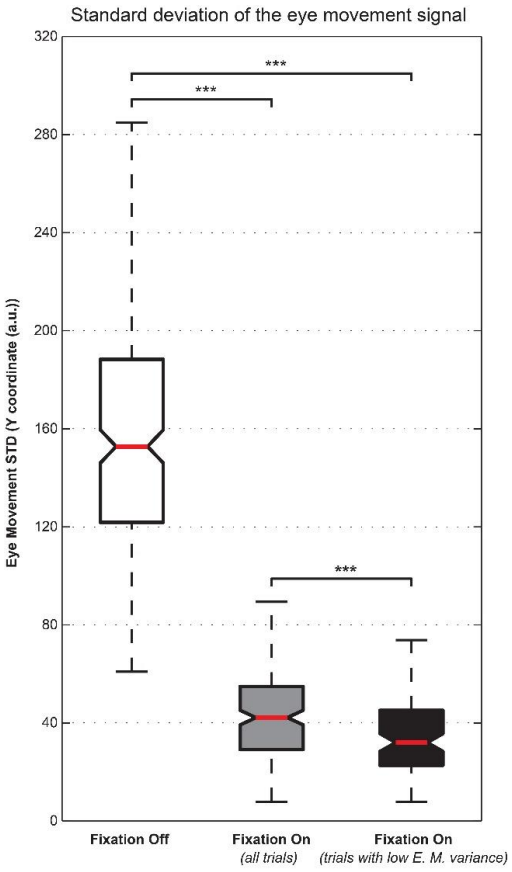




Supplementary Figure 5.2

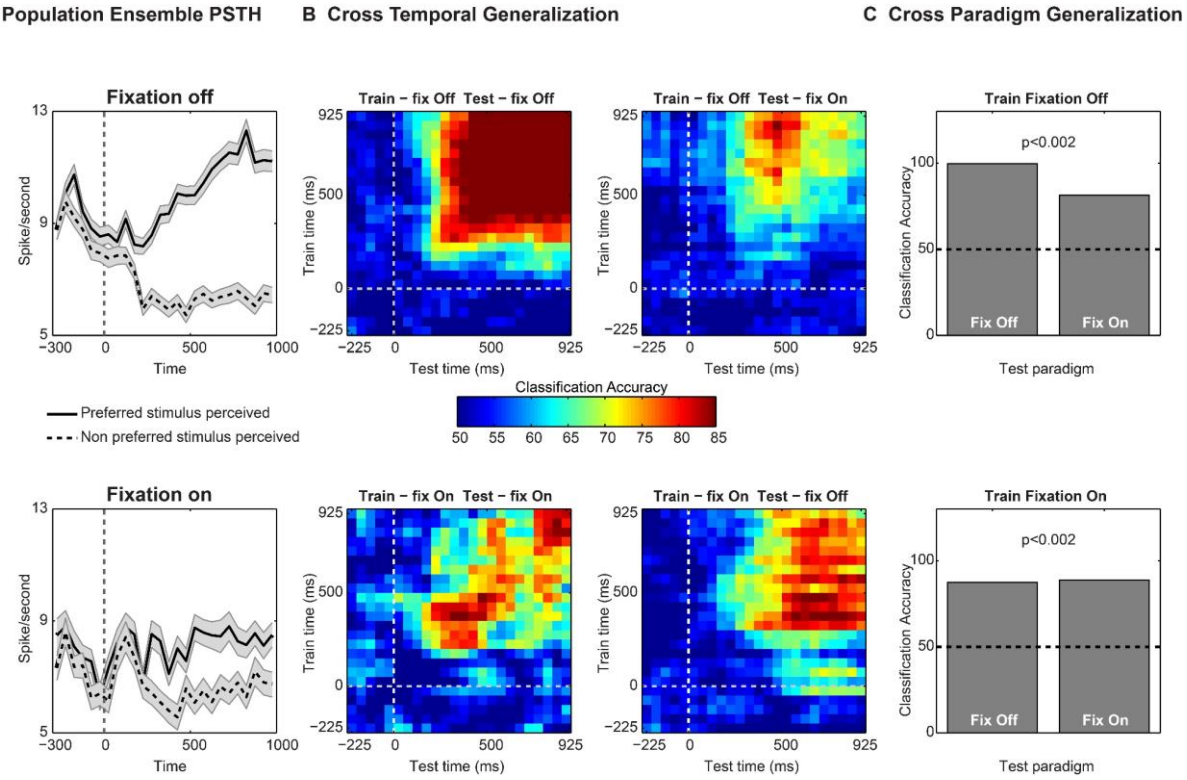
Spike density functions overlaid on spike raster plots for two units during fixation Off and fixation On. Unit activity is presented in response to eight different motion directions. In the middle are polar plots, displaying the tuning curves of each unit (average response of the unit in Hz to drifting gratings in different directions). The presented motion direction was

pseudorandomized across trials. Spike rasters are displayed for the first 'n' trials presented for every motion direction. Here, n is the minimum number of trials presented to the animal across any motion direction during a given control paradigm. PSTHs and tuning curves were computed taking all trials (of a given motion direction) into account. (A) Unit 36 displays a stronger response to stimulus with motion drifting downwards during both conditions. (B) The unit responds strongly to two opposite directions of motion, thus displaying orientation preference. Note that although the firing rate was higher during the fixation off paradigm, the unit displayed remarkably similar preference in its responses across both paradigms.



Supplementary Figure 5.3

Whisker box plots displaying the distribution of standard deviations (STD) estimated from the y-coordinate of the eye movement signal elicited during individual trials of the two control experiment during fixation Off and fixation On. The STD was computed from the eye movement signal elicited during the time window between 0 (stimulus onset) and 1000 ms (stimulus offset). For fixation On, we used either all trials, or selected trials, which displayed lower variance in the eye movement signal (see methods). The STD of the eye movement signal was significantly reduced (Wilcoxon rank sum test, *** denotes $p \leq 0.001$) during fixation On as compared to the fixation Off. The box denotes the 25th (Q1) and 75th percentiles (Q3) of the data, while the red line denotes the median. All adjacent values within $Q3 + 1.5\times(Q3-Q1)$ and $Q1-1.5\times(Q3-Q1)$ are contained within the upper and lower whisker lengths, respectively. The 95% confidence interval around the median is approximated by the notches, whose edges are calculated as median $\pm$ 1.57 (Q3−Q1)/(square root of number of samples).



Supplementary Figure 5.4

Similar to Figure 5, the figure summarizes the results pertaining to the multivariate pattern analysis assessing the invariance of the population code to motion content during the control experiment. However, only a small selection of the trials from the fixation On paradigm are included, where eye movements were further controlled (see methods and Supplementary Figure 5.3). (A) Population spiking activity (see methods) of prefrontal ensembles is presented during fixation off and fixation On paradigm. The population consisted of units which were significantly modulated in either of the two paradigms, and preferred the same motion direction (B) Cross-temporal decoding of stimulus contents during the two paradigms. Decoding accuracy was tested for each pair of train and test time windows similar to Figure 4 and 5 (C) The cross paradigm invariance of the population code was tested by training a classifier on activity in one paradigm and testing on the other, for a single bin of 400 ms (starting 400 ms post stimulus onset) during the presentation of the visual stimulus. We observed significant

(permutation test, $p < 0.002$) cross-task generalization accuracy, thus suggesting that the underlying code is largely invariant to the presence of large OKN, and encodes stimulus motion contents.

ACKNOWLEDGMENTS

This study was supported by the Max Planck Society. We would like to thank Prof. Nicho Hatsopoulos and Dr. Yusuke Murayama for help with animal surgery and Axel Oeltermann for his excellent technical help.

AUTHOR CONTRIBUTIONS

V.K., A.D. and T.I.P. designed the study. V.K., A.D. and S.S. trained animals. V.K. and A.D. performed experiments and collected data, with occasional help from S.S. V.K. and A.D. analyzed the data. S.S. contributed to spike sorting and selectivity analysis of control experiments. M.B. contributed to the decoding analysis. V.K. prepared and arranged the figures in the final format. S.S. provided the MATLAB generated version of the figures displayed in figure 5A, 5.2, 5.3 and 5.4 A. T.I.P. and N.K.L. supervised the study. N.K.L. and J.W. contributed unpublished reagents/analytical tools. N.K.L. provided the support to the group. V.K. and T.I.P. wrote the original manuscript draft. All authors participated in discussion and interpretation of the results and editing the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

Bibliography

1. R. Adolphs, The unsolved problems of neuroscience. *Trends Cogn Sci (Regul Ed)*. **19**, 173–175 (2015).

2. F. C. Crick, C. Koch, *23 problems in systems neuroscience* (Oxford University Press, 2006).

3. G. Miller, What is the biological basis of consciousness? *Science*. **309**, 79 (2005).

4. F. Crick, C. Koch, Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences* (1990).

5. B. J. Baars, Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog. Brain Res.* **150**, 45–53 (2005).

6. F. Crick, C. Koch, Consciousness and neuroscience. *Cereb. Cortex*. **8**, 97–107 (1998).

7. D. A. Leopold, N. K. Logothetis, Multistable phenomena: changing views in perception. *Trends Cogn Sci (Regul Ed)*. **3**, 254–264 (1999).

8. H. Lau, D. Rosenthal, Empirical support for higher-order theories of conscious awareness. *Trends Cogn Sci (Regul Ed)*. **15**, 365–373 (2011).

9. S. Dehaene, J.-P. Changeux, Experimental and theoretical approaches to conscious processing. *Neuron*. **70**, 200–227 (2011).

10. G. Rees, G. Kreiman, C. Koch, Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* **3**, 261–270 (2002).

11. J. P. Vignal, P. Chauvel, E. Halgren, Localised face processing by the human prefrontal cortex: stimulation-evoked hallucinations of faces. *Cogn. Neuropsychol.* **17**, 281–291 (2000).

12. O. Blanke, T. Landis, M. Seeck, Electrical cortical stimulation of the human prefrontal cortex evokes complex visual hallucinations. *Epilepsy Behav.* **1**, 356–361 (2000).

13. B. Odegaard, R. T. Knight, H. Lau, Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* **37**, 9593–9602 (2017).

14. A. Del Cul, S. Dehaene, P. Reyes, E. Bravo, A. Slachevsky, Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*. **132**, 2531–2540 (2009).

15. S. M. Fleming, J. Ryu, J. G. Golfinos, K. E. Blackmon, Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*. **137**, 2811–2822 (2014).

16. S. M. Szczepanski, R. T. Knight, Insights into human behavior from lesions to the prefrontal cortex. *Neuron*. **83**, 1002–1018 (2014).

17. R. K. Nakamura, M. Mishkin, Chronic "blindness" following lesions of nonvisual cortex in the monkey. *Exp. Brain Res.* **63**, 173–184 (1986).

18. I. Colás *et al.*, Conscious perception in patients with prefrontal damage. *Neuropsychologia*. **129**, 284–293 (2019).

19. B. van Vugt *et al.*, The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*. **360**, 537–542 (2018).

20. T. I. Panagiotaropoulos, G. Deco, V. Kapoor, N. K. Logothetis, Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron*. **74**, 924–935 (2012).

21. C. Libedinsky, M. Livingstone, Role of prefrontal cortex in conscious visual perception. *J. Neurosci.* **31**, 64–69 (2011).

22. H. Gelbard-Sagiv, L. Mudrik, M. R. Hill, C. Koch, I. Fried, Human single neuron activity precedes emergence of conscious perception. *Nat. Commun.* **9**, 2057 (2018).

23. G. Tononi, M. Boly, M. Massimini, C. Koch, Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* **17**, 450–461 (2016).

24. C. Koch, What Is Consciousness? *Nature*. **557**, S8–S12 (2018).

25. M. Boly *et al.*, Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? clinical and neuroimaging evidence. *J. Neurosci.* **37**, 9603–9613 (2017).

26. C. Koch, M. Massimini, M. Boly, G. Tononi, Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* **17**, 307–321 (2016).

27. S. Frässle, J. Sommer, A. Jansen, M. Naber, W. Einhäuser, Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* **34**, 1738–1747 (2014).

28. T. Knapen, J. Brascamp, J. Pearson, R. van Ee, R. Blake, The role of frontal and parietal brain areas in bistable perception. *J. Neurosci.* **31**, 10293–10301 (2011).

29. J. Brascamp, R. Blake, T. Knapen, Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nat. Neurosci.* **18**, 1672–1678 (2015).

30. S. Safavi, V. Kapoor, N. K. Logothetis, T. I. Panagiotaropoulos, Is the frontal lobe involved in conscious perception? *Front. Psychol.* **5**, 1063 (2014).

31. N. Zaretskaya, M. Narinyan, Introspection, attention or awareness? The role of the frontal lobe in binocular rivalry. *Front. Hum. Neurosci.* **8**, 527 (2014).

32. J. Brascamp, P. Sterzer, R. Blake, T. Knapen, Multistable perception and the role of the frontoparietal cortex in perceptual inference. *Annu. Rev. Psychol.* **69**, 77–103 (2018).

33. C. Wheatstone, Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*. **128**, 371–394 (1838).

34. R. Blake, N. K. Logothetis, Visual competition. *Nat. Rev. Neurosci.* **3**, 13–21 (2002).

35. N. Tsuchiya, M. Wilke, S. Frässle, V. A. F. Lamme, No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends Cogn Sci (Regul Ed)*. **19**, 757–

770 (2015).

36. M. A. Pitts, S. Metzler, S. A. Hillyard, Isolating neural correlates of conscious perception from neural correlates of reporting one's perception. *Front. Psychol.* **5**, 1078 (2014).

37. J. Aru, T. Bachmann, W. Singer, L. Melloni, Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* **36**, 737–746 (2012).

38. N. K. Logothetis, J. D. Schall, Binocular motion rivalry in macaque monkeys: eye dominance and tracking eye movements. *Vision Res.* **30**, 1409–1419 (1990).

39. M. Wei, F. Sun, The alternation of optokinetic responses driven by moving stimuli in humans. *Brain Res.* **813**, 406–410 (1998).

40. M. Naber, S. Frässle, W. Einhäuser, Perceptual rivalry: reflexes reveal the gradual nature of visual awareness. *PLoS ONE*. **6**, e20910 (2011).

41. R. Fox, S. Todd, L. A. Bettinger, Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vision Res.* **15**, 849–853 (1975).

42. J. M. Wolfe, Reversing ocular dominance and suppression in a single flash. *Vision Res.* **24**, 471–478 (1984).

43. R. W. Lansing, Electroencephalographic correlates of binocular rivalry in man. *Science*. **146**, 1325–1327 (1964).

44. W. J. Levelt, Note on the distribution of dominance times in binocular rivalry. *Br. J. Psychol.* **58**, 143–145 (1967).

45. I. N. Pigarev, G. Rizzolatti, C. Scandolara, Neurons responding to visual stimuli in the frontal lobe of macaque monkeys. *Neurosci. Lett.* **12**, 207–212 (1979).

46. F. A. Wilson, S. P. Scalaidhe, P. S. Goldman-Rakic, Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*. **260**, 1955–1958 (1993).

47. D. Zaksas, T. Pasternak, Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* **26**, 11726–11742 (2006).

48. G. A. Keliris, N. K. Logothetis, A. S. Tolias, The role of the primary visual cortex in perceptual suppression of salient visual stimuli. *J. Neurosci.* **30**, 12353–12365 (2010).

49. N. K. Logothetis, Single units and conscious vision. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **353**, 1801–1818 (1998).

50. D. L. Sheinberg, N. K. Logothetis, The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci USA*. **94**, 3408–3413 (1997).

51. E. Meyers, G. Kreiman, in *Visual Population Codes*, N. Kriegeskorte, G. Kreiman, Eds. (2011), pp. 517–538.

52. E. M. Meyers, D. J. Freedman, G. Kreiman, E. K. Miller, T. Poggio, Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J.*

Neurophysiol. **100**, 1407–1419 (2008).

53. M. Dieterich, S. F. Bucher, K. C. Seelos, T. Brandt, Horizontal or vertical optokinetic stimulation activates visual motion-sensitive, ocular motor and vestibular cortex areas with right hemispheric dominance. An fMRI study. *Brain*. **121 (Pt 8)**, 1479–1495 (1998).

54. J. N. Kim, M. N. Shadlen, Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat. Neurosci.* **2**, 176–185 (1999).

55. E. Lowet *et al.*, Enhanced Neural Processing by Covert Attention only during Microsaccades Directed toward the Attended Stimulus. *Neuron*. **99**, 207-214.e3 (2018).

56. G. A. Mashour, The controversial correlates of consciousness. *Science*. **360**, 493–494 (2018).

57. T. I. Panagiotaropoulos, V. Kapoor, N. K. Logothetis, Subjective visual perception: from local processing to emergent phenomena of brain activity. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **369**, 20130534 (2014).

58. M. C. Schmid, A. Maier, To see or not to see--thalamo-cortical networks during blindsight and perceptual suppression. *Prog. Neurobiol.* **126**, 36–48 (2015).

59. D. A. Leopold, N. K. Logothetis, Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*. **379**, 549–553 (1996).

60. N. K. Logothetis, J. D. Schall, Neuronal correlates of subjective visual perception. *Science*. **245**, 761–763 (1989).

61. S. P. O Scalaidhe, F. A. Wilson, P. S. Goldman-Rakic, Areal segregation of face-processing neurons in prefrontal cortex. *Science*. **278**, 1135–1138 (1997).

62. S. P. Scalaidhe, F. A. Wilson, P. S. Goldman-Rakic, Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cereb. Cortex*. **9**, 459–475 (1999).

63. M. J. Webster, J. Bachevalier, L. G. Ungerleider, Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb. Cortex*. **4**, 470–483 (1994).

