

BRAIN AS A COMPLEX SYSTEM

harnessing systems neuroscience tools & notions for an empirical approach

Shervin safavi

*Max Planck Institute for Biological Cybernetics
Department of Physiology of Cognitive Processes*

*IMPRS for Cognitive and Systems Neuroscience
School of Neural Information Processing*

For the full version of this PhD thesis please refer to the original reference [1]:

Shervin Safavi. "Brain as a Complex System, Harnessing Systems Neuroscience Tools & Notions for an Empirical Approach." PhD thesis. Universität Tübingen, 2022. doi: 10.15496/publikation-69434

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



Content of this thesis is licensed under a Creative Commons Attribution 3.0, except the scientific papers reprinted in the thesis (see part iv), that are subject to their own copyright protection.

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.

زمر کمر صحنه پیدا نمی شود
کسر غم غم خواهد از اینجا نمی شود
صحنه پوستی بگیر

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 Gustavo 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

| | |
|--|------|
| Acknowledgements | xiii |
| Preface | xix |
| Summary | xxi |
| I Synopsis | |
| 1 Brain as a complex system | 3 |
| 1.1 Complex systems | 3 |
| 1.2 Complex system tools in neuroscience | 4 |
| 1.3 Novel complementary approaches | 6 |
| 2 Approaching through neural data analysis | 9 |
| 2.1 Necessity of investigating across scales | 9 |
| 2.2 Available tools for investigating cross-scale relationships | 12 |
| 2.3 Need for new tools for investigating cross-scale relationships | 13 |
| 2.3.1 Tools to explore micro-meso relationships | 14 |
| 2.3.2 Tools to explore meso-macro relationships | 15 |
| 3 Approaching through neural theories | 17 |
| 3.1 Criticality hypothesis of the brain | 17 |
| 3.2 Signatures of criticality in neural systems | 18 |
| 3.3 Seeking for a bridge: a complementary approach | 19 |
| 3.3.1 Efficient coding as the computational objective | 20 |
| 3.3.2 Signature of criticality in efficient coding networks | 20 |
| 4 Approaching through cognition | 21 |
| 4.1 Visual awareness | 21 |
| 4.1.1 Binocular rivalry | 21 |
| 4.1.2 Neural correlate of binocular rivalry | 22 |
| 4.2 Why is appealing from a complex system perspective | 23 |
| 4.3 Experimental considerations | 24 |
| 4.4 Toward a meso-scale understanding | 25 |
| 4.4.1 Meso-scale dynamics | 25 |
| 4.4.2 Micro-Meso relationship | 25 |
| II Manuscripts Information | |
| 5 Paper 1 | 29 |
| 6 Paper 2 | 31 |
| 7 Paper 3 | 33 |
| 8 Paper 4 | 37 |
| 9 Paper 5 | 39 |
| 10 Paper 6 | 41 |
| 11 Paper 7 | 43 |
| 12 Paper 8 | 47 |
| III Outlook | |
| 13 Brain as a complex & adaptive system | 51 |
| 13.1 Complex adaptive systems | 52 |
| 13.2 Brain computational objectives | 52 |
| 13.3 Relating behavior to multi-scale brain dynamics | 53 |

| | |
|---|-----|
| 13.3.1 Relating neural dynamics and neural computation | 53 |
| 13.3.2 Exploiting models of pivotal tasks | 54 |
| 13.3.3 A principled framework for data fusion | 56 |
| 13.4 Understating the neuro-principles through dysfunctions | 56 |
| Acronyms | 61 |
| Bibliography | 63 |
| iv Manuscripts | |
| Paper 1 | 93 |
| Paper 2 | 160 |
| Paper 3 | 206 |
| Paper 4 | 254 |
| Paper 5 | 267 |
| Paper 6 | 269 |
| Paper 7 | 279 |
| Paper 8 | 337 |

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 methodology for detecting cooperative neural activities (neural events) in local field potentials, that can be used as a trigger to investigate simultaneous 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.

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.

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.

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 sdudyng *complex systems* [2–5]. Mitchell [3, 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 $\bar{\omega}$ and standard deviation β ,

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

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,

$$\theta'_j = \omega_j . \quad (1.1)$$

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:

$$\theta'_j = \omega_j + \kappa \frac{1}{N} \sum_i^N \sin(\theta_i - \theta_j) , \quad (1.2)$$

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 [6, 7] 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 right, 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 [[8](#), Chapter 1]. Indeed, this perspective toward the brain has been extensively articulated [[3](#), [9–20](#)].

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.

The tools and frameworks adapted from the field of complex systems to address neuroscientific questions can be divided into four categories

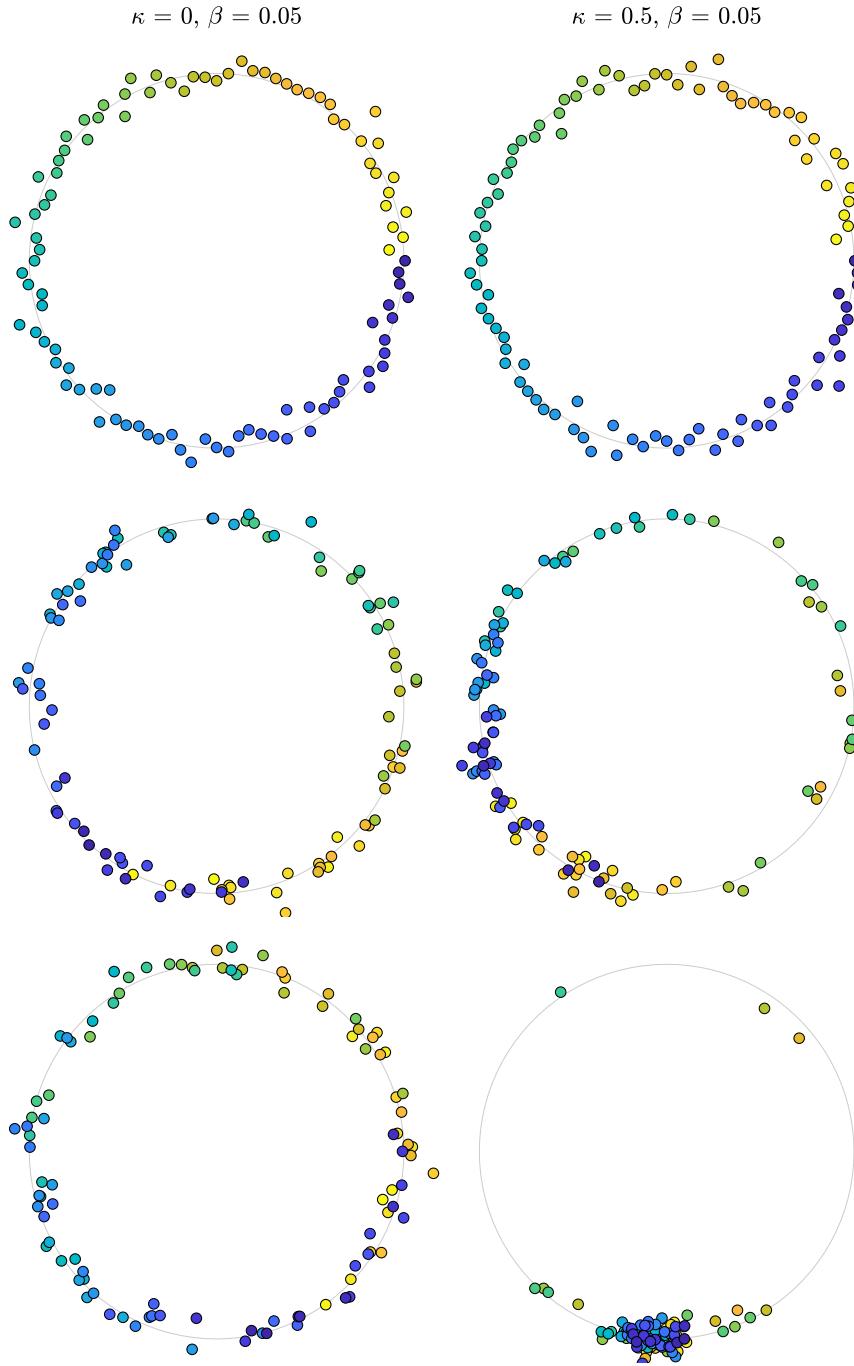


Figure 1.2: Kuramoto model (snapshots)

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 right, 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.

(of course, 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 field 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 [15, 21–25].

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 [26–28]. There have been various attempts to model single neuron [29], neuronal populations [30, Part 3][31], large-scale brain networks [32, 33] and even brain-environment system as dynamical systems [27].

INFORMATION THEORY: Information-theoretic tools have been extensively used in neuroscience, for purposes, as simple as studying neural coding in a single neuron [34–38] all the way to quantifying the level of consciousness [39–41] and providing a mathematical framework to represent the content of the conscious experience [42] (for a review see Tononi et al. [43]).

STATISTICAL PHYSICS: Statistical physics is a branch of physics which seeks for simple behaviors in systems consisting of many interacting components [44]. Such systems can be atoms of water in a glass [44], all the way to collective activity of a flock of birds [45, 46] and pattern of tweets in the Twitter network [47]. 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 [48] (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 [49]). Such caveats become even more critical for functionally relevant information-theoretic measures such as integrated information [41]. Computing or estimating the amount of integrated information in a system for more than a handful of units is challenging [43]. 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 [50, 51], communication through coherence [52, 53], and memory consolidation [54, 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) [55] 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 [56–58].

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 [45], 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 [5]. 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

*Approaching through
neural theories*

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 [59, 60], and “criticality hypothesis of the brain” [48], 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 [48, 61, 62], as it has been shown that general information processing capabilities such as sensitivity to input [63, 64], dynamic range [63, 65, 66], and information transmission and storage [67–70], 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
behavior and
cognition*

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 [71, 72] or the dynamics of the neural population can be explained by network models that are operating on the edge of a bifurcation [73, 74]. 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.

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* [5, 75]. Certainly, the brain is a prominent example of such systems [75]. 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[19]".

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 [8, 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 reorganization. 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 et al. [8, 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 [77, 78], from microcircuits to a single brain area [79], from a single area to the

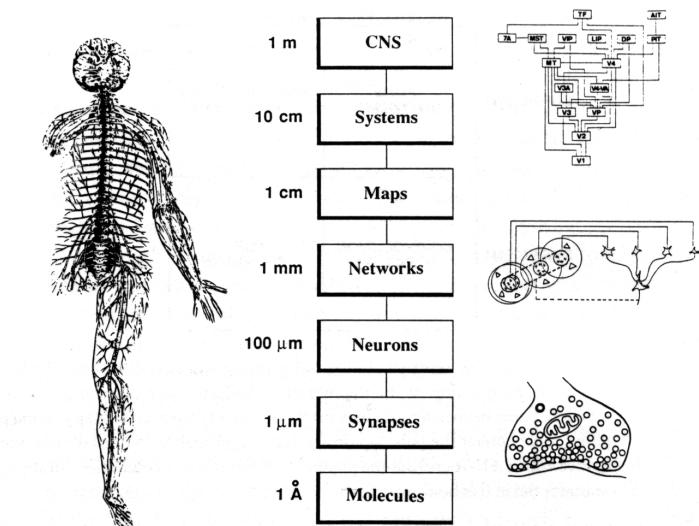


Figure 2.1: Schematic depiction of levels of organization

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 et al. [76] with permission.

whole brain [80, 81]. 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. [79]. 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 [82, 83]. 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 [52, 53, 84]. 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 [85]. 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 [6, 7] (for an integrative

review see [86]). 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 [87]). 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 [85, 88, 89].

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 [8, Chapter 1][90]. 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 et al. [8, 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 null 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 millisecond (10^{-3}) as opposed to nano second (10^{-9}) in electronic computers [90]. Other such examples are, spatial constraints (limitation by available space), energy consumption, and metabolism [8, 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 [91, 92] 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 [92] stated:

“There is thus no “functional 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 [93] and modeling [93–95] on the possibility of having ephaptic interactions in the brain (for a review also see [96]). Indeed, this evidence that electrical fields in the brain can functionally modulate the activity of neurons is in line with Bell [91, 92] 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 understanding 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-

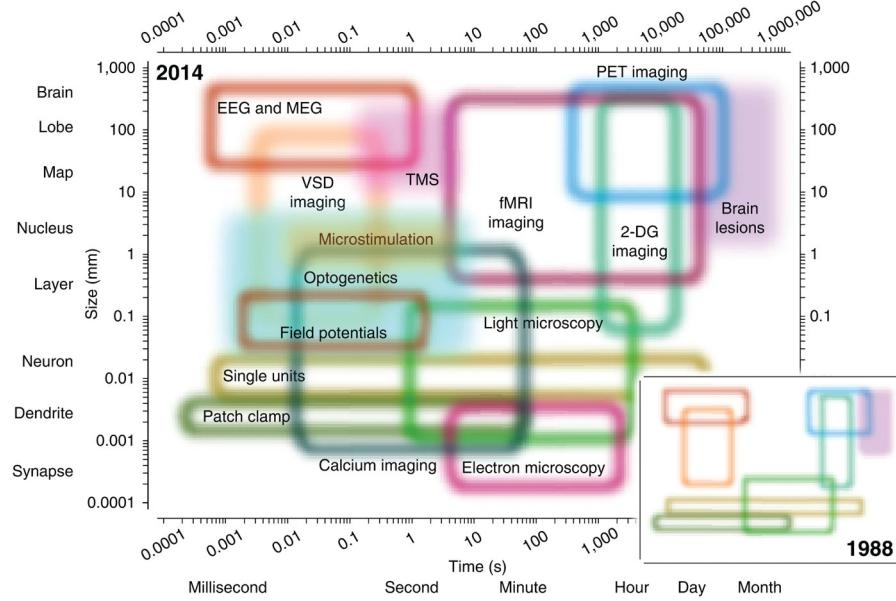


Figure 2.2: Spatio-temporal resolution of measurement methods in neuroscience
Demonstrate the spatial and temporal resolution of measurement methods being used in neuroscience (up to 2014). Each box depict the spatial (y-axis) and temporal (x-axis) of one measurement method. Open regions represent measurement 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. [97]. 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. [97] 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 have been developed for investigating the relationship between spiking activity and LFP [98–104]. 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 [105–109]. Methods used for exploring the relationship between these signals were conventional correlation analysis [105], system identification [105], Canonical Correlation Analysis (CCA) and its time-resolved kernelized version [110, 111]. 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 [57].

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 [112], 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 [112–116]. 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 [117, 118]. 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 [57].

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 relationships

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 [119–121] with hundreds or even thousands of recording sites [116, 120, 122, 123]. 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 [124, 125]), 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 [126–128].

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. [129] (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. [129] (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 [130].

¹ This is a well-established asymptotic behavior in random matrix theory for matrices with independent normally distributed entries [131]. 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 relationships

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 [57, 58, 132]. 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 capability of our method by identifying neural events in Hippocampus and LGN and also 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. [57] that was focused on sharp wave-ripples, the identified events in Hippocampus and LGN reflect

¹ Marčenko et al. [130] in not written in English, but is the original publication. The reader can refer to Anderson et al. [131, Chapter 2] instead.

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. [133], 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.

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 [[134](#), Chapter 5] as well as in wider ranges of natural phenomenon such as human social behavior [[135](#)] (see Mathis et al. [[136](#)] and Bar-Yam [[5](#)] 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 [[5](#)]. 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 [[48](#)]. 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 et al. [[137](#)]. 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. [[138](#)] at the scale of actin in dendrites, Johnson et al. [[139](#)] at the scale of neuronal membranes, Varley et al. [[140](#)] at the scale of the entire brain (for

more references see [48, 141]). Moreover, being close to this state is beneficial for the brain [48, 61, 62], as it has been shown that general information processing capabilities such as sensitivity to input [63, 64], dynamic range [63, 65, 66], or information transmission and storage [67–70], and various other computational characteristics has been also considered to be relevant [142–152] (also see [153–155] 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 [48, 141]).

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 [156]: 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 [157], forest fires [158], nuclear chain reactions [159], 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 et al. [137] 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 [48, 141]). 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. [160], that were observed in neural data [161] as well¹.

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* [29, 164, 165]. The point where the transition happens is also denoted as the critical point. There are various kinds of bifurcations (see Izhikevich [29]), 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 [166], and transitioning from an asynchronous to a synchronous state [167] have been considered as two important bifurcations for the brain (for further elaboration see Muñoz [48] and Cocchi et al. [165] and references therein). Avalanche criticality and bifurcation criticality can co-occur, when there is a continuous phase transition [165] (for example see [168, 169]), 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 [162, 163].

[170] 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 [44]. Tools from statistical mechanic have been applied in neural networks in order to understand their collective dynamics [171]. Along the same line Tkacik et al. [172] 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 [172] (but also see [173]).

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 computation². For instance, being in a state which is optimized to have the maximum sensitivity to input [63, 64], and maximum dynamic range [63, 65, 66] 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 criticality that can bridge the gap between criticality

² See also Lizier [174] (in particular chapter 6) that argue closeness to criticality is a state 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 [156, 175].

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) [176, Chapter 13][38, 177]. Indeed, models using this simple principle made various verified predictions about neural responses (e.g. receptive field in V1 [178, 179]).

Several variants of efficient coding have been developed (for a brief overview see [180]). 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. [180] 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 [180]. 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. [181]. 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. [182] 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. [181] and Chalk et al. [182] 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 [137], 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 [137]. 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 [[183](#)]. 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 [[184–186](#)].

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" [[187](#)]), 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 [188]. 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 [107]. 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 Tononi et al. [185], Blake et al. [189], Panagiotaropoulos et al. [190], and Koch et al. [191]).

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 [190]. From no modulated cell in Lateral Geniculate Nucleus (LGN) [192], to superior temporal sulcus (STS) and inferotemporal cortex (IT) [193], and Lateral Prefrontal Cortex (LPFC) [194, 195] 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 [196] and prefrontal-parietal networks [197] and causal interactions in prefrontal-occipital [198] 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* [189, 190]. 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 [185, 189–191, 194, 196–202]. 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) [73, 74, 203, 204]. Perhaps, the most appealing theoretical explanation is provided by Pastukhov et al. [74] 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 properties of the phenomena. Altogether, these finding make this phenomenon

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

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) [205] that play a crucial role in cognitive computations [206], 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 [193] and neurons in PFC have been also shown to be perceptually modulated in similar tasks [194, 207]. 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. [208], it was suggested that “frontal areas are associated with active report and introspection rather than with rivalry per se.”. In Safavi et al. [209] (also can refer to the corresponding summary,), based on a broad set of evidence, we argue that evidence provided by Frassle et al. [208] 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 [210]).

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. [194]) that used non-chronic recording with tetrodes where the experimenter could explore to find the neuron by moving the electrodes, Utah arrays are fixed and almost permanent. In Safavi et al. [126] and Kapoor et al. [195] (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 [126] in contrast to most of sensory cortices [211–215] (but also see [216]). Interestingly, we also found that spatial structure of functional connectivity

in ventro-lateral PFC is generally² 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 [217–222]. The finding on the spatial structure of noise correlation in vIPFC was not relevant for the binocular rivalry experiment as the spatial structures were not taken 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 mesoscopic-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 [126, 195] (also can refer to the corresponding summaries, and) and previously it was shown this is the case for face-selective neurons as well [194]. 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 [223] (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 et al. [224] 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 [225], which is important given than various studies argue that PFC is strongly involved in task monitoring [208]. Second, we know that the activity of neural populations in IT cortex is also correlated with perception in the absence of behavioral reports [207]. On the other side, from studies with causal intervention, we know that the activity of PFC is needed for difficult object recognition tasks [226]. 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 [227] 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. [228]

PRESENTATION AT SCIENTIFIC MEETINGS: NeurIPS 2019 Workshop: Learning with Temporal Point Processes [229], Bernstein 2021 [230]

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 [231], such as neurons' action potentials, or by members in social networks [232, 233], 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 [130]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 [130], is not written in English, but it is the original publication. Reader can refer to Anderson et al. [131, 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. [234]

PRESENTATION AT SCIENTIFIC MEETINGS: ESI-SyNC 2017 [235], AREADNE 2018 [236], Cosyne 2019 [237], Cosyne 2020 [238], Bernstein 2021 [230]

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 [239] Random Matrix Theory (RMT) [131] to approximate the distribution of singular values under the null hypothesis (see Safavi et al. [240] 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 [124, 125]), 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.

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
[[241](#)]

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 [[112](#)]. As LFPs are rich signals [[112–114](#)], 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) [[117, 118](#)], as well as larger scale (entire brain) [[57, 242](#)]. 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 [[113](#)]) [[243, 244](#)], we used Itakura-Saito divergence in the

optimization procedure of NMF [245] 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. [246]). 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. [133] 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. [57] 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. [133]. 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 \[132\]](#).

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 [247], DPG-Frühjahrstagung 2019 [248], Cosyne 2020 [249]

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 [8, Chapter 1]. Ideally, we need a framework that can accommodate both aspects of the brain in one framework [8, 250]. 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 [143, 181, 182, 251–257] and also models of neural dynamics with implications for neural computation [144, 149, 151, 166, 258–263]. 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 [181, 182, 257, 260, 264–267]. In particular, multiple recent studies provide qualitative or quantitative evidence on the usefulness of operating close to a phase transition for coding [182, 260, 266, 267]. Interestingly, the phase transition is also one of the pillars of the criticality hypothesis of the brain [48, 61, 62]. 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 [182]. 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 [181]). Given an idealized network with instantaneous synapses, the optimal network could be derived analytically from first principles [181]. 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. [182] 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 [137] 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.

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. [209]

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 [268], 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 [268]. Moreover, PFC is also a central sub-network [in a graph-theoretic sense] [205] that plays a crucial role in various cognitive functions [206]. 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 [269–271]. In particular, one of these studies focused on resolving the role of frontal lobe in conscious perception [208]. In this study, Frassle et al. [208] 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 et al. [272], in response to Frassle et al. [208], reviewed the experimental evidence based on fMRI BOLD activity in frontal lobe which suggests even with contrastive analysis (similar to Frassle et al. [208]), 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 [194, 195, 223]. In particular, two recent studies [195, 223], (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. [208]. Moreover, a recent study

by Kapoor et al. [225] based on analysis of a wider range of single units in vIPFC (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 [273–276]. Ultimately, measuring both electrophysiological activity and the BOLD signal in the same macaques engaged in an identical task of perceptual suppression settled the debate [274, 276]. 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).

PAPER INFORMATION

TITLE: Nonmonotonic spatial structure of interneuronal correlations in pre-frontal microcircuits

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

STATUS: Published in PNAS, see Safavi et al. [126]

PRESENTATION AT SCIENTIFIC MEETINGS: NeNa 2015 [277], AREADNE 2016 [278]

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 [279–282]. 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 [212], rapidly decay as a function of lateral distance in most of the sensory areas of the brain [211–215, 283]. 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] [205] that play a crucial role in cognitive computations [206], 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 [218, 220, 222, 284, 285] lateral connections are considerably expanded [217–222]. 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 vIPFC. 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 [212] 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 [286]. 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 vIPFC is different from most of the sensory cortices. In most sensory cortices, noise correlations decay monotonically as a function of distance; nevertheless, in vIPFC 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 vIPFC, 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 vIPFC. 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.

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. [195])

PRESENTATION AT SCIENTIFIC MEETINGS: FFRM 2015 [287], SfN 2018 [288], FENS 2018 [289], ASSC 2019 [290]

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 [291], higher order theory [292] and global neuronal workspace framework [293, 294] 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 [39–42] (for a review see Tononi et al. [43]), 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 [269, 270].

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) [194]. 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 [191, 271]). 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 [192–194, 202, 275], or the whole-brain dynamics measured with imaging techniques (EEG/MEG, fMRI) [185, 196–201] (for reviews see [189–191]). 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 [295]. 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 [194], 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 vIPFC. Lastly, the decoding algorithm that we used for decoding the content of the perception [296], 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 vIPFC 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).

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. [223]

PRESENTATION AT SCIENTIFIC MEETINGS: FFRM 2015 [287], SfN 2018 [288], AREADNE 2018 [297]

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 [19, 298]. 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) [299] has been used to extract spectral content of LFPs and Chronux toolbox [300] 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).

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 [302], reinforcement learning [303], and even understanding the emotion [304] that all inform us about the brain computations [305]. 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 [306], Chapter 1 and Siettos et al. [307]).

From a broader perspective, particularly in terms of Marr's levels of understating [308], 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 et al. [8, Chapter 2, Section 2] for a broad perspective on the importance of this bridge and Stephan et al. [309] and Forstmann et al. [310, 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.

*Approaching the
brain as a complex
and adaptive system*

*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*

*An integrative
understanding 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" [4, Chapter 3]. CAS have been studied for decades (see Morowitz et al. [311] 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 [312], financial market [4] and even language [313].

Different sets of properties and mechanisms which are considered to be common between different CAS have been suggested [314]. We outline the 4 features proposed by Holland [301]. 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 [315] and Arthur et al. [316, Chapter 1]).

Holland [301] 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 [4, Chapter 7] and [317, 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 [3, 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* 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 [[318](#)].

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 understanding such computational objectives in human/animals and multi-scale dynamics is a rich source for understanding 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 [[143](#), [181](#), [182](#), [251–257](#)] or neural dynamics [[144](#), [149](#), [151](#), [166](#), [258–263](#)] but also have some implications for the other one (also see Maass [[261](#)] 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 [181, 182, 254, 255, 257, 264, 319, 320]. 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. [257] and Lengyel et al. [251] 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 [177] and there have been various efforts to connect neural coding and neural dynamics [181, 182, 257, 264]. 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 [182, 321]). Nevertheless, the brain dynamics has been shown to be more complex than the reach of normative models so far [31, 164]. Not only in terms of complexity of the observed dynamics, but also in terms of scale, particularly large scale dynamics and multi-scale dynamics [141, 243]. 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 [48] – also briefly discussed in [Section 3.1](#)). Certainly, frameworks like criticality are insightful for brain dynamics [48] in particular because they provide explanations for observed multi-scale dynamics of the brain [141].

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 [322] and criticality [323–326] – 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 [327] and indeed the notion of cell assemblies are closely connected to computations implemented in the brain [328–332].

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 ex-

haustively 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. [333] 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 [334] 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 understanding 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 [74, 335–338]; 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 [71, 339–342]. Notably, most of these studies are centered around Bayesian model of the brain [343, 344].

Next to this extensive computational models (which include both normative and biophysical models) there is a large body of psychophysical (for review see [345]), electrophysiological and imaging (for review see [189, 190]), pharmacological [346, 347], and genetic studies [348–351]. 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. [352] 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 [71, 339–342]. Some of these models can even explain many aspects of binocular rivalry psychophysics and some aspects of neural dynamics [353, 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 [110, 111, 354, 355] and topological data analysis [356] 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 [357–360], and for a review see Park et al. [361]).

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 mechanistic 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 [362]. 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 [8, Chapter 1], psychiatric disorders have been studied from their genetic basis [363–365] all the way to their roots in the social interactions [366–368]. In terms of [Marr] levels of understanding [308], psychiatric disorders have been attacked in all three levels [369, Chapter 5][362].

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 [370–377] and more generally the interaction between the immune system and the brain has been receiving more attention and support recently ([371, 378–385]).

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, [386], 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.

LIST OF FIGURES

- 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 right, the oscillators do interact with each other as the coupling parameter is non-zero ($\kappa = 0.5$). 4
- 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 right, 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. 5
- 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 et al. [76] with permission. 10

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. [97]. Abbreviations used in the figure: EEG, electroencephalography; MEG, magnetoencephalography; PET, positron emission tomography; VSD, voltage-sensitive dye; TMS, transcranial magnetic stimulation; ^{2}DG , $^{2}\text{-deoxyglucose}$. Figure is adopted from Sejnowski et al. [97] with permission. [12](#)

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 Nucleus
- 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] Shervin Safavi. "Brain as a Complex System, Harnessing Systems Neuroscience Tools & Notions for an Empirical Approach." PhD thesis. Universität Tübingen, 2022. doi: 10.15496/publikation-69434 (cit. on p. 1).
- [2] Yaneer Bar-Yam. *Dynamics of Complex Systems*. Studies in Nonlinearity. Boulder, CO: Westview Press, 2003 (cit. on p. 3).
- [3] Melanie Mitchell. *Complexity: A Guided Tour*. 1 edition. Oxford: Oxford University Press, 2011 (cit. on pp. 3, 4, 52).
- [4] John H. Holland. *Complexity: A Very Short Introduction*. 1 edition. Oxford, United Kingdom: Oxford University Press, 2014 (cit. on pp. 3, 52).
- [5] Yaneer Bar-Yam. *Why Complexity Is Different*. 2017 (cit. on pp. 3, 7, 9, 17).
- [6] Yoshiki Kuramoto. "Self-Entrainment of a Population of Coupled Non-Linear Oscillators." *Int. Symp. Math. Probl. Theor. Phys.* Ed. by Huzihiro Araki. Lecture Notes in Physics. Berlin, Heidelberg: Springer, 1975, pp. 420–422. doi: 10.1007/BFb0013365 (cit. on pp. 3, 10).
- [7] Yoshiki Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Courier Corporation, 2003 (cit. on pp. 3, 10).
- [8] Patricia Smith Churchland and Terrence J. Sejnowski. *The Computational Brain*. Computational Neuroscience. Cambridge, Mass: MIT Press, 1992 (cit. on pp. 4, 9, 11, 37, 51, 57).
- [9] H. T. Siegelmann. "Complex Systems Science and Brain Dynamics." *Frontiers in computational neuroscience* 4 (2010). doi: 10.3389/fncom.2010.00007 (cit. on p. 4).
- [10] G. Werner. "Consciousness Viewed in the Framework of Brain Phase Space Dynamics, Criticality, and the Renormalization Group." *Chaos Soliton Fract* 55 (2013), pp. 3–12. doi: DOI10.1016/j.chaos.2012.03.014 (cit. on p. 4).
- [11] O. Sporns, G. Tononi, and G. M. Edelman. "Connectivity and Complexity: The Relationship between Neuroanatomy and Brain Dynamics." *Neural Networks* 13 (2000), pp. 909–922. doi: Doi10.1016/S0893-6080(00)00053-8 (cit. on p. 4).
- [12] W. Singer. "The Brain, a Complex Self-organizing System." *Eur. Rev.* 17.2 (2009), pp. 321–329. doi: 10.1017/S1062798709000751 (cit. on p. 4).
- [13] Eckehard Olbrich, Peter Achermann, and Thomas Wennekers. "The Sleeping Brain as a Complex System." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1952 (2011), pp. 3697–3707. doi: 10.1098/rsta.2011.0199 (cit. on p. 4).
- [14] C. Koch. "Systems Biology. Modular Biological Complexity." *Science* 337 (2012), pp. 531–2. doi: 10.1126/science.1218616 (cit. on p. 4).

- [15] E. Bullmore and O. Sporns. "Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems." *Nature reviews Neuroscience* 10.3 (2009), pp. 186–98. doi: 10.1038/nrn2575 (cit. on pp. 4, 6).
- [16] Christopher W. Lynn et al. "Human Information Processing in Complex Networks." *Nat. Phys.* (2020), pp. 1–9. doi: 10.1038/s41567-020-0924-7 (cit. on p. 4).
- [17] Richard F. Betzel and Danielle S. Bassett. "Multi-Scale Brain Networks." *NeuroImage. Functional Architecture of the Brain* 160 (2017), pp. 73–83. doi: 10.1016/j.neuroimage.2016.11.006 (cit. on p. 4).
- [18] D. S. Bassett and M. S. Gazzaniga. "Understanding Complexity in the Human Brain." *Trends in cognitive sciences* 15 (2011), pp. 200–9. doi: 10.1016/j.tics.2011.03.006 (cit. on p. 4).
- [19] Gyorgy Buzsaki. *Rhythms of the Brain*. New York, USA: Oxford University Press, 2011 (cit. on pp. 4, 9, 47).
- [20] D. R. Chialvo. "Emergent Complex Neural Dynamics." *Nat Phys* 6.10 (2010), pp. 744–750. doi: Doi10.1038/Nphys1803 (cit. on pp. 4, 9).
- [21] Denny Borsboom et al. "The Small World of Psychopathology." *PLOS ONE* 6.11 (2011), e27407. doi: 10.1371/journal.pone.0027407 (cit. on p. 6).
- [22] Martijn P. van den Heuvel and B. T. Thomas Yeo. "A Spotlight on Bridging Microscale and Macroscale Human Brain Architecture." *Neuron* 93.6 (2017), pp. 1248–1251. doi: 10.1016/j.neuron.2017.02.048 (cit. on p. 6).
- [23] Lianne H. Scholtens and Martijn P. van den Heuvel. "Multimodal Connectomics in Psychiatry: Bridging Scales From Micro to Macro." *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. Computational Methods and Modeling in Psychiatry* 3.9 (2018), pp. 767–776. doi: 10.1016/j.bpsc.2018.03.017 (cit. on p. 6).
- [24] Martijn P. van den Heuvel, Lianne H. Scholtens, and René S. Kahn. "Multi-Scale Neuroscience of Psychiatric Disorders." *Biological Psychiatry* (2019). doi: 10.1016/j.biopsych.2019.05.015 (cit. on p. 6).
- [25] Martijn P. van den Heuvel and Olaf Sporns. "A Cross-Disorder Connectome Landscape of Brain Dysconnectivity." *Nat. Rev. Neurosci.* 20.7 (2019), p. 435. doi: 10.1038/s41583-019-0177-6 (cit. on p. 6).
- [26] T. M. McKenna, T. A. McMullen, and M. F. Shlesinger. "The Brain as a Dynamic Physical System." *Neuroscience* 60.3 (1994), pp. 587–605. doi: 10.1016/0306-4522(94)90489-8 (cit. on p. 6).
- [27] Randall D. Beer. "A Dynamical Systems Perspective on Agent-Environment Interaction." *Artificial Intelligence* 72.1 (1995), pp. 173–215. doi: 10.1016/0004-3702(94)00005-L (cit. on p. 6).
- [28] Mikhail I. Rabinovich et al. "Dynamical Principles in Neuroscience." *Rev. Mod. Phys.* 78.4 (2006), pp. 1213–1265. doi: 10.1103/RevModPhys.78.1213 (cit. on p. 6).
- [29] Eugene M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting (Computational Neuroscience)*. Cambridge, Massachusetts, USA: The MIT Press, 2010 (cit. on pp. 6, 18).

- [30] Wulfram Gerstner et al. *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. 6).
- [31] G. Deco et al. "The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields." *PLoS computational biology* 4.8 (2008), e1000092. doi: 10.1371/journal.pcbi.1000092 (cit. on pp. 6, 54).
- [32] E. M. Izhikevich and G. M. Edelman. "Large-Scale Model of Mammalian Thalamocortical Systems." *Proceedings of the National Academy of Sciences of the United States of America* 105.9 (2008), pp. 3593–8. doi: 10.1073/pnas.0712231105 (cit. on p. 6).
- [33] G. Deco, V. K. Jirsa, and A. R. McIntosh. "Emerging Concepts for the Dynamical Organization of Resting-State Activity in the Brain." *Nature reviews. Neuroscience* 12 (2011), pp. 43–56. doi: 10.1038/nrn2961 (cit. on p. 6).
- [34] W. Bialek et al. "Reading a Neural Code." *Science* 252.5014 (1991), pp. 1854–1857. doi: 10.1126/science.2063199 (cit. on p. 6).
- [35] Rob R. de Ruyter van Steveninck et al. "Reproducibility and Variability in Neural Spike Trains." *Science* 275.5307 (1997), pp. 1805–1808. doi: 10.1126/science.275.5307.1805 (cit. on p. 6).
- [36] S. P. Strong et al. "Entropy and Information in Neural Spike Trains." *Phys. Rev. Lett.* 80.1 (1998), pp. 197–200. doi: 10.1103/PhysRevLett.80.197 (cit. on p. 6).
- [37] Alexander Borst and Frédéric E. Theunissen. "Information Theory and Neural Coding." *Nat. Neurosci.* 2.11 (1999), pp. 947–957. doi: 10.1038/14731 (cit. on p. 6).
- [38] Fred Rieke et al. *Spikes: Exploring the Neural Code*. A Bradford Book, 1999 (cit. on pp. 6, 20).
- [39] G. Tononi. "An Information Integration Theory of Consciousness." *BMC Neurosci.* 5 (2004), p. 42. doi: 10.1186/1471-2202-5-42 (cit. on pp. 6, 43).
- [40] D. Balduzzi and G. Tononi. "Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework." *PLoS computational biology* 4.6 (2008), e1000091. doi: 10.1371/journal.pcbi.1000091 (cit. on pp. 6, 43).
- [41] M. Oizumi, L. Albantakis, and G. Tononi. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0." *PLoS computational biology* 10 (2014), e1003588. doi: 10.1371/journal.pcbi.1003588 (cit. on pp. 6, 43).
- [42] D. Balduzzi and G. Tononi. "Qualia: The Geometry of Integrated Information." *PLoS computational biology* 5.8 (2009), e1000462. doi: 10.1371/journal.pcbi.1000462 (cit. on pp. 6, 43).
- [43] G. Tononi et al. "Integrated Information Theory: From Consciousness to Its Physical Substrate." *Nature reviews. Neuroscience* (2016). doi: 10.1038/nrn.2016.44 (cit. on pp. 6, 43).
- [44] James Sethna and Laboratory of Atomic and Solid State Physics James P. Sethna. *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. OUP Oxford, 2006 (cit. on pp. 6, 19).