64. K. G. Thompson, J. D. Schall, The detection of visual signals by macaque frontal eye field during masking. *Nat. Neurosci.* **2**, 283–288 (1999).

65. M. Michel, J. Morales, MINORITY REPORTS: CONSCIOUSNESS AND THE PREFRONTAL CORTEX.

66. J. Morales, H. Lau, The neural correlates of consciousness.

67. M. Rigotti *et al.*, The importance of mixed selectivity in complex cognitive tasks. *Nature*. **497**, 585–590 (2013).

68. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. **503**, 78–84 (2013).

69. V. Kapoor, M. Besserve, N. K. Logothetis, T. I. Panagiotaropoulos, Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. *Commun. Biol.* **1**, 215 (2018).

70. M. Schneider, V. G. Kemper, T. C. Emmerling, F. De Martino, R. Goebel, Columnar clusters in the human motion complex reflect consciously perceived motion axis. *Proc Natl Acad Sci USA*. **116**, 5096–5101 (2019).

71. S. Liu, Q. Yu, P. U. Tse, P. Cavanagh, Neural correlates of the conscious perception of visual location lie outside visual cortex. *Curr. Biol.* **29**, 4036-4044.e4 (2019).

72. R. Brown, H. Lau, J. E. LeDoux, Understanding the Higher-Order Approach to Consciousness. *Trends Cogn Sci (Regul Ed)*. **23**, 754–768 (2019).

73. Comparing the major theories of consciousness. - PsycNET, (available at https://psycnet.apa.org/record/2009-19897-077).

74. D. Pal *et al.*, Differential role of prefrontal and parietal cortices in controlling level of consciousness. *Curr. Biol.* **28**, 2145-2152.e5 (2018).

75. J. H. Marshel *et al.*, Cortical layer-specific critical dynamics triggering perception. *Science*. **365** (2019), doi:10.1126/science.aaw5202.

76. N. Block, What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends Cogn Sci (Regul Ed)*. **23**, 1003–1013 (2019).

77. I. I. Goldberg, M. Harel, R. Malach, When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron*. **50**, 329–339 (2006).

78. M. Watanabe *et al.*, Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science*. **334**, 829–831 (2011).

79. M. N. Shadlen, R. Kiani, in *Characterizing Consciousness: From Cognition to the Clinic?*, S. Dehaene, Y. Christen, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), *Research and Perspectives in Neurosciences*, pp. 27–46.

80. Y. H. R. Kang, F. H. Petzschner, D. M. Wolpert, M. N. Shadlen, Piercing of Consciousness as a Threshold-Crossing Operation. *Curr. Biol.* **27**, 2285-2295.e6 (2017).

81. N. Block, in *The Future of the Brain*, G. MARCUS, J. FREEMAN, Eds. (Princeton University Press, 2015), *Essays by the World's Leading Neuroscientists*, p. 161.

82. N. Logothetis, H. Merkle, M. Augath, T. Trinath, K. Ugurbil, Ultra high-resolution fMRI in monkeys with implanted RF coils. *Neuron*. **35**, 227–242 (2002).

83. E. M. Maynard, C. T. Nordhausen, R. A. Normann, The Utah intracortical Electrode Array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* **102**, 228–239 (1997).

84. R. Q. Quiroga, Z. Nadasdy, Y. Ben-Shaul, Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* **16**, 1661–1687 (2004).

85. A. S. Tolias *et al.*, Recording chronically from the same neurons in awake, behaving primates. *J. Neurophysiol.* **98**, 3780–3790 (2007).

86. E. M. Meyers, The neural decoding toolbox. *Front. Neuroinformatics*. **7**, 8 (2013).

87. Y. Zhang *et al.*, Object decoding with attention in inferior temporal cortex. *Proc Natl Acad Sci USA*. **108**, 8850–8855 (2011).

Prefrontal state fluctuations control access to consciousness

Abhilash Dwarakanath[1*], Vishal Kapoor[1*], Joachim Werner[1], Shervin Safavi[1,2], Leonid A. Fedorov[1], Nikos K. Logothetis[1, 3§], Theofanis I. Panagiotaropoulos [1,4§]

[1] Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany

[2] International Max Planck Research School, Tübingen 72076, Germany

[3] Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK

[4] Cognitive Neuroimaging Unit, CEA, DSV/I2BM, INSERM, Universite Paris-Sud, Universite Paris-Saclay, Neurospin Center, 91191 Gif/Yvette, France

[*] *These authors contributed equally to this work*

§ Co-senior authors

Correspondence:

abhilash.dwarakanath@tuebingen.mpg.de

theofanis.panagiotaropoulos@tuebingen.mpg.de

Abstract

In perceptual multistability, the content of consciousness alternates spontaneously between different interpretations of unchanged sensory input. The source of these internally driven transitions in conscious perception is unknown. Here we show that transient, low frequency (1-9 Hz) perisynaptic bursts in the macaque lateral prefrontal cortex precede spontaneous perceptual transitions in a no-report binocular motion rivalry task. These low-frequency transients suppress 20-40 Hz oscillatory bursts that selectively synchronise the discharge activity of neuronal ensembles signalling conscious content. Similar ongoing state changes, with dynamics resembling the temporal structure of spontaneous perceptual alternations during rivalry, dominate the prefrontal cortex during resting-state, thus pointing to their default, endogenous nature. Our results suggest that prefrontal state fluctuations control access to consciousness through a reorganisation in the activity of feature-specific neuronal ensembles.

One sentence summary

Prefrontal state transitions precede spontaneous transitions in the content of consciousness.

When the visual system is confronted with ambiguous sensory information, conscious awareness spontaneously fluctuates between different possible perceptual interpretations (*1*, *2*). In an unpredictable manner, one of the competing sensory representations temporarily gains access to consciousness while the others become perceptually suppressed, therefore dissociating sensory input from conscious perception. The mechanism that is responsible for this internally driven passage of sensory input from non-conscious processing to conscious access and vice versa is currently unknown.

Binocular rivalry (BR) is a well-known and commonly studied paradigm of multistable perception in the effort to identify the mechanism that underlies conscious awareness. In BR, the content of consciousness spontaneously alternates between two disparate stimuli that are continuously presented in each eye (*3–5*). Several mechanisms localised in and associated with different cortical areas, including competition between monocular neurons in the primary visual cortex (V1) and activation of a widespread cortical network driven from a non-linear ignition event in the prefrontal cortex (PFC) have been hypothesised to drive these spontaneous perceptual transitions and conscious access (*6*, *7*). However, monocular V1 neuron competition seems insufficient to explain BR (*8*) which is an instance of perceptual multistability involving competition between higher order perceptual representations, that are not bound to eye-specific input (*9*). Furthermore, although the PFC has been suggested to mediate visual consciousness (*10–16*), whether prefrontal signals contribute to or drive the emergence of conscious awareness, or simply reflect the cognitive consequences of reporting the content of perception like monitoring and introspection is still unclear (*17–22*). The confusion largely stems from probing the PFC using different signals (e.g. hemodynamic response using functional magnetic resonance imaging (fMRI) vs. electrophysiology) during voluntary perceptual report or no-report paradigms of exogenous or intrinsically driven changes in conscious perception (*15*, *19*, *20*, *23–26*).

Here we studied the mechanisms underlying conscious awareness in the macaque lateral PFC using a no-report BR paradigm. This allowed us to detect intrinsically driven transitions in conscious perception of opposing directions-of-motion. We combined this task with multielectrode recordings of local field potentials (LFPs) and simultaneously sampled direction-of-motion selective, competing ensembles. By using the optokinetic nystagmus (OKN) reflex as an objective criterion of perceptual state transitions, we removed any effects of voluntary motor reports on neural activity, identifying signals directly related to spontaneous transitions in the content of consciousness. We show here for the first time that prefrontal state changes in the low-frequency (1-9Hz) and beta regimes (20-40Hz) control access to conscious awareness.

Results

We used a no-report paradigm of binocular motion rivalry coupled with multielectrode extracellular recordings of LFPs and direction-of-motion selective neuronal ensembles in the inferior convexity of the macaque PFC (Fig. 1A). Two types of trials were employed: a) physical alternation (PA) of monocularly alternating, opposing directions of motion and b) binocular rivalry (BR), where the initial direction of motion was not removed but was followed by a flashed, opposing direction of motion in the contralateral eye (Fig. 1B, upper panel). Initially, this manipulation results in an externally-induced period of perceptual suppression of variable duration for the first stimulus (binocular flash suppression - BFS) which is followed by spontaneous perceptual transitions since the two competing representations start to rival for access to consciousness. In order to exclude the effect of voluntary perceptual reports on neural activity, the macaques were not trained to report their percept. Instead, the polarity of their motion-induced OKN elicited during passive observation of the stimuli (in both conditions, i.e. BR and PA), and previously shown to provide an accurate perceptual state read-out in both

humans and macaques (*25*, *27*), was used to infer perceptual dominance periods (Fig. 1B, lower panel). These dominance durations followed a gamma distribution, a hallmark of multistable perception, with a median dominance duration of 1.54±1.28s (median±SD) for spontaneous transitions in BR and 2.25±2.21s for transitions involving exogenous perceptual suppression in BFS (Fig. 1C, SI Fig. 1).

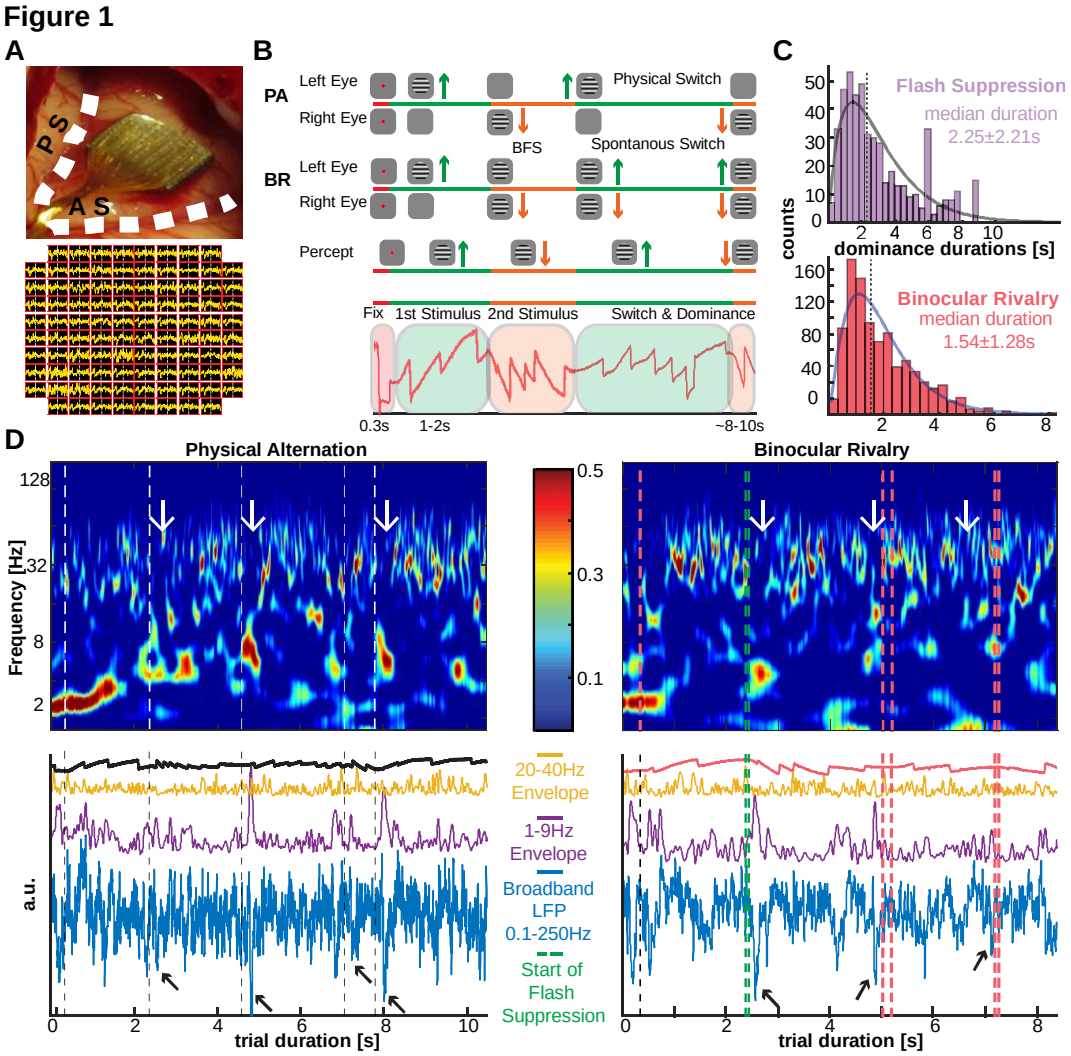


Figure 1 | Experiment and typical perisynaptic signals.

A. Multielectrode array in the inferior convexity of the PFC (top). AS: arcuate sulcus; PS: principal sulcus. Below, example spatial map of 0.1-250 Hz LFP signals around a spontaneous (white line) perceptual transition.

B. Task structure: For both PA and BR trials, following binocular fusion and fixation of a red dot (0.2°) for 300ms a grating (or 200 random dots at 100% coherence in one session) moving upward (90°, green) or downward (270°, orange) was initially presented in one eye (8°, 12-13°/sec, 100% contrast). After 1000-2000 ms the first stimulus

was removed in PA trials and followed by the presentation of the opposing motion direction and subsequently multiple monocular switches up to 8-10 seconds. In BR trials, after the presentation of the opposing motion direction (resulting in BFS) the two stimuli were let to compete for conscious access resulting in spontaneous perceptual switches. Different OKN patterns (red trace, highlighted in green and orange) elicited from two directions of motion allowed decoding of the conscious percept.

C. Histogram and gamma distribution function fit of perceptual dominance times during BFS (2.25±2.21s, median±SD) and spontaneous perceptual transitions during BR (1.54±1.28s, median±SD).

D. Channel-averaged normalized (z-score) time-frequency spectrograms for a single trial/observation period of PA (left) and BR (right) is shown at the top. White lines in PA reflect the manually marked change in the OKN polarity after the onset of the exogenous, monocular stimulus changes. Green lines in BR represent the start of the flash suppression phase (at 2.3sec) whereas the red lines represent the subsequent spontaneous perceptual transitions. 1-9 Hz bursts suppressing 20-40 Hz activity occurred following a switch in PA but before a transition in BR (white arrows). Bottom panels: Broadband LFP, instantaneous 1-9 Hz and 20-40 Hz signal amplitude and OKN traces for the same observation period. Black arrows point to negative deflections in the 0.1-250 Hz LFP trace (for display, all traces were normalised and plotted with an arbitrary shift to clearly delineate the different regimes).

Prefrontal state fluctuations precede conscious awareness

To identify mesoscopic prefrontal signals preceding spontaneous perceptual transitions, we first analysed the perisynaptic activity dynamics reflected in the LFPs (Figure 1D, SI Fig. 2). Transient negative deflections of the channel-averaged raw (0.1-250 Hz) LFPs (blue traces in Fig. 1D), suggestive of brief depolarising states, were observed to disrupt a default state of oscillatory bursts in the beta range (20-40 Hz) throughout the observation periods in both PA and BR trials. However, the strongest negative deflections appeared to occur just after the change in the OKN polarity induced by external stimulus changes in PA (Fig. 1D, left), but just before spontaneous perceptual transitions in BR (Fig. 1D, right). Pooling all transitions in PA (n=802) revealed that the power of these transient negative deflections was concentrated immediately after exogenous stimulus changes in a low-frequency (1-9 Hz) range that resulted

in a temporally transient suppression of the ongoing beta activity (20-40 Hz) (Fig. 2A, left).

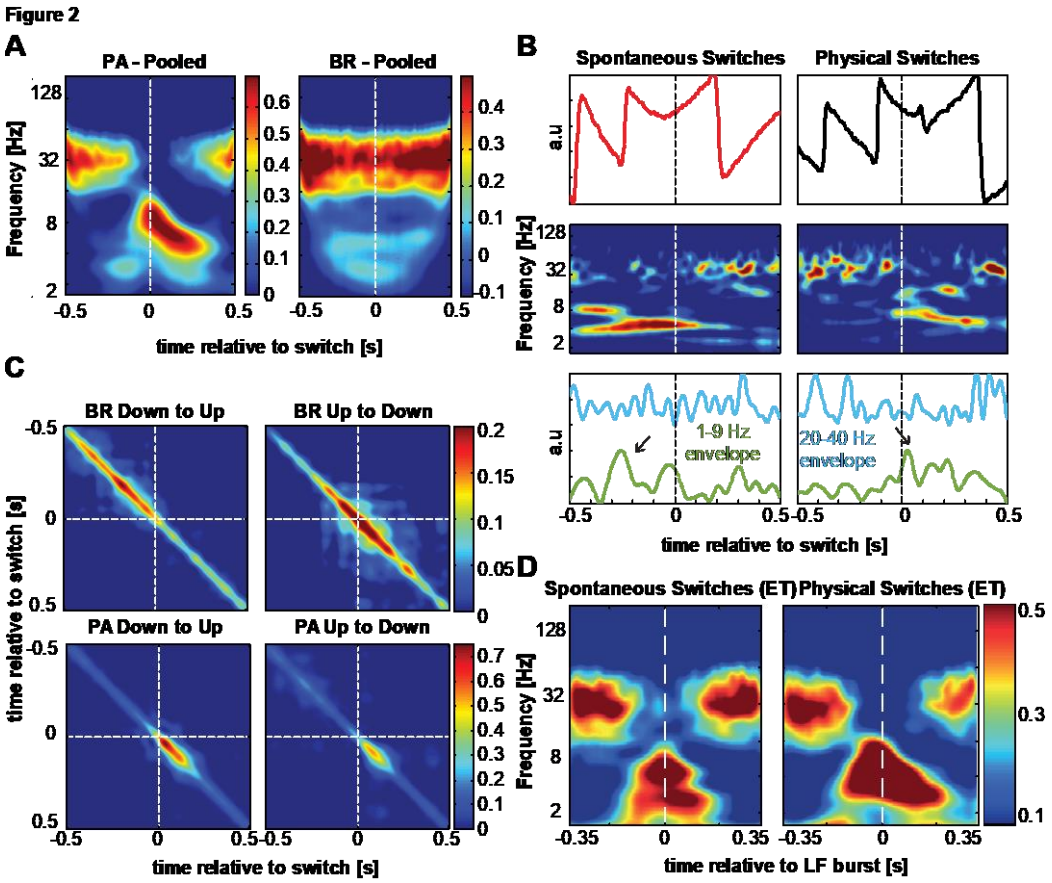


Figure 2| Time-frequency LFP analysis.