- [45] W. Bialek et al. "Statistical Mechanics for Natural Flocks of Birds." *Proceedings of the National Academy of Sciences of the United States of America* 109 (2012), pp. 4786–91. doi: 10.1073/pnas.1118633109 (cit. on pp. 6, 7).
- [46] W. Bialek et al. "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 (2014), pp. 7212–7. doi: 10.1073/pnas.1324045111 (cit. on p. 6).
- [47] Gavin Hall and William Bialek. "The Statistical Mechanics of Twitter Communities." *J. Stat. Mech.* 2019.9 (2019), p. 093406. doi: 10.1088/1742-5468/ab3af0 (cit. on p. 6).
- [48] Miguel A. Muñoz. "Colloquium: Criticality and Dynamical Scaling in Living Systems." *Rev. Mod. Phys.* 90.3 (2018), p. 031001. doi: 10.1103/RevModPhys.90.031001 (cit. on pp. 6, 8, 17, 18, 37, 54).
- [49] Mickael Zbili and Sylvain Rama. "A Quick and Easy Way to Estimate Entropy and Mutual Information for Neuroscience." *bioRxiv* (2020), p. 2020.08.04.236174. doi: 10.1101/2020.08.04.236174 (cit. on p. 6).
- [50] T. D. Sanger. "Neural Population Codes." *Curr. Opin. Neurobiol.* 13 (2003), pp. 238–49 (cit. on p. 7).
- [51] M. Shamir. "Emerging Principles of Population Coding: In Search for the Neural Code." *Curr. Opin. Neurobiol.* 25C (2014), pp. 140–148. doi: 10.1016/j.conb.2014.01.002 (cit. on p. 7).
- [52] P. Fries. "A Mechanism for Cognitive Dynamics: Neuronal Communication through Neuronal Coherence." *Trends in cognitive sciences* 9 (2005), pp. 474–480. doi: DOI10.1016/j.tics.2005.08.011 (cit. on pp. 7, 10).
- [53] P. Fries. "Rhythms for Cognition: Communication through Coherence." *Neuron* 88 (2015), pp. 220–35. doi: 10.1016/j.neuron.2015.09.034 (cit. on pp. 7, 10).
- [54] Christoph Von Der Malsburg, William A. Phillips, and W. Singer. *Malsburg, C: Dynamic Coordination in the Brain - From Neuron: From Neurons to Mind*. Illustrated edition. Cambridge, Mass: The MIT Press, 2010 (cit. on p. 7).
- [55] Partha Mitra and Hemant Bokil. *Observed Brain Dynamics*. Oxford University Press, USA, 2007 (cit. on p. 7).
- [56] N. K. Logothetis. "Intracortical Recordings and fMRI: An Attempt to Study Operational Modules and Networks Simultaneously." *NeuroImage* 62.2 (2012), pp. 962–9. doi: 10.1016/j.neuroimage.2012.01.033 (cit. on p. 7).
- [57] N. K. Logothetis et al. "Hippocampal–Cortical Interaction during Periods of Subcortical Silence." *Nature* 491.7425 (2012), pp. 547–553. doi: 10.1038/nature11618 (cit. on pp. 7, 13–15, 33, 34).
- [58] J. F. Ramirez-Villegas, N. K. Logothetis, and M. Besserve. "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 (2015), E6379–87. doi: 10.1073/pnas.1518257112 (cit. on pp. 7, 15).
- [59] Cole Mathis, Tanmoy Bhattacharya, and Sara Imari Walker. "The Emergence of Life as a First-Order Phase Transition." *Astrobiology* 17.3 (2017), pp. 266–276. doi: 10.1089/ast.2016.1481 (cit. on p. 8).

- [60] Dante R. Chialvo. "Life at the Edge: Complexity and Criticality in Biological Function." *ArXiv181011737 Q-Bio* (2018) (cit. on p. 8).
- [61] G. Tkacik and W. Bialek. "Information Processing in Living Systems." *Annu Rev Conden Ma P* 7 (2016), pp. 89–117. doi: 10.1146/annurev-conmatphys-031214-014803 (cit. on pp. 8, 18, 37).
- [62] Thierry Mora and William Bialek. "Are Biological Systems Poised at Criticality?" *J Stat Phys* 144 (2011), pp. 268–302. doi: 10.1007/s10955-011-0229-4 (cit. on pp. 8, 18, 37).
- [63] O. Kinouchi and M. Copelli. "Optimal Dynamical Range of Excitable Networks at Criticality." *Nat Phys* 2 (2006), pp. 348–352. doi: DOI10.1038/nphys289 (cit. on pp. 8, 18, 19).
- [64] Ludmila Brochini et al. "Phase Transitions and Self-Organized Criticality in Networks of Stochastic Spiking Neurons." *Sci. Rep.* 6 (2016), p. 35831. doi: 10.1038/srep35831 (cit. on pp. 8, 18, 19).
- [65] Daniel B. Larremore, Woodrow L. Shew, and Juan G. Restrepo. "Predicting Criticality and Dynamic Range in Complex Networks: Effects of Topology." *Phys. Rev. Lett.* 106.5 (2011), p. 058101. doi: 10.1103/PhysRevLett.106.058101 (cit. on pp. 8, 18, 19).
- [66] Tazima Nur et al. "Probing Spatial Inhomogeneity of Cholinergic Changes in Cortical State in Rat." *Sci. Rep.* 9.1 (2019), p. 9387. doi: 10.1038/s41598-019-45826-4 (cit. on pp. 8, 18, 19).
- [67] W. L. Shew et al. "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 (2011), pp. 55–63. doi: 10.1523/JNEUROSCI.4637-10.2011 (cit. on pp. 8, 18).
- [68] F. Vanni, M. Lukovic, and P. Grigolini. "Criticality and Transmission of Information in a Swarm of Cooperative Units." *Physical review letters* 107 (2011), p. 078103. doi: ARTN078103DOI10.1103/PhysRevLett.107.078103 (cit. on pp. 8, 18).
- [69] M. Lukovic et al. "Transmission of Information at Criticality." *Physica A* 416 (2014), pp. 430–438. doi: DOI10.1016/j.physa.2014.08.066 (cit. on pp. 8, 18).
- [70] D. Marinazzo et al. "Information Transfer and Criticality in the Ising Model on the Human Connectome." *PloS one* 9 (2014), e93616. doi: 10.1371/journal.pone.0093616 (cit. on pp. 8, 18).
- [71] W. Bialek and M. DeWeese. "Random Switching and Optimal Processing in the Perception of Ambiguous Signals." *Physical review letters* 74 (1995), pp. 3077–3080. doi: DOI10.1103/PhysRevLett.74.3077 (cit. on pp. 8, 55, 56).
- [72] G. S. Atwal. "Statistical Mechanics of Multistable Perception." *bioRxiv* (2014). doi: 10.1101/008177 (cit. on p. 8).
- [73] P. Theodoni et al. "Cortical Microcircuit Dynamics Mediating Binocular Rivalry: The Role of Adaptation in Inhibition." English. *Front Hum Neurosci* 5 (2011), p. 145. doi: 10.3389/fnhum.2011.00145 (cit. on pp. 8, 23).
- [74] A. Pastukhov et al. "Multi-Stable Perception Balances Stability and Sensitivity." *Frontiers in computational neuroscience* 7 (2013), p. 17. doi: 10.3389/fncom.2013.00017 (cit. on pp. 8, 23, 55).

- [75] Gaute T. Einevoll et al. "The Scientific Case for Brain Simulations." *Neuron* 102.4 (2019), pp. 735–744. doi: 10.1016/j.neuron.2019.03.027 (cit. on p. 9).
- [76] P. S. Churchland and T. J. Sejnowski. "Perspectives on Cognitive Neuroscience." *Science* 242.4879 (1988), pp. 741–745. doi: 10.1126/science.3055294 (cit. on pp. 10, 59).
- [77] M. J. Rasch et al. "Inferring Spike Trains from Local Field Potentials." *Journal of neurophysiology* 99.3 (2008), pp. 1461–76. doi: 10.1152/jn.00919.2007 (cit. on p. 9).
- [78] M. Rasch, N. K. Logothetis, and G. Kreiman. "From Neurons to Circuits: Linear Estimation of Local Field Potentials." *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29 (2009), pp. 13785–96. doi: 10.1523/JNEUROSCI.2390-09.2009 (cit. on p. 9).
- [79] C. Y. Li, M. M. Poo, and Y. Dan. "Burst Spiking of a Single Cortical Neuron Modifies Global Brain State." *Science* 324.5927 (2009), pp. 643–6. doi: 10.1126/science.1169957 (cit. on pp. 9, 10).
- [80] M. Schwalm et al. "Cortex-Wide BOLD fMRI Activity Reflects Locally-Recorded Slow Oscillation-Associated Calcium Waves." *eLife*. Vol. 6. 2017 (cit. on p. 10).
- [81] Valerio Zerbi et al. "Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation." *Neuron* 103.4 (2019), 702–718.e5. doi: 10.1016/j.neuron.2019.05.034 (cit. on p. 10).
- [82] M. Volgushev, S. Chauvette, and I. Timofeev. "Long-Range Correlation of the Membrane Potential in Neocortical Neurons during Slow Oscillation." *Progress in brain research* 193 (2011), pp. 181–99. doi: 10.1016/B978-0-444-53839-0.00012-0 (cit. on p. 10).
- [83] A. Hasenstaub et al. "Inhibitory Postsynaptic Potentials Carry Synchronized Frequency Information in Active Cortical Networks." *Neuron* 47.3 (2005), pp. 423–35. doi: 10.1016/j.neuron.2005.06.016 (cit. on p. 10).
- [84] T. Womelsdorf et al. "Modulation of Neuronal Interactions through Neuronal Synchronization." *Science* 316 (2007), pp. 1609–12. doi: 10.1126/science.1139597 (cit. on p. 10).
- [85] Michel Le Van Quyen. "The Brainweb of Cross-Scale Interactions." *New Ideas in Psychology* 29.2 (2011), pp. 57–63. doi: 10.1016/j.newideapsych.2010.11.001 (cit. on pp. 10, 11).
- [86] Juan A. Acebrón et al. "The Kuramoto Model: A Simple Paradigm for Synchronization Phenomena." *Rev. Mod. Phys.* 77.1 (2005), pp. 137–185. doi: 10.1103/RevModPhys.77.137 (cit. on p. 11).
- [87] Michael Breakspear, Stewart Heitmann, and Andreas Daffertshofer. "Generative Models of Cortical Oscillations: Neurobiological Implications of the Kuramoto Model." *Front. Hum. Neurosci.* 4 (2010). doi: 10.3389/fnhum.2010.00190 (cit. on p. 11).
- [88] Michel Le Van Quyen et al. "Exploring the Nonlinear Dynamics of the Brain." *Journal of Physiology-Paris*. Neuroscience and Computation 97.4 (2003), pp. 629–639. doi: 10.1016/j.jphysparis.2004.01.019 (cit. on p. 11).

- [89] Michel Le Van Quyen. "Disentangling the Dynamic Core: A Research Program for a Neurodynamics at the Large-Scale." *Biol. Res.* 36.1 (2003), pp. 67–88. doi: 10.4067/S0716-97602003000100006 (cit. on p. 11).
- [90] R. J. Douglas and K. A. Martin. "Recurrent Neuronal Circuits in the Neocortex." *Current biology : CB* 17 (2007), R496–500. doi: 10.1016/j.cub.2007.04.024 (cit. on p. 11).
- [91] Anthony J Bell. "Levels and Loops: The Future of Artificial Intelligence and Neuroscience." *Phil.Trans. R. Soc. Lond.B* (1999), p. 8 (cit. on p. 11).
- [92] Anthony J. Bell. "Towards a Cross-Level Theory of Neural Learning." *AIP Conference Proceedings* 954.1 (2007), pp. 56–73. doi: 10.1063/1.2821301 (cit. on p. 11).
- [93] Costas A. Anastassiou et al. "Ephaptic Coupling of Cortical Neurons." *Nat. Neurosci.* 14.2 (2011), pp. 217–223. doi: 10.1038/nn.2727 (cit. on p. 11).
- [94] Giulio Ruffini et al. "Realistic Modeling of Mesoscopic Ephaptic Coupling in the Human Brain." *PLOS Computational Biology* 16.6 (2020), e1007923. doi: 10.1371/journal.pcbi.1007923 (cit. on p. 11).
- [95] Hiba Sheheitli and Viktor K. Jirsa. "A Mathematical Model of Ephaptic Interactions in Neuronal Fiber Pathways: Could There Be More than Transmission along the Tracts?" *Netw. Neurosci.* 4.3 (2020), pp. 595–610. doi: 10.1162/netn_a_00134 (cit. on p. 11).
- [96] C. A. Anastassiou and C. Koch. "Ephaptic Coupling to Endogenous Electric Field Activity: Why Bother?" *Curr. Opin. Neurobiol.* 31C (2014), pp. 95–103. doi: 10.1016/j.conb.2014.09.002 (cit. on p. 11).
- [97] Terrence J. Sejnowski, Patricia S. Churchland, and J. Anthony Movshon. "Putting Big Data to Good Use in Neuroscience." *Nat. Neurosci.* 17.11 (2014), pp. 1440–1441. doi: 10.1038/nn.3839 (cit. on pp. 12, 60).
- [98] M. Zeitler, P. Fries, and S. Gielen. "Assessing Neuronal Coherence with Single-Unit, Multi-Unit, and Local Field Potentials." *Neural computation* 18 (2006), pp. 2256–81. doi: 10.1162/neco.2006.18.9.2256 (cit. on p. 13).
- [99] Go Ashida, Hermann Wagner, and Catherine E. Carr. "Processing of Phase-Locked Spikes and Periodic Signals." *Analysis of Parallel Spike Trains*. Springer Series in Computational Neuroscience. Springer, Boston, MA, 2010, pp. 59–74. doi: 10.1007/978-1-4419-5675-0_4 (cit. on p. 13).
- [100] M. Vinck et al. "The Pairwise Phase Consistency: A Bias-Free Measure of Rhythmic Neuronal Synchronization." *NeuroImage* 51 (2010), pp. 112–22. doi: 10.1016/j.neuroimage.2010.01.073 (cit. on p. 13).
- [101] M. Vinck et al. "Improved Measures of Phase-Coupling between Spikes and the Local Field Potential." *Journal of computational neuroscience* 33 (2012), pp. 53–75. doi: 10.1007/s10827-011-0374-4 (cit. on p. 13).
- [102] Haiteng Jiang et al. "Measuring Directionality between Neuronal Oscillations of Different Frequencies." *NeuroImage* 118 (2015), pp. 359–367. doi: 10.1016/j.neuroimage.2015.05.044 (cit. on p. 13).

- [103] Z. Li, D. Cui, and X. Li. "Unbiased and Robust Quantification of Synchronization between Spikes and Local Field Potential." *Journal of neuroscience methods* 269 (2016), pp. 33–8. doi: 10.1016/j.jneumeth.2016.05.004 (cit. on p. 13).
- [104] Mohammad Zarei, Mehran Jahed, and Mohammad Reza Daliri. "Introducing a Comprehensive Framework to Measure Spike-LFP Coupling." *Front. Comput. Neurosci.* 12 (2018). doi: 10.3389/fncom.2018.00078 (cit. on p. 13).
- [105] N. K. Logothetis et al. "Neurophysiological Investigation of the Basis of the fMRI Signal." *Nature* 412 (2001), pp. 150–7. doi: 10.1038/35084005 (cit. on p. 13).
- [106] N. K. Logothetis. "The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal." *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23 (2003), pp. 3963–71 (cit. on p. 13).
- [107] N. K. Logothetis. "What We Can Do and What We Cannot Do with fMRI." *Nature* 453 (2008), pp. 869–78. doi: 10.1038/nature06976 (cit. on pp. 13, 22).
- [108] J. B. Goense and N. K. Logothetis. "Neurophysiology of the BOLD fMRI Signal in Awake Monkeys." *Current biology : CB* 18 (2008), pp. 631–40. doi: 10.1016/j.cub.2008.03.054 (cit. on p. 13).
- [109] D. Zaldivar et al. "Dopamine-Induced Dissociation of BOLD and Neural Activity in Macaque Visual Cortex." *Current biology : CB* 24 (2014), pp. 2805–11. doi: 10.1016/j.cub.2014.10.006 (cit. on p. 13).
- [110] Felix Bießmann et al. "Temporal Kernel CCA and Its Application in Multimodal Neuronal Data Analysis." *Mach. Learn.* 79.1-2 (2009), pp. 5–27. doi: 10.1007/s10994-009-5153-3 (cit. on pp. 13, 56).
- [111] Y. Murayama et al. "Relationship between Neural and Hemodynamic Signals during Spontaneous Activity Studied with Temporal Kernel CCA." *Magnetic resonance imaging* 28 (2010), pp. 1095–103. doi: 10.1016/j.mri.2009.12.016 (cit. on pp. 13, 56).
- [112] Hans Liljenstroem. "Mesoscopic Brain Dynamics." *Scholarpedia* 7.9 (2012), p. 4601. doi: 10.4249/scholarpedia.4601 (cit. on pp. 13, 33).
- [113] 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 (2012), pp. 407–20. doi: 10.1038/nrn3241 (cit. on pp. 13, 33).
- [114] G. T. Einevoll et al. "Modelling and Analysis of Local Field Potentials for Studying the Function of Cortical Circuits." *Nature reviews. Neuroscience* 14.11 (2013), pp. 770–85. doi: 10.1038/nrn3599 (cit. on pp. 13, 33).
- [115] O. Herreras. "Local Field Potentials: Myths and Misunderstandings." *Front Neural Circuit* 10 (2016), p. 101. doi: 10.3389/fncir.2016.00101 (cit. on p. 13).
- [116] Bijan Pesaran et al. "Investigating Large-Scale Brain Dynamics Using Field Potential Recordings: Analysis and Interpretation." *Nat. Neurosci.* (2018), p. 1. doi: 10.1038/s41593-018-0171-8 (cit. on pp. 13, 14).

- [117] Jozsef Csicsvari et al. "Ensemble Patterns of Hippocampal CA3-CA1 Neurons during Sharp Wave–Associated Population Events." *Neuron* 28.2 (2000), pp. 585–594. doi: 10.1016/S0896-6273(00)00135-5 (cit. on pp. 14, 33).
- [118] A. Oliva et al. "Role of Hippocampal CA2 Region in Triggering Sharp-Wave Ripples." *Neuron* 91 (2016), pp. 1342–55. doi: 10.1016/j.neuron.2016.08.008 (cit. on pp. 14, 33).
- [119] Adam S. Dickey et al. "Single-Unit Stability Using Chronically Implanted Multielectrode Arrays." *J. Neurophysiol.* 102.2 (2009), pp. 1331–1339. doi: 10.1152/jn.90920.2008 (cit. on p. 14).
- [120] J. J. Jun et al. "Fully Integrated Silicon Probes for High-Density Recording of Neural Activity." *Nature* 551 (2017), pp. 232–236. doi: 10.1038/nature24636 (cit. on p. 14).
- [121] Ashley L Juavinett, George Bekheet, and Anne K Churchland. "Chronically Implanted Neuropixels Probes Enable High-Yield Recordings in Freely Moving Mice." *eLife* 8 (2019). Ed. by Laura L Colgin and Nick Steinmetz, e47188. doi: 10.7554/eLife.47188 (cit. on p. 14).
- [122] György Buzsáki. "Large-Scale Recording of Neuronal Ensembles." *Nat. Neurosci.* 7.5 (2004), pp. 446–451. doi: 10.1038/nn1233 (cit. on p. 14).
- [123] Makoto Fukushima, Zenas C Chao, and Naotaka Fujii. "Studying Brain Functions with Mesoscopic Measurements: Advances in Electrocorticography for Non-Human Primates." *Current Opinion in Neurobiology*. Large-Scale Recording Technology (32) 32 (2015), pp. 124–131. doi: 10.1016/j.conb.2015.03.015 (cit. on p. 14).
- [124] G. Buzsaki et al. *High-Frequency Network Oscillation in the Hippocampus*. 1992 (cit. on pp. 14, 32).
- [125] Juan F. Ramirez-Villegas et al. "Dissecting the Synapse- and Frequency-Dependent Network Mechanisms of In Vivo Hippocampal Sharp Wave-Ripples." *Neuron* 100.5 (2018), 1224–1240.e13. doi: 10.1016/j.neuron.2018.09.041 (cit. on pp. 14, 32).
- [126] Shervin Safavi et al. "Nonmonotonic Spatial Structure of Interneuronal Correlations in Prefrontal Microcircuits." *PNAS* (2018), p. 201802356. doi: 10.1073/pnas.1802356115 (cit. on pp. 15, 24, 25, 41, 91).
- [127] Jason S. Sherfey et al. "Flexible Resonance in Prefrontal Networks with Strong Feedback Inhibition." *PLOS Computational Biology* 14.8 (2018), e1006357. doi: 10.1371/journal.pcbi.1006357 (cit. on p. 15).
- [128] Jason Sherfey et al. "Prefrontal Oscillations Modulate the Propagation of Neuronal Activity Required for Working Memory." *Neurobiology of Learning and Memory* 173 (2020), p. 107228. doi: 10.1016/j.nlm.2020.107228 (cit. on p. 15).
- [129] Shervin Safavi et al. "Uncovering the Organization of Neural Circuits with Generalized Phase Locking Analysis." *PLOS Computational Biology* 19.4 (2023), e1010983. doi: 10.1371/journal.pcbi.1010983 (cit. on pp. 15, 91).
- [130] V A Marčenko and L A Pastur. "Distribution of Eigenvalues for Some Sets of Random Matrices." *Math. USSR Sb.* 1.4 (1967), pp. 457–483. doi: 10.1070/SM1967v001n04ABEH001994 (cit. on pp. 15, 30).

- [131] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge; New York: Cambridge University Press, 2010 (cit. on pp. 15, 30, 32).
- [132] N. K. Logothetis. “Neural-Event-Triggered fMRI of Large-Scale Neural Networks.” *Curr. Opin. Neurobiol.* 31C (2014), pp. 214–222. DOI: 10.1016/j.conb.2014.11.009 (cit. on pp. 15, 35).
- [133] Michael Schellenberger Costa et al. “A Thalamocortical Neural Mass Model of the EEG during NREM Sleep and Its Response to Auditory Stimulation.” *PLOS Computational Biology* 12.9 (2016), e1005022. DOI: 10.1371/journal.pcbi.1005022 (cit. on pp. 16, 34).
- [134] Daniel V. Schroeder. *An Introduction to Thermal Physics*. 1st edition. San Francisco, CA: Pearson, 1999 (cit. on p. 17).
- [135] Claudio Castellano, Matteo Marsili, and Alessandro Vespignani. “Nonequilibrium Phase Transition in a Model for Social Influence.” *Phys. Rev. Lett.* 85.16 (2000), pp. 3536–3539. DOI: 10.1103/PhysRevLett.85.3536 (cit. on p. 17).
- [136] Cole Mathis, Tanmoy Bhattacharya, and Sara Imari Walker. “The Emergence of Life as a First-Order Phase Transition.” *Astrobiology* 17.3 (2017), pp. 266–276. DOI: 10.1089/ast.2016.1481 (cit. on p. 17).
- [137] J. M. Beggs and D. Plenz. “Neuronal Avalanches in Neocortical Circuits.” *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23 (2003), pp. 11167–77 (cit. on pp. 17, 18, 20, 38).
- [138] Mayte Bonilla-Quintana et al. “Actin in Dendritic Spines Self-Organizes into a Critical State.” *bioRxiv* (2020), p. 2020.04.22.054577. DOI: 10.1101/2020.04.22.054577 (cit. on p. 17).
- [139] James K. Johnson et al. “Single-Cell Membrane Potential Fluctuations Evince Network Scale-Freeness and Quasicriticality.” *J. Neurosci.* (2019), pp. 3163–18. DOI: 10.1523/JNEUROSCI.3163-18.2019 (cit. on p. 17).
- [140] Thomas F. Varley et al. “Differential Effects of Propofol and Ketamine on Critical Brain Dynamics.” *bioRxiv* (2020), p. 2020.03.27.012070. DOI: 10.1101/2020.03.27.012070 (cit. on p. 17).
- [141] Vudit Agrawal et al. “Scale-Change Symmetry in the Rules Governing Neural Systems.” *iScience* 12 (2019), pp. 121–131. DOI: 10.1016/j.isci.2019.01.009 (cit. on pp. 18, 54).
- [142] A. M. Turing. “I.—Computing Machinery and Intelligence.” *Mind* LIX (1950), pp. 433–460. DOI: 10.1093/mind/LIX.236.433 (cit. on p. 18).
- [143] Takuma Tanaka, Takeshi Kaneko, and Toshio Aoyagi. “Recurrent Infomax Generates Cell Assemblies, Neuronal Avalanches, and Simple Cell-Like Selectivity.” *Neural Computation* 21.4 (2008), pp. 1038–1067. DOI: 10.1162/neco.2008.03-08-727 (cit. on pp. 18, 37, 53).
- [144] J. Hidalgo et al. “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 (2014), pp. 10095–100. DOI: 10.1073/pnas.1319166111 (cit. on pp. 18, 37, 53).
- [145] Jorge Hidalgo et al. “Cooperation, Competition and the Emergence of Criticality in Communities of Adaptive Systems.” *J. Stat. Mech.* 2016.3 (2016), p. 033203. DOI: 10.1088/1742-5468/2016/03/033203 (cit. on p. 18).

- [146] Pedro A. M. Mediano, Juan Carlos Farah, and Murray Shanahan. "Integrated Information and Metastability in Systems of Coupled Oscillators." *ArXiv160608313 Q-Bio* (2016) (cit. on p. 18).
- [147] Sina Khajehabdollahi et al. "The Emergence of Integrated Information, Complexity, and Consciousness at Criticality." *bioRxiv* (2019), p. 521567. doi: 10.1101/521567 (cit. on p. 18).
- [148] Heiko Hoffmann and David W. Payton. "Optimization by Self-Organized Criticality." *Sci. Rep.* 8.1 (2018), p. 2358. doi: 10.1038/s41598-018-20275-7 (cit. on p. 18).
- [149] L. Michiels van Kessenich et al. "Pattern Recognition with Neuronal Avalanche Dynamics." *Phys. Rev. E* 99.1 (2019), p. 010302. doi: 10.1103/PhysRevE.99.010302 (cit. on pp. 18, 37, 53).
- [150] Rong Wang et al. "Hierarchical Connectome Modes and Critical State Jointly Maximize Human Brain Functional Diversity." *Phys. Rev. Lett.* 123.3 (2019), p. 038301. doi: 10.1103/PhysRevLett.123.038301 (cit. on p. 18).
- [151] Kathleen Finlinson et al. "Optimal Control of Excitable Systems near Criticality." *Phys. Rev. Research* 2.3 (2020), p. 033450. doi: 10.1103/PhysRevResearch.2.033450 (cit. on pp. 18, 37, 53).
- [152] Roxana Zeraati et al. *Attentional Modulation of Intrinsic Timescales in Visual Cortex and Spatial Networks*. 2021. doi: 10.1101/2021.05.17.444537 (cit. on p. 18).
- [153] J. M. Beggs. "The Criticality Hypothesis: How Local Cortical Networks Might Optimize Information Processing." *Philos T R Soc A* 366 (2008), pp. 329–343. doi: DOI10.1098/rsta.2007.2092 (cit. on p. 18).
- [154] W. L. Shew and D. Plenz. "The Functional Benefits of Criticality in the Cortex." *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* 19 (2013), pp. 88–100. doi: 10.1177/1073858412445487 (cit. on p. 18).
- [155] Roxana Zeraati, Viola Priesemann, and Anna Levina. "Self-Organization Toward Criticality by Synaptic Plasticity." *Front. Phys.* 9 (2021), p. 103. doi: 10.3389/fphy.2021.619661 (cit. on p. 18).
- [156] Roxana Zeraati. *Studying Criticality and Its Different Measures in Neuroscience*. Tech. rep. Tuebingen, Germany: Max Planck Institute for Biological Cybernetics, 2017 (cit. on pp. 18, 19).
- [157] B. Gutenberg and C. F. Richter. "Earthquake Magnitude, Intensity, Energy, and Acceleration(Second Paper)." *Bulletin of the Seismological Society of America* 46.2 (1956), pp. 105–145 (cit. on p. 18).
- [158] Bruce D. Malamud, Gleb Morein, and Donald L. Turcotte. "Forest Fires: An Example of Self-Organized Critical Behavior." *Science* 281.5384 (1998), pp. 1840–1842. doi: 10.1126/science.281.5384.1840 (cit. on p. 18).
- [159] Theodore Edward Harris. *The Theory of Branching Processes*. Grundlehren Der Mathematischen Wissenschaften. Berlin Heidelberg: Springer-Verlag, 1963 (cit. on p. 18).
- [160] J. P. Sethna, K. A. Dahmen, and C. R. Myers. "Crackling Noise." *Nature* 410.6825 (2001), pp. 242–50. doi: 10.1038/35065675 (cit. on p. 18).

- [161] N. Friedman et al. "Universal Critical Dynamics in High Resolution Neuronal Avalanche Data." *Physical review letters* 108 (2012), p. 208102 (cit. on p. 18).
- [162] Laurence Aitchison, Nicola Corradi, and Peter E. Latham. "Zipf's Law Arises Naturally When There Are Underlying, Unobserved Variables." *PLOS Computational Biology* 12.12 (2016), e1005110. doi: 10.1371/journal.pcbi.1005110 (cit. on p. 18).
- [163] Jonathan Touboul and Alain Destexhe. "Power-Law Statistics and Universal Scaling in the Absence of Criticality." *Phys. Rev. E* 95.1 (2017), p. 012413. doi: 10.1103/PhysRevE.95.012413 (cit. on p. 18).
- [164] M. Breakspear. "Dynamic Models of Large-Scale Brain Activity." *Nature neuroscience* 20.3 (2017), pp. 340–352. doi: 10.1038/nn.4497 (cit. on pp. 18, 54).
- [165] Luca Cocchi et al. "Criticality in the Brain: A Synthesis of Neurobiology, Models and Cognition." *Progress in Neurobiology* 158 (2017), pp. 132–152. doi: 10.1016/j.pneurobio.2017.07.002 (cit. on p. 18).
- [166] N. Bertschinger and T. Natschläger. "Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks." *Neural computation* 16 (2004), pp. 1413–1436 (cit. on pp. 18, 37, 53).
- [167] Serena di Santo et al. "Landau–Ginzburg Theory of Cortex Dynamics: Scale-free Avalanches Emerge at the Edge of Synchronization." *PNAS* (2018), p. 201712989. doi: 10.1073/pnas.1712989115 (cit. on p. 18).
- [168] M. O. Magnasco, O. Piro, and G. A. Cecchi. "Self-Tuned Critical Anti-Hebbian Networks." *Physical review letters* 102 (2009), p. 258102 (cit. on p. 18).
- [169] Fabrizio Pittorino et al. "Chaos and Correlated Avalanches in Excitatory Neural Networks with Synaptic Plasticity." *Phys. Rev. Lett.* 118.9 (2017), p. 098102. doi: 10.1103/PhysRevLett.118.098102 (cit. on p. 18).
- [170] Karlis Kanders, Tom Lorimer, and Ruedi Stoop. "Avalanche and Edge-of-Chaos Criticality Do Not Necessarily Co-Occur in Neural Networks." *Chaos* 27.4 (2017), p. 047408. doi: 10.1063/1.4978998 (cit. on p. 18).
- [171] D. J. Amit and Daniel J. Amit. *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press, 1992 (cit. on p. 19).
- [172] G. Tkacik et al. "Thermodynamics and Signatures of Criticality in a Network of Neurons." *Proceedings of the National Academy of Sciences of the United States of America* (2015). doi: 10.1073/pnas.1514188112 (cit. on p. 19).
- [173] Marcel Nonnenmacher et al. "Signatures of Criticality Arise from Random Subsampling in Simple Population Models." *PLOS Computational Biology* 13.10 (2017), e1005718. doi: 10.1371/journal.pcbi.1005718 (cit. on p. 19).
- [174] Joseph T. Lizier. *The Local Information Dynamics of Distributed Computation in Complex Systems*. Springer Theses. Berlin: Springer, 2013 (cit. on p. 19).

- [175] W. P. Clawson et al. "Adaptation towards Scale-Free Dynamics Improves Cortical Stimulus Discrimination at the Cost of Reduced Detection." *PLoS computational biology* 13 (2017), e1005574. doi: 10.1371/journal.pcbi.1005574 (cit. on p. 19).
- [176] Walter A. Rosenblith, ed. *Sensory Communication*. The MIT Press, 2012. doi: 10.7551/mitpress/9780262518420.001.0001 (cit. on p. 20).
- [177] Rodrigo Quián Quiroga and Stefano Panzeri, eds. *Principles of Neural Coding*. Boca Raton: CRC Press, 2013 (cit. on pp. 20, 54).
- [178] B. A. Olshausen and D. J. Field. "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images." *Nature* 381 (1996), pp. 607–9. doi: 10.1038/381607a0 (cit. on p. 20).
- [179] Eero P Simoncelli and Bruno A Olshausen. "Natural Image Statistics and Neural Representation." *Annu. Rev. Neurosci.* 24.1 (2001), pp. 1193–1216. doi: 10.1146/annurev.neuro.24.1.1193 (cit. on p. 20).
- [180] Matthew Chalk, Olivier Marre, and Gašper Tkačik. "Toward a Unified Theory of Efficient, Predictive, and Sparse Coding." *Proc. Natl. Acad. Sci. U.S.A.* 115.1 (2018), pp. 186–191. doi: 10.1073/pnas.1711114115 (cit. on p. 20).
- [181] M. Boerlin, C. K. Machens, and S. Deneve. "Predictive Coding of Dynamical Variables in Balanced Spiking Networks." *PLoS computational biology* 9 (2013), e1003258. doi: 10.1371/journal.pcbi.1003258 (cit. on pp. 20, 37, 38, 53, 54).
- [182] M. Chalk, B. Gutkin, and S. Deneve. "Neural Oscillations as a Signature of Efficient Coding in the Presence of Synaptic Delays." *eLife* 5 (2016). doi: 10.7554/eLife.13824 (cit. on pp. 20, 37, 38, 53, 54).
- [183] Koch Christof. *The Quest for Consciousness: A Neurobiological Approach*. 1st edition. Denver, Colo.: Roberts and Company Publishers, 2004 (cit. on p. 21).
- [184] N. K. Logothetis. "Vision: A Window into Consciousness." *Sci Am* 16.3 (2006), pp. 4–11. doi: 10.1038/scientificamerican0906-4sp (cit. on p. 21).
- [185] G. Tononi and C. Koch. "The Neural Correlates of Consciousness: An Update." *Annals of the New York Academy of Sciences* 1124.1 (2008), pp. 239–61. doi: 10.1196/annals.1440.004 (cit. on pp. 21–23, 44).
- [186] Christof Koch. *Consciousness: Confessions of a Romantic Reductionist*. The MIT Press, 2012 (cit. on p. 21).
- [187] F. Crick. "Visual Perception: Rivalry and Consciousness." *Nature* 379.6565 (1996), pp. 485–6. doi: 10.1038/379485a0 (cit. on p. 21).
- [188] Olivia Carter et al. "Perceptual Rivalry across Animal Species." *J. Comp. Neurol.* n/a.n/a (2020). doi: 10.1002/cne.24939 (cit. on p. 22).
- [189] R. Blake and N. Logothetis. "Visual Competition." *Nat. Rev. Neurosci.* 3.1 (2002), pp. 13–21. doi: 10.1038/nrn701 (cit. on pp. 22, 23, 44, 55).
- [190] Theofanis I. Panagiotaropoulos, Vishal Kapoor, and Nikos K. Logothetis. "Subjective Visual Perception: From Local Processing to Emergent Phenomena of Brain Activity." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1641 (2014), p. 20130534. doi: 10.1098/rstb.2013.0534 (cit. on pp. 22, 23, 44, 55).

- [191] Christof Koch et al. "Neural Correlates of Consciousness: Progress and Problems." *Nat. Rev. Neurosci.* 17.5 (2016), pp. 307–321. DOI: 10.1038/nrn.2016.22 (cit. on pp. 22, 23, 44).
- [192] S. R. Lehky and J. H. R. Maunsell. "No Binocular Rivalry in the LGN of Alert Macaque Monkeys." *Vision research* 36.9 (1996), pp. 1225–1234. DOI: Doi10.1016/0042-6989(95)00232-4 (cit. on pp. 22, 44).
- [193] D. L. Sheinberg and N. K. Logothetis. "The Role of Temporal Cortical Areas in Perceptual Organization." *Proceedings of the National Academy of Sciences of the United States of America* 94.7 (1997), pp. 3408–13 (cit. on pp. 22, 24, 44).
- [194] T. I. Panagiotaropoulos et al. "Neuronal Discharges and Gamma Oscillations Explicitly Reflect Visual Consciousness in the Lateral Prefrontal Cortex." *Neuron* 74.5 (2012), pp. 924–35. DOI: 10.1016/j.neuron.2012.04.013 (cit. on pp. 22–25, 39, 44, 45).
- [195] Vishal Kapoor et al. "Decoding Internally Generated Transitions of Conscious Contents in the Prefrontal Cortex without Subjective Reports." *Nat Commun* 13.1 (2022), p. 1535. DOI: 10.1038/s41467-022-28897-2 (cit. on pp. 22, 24, 25, 39, 43, 91).
- [196] J. F. Hipp, A. K. Engel, and M. Siegel. "Oscillatory Synchronization in Large-Scale Cortical Networks Predicts Perception." *Neuron* 69 (2011), pp. 387–96. DOI: 10.1016/j.neuron.2010.12.027 (cit. on pp. 22, 23, 44).
- [197] S. M. Doesburg et al. "Rhythms of Consciousness: Binocular Rivalry Reveals Large-Scale Oscillatory Network Dynamics Mediating Visual Perception." *PLoS One* 4.7 (2009), e6142. DOI: 10.1371/journal.pone.0006142 (cit. on pp. 22, 23, 44).
- [198] F. Imamoglu et al. "Changes in Functional Connectivity Support Conscious Object Recognition." *NeuroImage* 63.4 (2012), pp. 1909–17. DOI: 10.1016/j.neuroimage.2012.07.056 (cit. on pp. 22, 23, 44).
- [199] M. Wang, D. Arteaga, and B. J. He. "Brain Mechanisms for Simple Perception and Bistable Perception." *Proceedings of the National Academy of Sciences of the United States of America* 110.35 (2013), E3350–9. DOI: 10.1073/pnas.1221945110 (cit. on pp. 23, 44).
- [200] E. D. Lumer, K. J. Friston, and G. Rees. "Neural Correlates of Perceptual Rivalry in the Human Brain." *Science* 280.5371 (1998), pp. 1930–4. DOI: 10.1126/science.280.5371.1930 (cit. on pp. 23, 44).
- [201] R. Srinivasan et al. "Increased Synchronization of Neuromagnetic Responses during Conscious Perception." *The Journal of neuroscience : the official journal of the Society for Neuroscience* 19.13 (1999), pp. 5435–48 (cit. on pp. 23, 44).
- [202] H. Bahmani, N. Logothetis, and G. Keliris. *Neural Correlates of Binocular Rivalry in Parietal Cortex*. Freiburg, Germany, 2011 (cit. on pp. 23, 44).
- [203] T. Ditzinger and H. Haken. "Oscillations in the Perception of Ambiguous Patterns - a Model Based on Synergetics." *Biological cybernetics* 61.4 (1989), pp. 279–287. DOI: Doi10.1007/Bf00203175 (cit. on p. 23).
- [204] J. Braun and M. Mattia. "Attractors and Noise: Twin Drivers of Decisions and Multistability." *NeuroImage* 52 (2010), pp. 740–751. DOI: DOI10.1016/j.neuroimage.2009.12.126 (cit. on p. 23).

- [205] D. S. Modha and R. Singh. "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 (2010), pp. 13485–90. doi: 10.1073/pnas.1008054107 (cit. on pp. 24, 39, 41).
- [206] E. K. Miller and J. D. Cohen. "An Integrative Theory of Prefrontal Cortex Function." *Annual review of neuroscience* 24 (2001), pp. 167–202. doi: 10.1146/annurev.neuro.24.1.167 (cit. on pp. 24, 39, 41).
- [207] Janis Karan Hesse and Doris Y Tsao. "A New No-Report Paradigm Reveals That Face Cells Encode Both Consciously Perceived and Suppressed Stimuli." *eLife* 9 (2020). Ed. by Ming Meng, e58360. doi: 10.7554/eLife.58360 (cit. on pp. 24, 26).
- [208] S. Frassle et al. "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 (2014), pp. 1738–47. doi: 10.1523/JNEUROSCI.4403-13.2014 (cit. on pp. 24, 26, 39).
- [209] Shervin Safavi et al. "Is the Frontal Lobe Involved in Conscious Perception?" *Front. Psychol.* 5 (2014). doi: 10.3389/fpsyg.2014.01063 (cit. on pp. 24, 39, 91).
- [210] David Leopold, A. Maier, and N.K. Logothetis. "Measuring Subjective Visual Perception in the Nonhuman Primate." *Journal of Consciousness Studies* 10.9-10 (2003), pp. 115–130 (cit. on p. 24).
- [211] Gideon Rothschild, Israel Nelken, and Adi Mizrahi. "Functional Organization and Population Dynamics in the Mouse Primary Auditory Cortex." *Nat. Neurosci.* 13.3 (2010), pp. 353–360. doi: 10.1038/nn.2484 (cit. on pp. 24, 41).
- [212] M. R. Cohen and A. Kohn. "Measuring and Interpreting Neuronal Correlations." *Nature neuroscience* 14.7 (2011), pp. 811–9. doi: 10.1038/nn.2842 (cit. on pp. 24, 41, 42).
- [213] M. A. Smith and A. Kohn. "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 (2008), pp. 12591–603. doi: 10.1523/JNEUROSCI.2929-08.2008 (cit. on pp. 24, 41).
- [214] M. A. Smith and M. A. Sommer. "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 (2013), pp. 5422–32. doi: 10.1523/JNEUROSCI.4782-12.2013 (cit. on pp. 24, 41).
- [215] Daniel J. Denman and Diego Contreras. "The Structure of Pairwise Correlation in Mouse Primary Visual Cortex Reveals Functional Organization in the Absence of an Orientation Map." *Cereb Cortex* 24.10 (2014), pp. 2707–2720. doi: 10.1093/cercor/bht128 (cit. on pp. 24, 41).
- [216] R. Rosenbaum et al. "The Spatial Structure of Correlated Neuronal Variability." *Nature neuroscience* 20 (2017), pp. 107–114. doi: 10.1038/nn.4433 (cit. on p. 24).
- [217] Jonathan B. Levitt et al. "Topography of Pyramidal Neuron Intrinsic Connections in Macaque Monkey Prefrontal Cortex (Areas 9 and 46)." *J. Comp. Neurol.* 338.3 (1993), pp. 360–376. doi: 10.1002/cne.903380304 (cit. on pp. 25, 41).

- [218] Y. Amir, M. Harel, and Rafael Malach. "Cortical Hierarchy Reflected in the Organization of Intrinsic Connections in Macaque Monkey Visual Cortex." *J. Comp. Neurol.* 334.1 (1993), pp. 19–46. DOI: 10.1002/cne.903340103 (cit. on pp. 25, 41).
- [219] Jennifer S. Lund, Takashi Yoshioka, and Jonathan B. Levitt. "Comparison of Intrinsic Connectivity in Different Areas of Macaque Monkey Cerebral Cortex." *Cereb Cortex* 3.2 (1993), pp. 148–162. DOI: 10.1093/cercor/3.2.148 (cit. on pp. 25, 41).
- [220] M. F. Kritzer and P. S. Goldman-Rakic. "Intrinsic Circuit Organization of the Major Layers and Sublayers of the Dorsolateral Prefrontal Cortex in the Rhesus Monkey." *J. Comp. Neurol.* 359.1 (1995), pp. 131–143. DOI: 10.1002/cne.903590109 (cit. on pp. 25, 41).
- [221] Ichiro Fujita and Taeko Fujita. "Intrinsic Connections in the Macaque Inferior Temporal Cortex." *J. Comp. Neurol.* 368.4 (1996), pp. 467–486. DOI: 10.1002/(SICI)1096-9861(19960513)368:4<467::AID-CNE1>3.0.CO;2-2 (cit. on pp. 25, 41).
- [222] Hisashi Tanigawa, QuanXin Wang, and Ichiro Fujita. "Organization of Horizontal Axons in the Inferior Temporal Cortex and Primary Visual Cortex of the Macaque Monkey." *Cerebral Cortex* 15.12 (2005), pp. 1887–1899. DOI: 10.1093/cercor/bhi067 (cit. on pp. 25, 41).
- [223] Abhilash Dwarakanath et al. "Bistability of Prefrontal States Gates Access to Consciousness." *Neuron* (2023). DOI: 10.1101/j.neuron.2023.02.027 (cit. on pp. 25, 39, 47, 91).
- [224] G. Buzsaki and E. W. Schomburg. "What Does Gamma Coherence Tell Us about Inter-Regional Neural Communication?" *Nature neuroscience* 18 (2015), pp. 484–9. DOI: 10.1038/nn.3952 (cit. on p. 26).
- [225] Vishal Kapoor et al. "Parallel and Functionally Segregated Processing of Task Phase and Conscious Content in the Prefrontal Cortex." *Commun. Biol.* 1.1 (2018), pp. 1–12. DOI: 10.1101/s42003-018-0225-1 (cit. on pp. 26, 40).
- [226] Kohitij Kar and James J. DiCarlo. "Fast Recurrent Processing via Ventral Prefrontal Cortex Is Needed by the Primate Ventral Stream for Robust Core Visual Object Recognition." *bioRxiv* (2020), p. 2020.05.10.086959. DOI: 10.1101/2020.05.10.086959 (cit. on p. 26).
- [227] Amy Brand et al. "Beyond Authorship: Attribution, Contribution, Collaboration, and Credit." *Learn. Publ.* 28.2 (2015), pp. 151–155. DOI: 10.1087/20150211 (cit. on p. 27).
- [228] Shervin Safavi, Nikos K. Logothetis, and Michel Besserve. "From Univariate to Multivariate Coupling between Continuous Signals and Point Processes: A Mathematical Framework." *Neural Computation* (2021), pp. 1–67. DOI: 10.1162/neco_a_01389 (cit. on pp. 29, 91).
- [229] S. Safavi, N. K. Logothetis, and M. Besserve. "Multivariate Coupling Estimation between Continuous Signals and Point Processes." *NeurIPS 2019 Workshop: Learning with Temporal Point Processes.* 2019 (cit. on p. 29).
- [230] Shervin Safavi et al. "Generalized Phase Locking Analysis: A Multivariate Technique for Investigating Spike-Field Coupling." *Bernstein Conference. G-Node,* 2021. DOI: 10.12751/NNCN.BC2021.P109 (cit. on pp. 29, 31).

- [231] Don H. Johnson. "Point Process Models of Single-Neuron Discharges." *J Comput Neurosci* 3.4 (1996), pp. 275–299. doi: 10.1007/BF00161089 (cit. on p. 29).
- [232] Hanjun Dai et al. "Recurrent Coevolutionary Latent Feature Processes for Continuous-Time Recommendation." *Proc. 1st Workshop Deep Learn. Recomm. Syst.* DLRS 2016. New York, NY, USA: Association for Computing Machinery, 2016, pp. 29–34. doi: 10.1145/2988450.2988451 (cit. on p. 29).
- [233] Abir De et al. "Learning and Forecasting Opinion Dynamics in Social Networks." *Proc. 30th Int. Conf. Neural Inf. Process. Syst. NIPS'16*. Red Hook, NY, USA: Curran Associates Inc., 2016, pp. 397–405 (cit. on p. 29).
- [234] Shervin Safavi et al. "Uncovering the Organization of Neural Circuits with Generalized Phase Locking Analysis." *bioRxiv* (2020), p. 2020.12.09.413401. doi: 10.1101/2020.12.09.413401 (cit. on p. 31).
- [235] S. Safavi et al. "Generalized Phase Locking Analysis of Electrophysiology Data." *ESI Systems Neuroscience Conference (ESI-SyNC 2017): Principles of Structural and Functional Connectivity*. 2017 (cit. on p. 31).
- [236] S. Safavi et al. "Generalized Phase Locking Analysis of Electrophysiology Data." *AREADNE 2018: Research in Encoding And Decoding of Neural Ensembles*. AREADNE Foundation, 2018, p. 88 (cit. on p. 31).
- [237] M. Besserve et al. "Generalized Phase Locking Analysis of Electrophysiology Data." *Computational and Systems Neuroscience Meeting (COSYNE 2019)*. 2019, pp. 184–185 (cit. on p. 31).
- [238] S. Safavi et al. "Uncovering the Organization of Neural Circuits with Generalized Phase-Locking Analysis." *Computational and Systems Neuroscience Meeting (COSYNE 2020)*. 2020, pp. 150–151 (cit. on p. 31).
- [239] Odd O. Aalen, Ørnulf Borgan, and Håkon K. Gjessing. *Survival and Event History Analysis: A Process Point of View*. Statistics for Biology and Health. New York, NY: Springer, 2008 (cit. on p. 32).
- [240] Shervin Safavi, Nikos K. Logothetis, and Michel Besserve. "From Univariate to Multivariate Coupling between Continuous Signals and Point Processes: A Mathematical Framework." *ArXiv200504034 Q-Bio Stat* (2020) (cit. on p. 32).
- [241] M. Besserve and S. Safavi. "Practical on Machine Learning for Neuroscience." *Machine Learning Summer School (MLSS 2016)*. 2016 (cit. on p. 33).
- [242] J Karimi Abadchi et al. "Spatiotemporal Patterns of Neocortical Activity around Hippocampal Sharp-Wave Ripples." *eLife* 9 (2020). Ed. by Sachin Deshmukh, Laura L Colgin, and Sachin Deshmukh, e51972. doi: 10.7554/eLife.51972 (cit. on p. 33).
- [243] Walter J. Freeman and Michael Breakspear. "Scale-Free Neocortical Dynamics." *Scholarpedia* 2.2 (2007), p. 1357. doi: 10.4249/scholarpedia.1357 (cit. on pp. 33, 54).
- [244] B. J. He. "Scale-Free Brain Activity: Past, Present, and Future." *Trends in cognitive sciences* 18 (2014), pp. 480–7. doi: 10.1016/j.tics.2014.04.003 (cit. on p. 33).

- [245] C. Févotte, N. Bertin, and J. Durrieu. “Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis.” *Neural Comput.* 21.3 (2009), pp. 793–830. doi: 10.1162/neco.2008.04-08-771 (cit. on p. 34).
- [246] B. Mailhé et al. “Shift-Invariant Dictionary Learning for Sparse Representations: Extending K-SVD.” *2008 16th Eur. Signal Process. Conf.* 2008, pp. 1–5 (cit. on p. 34).
- [247] S. Safavi et al. “From Optimal Efficient Coding to Criticality.” *Conference on Complex Systems (CCS 2018) Satellite: Complexity from Cells to Consciousness: Free Energy, Integrated Information, and Epsilon Machines.* 2018 (cit. on p. 37).
- [248] S. Safavi et al. “Signatures of Criticality in Efficient Coding Networks.” *DPG-Frühjahrstagung 2019.* 2019 (cit. on p. 37).
- [249] A. Levina et al. “Signatures of Criticality Observed in Efficient Coding Networks.” *Computational and Systems Neuroscience Meeting (COSYNE 2020).* 2020, p. 109 (cit. on p. 37).
- [250] C. W. Eurich. *Neural Dynamics and Neural Coding Two Complementary Approaches.* Tech. rep. 2003 (cit. on p. 37).
- [251] Máté Lengyel et al. “Matching Storage and Recall: Hippocampal Spike Timing-Dependent Plasticity and Phase Response Curves.” *Nat. Neurosci.* 8.12 (2005), pp. 1677–1683. doi: 10.1038/nn1561 (cit. on pp. 37, 53, 54).
- [252] S. Deneve. “Bayesian Spiking Neurons I: Inference.” *Neural computation* 20 (2008), pp. 91–117. doi: 10.1162/neco.2008.20.1.91 (cit. on pp. 37, 53).
- [253] S. Deneve. “Bayesian Spiking Neurons II: Learning.” *Neural computation* 20 (2008), pp. 118–45. doi: 10.1162/neco.2008.20.1.118 (cit. on pp. 37, 53).
- [254] L. Buesing et al. “Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons.” *PLoS computational biology* 7 (2011), e1002211. doi: 10.1371/journal.pcbi.1002211 (cit. on pp. 37, 53, 54).
- [255] Johannes Bill et al. “Distributed Bayesian Computation and Self-Organized Learning in Sheets of Spiking Neurons with Local Lateral Inhibition.” *PLOS ONE* 10.8 (2015), e0134356. doi: 10.1371/journal.pone.0134356 (cit. on pp. 37, 53, 54).
- [256] Fleur Zeldenrust, Boris Gutkin, and Sophie Denévé. “Efficient and Robust Coding in Heterogeneous Recurrent Networks.” *bioRxiv* (2019), p. 804864. doi: 10.1101/804864 (cit. on pp. 37, 53).
- [257] Rodrigo Echeveste et al. “Cortical-like Dynamics in Recurrent Circuits Optimized for Sampling-Based Probabilistic Inference.” *Nat. Neurosci.* 23.9 (2020), pp. 1138–1149. doi: 10.1038/s41593-020-0671-1 (cit. on pp. 37, 53, 54).
- [258] Chris Eliasmith. “A Unified Approach to Building and Controlling Spiking Attractor Networks.” *Neural Computation* 17.6 (2005), pp. 1276–1314. doi: 10.1162/0899766053630332 (cit. on pp. 37, 53).
- [259] D. Sussillo. “Neural Circuits as Computational Dynamical Systems.” *Curr Opin Neurobiol* 25 (2014), pp. 156–63. doi: 10.1016/j.conb.2014.01.008 (cit. on pp. 37, 53).

- [260] Oren Shriki and Dovi Yellin. "Optimal Information Representation and Criticality in an Adaptive Sensory Recurrent Neuronal Network." *PLOS Computational Biology* 12.2 (2016), e1004698. doi: 10.1371/journal.pcbi.1004698 (cit. on pp. 37, 53).
- [261] W. Maass. "Searching for Principles of Brain Computation." *Curr. Opin. Behav. Sci.* 11 (2016), pp. 81–92. doi: 10.1016/j.cobeha.2016.06.003 (cit. on pp. 37, 53).
- [262] Christopher M Kim and Carson C Chow. "Learning Recurrent Dynamics in Spiking Networks." *eLife* 7 (2018), e37124. doi: 10.7554/eLife.37124 (cit. on pp. 37, 53).
- [263] Guozhang Chen and Pulin Gong. "Computing by Modulating Spontaneous Cortical Activity Patterns as a Mechanism of Active Visual Processing." *Nat Commun* 10.1 (2019), pp. 1–15. doi: 10.1038/s41467-019-12918-8 (cit. on pp. 37, 53).
- [264] G. Bard Ermentrout, Roberto F. Galán, and Nathaniel N. Urban. "Relating Neural Dynamics to Neural Coding." *Phys. Rev. Lett.* 99.24 (2007), p. 248103. doi: 10.1103/PhysRevLett.99.248103 (cit. on pp. 37, 54).
- [265] Andrea Alamia and Rufin VanRullen. "Alpha Oscillations and Traveling Waves: Signatures of Predictive Coding?" *PLOS Biology* 17.10 (2019), e3000487. doi: 10.1371/journal.pbio.3000487 (cit. on p. 37).
- [266] Jonathan Kadmon, Jonathan Timcheck, and Surya Ganguli. "Predictive Coding in Balanced Neural Networks with Noise, Chaos and Delays." *ArXiv200614178 Cond-Mat Q-Bio Stat* (2020) (cit. on p. 37).
- [267] Kai Roeth, Shuai Shao, and Julijana Gjorgjieva. "Efficient Population Coding Depends on Stimulus Convergence and Source of Noise." *bioRxiv* (2020), p. 2020.06.15.151795. doi: 10.1101/2020.06.15.151795 (cit. on p. 37).
- [268] Nikos Logothetis. *Studies of Large-Scale Networks with DES- & NETfMRI*. Tech. rep. Max Planck Institute for Biological Cybernetics, 2014 (cit. on p. 39).
- [269] Jaan Aru et al. "Distilling the Neural Correlates of Consciousness." *Neuroscience & Biobehavioral Reviews* 36.2 (2012), pp. 737–746. doi: 10.1016/j.neubiorev.2011.12.003 (cit. on pp. 39, 43).
- [270] Tom A. de Graaf, Po-Jang Hsieh, and Alexander T. Sack. "The 'correlates' in Neural Correlates of Consciousness." *Neurosci Biobehav Rev* 36.1 (2012), pp. 191–197. doi: 10.1016/j.neubiorev.2011.05.012 (cit. on pp. 39, 43).
- [271] Naotsugu Tsuchiya et al. "No-Report Paradigms: Extracting the True Neural Correlates of Consciousness." *Trends in Cognitive Sciences* 19.12 (2015), pp. 757–770. doi: 10.1016/j.tics.2015.10.002 (cit. on pp. 39, 44).
- [272] N. Zaretskaya and M. Narinyan. "Introspection, Attention or Awareness? The Role of the Frontal Lobe in Binocular Rivalry." *Frontiers in human neuroscience* 8 (2014), p. 527. doi: 10.3389/fnhum.2014.00527 (cit. on p. 39).
- [273] D. A. Leopold and N. K. Logothetis. "Activity Changes in Early Visual Cortex Reflect Monkeys' Percepts during Binocular Rivalry." *Nature* 379.6565 (1996), pp. 549–53. doi: 10.1038/379549a0 (cit. on p. 40).

- [274] A. Maier et al. "Divergence of fMRI and Neural Signals in V1 during Perceptual Suppression in the Awake Monkey." *Nature neuroscience* 11.10 (2008), pp. 1193–200. DOI: 10.1038/nn.2173 (cit. on p. 40).
- [275] G. A. Keliris, N. K. Logothetis, and A. S. Tolias. "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 (2010), pp. 12353–65. DOI: 10.1523/JNEUROSCI.0677-10.2010 (cit. on pp. 40, 44).
- [276] David A. Leopold. "Primary Visual Cortex: Awareness and blindsight." *Annu. Rev. Neurosci.* 35.1 (2012), pp. 91–109. DOI: 10.1146/annurev-neuro-062111-150356 (cit. on p. 40).
- [277] A. Dwarakanath et al. "Temporal Regimes of State-Dependent Correlated Variability in the Macaque Ventrolateral Prefrontal Cortex." 2015, p. 18 (cit. on p. 41).
- [278] S. Safavi et al. "A Non-Monotonic Correlation Structure in the Macaque Ventrolateral Prefrontal Cortex." AREADNE. The AREADNE Foundation, 2016, p. 53 (cit. on p. 41).
- [279] Rodney J. Douglas, Kevan A.C. Martin, and David Whitteridge. "A Canonical Microcircuit for Neocortex." *Neural Comput.* 1.4 (1989), pp. 480–488. DOI: 10.1162/neco.1989.1.4.480 (cit. on p. 41).
- [280] Rodney J. Douglas and Kevan A.C. Martin. "Neuronal Circuits of the Neocortex." *Annu. Rev. Neurosci.* 27.1 (2004), pp. 419–451. DOI: 10.1146/annurev.neuro.27.070203.144152 (cit. on p. 41).
- [281] Rodney J. Douglas and Kevan A. C. Martin. "Mapping the Matrix: The Ways of Neocortex." *Neuron* 56.2 (2007), pp. 226–238. DOI: 10.1016/j.neuron.2007.10.017 (cit. on p. 41).
- [282] K. D. Harris and T. D. Mrsic-Flogel. "Cortical Connectivity and Sensory Coding." *Nature* 503:7474 (2013), pp. 51–8. DOI: 10.1038/nature12654 (cit. on p. 41).
- [283] C. Constantinidis and P. S. Goldman-Rakic. "Correlated Discharges among Putative Pyramidal Neurons and Interneurons in the Primate Prefrontal Cortex." *Journal of neurophysiology* 88.6 (2002), pp. 3487–3497. DOI: DOI10.1152/jn.00188.2002 (cit. on p. 41).
- [284] Alessandra Angelucci et al. "Circuits for Local and Global Signal Integration in Primary Visual Cortex." *J. Neurosci.* 22.19 (2002), pp. 8633–8646. DOI: 10.1523/JNEUROSCI.22-19-08633.2002 (cit. on p. 41).
- [285] Nicole Voges et al. "A Modeler's View on the Spatial Structure of Intrinsic Horizontal Connectivity in the Neocortex." *Progress in Neurobiology* 92.3 (2010), pp. 277–292. DOI: 10.1016/j.pneurobio.2010.05.001 (cit. on p. 41).
- [286] W. Bair, E. Zohary, and W. T. Newsome. "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 (2001), pp. 1676–97 (cit. on p. 42).
- [287] R. Antoniou et al. "Perceptual Modulation of Pupillary Reflex in Macaque Monkeys." *Federation of European Neuroscience Society Featured Regional Meeting (FFRM 2015)*. 2015 (cit. on pp. 43, 47).

- [288] F. Panagiotaropoulos et al. "Modulation of Neural Discharges and Local Field Potentials in the Macaque Prefrontal Cortex during Binocular Rivalry." *48th Annual Meeting of the Society for Neuroscience (Neuroscience 2018)*. 2018 (cit. on pp. 43, 47).
- [289] V. Kapoor et al. "Spiking Activity in the Prefrontal Cortex Reflects Spontaneous Perceptual Transitions during a No Report Binocular Rivalry Paradigm." *11th FENS Forum of Neuroscience*. 2018 (cit. on p. 43).
- [290] V. Kapoor et al. "Neuronal Discharges in the Prefrontal Cortex Reflect Changes in Conscious Perception during a No Report Binocular Rivalry Paradigm." *Association for the Scientific Study of Consciousness* 23. 2019 (cit. on p. 43).
- [291] F. Crick and C. Koch. "Consciousness and Neuroscience." *Cerebral cortex* 8.2 (1998), pp. 97–107 (cit. on p. 43).
- [292] H. Lau and D. Rosenthal. "Empirical Support for Higher-Order Theories of Conscious Awareness." *Trends in cognitive sciences* 15 (2011), pp. 365–73. DOI: 10.1016/j.tics.2011.05.009 (cit. on p. 43).
- [293] Bernard J. Baars. "Global Workspace Theory of Consciousness: Toward a Cognitive Neuroscience of Human Experience." *Prog Brain Res* 150 (2005), pp. 45–53. DOI: 10.1016/S0079-6123(05)50004-9 (cit. on p. 43).
- [294] S. Dehaene and J. P. Changeux. "Experimental and Theoretical Approaches to Conscious Processing." *Neuron* 70.2 (2011), pp. 200–27. DOI: 10.1016/j.neuron.2011.03.018 (cit. on p. 43).
- [295] W. J. M. Levelt. "Note on the Distribution of Dominance Times in Binocular Rivalry." *Br. J. Psychol.* 58.1-2 (1967), pp. 143–145. DOI: 10.1111/j.2044-8295.1967.tb01068.x (cit. on p. 45).
- [296] E. M. Meyers. "The Neural Decoding Toolbox." *Frontiers in neuroinformatics* 7 (2013), p. 8. DOI: 10.3389/fninf.2013.00008 (cit. on p. 45).
- [297] A. Dwarakanath et al. "Perisynaptic Activity in the Prefrontal Cortex Reflects Spontaneous Transitions in Conscious Visual Perception." *AREADNE 2018: Research in Encoding And Decoding of Neural Ensembles*. AREADNE Foundation, 2018, p. 58 (cit. on p. 47).
- [298] G. Buzsaki, N. Logothetis, and W. Singer. "Scaling Brain Size, Keeping Timing: Evolutionary Preservation of Brain Rhythms." *Neuron* 80.3 (2013), pp. 751–64. DOI: 10.1016/j.neuron.2013.10.002 (cit. on p. 47).
- [299] S. G. Mallat. *A Wavelet Tour of Signal Processing*. 2nd ed. San Diego: Academic Press, 1999 (cit. on p. 47).
- [300] H. Bokil et al. "Chronux: A Platform for Analyzing Neural Signals." *Journal of neuroscience methods* 192.1 (2010), pp. 146–51. DOI: 10.1016/j.jneumeth.2010.06.020 (cit. on p. 47).
- [301] John H. Holland. "Studying Complex Adaptive Systems." *Jrl Syst Sci & Complex* 19.1 (2006), pp. 1–8. DOI: 10.1007/s11424-006-0001-z (cit. on pp. 51, 52).
- [302] Edward S. Reed. *Encountering the World: Toward an Ecological Psychology*. 1 edition. New York: Oxford University Press, 1996 (cit. on p. 51).

- [303] Yael Niv. "Reinforcement Learning in the Brain." *Journal of Mathematical Psychology*. Special Issue: Dynamic Decision Making 53.3 (2009), pp. 139–154. doi: 10.1016/j.jmp.2008.12.005 (cit. on p. 51).
- [304] Dominik R. Bach and Peter Dayan. "Algorithms for Survival: A Comparative Perspective on Emotions." *Nat Rev Neurosci* 18.5 (2017), pp. 311–319. doi: 10.1038/nrn.2017.35 (cit. on p. 51).
- [305] Yael Niv. *The Primacy of Behavioral Research for Understanding the Brain*. Tech. rep. PsyArXiv, 2020. doi: 10.31234/osf.io/y8mxe (cit. on p. 51).
- [306] Misha Z. Pesenson, ed. *Multiscale Analysis and Nonlinear Dynamics: From Genes to the Brain*. Reviews of Nonlinear Dynamics and Complexity. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2013 (cit. on p. 51).
- [307] Constantinos Siettos and Jens Starke. "Multiscale Modeling of Brain Dynamics: From Single Neurons and Networks to Mathematical Tools." *Wiley Interdiscip Rev Syst Biol Med* 8.5 (2016), pp. 438–458. doi: 10.1002/wsbm.1348 (cit. on p. 51).
- [308] D. Marr and T. Poggio. "From Understanding Computation to Understanding Neural Circuitry." *Neurosci. Res. Program Bull.* 15.3 (1979), pp. 470–488 (cit. on pp. 51, 57).
- [309] Klaas E. Stephan et al. "Translational Perspectives for Computational Neuroimaging." *Neuron* 87.4 (2015), pp. 716–732. doi: 10.1016/j.neuron.2015.07.008 (cit. on p. 51).
- [310] Birte U. Forstmann and Eric-Jan Wagenmakers, eds. *An Introduction to Model-Based Cognitive Neuroscience*. New York: Springer, 2015 (cit. on p. 51).
- [311] Harold J. Morowitz and Jerome L. Singer. *The Mind, The Brain And Complex Adaptive Systems*. Reading, Mass: Westview Press, 1995 (cit. on p. 52).
- [312] Debasish Chowdhury. "Immune Network: An Example of Complex Adaptive Systems." *Artificial Immune Systems and Their Applications*. Ed. by Dipankar Dasgupta. Berlin, Heidelberg: Springer, 1999, pp. 89–104. doi: 10.1007/978-3-642-59901-9_5 (cit. on p. 52).
- [313] Nick C. Ellis and Diane Larsen-Freeman. *Language as a Complex Adaptive System*. John Wiley & Sons, 2009 (cit. on p. 52).
- [314] Jason Brownlee. *Complex Adaptive Systems*. Tech. rep. 070302A. 2007 (cit. on p. 52).
- [315] Murray Gell-Mann. "Complex Adaptive Systems." *Complexity: Metaphors, Models, and Reality*. Ed. by G. Cowan, D. Pines, and D. Meltzer. 19. Reading, MA: Addison-Wesley, 1994, pp. 17–45 (cit. on p. 52).
- [316] W. Brian Arthur et al. *The Economy as an Evolving Complex System II*. Addison-Wesley, 1997 (cit. on p. 52).
- [317] John H. Holland. *Signals and Boundaries: Building Blocks for Complex Adaptive Systems*. English. Illustrated edition. The MIT Press, 2012 (cit. on p. 52).
- [318] Florentin Woergoetter and Bernd Porr. "Reinforcement Learning." *Scholarpedia* 3.3 (2008), p. 1448. doi: 10.4249/scholarpedia.1448 (cit. on p. 53).

- [319] Francesca Mastrogiuseppe and Srdjan Ostojic. "Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks." *Neuron* 99.3 (2018), 609–623.e29. doi: [10.1016/j.neuron.2018.07.003](https://doi.org/10.1016/j.neuron.2018.07.003) (cit. on p. 54).
- [320] Alexis M. Dubreuil et al. "Complementary Roles of Dimensionality and Population Structure in Neural Computations." *bioRxiv* (2020), p. 2020.07.03.185942. doi: [10.1101/2020.07.03.185942](https://doi.org/10.1101/2020.07.03.185942) (cit. on p. 54).
- [321] Erik J. Peterson and Bradley Voytek. "Healthy Oscillatory Coordination Is Bounded by Single-Unit Computation." *bioRxiv* (2018), p. 309427. doi: [10.1101/309427](https://doi.org/10.1101/309427) (cit. on p. 54).
- [322] Xue-Xin Wei and Alan A. Stocker. "Mutual Information, Fisher Information, and Efficient Coding." *Neural Computation* 28.2 (2015), pp. 305–326. doi: [10.1162/NECO_a_00804](https://doi.org/10.1162/NECO_a_00804) (cit. on p. 54).
- [323] Mikhail Prokopenko et al. "Relating Fisher Information to Order Parameters." *Phys. Rev. E* 84.4 (2011), p. 041116. doi: [10.1103/PhysRevE.84.041116](https://doi.org/10.1103/PhysRevE.84.041116) (cit. on p. 54).
- [324] B. C. Daniels et al. "Quantifying Collectivity." *Curr Opin Neurobiol* 37 (2016), pp. 106–113. doi: [10.1016/j.conb.2016.01.012](https://doi.org/10.1016/j.conb.2016.01.012) (cit. on p. 54).
- [325] Alexander C. Kalloniatis, Mathew L. Zuparic, and Mikhail Prokopenko. "Fisher Information and Criticality in the Kuramoto Model of Non-identical Oscillators." *Phys. Rev. E* 98.2 (2018), p. 022302. doi: [10.1103/PhysRevE.98.022302](https://doi.org/10.1103/PhysRevE.98.022302) (cit. on p. 54).
- [326] Eric S. Kuebler et al. "Optimal Fisher Decoding of Neural Activity Near Criticality." *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. doi: [10.1007/978-3-030-20965-0_9](https://doi.org/10.1007/978-3-030-20965-0_9) (cit. on p. 54).
- [327] D. Plenz and T. C. Thiagarajan. "The Organizing Principles of Neuronal Avalanches: Cell Assemblies in the Cortex?" *Trends in neurosciences* 30 (2007), pp. 101–10. doi: [10.1016/j.tins.2007.01.005](https://doi.org/10.1016/j.tins.2007.01.005) (cit. on p. 54).
- [328] W. Singer et al. "Formation of Cortical Cell Assemblies." *Cold Spring Harb Sym* 55 (1990), pp. 939–52 (cit. on p. 54).
- [329] K. D. Harris et al. "Organization of Cell Assemblies in the Hippocampus." *Nature* 424 (2003), pp. 552–6. doi: [10.1038/nature01834](https://doi.org/10.1038/nature01834) (cit. on p. 54).
- [330] K. D. Harris. "Neural Signatures of Cell Assembly Organization." *en. Nat Rev Neurosci* 6.5 (2005), pp. 399–407. doi: [10.1038/nrn1669](https://doi.org/10.1038/nrn1669) (cit. on p. 54).
- [331] G. Buzsaki. "Neural Syntax: Cell Assemblies, Synapsemes, and Readers." *Neuron* 68.3 (2010), pp. 362–85. doi: [10.1016/j.neuron.2010.09.023](https://doi.org/10.1016/j.neuron.2010.09.023) (cit. on p. 54).
- [332] Christian Tetzlaff et al. "The Use of Hebbian Cell Assemblies for Nonlinear Computation." *Sci. Rep.* 5.1 (2015), p. 12866. doi: [10.1038/srep12866](https://doi.org/10.1038/srep12866) (cit. on p. 54).
- [333] Sean E. Cavanagh et al. "A Circuit Mechanism for Irrationalities in Decision-Making and NMDA Receptor Hypofunction: Behaviour, Computational Modelling, and Pharmacology." *bioRxiv* (2019), p. 826214. doi: [10.1101/826214](https://doi.org/10.1101/826214) (cit. on p. 55).