A. Grand average time-frequency analysis of all physical (left) and spontaneous (right) perceptual transitions. Spectrograms are aligned (t=0) to manually marked OKN changes in both PA and BR for periods of stable perceptual dominance before and after the switch.

B. Upper panel: OKN traces around a single physical (black) and spontaneous (red) transition. Middle panel: channel-averaged normalised spectrograms aligned to the OKN slope change for the two conditions. Lower panel: Normalised instantaneous amplitudes of the two modulated frequency bands (i.e. 1-9 Hz, green and 20-40 Hz, cyan) identified from the spectrograms. Low-frequency bursts occur after the physical switch but before the spontaneous transition (black arrows).

C. Differences in the onset of the low-frequency activity across physical and spontaneous transitions in direction of motion are reflected in the temporal auto-covariance of the low frequency envelopes across the array recorded simultaneously for every transition type (down to up, left and up to down, right). Most of the similarity is observed before a transition in BR (upper panel) but after a transition in PA (bottom panel).

D. Grand average low-frequency peak aligned spectrograms after a physical and before a spontaneous transition. Low frequency activation results in beta power suppression for around 300ms both PA and BR.

Low frequency induced beta suppression was also observed for intrinsically generated perceptual transitions in BR (n=573); however, it started well before (~ 400ms) the spontaneous OKN change (Fig. 2B, right). In BR, the absence of a consistent feedforward response, locked to an external change of the sensory input, resulted in a temporal jitter of the low-frequency transients across different transitions and neuronal sites (SI Fig. 3). In individual transitions, the low-frequency-triggered beta-suppression started clearly before the spontaneous OKN transition (Fig. 2B). Indeed, the temporal-covariance of the channel-averaged low-frequency signal across transitions was concentrated well before the perceptual change in BR, whereas it was heavily concentrated in the post-transition period during PA (Fig. 2C). Computing spectrograms aligned to the low-frequency event peaks detected before and after the transition in BR and PA respectively showed the similarity in the coupling of low-frequency transients and beta-burst suppression between the two conditions (Fig. 2D). Therefore, 1-9 Hz suppression of 20-40 Hz activity follows exogenous stimulus changes in PA, but precedes spontaneous OKN- inferred perceptual transitions in BR.

Since both low frequency and beta activity were not sustained oscillations but appeared to occur in bursts, we quantified the burst-rate of low-frequency and beta activity before and after the time of exogenous (PA) and endogenous (BR) perceptual transitions using a burst-rate metric (described in methods). Low frequency burst rate (bursts/transition/channel) was significantly higher after the OKN change in PA (0.35±0.0045, n = 26812, post-transition, vs. 0.09±0.0016, pre-transition, n = 6612; $p<10^{-185}$ mean ± SEM), but before the OKN change in BR (0.17 ± 0.002, n = 9667, pre-transition vs 0.14 ± 0.002, n = 7730, post-transition, $p<10^{-43}$ mean ± SEM) (Fig. 3A and SI Figure 4). Furthermore, low frequency bursts were significantly

more before a spontaneous perceptual transition than before a physical transition (0.17 ± 0.002, n = 9667, pre-transition BR vs. 0.09 ± 0.0016, n = 6612, pre-transition PA; $p<10^{-147}$ mean ± SEM). Importantly, low-frequency bursts, including the last-detected burst before a transition, occurred on average even before the end of the last dominance period preceding a transition in BR (end of dominance: -97.4 ± 140ms, low-frequency bursts: -198 ± 133ms, median±SD, $p<10^{-67}$, Fig. 3B, and Fig. 2B for an example transition). As expected, low frequency bursts occurred predominantly and significantly after the OKN change in PA (64ms ± 147ms, median±SD).

Figure 3

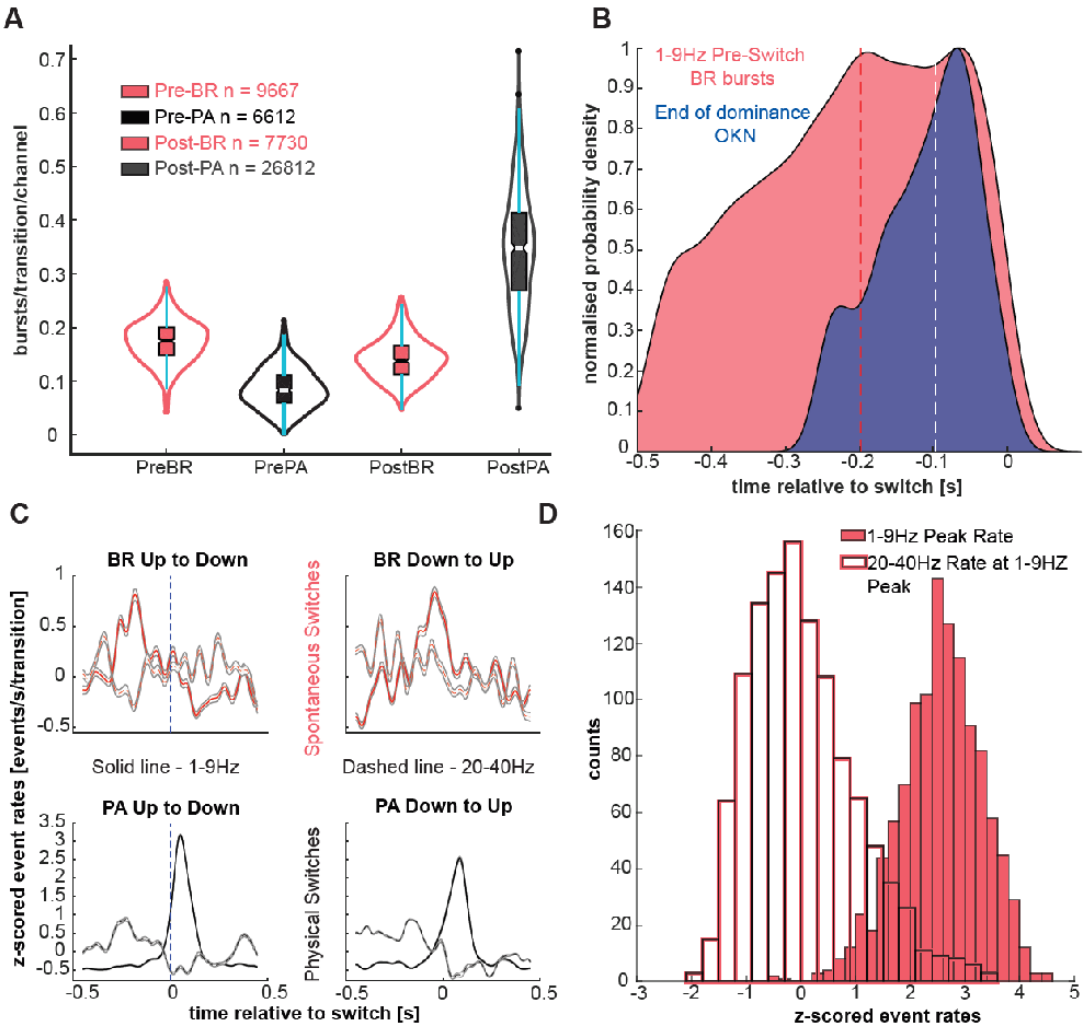


Figure 3| LFP burst rate analysis

A. Burst rate per transition per channel for periods before and after transitions in PA (red) and BR (black). In

order from left to right: Pre-BR (dark pink), Pre-PA (black), Post-BR (light pink), Post-PA (grey). The whiskers of the box-plots show the dispersion of the data. White lines depict median values. Matching colour dots represent outliers. More low frequency bursts occur before spontaneous, but after physical switches. Burst rate before a physical switch is very low, suggesting noise levels. This baseline burst-rate needs to be ramped up for a switch to occur. PA: (0.35 ± 0.0045, n = 26812, post-transition, vs. 0.09 ± 0.0016, pre-transition, n = 6612; $p<10^{-185}$ mean ± SEM) BR: (0.17 ± 0.002, n = 9667, pre-transition vs 0.14 ± 0.002, n = 7730, post-transition, $p<10^{-43}$ mean ± SEM).

B. Distribution of low frequency burst-times and OKN times marking the end of the previous dominance period before a spontaneous transition. We fit the probability distribution using a kernel density estimate with a variable width. These functions were then normalised for direct comparison. Low-frequency bursts on average occurred at -198±133ms before a switch, whereas the end of dominance occurred at -97.4±140ms before a transition to the competing representation ($p<10^{-67}$ median±SD).

C. Normalised (z-score) burst rate in time (events/s/transition) during BR (red lines) and PA trials (white lines) for low-frequency (solid lines) and beta activity (dashed lines).

D. Distribution of beta-band rate at the peak of low-frequency rate before spontaneous perceptual transitions in BR. (r = -0.08, p=0.0071; pooled across both transition types)

The low-frequency activation before a spontaneous perceptual reversal in BR is better observed in the evolution of bursting activity in time (quasi-PSTH, i.e. the detected bursts are converted into binary event trains, smoothed and then averaged). In BR, the peak-rate of 1-9 Hz bursts occurred at -160 ± 237ms and -28 ± 199ms (median±SD) before the spontaneous perceptual transitions for the two transition types respectively (Figure 3C, top row), while in PA they occurred at 52 ± 28ms (median±SD) and 82 ± 64.5ms (median±SD) following marked OKN change (Fig. 3C, bottom row). These differences were further enhanced when the bursts towards the end of the post-switch window (bursts occurring 150ms after the transition) were discarded (SI Fig. 5). Confirming the time-frequency analysis pattern in Fig. 2A and suggesting a frequency-specific competitive process (i.e. cortical state fluctuations) in the PFC, low-

frequency and beta burst-rates were significantly anti-correlated in BR (r = -0.08, p = 0.0071; pooled across both transition types; Fig. 3D).

Spatiotemporal build-up of prefrontal activity

Are the perisynaptic transients preceding a spontaneous change in the content of consciousness random large excursions from baseline (noise) activity or do they reflect a gradual spatiotemporal build-up process that is critical for inducing a spontaneous transition? Indeed, we noticed that in many instances before a spontaneous transition, the last transient low-frequency (1-9 Hz) burst was frequently preceded by similar but of lower amplitude bursts (Fig. 1D and SI Fig. 1, 6). When the instantaneous amplitude of the low-frequency activity triggered at every switch was averaged first across populations and then across transitions, we observed a gradual increase approaching spontaneous but not physical transitions (Fig. 4A). Fitting a linear model to the relationship between the transition-averaged low frequency burst amplitudes at every time point before a transition revealed a linearly increasing relationship between the two variables before a spontaneous (adjusted $R^2 = 0.34$) but not before a physical switch (adjusted $R^2 = - 0.003$) (Fig. 4B). While this low frequency burst amplitude exhibited a gradual linear increase, the number of activated prefrontal sites abruptly increased just before a spontaneous reversal, suggesting a non-linear increase in the spatial spread of prefrontal activation just before a spontaneous change in the content of consciousness (Fig. 4C and SI Fig. 7). Taken together, these results point to a mesoscopic, spatio-temporal spread of low frequency prefrontal bursts before spontaneous perceptual reversals. Both linear and non-linear increases in the amplitude and spatial spread of activation respectively suggest the operation of both effects in a prefrontal ignition (*6*) process during BR.

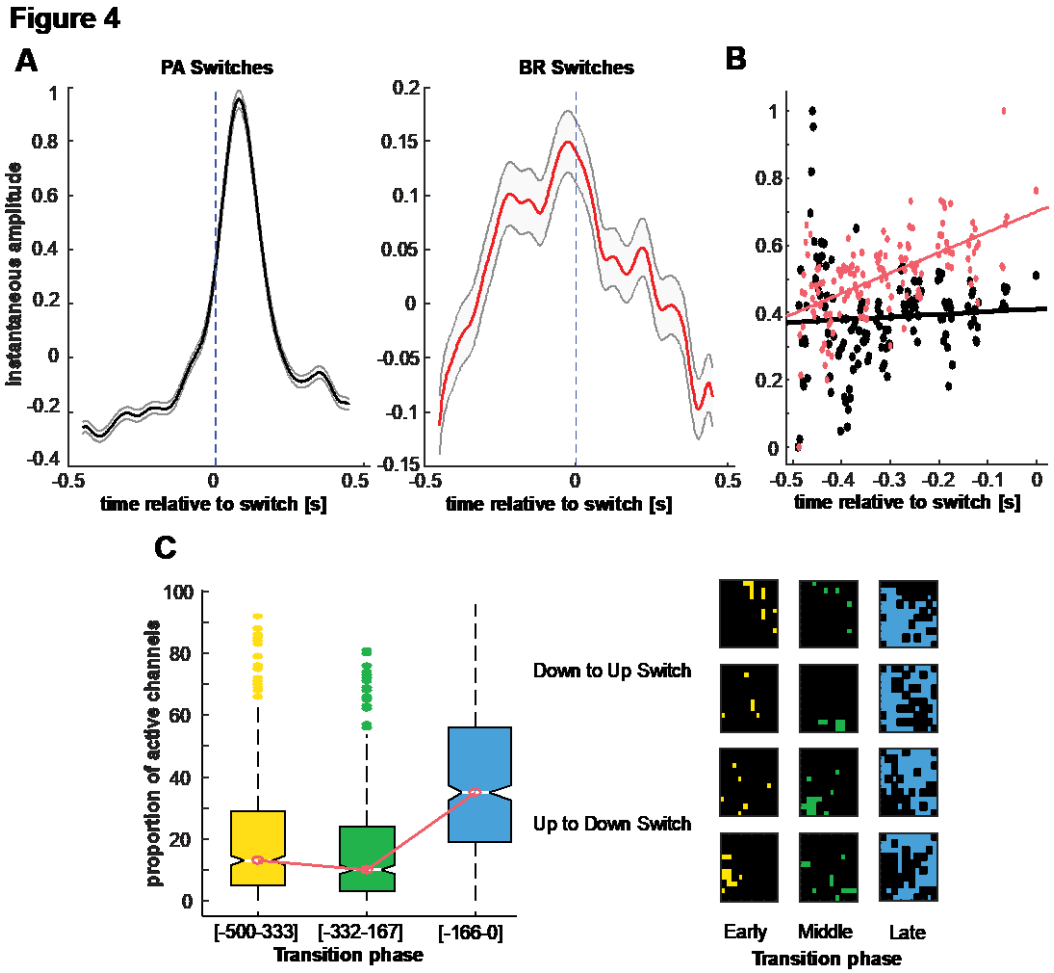


Figure 4| Spatiotemporal build-up preceding spontaneous transitions

A. Low-frequency instantaneous amplitude shows a slow climbing activity before a perceptual transition (right) but not before a physical transition (left). Curves reflect an average ± SEM across transitions of the channel-averaged activity for each collected transition.

B. Average build-up of low-frequency activity in time by fitting a linear model to the pooled and averaged amplitude at every time-point (red – BR, black – PA) across all measured neuronal sites. While before a spontaneous transition, the recorded PFC area ramps up its low-frequency activity in time, before a physical transition, it remains flat.

C. Spatio-temporal spread of low frequency activity shown in the proportion of channels for a given transition that displayed low-frequency burst peaks in early ([-500 to -333ms]), middle ([-332 to -167ms]) and late temporal windows ([-166 to 0ms]). Data are pooled across the two macaques. A significantly larger proportion ($p<0.01$) of neuronal sites peak closer to the switch, showing a sharp, non-continuous increase in the number of recruited channels (left). Activation of sites on the array shown for four examples of different types of spontaneous

transitions from 2 macaques (right).

We further hypothesised that if an increase in low-frequency bursting is critical for inducing spontaneous perceptual reversals, then low-frequency amplitude should be significantly weaker when perceptual transitions were not complete, but resulted in piecemeal (PM) periods in which perception did not unambiguously favour either of the two competing directions of motion (Fig. 5A). Subtracting the time-frequency decomposition of transitions to a PM percept from that of clean BR perceptual transitions revealed a preponderance of low-frequency activity before a transition, suggesting that the amplitude of low frequency bursts is critical for completing a perceptual transition to another period of clear dominance. Moreover, when we subtracted the averaged spectrogram of randomly triggered periods during sustained perceptual dominance from the spectrogram of activity around clear switches, we could recover the general pattern observed during these spontaneous transitions (Fig. 5B).

. This suggests that low-frequency activity should be significantly up-modulated from noise level, potentially crossing a threshold to cause a perceptual transition (SI Fig. 8).