- [334] Michael J. Frank. "Linking Across Levels of Computation in Model-Based Cognitive Neuroscience." *An Introduction to Model-Based Cognitive Neuroscience*. Ed. by Birte U. Forstmann and Eric-Jan Wagenmakers. New York, NY: Springer, 2015, pp. 159–177. doi: 10.1007/978-1-4939-2236-9_8 (cit. on p. 55).
- [335] R. Moreno-Bote, J. Rinzel, and N. Rubin. "Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability." *Journal of neurophysiology* 98.3 (2007), pp. 1125–39. doi: 10.1152/jn.00116.2007 (cit. on p. 55).
- [336] A. Shapiro et al. "Balance between Noise and Adaptation in Competition Models of Perceptual Bistability." *Journal of computational neuroscience* 27.1 (2009), pp. 37–54. doi: 10.1007/s10827-008-0125-3 (cit. on p. 55).
- [337] Shashaank Vattikuti et al. "Canonical Cortical Circuit Model Explains Rivalry, Intermittent Rivalry, and Rivalry Memory." *PLOS Computational Biology* 12.5 (2016), e1004903. doi: 10.1371/journal.pcbi.1004903 (cit. on p. 55).
- [338] Benjamin P. Cohen, Carson C. Chow, and Shashaank Vattikuti. "Dynamical Modeling of Multi-Scale Variability in Neuronal Competition." *Commun Biol* 2.1 (2019), pp. 1–11. doi: 10.1038/s42003-019-0555-7 (cit. on p. 55).
- [339] Peter Dayan. "A Hierarchical Model of Binocular Rivalry." *Neural Comput.* 10.5 (1998), pp. 1119–1135. doi: 10.1162/089976698300017377 (cit. on pp. 55, 56).
- [340] J. Hohwy, A. Roepstorff, and K. Friston. "Predictive Coding Explains Binocular Rivalry: An Epistemological Review." *Cognition* 108.3 (2008), pp. 687–701. doi: 10.1016/j.cognition.2008.05.010 (cit. on pp. 55, 56).
- [341] G. S. Atwal. "Statistical Mechanics of Multistable Perception." *BioRxiv* (2014). doi: 10.1101/008177 (cit. on pp. 55, 56).
- [342] Samuel Gershman, Ed Vul, and Joshua B. Tenenbaum. "Perceptual Multistability as Markov Chain Monte Carlo Inference." *Advances in Neural Information Processing Systems*. 2014, pp. 611–619 (cit. on pp. 55, 56).
- [343] D. C. Knill and A. Pouget. "The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation." *Trends in neurosciences* 27 (2004), pp. 712–9. doi: 10.1016/j.tins.2004.10.007 (cit. on p. 55).
- [344] Kenji Doya et al. *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press, 2007 (cit. on p. 55).
- [345] P. C. Klink, R. J. A. van Ee, and R. van Ee. "United We Sense, Divided We Fail: Context-Driven Perception of Ambiguous Visual Stimuli." *Philos Trans R Soc Lond B Biol Sci* 367.1591 (2012), pp. 932–941. doi: 10.1098/rstb.2011.0358 (cit. on p. 55).
- [346] O Carter et al. "Psilocybin Slows Binocular Rivalry Switching through Serotonin Modulation." en (), p. 1 (cit. on p. 55).
- [347] Jeff Mentch et al. "GABAergic Inhibition Gates Perceptual Awareness During Binocular Rivalry." *J. Neurosci.* 39.42 (2019), pp. 8398–8407. doi: 10.1523/JNEUROSCI.0836-19.2019 (cit. on p. 55).

- [348] S. M. Miller et al. "Genetic Contribution to Individual Variation in Binocular Rivalry Rate." *Proceedings of the National Academy of Sciences* 107.6 (2010), pp. 2664–2668. doi: 10.1073/pnas.0912149107 (cit. on p. 55).
- [349] Trung T. Ngo et al. "Psychiatric and Genetic Studies of Binocular Rivalry: An Endophenotype for Bipolar Disorder?" *Acta Neuropsychiatr.* 23.1 (2011), pp. 37–42. doi: 10.1111/j.1601-5215.2010.00510.x (cit. on p. 55).
- [350] Phillip C.F. Law, Steven M. Miller, and Trung T. Ngo. "The Effect of Stimulus Strength on Binocular Rivalry Rate in Healthy Individuals: Implications for Genetic, Clinical and Individual Differences Studies." *Physiology & Behavior* 181 (2017), pp. 127–136. doi: 10.1016/j.physbeh.2017.08.023 (cit. on p. 55).
- [351] Biqing Chen et al. "Genomic Analyses of Visual Cognition: Perceptual Rivalry and Top-Down Control." *J. Neurosci.* 38.45 (2018), pp. 9668–9678. doi: 10.1523/JNEUROSCI.1970-17.2018 (cit. on p. 55).
- [352] T. Womelsdorf et al. "Burst Firing Synchronizes Prefrontal and Anterior Cingulate Cortex during Attentional Control." *Current biology : CB* 24 (2014), pp. 2613–21. doi: 10.1016/j.cub.2014.09.046 (cit. on p. 56).
- [353] Pantelis Leptourgos. "Dynamical Circular Inference in the General Population and the Psychosis Spectrum : Insights from Perceptual Decision Making." Thesis. Paris Sciences et Lettres, 2018 (cit. on p. 56).
- [354] F. Biessmann et al. "Analysis of Multimodal Neuroimaging Data." *IEEE Rev Biomed Eng* 4 (2011), pp. 26–58. doi: 10.1109/RBME.2011.2170675 (cit. on p. 56).
- [355] S. Fazli et al. "Learning From More Than One Data Source: Data Fusion Techniques for Sensorimotor Rhythm-Based Brain-Computer Interfaces." *P IEEE* 103 (2015), pp. 891–906. doi: 10.1109/Jproc.2015.2413993 (cit. on p. 56).
- [356] Mengsen Zhang et al. "Topological Portraits of Multiscale Coordination Dynamics." *Journal of Neuroscience Methods* 339 (2020), p. 108672. doi: 10.1016/j.jneumeth.2020.108672 (cit. on p. 56).
- [357] L. Shpigelman et al. "Spikernels: Predicting Arm Movements by Embedding Population Spike Rate Patterns in Inner-Product Spaces." *Neural computation* 17.3 (2005), pp. 671–90. doi: 10.1162/0899766053019944 (cit. on p. 56).
- [358] A. R. Paiva, I. Park, and J. C. Principe. "A Reproducing Kernel Hilbert Space Framework for Spike Train Signal Processing." *Neural computation* 21.2 (2009), pp. 424–49. doi: 10.1162/neco.2008.09-07-614 (cit. on p. 56).
- [359] A. R. C. Paiva, I. Park, and J. C. Principe. "Inner Products for Representation and Learning in the Spike Train Domain." *Stat. Signal Process. Neurosci. Neurotechnology* (2010), pp. 265–309. doi: 10.1016/B978-0-12-375027-3.00008-9 (cit. on p. 56).
- [360] L. Li et al. "A Tensor-Product-Kernel Framework for Multiscale Neural Activity Decoding and Control." *Computational intelligence and neuroscience* 2014 (2014), p. 870160. doi: 10.1155/2014/870160 (cit. on p. 56).

- [361] Il Memming Park et al. "Kernel Methods on Spike Train Space for Neuroscience: A Tutorial." *arXiv* (2013) (cit. on p. 56).
- [362] Quentin J. M. Huys et al. "Advances in the Computational Understanding of Mental Illness." *Neuropsychopharmacology* (2020), pp. 1–17. doi: 10.1038/s41386-020-0746-4 (cit. on pp. 56, 57).
- [363] Margit Burmeister, Melvin G. McInnis, and Sebastian Zöllner. "Psychiatric Genetics: Progress amid Controversy." *Nat. Rev. Genet.* 9.7 (2008), pp. 527–540. doi: 10.1038/nrg2381 (cit. on p. 57).
- [364] Orna Issler and Alon Chen. "Determining the Role of microRNAs in Psychiatric Disorders." *Nat. Rev. Neurosci.* 16.4 (2015), pp. 201–212. doi: 10.1038/nrn3879 (cit. on p. 57).
- [365] Olav B. Smeland et al. "The Polygenic Architecture of Schizophrenia — Rethinking Pathogenesis and Nosology." *Nat. Rev. Neurol.* (2020), pp. 1–14. doi: 10.1038/s41582-020-0364-0 (cit. on p. 57).
- [366] Leonhard Schilbach. "Towards a Second-Person Neuropsychiatry." *Philos Trans R Soc Lond B Biol Sci* 371.1686 (2016). doi: 10.1098/rstb.2015.0081 (cit. on p. 57).
- [367] Victoria Leong and Leonhard Schilbach. "The Promise of Two-Person Neuroscience for Developmental Psychiatry: Using Interaction-Based Sociometrics to Identify Disorders of Social Interaction." *Br. J. Psychiatry* 215.5 (2019), pp. 636–638. doi: 10.1192/bj.p.2019.73 (cit. on p. 57).
- [368] Meltem Sevgi et al. "Social Bayes: Using Bayesian Modeling to Study Autistic Trait-Related Differences in Social Cognition." *Biological Psychiatry. Molecular Mechanisms of Neurodevelopmental Disorders* 87.2 (2020), pp. 185–193. doi: 10.1016/j.biopsych.2019.09.032 (cit. on p. 57).
- [369] A. David Redish and Joshua A. Gordon, eds. *Computational Psychiatry: New Perspectives on Mental Illness*. Strüngmann Forum Reports. Cambridge, Massachusetts: The MIT Press, 2016 (cit. on p. 57).
- [370] Golam M Khandaker et al. "Inflammation and Immunity in Schizophrenia: Implications for Pathophysiology and Treatment." *The Lancet Psychiatry* 2.3 (2015), pp. 258–270. doi: 10.1016/S2215-0366(14)00122-9 (cit. on p. 57).
- [371] Edward Bullmore. *The Inflamed Mind: A Radical New Approach to Depression*. 2018 (cit. on p. 57).
- [372] Antonio L. Teixeira and Moises E. Bauer. *Immunopsychiatry: A Clinician's Introduction to the Immune Basis of Mental Disorders*. Oxford University Press, 2019 (cit. on p. 57).
- [373] Ning Yuan et al. "Inflammation-Related Biomarkers in Major Psychiatric Disorders: A Cross-Disorder Assessment of Reproducibility and Specificity in 43 Meta-Analyses." *Transl Psychiatry* 9.1 (2019), pp. 1–13. doi: 10.1038/s41398-019-0570-y (cit. on p. 57).
- [374] Andreas Mayer. "Optimal Immune Systems : A Ressource Allocation and Information Processing View of Immune Defense." Theses. PSL Research University, 2017 (cit. on p. 57).
- [375] Maya Schiller, Tamar L. Ben-Shaanan, and Asya Rolls. "Neuronal Regulation of Immunity: Why, How and Where?" *Nat. Rev. Immunol.* (2020), pp. 1–17. doi: 10.1038/s41577-020-0387-1 (cit. on p. 57).

- [376] Faraj Haddad, Salonee Patel, and Susanne Schmid. "Maternal Immune Activation by Poly I:C as a Preclinical Model for Neurodevelopmental Disorders: A Focus on Autism and Schizophrenia." *Neuroscience & Biobehavioral Reviews* (2020). doi: 10.1016/j.neubiorev.2020.04.012 (cit. on p. 57).
- [377] Golam M Khandaker, Urs Meyer, and Peter B Jones. *Neuroinflammation and Schizophrenia*. 2020 (cit. on p. 57).
- [378] Murilo S. de Abreu et al. "Psychoneuroimmunology and Immunopsychiatry of Zebrafish." *Psychoneuroendocrinology* 92 (2018), pp. 1–12. doi: 10.1016/j.psyneuen.2018.03.014 (cit. on p. 57).
- [379] Ana Badimon et al. "Negative Feedback Control of Neuronal Activity by Microglia." *Nature* (2020), pp. 1–7. doi: 10.1038/s41586-020-2777-8 (cit. on p. 57).
- [380] Thomas Pfeiffer and David Attwell. "Brain's Immune Cells Put the Brakes on Neurons." *Nature* (2020). doi: 10.1038/d41586-020-02713-7 (cit. on p. 57).
- [381] Grant S. Shields, Chandler M. Spahr, and George M. Slavich. "Psychosocial Interventions and Immune System Function: A Systematic Review and Meta-analysis of Randomized Clinical Trials." *JAMA Psychiatry* 77.10 (2020), pp. 1031–1043. doi: 10.1001/jamapsychiatry.2020.0431 (cit. on p. 57).
- [382] Donovan M et al. "Social Isolation Alters Behavior, the Gut-Immune-Brain Axis, and Neurochemical Circuits in Male and Female Prairie Voles." *Neurobiology of Stress* (2020), p. 100278. doi: 10.1016/j.ynstr.2020.100278 (cit. on p. 57).
- [383] Josephine Heine et al. "Transdiagnostic Hippocampal Damage Patterns in Neuroimmunological Disorders." *NeuroImage: Clinical* 28 (2020), p. 102515. doi: 10.1016/j.nicl.2020.102515 (cit. on p. 57).
- [384] Tamar Koren et al. "Remembering Immunity: Neuronal Ensembles in the Insular Cortex Encode and Retrieve Specific Immune Responses." *bioRxiv* (2020), p. 2020.12.03.409813. doi: 10.1101/2020.12.03.409813 (cit. on p. 57).
- [385] Adi Kol and Inbal Goshen. "The Memory Orchestra: The Role of Astrocytes and Oligodendrocytes in Parallel to Neurons." *Current Opinion in Neurobiology* 67 (2021), pp. 131–137. doi: 10.1016/j.conb.2020.10.022 (cit. on p. 57).
- [386] Sophie Deneve and Renaud Jardri. "Circular Inference: Mistaken Belief, Misplaced Trust." *Current Opinion in Behavioral Sciences. Computational Modeling* 11 (2016), pp. 40–48. doi: 10.1016/j.cobeha.2016.04.001 (cit. on p. 57).
- [387] Shervin Safavi et al. *Signatures of Criticality in Efficient Coding Networks*. 2023. doi: 10.1101/2023.02.14.528465 (cit. on p. 91).

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. [[228](#), Neural Computation 2021]
2. Safavi et al. [[129](#), PLoS Computational Biology 2023]
3. Besserve et al.; preliminary manuscript is available in the appendix, ([Paper 3](#))
4. Safavi et al. [[387](#), BioRxiv 2023]
5. Safavi et al. [[209](#), Front. Psychol. 2014]
6. Safavi et al. [[126](#), PNAS 2018]
7. Kapoor et al. [[195](#), Nature Communications 2022]
8. Dwarakanath et al. [[223](#), Neuron 2023]

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

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 $j \in \{1, 2, \dots, N\}$ (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

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

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

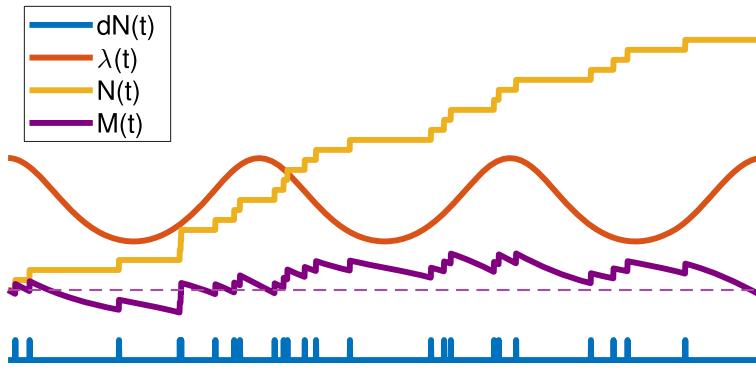


Figure 1: Doob-Meyer decomposition for an example inhomogeneous Poisson process with oscillatory rate $\lambda(t)$ of frequency $f = 1$ Hz; average firing rate $\lambda_0 = 5$ Hz (dashed line indicates the reference 0). See section 3.3 for details of the simulation.

$M = \{M(t); t \in [0, \tau]\}$ is a zero-mean martingale relative¹ 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:

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

Consider now a (univariate) counting process $\{(N(t), \mathcal{F}_t); t \geq 0\}$, 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,

$$N(t) = M(t) + \int_0^t \lambda(t)dt, \quad (2.3)$$

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),

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

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, \hat{c} , can then be decomposed as

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

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 are²

$$\begin{aligned} c^* &\triangleq \mathbb{E}[\hat{c}] = \int_0^T x(t)\lambda(t)dt \quad \text{and} \\ \text{Var}[\hat{c}] &= \mathbb{E}[|\hat{c} - c^*|^2] = \int_0^T |x^2(t)|\lambda(t)dt. \end{aligned} \quad (2.6)$$

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 \hat{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 $\hat{\mathbf{C}}$ that gathers all pairwise measurements. However, such task is nontrivial due to the martingale fluctuations affecting $\hat{\mathbf{C}}$, 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

²See section B.1.1 for more details.

real symmetric matrix 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),

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

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

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

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 $\{X_n\}_{n>0}$ of dimension $p \times n$ such that $\frac{p}{n} \xrightarrow[n \rightarrow +\infty]{} \alpha \leq 1$, with coefficients sampled independently and identically distributed (i.i.d.) from a (possibly complex) standard Normal distribution, the ESD of the Wishart matrix $S_n = \frac{1}{n} X_n X_n^H$ (where \cdot^H indicates the transposed complex conjugate) converges to the Marchenko-Pastur (MP) law $\mu_{MP}(x)$ (Marchenko & Pastur, 1967) with density

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

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 $\{X_n\}_{n>0}$ 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

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

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 \hat{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,*

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

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

$$\sqrt{K} (\widehat{c}_K - c^*) \xrightarrow[K \rightarrow +\infty]{} \mathcal{N} \left(0, \int_0^T x^2(t) \lambda(t) dt \right).$$

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. \square

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

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

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 multi-trial 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 $\widehat{\text{PLV}}_K$ estimated from K trials of measurements on $[0, T]$ tends to the limit

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

Moreover, as $K \rightarrow +\infty$, the residual,

$$\sqrt{K} (\widehat{\text{PLV}}_K - \text{PLV}^*), \quad (3.3)$$

³This allows the normalization factor to converge to a deterministic quantity as $K \rightarrow +\infty$ 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

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

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 $\tilde{V}(t)$ in theorem 1 by a covariance matrix that can be assessed with martingale results given in section B.1.1. \square

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

$$\lambda(t) = \lambda_0 (1 + \varkappa \cos(\phi(t) - \varphi_0)), \quad \lambda_0 > 0, \quad 0 \leq \varkappa \leq 1,$$

we get easily with trigonometric identities that $\text{PLV}^* = \frac{1}{2} \varkappa e^{i\varphi_0}$ and the residual of equation 3.3 converges to an isotropic complex gaussian of total variance⁴ $\frac{1}{\lambda_0 T}$ 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

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

the PLV^* vanishes and the residual's variance remains the same. These properties make PLV straightforward 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 $\kappa \geq 0$ 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

⁴The 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,

$$\lambda(t) = \lambda_0 \exp(\kappa \cos(\phi(t) - \varphi_0)). \quad (3.4)$$

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

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

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 non-linearly increasing phase $\phi(t)$ and a point process with intensity

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

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).

⁵They 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 $K \rightarrow +\infty$) to

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

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

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

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

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

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. \square

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 $\widehat{\text{PLV}}_K$ converges (for $K \rightarrow +\infty$) to*

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

and the scaled residual,

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

converges to a zero mean complex gaussian:

$$\sqrt{K} (\widehat{\text{PLV}}_K - \text{PLV}^*) \xrightarrow{K \rightarrow +\infty} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{Cov}(Z) \right),$$

with the following covariance:

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

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. \square

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.

⁶In 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 ,

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

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

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

where $\gamma_T = T f$ 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:

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

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

⁷To 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,

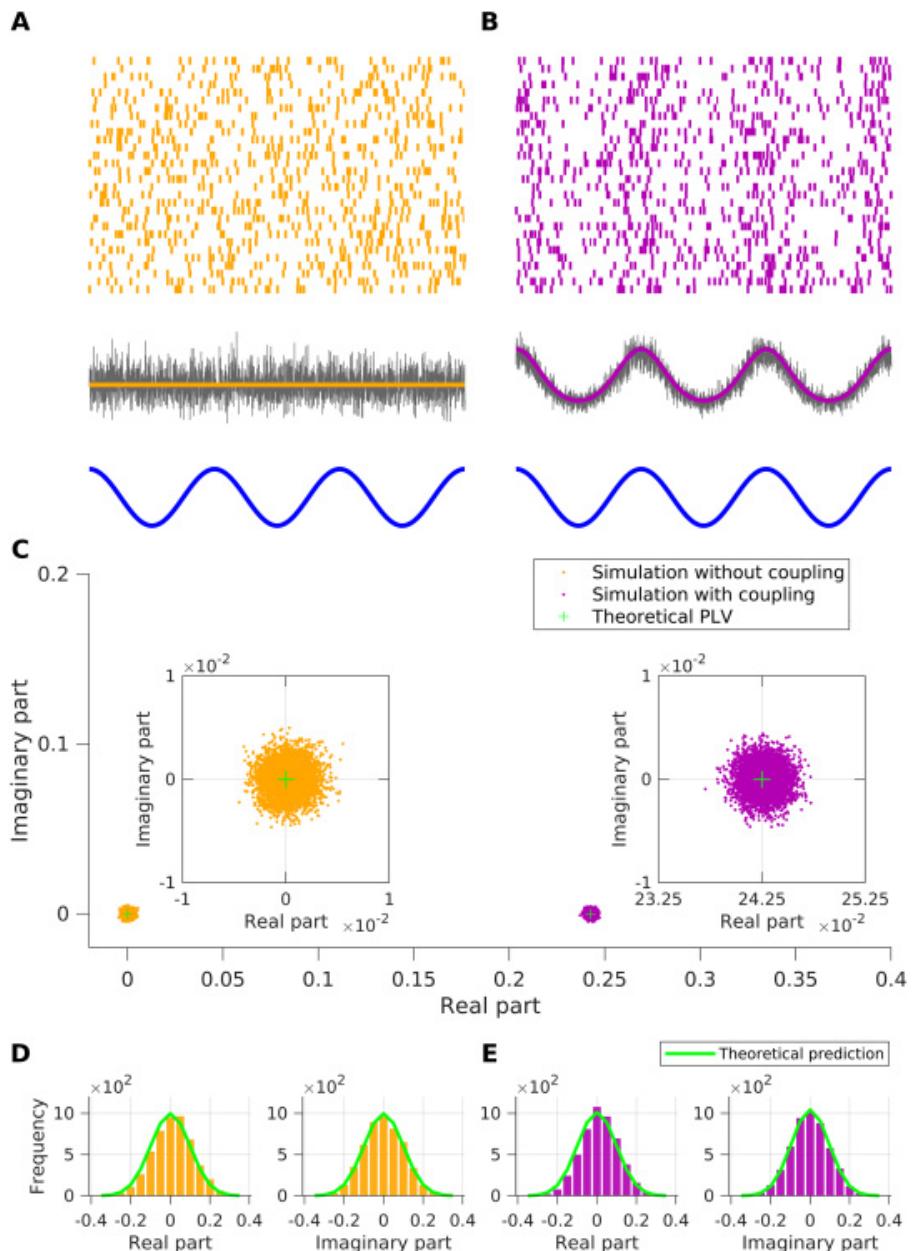


Figure 2: Simulation of (A) homogeneous Poisson spike trains and (B) phase-locked spike train with Poisson statistics (von Mises model with $\kappa = 0.5$). First row: Example raster plot of the spikes. Second row: Empirical firing rate (gray line) and ground truth firing rate (orange and purple traces). 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.

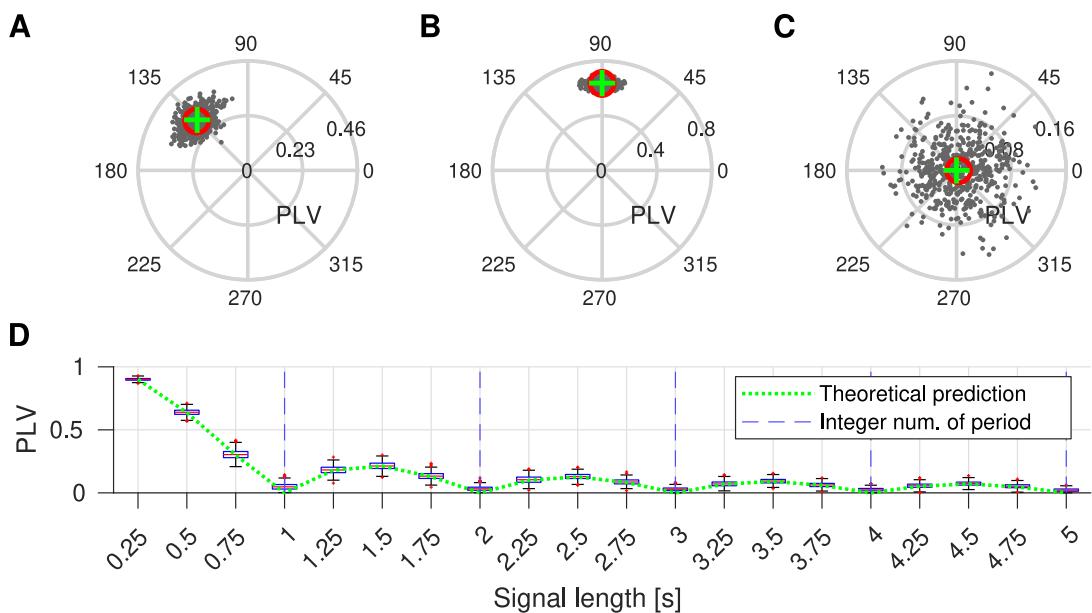


Figure 3: (A–C) Distribution of simulated complex-valued PLVs (gray dots), average of the simulated PLVs (red circle), and theoretical prediction based on equation 3.11 (green crosses) for (A) $\gamma_T = 0.75$, (B) $\gamma_T = 0.5$, and (C) $\gamma_T = 1$. All complex-valued PLVs are represented in 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; Mastrogiossepe & Ostoic, 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 C^* between an n -dimensional vector of counting processes N with associated intensity vector $\lambda(t)$ and a multivariate p -dimensional signal $x(t)$, and its estimate based on independent trials \widehat{C}_K , respectively defined as

$$C^* = \int_0^T x(t)\lambda(t)^\top dt \quad \text{and} \quad \widehat{C}_K = \frac{1}{K} \sum_{k=1}^K \int_0^T x(t)dN^{(k)}(t)^\top. \quad (4.1)$$

In this multivariate setting, the coupling matrix between the point process and continuous signal can be characterized by the singular value(s) of C^* ,

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0,$$

and associated orthonormal singular vectors $\{(u_k, v_k)\}$, such that

$$C^* = \sum_{k=1}^p u_k \sigma_k v_k^H.$$

When the dimension of the coupling matrix gets large, recovering the entire structure of C^* using its estimate \widehat{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{C}_K by analyzing the eigenvalues of the hermitian matrix $\frac{1}{n}\widehat{C}_K\widehat{C}_K^H$ denoted

$$\ell_1 \geq \ell_2 \geq \dots \geq \ell_p \geq 0.$$

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:

$$\delta_{lj} = \begin{cases} 1, & \text{if } l = j, \\ 0, & \text{otherwise.} \end{cases} \quad (4.2)$$

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 $\mathbf{x}(t) = (x_j)_{j=1 \dots p}$ and K independent copies of multivariate Poisson process $\mathbf{N}(t) = (N_i)_{i=1 \dots n}$ with rate vector $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_n]^\top$. Then the normalized coupling matrix $\sqrt{K} \widehat{\mathbf{C}}_K \text{diag}(\sqrt{T} \boldsymbol{\lambda})^{-1}$, with $\widehat{\mathbf{C}}_K$ given by equation 4.1, converges in distribution for $K \rightarrow +\infty$ 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 \mathbf{x} entail vanishing correlations between all matrix coefficients and lead to the analytical expression of the covariance matrix. \square

This result suggests that for large n and $p = p(n)$, coupling matrices $\widehat{\mathbf{C}}_K^n$ of increasing size can be used to build the Wishart-like matrix sequence,

$$S_n \triangleq \frac{K}{n} \widehat{\mathbf{C}}_K^n \text{diag}(T \boldsymbol{\lambda})^{-1} (\widehat{\mathbf{C}}_K^n)^H, \quad (4.3)$$

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 $\frac{p(n)}{n} \xrightarrow[n \rightarrow +\infty]{} \alpha \in (0, +\infty)$, and*

$$\frac{1}{n^2 K(n)^2} \sum_{\Gamma} \left(\int_0^T \bar{x}_j x_l x_{j'} \bar{x}_{l'} dt \right)^2 \rightarrow 0, \text{ uniformly in } k \leq n, \quad (4.4)$$

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 $\{\widehat{C}_{K(n)}^n\}_{n \in \mathbb{N}^*}$ built as in theorem 2 for $p = p(n)$; then the corresponding sequence $\{S_n\}$ 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). \square

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 $x(t)$ and given the number of terms in the sum bounded by n^4 , we can already see that $\frac{n}{K(n)} \rightarrow 0$, 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, $x_l(t) = \frac{1}{\sqrt{T}} \exp(i2\pi lt/T)$. 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 $\frac{\sqrt{n}}{K(n)} \rightarrow 0$, 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*

$$\frac{p(n)}{n} \rightarrow \alpha \in (0, +\infty) \quad \text{and} \quad \frac{1}{K(n)} \sum_{1 \leq i, k \leq p(n)} \int_0^T |x_i x_j|^2(t) dt < B, \quad (4.5)$$

for some constant B . Then 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 $\{S_n\}$, respectively. Then in probability

$$\ell_1(n) \rightarrow (1 + \sqrt{\alpha})^2 \quad \text{and} \quad \ell_p(n) \rightarrow (1 - \sqrt{\alpha})^2 \mathbf{1}_{\alpha < 1}.$$

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. \square

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 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 $M_n = (\frac{1}{\sqrt{n}}X_n + A_n)(\frac{1}{\sqrt{n}}X_n + A_n)^H$. Then as $n \rightarrow \infty$ and $\frac{p}{n} \rightarrow \alpha \in (0, 1)$, almost surely,*

$$\lambda_i(M_n) \rightarrow \begin{cases} \frac{(1+\theta_i)(c+\theta_i)}{\theta_i}, & \text{if } \theta_i > \sqrt{\alpha}, \\ (1 + \sqrt{\alpha})^2, & \text{otherwise.} \end{cases}$$

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, x may represent LFP measurements collected on each recording channel and 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{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 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 neurophysiology data). We assume the rows of 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 X , the discrete time approximation of $x(t)$, according to

$$X = WL, \quad \text{with} \quad W = \left(\frac{1}{q} LL^H \right)^{-\frac{1}{2}}, \quad (4.6)$$

where the power in the expression of the whitening matrix 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.

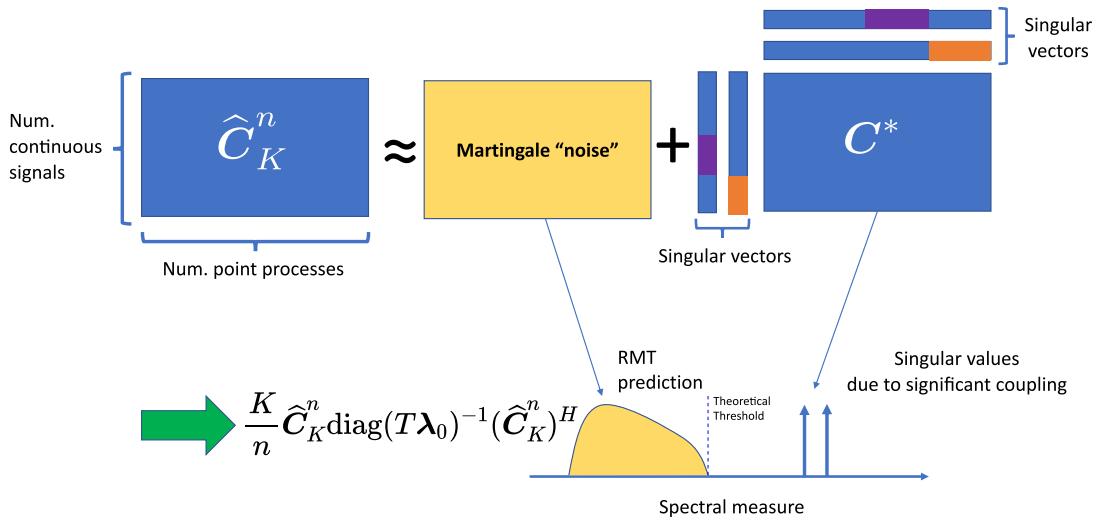


Figure 4: We assume \widehat{C}_K^n is a superposition of martingale noise and a low-rank deterministic matrix C^* reflecting the actual coupling. If the singular values of a normalized version of C^* are large enough (larger than the upper end of the MP law support), theory suggests that they will correspond 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,

$$S_n = \frac{K}{n} \widehat{C}_K^n \text{diag}(T\lambda_0)^{-1} (\widehat{C}_K^n)^H.$$

The null hypothesis of nonsignificance of the k th largest singular value $\widehat{\sigma}_k$ of the normalized coupling matrix,

$$\sqrt{K} \widehat{C}_K^n \text{diag}(\sqrt{T\lambda_0})^{-1},$$

should thus be rejected if the corresponding k th largest eigenvalue ℓ_k of S_n is superior to the significance threshold, leading to the condition

$$\widehat{\sigma}_k = \sqrt{n\ell_k} > \sqrt{n\theta_{DET}} = \sqrt{n}(1 + \sqrt{\alpha}). \quad (4.7)$$

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 multivariate 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 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 $O_j(t) = e^{2\pi i f_j t}$, $j \in \{1, \dots, N_{osc}\}$, 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 $t = \{k\Delta\}_{k=1, \dots, q}$. The synthesized discrete time multichannel LFP ($\Psi[t] = \{\psi_l[t]\}_{l=1, \dots, n_c}$) can be written as

$$\Psi_l[t] = O_{j(l)}[t] \odot \exp(i\eta_l[t]), \quad (4.8)$$

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{C}_K , we first preprocess $\Psi[t]$ 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 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 X , the discrete time version of $x(t)$. The coupling matrix \widehat{C}_K is then computed according to equation 4.1 using 10 trials (barring trivial approximation to the closest time sample in X).

Then in order to approximate the normalization $\sqrt{K} \widehat{C}_K^n \text{diag}(\sqrt{T \lambda_0})^{-1}$ based on empirical data, we use the total number of events for unit u occurring across all K trials $N_{\text{tot}}^u = \sum_{k=1}^K N_k^u$ and multiply each column u of the coupling matrix by

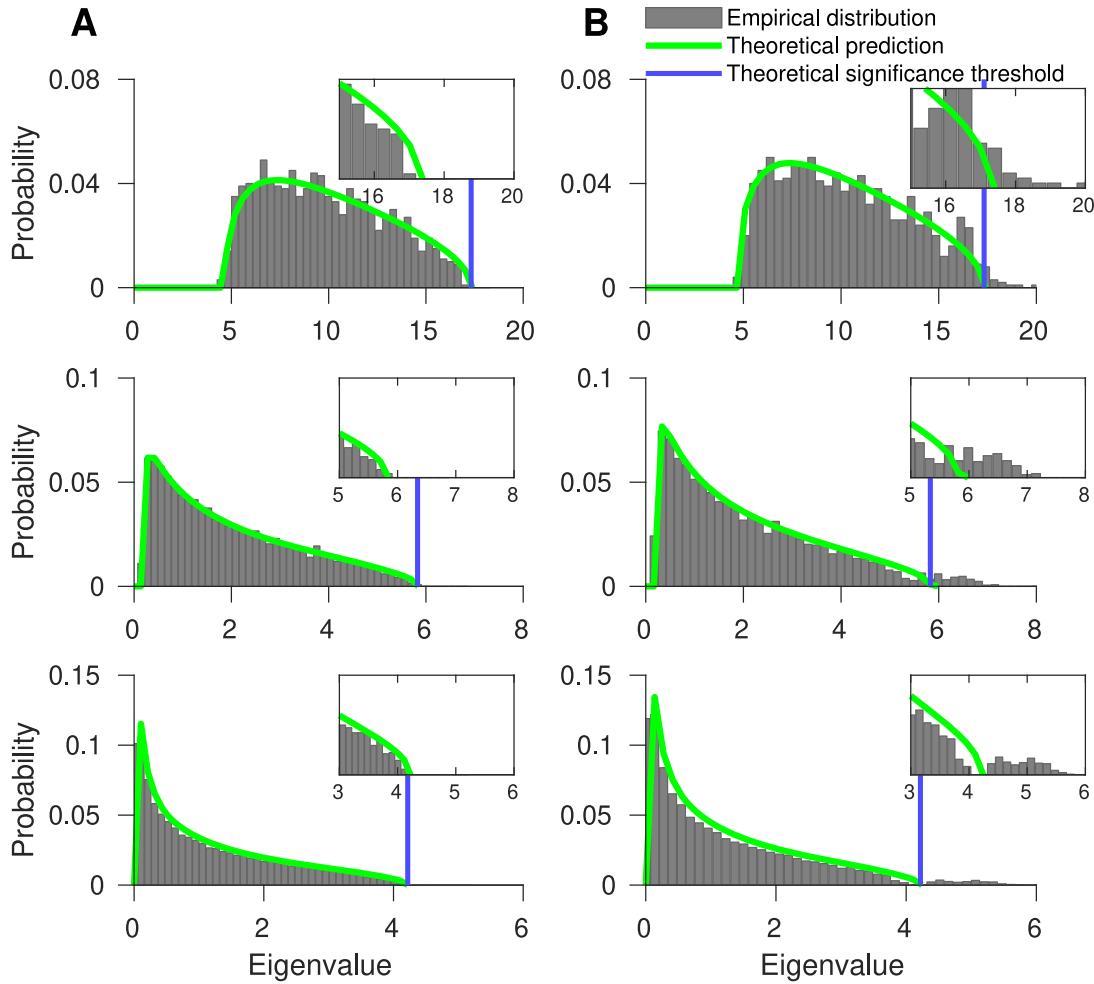


Figure 5: Theoretical Marchenko-Pastur distribution (green lines) and empirical distribution (gray bars) for (A) simulation without coupling ($\kappa = 0$) and (B) with coupling ($\kappa = 0.15$) between multivariate spikes and LFP. Rows represent the spectral distribution of simulations with different 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.

$$\frac{K}{\sqrt{N_{\text{tot}}^u}} \approx \frac{K}{\sqrt{K \int_0^T \lambda_u(t) dt}}$$

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 $x(t)$, 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 (Kruskin 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 recentered⁸ 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

$$\hat{c}_K = \frac{1}{K} \sum_{k=1}^K \int_0^T x(t) dM^{(k)}(t) + \frac{1}{K} \sum_{k=1}^K \int_0^T x(t) \lambda(t) dt(t). \quad (\text{A.1})$$

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

⁸The 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:

$$\mathbb{E}[\widehat{c}_K] = \int_0^T x(t)\lambda(t)dt(t).$$

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 $\tilde{M}^{(K)}$ sequence that will satisfy the conditions described in Aalen et al. (2008, p. 63) (\xrightarrow{P} indicate convergence in probability):

1. $\text{Var}(\tilde{M}^{(K)}(t)) \xrightarrow[K \rightarrow +\infty]{P} \tilde{V}(t)$ for all $t \in [0, T]$, with \tilde{V} increasing and $\tilde{V}(0) = 0$.
2. Informally, the size of the jumps of $\tilde{M}^{(K)}$ tends to zero (see Aalen et al., 2008, p. 63). Formally, for any $\epsilon > 0$, the martingale $\tilde{M}_\epsilon^{(K)}(t)$ gathering the jumps $> \epsilon$ satisfies $\text{Var}(\tilde{M}_\epsilon^{(K)}(t)) \xrightarrow[K \rightarrow +\infty]{P} 0$.