Figure 5

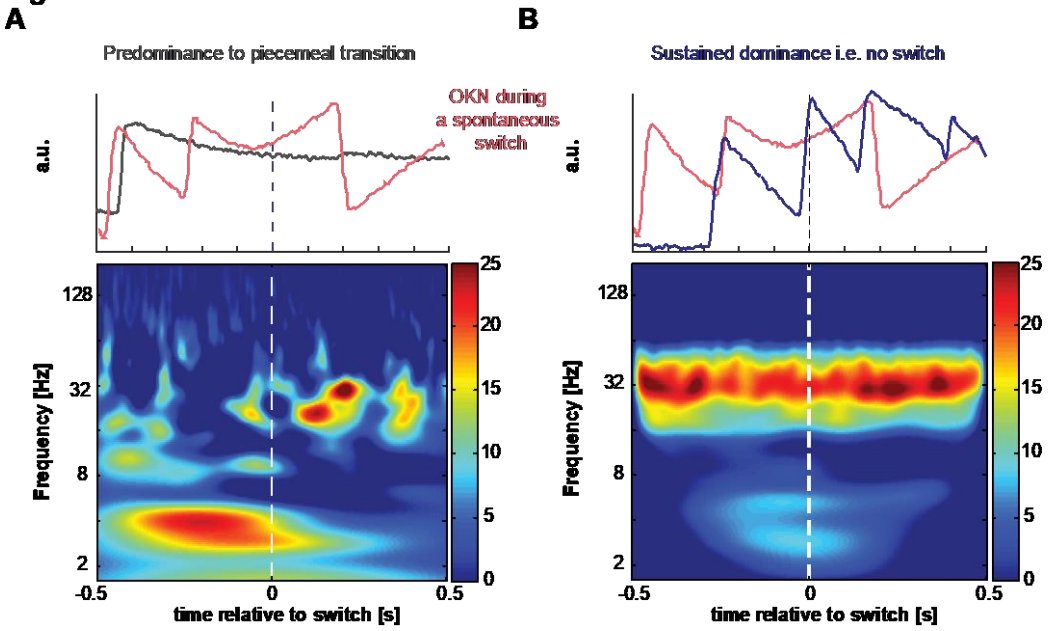


Figure 5| Low frequency amplitude is critical for inducing clear switches.

A. Top panel shows two typical OKN patterns elicited during a spontaneous transition (red) and during a transition

to piecemeal (grey). The subtracted spectrogram (BR minus PM) below shows a large difference in low-frequency activity before a spontaneous switch suggesting very weak low frequency activation during a transition to piecemeal perception.

B. Top panel shows two typical OKN patterns elicited during a spontaneous transition (red) and during sustained dominance where no switches occurred (blue). The subtracted spectrogram (BR minus no switches) recovers the time-frequency pattern observed during BR suggesting that the LFP activity during sustained perceptual dominance periods is at noise level.

Conscious content specific ensemble activity transitions and LFPs

To understand the temporal relationship between perisynaptic state fluctuations and neuronal populations reflecting the content of consciousness, we compared the convergence times (see Methods) of the normalised discharge activity of simultaneously recorded ensembles selective for the rivalling directions-of-motion (Fig. 6A, SI Fig. 9), to the low-frequency burst and peak-rate distributions. We found that discharge activity converged significantly later compared to both the low-frequency peak-rates (-60 ± 222ms for LFP event/s/transition vs. 77 ± 76ms median±SD, convergence of spiking; $p<10^{-88}$) and burst-times (-112 ± 190ms median burst time (truncated at 250ms post-switch) vs. 77 ± 76ms, median±SD convergence of spiking, $p<10^{-74}$) in spontaneous perceptual transitions. In a large number of spontaneous transitions (86.2% compared to peak-rates and 89% compared to burst times), spiking activity crossovers occurred after the median truncated low-frequency peak-rates and burst-times (Fig. 6B). Therefore, low-frequency transients could reflect a pre-conscious process preceding an intrinsically generated perceptual transition, i.e. a change in the content of consciousness by feature-selective neural ensembles.

Are these low-frequency LFP modulations non-specific or are they related to the patterns seen in the spiking activity of selective ensembles that reflect the current content of consciousness? To answer this question, we contrasted the selective-ensemble-summed low-

frequency instantaneous amplitude to their respective summed spiking PSTHs. For the same transitions, the respective ensemble spiking activity showed a clear crossover around the OKN transition (both in BR and PA, Fig. 6C), pointing to a change in encoding of the dominant percept, while no such divergence was observed in the pattern of low frequency activity. This activity is therefore distinguished from the spiking of selective neurons, and most likely reflects a pre-conscious process.

Finally, to understand how prefrontal state fluctuations could result in spiking-network and therefore perceptual reorganisation, we computed the spike-field coherence (SFC) (*28, 29*) of the simultaneously recorded, feature-selective ensemble activity and the global broadband LFP across all transitions. After a spontaneous perceptual transition, when the negative LFP deflections and therefore the low frequency (1-9 Hz) transients were less prevalent, the perceptually dominant ensemble was more coherent in the beta range (~25-40 Hz) compared to the suppressed ensemble ($p < 0.03$; Fig. 6D). However, in the period approaching a spontaneous transition when low frequency transients suppressed beta bursts, there were no differences between suppressed and dominant populations. These results suggest that prefrontal ensembles signalling the current content of consciousness are synchronised in the beta-band of the LFP. Low-frequency transients dissolve these beta-coherent ensembles, potentially increasing the likelihood for spiking in the suppressed population and therefore the likelihood for perceptual reorganisation.

Intrinsic nature of prefrontal state fluctuations

If the competition between low-frequency transients and beta-bursts found to control access to visual awareness is intrinsically generated, reflecting waking state fluctuations, traces of this process should also be observed during resting-state: i.e., in the absence of structured visual input. Indeed, in resting-state, low-frequency bursts suppressed beta activity (Fig. 6E).

Periods of uninterrupted beta activity exhibited a gamma distribution with a duration of 1.2 ± 1.44s (median±SD), which is close to the psychophysical distribution of stable perceptual dominance durations (1.54 ± 1.28s) that are also characterised by uninterrupted beta activity (Fig. 6F). Therefore, prefrontal state fluctuations appear to reflect an intrinsically generated process in the macaque PFC.

Figure 6

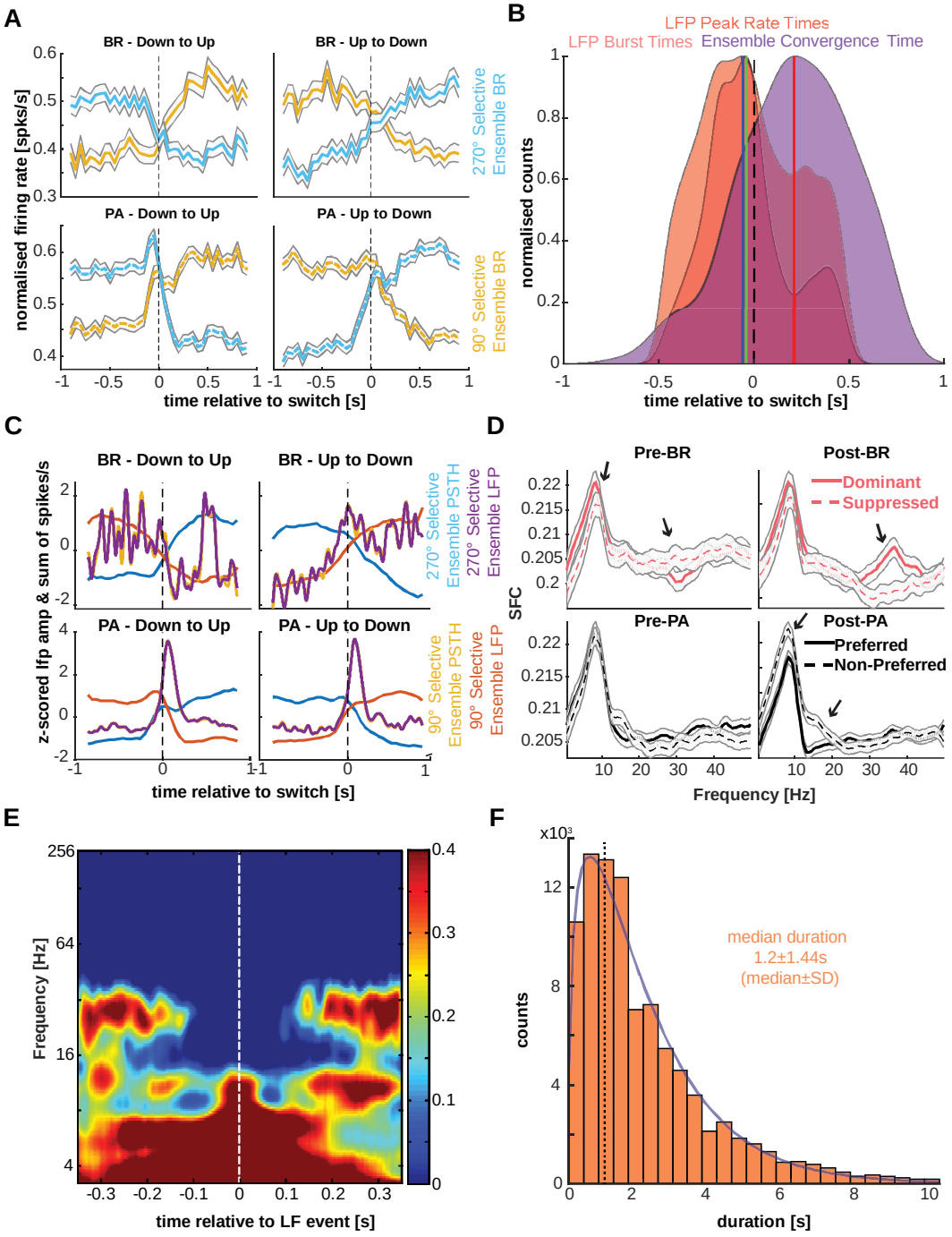


Figure 6| Ensemble spiking activity-LFP relationship and resting state analysis.

A. Top row: Average (across all collected transitions) population PSTHs of simultaneously recorded feature specific neuronal ensembles during BR. For each transition, we summed spiking activity across all selective single neurons belonging to each competing ensemble (upward and downward selective, blue and yellow, respectively). Bottom row: same for PA.

B. Distribution of the truncated (-500 to 250ms around the OKN transition) low-frequency burst times (pink), event-rate peak times (red) and the time of convergence of the PSTHs of each competing ensemble for each transition estimated by LOWESS smoothing and interpolation. Low-frequency activation preceded a change in the content of consciousness from neuronal populations (-60 ± 222ms for LFP event/s/transition and -112 ± 190ms vs. 77±76ms median±SD, convergence of spiking; $p<10^{-74}$ for both comparisons).

C. Although the PSTHs of the two competing ensembles for both BR and PA show a clear divergence in their activity after the transitions, the low-frequency instantaneous amplitudes of the same ensembles show no significant differences in their pattern, pointing to a non-specific and global role of these LFPs in driving perceptual transitions.

D. After a BR transition, spiking of the dominant population was significantly more coherent with the beta LFP band compared to suppression (black arrows point to frequency bins with statistically significant differences between dominant and suppressed $p<0.05$). SFC during pre-switch BR periods when low-frequency transient bursts are more prevalent did not exhibit similar differences in the beta band. These effects in beta band were absent in physical transitions where SFC for a dominant preferred stimulus was significantly reduced in a lower frequency range.

E. Grand-average (n=480), low frequency burst-triggered spectrogram during resting state showing cortical state fluctuations between low frequency and beta activity in the absence of structured sensory input.

F. Periods of sustained beta activation (1.2 ± 1.44s, median±SD) follow a gamma distribution (BIC (Bayesian information criterion (*30*)) = $2.3x10^5$ for a gamma distribution vs BIC = $2.9x10^5$ for an exponential) during resting state (activity periods longer than 10s were discarded to maintain equivalence with the experimental trial durations). The median duration is remarkably similar to the psychophysical gamma distribution (1.54 ± 1.28s, median±SD).

Discussion

Contrary to the notion that PFC activity reflects the consequences of conscious perception (*18*, *19*, *21*, *26*, *31–33*), our results reveal that perisynaptic prefrontal state fluctuations precede spontaneous changes in the content of consciousness, with the latter reflected in feature specific prefrontal ensembles, in the absence of any voluntary report requirements. Therefore, the PFC appears to have a centrally important role in the emergence of conscious perception. Previous electrophysiological studies in the PFC have revealed representations of the content of consciousness using exogenous perceptual manipulations (*23*, *34*) and preparatory activity before spontaneous perceptual changes, that could not however be dissociated from signals related to voluntary motor reports (*24*, *35*). Furthermore, studies using BOLD fMRI that indirectly reflects neural activity disagree on the impact of voluntary reports on PFC activity during spontaneous perceptual transitions (*12*, *19*). Here we used a no-report spontaneous perceptual transition paradigm coupled with multielectrode recordings of local field activity and feature-specific neuronal ensembles that was instrumental for revealing the mechanistic details of conscious awareness emergence.

Our findings reveal a gating-like mechanism where intrinsically generated cortical state fluctuations between low-frequency transients and periods of sustained beta-bursts control the access of competing perceptual representations to consciousness. The spatiotemporal build-up of 1-9 Hz bursts before spontaneous perceptual transitions is reminiscent of the Readiness Potential or *Bereitschaftspotential*; i.e., a steady accumulation of activity in the Anterior Cingulate Cortex (ACC) and Supplementary Motor Area (SMA) preceding the awareness of the volition to initiate a report (*36–39*). This motor-related process is reflected in a spiking activity build-up before the voluntary motor report of a perceptual transition in BR (*24*). Our results show that both linear and non-linear increases in the instantaneous amplitude and spatial-spread of low-frequency prefrontal bursts respectively precede spontaneous transitions

in the content of consciousness without voluntary behavioural reports.

Periods of beta burst suppression are thought to reflect a decrease in endogenous cortical processing, since beta is suppressed during cognitive processes like attention, decision-making and movement-planning (*40–43*). Beta activity could therefore reflect an intrinsic mode of cortical operation that shields ongoing behavioural and processing states ("status-quo") from interference and distractors (*40*, *41*). As a result, transient decreases in cortical beta activity could increase sensory information relay (*44*, *45*, *46*, *47*) providing a mechanism for controlling bottom-up sensory processing through top-down knowledge (*48*). Similar desynchronised brain states could mediate low-level awareness (*49*). Indeed, we found that a dissolution of beta-coherent dominant ensembles from low-frequency transient bursts precedes intrinsically generated changes between the two perceptual states. This is reminiscent of rule-selective prefrontal ensembles that are coherent in the beta band with low-frequency activity inhibiting a rule that is about to be deselected (*50*). This suggests that the underlying prefrontal mechanism for the emergence of conscious awareness and cognitive control might be the same. We propose that the intrinsically-generated prefrontal beta activity could reflect the neural correlate of the prefrontal threshold mechanism that has been long hypothesised to control access to consciousness (*34*, *51–54*). Disruption of intrinsic prefrontal beta from transient depolarisation events could provide temporal windows for a reorganisation in the discharge activity of neuronal ensembles that encode the competing representations and consequently increase the likelihood for a perceptual transition to happen. Indeed, beta-suppression occurred earlier than the flip in the population discharge activity encoding the content of consciousness (SI Fig. 10). This top-down mechanism of perceptual reorganisation is fundamentally different from bottom-up mechanisms proposing that competition between monocular neurons in primary visual cortex (V1) resolves BR (*7*, *8*). Neuronal activity in V1 is indeed only weakly modulated during BR in both monocular and binocular neurons (*8*),

while BOLD modulation of V1 is detected in superficial layers suggesting feedback from higher cortical areas (*55*). Furthermore, optical imaging signals in V1 during BR can also be observed during anaesthesia (*56*), suggesting that V1 activity is not sufficient for conscious visual perception.

Finally, spontaneous cortical activity can attain various states during wakefulness, and mimic sensory-driven activity (*57–59*). We observed suppression of beta bursts by low-frequency transients also during periods of resting state. This suggests that the source of spontaneous transitions in the content of consciousness is the operation of waking state fluctuations in the PFC. Taken together, our results reveal a pivotal role of prefrontal state fluctuations in the emergence of conscious awareness.

Methods

Electrophysiological recordings.

We performed extracellular electrophysiological recordings in the inferior convexity of the lateral PFC of 2 awake adult, male rhesus macaques (*Macaca mulatta*) using chronically implanted Utah microelectrode arrays (*60*) (Blackrock Microsystems, Salt Lake City, Utah USA). We implanted the arrays 1 - 2 millimetres anterior to the bank of the arcuate sulcus and below the ventral bank of the principal sulcus, thus covering a large part of the inferior convexity in the ventrolateral PFC, where neurons selective for direction of motion have been previously found (*61*, *62*). The arrays were 4x4mm wide, with a 10 by 10 electrode configuration and inter-electrode distance of 400μm. Electrodes were 1mm long therefore recording from the middle cortical layers. The monkeys were implanted with form-specific titanium head posts on the cranium after modelling the skull based on an anatomical MRI scan acquired in a vertical 7T scanner with a 60cm diameter bore (Biospec 47/40c; Bruker Medical,

Ettlingen, Germany). All experiments were approved by the local authorities (Regierungspräsidium, protocol KY6/12 granted to TIP as the principal investigator) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

Data acquisition, spike sorting and local field potentials.