Then $\tilde{M}^{(K)}(t)$ converges in distribution to a gaussian martingale of variance $\tilde{V}(t)$.

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 $\tilde{M}^{(K)} = 1/\sqrt{K} \sum_{k=1}^K M_x^{(k)}$.

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,

$$\tilde{V}(t) = \text{Var}(\tilde{M}^{(K)}(t)) = \text{Var}\left(\frac{1}{\sqrt{K}} \sum_{k=1}^K M_x^{(k)}(t)\right) = \sum_{k=1}^K \text{Var}\left(\frac{1}{\sqrt{K}} M_x(t)\right), \quad (\text{A.2})$$

and using equation B.4, we get

$$\tilde{V}(t) = \frac{1}{K} \sum \int_0^t x^2(t)\lambda(t)dt = \int_0^t x^2(t)\lambda(t)dt. \quad (\text{A.3})$$

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 , $\tilde{M}_\epsilon^{(K)}(t)$ 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 $\tilde{V}(t)$;

$$\tilde{M}^{(K)} \xrightarrow[K \rightarrow +\infty]{} \mathcal{N} \left(0, \int_0^T x^2(t) \lambda(t) dt \right). \quad (\text{A.4})$$

Finally, using equation A.1, we conclude the proof by noticing that the above martingale corresponds exactly to the quantity $\sqrt{K}(\hat{c}_K - c^*)$. Therefore,

$$\sqrt{K}(\hat{c}_K - c^*) \xrightarrow[K \rightarrow +\infty]{} \mathcal{N} \left(0, \int_0^T x^2(t) \lambda(t) dt \right). \quad (\text{A.5})$$

□

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 $\tilde{V}(t)$. The covariance between a martingale's real part,

$$M_{\text{Re}}(t) = \int_0^t \text{Re}(e^{i\phi(s)}) dM(s) ds,$$

and imaginary part,

$$M_{\text{Im}}(t) = \int_0^t \text{Im}(e^{i\phi(s)}) dM(s) ds,$$

is given by

$$\int_0^t \text{Re}(e^{i\phi(s)}) \text{Im}(e^{i\phi(s)}) \lambda(s) ds. \quad (\text{A.6})$$

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

$$\begin{aligned} & \text{Cov} \left(\begin{bmatrix} \text{Re}\{Z\} \\ \text{Im}\{Z\} \end{bmatrix} \right) \\ &= \begin{bmatrix} \int_0^t (\text{Re}(e^{i\phi(s)}))^2 \lambda(s) ds & \int_0^t \text{Re}(e^{i\phi(s)}) \text{Im}(e^{i\phi(s)}) \lambda(s) ds \\ \int_0^t \text{Re}(e^{i\phi(s)}) \text{Im}(e^{i\phi(s)}) \lambda(s) ds & \int_0^t (\text{Im}(e^{i\phi(s)}))^2 \lambda(s) ds \end{bmatrix} \quad (\text{A.7}) \end{aligned}$$

$$= \int_0^t \begin{bmatrix} \cos^2(\phi(s)) & \sin(2\phi(s))/2 \\ \sin(2\phi(s))/2 & \sin^2(\phi(s)) \end{bmatrix} \lambda(s) ds. \quad (\text{A.8})$$

Similar to theorem 1, as $K \rightarrow +\infty$, 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):

$$\sqrt{K}(\widehat{c}_K - c^*) \xrightarrow[K \rightarrow +\infty]{} \mathcal{N}(0, \text{Cov}(Z)).$$

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{c}_K is not exactly the multitrial PLV estimate—more precisely,

$$\widehat{c}_K = \frac{1}{K} \sum_{k=1}^K \int_0^T e^{i\phi(t)} dN^{(k)}(t) = \frac{1}{K} \sum_{k=1}^K \sum_{j=1}^{N_k} e^{i\phi(t_j^k)} = \frac{\left(\sum_{k=1}^K N_k\right)}{K} \widehat{\text{PLV}}_K.$$

Thus, we can write $\widehat{\text{PLV}}_K = \nu_K \cdot \widehat{c}_K$, with $\nu_K = \frac{K}{(\sum_{k=1}^K N_k)}$. With the same techniques (using $x(t) = 1$), we can show convergence in the distribution of ν_K to a constant:

$$\frac{1}{\nu_K} = \frac{\left(\sum_{k=1}^K N_k\right)}{K} = 1/K \sum_k \int_0^T 1 \cdot dN^{(k)} \xrightarrow[K \rightarrow +\infty]{} \int_0^T \lambda(t) dt = \Lambda(T).$$

This leads to

$$\nu_K \xrightarrow[K \rightarrow +\infty]{} \frac{1}{\Lambda(T)}.$$

Following a version of Slutsky's theorem (Mittelhammer, 1996, theorem 5.10), since ν_K and \widehat{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

$$\text{PLV}^* = \lim_{K \rightarrow \infty} \nu_K \cdot \widehat{c}_K = \frac{c^*}{\Lambda(T)}$$

and can decompose the PLV residual as follows:

$$\sqrt{K}(\widehat{\text{PLV}}_K - \text{PLV}^*) = \sqrt{K}\nu_K(\widehat{c}_K - c^*) + \sqrt{K}(\nu_K c^* - \text{PLV}^*).$$

Taking the limit of the above equation, the second term clearly vanishes (see the above limit of v_K), and the first term, using again the limit of products, leads to the final result:

$$\sqrt{K} \left(\widehat{\text{PLV}}_K - \text{PLV}^* \right) \xrightarrow{K \rightarrow +\infty} \mathcal{N} \left(0, \frac{1}{\Lambda(T)^2} \text{Cov}(Z) \right).$$

□

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:

$$\text{PLV}^* = \frac{\int_0^T e^{i\phi(t)} \lambda(t) dt}{\int_0^T \lambda(t) dt} \quad (\text{A.9})$$

$$= \frac{r_o \int_0^T e^{i\phi(t)} \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt}{r_o \int_0^T \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt}. \quad (\text{A.10})$$

We change the integration variable from $\phi(t)$ to θ :

$$\text{PLV}^* = \frac{\int_{\phi(0)}^{\phi(T)} e^{i\theta} \exp(\kappa \cos(\theta - \varphi_0)) d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa \cos(\theta - \varphi_0)) d\theta}. \quad (\text{A.11})$$

To simplify the integral (bring the φ_0 out of the integral), we change the integration variable again, from θ to ψ , ($\psi = \theta - \varphi_0$):

$$\text{PLV}^* = \frac{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} e^{i(\psi+\varphi_0)} \exp(\kappa \cos(\psi)) d\psi}{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} \exp(\kappa \cos(\psi)) d\psi} \quad (\text{A.12})$$

$$= e^{i\varphi_0} \frac{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} e^{i\psi} \exp(\kappa \cos(\psi)) d\psi}{\int_{\phi(0)-\varphi_0}^{\phi(T)-\varphi_0} \exp(\kappa \cos(\psi)) d\psi}. \quad (\text{A.13})$$

Given that that integrand is a 2π -periodic functions (thus, the integral is invariant to translations of the integration interval), we get

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_{-\pi}^{\pi} e^{i\psi} \exp(\kappa \cos(\psi)) d\psi}{\int_{-\pi}^{\pi} \exp(\kappa \cos(\psi)) d\psi}.$$

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

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_0^\pi \cos(\psi) \exp(\kappa \cos(\psi)) d\psi}{\int_0^\pi \exp(\kappa \cos(\psi)) d\psi}.$$

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):

$$I_k(\kappa) = \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta + \frac{\sin(k\pi)}{\pi} \int_0^{+\infty} e^{-\kappa \cosh t - kt} dt \quad (\text{A.14})$$

$$= \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta, \quad (\text{A.15})$$

we can derive the compact form:

$$\text{PLV}^* = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}. \quad (\text{A.16})$$

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 $[0, T]$:

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(\phi(t)) \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt. \quad (\text{A.17})$$

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

$$\begin{aligned} (\text{Cov}(Z))_{11} &= \frac{1}{\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2} \\ &\times \int_0^T \cos^2(\phi(t)) \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt. \end{aligned} \quad (\text{A.18})$$

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 A.18 becomes

$$\begin{aligned} (\text{Cov}(Z))_{11} &= \frac{1}{\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2} \\ &\times \int_0^T \cos^2(\phi(t) - \varphi_0) \exp(\kappa \cos(\phi(t) - \varphi_0)) \phi'(t) dt. \end{aligned} \quad (\text{A.19})$$

We change the variable of the integral from $\phi(t) - \varphi_0$ to θ and use the following trigonometric identity,

$$\cos^2(\theta) = \frac{1}{2} (1 + \cos(2\theta)), \quad (\text{A.20})$$

to obtain

$$\begin{aligned} (\text{Cov}(Z))_{11} &= \frac{1}{2\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2} \int_{\phi(0)}^{\phi(T)} (1 + \cos(2\theta)) \exp(\kappa \cos(\theta)) d\theta \\ &= \frac{1}{2\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2} \\ &\quad \times \int_{\phi(0)}^{\phi(T)} (\exp(\kappa \cos(\theta)) + \cos(2\theta) \exp(\kappa \cos(\theta))) d\theta. \end{aligned}$$

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

$$\begin{aligned} (\text{Cov}(Z))_{11} &= \frac{1}{2\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2} \left[\int_0^{2\pi\gamma_T} \exp(\kappa \cos(\theta)) d\theta \right. \\ &\quad \left. + \int_0^{2\pi\gamma_T} \cos(2\theta) \exp(\kappa \cos(\theta)) d\theta \right], \end{aligned}$$

$$(\text{Cov}(Z))_{11} = \frac{1}{2\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2} [2\gamma_T \pi I_0(\kappa) + 2\gamma_T \pi I_2(\kappa)] \quad (\text{A.21})$$

$$= \frac{2\pi\gamma_T}{2\lambda_0 (\phi(T) - \phi(0))^2 I_0(\kappa)^2} [I_0(\kappa) + I_2(\kappa)] \quad (\text{A.22})$$

$$= \frac{1}{2\lambda_0 (\phi(T) - \phi(0)) I_0(\kappa)^2} [I_0(\kappa) + I_2(\kappa)], \quad (\text{A.23})$$

where γ_T is the number of oscillation periods contained in $[0, T]$.

We can have a similar calculation for the imaginary part, $(\text{Cov}(Z))_{22}$, 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 symmetries of integrand.

Therefore, we showed that for a given $\kappa \geq 0$, scaled residual

$$Z' = e^{-i\varphi_0} \sqrt{K} (\widehat{\text{PLV}}_K - \text{PLV}^*)$$

converges to a zero mean complex gaussian with the following covariance:

$$\begin{aligned} \text{Cov} \begin{bmatrix} \text{Re}\{Z'\} \\ \text{Im}\{Z'\} \end{bmatrix} &= \begin{bmatrix} \text{Re}\{Ze^{-i\varphi_0}\} \\ \text{Im}\{Ze^{-i\varphi_0}\} \end{bmatrix} \\ &= \frac{1}{2\lambda_0(\phi(T) - \phi(0))I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}. \end{aligned}$$

□

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 $[0, T]$:

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(\phi(t))dt. \quad (\text{A.24})$$

By changing the variable from $\phi(t)$ to θ , we get

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta. \quad (\text{A.25})$$

As $\Lambda(T) = \int_0^T \lambda_0 dt = \lambda_0 T$, we have

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta \quad (\text{A.26})$$

$$= \frac{1}{\lambda_0 T^2} \int_{\phi(0)}^{\phi(T)} \cos^2(\theta)\tau'(\theta)d\theta. \quad (\text{A.27})$$

With a similar calculation for other coefficients of the covariance matrix, we get

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

Therefore, we showed that the scaled residual,

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

converges to a zero mean complex gaussian:

$$\sqrt{K} \left(\widehat{\text{PLV}}_K - \text{PLV}^* \right) \xrightarrow[K \rightarrow +\infty]{} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \text{Cov}(Z) \right).$$

□

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:

$$C = \int_0^t x(t) dN(t)^\top. \quad (\text{A.28})$$

As for the univariate case, under mild assumptions, we can associate a martingale with a vector-valued counting process $N(t)$:

$$M(t) = N(t) - \int_0^t \lambda(s) ds. \quad (\text{A.29})$$

As in this theorem, we assume $\lambda(t) = \lambda_0$, $t \in [0, T]$, we get

$$M(t) = N(t) - \lambda_0 t. \quad (\text{A.30})$$

The $(p \times n)$ matrix-valued martingale for the empirical coupling matrix of equation 4.1, resulting from stochastic integration, is

$$M_x(t) = \int_0^t x(s) dM^\top(s) ds \quad (\text{A.31})$$

and can be decomposed similarly to equation B.3 as

$$M_x(t) = \int_0^t x(s) dN(s)^\top - \int_0^t x(s) \lambda_0 ds. \quad (\text{A.32})$$

By generalizing the steps of theorem 1, we introduced the $(p \times n)$ -variate martingale:

$$\tilde{M}^{(K)}(t) = 1/\sqrt{K} \sum_{k=1}^K M_x^{(k)}(t) \quad (\text{A.33})$$

$$= 1/\sqrt{K} \sum_{k=1}^K \int_0^t x(s) \left(dM^{(k)} \right)^\top(s) ds. \quad (\text{A.34})$$

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 $\mathbf{H}^{(K)}(t)$, consider the multivariate stochastic integral of multivariate martingale $\mathbf{M}^{(K)}$ with intensity vector $\boldsymbol{\lambda}^{(K)}(t)$:*

$$\int_0^t \mathbf{H}^{(K)}(u) d\mathbf{M}^{(K)}(u).$$

Assume:

1. $\int_0^t \mathbf{H}^{(K)}(u) \text{diag}\{\boldsymbol{\lambda}^{(K)}(u)\} \mathbf{H}^{(K)}(u)^\top du \xrightarrow{P} \mathbf{V}(t)$.
2. $\sum_{j=1}^k \int_0^t (\mathbf{H}^{(K)}(u))^2 \mathbf{1}_{|\mathbf{H}^{(K)}(u)| > \epsilon} \boldsymbol{\lambda}_j^{(K)}(u) du \xrightarrow{P} 0$, for all $t \in [0, T]$ and $\epsilon > 0$.

The above stochastic integral converges in distribution to a mean-zero gaussian martingale of covariance $\mathbf{V}(t)$.

We notice that when summing across K trials (see equation A.34), deterministic signals \mathbf{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 $\sum_{k=1}^K N^{(k)}(t)$ are distributed as multivariate Poisson processes with intensity vector $K\boldsymbol{\lambda}_0$, such that

$$\tilde{\mathbf{M}}^{(K)}(t) = 1/\sqrt{K} \int_0^t \mathbf{x}(s) d\mathbf{P}^\top(s) ds, \quad (\text{A.35})$$

where \mathbf{P} is the martingale associated with the pooled process,

$$\mathbf{P}(t) = \left(\sum_{k=1}^K N^{(k)}(t) \right) - \int_0^t K\boldsymbol{\lambda}(s) ds. \quad (\text{A.36})$$

Given that the coupling matrix is matrix valued, we have to vectorize it in order to apply the above CLT. Let $\text{Vec}\{\cdot\}$ be the operator that concatenates the successive columns of a matrix into a larger column vector. $\tilde{\mathbf{M}}^{(K)}(t)$ is a $(p \times n)$ -variate matrix-valued process, and its vectorized version, $\text{Vec}\{\tilde{\mathbf{M}}^{(K)}(t)\}$, is a $(pn \times 1)$ -variate vector process. We can write equation A.35 in vectorized form as

$$\text{Vec}\{\tilde{\mathbf{M}}^{(K)}(t)\} = \int_0^t \mathbf{H}(s) d\mathbf{P}^\top(s) ds,$$

with the $(pn \times n)$ -variate block diagonal matrix:

$$H(s) = \frac{1}{\sqrt{K}} \begin{bmatrix} x(s) & 0 & \cdots & \cdots & 0 \\ 0 & x(s) & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & x(s) \end{bmatrix}. \quad (\text{A.37})$$

The variance of $\text{Vec}\{\tilde{M}^{(K)}(t)\}$ (a $(pn \times pn)$ -variate covariance matrix which is also called predictable variation process) can be written, based on proposition 1, as

$$\tilde{V}(t) = \int_0^t H(s) \text{diag}\{\lambda(s)\} H(s)^\top ds. \quad (\text{A.38})$$

Since we assume a constant intensity function, $\lambda(t) = \lambda_0 = \{\lambda_k\}_k$ ($(n \times 1)$ -variate matrix), we can simplify equation A.38 as follows:

$$\tilde{V}(t) = \int_0^t H(s) \text{diag}\{K\lambda_0\} H(s)^\top ds. \quad (\text{A.39})$$

Replacing $H(s)$ with the block diagonal matrix defined in equation A.37 leads us to

$$\tilde{V}(t) = \frac{1}{K} \begin{bmatrix} \int_0^t K\lambda_1 x(s)x(s)^H ds & 0 & \cdots & \cdots & 0 \\ 0 & \int_0^t K\lambda_2 x(s)x(s)^H ds & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \int_0^t K\lambda_n x(s)x(s)^H ds \end{bmatrix} \quad (\text{A.40})$$

$$= \begin{bmatrix} \lambda_1 \int_0^t x(s)x(s)^H ds & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 \int_0^t x(s)x(s)^H ds & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \lambda_n \int_0^t x(s)x(s)^H ds \end{bmatrix}. \quad (\text{A.41})$$

This fulfills condition 1 of the CLT for all $t \in [0, T]$. For the second condition, it is enough to see that the coefficients of \mathbf{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 $\tilde{\mathbf{V}}(t)$ for each t . Specializing the result for $t = T$, we get, based on assumption 2, a diagonal covariance matrix with block-constant diagonal coefficients,

$$\tilde{\mathbf{V}}(T) = \begin{bmatrix} T\lambda_1 \mathbf{I}_p & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & T\lambda_2 \mathbf{I}_p & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & T\lambda_n \mathbf{I}_p \end{bmatrix}, \quad (\text{A.42})$$

where \mathbf{I}_p indicates the $(p \times p)$ identity matrix, which provides the covariance matrix of the (vectorized) coefficients of matrix $\sqrt{K}\widehat{\mathbf{C}}_K$.

Therefore, for the normalized coupling matrix, $\widehat{\mathbf{C}}_K \text{diag}(\sqrt{T\lambda_0})^{-1}$, 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 $K \rightarrow +\infty$ to a random matrix with i.i.d. unit variance complex gaussian coefficients (because lack of correlations implies independence in the gaussian case):

$$\sqrt{K} \text{Vec}\{\widehat{\mathbf{C}}_K \text{diag}(\sqrt{T\lambda_0})^{-1}\} \xrightarrow[K \rightarrow +\infty]{} \mathcal{N}(\mathbf{0}_{pn}, \mathbf{I}_{pn}). \quad (\text{A.43})$$

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. $\frac{1}{n} \max_{j \neq l} \mathbb{E} |\bar{X}_{jk}X_{lk}|^2 \rightarrow 0$ uniformly in $k \leq n$.
3. $\frac{1}{n^2} \sum_{\Gamma} (\mathbb{E}[(\bar{X}_{jk}X_{lk} - \delta_{lj})(X_{j'k}\bar{X}_{l'k} - \delta_{j'l'})])^2 \rightarrow 0$ 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. $p/n \rightarrow \alpha \in (0, \infty)$.

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 $\mathbb{E}[\bar{X}_{jk}X_{lk}]$ is the covariance between the coefficients of the normalized coupling matrix.

For condition 2, let us first evaluate

$$\mathbb{E} |\bar{X}_{jk}X_{lk} - \delta_{lj}|^2.$$

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

$$\begin{aligned} \bar{X}_{jk}X_{lk} &= - \int_0^T (X_{lk}\bar{H}_{jk}(s) + \bar{X}_{jk}H_{lk}(s)) K\lambda_k ds \\ &\quad + \int_0^T [(\bar{X}_{jk}(s_-) + \bar{H}_{jk}(s_-))(X_{lk}(s_-) + H_{lk}(s_-)) \\ &\quad - \bar{X}_{jk}X_{lk}(s_-)] (dP_k(s) + K\lambda_k dt), \\ &= \int_0^T (X_{lk}\bar{H}_{jk}(s_-) + \bar{X}_{jk}H_{lk}(s_-)) dP_k(s) \\ &\quad + \int_0^T [\bar{H}_{jk}(s_-)H_{lk}(s_-)] (dP_k(s) + K\lambda_k ds). \end{aligned} \quad (\text{A.44})$$

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_{lj} . As a consequence, $\mathbb{E} |\bar{X}_{jk}X_{lk} - \delta_{lj}|^2$ is the variance of the above expression, which is (by stochastic integral formula)

$$\begin{aligned} \mathbb{E} |\bar{X}_{jk}X_{lk} - \delta_{lj}|^2 &= - \int_0^T \mathbb{E} [(X_{lk}(s_-)\bar{H}_{jk}(s_-) + \bar{X}_{jk}(s_-)H_{lk}(s_-))^2] K\lambda_k ds \\ &\quad + \int_0^T [\bar{H}_{jk}(s_-)H_{lk}(s_-)]^2 K\lambda_k ds. \end{aligned} \quad (\text{A.45})$$

Applying again the formula for predictable variation process, we obtain

$$\begin{aligned} \mathbb{E} |\bar{X}_{jk} X_{lk} - \delta_{lj}|^2 &= - \int_0^T \left[\int_0^s (H_{lk}(u) \bar{H}_{jk}(s_-) + \bar{H}_{jk}(u) H_{lk}(s_-))^2 K \lambda_k du \right] K \lambda_k ds \\ &\quad + \int_0^T [\bar{H}_{jk}(s_-) H_{lk}(s_-)]^2 K \lambda_k ds. \end{aligned} \quad (\text{A.46})$$

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:

$$\begin{aligned} \frac{1}{K^2 \lambda_k^2} \mathbb{E} [(\bar{X}_{jk} X_{lk}) (X_{j'k} \bar{X}_{l'k})] &= \int_0^T H_{lk} H_{j'k} ds \int_0^T \bar{H}_{jk} \bar{H}_{l'k} ds \\ &\quad + \int_0^T H_{lk} \bar{H}_{jk} ds \int_0^T H_{j'k} \bar{H}_{l'k} ds + \int_0^T H_{lk} \bar{H}_{l'k} ds \int_0^T \bar{H}_{jk} H_{j'k} ds \\ &\quad + \frac{1}{K \lambda_k} \int_0^T H_{lk} \bar{H}_{jk} H_{j'k} \bar{H}_{l'k} ds \\ &= \frac{1}{\lambda_k^2 T^2 K^2} \left[\int_0^T x_l x_{j'} ds \int_0^T \bar{x}_j \bar{x}_{l'} ds + \int_0^T x_l \bar{x}_j ds \int_0^T x_{j'} \bar{x}_{l'} ds \right. \\ &\quad \left. + \int_0^T x_l \bar{x}_{l'} ds \int_0^T \bar{x}_{jk} x_{j'k} ds \right] + \frac{1}{K^3 \lambda_k^3 T^2} \int_0^T x_l \bar{x}_j x_{j'} \bar{x}_{l'} ds. \end{aligned}$$

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

$$\begin{aligned} \frac{1}{K^2 \lambda_k^2} \mathbb{E} [(\bar{X}_{jk} X_{lk} - 1) (X_{j'k} \bar{X}_{l'k})] \\ = \frac{1}{K^2 \lambda_k^2} \mathbb{E} [(\bar{X}_{jk} X_{lk}) (X_{j'k} \bar{X}_{l'k})] - \frac{1}{K^2 \lambda_k^2} \mathbb{E} [X_{j'k} \bar{X}_{l'k}] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\lambda_k^2 T^2 K^2} \left[\int_0^T x_l x_{j'} ds \int_0^T \bar{x}_j \bar{x}_{l'} ds + \int_0^T (x_l \bar{x}_j - T) ds \int_0^T x_{j'} \bar{x}_{l'} ds \right. \\
&\quad \left. + \int_0^T x_l \bar{x}_{l'} ds \int_0^T \bar{x}_{jk} x_{j'k} ds \right] + \frac{1}{K^3 \lambda_k^3 T^2} \int_0^T x_l \bar{x}_j x_{j'} \bar{x}_{l'} ds,
\end{aligned}$$

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

$$\begin{aligned}
&\lim \frac{1}{n^2} \sum_{\Gamma} (\mathbb{E} [(\bar{X}_{jk} X_{lk} - \delta_{lj}) (X_{j'k} \bar{X}_{l'k} - \delta_{j'l'})])^2 \\
&= \lim \frac{1}{n^2 K^2 \lambda_k^2} \sum_{\Gamma} (\mathbb{E} [(\bar{X}_{jk} X_{lk} - \delta_{lj}) (X_{j'k} \bar{X}_{l'k} - \delta_{j'l'})])^2. \tag{A.47}
\end{aligned}$$

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. \square

Proof of Theorem 4. Let us use the result of Chafaï and Tikhomirov (2018) adapted to our complex case and adapt the dimension notation ($n \rightarrow p(n)$, $m_n \rightarrow n$, 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 ,

$$\widehat{\Sigma}_n = \frac{1}{n} \sum_{k=1}^n X_n^{(k)} X_n^{(k)H}.$$

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 $f : \mathbb{N} \rightarrow [0, 1]$, $g : \mathbb{N} \rightarrow \mathbb{R}^+$ such that $f(r) \rightarrow 0$ and $g(r) \rightarrow 0$ as $r \rightarrow \infty$, and for every $p \in \mathbb{N}$, for any orthogonal projection $P : \mathbb{C}^p \rightarrow \mathbb{C}^p$ of rank $r > 0$, for any real $t > f(r).r$ we have

$$\mathbb{P} (\|P X_n\|^2 - r \geq t) \leq \frac{g(r)r}{t^2}.$$

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 $\|PX_n\|^2$. The inequality leads to, for any t ,

$$\mathbb{P}(\|PX_n\|^2 - r \geq \sigma t) \leq \mathbb{P}(|\|PX_n\|^2 - r| \geq \sigma t) \leq \frac{1}{t^2},$$

so we get $\mathbb{P}(\|PX_n\|^2 - r \geq t) \leq \sigma^2/t^2$ 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

$$X_n = \int_0^T \frac{x(t)}{\sqrt{K\lambda T}} dP(t),$$

with P the compensated Poisson process martingale of rate $K\lambda$. In an orthonormal basis adapted to the orthogonal projection P with rank r , we can rewrite

$$\|PX_n\|^2 = \sum_{k=1}^r |\langle w_k, X_n \rangle|^2,$$

where $\{w_k\}$ are r orthonormal vectors in \mathbb{C}^p . Then we have

$$\sigma^2 = \sum_{k,l \leq r} \mathbb{E} [|\langle w_k, X_n \rangle|^2 |\langle w_l, X_n \rangle|^2 - 1].$$

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

$$\sigma^2 = \frac{1}{K\lambda T^2} \sum_{k,l \leq r} \int_0^T \langle w_k, x(t) \rangle \langle x(t), w_k \rangle \langle w_l, x(t) \rangle \langle x(t), w_l \rangle dt,$$

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,

$$\mathcal{X}(U, V) = \int_0^T \langle V, x x^H(t) \rangle \langle x x^H(t), U \rangle dt,$$

with associated eigenvalues $\xi_1 \geq \dots \geq \xi_{p^2} \geq 0$ such that

$$\sigma^2 = \frac{1}{K\lambda T^2} \sum_{k,l \leq r} \mathcal{X}(\mathbf{w}_k \mathbf{w}_l^H, \mathbf{w}_k \mathbf{w}_l^H).$$

This sum is maximized when the r^2 unitary tensor matrices of the sum $\mathbf{w}_k \mathbf{w}_l$ 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

$$V(t) = \mathbb{E}[M(t)^2] = \int_0^t \lambda(s) ds.$$

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

$$M_H(t) = \int_0^t H(s)dM(s)ds. \quad (\text{B.2})$$

Using equation B.1, we can write

$$M_H(t) = \int_0^t H(s)dN(s) - \int_0^t H(s)\lambda(s)ds, \quad (\text{B.3})$$

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

$$V_H(t) = \mathbb{E}[M_H(t)^2] = \int_0^t H(s)^2\lambda(s)ds, \quad (\text{B.4})$$

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

$$V_{H,G}(t) = \mathbb{E}[M_H(t)M_G(t)] = \int_0^t H(s)G(s)\lambda(s)ds. \quad (\text{B.5})$$

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

$$X(t) = X(0) + \int_0^t f(X(s), s)ds + \int_0^t h(X(s), s)dN(s), \quad (\text{B.6})$$

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:

$$\mathbb{E}[X(t)] = \mathbb{E}[X(0)] + \int_0^t f(X(s), s)ds + \int_0^t \mathbb{E}[h(X(s), s)]\lambda(s)ds. \quad (\text{B.7})$$

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):

$$\begin{aligned} Y(t) &= Y(0) + \int_0^t \frac{dF}{dx}(X(s))f(X(s), s)ds \\ &\quad + \int_0^t [F(X(s_-) + h(X(s_-), s)) - F(X(s_-))]dN(s), \end{aligned} \quad (\text{B.8})$$

where $X(s_-) = \lim_{t \rightarrow s_-} X(t)$ indicates the left limit.

For a scalar function of a multivariate process $Y(t) = F(\mathbf{X}(t))$ with

$$Y(t) = Y(0) + \int_0^t f(\mathbf{X}(s), s)ds + \int_0^t \mathbf{h}(\mathbf{X}(s), s)dN(s), \quad (\text{B.9})$$

the generalization is straightforward:

$$\begin{aligned} Y(t) &= Y(0) + \int_0^t \sum_k \frac{dF}{dx_k}(\mathbf{X}(s))f_k(\mathbf{X}(s), s)ds \\ &\quad + \int_0^t [F(\mathbf{X}(s_-) + \mathbf{h}(\mathbf{X}(s_-), s)) - F(\mathbf{X}(s_-))]dN(s). \end{aligned} \quad (\text{B.10})$$

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

$$\begin{aligned} \mathbb{E}[WXYZ](t) &= \int_0^t ABCD(s_-)\lambda(s)ds \\ &\quad + \left(\int_0^t AB(s)\lambda(s)ds \right) \left(\int_0^t CD(s)\lambda(s)ds \right) \\ &\quad + \left(\int_0^t AC(s)\lambda(s)ds \right) \left(\int_0^t BD(s)\lambda(s)ds \right) \\ &\quad + \left(\int_0^t AD\lambda(s)ds \right) \left(\int_0^t BC(s)\lambda(s)ds \right). \end{aligned} \quad (\text{B.11})$$

Proof. We apply the above formula to $F(W, X, Y, Z) = WXYZ$, yielding

$$\begin{aligned} WXYZ(t) &= - \int_0^t (AXYZ(s) + WBYZ(s) + WXCZ(s) + WXYD(s))\lambda ds \\ &\quad + \int_0^t [(W(s_-) + A)(X(s_-) + B)(Y(s_-) + C)(Z(s_-) + D) \\ &\quad - WXYZ(s_-)]dN(s). \end{aligned}$$

Expanding the second term, we obtain the formula

$$\begin{aligned} WXYZ(t) &= \int_0^t (AXYZ(s) + WBYZ(s) + WXCZ(s) + WXYD(s)) dM(s) \\ &\quad + \int_0^t (ABYZ(s_-) + AXCZ(s_-) + AXYD(s_-) + WBCZ(s_-) \\ &\quad + WBYZ(s_-) + WXCD(s_-))dN(s) + \int_0^t ABCD(s_-)dN(s) \\ &\quad + \int_0^t (ABCZ(s_-) + AXCD(s_-) + ABYD(s_-) + WBCD(s_-))dN(s). \end{aligned}$$

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

$$\begin{aligned}\mathbb{E}[WXYZ](t) &= \int_0^t ABCD(s_-)d\lambda(s) + \int_0^t (AB\mathbb{E}YZ(s_-) + AC\mathbb{E}XZ(s_-) \\ &\quad + AD\mathbb{E}XY(s_-) + BC\mathbb{E}WZ(s_-) + BD\mathbb{E}WY(s_-) \\ &\quad + CD\mathbb{E}WX(s_-))\lambda(s)ds.\end{aligned}\tag{B.12}$$

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,

$$\begin{aligned}\mathbb{E}[WXYZ](t) &= \int_0^t ABCD(s_-)d\lambda(s) + \int_0^t \left(AB(s_-) \int_0^s CD(u_-)\lambda(u)du \right. \\ &\quad + CD(s_-) \int_0^s AB(u_-)\lambda(u)du + AC(s_-) \int_0^s BD(u_-)\lambda(u)du \\ &\quad + BD(s_-) \int_0^s AC(u_-)\lambda(u)du + AD(s_-) \int_0^s BC(u_-)\lambda(u)du \\ &\quad \left. + BC(s_-) \int_0^s AD(u_-)\lambda(u)du \right) \lambda(s)ds.\end{aligned}\tag{B.13}$$

We then observe that the terms inside the integral can be paired such that the integral form of the product derivative formula ($\int f \int g = \int (g \int f + f \int g)$) can be applied, leading directly to equation B.11. \square

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*

$$F_{X_n}(x) \xrightarrow{n \rightarrow +\infty} F_X,$$

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 ,

$$\mathbb{E}[f(X_n)] = \int f d\mu_{X_n} \rightarrow \mathbb{E}[f(X)] = \int f d\mu_X,$$

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, $F_X(x) = P(X_1 < x_1, \dots, X_n < x_n)$, in definitions. A simple necessary and sufficient condition for $X \rightarrow Y$ is that for all vectors t , $t^\top X \rightarrow t^\top Y$ (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 $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0. \quad (\text{B.14})$$

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 μ takes the deterministic value $\mu(\omega)$.

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 μ_n converges weakly in probability to the deterministic measure ν for any $\epsilon > 0$:*

$$P(\rho(\mu_n, \nu) > \epsilon) \xrightarrow{n \rightarrow +\infty} 0. \quad (\text{B.15})$$

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 μ_n converges weakly with probability one to the deterministic measure ν for any $\epsilon > 0$:*

$$P\left(\rho(\mu_n(\omega), \nu) \xrightarrow{n \rightarrow +\infty} 0\right) = 1. \quad (\text{B.16})$$

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; Doussal, 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 X be a $p \times n$ data matrix. Assume that the coefficients of 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 $x_{ij} = x_{ij}^{\text{real}} + ix_{ij}^{\text{imag}}$, where $x_{ij} = x_{ij}^{\text{real}}$ and x_{ij}^{imag} are independent (real) $\mathcal{N}(0, \frac{1}{2})$. This implies that columns of X are i.i.d. $\mathcal{N}_{\mathbb{C}}(\mathbf{0}_p, I_p)$ and, similarly, the real and imaginary parts are $\mathcal{N}(\mathbf{0}_p, I_p/2)$.

As n grows and $\frac{p}{n} \xrightarrow{n \rightarrow +\infty} \alpha \in (0, +\infty)$, the ESD of the so-called Wishart ensemble, $S_n = \frac{1}{n}XX^H$, converges to the Marchenko-Pastur law $\mu_{MP}(x)$ (Marchenko & Pastur, 1967) with density

$$\frac{d\mu_{MP}}{dx}(x) = \frac{1-\alpha}{\alpha} \mathbf{1}_{\alpha>1} \delta_0 + \frac{1}{2\pi\alpha x} \sqrt{(b-x)(x-a)} \mathbf{1}_{[a,b]}, \quad (\text{B.17})$$

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).