Broadband neural signals (0.0001–30 kHz) were recorded using Neural Signal Processors (NSPs) (Blackrock Microsystems). Signals from the Utah array were digitised, amplified, and then routed to the NSPs for acquisition. For the offline detection of action potentials, broadband data were filtered between 0.6 and 3 kHz using a second-order Butterworth filter (the filter was chosen such that it allowed a flat response in the passband while contributing the least phase distortion due to its low order, yet having an acceptable attenuation in the stop band, i.e. a roll-off starting at -20dB). The amplitude for spike detection was set to five times the median absolute deviation (MAD) (*63*). Spikes were rejected if they occurred within 0.5 ms of each other or if they were larger than 50 times the MAD. All of the collected spikes were aligned to the minimum. Automatic clustering to detect putative single neurons was performed by a Split and Merge Expectation-Maximisation (SMEM) algorithm that fits a mixture of Gaussians to the spike feature data which consisted of the first three principal components (*64*) (Klustakwik). The clusters were finalised manually using a cut-and-merge software (*65*) (Klusters). For the analysis of perisynaptic LFP activity, the broadband signal was decimated to 500 Hz sampling rate using a Type I Chebyshev Filter, preserving frequency components up to 200 Hz.

Visual stimulation and experimental paradigm.

Visual stimuli were generated by in-house software written in C/Tcl and used OpenGL implementation. Stimuli were displayed using a dedicated graphics workstation (TDZ 2000;

Intergraph Systems, Huntsville, AL, USA) with a resolution of 1,280 × 1,024 and a 60 Hz refresh rate. An industrial PC with one Pentium CPU (Advantech) running the QNX real-time operating system (QNX Software Systems) controlled the timing of stimulus presentation, and the digital pulses to the electrophysiological data acquisition system. Eye movements were captured using an IR camera at 1kHz sampling rate using the software iView (SensoriMotoric Instruments GmBH, Germany). They were monitored online and stored for offline analysis using both the QNX-based acquisition system and the Blackrock data acquisition system. We were able to capture reliably the eye movements of the animals by positioning the IR camera in front of a cold mirror stereoscope.

Initially, the two monkeys (A11 and H07) were trained to fixate on a red square of size 0.2° of visual angle about 45cm away from the monitors that could be viewed through the stereoscope. This dot was first presented in one eye (the location of the red fixation square was adjusted to the single eye vergence of each individual monkey) and the eye-position was centred using a self-constructed linear offset amplifier. While the monkey was fixating the dot was removed and immediately presented in the other eye. Over multiple presentations, the offset between the two eyes was averaged to provide a horizontal correction factor to allow the two dots to be perfectly fused within the resolution limitations of the recording device (1/100th of a degree). The monkeys were trained to maintain fixation within a window of 2° of visual angle during initiation. After 300ms of fixation, a moving grating of size 8°, moving in the vertical direction (90° or 270°) at a speed of 12° (monkey H) and 13° (monkey A) per second, with a spatial frequency of 0.5 cycles/degree of visual angle and at 100% contrast was presented for 1000-2000ms. This marked the first monocular stimulus epoch in both conditions, viz. Binocular Rivalry (BR) and Physical Alternation (PA). At the end of 1000-2000ms, the second stimulus with the same properties as above but moving in the opposite direction was presented to the other eye. In the BR trials, this marked the “Flash Suppression” phase. These two

competing stimuli were allowed to rival with each other for a period of 6000-8000ms. In the PA trials, switches in the percept were mimicked by alternatively removing one stimulus based on the mean dominance time computed from the Gamma Distributions (tailored to each monkey's performance and statistics) acquired during multiple training sessions, and adjusted to be closer to a mean of 2000ms. Free viewing within the 8° window elicited the Optokinetic Nystagmus (OKN) reflex concomitant to the perceived direction of motion which served in lieu of a voluntary report, fulfilling the criterion of a "no-report paradigm". The monkeys were given a liquid reward (either water or juice) at the end of the trial, if their OKN successfully remained within the specified viewing window during the entire duration of the trial. Every successful trial was followed by a 2000-2500ms inter-trial period.

Detection of spontaneous transitions

The recorded eye-movement signal in the Y-coordinate was first low-pass filtered using a 3rd order Butterworth Filter below 20 Hz to remove involuntary jitter-induced high-frequency noise. A custom GUI written in MATLAB allowed us to manually identify the end of a dominance period and the beginning of the subsequent one. Manual marking (performed by two authors, AD and VK) was necessitated due to the large variability in the shapes that comprised the OKN complex. These events were based on the change in the slope of the slow-phase of the OKN. Such spontaneous switches were identified by the difference in the end of a dominance and the beginning of the next one; specifically, if this difference was less than 250ms (a fast switch). A "clean" transition was designated if the previous dominance and the subsequent one lasted for at least 500ms without being broken. Analogous to subjective reports, we aligned the LFP and the spiking activity on the beginning of the subsequent dominance period. This was performed in the same way for both BR and PA trials

Treatment of the LFP data

Firstly, the decimated LFP signal (0.1-500 Hz) around the OKN transitions was decomposed into a time-frequency representation using a Continuous Wavelet Transform (CWT, MATLAB 2016b) with a Morse wavelet of 7 cycles. This allowed us to resolve 169 frequencies from 0.5 to 256 Hz (500 Hz sampling rate) while preserving the full temporal resolution. The CWT for each channel in each transition (BR, PA, PM and RT) was first z-scored in the frequency domain to visualise the relative changes in power and then pooled across all channels and averaged. To visualise the differences between spontaneous transitions, piecemeals and randomly-triggered periods, the latter two spectrograms were subtracted from the former, respectively.

To understand the evolution of the LFP activity, we first filtered the broadband LFP trace into two constituent bands that were identified to be modulated during the task from the time-frequency analysis, i.e., the low-frequency (1-9 Hz) and the beta band (20-40 Hz). We used a 4th and 8th order Chebyshev Type I filter respectively, with a maximum passband ripple of 0.001dB. To get the instantaneous amplitude in time, we transformed the signal into the Hilbert space and then computed the absolute value. Bursting events were detected at each transition in each channel using a threshold which was 4 times the standard deviation of the noise modelled as a Gaussian distribution. The minimum duration of each event to be detected was set as one full cycle of the highest frequency in that band, i.e. 111ms for the 1-9 Hz band and 25ms for the 20-40 Hz band (*66*). The event-rate in time was computed as a quasi-PSTH by turning the detected bursts into a binary spike-train and smoothed with a Gaussian kernel of width 25ms, and then averaged across all channels. The burst rate was computed as the sum of low-frequency bursts normalised by the number of transitions and channels (bursts/transition/channel). To compute the build-up in the low-frequency activity, the amplitude at each detected time-point was averaged first across all channels for a given

transition, and then averaged across all transitions. A line was then fit to this mean scatter-plot using the CurveFit Toolbox in MATLAB.

Construction of direction of motion specific neural ensembles

Single neuron selectivity was assessed during perceptual transition periods of binocular rivalry (perceptual switches) and physical alternation (stimulus switches). During binocular rivalry trials, these periods were selected according to the following criteria: 1. Perceptual dominance (judged from the OKN signal) must be maintained for at least 1000 milliseconds post a perceptual switch 2. A preceding perceptual dominance for the competing stimulus must be maintained for 1000 milliseconds, and finally 3. The delay between the end and the beginning of the two dominance phases was not more than 250 milliseconds. For physical alternations, we selected trials, wherein a stimulus was presented for at least 1000 milliseconds before and after a stimulus switch. The spiking activity was triggered at the beginning of a forward dominance (BR) and stimulus change (PA).

Selectivity was assessed by comparing the distributions of the total number of spike counts across trials where the upward drifting grating was perceived, post (0 to 1000 ms) or pre switch (-1000 to 0 ms), with trials where a downward drifting grating was perceived. We used a Wilcoxon rank sum test and all neurons where $p<0.05$ were considered as selective. For a given transition, spikes were binned in 50ms bins for each selective neuron, and the resultant spike-count histograms were summed across the neurons that make up each selective ensemble to represent a population vector.

To analyse the crossover times between the two competing populations, we computed the trend in these normalised direction-selective population sum PSTH activity for every transition in a 900ms window around the time of the marked smooth pursuit OKN change [-900 to 900] by smoothing the raw ensemble population vectors for the two competing populations using a LOWESS filter. Next, we detected each intersection between these two

given vectors using standard interpolation. Where multiple intersection points were detected, only that point was considered which was followed by divergences for a minimum of 200ms before and after the intersection point, denoting distinct encoding of the currently active percept.

Spike-field Coherence

The spike-field coherence (SFC) was computed between the spiking activity of selective ensembles for each transition, and the global LFP for that particular transition averaged over all electrodes. A rate adjustment and a finite-size correction was applied before computing the SFC via a multi-taper method (*67*) (Chronux Toolbox).

Treatment of resting-state activity

LFPs from two continuously-recorded resting state sessions on days when no task-recording was performed, were decimated to 500 Hz as mentioned above. In each channel, the beta bursts were detected using the previously-mentioned LFP event-detection algorithm. The mean of the inter-event interval was used as a threshold to decide which collection of events constituted a phase of sustained activity. These epochs were collected across all channels and pooled across the two monkeys. Both a gamma and an exponential distribution were fit to the observations, with the gamma distribution clearly revealing lower AIC and BIC measures, thereby pointing to this distribution being a better fit than the exponential.

Statistical methods

All statistical comparisons were performed using a Wilcoxon ranksum test (*68*) due to the unknown nature of the underlying distribution from which the data originated. Distributions were fit using the MATLAB statistical toolbox using a Maximum-Likelihood-Estimate

method. For model comparisons, the allFitDist.m toolbox was used that also generated metrics for appropriate model selection. For non-parametric fitting of distributions with widely different sample numbers, the kernel density estimate method implemented in the MATLAB Statistics Toolbox was used to generate the best-fit function, which was then normalised for visualisation.

SI Section - Controlling for failed and non-occurrence of switches (SI Figure 8)

Is there a difference in the type and magnitude of low-frequency activity leading to a perceptual transition as compared to when it never happens? To this end, we randomly triggered the LFP activity in observation periods where no such spontaneous transitions occurred, i.e. during periods of sustained predominance. Subtracting the time-frequency decomposition of these randomly triggered (RT) periods from that of BR preserved the pattern observed in the latter, suggesting that weak, low-frequency activity occurs as baseline noise, which only leads to a perceptual change when it is spatiotemporally ramped up in a structured manner (Figure 4B vs Figure 2A). Indeed, we computed a mean rate of 0.015±0.00005 (n = 55026 after considering 100 iterations, i.e. only 550.26 bursts per iteration) bursts per transition; an order of magnitude lower than the corresponding periods during BR (0.17±0.002, n = 9667, bursts/transition, SI Figure 8D). Furthermore, the proportion of sites that displayed low-frequency bursting activity across all BR periods was 100%, compared to only 51% in RT periods. These results further confirm that the low-frequency burst-rate, build-up, and a larger spatial spread of activation is necessary to drive spontaneous transitions.

Furthermore, the low-frequency burst-rate was higher before a clear spontaneous transition compared to the period before transition to a piecemeal percept (0.17±0.002, n = 9667, pre-BR, vs. 0.14±0.004, pre-PM, n = 2486, bursts/transition; $p<10^{-25}$) (SI Figure 8A), with the low-frequency peak rate occurring after the transition to piecemeal (SI Figure 8B).

Furthermore, the burst rate was significantly higher after the transition to a piecemeal period (0.16±0.004, PM, n = 3531, vs 0.14±0.004 pre-PM, n = 2486, bursts/transition; $p<10^{-5}$), while the anti-correlation between low-frequency and beta was significant but weaker compared to clear spontaneous transitions (r = -0.009, $p<10^{-137}$ vs r = -0.05, $p<10^{-295}$, $p < 10^{-14}$, Figure 3D and SI Figure 8C).

SI Section – PA Switches aligned to the TTL pulses (SI Figure 6)

The PA switches used for analysis in the manuscript were aligned to the manual marks. In addition, we also aligned them to the experimental TTL pulses. The results do not change either qualitatively or quantitatively.

SI Section – Automatic detection of switches after Aleshin et al 2019 (SI Figure 11)

To compare our manually marked switches to those detected automatically, we adapted the Cumulative Smooth Pursuit algorithm designed by Aleshin et al (2019) (*69*). From visual inspection and comparison, manual marking was more robust than the automatic detection, which was contaminated with multiple false positives and negatives. In light of this performance, we continued to use data aligned to the manual marks.

References

1. L. A. Necker, LXI.*Observations on some remarkable optical phænomena seen in Switzerland; and on an optical phænomenon which occurs on viewing a figure of a crystal or geometrical solid*. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. **1**, 329–337 (1832).

2. E. G. Boring, A new ambiguous figure. *Am. J. Psychol.* **42**, 444 (1930).

3. R. Blake, N. K. Logothetis, Visual competition. *Nat. Rev. Neurosci.* **3**, 13–21 (2002).

4. C. W. G. Clifford, Binocular rivalry. *Curr. Biol.* **19**, R1022–3 (2009).

5. R. Blake, A Primer on Binocular Rivalry, Including Current Controversies. *Brain and Mind*. **2**, 5–38 (2001).

6. C. Moutard, S. Dehaene, R. Malach, Spontaneous Fluctuations and Non-linear Ignitions: Two Dynamic Faces of Cortical Recurrent Loops. *Neuron*. **88**, 194–206 (2015).

7. R. Blake, A neural theory of binocular rivalry. *Psychol. Rev.* **96**, 145–167 (1989).

8. D. A. Leopold, N. K. Logothetis, Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*. **379**, 549–553 (1996).

9. N. K. Logothetis, D. A. Leopold, D. L. Sheinberg, What is rivalling during binocular rivalry? *Nature*. **380**, 621–624 (1996).

10. P. Sterzer, A. Kleinschmidt, G. Rees, The neural bases of multistable perception. *Trends Cogn. Sci. (Regul. Ed.)*. **13**, 310–318 (2009).

11. V. A. Weilnhammer, K. Ludwig, G. Hesselmann, P. Sterzer, Frontoparietal cortex mediates perceptual transitions in bistable perception. *J. Neurosci.* **33**, 16009–16015 (2013).

12. E. D. Lumer, K. J. Friston, G. Rees, Neural correlates of perceptual rivalry in the human brain. *Science*. **280**, 1930–1934 (1998).

13. T. Knapen, J. Brascamp, J. Pearson, R. van Ee, R. Blake, The role of frontal and parietal brain areas in bistable perception. *J. Neurosci.* **31**, 10293–10301 (2011).

14. F. Tong, M. Meng, R. Blake, Neural bases of binocular rivalry. *Trends Cogn. Sci. (Regul. Ed.)*. **10**, 502–511 (2006).

15. T. A. de Graaf, M. C. de Jong, R. Goebel, R. van Ee, A. T. Sack, On the functional relevance of frontal cortex for passive and voluntarily controlled bistable vision. *Cereb. Cortex*. **21**, 2322–2331 (2011).

16. J. F. Hipp, A. K. Engel, M. Siegel, Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*. **69**, 387–396 (2011).

17. B. Odegaard, R. T. Knight, H. Lau, Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* **37**, 9593–9602 (2017).

18. M. Boly *et al.*, Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? clinical and neuroimaging evidence. *J. Neurosci.* **37**, 9603–9613 (2017).

19. S. Frässle, J. Sommer, A. Jansen, M. Naber, W. Einhäuser, Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* **34**, 1738–1747 (2014).

20. M. C. de Jong *et al.*, Intracranial recordings of occipital cortex responses to illusory visual events. *J. Neurosci.* **36**, 6297–6311 (2016).

21. C. Koch, M. Massimini, M. Boly, G. Tononi, Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* **17**, 307–321 (2016).

22. J. Brascamp, R. Blake, T. Knapen, Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nat. Neurosci.* **18**, 1672–1678 (2015).

23. T. I. Panagiotaropoulos, G. Deco, V. Kapoor, N. K. Logothetis, Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron*. **74**, 924–935 (2012).

24. H. Gelbard-Sagiv, L. Mudrik, M. R. Hill, C. Koch, I. Fried, Human single neuron activity precedes emergence of conscious perception. *Nat. Commun.* **9**, 2057 (2018).

25. N. K. Logothetis, J. D. Schall, Binocular motion rivalry in macaque monkeys: eye dominance and tracking eye movements. *Vision Res.* **30**, 1409–1419 (1990).

26. N. Tsuchiya, M. Wilke, S. Frässle, V. A. F. Lamme, No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends Cogn. Sci. (Regul. Ed.)*. **19**, 757–770 (2015).

27. R. Fox, S. Todd, L. A. Bettinger, Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vision Res.* **15**, 849–853 (1975).

28. C. Chandrasekaran, A. Trubanova, S. Stillittano, A. Caplier, A. A. Ghazanfar, The natural statistics of audiovisual speech. *PLoS Comput. Biol.* **5**, e1000436 (2009).

29. M. A. A. van der Meer, A. D. Redish, Low and High Gamma Oscillations in Rat Ventral Striatum have Distinct Relationships to Behavior, Reward, and Spiking Activity on a Learned Spatial Decision Task. *Front Integr Neurosci*. **3**, 9 (2009).

30. G. Schwarz, Estimating the Dimension of a Model. *Ann. Statist.* **6**, 461–464 (1978).

31. V. A. F. Lamme, Towards a true neural stance on consciousness. *Trends Cogn. Sci. (Regul. Ed.)*. **10**, 494–501 (2006).

32. K. Sandberg, S. Frässle, M. Pitts, Future directions for identifying the neural correlates of consciousness. *Nat. Rev. Neurosci.* (2016), doi:10.1038/nrn.2016.104.

33. M. Boly *et al.*, Consciousness in humans and non-human animals: recent advances and future directions. *Front. Psychol.* **4**, 625 (2013).