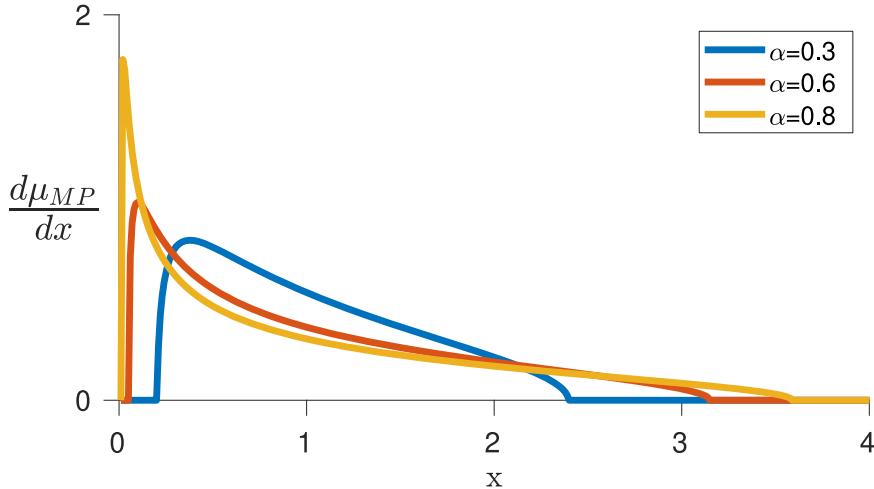


Figure 6: Density of the Marchenko-Pastur law for different values of the aspect ratio of the matrices, α , in equation 2.7.

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 $\{\mu_k\}$ and $\{\nu_k\}$ 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 $\{X_n\}$ with independent columns and study the ESD of

$$S_n = \frac{1}{n} X_n X_n^H.$$

In the following proposition, we use the Kronecker delta symbol δ_{ij} (see equation 4.2) and denote by \bar{X} the complex conjugate of X .

Proposition 6. *Let As $n \rightarrow \infty$, and assume the following. Let*

1. $\mathbb{E} \bar{X}_{jk} X_{lk} = \delta_{lj}$, for all k .
2. $\frac{1}{n} \max_{j \neq l} \mathbb{E} |\bar{X}_{jk} X_{lk} - \delta_{lj}|^2 \rightarrow 0$ uniformly in $k \leq n$.
3. $\frac{1}{n^2} \sum_{\Gamma} (\mathbb{E} (\bar{X}_{jk} X_{lk} - \delta_{lj}) (X_{j'k} \bar{X}_{l'k} - \delta_{j'l'}))^2 \rightarrow 0$ 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. $p/n \rightarrow \alpha \in (0, \infty)$.

Then, with probability 1, the ESD of 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 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. \square

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 :

$$\gamma_T = \frac{T}{\tau} = \frac{\phi(T) - \phi(0)}{2\pi}.$$

⁹This 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

$$\lambda(t) = \lambda_0 \exp(\kappa \cos(\phi(t) - \varphi_0)), \quad (\text{C.2})$$

for a given $\kappa \geq 0$. Then the expectation of the multitrial PLV estimate converges (for $K \rightarrow +\infty$) to

$$\text{PLV}^* = \frac{\int_0^T e^{i2\pi ft} \exp(\kappa \cos(2\pi ft - \varphi_0)) dt}{\int_0^T \exp(\kappa \cos(2\pi ft - \varphi_0)) dt}. \quad (\text{C.3})$$

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

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_{\phi(0)}^{\phi(T)} \cos(\theta) \exp(\kappa \cos(\theta)) d\theta}{\int_{\phi(0)}^{\phi(T)} \exp(\kappa \cos(\theta)) d\theta} = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}, \quad (\text{C.4})$$

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

$$\text{Cov} \begin{bmatrix} \text{Re}\{Ze^{-i\varphi_0}\} \\ \text{Im}\{Ze^{-i\varphi_0}\} \end{bmatrix} = \frac{1}{2\lambda_0 T I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}. \quad (\text{C.5})$$

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:

$$\text{PLV}^* = \frac{\int_0^T e^{i\phi(t)} \lambda(t) dt}{\int_0^T \lambda(t) dt} \quad (\text{C.6})$$

$$= \frac{\lambda_0 \int_0^T e^{i\phi(t)} \exp(\kappa \cos(\phi(t) - \varphi_0)) dt}{\lambda_0 \int_0^T \exp(\kappa \cos(\phi(t) - \varphi_0)) dt} \quad (\text{C.7})$$

$$= \frac{\lambda_0 \int_0^T e^{imt} \exp(\kappa \cos(mt - \varphi_0)) dt}{\lambda_0 \int_0^T \exp(\kappa \cos(mt - \varphi_0)) dt}. \quad (\text{C.8})$$

We change the integration variable from mt to θ :

$$\text{PLV}^* = \frac{\int_{\theta(0)}^{\theta(T)} e^{i\theta} \exp(\kappa \cos(\theta - \varphi_0)) d\theta}{\int_{\theta(0)}^{\theta(T)} \exp(\kappa \cos(\theta - \varphi_0)) d\theta}. \quad (\text{C.9})$$

To simplify the integral (bring the φ_0 out of the integral), we change the integration variable again, from θ to ψ , ($\psi = \theta - \varphi_0$),

$$\text{PLV}^* = \frac{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} e^{i(\psi+\varphi_0)} \exp(\kappa \cos(\psi)) d\psi}{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} \exp(\kappa \cos(\psi)) d\psi} \quad (\text{C.10})$$

$$= e^{i\varphi_0} \frac{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} e^{i\psi} \exp(\kappa \cos(\psi)) d\psi}{\int_{\theta(0)-\varphi_0}^{\theta(T)-\varphi_0} \exp(\kappa \cos(\psi)) d\psi}. \quad (\text{C.11})$$

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

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_{-\pi}^{\pi} e^{i\psi} \exp(\kappa \cos(\psi)) d\psi}{\int_{-\pi}^{\pi} \exp(\kappa \cos(\psi)) d\psi}.$$

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

$$\text{PLV}^* = e^{i\varphi_0} \frac{\int_0^\pi \cos(\psi) \exp(\kappa \cos(\psi)) d\psi}{\int_0^\pi \exp(\kappa \cos(\psi)) d\psi}.$$

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)

$$I_k(\kappa) = \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta + \frac{\sin(k\pi)}{\pi} \int_0^{+\infty} e^{-\kappa \cosh t - kt} dt \quad (\text{C.12})$$

$$= \frac{1}{\pi} \int_0^\pi \cos(k\theta) \exp(\kappa \cos(\theta)) d\theta, \quad (\text{C.13})$$

we can derive the compact form:

$$\text{PLV}^* = e^{i\varphi_0} \frac{I_1(\kappa)}{I_0(\kappa)}. \quad (\text{C.14})$$

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 $[0, T]$

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(\phi(t)) \exp(\kappa \cos(\phi(t) - \varphi_0)) dt. \quad (\text{C.15})$$

As we have

$$\Lambda(T) = \lambda_0 T I_0(\kappa),$$

we can continue with equation C.15 as,

$$(\text{Cov}(Z))_{11} = \frac{1}{\lambda_0 T^2 I_0(\kappa)^2} \int_0^T \cos^2(\phi(t)) \exp(\kappa \cos(\phi(t) - \varphi_0)) dt. \quad (\text{C.16})$$

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

$$(\text{Cov}(Z))_{11} = \frac{1}{\lambda_0 T^2 I_0(\kappa)^2} \int_0^T \cos^2(\phi(t) - \varphi_0) \exp(\kappa \cos(\phi(t) - \varphi_0)) dt. \quad (\text{C.17})$$

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,

$$\cos^2(\theta) = \frac{1}{2} (1 + \cos(2\theta)), \quad (\text{C.18})$$

to obtain

$$\begin{aligned} (\text{Cov}(Z))_{11} &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \int_{\theta(0)}^{\theta(T)} (1 + \cos(2\theta)) \exp(\kappa \cos(\theta)) d\theta \\ &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \int_{\theta(0)}^{\theta(T)} (\exp(\kappa \cos(\theta)) + \cos(2\theta) \exp(\kappa \cos(\theta))) d\theta. \end{aligned}$$

Given that the integral is invariant to translations of the integration, we get

$$\begin{aligned} (\text{Cov}(Z))_{11} &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} \left[\int_0^{2\pi\gamma_T} \exp(\kappa \cos(\theta)) d\theta \right. \\ &\quad \left. + \int_0^{2\pi\gamma_T} \cos(2\theta) \exp(\kappa \cos(\theta)) d\theta \right] \\ (\text{Cov}(Z))_{11} &= \frac{1}{2m\lambda_0 T^2 I_0(\kappa)^2} [2\gamma_T \pi I_0(\kappa) + 2\gamma_T \pi I_2(\kappa)] \end{aligned} \quad (\text{C.19})$$

$$= \frac{2\pi\gamma_T}{2m\lambda_0 T^2 I_0(\kappa)^2} [I_0(\kappa) + I_2(\kappa)] \quad (\text{C.20})$$

$$= \frac{mT}{2m\lambda_0 T^2 I_0(\kappa)^2} [I_0(\kappa) + I_2(\kappa)] \quad (\text{C.21})$$

$$= \frac{1}{2\lambda_0 T I_0(\kappa)^2} [I_0(\kappa) + I_2(\kappa)]. \quad (\text{C.22})$$

We can have a similar calculation for the imaginary part, $(\text{Cov}(Z))_{22}$, 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

$$Z' = e^{-i\varphi_0} \sqrt{K} (\widehat{\text{PLV}}_K - \text{PLV}^*)$$

converges to a zero mean complex gaussian with the following covariance:

$$\text{Cov} \begin{bmatrix} \text{Re}\{Z'\} \\ \text{Im}\{Z'\} \end{bmatrix} = \begin{bmatrix} \text{Re}\{Ze^{-i\varphi_0}\} \\ \text{Im}\{Ze^{-i\varphi_0}\} \end{bmatrix} = \frac{1}{2\lambda_0 T I_0(\kappa)^2} \begin{bmatrix} I_0(\kappa) + I_2(\kappa) & 0 \\ 0 & I_0(\kappa) - I_2(\kappa) \end{bmatrix}. \quad \square$$

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

$$\lambda(t) = \lambda_0 (1 + \varkappa \cos(2\pi mt/T - \varphi_0)), \quad \lambda_0 > 0, \quad 0 \leq \varkappa \leq 1, \quad (\text{C.23})$$

and the point process is homogeneous Poisson with rate λ_0 . Then the expectation of the PLV estimate converges (for $K \mapsto +\infty$) to

$$\text{PLV}^* = \frac{1}{2} \varkappa e^{i\varphi_0} \delta_{km}, \quad (\text{C.24})$$

where δ_{km} denotes the Kronecker symbol. Moreover, the asymptotic covariance of $Z = \sqrt{K} (\widehat{\text{PLV}}_K - \text{PLV}^*)$ is

$$\text{Cov} \begin{bmatrix} \text{Re}\{Z\} \\ \text{Im}\{Z\} \end{bmatrix} = \frac{1}{2\lambda_0 T} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (\text{C.25})$$

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:

$$\text{PLV}^* = \frac{\int_0^T e^{i\phi(t)} \lambda(t) dt}{\int_0^T \lambda(t) dt} \quad (\text{C.26})$$

$$= \frac{\int_0^T e^{i2\pi kt/T} (1 + \varkappa \cos(2\pi mt/T - \varphi_0)) dt}{\int_0^T (1 + \varkappa \cos(2\pi mt/T - \varphi_0)) dt}. \quad (\text{C.27})$$

By using Euler's formula, we can write the second term in the numerator as weighted sum of exponentials ($\cos(x) = \frac{1}{2}(e^{ix} + e^{-ix})$),

$$\text{PLV}^* = \frac{1}{2} \frac{\int_0^T e^{i2\pi kt/T} + \varkappa \int_0^T e^{i2\pi kt/T} (e^{i(2\pi mt/T - \varphi_0)} + e^{-i(2\pi mt/T - \varphi_0)}) dt}{\int_0^T dt + \int_0^T \varkappa \cos(2\pi mt/T - \varphi_0) dt} \quad (\text{C.28})$$

$$= \frac{1}{2} \frac{\int_0^T e^{i2\pi kt/T} + \varkappa \int_0^T e^{i2\pi(k+m)t/T} e^{i\varphi_0} + \varkappa \int_0^T e^{-i2\pi(k-m)t/T} e^{i\varphi_0} dt}{\int_0^T dt + \int_0^T \varkappa \cos(2\pi mt/T - \varphi_0) dt} \quad (\text{C.29})$$

$$= \frac{1}{2} \frac{\int_0^T e^{i2\pi kt/T} + \varkappa e^{i\varphi_0} \int_0^T e^{i2\pi(k+m)t/T} + \varkappa e^{i\varphi_0} \int_0^T e^{-i2\pi(k-m)t/T} dt}{\int_0^T dt + \varkappa \int_0^T \cos(2\pi mt/T - \varphi_0) dt}. \quad (\text{C.30})$$

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,

$$\text{PLV}^* = \frac{1}{2} \frac{\varkappa e^{i\varphi_0} \int_0^T e^{-i2\pi(k-m)t/T} dt}{\int_0^T dt} \quad (\text{C.31})$$

$$= \frac{1}{2} \varkappa e^{i\varphi_0} \delta_{km}. \quad (\text{C.32})$$

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 $[0, T]$,

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{\Lambda(T)^2} \int_0^T \cos^2(2\pi kt/T) (1 + \varkappa \cos(2\pi mt/T - \varphi_0)) dt, \quad (\text{C.33})$$

and use the trigonometric identity, equation A.20, to get

$$(\text{Cov}(Z))_{11} = \frac{\lambda_0}{2\Lambda(T)^2} \int_0^T (1 + \cos(4\pi kt/T))(1 + \varkappa \cos(2\pi mt/T - \varphi_0)) dt. \quad (\text{C.34})$$

In the resulting equation,

$$\begin{aligned} & (\text{Cov}(Z))_{11} \\ &= \frac{\lambda_0}{2\Lambda(T)^2} \left[\int_0^T dt + \varkappa \int_0^T \cos(2\pi mt/T - \varphi_0) dt + \int_0^T \cos(4\pi kt/T) dt \right. \\ &\quad \left. + \varkappa \int_0^T \cos(4\pi kt/T) \cos(2\pi mt/T - \varphi_0) dt \right], \end{aligned} \quad (\text{C.35})$$

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

$$\begin{aligned} & \int_0^T \cos(4\pi kt/T) \cos(2\pi mt/T - \varphi_0) dt \\ &= \cos(\varphi_0) \int_0^T \cos(4\pi kt/T) \cos(2\pi mt/T) dt \\ &\quad + \sin(\varphi_0) \int_0^T \cos(4\pi kt/T) \sin(2\pi mt/T) dt, \end{aligned} \quad (\text{C.36})$$

and k and m are integers.

Finally, given that $\Lambda(T) = \int_0^T \lambda(t) dt = \lambda_0 T$, we have

$$(\text{Cov}(Z))_{11} = \frac{1}{2\lambda_0 T}. \quad (\text{C.37})$$

We have a similar calculation for the imaginary part, $(\text{Cov}(Z))_{22}$. The off-diagonal elements of the covariance matrix vanish due to the symmetry of the integrand.

Therefore, we showed that for the scaled residual,

$$Z = \sqrt{K} (\widehat{\text{PLV}}_K - \text{PLV}^*)$$

converges to a zero mean isotropic complex gaussian:

$$\sqrt{K} (\widehat{\text{PLV}}_K - \text{PLV}^*) \xrightarrow{K \rightarrow +\infty} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{2\lambda_0 T} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

□

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 $O^{orig}[t] = e^{2\pi i ft}$, 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

$$O[t] = e^{2\pi i ft} \exp(i\eta[t]), \quad (\text{D.1})$$

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, $O^{orig}[t] = \{O_j^{orig}[t]\}_{j=1,\dots,N_{osc}}$, and sample i.i.d. the noise for each oscillation, leading to the vector time series $\eta[t]$. In this case, the noisy oscillations can be written as

$$O[t] = O^{orig}[t] \odot \exp(i\eta[t]), \quad (\text{D.2})$$

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.
- Mastrogiovanni, 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., Schöffelen, 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.

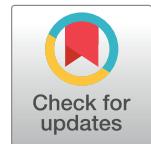
RESEARCH ARTICLE

Uncovering the organization of neural circuits with Generalized Phase Locking Analysis

Shervin Safavi^{1,2}, Theofanis I. Panagiotaropoulos^{1,3}, Vishal Kapoor^{1,4}, Juan F. Ramirez-Villegas^{1,5}, Nikos K. Logothetis^{1,4,6}, Michel Besserve^{1,7*}

1 Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, **2** IMPRS for Cognitive and Systems Neuroscience, University of Tübingen, Tübingen, Germany, **3** Cognitive Neuroimaging Unit, INSERM, CEA, CNRS, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France, **4** International Center for Primate Brain Research (ICPBR), Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Chinese Academy of Sciences (CAS), Shanghai 201602, China, **5** Institute of Science and Technology Austria (IST Austria), Klosterneuburg, Austria, **6** Centre for Imaging Sciences, Biomedical Imaging Institute, The University of Manchester, Manchester, United Kingdom, **7** Department of Empirical Inference, Max Planck Institute for Intelligent Systems and MPI-ETH Center for Learning Systems, Tübingen, Germany

* michel.besserve@tuebingen.mpg.de



OPEN ACCESS

Citation: Safavi S, Panagiotaropoulos TI, Kapoor V, Ramirez-Villegas JF, Logothetis NK, Besserve M (2023) Uncovering the organization of neural circuits with Generalized Phase Locking Analysis. PLoS Comput Biol 19(4): e1010983. <https://doi.org/10.1371/journal.pcbi.1010983>

Editor: Daniele Marinazzo, Ghent University, BELGIUM

Received: June 22, 2022

Accepted: February 27, 2023

Published: April 3, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1010983>

Copyright: © 2023 Safavi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The present work is a methodological study, and no new data have been generated for it. We provide the MatLab codes to regenerate all the figures from scratch (except the

Abstract

Despite the considerable progress of *in vivo* neural recording techniques, inferring the biophysical mechanisms underlying large scale coordination of brain activity from neural data remains challenging. One obstacle is the difficulty to link high dimensional functional connectivity measures to mechanistic models of network activity. We address this issue by investigating spike-field coupling (SFC) measurements, which quantify the synchronization between, on the one hand, the action potentials produced by neurons, and on the other hand mesoscopic “field” signals, reflecting subthreshold activities at possibly multiple recording sites. As the number of recording sites gets large, the amount of pairwise SFC measurements becomes overwhelmingly challenging to interpret. We develop *Generalized Phase Locking Analysis* (GPLA) as an interpretable dimensionality reduction of this multivariate SFC. GPLA describes the dominant coupling between field activity and neural ensembles across space and frequencies. We show that GPLA features are *biophysically interpretable* when used in conjunction with appropriate network models, such that we can identify the influence of underlying circuit properties on these features. We demonstrate the statistical benefits and interpretability of this approach in various computational models and Utah array recordings. The results suggest that GPLA, used jointly with biophysical modeling, can help uncover the contribution of recurrent microcircuits to the spatio-temporal dynamics observed in multi-channel experimental recordings.

Author summary

Modern neural recording techniques give access to increasingly highly multivariate spike data, together with spatio-temporal activities of local field potentials reflecting integrative processes. We introduce GPLA as a generalized coupling measure between these point-

schematic figures and figures contain raw experimental data) in the public repository <https://github.com/shervinsafavi/gpla.git>.

Funding: All authors were supported by the Max Planck Society. M.B. was supported by the German Federal Ministry of Education and Research (BMBF) through the funding scheme received by the Tübingen AI Center, FKZ: 01IS18039B. N.K.L. and V.K. acknowledge the support from the Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX02). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

process and continuous-time activities to help neuroscientists uncover the distributed organization of neural networks. We develop statistical analysis and modeling methodologies for this measure and demonstrate its interpretability in simulated and experimental multi-electrode recordings.

This is a *PLOS Computational Biology* Methods paper.

Introduction

Understanding brain function requires uncovering the relationships between neural mechanisms at different scales [1–3]: from single neurons to microcircuits [4, 5], from microcircuits to a single brain area [6], and from a single area to the whole brain [7, 8]. Therefore, it is crucial to develop tools to investigate the cooperative phenomena that can connect different levels of organization, such as oscillatory neuronal dynamics [3]. These oscillations are hypothesized to support neural computations [9–13] and various cognitive functions [14–17] and manifest themselves in Local Field Potentials (LFP), a mesoscopic extracellular signal [18] 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 [19–22]. As a consequence, a large body of empirical investigations support the functional relevance of LFP oscillations (for reviews, see [19, 20, 22–24]).

The synchronization between spiking activity and LFP has been observed experimentally and led to different theories of cognitive functions. Notably, attention has been hypothesized to rely on interactions between various neural populations coordinated by network oscillations [25–28]. However, the exact network mechanisms that govern these spike-field coupling (SFC) phenomena remain elusive, and notably the role played by the phase of the coupling [29–31]. Arguably two major obstacles are (1) that common SFC measures are pairwise [32–38], which makes the information conveyed about the network increasingly difficult to grasp as the number of pairs of channels increases and (2) the link between SFC measurements and the underlying neural circuit mechanisms is not well understood, a problem that we will call *biophysical interpretability*.

Elaborating on the first obstacle, pairwise analyses are arguably suboptimal for modern neural recording datasets, as state-of-the-art multichannel electrophysiology systems [39–41] allow simultaneous recording of hundreds or even thousands of sites [24, 40, 42, 43]. This represents an unprecedented opportunity to study the large scale collective organization binding spiking activity of individual units with mesoscopic spatio-temporal dynamics (e. g. wave patterns [44]), but at the same time generates high dimensional matrices of pairwise connectivity measurements from which extracting interpretable information is a challenge in itself. Moreover, statistical analysis and significance assessment of parallel spike trains is also challenging (see [45] for a review) and requires novel, more computationally *efficient* approaches in the high dimensional setting.

Elaborating on the second obstacle, this shift in data dimensionality also offers the opportunity to go beyond the phenomenological model of synchronization between these activities, to achieve a precise account of how network properties shape the detailed spatio-temporal

characteristics of collective phenomena. Such account fits the general framework of *constitutive mechanistic explanation* in Science [46, 47], where the occurrence of a phenomenon is explained based on mechanisms governing components of the system under study. However, this requires simple enough links to be established between high-dimensional observations and properties of biophysical models, which remains largely unaddressed for large-scale neural networks [48–51]. To reflect this issue, we will call *biophysical interpretability* the extent to which quantities derived from brain activity measurements can be related to properties of a given biophysical model. While methodologies have been developed that elegantly combine frequency-based analysis, multi-variate methods and dimensionality reduction techniques [52–54], the outcomes typically remain only interpreted as mere phenomenological models of brain activity, broadly built on the idea that the brain processes information through networks of oscillators coupled at different frequencies. As such, these approaches suffer from critical limitations when it comes to providing mechanistic insights that relate to a biologically realistic understanding of the underlying neural circuits. To go beyond this limitation, model reduction approaches are used in physics and biology, including neuroscience [55, 56], to simplify complex models. While these methods have initially mostly been developed to reduce computational complexity, they have also started to be used to foster interpretability [57]. The characteristics of biophysical model reductions and advanced multivariate data analyses need to be carefully chosen to allow biophysical interpretability of high-dimensional measurements, and have not yet been explored for the case of multivariate spike-field coupling.

Thus, we develop a “Generalized Phase Locking Analysis” (GPLA) to address the need for an efficient multivariate method that, in conjunction with suitable neural models, allows biophysical interpretations of spike-field coupling data. GPLA characterizes and assesses statistically the coupling between the spiking activity of large populations of units and large-scale spatio-temporal patterns of LFP.

The benefits of this approach are demonstrated in detail with network models with increasing levels of complexity and biophysical realism, and ultimately with neural data. Each of these settings is designed to demonstrate certain strengths of GPLA. First we provide a theoretical motivation and illustrate how to interpret the outcome of the analysis with toy models. Then we illustrate the statistical benefits of GPLA over uni-variate methods with several simple generative models of spike and LFP. Thereafter, we turn to biophysical interpretability of GPLA using an analytical reduction of two population (excitatory-inhibitory) neural field models. This mechanistic interpretation is exemplified in computational models of hippocampal and cortical neural networks. In particular, we show how studying the phase of GPLA can untangle the contribution of recurrent interactions to the observed spatio-temporal dynamics. Based on these results, application of GPLA to Utah array recordings finally provides evidence of strong feedback inhibition in the macaque prefrontal cortex.

Results

The overarching motivation of this work is to foster a neuroscientific understanding of experimental data by leveraging biophysical models, i.e. models that comprise equations accounting for the biophysics of neural activity and measurements. Broadly construed, models can be ranked according to the chosen trade-off between realism and complexity, with on one end simplified (e.g. low dimensional, linear, ...) analytically tractable models, whose dependency on biophysical properties is easily characterized, and on the opposite end, highly detailed computational models, where the role of biophysical parameters can only be assessed by running costly simulations. However, when it comes to the use of models for interpreting data, another key aspect coming into play is the choice of *quantities of interest* (QoI) (following

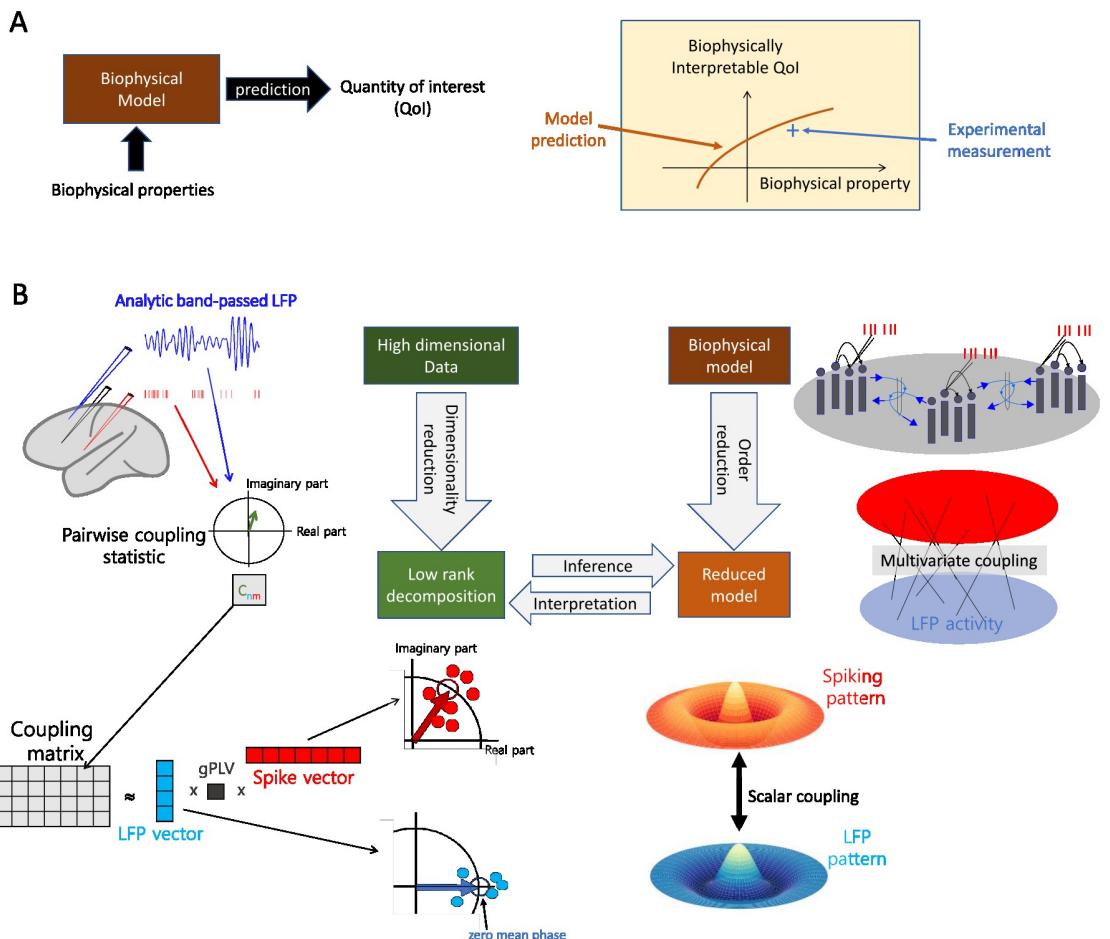


Fig 1. Interpretability of multivariate SFC through GPLA. (A) Schematic for the concept of *biophysical interpretability*. A biophysical model allows to make predictions about some observable quantity derived from neural data, that we call *Quantity of Interest* (QoI). The QoI is biophysically interpretable whenever its variations can be explained by changes in some property of the model. (B) (Top-left) A coupling matrix is estimated from electrophysiology data by gathering complex SFC estimates 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. (Bottom-left) The coupling matrix can be approximated using its largest singular value and the corresponding singular vectors. 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 to the LFP pattern. (Top-right) A biophysical model accounts for the underlying network connectivity and dynamics, as well as the measurement process that leads to the collected data. This leads to a theoretical account of multivariate spike-field coupling. (Bottom-right) Model reduction entails simplifying assumption, leading to a low-rank description of the coupling in the model based on the key mechanistic parameters. This description is compared to the left-hand side low-rank decomposition, obtained from experimental data, to infer parameters and interpret the data. All clip art in this figure was designed and drawn by authors M.B. and S.S..

<https://doi.org/10.1371/journal.pcbi.1010983.g001>

[57]), which are used to assess qualitatively or quantitatively the match between experimental data and candidate models. Indeed, some QoIs may be easier to interpret with a given biophysical model than others, in the sense that they have a more straightforward dependency on biophysical parameters. This leads to us introducing the notion of *biophysical interpretability*, illustrated in Fig 1A and defined as follows: *a QoI is biophysically interpretable for a particular*

model whenever we can identify biophysical properties of the model that influence the QoI in a simple way (the simpler, the more interpretable). By “simple”, we refer to the low complexity of the functional relationship between the properties and the QoI. To fix ideas, functional relations containing only few biophysical parameters and monotonous functions will be considered simple, although the choice of notion of complexity may be adapted to the case at hand. This interpretability property clearly depends on the model employed, and its benefits for analyzing experimental data relies on the assumption that the chosen model captures well relevant properties of the ground truth mechanism, as it is the case for any attempt at understanding empirical data through mechanistic models. The validity of this assumption can never be fully guaranteed and which aspects to incorporate in the model to address a particular question should be assessed based on the literature. In this work, we rely on neural field models [55], which lend themselves to analytical treatment, while still accounting for key aspects of the underlying biophysical network mechanisms. As neurophysiology experiments rely on an increasing number of recording channels, the choice of QoI has to be made from a space of increasingly large dimensions, and we argue that the notion of biophysical interpretability can guide this choice. Specifically, we will consider QoIs that quantify SFC, and a multidimensional generalization of it.

Generalizing SFC to the multivariate setting

QoIs characterizing the coupling between signals originating from a pair of recording channels are commonly used in Neuroscience. On the one hand, we consider the instantaneous 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 narrow band of center frequency f . We assume $L_f(t)$ is the complex analytic signal representation of this oscillation, computed using the Hilbert transform [58], 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. The coupling between these signals can be characterized by the covariance

$$c(f) = \langle \lambda(t)L_f(t) \rangle = \langle \lambda(t)a_f(t)e^{i\phi_f(t)} \rangle = |c|e^{i\Phi_c} = |c|(\cos(\Phi_c) + i \sin(\Phi_c)), \quad (1)$$

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 Fig 1(Top-right)). This coupling measure is a modification of the Phase-Locking Value (PLV) [33] (see Eq 13), 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. We consider the coupling defined in Eq 1 as a base quantity to explain our approach, while normalization will be addressed at the end of this section. Although $\lambda(t)$ is a priori unknown, $c(f)$ is straightforward to estimate based on observed spike times, leading to the empirical estimate denoted $\hat{c}(f)$ (see [59]). However, as more channels are recorded, the number of PLV values to consider increases dramatically, which poses a challenge to their interpretation. Alternatively, using dimensionality reduction to synthesize the information provided by this large number of couplings may provide a more interpretable picture of the functioning of the underlying circuits.

As illustrated in Fig 1B, Generalized Phase Locking Analysis (GPLA) is introduced as a dimensionality reduction technique to estimate the key properties of the coupling matrix $\mathbf{C}(f)$ consisting of the pairwise couplings between a large number of units and LFP channels at frequency f . The estimate $\hat{\mathbf{C}}(f)$ of the coupling matrix based on spiking activity is defined as follows. Given N_m 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's estimate $\hat{\mathbf{C}}(f)$ is computed by summing the values taken by the analytic signal at all spike times (see Fig 1B (Top-left)),

$$\hat{\mathbf{C}}(f)_{n,m} = \sum_k L_f^n(t_k^m). \quad (2)$$

Next, as schematized in Fig 1B(Bottom-left), the coupling matrix is approximated by the term with the largest singular value d_1 of its Singular Value Decomposition (SVD) leading to

$$\hat{\mathbf{C}} = UDV^H = \sum_k d_k \mathbf{u}_k \mathbf{v}_k^H \approx d_1 \mathbf{u}_1 \mathbf{v}_1^H, \quad (3)$$

where \mathbf{v}_k^H indicates the transpose conjugate of the vector \mathbf{v}_k . In this expression, the singular value d_1 is a positive scalar, that we will call generalized Phase Locking Value (gPLV), and which quantifies the magnitude of the coupling. In order to assess the effect size of d_1 , but also to perform significance analysis, normalization of the coupling matrix $\hat{\mathbf{C}}(f)$ is typically performed, as described in section *GPLA for electrophysiology data* of Materials and methods. In particular, the LFP time series can be whitened beforehand, such that the outcome of GLPA is invariant to the LFP power at each frequency. Table 1 indicates for which experiments such a normalization is applied. The associated complex valued singular vectors in this factorization will be respectively called the *LFP vector*, defined as $\mathbf{u} = \mathbf{u}_1$ and the *spike vector*, defined as $\mathbf{v} = \mathbf{v}_1$. As illustrated in Fig 1B (bottom), 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, based on Eq 3, the difference between the phases of each component of \mathbf{u} and \mathbf{v} reflects the phase lag between spiking and LFP activities for the respective channels and units. Notably, this implies that all units and all LFP channels with non-vanishing coefficients in spike and LFP vectors have correlated activities at this frequency, as will be further illustrated in the next section. In particular, two units with non-zero coefficients in the spike vector typically have correlated spike rates at this frequency. Multiplication of both singular vectors by the same unit complex number leads to the exact same approximation as Eq 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 across all components of the LFP vector $\langle \mathbf{u} \rangle = \frac{1}{n_c} \sum_k \mathbf{u}_k$ to zero, as illustrated in Fig 1B (bottom). As a consequence, the phase of the mean of the spike vector coefficients $\langle \mathbf{v} \rangle = \frac{1}{n_u} \sum_k \mathbf{v}_k$ reflects the difference of mean phases between spiking and LFP activities. See section *GPLA for electrophysiology data* in Materials and methods for more details.

Table 1. Summary of normalization by spike count and whitening application in all figures.

| Figure num. | Spike count normalization type | Whitening applied | Equations |
|-------------|--------------------------------|-------------------|--------------------|
| Fig 2 | 1/N | No | 16 |
| Fig 3 | 1/N | No | 16 |
| Fig 4 | 1/ \sqrt{N} | Yes | 17 |
| Fig 6 | 1/ \sqrt{N} | Yes | 17 |
| Fig 7 | 1/ \sqrt{N} | Yes | 17 |
| Fig 8 | 1/ \sqrt{N} | Yes | 17 |

<https://doi.org/10.1371/journal.pcbi.1010983.t001>

Importantly, we demonstrate GPLA can simultaneously be applied to a neural field model (analytically or in simulations), to yield a reduced biophysical model (see Fig 1B(Right)). The outcome of GPLA applied to neural data can then be interpreted based on this reduction, as explained in section *Reduction of complex models based on linear response theory*. Notably, this can be exploited to study the key network characteristics giving rise to the observed spike-field coupling. Here, we demonstrate the possibility of this hybrid approach merging modeling and analysis for a certain class of generative models, while further development is needed to extend it to a more general setting.

Illustration of GPLA and statistical benefits over univariate SFC

We first illustrate how GPLA provides an intuitive phenomenological model of the coupling between the population of spiking units and LFPs. We use three simulations in which a transient global LFP oscillation recorded in a single channel (Fig 2A) modulates the firing probability in 18 spike trains (attributed to neuron-like units). As described on the left column of Fig 2C–2F, models instantiate (1) a global oscillation driving a synchronous population of neurons (2) wave-like discharges of neurons (similar to the case of “delayed excitation from a single oscillator” described by [44]) (3) groups of cells that fire together predominantly at three distinct phase values of the LFP. For comparison, a fourth simulation is performed with no coupling. Exemplary spike trains for each model are displayed in the second column of Fig 2C–2F overlaid on the magnified version of the LFP oscillation.

For all models, the coupling is well reflected by the gPLV magnitude obtained from these simulations, as shown in Fig 2B. Moreover, the phase of the spike vector components resulting from GPLA summarizes the coupling structure in an intuitive way in Fig 2C–2F (right column), showing: (1) all components collapse to a single phase, (2) evenly distributed phases of the spike vector coefficients over a 180 degrees interval, (3) three distinct phases, (4) an isotropic phase distribution, as predicted by mathematical analysis [59].