34. B. van Vugt *et al.*, The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*. **360**, 537–542 (2018).

35. C. Libedinsky, M. Livingstone, Role of prefrontal cortex in conscious visual perception. *J. Neurosci.* **31**, 64–69 (2011).

36. C. H. M. Brunia, G. J. M. van Boxtel, K. B. E. Böcker, *Negative Slow Waves as Indices of Anticipation: The Bereitschaftspotential, the Contingent Negative Variation, and the Stimulus-Preceding Negativity* (Oxford University Press, 2011), *Oxford Handbooks Online*.

37. A. Schurger, Specific Relationship between the Shape of the Readiness Potential, Subjective Decision Time, and Waiting Time Predicted by an Accumulator Model with Temporally Autocorrelated Input Noise. *Eneuro*. **5** (2018), doi:10.1523/ENEURO.0302-17.2018.

38. H. H. Kornhuber, L. Deecke, [changes in the brain potential in voluntary movements and passive movements in man: readiness potential and reafferent potentials]. *Pflugers Arch. Gesamte Physiol. Menschen Tiere*. **284**, 1–17 (1965).

39. B. Libet, C. A. Gleason, E. W. Wright, D. K. Pearl, Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*. **106 (Pt 3)**, 623–642 (1983).

40. C. Tzagarakis, N. F. Ince, A. C. Leuthold, G. Pellizzer, Beta-band activity during motor planning reflects response uncertainty. *J. Neurosci.* **30**, 11270–11277 (2010).

41. W. J. Ray, H. W. Cole, EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*. **228**, 750–752 (1985).

42. J. Alayrangues, F. Torrecillos, A. Jahani, N. Malfait, Error-related modulations of the sensorimotor post-movement and foreperiod beta-band activities arise from distinct

neural substrates and do not reflect efferent signal processing. *Neuroimage*. **184**, 10–24 (2019).

43. V. Piai, A. Roelofs, J. Rommers, K. Dahlslätt, E. Maris, Withholding planned speech is reflected in synchronized beta-band oscillations. *Front. Hum. Neurosci.* **9**, 549 (2015).

44. F. David, E. Courtiol, N. Buonviso, N. Fourcaud-Trocmé, Competing mechanisms of gamma and beta oscillations in the olfactory bulb based on multimodal inhibition of mitral cells over a respiratory cycle. *Eneuro*. **2** (2015), doi:10.1523/ENEURO.0018-15.2015.

45. A. K. Engel, P. Fries, Beta-band oscillations--signalling the status quo? *Curr. Opin. Neurobiol.* **20**, 156–165 (2010).

46. T. I. Panagiotaropoulos, V. Kapoor, N. K. Logothetis, Desynchronization and rebound of beta oscillations during conscious and unconscious local neuronal processing in the macaque lateral prefrontal cortex. *Front. Psychol.* **4**, 603 (2013).

47. B. Spitzer, S. Haegens, Beyond the status quo: A role for beta oscillations in endogenous content (re)activation. *Eneuro*. **4** (2017), doi:10.1523/ENEURO.0170-17.2017.

48. E. K. Miller, M. Lundqvist, A. M. Bastos, Working Memory 2.0. *Neuron*. **100**, 463–475 (2018).

49. Y. Zerlaut, A. Destexhe, Enhanced Responsiveness and Low-Level Awareness in Stochastic Network States. *Neuron*. **94**, 1002–1009 (2017).

50. O. Jensen, M. Bonnefond, Prefrontal α- and β-band oscillations are involved in rule selection. *Trends Cogn. Sci. (Regul. Ed.)*. **17**, 10–12 (2013).

51. H. Lau, D. Rosenthal, Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci. (Regul. Ed.)*. **15**, 365–373 (2011).

52. S. Dehaene, L. Naccache, Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*. **79**, 1–37 (2001).

53. In the theatre of consciousness. Global Workspace Theory, a rigor...: Ingenta Connect, (available at https://www.ingentaconnect.com/content/imp/jcs/1997/00000004/00000004/776).

54. A. Del Cul, S. Dehaene, P. Reyes, E. Bravo, A. Slachevsky, Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*. **132**, 2531–2540 (2009).

55. Program Planner, (available at http://www.abstractsonline.com/pp8/index.html#!/4649/presentation/11577).

56. H. Xu *et al.*, Rivalry-Like Neural Activity in Primary Visual Cortex in Anesthetized Monkeys. *J. Neurosci.* **36**, 3231–3242 (2016).

57. L. Mazzucato, A. Fontanini, G. La Camera, Dynamics of multistable states during ongoing and evoked cortical activity. *J. Neurosci.* **35**, 8214–8231 (2015).

58. M. Tsodyks, T. Kenet, A. Grinvald, A. Arieli, Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*. **286**, 1943–1946 (1999).

59. M. J. McGinley *et al.*, Waking state: rapid variations modulate neural and behavioral responses. *Neuron*. **87**, 1143–1161 (2015).

60. E. M. Maynard, C. T. Nordhausen, R. A. Normann, The Utah intracortical Electrode Array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* **102**, 228–239 (1997).

61. C. R. Hussar, T. Pasternak, Flexibility of sensory representations in prefrontal cortex depends on cell type. *Neuron*. **64**, 730–743 (2009).

62. S. Safavi *et al.*, Nonmonotonic spatial structure of interneuronal correlations in prefrontal microcircuits. *Proc. Natl. Acad. Sci. USA*. **115**, E3539–E3548 (2018).

63. R. Q. Quiroga, Z. Nadasdy, Y. Ben-Shaul, Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* **16**, 1661–1687 (2004).

64. S. N. Kadir, D. F. M. Goodman, K. D. Harris, High-dimensional cluster analysis with the masked EM algorithm. *Neural Comput.* **26**, 2379–2394 (2014).

65. L. Hazan, M. Zugaro, G. Buzsáki, Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. *J. Neurosci. Methods*. **155**, 207–216 (2006).

66. N. K. Logothetis *et al.*, Hippocampal-cortical interaction during periods of subcortical silence. *Nature*. **491**, 547–553 (2012).

67. H. Bokil, P. Andrews, J. E. Kulkarni, S. Mehta, P. P. Mitra, Chronux: a platform for analyzing neural signals. *J. Neurosci. Methods*. **192**, 146–151 (2010).

68. J. L. Hodges, E. L. Lehmann, Estimates of location based on rank tests. *Ann. Math. Statist.* **34**, 598–611 (1963).

69. S. Aleshin, G. Ziman, I. Kovács, J. Braun, Perceptual reversals in binocular rivalry: Improved detection from OKN. *J. Vis.* **19**, 5 (2019).

Acknowledgements

We thank Dr. Yusuke Murayama and the other technical and animal care staff for excellent technical assistance, Prof. Nicho Hatsopoulos for help with the implantations of the Utah arrays, Prof. Stanislas Dehaene for his inputs and insights, and finally Mr. Akshat Jain for his assistance in writing code for the resting state analysis and preparing publishable quality figures.

Funding

Funding was provided by the Max Planck Society for this research project.

Author contributions

Conceptualisation: AD, VK, TIP (lead), NKL; Data curation: AD (lead), VK and JW; Formal analysis: AD (lead), VK, JW, LAF; Funding acquisition: NKL; Investigation: AD (equal), VK (equal), TIP (supporting); Methodology: AD (equal), VK (equal), JW & SS (supporting), TIP (equal); Project administration: TIP; Resources: JW, NKL (lead); Software: AD (lead), VK, JW, LAF & SS (supporting); Supervision: TIP; Visualisation: AD (lead), TIP (supporting); Writing - original draft: AD, TIP (lead); Writing - review & editing: AD, VK, LAF, SS, TIP (lead), NKL.

Competing interests

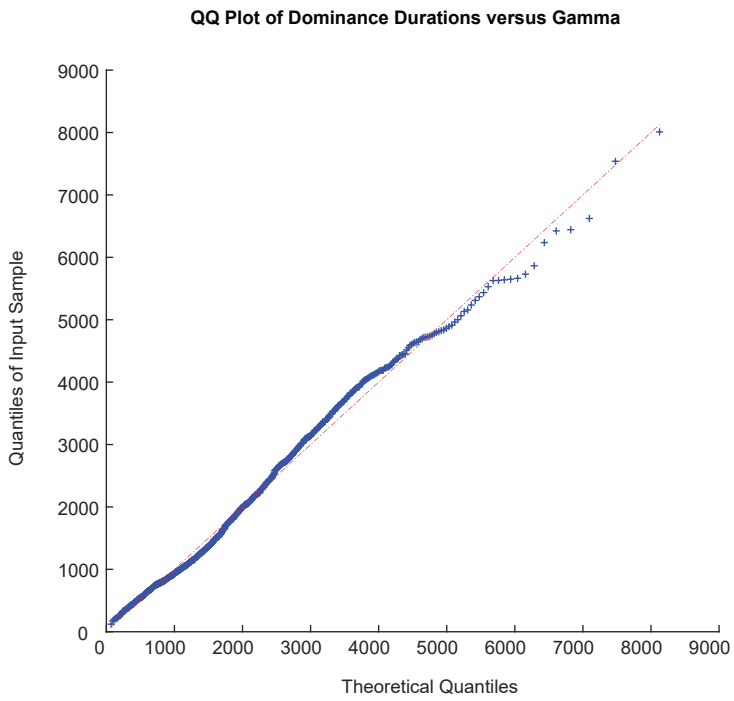
The authors declare no competing interests.

Data and materials availability

Raw data and analysis codes are available upon request.

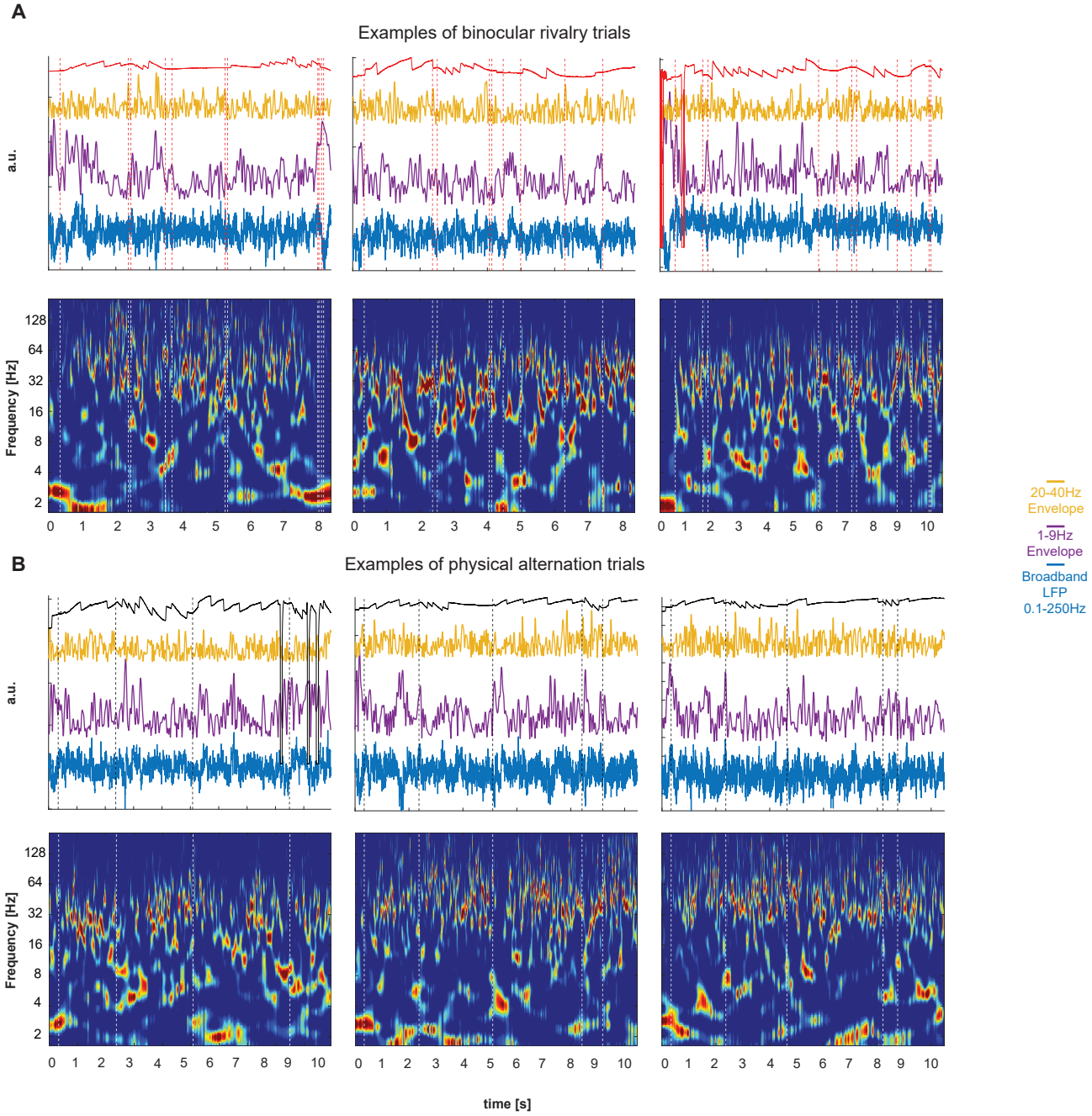
Supplementary materials

Methods, Supporting Information, SI Figures 1-9, References #61-69



SI Figure 1 | Quantile-Quantile diagnostics

We qualitatively confirmed that the gamma distribution was the best fit for the dominance durations using a Quantile-Quantile plot. The red dashed line shows the theoretical distribution whereas the blue crosses show the fidelity of the data to the theoretical distribution. Furthermore, a comparison of the Bayesian Information Criterion (BIC) yielded the lowest value for a gamma distribution (1.92x10^4) as compared to an exponential (1.96x10^4) or a logistic function (1.962x10^4), pointing to the gamma distribution as the best-fit model.

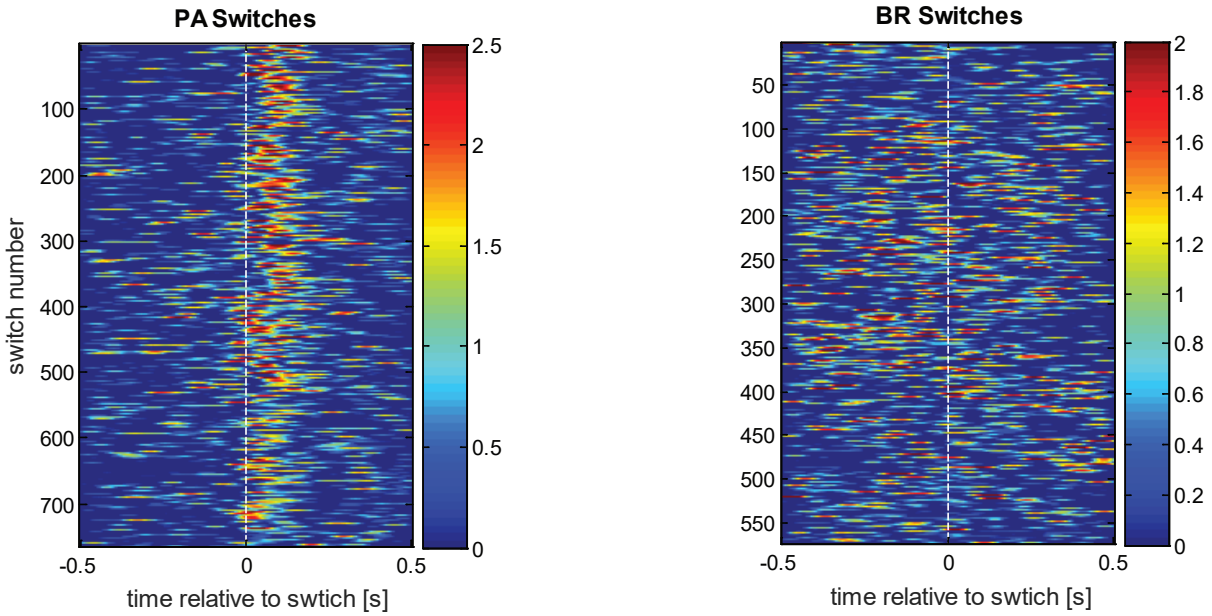


SI Figure 2 | Low Frequency and Beta Instataneous Amplitudes corresponding to a percept switch

A. The low frequency activity and the concomitant beta activity averaged across all the 96 channels for each trial in BR shows a large amplitude deflection and putative climbing activity just before a spontaneous percept switch (marked by the change in the polarity of the Optokinetic Nystagmus(OKN)), or during a piecemeal period before an upcoming transition

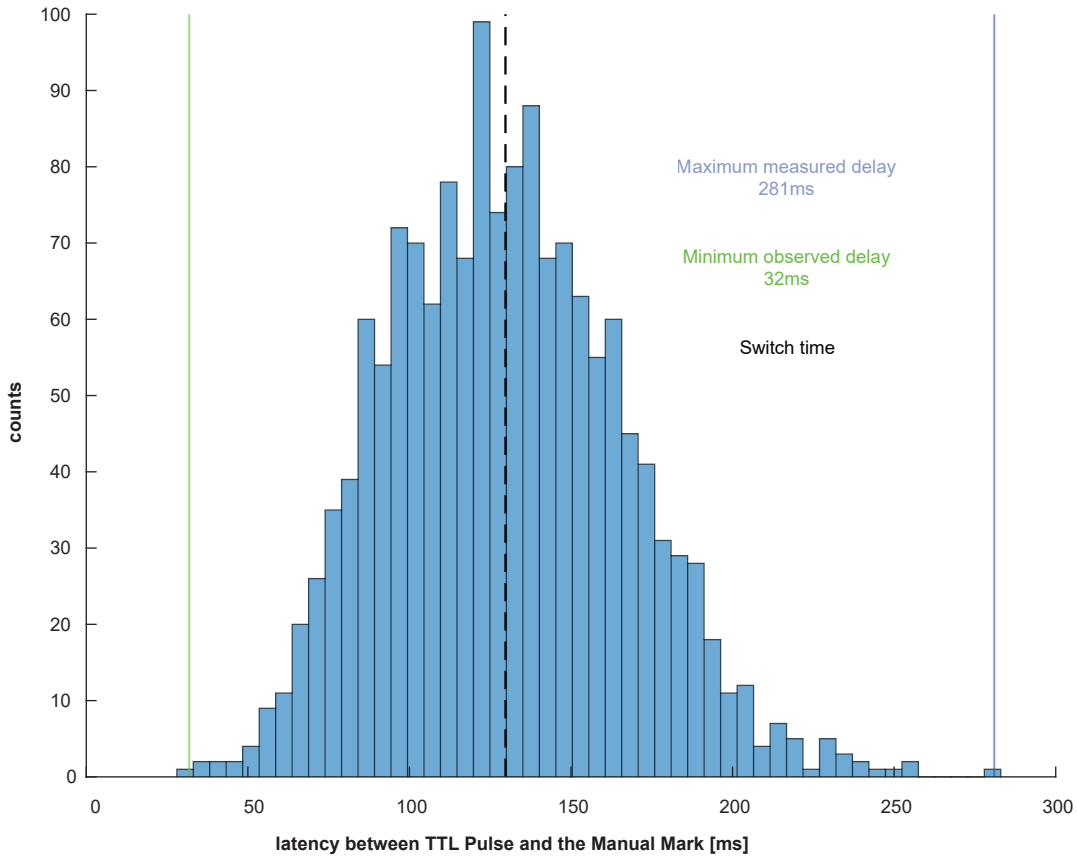
B. The low frequency activity and the concomitant beta activity in Physical Alternation(PA) trials shows the same kind of large amplitude deflection but this time at the point of or after the stimulus has been switched (manually marked transitions succeed TTL pulses). The amplitude deflection is even greater in magnitude because it is a visually evoked potential.

Random deviations from baseline are also observed; however, they are far fewer in number and lesser in magnitude



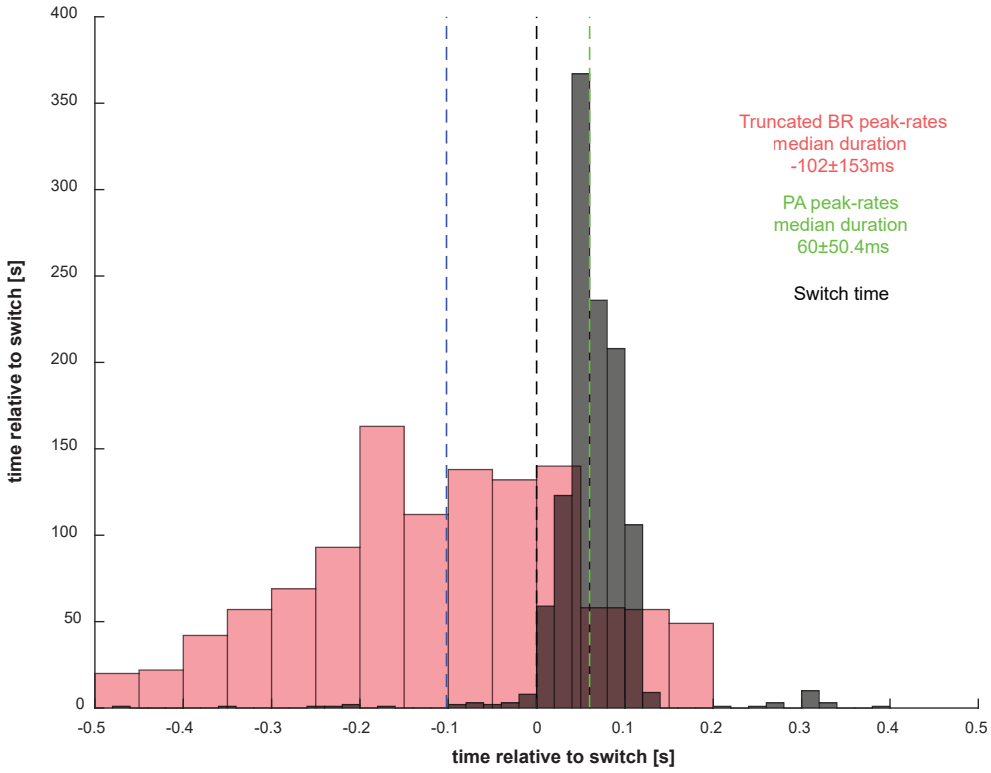
SI Figure 3 | Qualitative analysis of the temporal jitter in the 1-9Hz instantaneous amplitude

This figure depicts the 1-9Hz instantaneous amplitude extracted around clean switches (i.e. a minimum dominance of 500ms before and after a transition) in both the physical alternation (left) and binocular rivalry (right) conditions. While after the transition in PA, a strong and consistent visually-evoked potential is observed, the activity around a spontaneous switch is rather diffuse, yet shows a preponderance of low-frequency activity rising and concentrated in the pre-switch period.



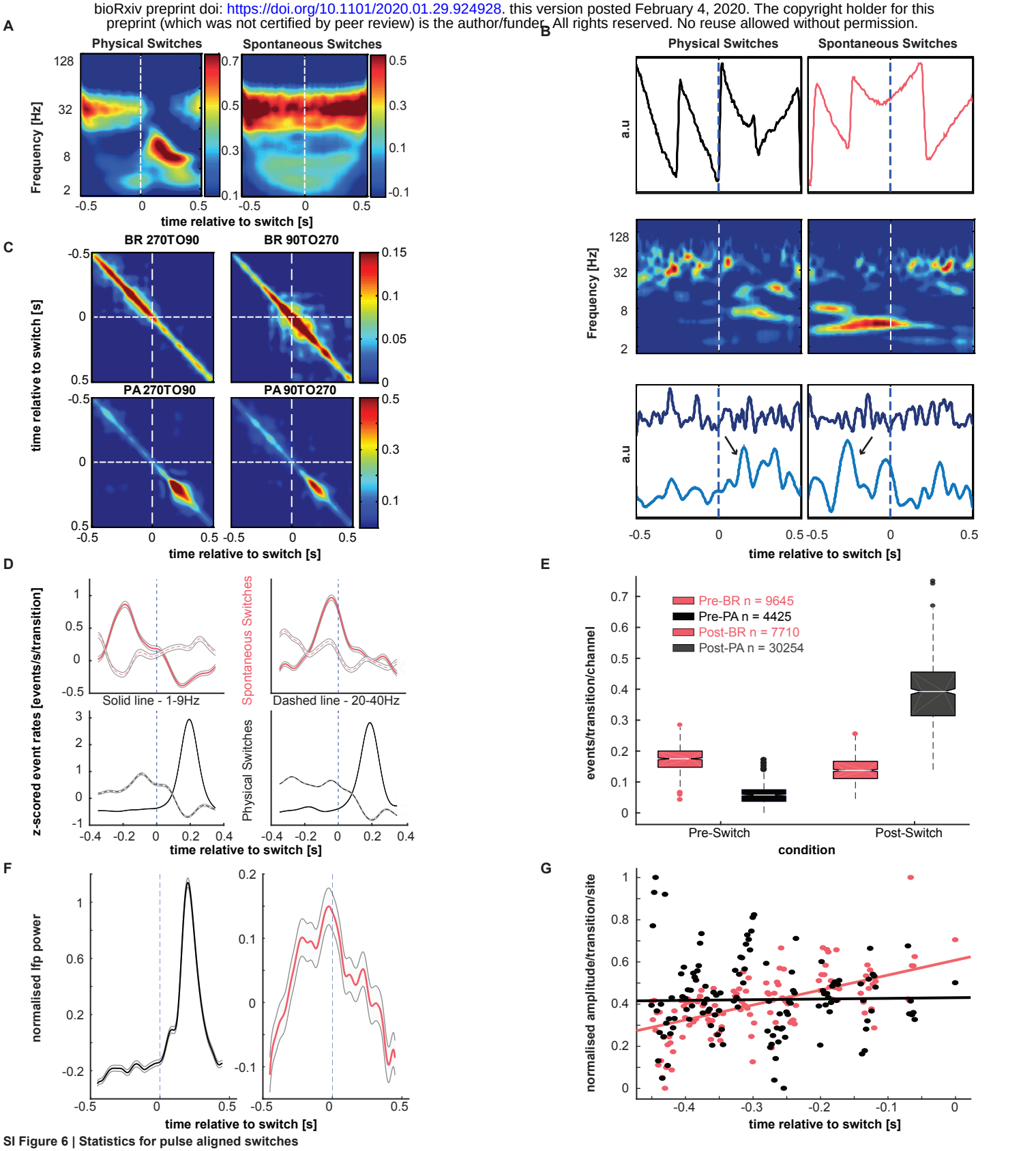
SI Figure 4 | Comparison of the difference between the TTL pulse and the manually marked change in Physical Alternation

We computed the latency between the onset of the subsequent stimulus and the succeeding change in the polarity of the induced OKN for all PA trials. We found an average latency of 129.4±36.5 ms (mean±SD), which indicates that the change in the eye-movement is induced within a very short interval. The minimum latency observed was 32ms whereas the maximum observed latency was 281ms.



SI Figure 5 | Truncated low-frequency peak rate comparison

Because events towards the end of the post-switch window in BR could signal an upcoming transition, we discarded these events that occurr after 150ms (timing of the end of the VEP in PA). The difference between the median timing of the peak rate in BR and the VEP in PA was further enhanced.



SI Figure 6 | Statistics for pulse aligned switches

A. Grand average time-frequency analysis of all physical (left) and spontaneous (right) perceptual transitions. Spectrograms are aligned (t=0) to the experimental TTL pulse for PA and to the OKN change for BR for periods of stable perceptual dominance before and after the switch.

B. Upper panel: OKN traces around a single physical (black) and spontaneous (red) transition. Middle panel: channel-averaged normalised spectrograms around the single transition event for the two conditions. Lower panel: Normalised instantaneous amplitudes of the two modulated frequency bands (i.e 1-9Hz and 20-40Hz) identified from the spectrograms. Low-frequency bursts occur after the physical switch but before the spontaneous transition.

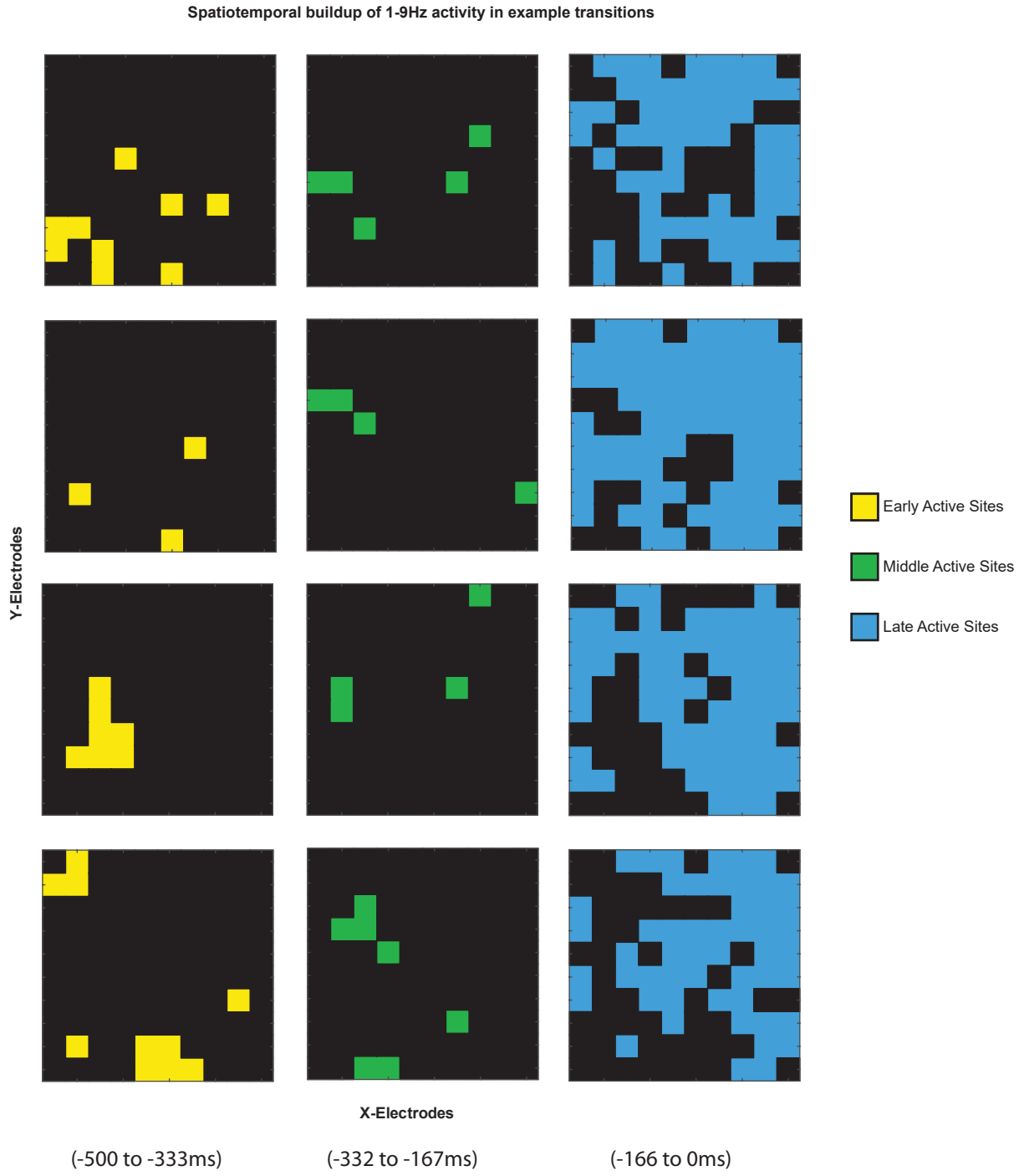
C. Differences in the onset of the low-frequency activity across physical and spontaneous transitions are reflected in the temporal auto-covariance of the low-frequency envelopes across the array recorded simultaneously for every transition type (270 to 90, left and 90 to 270, right). Most of the similarity is observed after a transition in PA (bottom panel) but before a transition in BR (upper panel).

D. Normalised (z-score) event rate in time (events/s/transition) during BR (red lines) and PA trials (white lines) for low-frequency (solid lines) and beta activity (dashed lines).

E. Burst rate per transition per channel for periods before and after transitions in PA (red) and BR (black). In order from left to right: Pre-BR (dark pink), Pre-PA (black), Post-BR (light pink), Post-PA (grey). The whiskers of the box-plots show the dispersion of the data. The white line depicts the median. Outliers are represented by appropriately coloured dots. A significantly higher number of low-frequency bursts occur before spontaneous, but after physical switches. Burst rate before a physical switch is very low, suggesting noise levels. This baseline burst-rate needs to be ramped up for a switch to occur. PA: (0.39±0.004, n = 30254, post-transition, vs. 0.06±0.001, pre-transition, n = 4425; p<10-188 median±SEM) BR: (0.18±0.002, n = 9645, pre-transition vs 0.14±0.002, n = 7710, post-transition, p<10-43 median±SEM).

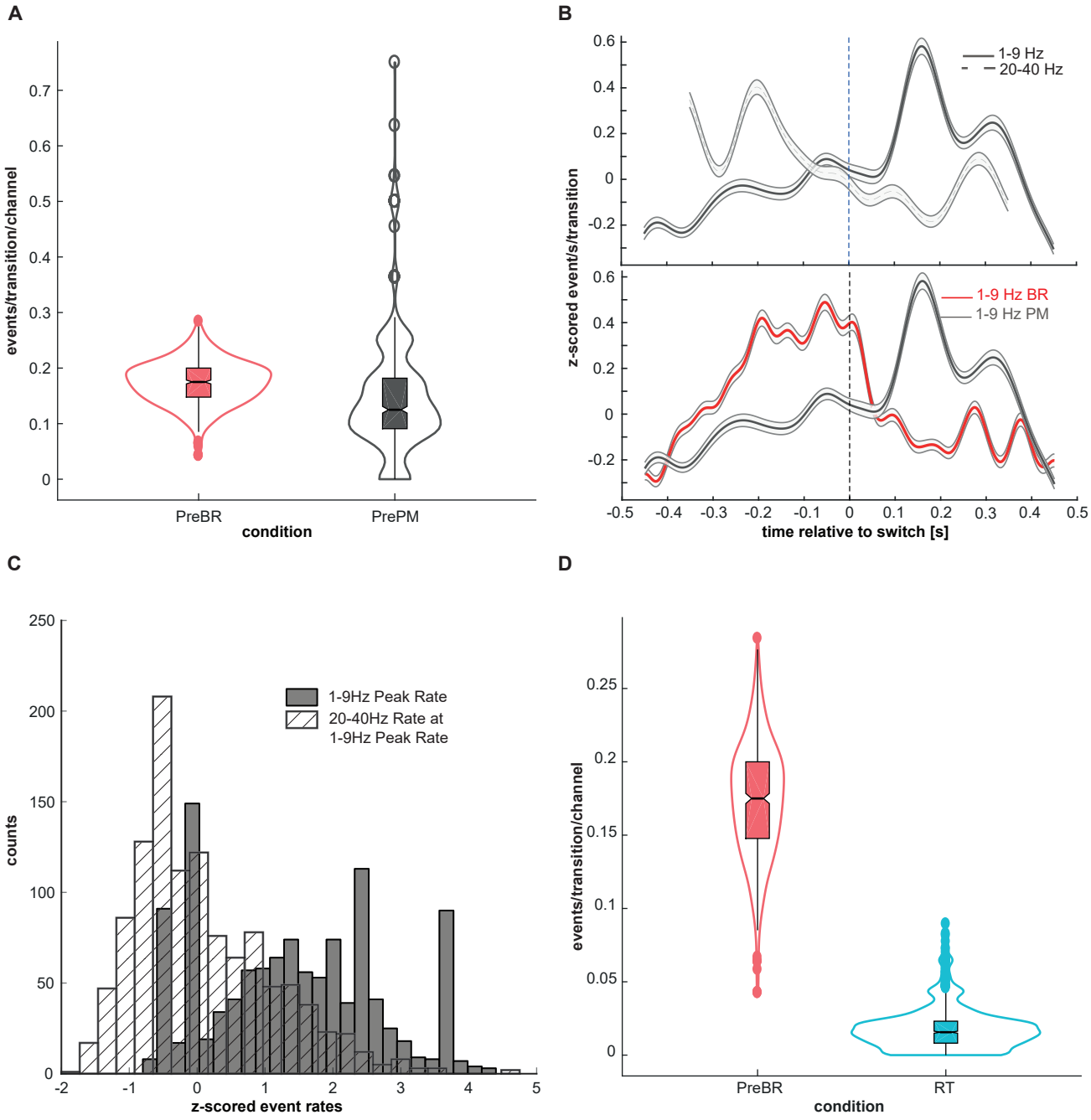
F. Low-frequency instantaneous amplitude shows a slow climbing activity before a perceptual transition (right) but not before a physical transition (left). Curves reflect an average across transitions of the channel-averaged activity for each transition

G. Average build-up of low-frequency activity in time is shown by fitting a line to the pooled and averaged amplitude at every time-point (red – BR, black – PA) across all measured neuronal sites. While before a spontaneous transition, the recorded PFC area ramps up its low-frequency activity in time, before a physical transition, it remains flat.