These simple simulations demonstrate how to interpret the spike vector. Because there is a single LFP channel in this setting, GPLA straightforwardly combines univariate coupling measures of each unit. However, statistical analysis of GPLA is different from the univariate case, as we show next with a setting similar to the above model (3) of Fig 2E but with weaker coupling of individual neurons to the oscillation, leading to values at the edge of significance (assessed with the surrogate-based test, see section *Significance assessment of gPLV* in Materials and methods). An illustrative simulation in the case of low noise and large number of observed spikes is shown in Fig 3B, together with the corresponding spike vector in Fig 3C, providing results similar to Fig 2E.

For quantitative analysis, we consider the setting of a single LFP channel and a handful of neurons are the focus of the analysis (Fig 3A–3E). Such recordings are still common and valuable in human electrophysiology experiments for understanding cognition [60, 61]. While pooling the spikes from all units into a single spike train to get a *pooled Phase-Locking-Value* (pPLV) may 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 Fig 2C, but not for Fig 2D and 2E). In contrast, GPLA exploits the spike times from multiple neurons to assess the global coupling between spikes and LFPs without requiring such homogeneity. We ran 5000 simulations with only 3 units and compared the coupling assessment based on PLV, pPLV, and gPLV. Fig 3D represents the estimated PLVs, with averages matching the couplings obtained with a larger number of spikes in Fig 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 Materials and methods section

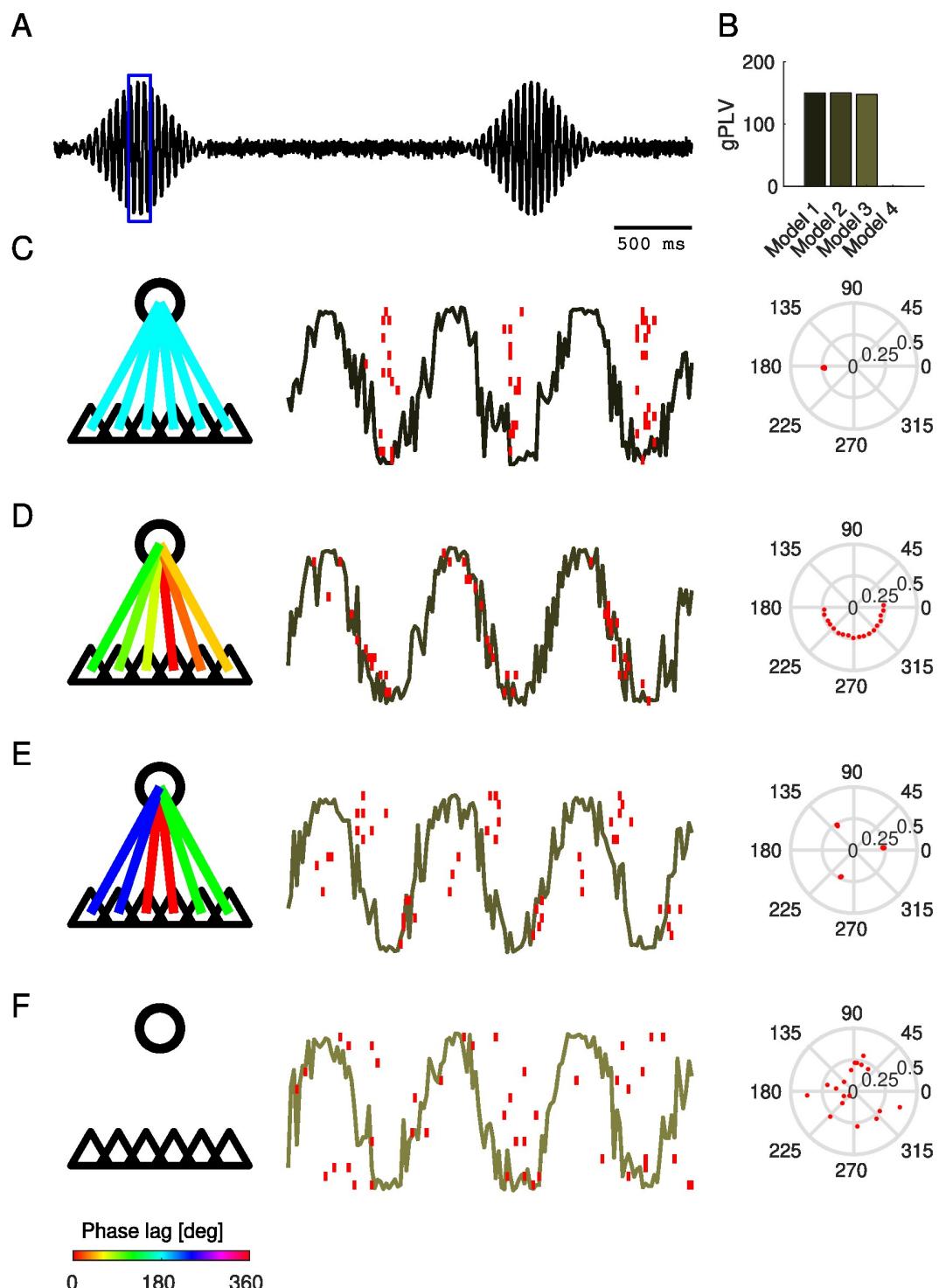


Fig 2. Illustration of GPLA on simple simulations. (A) Normalized amplitude of LFP-like oscillatory signals. (B) gPLVs for different models demonstrated in C-F (C-F) Various scenarios of spike-LFP coupling. Left: schematic representation of the modulating LFP oscillation (circle), and 6 representative 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 Fig 3). (F) Spiking activity uncoupled to LFP oscillation (independent homogeneous spike trains). Also see Table A in S1 Appendix for a methodological summary.

<https://doi.org/10.1371/journal.pcbi.1010983.g002>

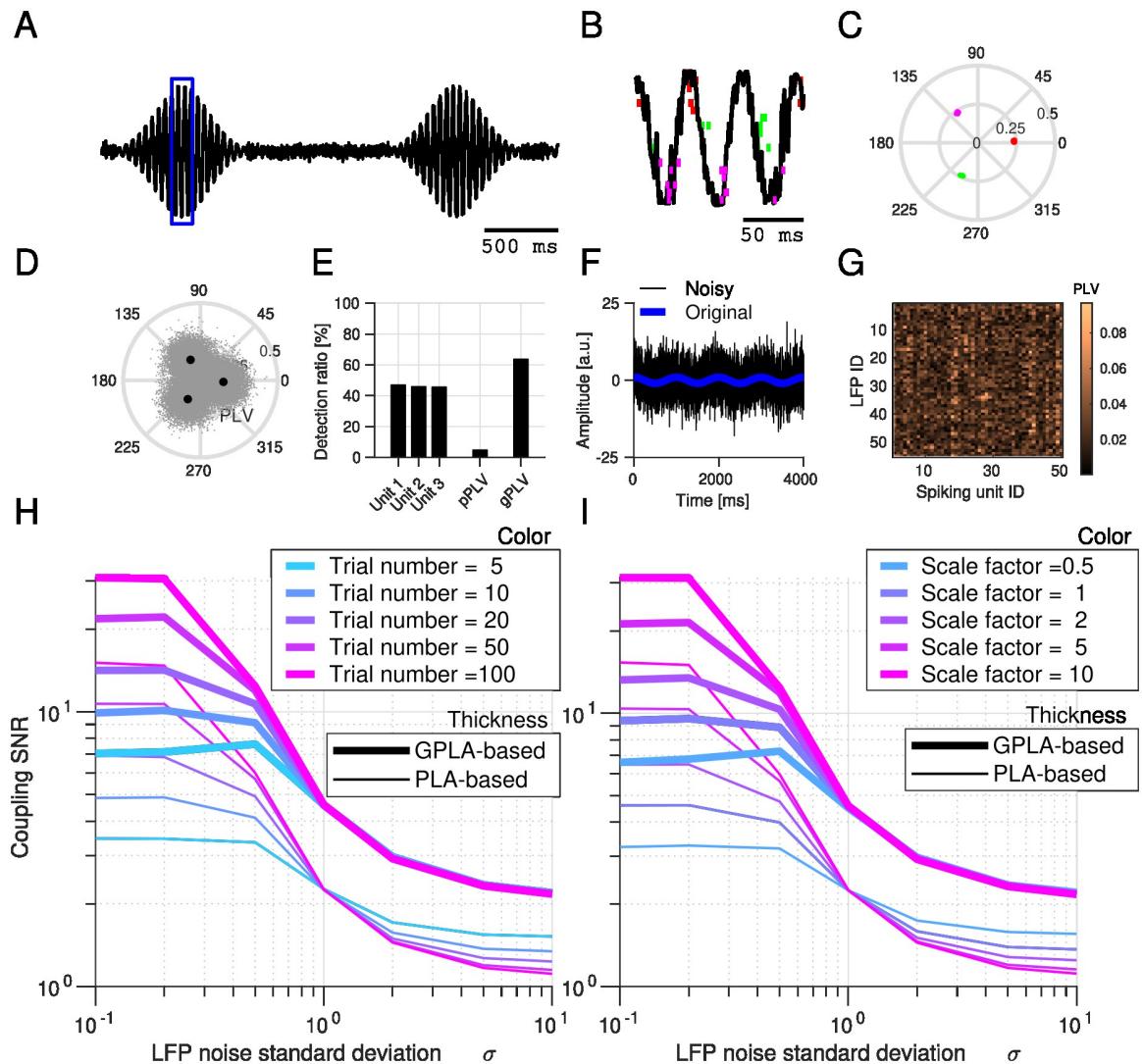


Fig 3. Comparison of GPLA and uni-variate spike-field coupling. (A) Normalized amplitude of LFP-like transient oscillatory signal with additive Gaussian white noise (used in the first simulation). (B) LFP-like signal and overlaid spike raster (colored vertical lines—colors indicate each population of units with common 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 of 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) used in the second simulation. (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 ratio of coupling strength (PLV) to estimation error (the difference between estimated PLV and the ground truth). Also see Table A in S1 Appendix for a methodological summary.

<https://doi.org/10.1371/journal.pcbi.1010983.g003>

Significance assessment of gPLV) and with a significance threshold of 5%. As it is demonstrated in Fig 3E, gPLV detection outperforms the competing approaches (PLV and pPLV).

Beyond improved detection of a significant overall coupling, GPLA-based estimation of pairwise couplings based on the approximation of Eq 3 may also 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 (see Eq 16 for the expression of normalized coupling matrix used here). To demonstrate this, we performed another simulation (Fig 3F–3I), 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 section *Simulation of phase-locked spike trains* in S1 Appendix for details), and modulating the firing rates of the units, lower firing rates leading to a larger amount of estimation variance for the PLV [59]. An example LFP trace with (black) and without (blue) noise is exemplified in Fig 3F and an example coupling matrix in the presence of noise is also illustrated in Fig 3G. In this case, the ground-truth coupling matrix has rank one, as all the units are locked to a single frequency (coupling matrices with higher ranks can also be achieved and analyzed in a similar way, see Fig 4). We ran the simulations with different amounts of LFP noise (indicated on the x-axis of Fig 3H and 3I), computed the coupling coefficients (similar to Fig 3G) and compared it to ground truth (based on Equation S2 in S1 Appendix). The Signal-to-Noise Ratio (SNR) was defined as the ratio of coupling strength (PLV) to estimation error (the difference between estimated PLV and the ground truth—for more details see section *Computing Signal-to-Noise Ratio* in S1 Appendix) and was used to compare the quality of GPLA-based and univariate estimation (indicated in the y-axis of Fig 3H and 3I). 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 (Fig 3H and 3I). Additionally, we can observe a sharp drop of the estimation quality of the GPLA-based approach as the noise increases, likely reflecting a phase transition phenomenon in high-dimensional random matrices reported in [59]: above some noise level threshold, singular value and vector information cannot be retrieved from noisy observations, while they can be recovered with very good accuracy above it. This property is further exploited in the next section.

Random matrix theory based fast significance assessment

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 also investigated this question using mathematical analysis, and exploited 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 [62, 63]. In the absence of coupling between spikes and LFP, appropriate preprocessing allows deriving analytically the asymptotic distribution of univariate and multivariate coupling measures [59], including the convergence of the squared singular values to the classical Marchenko-Pastur (MP) law [64]. Based on the MP law, we can define an upper bound on the largest singular values of the coupling matrix that depends only on its dimensions, such that exceeding this bound indicates the significance of the coupling (for more details see Materials and methods section *Analytical test* and [59]), leading to a fast 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 (denoted $O_k(t)$ in Fig 4A) with different multiplicative weights applied for

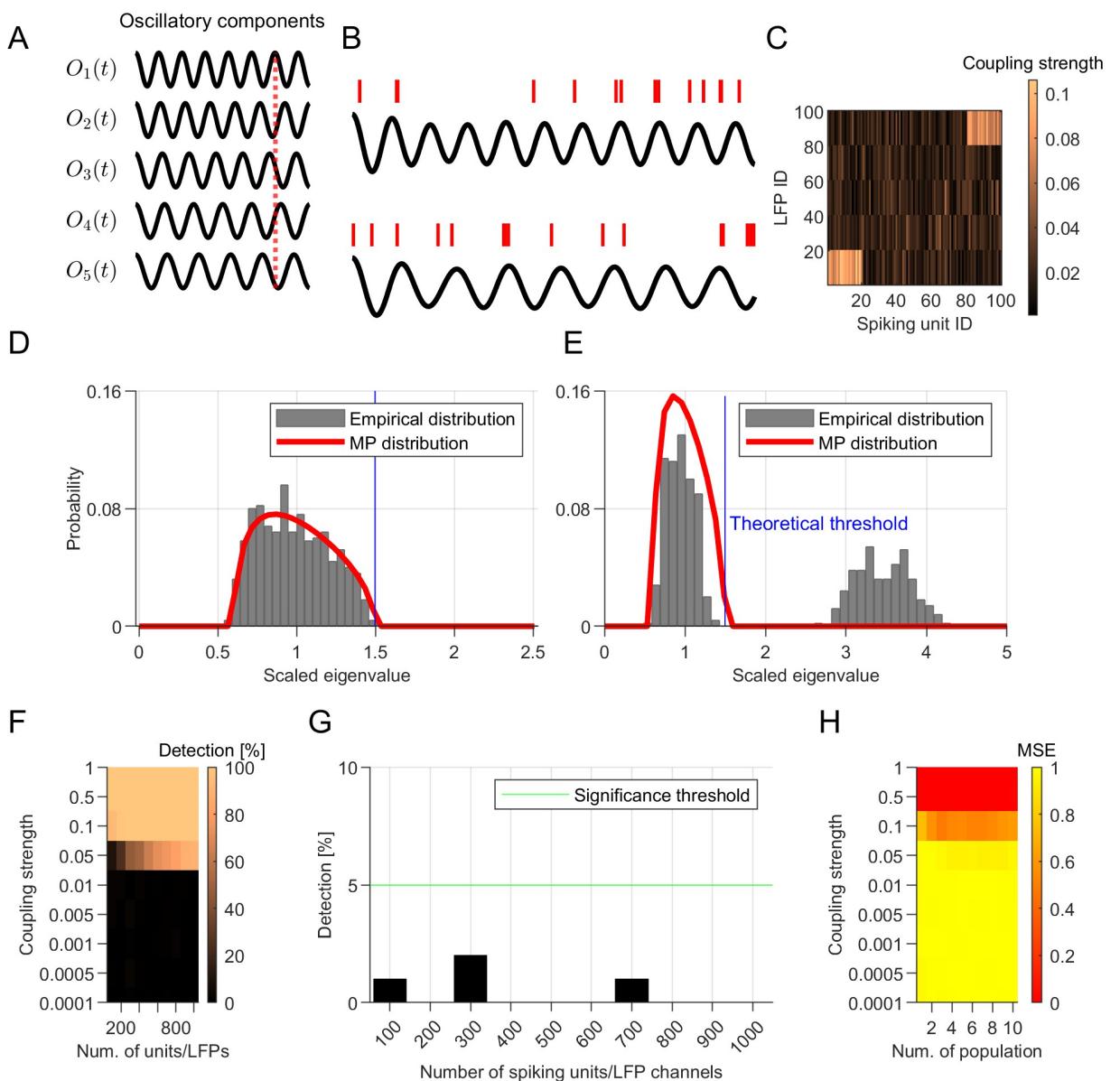


Fig 4. Statistical analysis of GPLA with a theoretical significance test. (A) LFPs are synthesized by mixing several oscillatory components ($O_k(t)$). The vertical red line evidences the phase shift between them. (B) Two exemplary spike trains (each from one of the coupled populations) and the corresponding LFPs. In the LFP trace on the top, the oscillatory component with the highest frequency is dominant while the bottom one is dominated by the lowest frequency component. (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). See also Table A for a methodological summary.

<https://doi.org/10.1371/journal.pcbi.1010983.g004>

each LFP channel and generated the spike trains of each unit with Poisson statistics. As for the coupling between spikes and LFPs, 2/5th of the units were coupled to the LFP oscillations (exemplified in Fig 4B), while the remaining units had homogeneous Poisson spike trains (for details see S1 Appendix, section *Simulation of phase-locked spike trains*). The estimated

coupling matrix computed based on Eq 17 for a simulation with 100 spike trains and 100 LFPs is exemplified in Fig 4C, where we have two coupled populations, one coupled to the lowest-frequency and one coupled to the highest-frequency oscillatory component of the LFP (reflected by the top-right and bottom-left bright blocks of the coupling matrix in Fig 4C and sample spike trains and LFP in Fig 4B).

By computing the SVD of the coupling matrix after application of the preprocessing explained in S1 Appendix, section *LFP pre-processing*, we can obtain a spectral distribution for the squared singular values, which matches the prediction of the theory (Fig 4D and 4E). In the absence of coupling between spikes and LFP signals (Fig 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 section *Analytical test* in Materials and methods for more details).

We further quantified the type I and II error of this analytical test. For type II error, we ran the simulations with non-zero coupling between spikes and LFP signals. As shown in Fig 4F, GPLA was able to detect a significant coupling between spike and LFP even when the coupling strength was as small as 0.05 (no coupling corresponds to 0 strength and perfect coupling corresponds to 1). These results also show the performance of the test does not degrade with the increasing dimension of the data through the number of recording channels (Fig 4F). This is in contrast with assessing individually the significance of pairwise couplings, for which correction for multiple comparisons (e.g., Bonferroni) would typically lead to a degradation of the power of the test as the number of units/LFPs increases. This is particularly relevant for weaker couplings, as they may lose significance after correction for multiple comparisons. Additionally, we quantified the type I error of the test by running simulations with no coupling between spikes and LFP and quantified the number of false positives. Our results show that our analytical test has a small (< 5%) false positive rate (Fig 4G).

We also quantified the performance of the method for estimating the number of populations coupled to different rhythms. In this simulation, the number of coupled populations can be determined by the number of significant singular values (see the section *Simulation of phase-locked spike trains* in S1 Appendix and section *Analytical test* in Materials and methods for more details). Similar to the simulation explained earlier (Fig 4A–4C), we simulated 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 was larger than a minimum strength of 0.5, the method was able to capture the number of populations with very low error, $MSE < 0.015$ (Fig 4H).

Neural field modeling of SFC

While the above results have addressed GPLA's outcome from a statistical perspective, its biophysical interpretation requires modeling the underlying neural network dynamics. The basis for this interpretation will be 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) [65, 66]. The model is governed by three basic input-output relations (see S1 Appendix, section *Analytical neural field modeling of spike-field coupling*) and depicted in Fig 5A: (1) the dynamics of the average somatic membrane potentials V_E and V_I of each population is governed by exogenous 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). In the

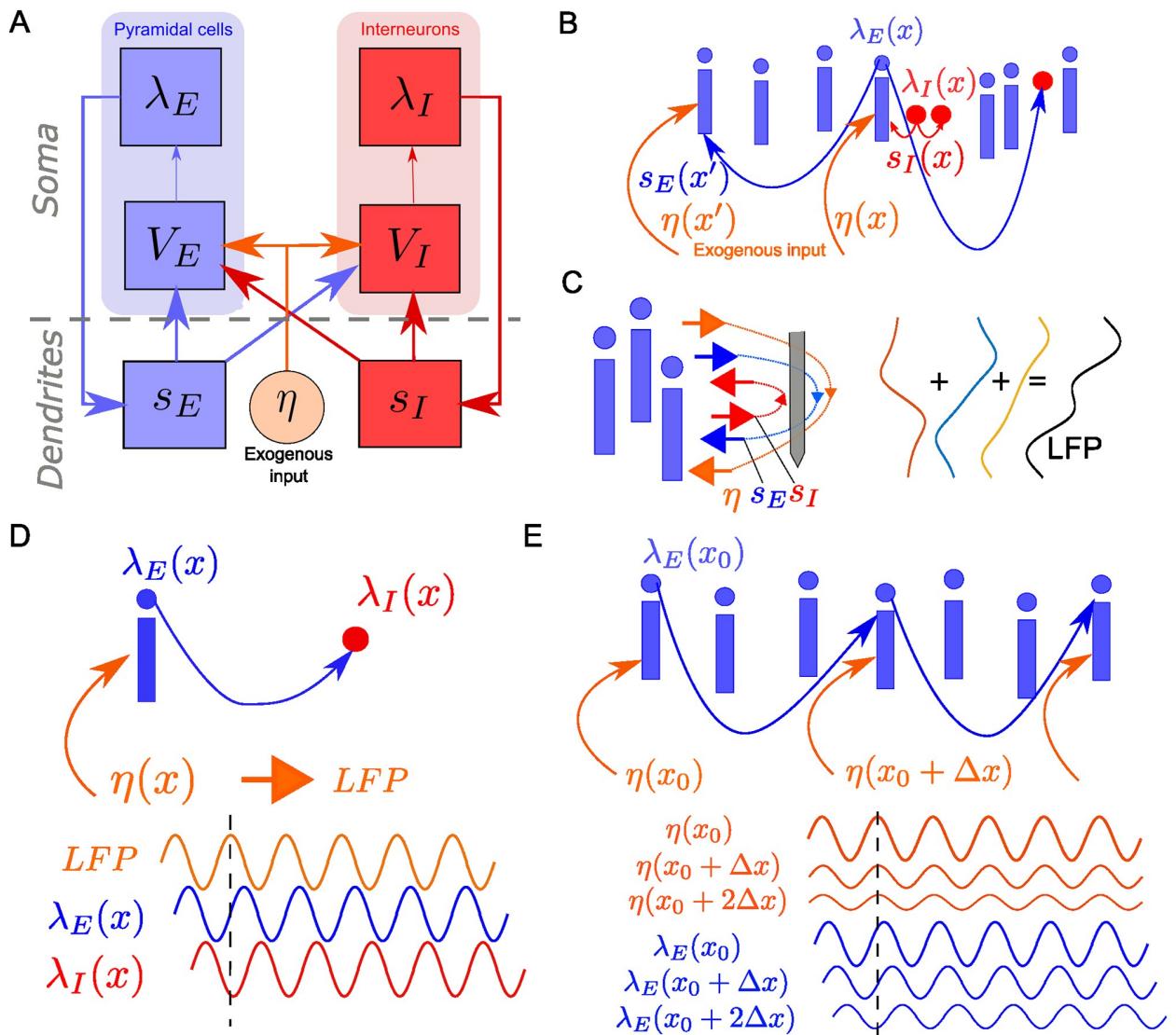


Fig 5. Generative model of spike-LFP coupling. (A) A two-population neural field model of neural dynamics. V_k , λ_k and s_k indicate respectively somatic membrane potential, firing rate and post-synaptic current for Excitatory ($k = E$) and Inhibitory ($k = I$) populations. η indicates the exogenous input to the circuit. Arrows indicate 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.

<https://doi.org/10.1371/journal.pcbi.1010983.g005>

context of large-scale recordings, the neural population can be distributed across one or several spatial directions. Following a classical approximation depicted in Fig 5B, inhibitory connections are assumed local [67–69], such that coupling between cells surrounding distinct recording locations happens exclusively through excitatory axons ($s_E(x)$ may depend on λ_E at other spatial locations than x) as well as through common exogenous input current η .

For the LFP $L(t)$, resulting from the conduction of trans-membrane currents in the extra-cellular space, we assume the contribution of currents flowing through the membrane of inter-neurons is negligible, based on the weakness of the anisotropy induced by their dendritic geometry across the population [70–72]. 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 [19, 20, 23]. Fig 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.

In the following simplistic but biophysically interpretable connectivity scenarios, this model provides insights on how the underlying microcircuit parameters influence SFC properties. Fig 5D depicts a microcircuit receiving exogenous inputs exclusively onto the pyramidal cells' dendrites (no feedforward inhibition), while I cells receive local excitatory inputs, but do not synapse back onto E cells (no feedback inhibition). If we assume additionally that sub-threshold activity is dominated by the exogenous input currents and proportional to the measured LFP, then the lag induced by the membrane potential dynamics then results in a positive (frequency-dependent) lag 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 an arbitrary frequency, this implies a phase lag configuration between the (oscillatory) responses of these variables.

Circuit assumptions may also provide insights on how the same variable varies across spatial locations, which we illustrate by extending spatially the previous microcircuit scenario (with no feedforward and feedback inhibition), by adding horizontal E-E connectivity with a decreasing strength as a function of distance (see Fig 5E). If we assume the activity results from a spatially inhomogeneous oscillatory input, with larger input amplitude at a given side (on the left in Fig 5E), the delay induced by membrane dynamics entails the propagation of the activities from one side of the circuit to the opposite. This leads to an interesting relationship between the phase and amplitude of oscillatory activity: the location of the largest amplitude is ahead of time with respect to the neighboring locations with smaller amplitudes. Interestingly, these propagation-like patterns are induced by the assumed network horizontal connectivity, while the input to the structure does not have phase lags at different locations [44]. These simple connectivity scenarios indicate that phase and amplitude of oscillatory activities, which GPLA captures through the spike and LFP vectors, are informative about the underlying microcircuit structure and dynamics. More realistic scenarios must take into account recurrent interactions between cell populations, as we will see in the next sections.

Note that up to this point, the developed neural field models can be used to interpret univariate as well as multivariate SFC. When the number of pairs for which SFC can be computed becomes large, a difficulty of a different nature appears: how can we synthesize the interpretations that we get from all these pairs? While ad hoc approaches for selecting relevant pairs to derive interpretations from is an option, we can try instead to establish interpretability of QoIs derived from GPLA, as we found support for its relevance for describing coupling properties of the system as a whole.

Reduction of complex models based on linear response theory

In order to analyze more complex circuits, a systematic and quantitative way to link model parameters to the coupling between network activities at a given frequency is required. We assume small-amplitude perturbations in the neighborhood of an operating point, such that

the static sigmoidal conversion of membrane potentials into spike rates can be linearized (see section *Analytical neural field modeling of spike-field coupling* in [S1 Appendix](#) and [55, 73–75]). 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. When considering the coupling between firing rate ($\lambda_E(x_1, t)$) and field ($L(x_2, t)$) at two locations x_1 and x_2 , linearity and time invariance entails the existence of transfer functions (denoted H_{λ_E} and H_L respectively), linking the spatial distributions of the time domain Fourier transforms of network activities, denoted $\hat{\lambda}_E(x, f)$ and $\hat{L}(x, f)$, to the one of the exogenous input $\hat{\eta}(x, f)$, as follows:

$$\hat{\lambda}_E(x, f) = \int H_{\lambda_E}(x, x', f) \hat{\eta}(x', f) dx' \quad \text{and} \quad \hat{L}(x, f) = \int H_L(x, x', f) \hat{\eta}(x', f) dx'. \quad (4)$$

Next, this model can be simplified by assuming an approximate space-frequency separability:

$$\hat{\eta}(x, f) \approx n(x)\hat{\epsilon}(f), \text{ at each location } x \text{ and each frequency } f. \quad (5)$$

Using the above transfer functions, this leads to both spike rate and LFP being proportional to the exogenous input, with respective multiplicative coefficients $\psi_E(x, f)$ and $\psi_L(x, f)$ defined as follows:

$$\hat{\lambda}_E(x, f) \approx \int H_{\lambda_E}(x, x', f) n(x') \hat{\epsilon}(f) dx' = \psi_E(x, f) \hat{\epsilon}(f) \quad \text{and} \quad \hat{L}(x, f) \approx \psi_L(x, f) \hat{\epsilon}(f). \quad (6)$$

As a consequence, the coupling between LFP and E spikes at respective locations x_1 and x_2 writes (up to a multiplicative constant, see section *Details for the low rank approximation of Equation 7* in [S1 Appendix](#))

$$C_{x_1, x_2}(f) \approx \langle \hat{\lambda}_E(x_2, f)^* \hat{L}(x_1, f) \rangle \approx \psi_L(x_1, f) \psi_E(x_2, f)^* \langle |\hat{\epsilon}(f)|^2 \rangle, \quad (7)$$

where z^* denotes the complex conjugate of z . This shows that the coupling between L at x_1 and λ_E at x_2 is separable in the spatial variables (x_1, x_2) , and characterized by two functions of space: one for the field, ψ_L , and one for the excitatory spiking, ψ_E . In particular, as $\langle |\hat{\epsilon}(f)|^2 \rangle$ is a positive number, the phase of C_{x_1, x_2} reflects a property of the underlying circuit irrespective of its input, and given by the phase difference between ψ_L and ψ_E at the considered frequency and locations. Importantly, the functions ψ_L and ψ_E also describe the coupling between the same variables at different locations, e.g. $\langle \hat{\lambda}_E(x_2, f)^* \hat{\lambda}_E(x_1, f) \rangle \approx \psi_E(x_1, f) \psi_E(x_2, f)^* \langle |\hat{\epsilon}(f)|^2 \rangle$, such that their phase distribution across locations informs about the spatial functional connectivity of the network. Likewise, ψ_I can be defined for inhibitory activity and merged with ψ_E to describe the rates of all units of both populations.

In practice, $C_{x_1, x_2}(f)$ can be measured at only a finite number of locations, corresponding to electrode channels where L , λ_E and λ_I are recorded. This leads to a rectangular matrix $C(f)$ estimated by multiple pairwise SFC estimations, combining excitatory and inhibitory units. The above separability Eq 5 then implies that $C(f)$ is a rank-one matrix, such that it can be decomposed exactly according to GPLA, where the LFP vector reflects ψ_L while the spike vector concatenating E and I units reflects ψ_E and ψ_I . Overall, Eq 6 imply that the spatial distribution of the phase and magnitude spike and LFP vectors is influenced by the underlying network interactions (shaping the transfer functions such as H_{λ_E} and H_L), as well as by the type of currents that dominate the LFP. As we will illustrate in the next sections, the analysis of these GPLA features across frequencies is thus 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 [76]. 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 each cell-type firing at a specific phase of the ripple oscillation [76], but also with further spike-field couplings at lower frequencies [77].

We use simulations of in-vivo SWR described in [78] 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 distant 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 $I - I$ synapses (see Fig 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 S1 Appendix, section *Simulation of hippocampal sharp wave-ripples* for more details).

We first apply GPLA to a single hippocampal subfield, CA1, as various studies suggest SWRs emerge from it in response to afferent CA2- and CA3-ensemble synchronous discharges [79, 80]. 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 Fig 6B.

GPLA results for representative frequency bands are provided in Fig 6C–6E and for all bands covering the 1–180Hz interval in S2 Fig. The overall coupling magnitude (g_{PLV}) was significant for all frequencies (Fig 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 was detected in the ripple band (80–180 Hz), in line with results obtained with classical univariate techniques on experimental data [76].

The LFP vectors strongly overlap across frequency bands, and exhibit a biphasic electric potential profile typical of laminar recordings (Fig 6D). This corresponds 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. 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 [78], and assuming all cells have identical trans-membrane currents flowing through their somatic and dendritic compartments (see S1 Appendix, section *Simulation of hippocampal sharp wave-ripples*). 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 populations are highly similar to the LFP vector (cosine similarity > 0.97 for LFP vector of all three frequencies in Fig 6D). Note that the sign of the LFP vectors' coefficients results from our convention of setting the phase of

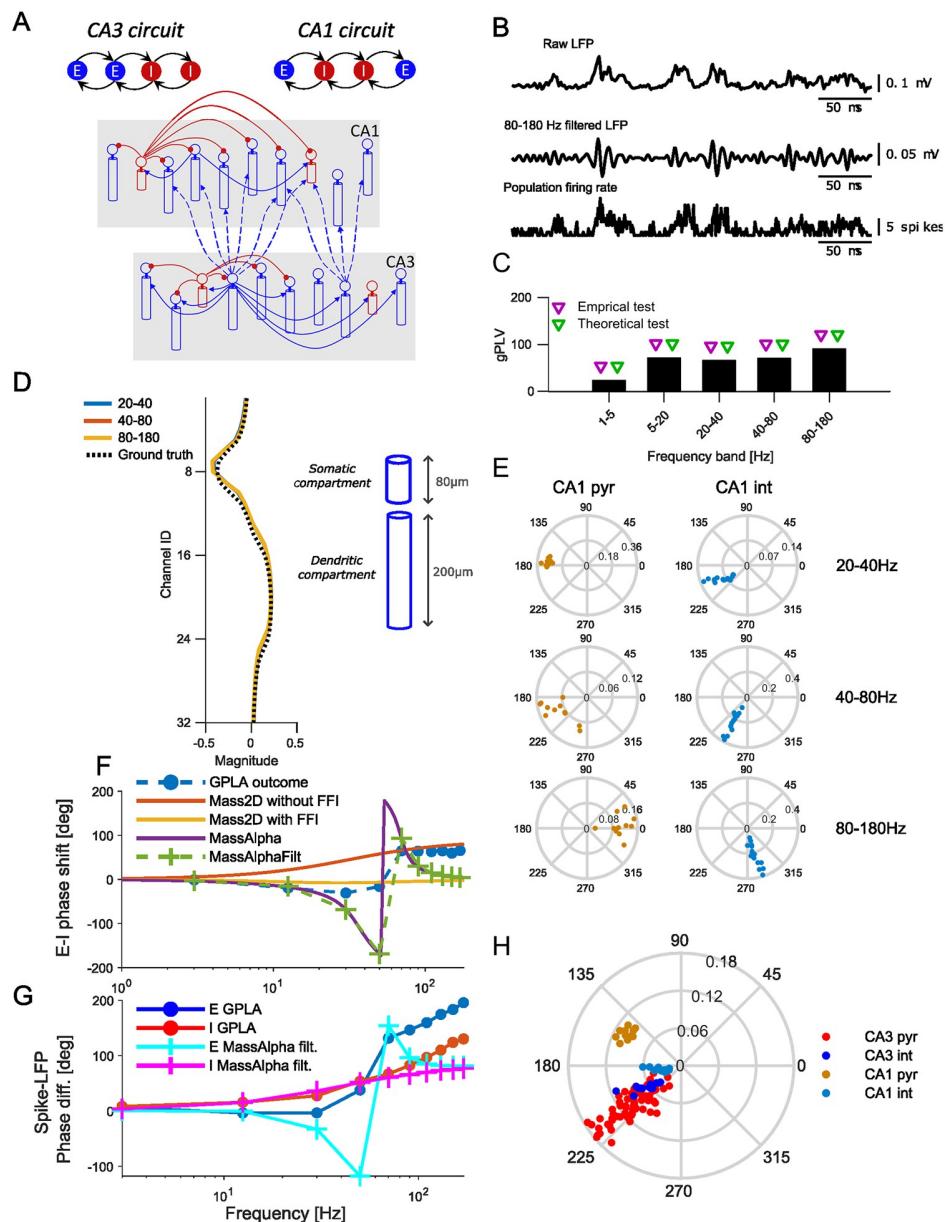


Fig 6. GPLA of hippocampal SWRs generated by a biophysical model of [78]. (A) Hippocampal multi-compartment model. Top: Canonical circuits of CA1 and CA3. Bottom: Schematic of the whole model (blue, excitatory connections; red, inhibitory). (B) From top to bottom: Example broad band CA1 LFP trace, band-pass 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) LFP vectors for GPLA of CA1 (blue and red curves are overlapping), superimposed to ground truth dipolar LFP profile passively generated by the two compartment models of the pyramidal cell population. The right-hand side schematic illustrates the vertical dimensions of one cell's compartments. (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 from GPLA jointly applied to CA1 and CA3 in the gamma band (20–40 Hz). Related Supplementary Figures: S1 Fig, Use of EPSP as LFP proxy; S2 Fig, Joint GPLA of CA3 and CA1 activities; S3 Fig, Joint GPLA of CA3 and CA1 activities.

<https://doi.org/10.1371/journal.pcbi.1010983.g006>

its mean to zero (see Fig 1C). 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 coefficients located in the peri-somatic layer (named *stratum pyramidale* in CA1). This would then lead to a sign consistent with classical analyses, e. g., triggered averaging based on spikes or oscillatory peak [81]. These results, inline with recent studies [82], overall 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 [83, 84]. Notably, the result of a similar analysis based on uni-variate phase locking analysis leads to a profile which is incompatible with the ground truth (see S4 Fig).

Moreover, the spike vector components' distribution in the complex plane (Fig 6E) supports that both E and I cells are synchronized to CA1 LFP in the ripple band (80–180 Hz), but at different phases, in line with experimental and simulation results [76, 78]. This extends the observation made for one-directional $E \rightarrow I$ coupling in Fig 5D to a more realistic case of recurrent $E - I$ interactions. Interestingly, pyramidal cells can be clearly differentiated from interneurons based only on their components' respective phase in the spike vector, showing that interneurons lead pyramidal cells in lower frequency bands, while drastically switch to the converse in high frequencies (see also Fig 6F). This direct outcome of GPLA avoids the task of choosing a reference LFP channel on an ad hoc basis to compare the phase of univariate couplings of each units relative to it. Moreover, it can be used not only for inferring cell types from experimental data [85], but also, based on its biophysical interpretability, to address mechanistic questions, as we illustrate next.