SI Figure 7 | Spatial buildup of low-frequency activity.

This figure shows the spatial buildup of 1-9Hz activity in 3 temporal windows viz. in an early window ([-500 to -333ms]), a middle window ([-334 to -166ms]) and a late window ([-167 to 0ms]) preceding four typical spontaneous switches. Progressively, more sites are activated approaching a switch.



SI Figure 8 | Controlling for failed and non-occurrence of switches

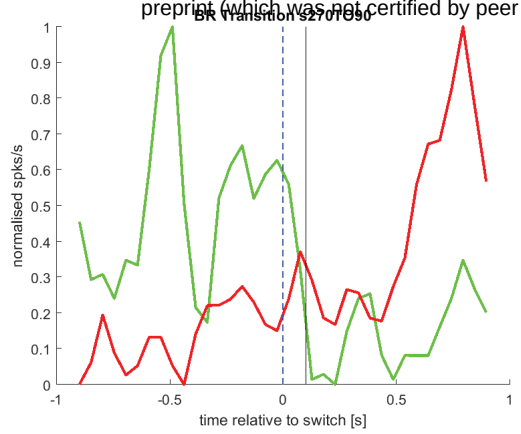
a. The low frequency burst rate per transition per channel before a spontaneous transition (Pre-BR, n= 9667, 0.17±0.002) when compared to the period before transition to the piecemeal percept (Pre-PM, n= 2486, 0.147±0.004) is higher, but not significantly (p=0.08).

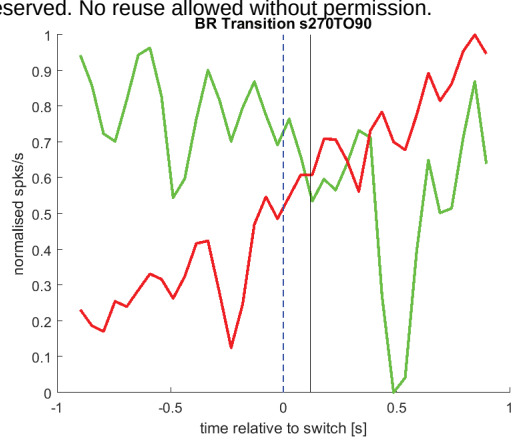
b. Top row panel: z-scored event rates in time (events/transition/s) before and after the transition to piecemeal percepts, the solid line being the low-frequency component and the dashed line being the beta-band activity. While before the transition to piecemeal, the beta dominates, signalling the active percept, after the transition the low-frequency inhibits the beta thus signalling an upcoming percept.

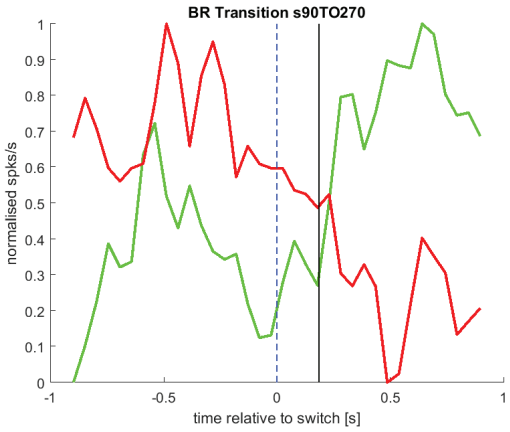
Bottom row panels: z-scored event rates in time (events/transition) for the low frequency activity for BR(red) and PM(grey). The low frequency activity burst rate peaks before transition to another clear dominance during BR (red) whereas it peaks after the transition to a piecemeal percept (grey).

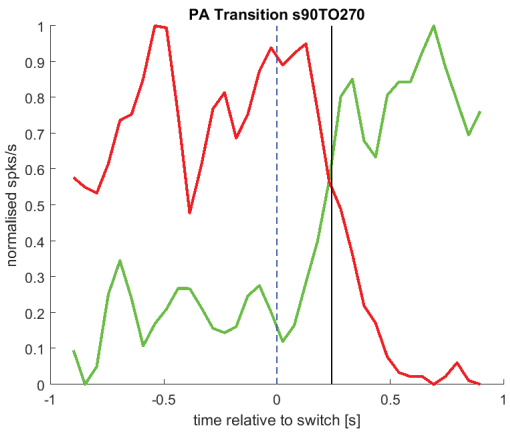
c. The distribution of the peak low-frequency rates vis a vis the rate of the beta activity at the timing of the low-frequency peak reveals no significant antagonism before a transition to a piecemeal (r = -0.0073, p=0.805) as compared to before clear spontaneous transitions, where a significant decoupling is observed (r = -0.08, p=0.0071).

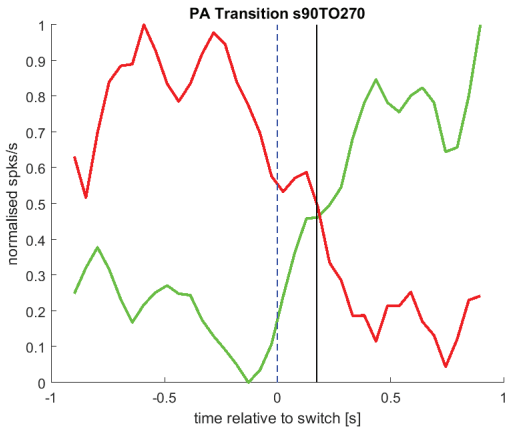
d. Low-frequency event rate per transition per channel for the two conditions, spontaneous switch and randomly triggered periods. The burst rate before a spontaneous switch (pre-BR, 0.17±0.002, n = 9667) is significantly higher than during randomly triggered epochs (RT, 0.018±0.00007, n = 55026, 100 iterations, 550.26 bursts per iteration).





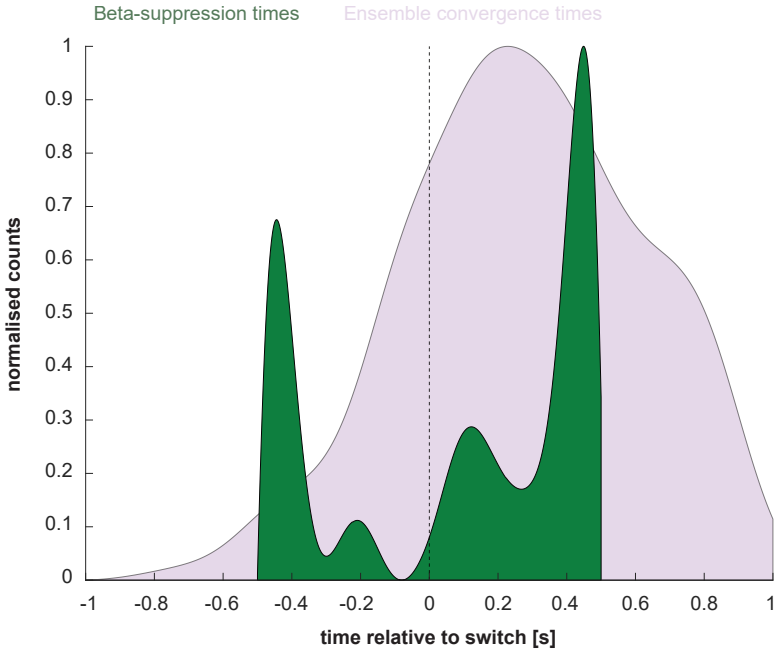






SI Figure 9 | Estimation of change in encoding in selective ensembles

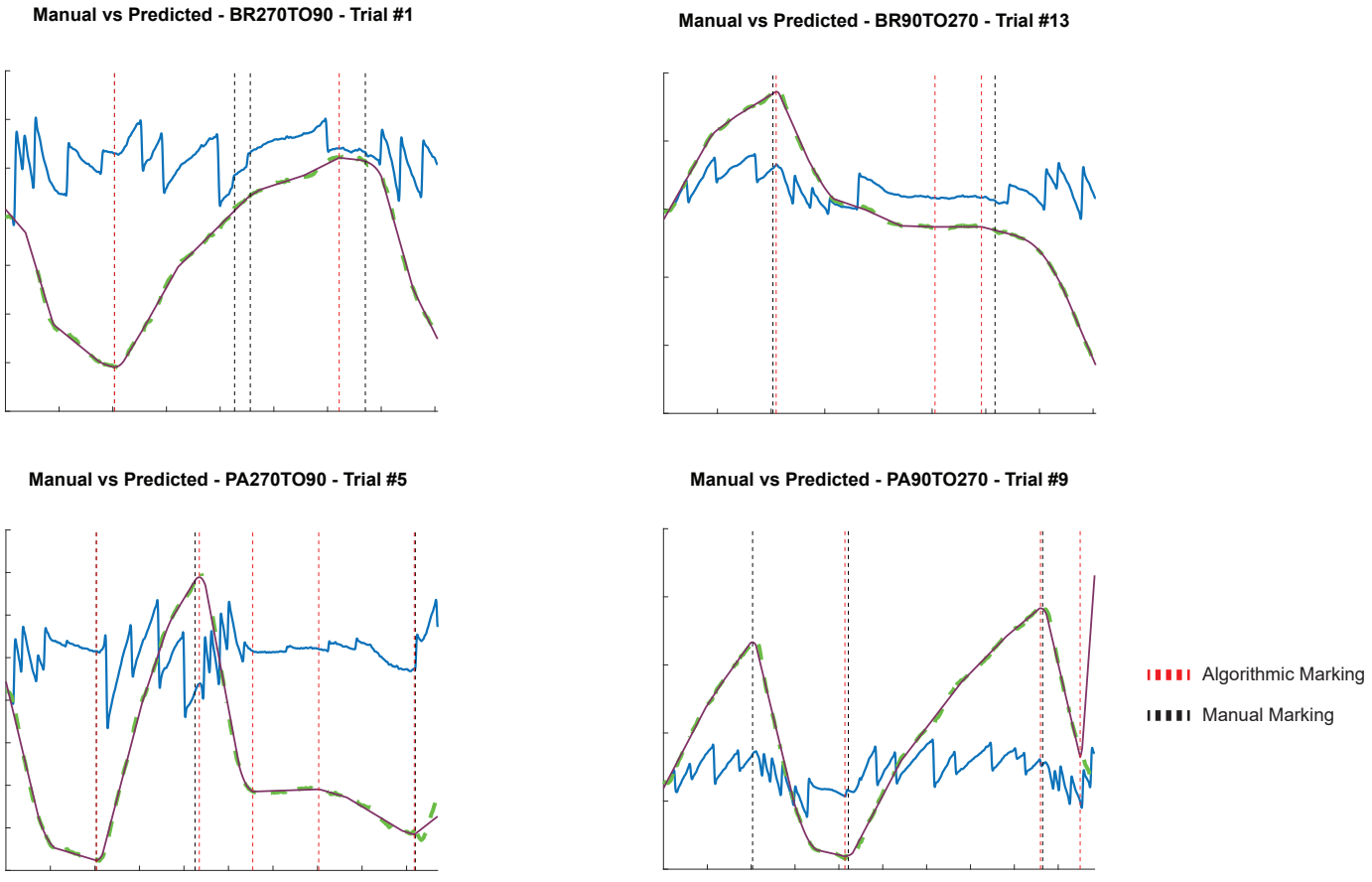
Cross-over points of the normalised population vectors (summed spike counts per bin) describe the point of change in the encoding of the dominant percept by the selective ensembles of neurons. These points are computed by first estimating the best-fit trend of the population vectors LOWESS smoothing, and then estimating the intersection using interpolation. If multiple intersections are estimated, then that intersection point is chosen such that for 200ms before and after it, there are no other crossings, i.e. stable divergence is observed. Three examples of BR and PA transitions each are shown above (green = downward selective population; red = upward selective population). These cross-over points are then collected across all transitions.



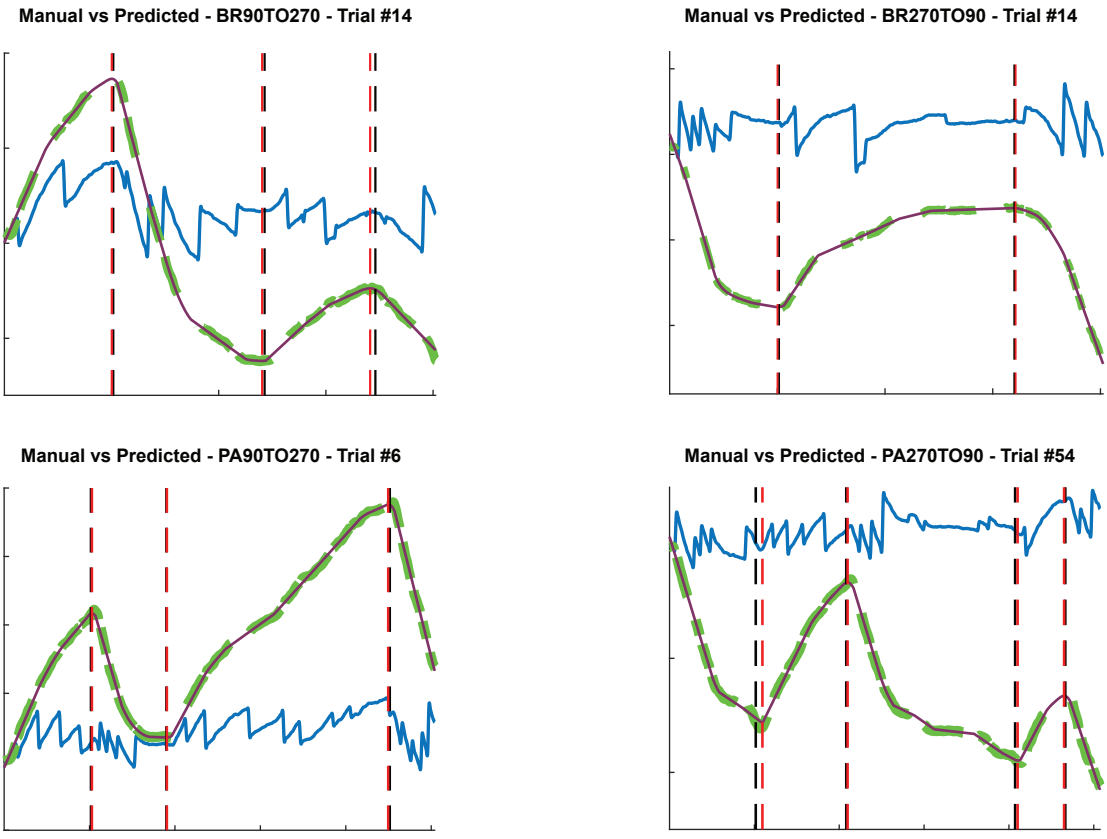
SI Figure 10 | Comparison of Beta-Suppression and Ensemble convergence times

The green probability density function shows the distribution of times of beta-activity while the purple function shows the ensemble convergence times of the two competing populations around a switch. The ongoing beta is suppressed even before the a spontaneous transition happens inferred by the change in the polarity of the OKN.

A



B



SI Figure 11 | Comparison of Automatically detected and manually marked percept switches from the OKN readout for BR and PA trials in 2 example sessions

The percept switches algorithmically detected (red) using extremum points in the the computed cumulative smooth pursuit were compared to the manually marked (black) ones using the inflection point of the polarity reversal of the OKN readout. The results were inconsistent across across sessions and conditions, only the manual markings were used to align all the data.

A. The automatic marking method in this dataset inaccurately marks switches determined by falling within a threshold window of the manually marked switches (black), along with a significant number of outright false positives.
B. The automatically marked switches in this dataset are equally accurate or slightly worse in accuracy than manual marked switches.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`http://code.google.com/p/classicthesis/`