We focus on the classical question of oscillogensis, aiming at uncovering the circuit mechanisms responsible for the emergence of fast oscillations. The way it is addressed in the literature is paradigmatic of mechanistic questions: scientists resort to experimentation and modelling to chose between a restricted number of candidate hypotheses. Two classical candidate mechanisms are the Interneuron Network Gamma (ING) relying on the coupling between inhibitory interneurons under tonic excitation [86], and Pyramidal Interneuron Network Gamma (PING) relying on the interaction between excitatory principal cells and inhibitory interneurons [87, 88]. We take advantage of these biophysically realistic hippocampal simulations, for which the ING mechanism has been shown to be the generator of high frequency activity [78], to assess how biophysical interpretability can help decide which of the two above hypothesis is the right one. To do that, we will derive analytically SFC's phase in linearized neural mass models of the microcircuit activity with different levels of complexity: the simplest accounting for PING, and the more complex also accounting for ING.

In line with [74], we first designed the *Mass2D* model, taking into account somatic time constants (resulting from membrane capacitance and leak currents), but neglecting synaptic dynamics (see S1 Appendix, section *Analysis and simulation of two population neural mass models*). As a result, *Mass2D* is a 2 dimensional dynamical system, allowing only PING resonance through the interactions between pyramidal cells and interneurons. As shown in Fig 6F for typical parameters, and demonstrated analytically (see S1 Appendix, section *Analysis and simulation of two population neural mass models*) the predicted phase shift across frequencies could neither account for the driving by interneurons in CA1, nor for phase changes in high frequencies ($> 30\text{Hz}$). 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 [89], as *Mass2D* does not account for them.

The emergence of oscillations through $I - I$ interactions is well understood mathematically, showing that sufficiently strong delayed recurrent inhibition gives rise to resonance or sustained oscillations [90]. We account for this ING mechanism in an extension of the *Mass2D* model, the *MassAlpha* model, by including an additional synaptic delay and/or a synaptic time constant for $I - I$ synapses [74], through the use of so-called *alpha synapses*. (see [S1 Appendix](#), section *Analysis and simulation of two population neural mass models* for details). Interestingly, the resulting sign of the phase shift between E and I populations of this model is now in qualitative agreement with GPLA estimation ([Fig 6F](#)), exhibiting a reversal in the lead-lag relation between populations as frequency grows, thereby providing more support for the ING oscillogenesis hypothesis than for PING, in line with evidence provided in the original study [78]. The SFC phase is thus biologically interpretable for the chosen family of neural mass models, in the sense that a phase reversal across the frequency axis appears when lagged $I - I$ interactions responsible for ING are introduced. Because this phase reversal also appears in simulations exhibiting ING that rely on a much more complex model (Hodgkin-Huxley neurons instead of neural masses), this supports the idea that biophysical interpretations of SFC based on our simplified models may generalize to more realistic settings and to experimental recordings.

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), due to our chosen phase convention (see [Eq 23](#)). Given that the LFP is a linear combination of all post-synaptic currents of the network, we can leverage biological interpretability of GPLA to evaluate which of these currents is the most representative of the observed spike-LFP phase relation. As shown in [Fig 6G](#), 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 (see [S1 Fig](#) for a comparison with using the EPSP as LFP proxy). In contrast, using EPSP as an LFP proxy still fails to reproduce these two aspects (see [S1 Fig](#)), illustrating how GPLA, beyond microcircuit dynamics, may also help address the cellular underpinnings of experimentally observed LFP [91]. 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. We however emphasize that we restricted ourselves to a qualitative comparison of GPLA features for choosing from a restricted set of biophysical models, which best matches the ground truth mechanisms. This approach holds potential for designing a full-fledged GPLA-based model selection tool, whose development is left to future work.

Importantly, GPLA can also provide further insights when concurrent recordings from multiple regions are available. It allows investigating the coordination of spiking activities across structures without relying on an arbitrary choice of reference LFP channel (also see the analysis of neural data for a realistic demonstration, [S8 Fig](#)), by automatically extracting a multi-channel LFP activity (reflected by the LFP vector) that relates the most to spiking activities at a given frequency. We illustrate this by running GPLA jointly on spikes and LFPs from both CA1 and its afferent structure CA3, using the exact same model as above. [Fig 6H](#) depicts coefficients of the resulting spike vector, 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 [S3 Fig](#)) that are consistent with the GPLA obtained from individual structures (see [S2 Fig](#)). This notably suggests that the gamma activity has a dominant coherent component spanning the two structures consistently with current hypotheses that this rhythm supports communication between sub-fields during memory trace replay [78, 92].

Application to spatio-temporal patterns of neural field models

One context where biophysically interpretable multivariate methods such as GPLA hold potential is the analysis of cortical spatio-temporal dynamics. Horizontal connectivity is believed to endow many regions with distributed information processing capabilities [30, 44, 93]. However, how underlying connectivity properties relate to experimentally observed multi-channel recordings remains largely elusive. We assessed the ability of GPLA to address this question by first simulating electrode array recordings of a piece of cortical surface with a 2D neural field model, as described in Fig 5. We used an exponentially decaying horizontal excitatory connectivity with a spatial scale constant $r_0 = 440\mu\text{m}$, following recent analyses of cortical recordings [91]. The spatio-temporal dynamics were down-sampled spatially on a grid with a step size $\Delta x = 800\mu\text{m}$, representing the inter-electrode distance of a putative electrode array of 1.2cm size (see S1 Appendix, section *Analysis and simulations of neural field models* for details). The field is stimulated by a synchronous excitatory exogenous input with a narrow (1.4 mm STD) isotropic Gaussian spatial amplitude distribution reaching its maximum at the center of the field. 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 (Fig 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 (Fig 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 (Fig 7B) [94]. 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 [95]. GPLA is then computed between this excitatory spiking activity and different LFP proxies. The results in Fig 7C–7H 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 S1 Appendix section *Analysis and simulations of neural field models*).

First, as reflected in the gPLV values (Fig 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 [96], 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 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 (Fig 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 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 ($p < 10^{-4}$, F-test on the linear regression model; $N = 69$), whereas, in the simulation with weak

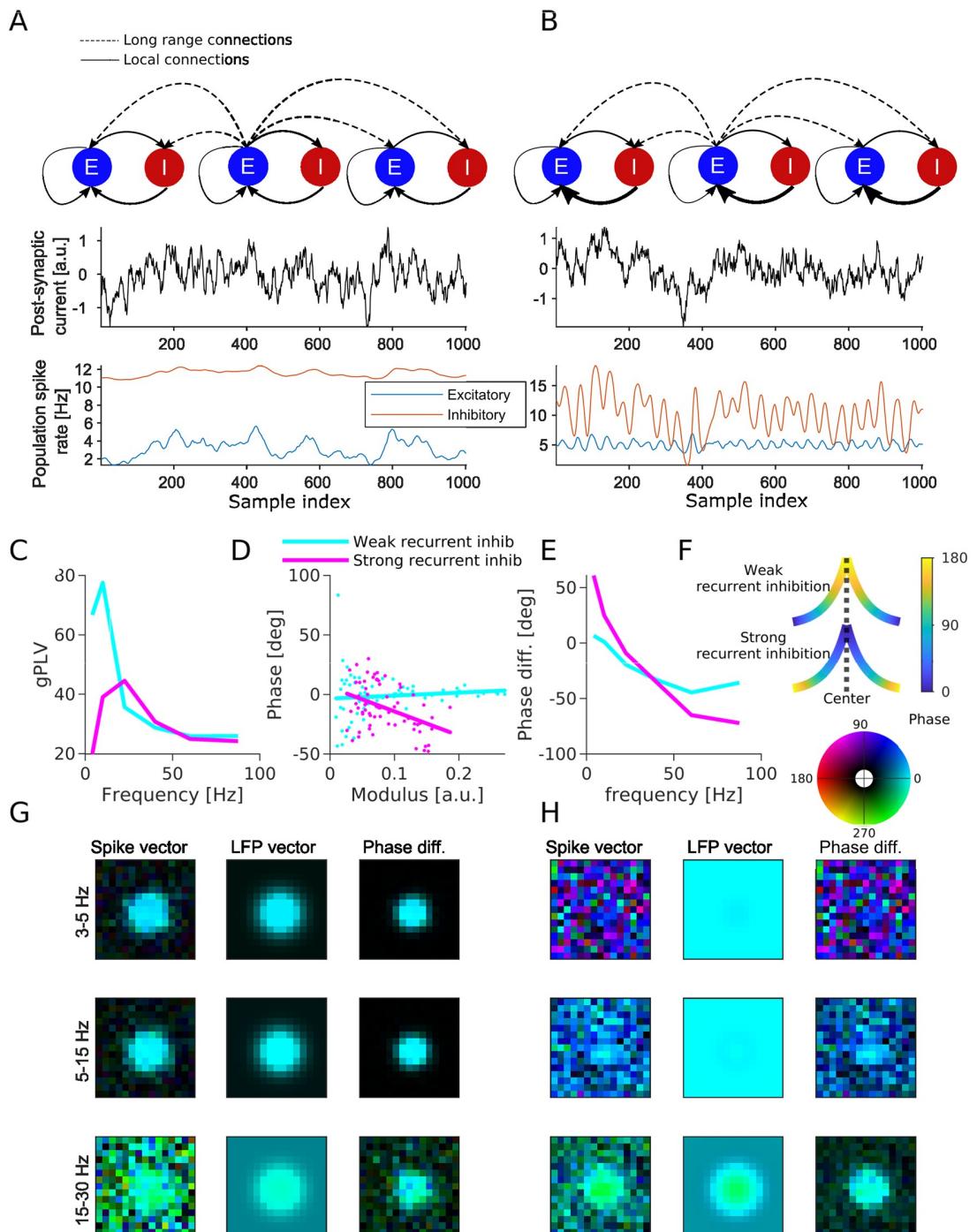


Fig 7. Neural field simulation using EPSP as LFP proxy. (A) Simulation with weak recurrent inhibition. Example time course at center location for exogenous input (top), E- and I- populations rates (bottom). (B) Same as A for strong recurrent inhibition. (C) gPLV as a function of frequency for both models. (D) Phase of spike vector coefficients as a function of their modulus for the frequency band yielding maximum gPLV for both models (each dot one coefficient, and the continuous lines are plotted based on linear regression). (E) Shift between 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 Eq 8. 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 weak recurrent inhibition (model schematized in A). (H) Same as G for 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). Related supplementary Figures: [S5 Fig](#), Phase-modulus relation dependency on level of inhibition; [S6 Fig](#), GPLA using IPSP as LFP proxy.

<https://doi.org/10.1371/journal.pcbi.1010983.g007>

recurrent inhibition, phase is not correlated with modulus ($p > 0.3$, F-test on the linear regression model; $N = 69$) ([Fig 7D](#)).

This last difference between the two connectivity 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 ([Fig 7G and 7H](#)), reflecting how phase changes as magnitude decreases when going away from the center (the location with the largest input). This gradient can be predicted by theoretical analysis of a one dimensional neural field, as we show in detail in [S1 Appendix](#), section *Spatio-temporal phase analysis in 1D*. Briefly, the spike vector can be approximated by the spatial convolution of the input spatial pattern at a given temporal frequency f by a kernel of the form

$$k(x) = e^{-|x|a(f)} = e^{-|x|\operatorname{Re}[a(f)]} e^{-i|x|\operatorname{Im}[a(f)]}. \quad (8)$$

The first term of this kernel has a negative real number multiplied by distance in the exponential that makes the activity decay away from the locations where exogenous input is the highest, as intuitively expected from the horizontal connectivity of the circuit. For the second term of the product in [Eq 8](#), the imaginary number in the argument of the exponential 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 the 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 [Fig 7F](#). On the contrary, if $\operatorname{Im}[a]$ is negative, locations with the highest input are lagging behind (bottom illustration of [Fig 7F](#)). Interestingly, these spatial features of the spike vector can be related to the biophysical parameters of the neural field model. Indeed, we can show that the frequency-dependent complex number a (f) that controls this behavior satisfies the approximate relation (valid at low frequencies, see [S1 Appendix](#) section *Spatio-temporal phase analysis in 1D*

$$a^2 \approx \frac{1}{r_0^2} [1 + v_{E \rightarrow I} v_{I \rightarrow E} - v_{E \rightarrow E} - i2\pi\tau f (2v_{E \rightarrow I} v_{I \rightarrow E} - v_{E \rightarrow E})]. \quad (9)$$

with r_0 the above defined spatial scale of excitatory horizontal connectivity, $v_{P \leftarrow Q}$ the magnitude of synaptic connectivity from population Q to P . It can be deduced from this expression that the sign of the imaginary part of a (same as for a^2) will depend on the relative strength of recurrent inhibition onto pyramidal cells, controlled by $v_{E \leftarrow I} v_{I \leftarrow E}$, with respect to recurrent excitation controlled by $v_{E \leftarrow E}$. Intuitively, having no recurrent inhibition leads to $\operatorname{Im}[a] > 0$, and classical propagation, mediated by excitatory horizontal connections, away from the location that received an input. In contrast, large recurrent inhibition leads to $2v_{E \leftarrow I} v_{I \leftarrow E} > > v_{E \leftarrow E}$ and $\operatorname{Im}[a] < 0$. This can be interpreted as a tendency of recurrent inhibition to “suppress” the input that created the response, generating a “wave” converging back to the points where the input was highest. The theory also predicts that large values of $v_{E \leftarrow I} v_{I \leftarrow E}$, as used in the strongly recurrent simulation, can generate strong phase gradients. In contrast, linear stability constrains the values of $v_{E \leftarrow E}$ to remain small, reflecting our choice for the simulations, and resulting in a comparatively moderate slope for the weakly recurrent case. More quantitatively, we further analyzed in [S5 Fig](#) the relation between the complex number a resulting from a linear approximation of our simulated neural field models, and the linear regression coefficient of the phase-modulus analysis performed in ([Fig 7D](#)), for four choices of recurrent inhibition parameters (see [Table 2](#)), ranging

Table 2. List of 2D neural field model parameters.

| Parameter name | Symbol | Value (for each level of recurrent inhibition) | | | |
|-------------------------------------|-------------------------------|--|------------|------------|--------|
| | | weak | lower med. | upper med. | strong |
| E membrane time constant | τ_E | 20ms | 20ms | 20ms | 20ms |
| I membrane time constant | τ_I | 20ms | 20ms | 20ms | 20ms |
| $E \rightarrow E$ synaptic strength | $\tilde{v}_{E \rightarrow E}$ | 0.2 | 0.2 | 0.2 | 0.2 |
| $I \rightarrow I$ synaptic strength | $\tilde{v}_{I \rightarrow I}$ | 0 | 0 | 0 | 0 |
| $E \rightarrow I$ synaptic strength | $\tilde{v}_{E \rightarrow I}$ | 0.2 | 0.2 | 0.2 | 0.2 |
| $I \rightarrow E$ synaptic strength | $\tilde{v}_{I \rightarrow E}$ | 1 | 1 | 1 | 1 |
| E excitability | χ_E | 1 | 1 | 1 | 1 |
| I excitability | χ_I | 0.1 | .33 | 1 | 3.33 |
| E sigmoid threshold | $V_{th,E}$ | 0 | 0 | 0 | 0 |
| I sigmoid threshold | $V_{th,I}$ | 0 | 1 | 5 | 5 |
| E maximum rate | Q_E | 20Hz | 20Hz | 20Hz | 20Hz |
| I maximum rate | Q_I | 20Hz | 20Hz | 20Hz | 20Hz |

<https://doi.org/10.1371/journal.pcbi.1010983.t002>

from weak to strong inhibition. The result exhibits a clear monotonous relation between the regression coefficient and $\text{Im}[a]$ as well as $\frac{\text{Im}[a]}{\text{Re}[a]}$. Note however that this relation is not one to one, as a characterizes the properties of a kernel that is convolved to the exogenous input to the structure to yield the spike vector, thereby resulting in a spatial smoothing of the phase.

Overall, contrasting multiple cases shows that modifications of the strength of feedback inhibition are reflected not only in the dominant frequency of spike-LFP synchronization (Fig 7C), but also in the spike-LFP shifts of the GPLA results (Fig 7E), and in the relationship between modulus and phase of spike vector coefficients (Fig 7D). Notably, these observations are being made in the 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, such traveling waves-like phase gradients, 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.

As it has been argued in the literature that LFP activity may in some cases reflect inhibitory activity [91], we also provide GPLA results when taking the IPSP activity as LFP proxy in S6 Fig. The variations of GPLA features across the frequency axis witness clear differences with respect to the results of Fig 7C–7E, in particular when it comes to the phase difference between spike and LFP phases. This suggests that GPLA also provides information that allow to infer which neural processes are reflected in LFP activity.

Analysis of Utah array data in the prefrontal cortex

The biophysical interpretability of GPLA features demonstrated in the context of neural field simulations suggests it can provide mechanistic insights about experimental recordings of spatio-temporal cortical activity. Indeed, electrode arrays are able to record the activity of hundreds of units and LFP channels spatially distributed along the cortical surface, and GPLA can be used to link these activities to recurrent cortical circuits, believed to play a key role in information processing. We apply GPLA to Utah array (10×10 electrodes, inter-electrode distance $400\mu\text{m}$) recordings from the ventrolateral prefrontal cortex of one anaesthetized rhesus monkey (see Fig 8A). LFP signals were preprocessed as described in S1 Appendix, section *Animal*

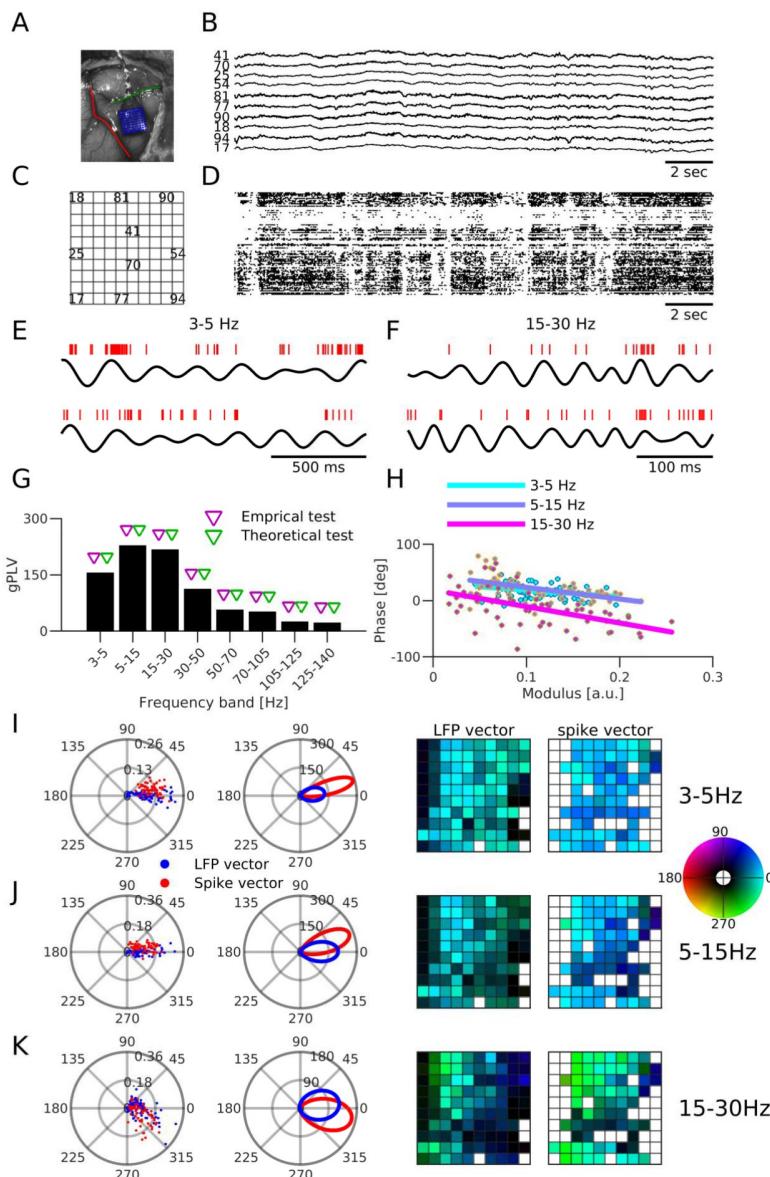


Fig 8. Application to electrophysiological recordings in non-human primate PFC. (A) Location of the Utah array, anterior to the arcuate sulcus (red line) and inferior to the principal sulcus (green line). (B) Broadband trace of the recorded LFP (from the recording channels indicated in C). (C) Utah array 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, (J) 5–15 Hz, and (K) 15–30 Hz. The first column depicts the LFP (blue dots) and spike (red dots) in the complex plane. The 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 fourth 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 ($r e^{i\phi}$) is represented by hue and brightness of a pixel. The 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. Related supplementary Figure: S7 Fig, Analysis of PFC Utah array data.

<https://doi.org/10.1371/journal.pcbi.1010983.g008>

preparation and intracortical recordings, and multi-unit activity with a minimum of 5 Hz firing rate was used. Recorded signals are exemplified in Fig 8B–8F. Exemplary LFP traces are illustrated in Fig 8B. Each trace is recorded from the location specified in Fig 8C. Spike trains are also displayed in Fig 8D (for the same epoch used in Fig 8B). As the analysis is performed in band-limited frequency ranges, we also exemplified band-passed LFP signals (together with spikes) in Fig 8E and 8F. The dataset consisted of 200 trials of visual stimulation (10 sec) and inter-trials (10 sec) each 20 sec.

Computing GPLA in different frequency bands revealed that the strongest coupling was in the alpha range (5–15Hz) (Fig 8G). Furthermore, we assessed the significance of coupling with both surrogate and analytical tests (see [Materials and 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 may be affected by the contamination of high frequency LFP bands by spike waveforms of units recorded in the same channel [97, 98]. This may bias spike-LFP coupling towards the specific relation between the spiking of those specific units and the surrounding field, instead of capturing the relation of the underlying population rate to this field, as assumed in neural field models.

Fig 8I–8K further shows the spike and LFP vectors for the three frequencies with the largest coupling according to their gPLVs (for other frequencies, see [S7 Fig](#)). Representing spike and LFP vectors in the complex plane (Fig 8I–8K 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 (Fig 8I–8K 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 (Fig 7G), when EPSP is taken as an LFP proxy.

The spatial mappings of the LFP and spike vectors on the Utah array (Fig 8I–8K, 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 Fig 8K, fourth column). We hypothesize this is due to a higher activation of spatially localized populations, as supported by anatomical studies of the PFC [99, 100] and electrophysiological [101] studies. Notably, capturing this aspect of the circuitry from the neural data based uni-variate phase locking analysis relies on finding a suitable choice of LFP reference channel, which is typically challenging (see [S8 Fig](#), for comparison of multi-variate analysis and examples of uni-variate based on two different choices of reference channel).

Furthermore, in 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 (Fig 8H, $p < 10^{-6}$, F-test on the linear regression model; $N = 66$). 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 (Fig 7D). Notably, the result of a similar analysis based on uni-variate phase locking analysis leads to a profile incompatible with our conclusion based on neural field simulation (see [S9 Fig](#)).

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 [96, 102]. This analysis illustrates how GPLA can support the mechanistic understanding of high-dimensional experimental recordings.

Discussion

In spite of the relevance of spike-field relationships for assaying coordination mechanisms in brain networks [19, 20, 22–24], they are still not systematically investigated in the context of highly multivariate recordings. Potential reasons could be the lack of multivariate methodologies for investigating such coupling beyond a single pair of spiking unit and a LFP channel, and interpretability challenges.

In this study, we developed Generalized Phase Locking Analysis (GPLA) as—to the best of our knowledge—the first *multivariate* method with *demonstrated biophysically interpretability* 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 concise way. At a given frequency, the spike and LFP vectors represent the dominant LFP and spiking spatio-temporal distribution, while the generalized Phase Locking Value (gPLV) characterizes the strength of the coupling between LFP and spike patterns.

Some of the conclusions we draw based on GPLA may to some extent also be achievable with univariate techniques, *but in contrast to GPLA, this typically requires ad hoc decisions or guiding univariate methods with considerable amount of prior knowledge on the structure under study*. For instance, univariate techniques can be used for analyses we described in Fig 6E and 6H, provided a suitable LFP reference channel is used to assess the coupling of all recorded units. Choosing such channel is not trivial unless it is justified, for example, by prior knowledge on the hippocampal circuitry. Even in such case, prior knowledge may not reflect accurately the properties of the recordings and bias the analysis. An arbitrary choice of reference channel will not faithfully reflect the dominant coherent activity with units primarily synchronize. Certainly, such caveats are even more pronounced when investigating structures with less prior knowledge and with recording techniques yielding a larger number of channels.

We demonstrated that GPLA’s outcome features, such as the overall spike-LFP phase shift, the phase shift between different cell types (excitatory and inhibitory), and the spatial phase gradients, provide information about the overall organization of the recorded structure that are not easily quantifiable with simpler measurements.

First, application to realistic simulations of hippocampal SWR revealed various characteristics of hippocampal circuitry with minimal prior knowledge. Second, in order to better interpret spike and LFP vectors’ spatial distribution, we also simulated spatially extended neural field models and demonstrated that phase gradients of spike and LFP vectors in these neural field models reflect properties of 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 gPLV were investigated to develop an empirical and theoretical framework for assessing the significance of coupling. The theoretical statistical test built upon Random Matrix Theory [59] makes the method applicable to high dimensional electrophysiology data with low run-time complexity, which is important for modern probes such as Neuro-pixel, featuring 960 recording sites [40]. In contrast, conventional statistical testing procedures based on the generation of surrogate data become computationally expensive as the number of recorded neurons increases.

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. In

particular, among approaches exploiting multiple LFP channels, none fully exploit the statistical relation between spiking activity recorded from multiple sites.

The Spike-Triggered Average (STA) of LFP is one of the common multivariate technique for characterizing spike-LFP relationship [103, 104]. It has moreover been interpreted as a measure of functional connectivity [104] (but also see [105] and [106]). Although STA can exploit multivariate LFP signals, it can only be computed based on a single spike train, thereby ignoring the information provided by the remaining units. Similarly, even sophisticated extensions of spike-triggered averaging of LFP [91] still rely on the information of individual spiking units. In a similar vein, the study of [107], which showed that the probability of spiking can be statistically related to the LFP phase in multiple distant regions, was also limited to spiking units taken individually.

This appears clearly as a limitation, because statistical relationships between the spiking activity of different units, such as lags between the activity of different types of neurons (e.g. excitatory and inhibitory neurons) [108] can inform us about the organization of the neural circuit. Notably, this is supported by our simulations and previous experimental work [85].

Apart from works that specifically target spike-field coupling, a body of methodological studies by *van der Meij* and colleagues use the idea of extracting a dominant frequency coupling structure with dimensionality reduction techniques [52–54], akin to GPLA's principle. In spite of the similarities between these methodologies from a data analysis perspective, GPLA-based investigation of spike-LFP coupling further leads to a biophysical interpretations in terms of underlying circuit properties, while this key question is left unaddressed by other approaches. This is due to the ability of GPLA to allow both dimensionality reduction (of experimental recordings) and model reduction (of neural field model) such that the outcome of both reductions can be related.

Limitations and potential extensions

One limitation of GPLA is that it considers the underlying network dynamics to be fixed for the analyzed data. Although the use of GPLA on simulation of Hippocampal Sharp Wave-Ripples demonstrates that an application on even such transient and aperiodic signals is insightful (Fig 6), but certainly due to the non-stationarity of neural dynamics, the time-resolved analysis of spike-LFP data (that likely required further methodological development) may improve our understanding of the underlying processes. As an alternative, it is however possible to apply the present methodology to portions of recordings containing identified transient phenomena, such as hippocampal Sharp Wave-Ripples, that are likely key to understand brain function [109, 110]. For example, as 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.

Another limitation comes from the nature of SVD, leading to orthogonal singular vectors that may or may not capture the properties of distinct physiological processes. We therefore mostly limited the scope of this study to capturing the dominant coupling between spikes and LFP, reflected in the largest SV and corresponding vectors. However, as simulations used in Fig 4 demonstrate in simple cases, the number of significant SVs may correctly identify the number of neuronal population coupled to different rhythms. In general, the amount of information neglected by limiting the analysis to the largest singular value, highly depends on the settings under study. We have demonstrated a variety of them in this manuscript. For instance, the simulation used in Fig 3 exemplifies a small loss, and for Fig 6 a large one. Certainly, more

quantitative approaches can also be taken, for instance, by quantifying this loss by the ratio of the largest singular value to the sum of all singular values.

Furthermore, GPLA can also be improved by exploiting a better univariate estimation method. Various novel methodologies for assessing pairwise spike-field coupling have been developed in recent years [34–38, 111, 112] each providing some improvements over classical measures such as PLV. For instance, [38] proposed a bias-free estimation of spike-LFP coupling in the low firing rate regime. Replacing the coefficients of coupling matrix (Eq 17) with these improved pairwise estimates may bring those benefits to GPLA as well. Nevertheless, the pairwise estimate used in the present paper has the benefit of yielding well behaved statistical properties as the number of recording channels gets large, allowing to quickly assess the significance of the coupling using Random Matrix Theory. Alternative pairwise coupling estimates would likely need to be adapted in order to preserve the statistical benefits of our approach. This typically requires calculating the asymptotic distribution of the coupling statistics and devising and appropriate normalization thereafter. In case the new coupling measures are not adaptable to the analytical test, the surrogate-based test 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, interpreting the outcome of statistically sound approaches such as GPLA in terms of biological mechanisms remains challenging. Thanks to the analysis of neural mass/field models, we could link several features of GPLA to a mechanistic interpretation. First, we applied this strategy to simulations from a biophysically realistic model of hippocampal ripples in order to use a system for which the underlying mechanism are well understood, but more complex than the neural field 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 interpretable. This allowed in particular (1) to relate the LFP vector to 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 cortical electrode array recordings. This is an important step as the mechanisms underlying spatio-temporal phenomena observed *in vivo* remain largely elusive. While keeping the complexity of these models minimal (using exponentially decaying horizontal excitation and local inhibition), we could already observe that altering the microcircuit structure resulted in interpretable qualitative modifications of GPLA's 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 into the underlying organization of the recorded circuits.

As mentioned when introducing the concept of biophysical interpretability, the reliability of mechanistic interpretations drawn from GPLA crucially depends on the ability of the reduced biophysical models that we use to approximate key ground truth mechanisms underlying the data. Although no absolute guarantees can be provided, we showed in two sets of simulated data that the linearized neural field approximations provided qualitative insights in line with ground truth mechanisms, which were based on more complex (notably non-linear) models. 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. Neural mass modeling has of course inherent limitations due to approximating local population activity by their mean rate, such as their typical inability to account for synchronization of spike times. However, multiple refinements of these models have been developed and offer potential for improving biological realism. Notably, next generation neural mass models are able to capture event-related synchronization between neurons [113, 114] and can incorporate the dynamics of intrinsic currents that are key to modeling complex phenomena empirically observed, such as spindle oscillations [115]. In addition, neural field models can be improved in light of the knowledge about the horizontal connectivity of the structure, which may not be monotonous (for example see recent findings on non-monotonic correlation structure in V1 [116] and PFC [101]), and heterogeneous [117].

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. One aspect that entails limitations is the linear response theory on which we base our interpretations in the present work. Linearization is typically justified for a stable system exhibiting low amplitude fluctuations around its equilibrium point. However, non-linear model reduction techniques such as the Galerkin method [118] allows to extend low dimensional, interpretable approximations of high-dimensional systems to more general settings.

The investigation of more complex models will benefit from incorporating systematic parameter estimation approaches, taking inspiration from inference techniques that have been developed for modeling the activity of one or several neurons [119, 120] and combining them with the present model and dimensionality reduction approaches to ensure tractable estimation.

Ultimately, our results support the relevance of GPLA for studying distributed information processing in higher-tier cortical areas such as PFC and hippocampus, where spike-LFP interactions have proven key to elucidating the neural basis of cognitive functions such as working memory [121, 122], memory consolidation and spatial navigation [123, 124]. This approach is likely to provide further insights about coordination mechanisms by shifting the focus from properties of individual units to characteristics of spatially extended networks taken as a whole.

Materials and methods

Ethics statement

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 [125]] (more details on these experiments are provided in a previous study exploiting this data by [101]). All experiments were approved by the local government authorities (Regierungspräsidium, Tübingen, Baden-Württemberg, Germany), and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals.

GPLA 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 [58]. On the other hand, the filtered signal should preserve the temporal dynamics in the frequency of interest. The second step extracts 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 signals recorded in different channels by applying a whitening operator. Whitening is only necessary to be able to use tools from Random Matrix Theory [62] for the purpose of statistical analysis, otherwise generalized phase locking value, spike and LFP vectors can all be calculated in the original channel space. In both cases, GPLA outputs can be interpreted in the channel space (by inverting the whitening operation if it has been applied). For more detail on the rationale for the inclusion of the whitening step, see section *Analytical test* and [59].

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 decorrelates 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 account for 99% of the total variance of the 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,

$$\text{Cov}(L) = \frac{1}{T} LL^H, \quad (10)$$

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 $\text{Cov}^{\text{red}}(L^{(k)})$, and computed as follows,

$$\text{Cov}^{\text{red}}(L^{(k)}) = \sum_{p=1}^{n_c^{\text{eff}}} \lambda_p^{(k)} x_p^{(k)} (x_p^{(k)})^H, \quad (11)$$

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,

$$L_w^{(k)} = (\Lambda^{(k)})^{-\frac{1}{2}} (X^{(k)})^H L^{(k)}, \quad (12)$$

where $\Lambda^{(k)}$ is a $n_c^{\text{eff}} \times n_c^{\text{eff}}$ diagonal matrix containing the eigenvalues of the above single-trial reduced rank LFP covariance matrix, and $X^{(k)}$ is a $n_c \times n_c^{\text{eff}}$ matrix containing the eigenvectors $x_k^{(k)}$.

Coupling matrix. Given the spike times of a single spike train $\{t_j^{(k)}\}_{j=1 \dots N^{(k)}}$ and $L_w^{(k)}$ a single channel pre-processed LFP analytic signal (as explained in section *LFP pre-processing*) and its phase $\phi (= \angle L)$, the conventional measure of spike-LFP coupling, Phase Locking Value (PLV), defined as follows:

$$\text{PLV} = \frac{1}{N_{\text{tot}}} \sum_{k=1}^K \sum_{j=1}^{N^{(k)}} \exp\left(\mathbf{i}\phi_{t_j^{(k)}}^{(k)}\right), \quad (13)$$

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.

$$N_{\text{tot}} = \sum_{k=1}^K N^{(k)}. \quad (14)$$

In addition to PLV, we introduce a similar coupling statistics, denoted by c ,

$$c = \frac{1}{\sqrt{N_{\text{tot}}}} \sum_{k=1}^K \sum_{j=1}^{N^{(k)}} I_{t_j^{(k)}}, \quad (15)$$

to be used when the theoretical significance test is intended to be used (see section *Analytical test*) (for summary on type of normalization used in different figures see [Table 1](#)). 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 [126, 127] 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 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 $\sqrt{N_{\text{tot}}}$ in [Eq 15](#) versus N_{tot} in [Eq 13](#)). Basically, a scaling by $\sqrt{N_{\text{tot}}}$ is needed to normalize the variance of entries of the coupling matrix to 1, in order to be able to use tools from Random Matrix Theory [62] (see [59] 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 $\{t_j^{m,(k)}\}_{j=1 \dots N_m^{(k)}}$, $\phi_w^{(k)}$ LFP phase, and $L_w^{(k)}$ the analytical LFP, we can define the coupling matrix \mathbf{C} , based on PLV ([Eq 13](#), also similar to [82]) as follows,

$$(\mathbf{C})_{n,m} = \frac{1}{N_m^{(k)}} \sum_{k=1}^K \sum_{j=1}^{N_m^{(k)}} \exp\left(\mathbf{i}(\phi^{(k)})_{n,t_j^{(k)}}\right), \quad (16)$$

or based on c ([Eq 15](#)),

$$(\mathbf{C})_{n,m} = \frac{1}{\sqrt{N_m^{(k)}}} \sum_{k=1}^K \sum_{j=1}^{N_m^{(k)}} (L^{(k)})_{n,t_j^{(k)}}, \quad (17)$$

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 . Readers can also refer to [59, Section 4] for a different formulation.

Let n_c and n_s be the number of LFP channels and number of spiking units, respectively, C is thus a $n_c \times n_s$ complex-valued matrix (or $n_c^{\text{eff}} \times n_s$ 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 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

$$C = UDV^H = \sum_{k=1}^p d_k u_k v_k^H, \quad (18)$$

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 Eq (18), 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 [128, 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 Eq 3, U and V are unitary matrices i. e. $U^H U = I$ and $V^H V = I$. 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 *Reduction of complex models based on linear response theory* and *Generalizing SFC 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:

$$C \sim (gPLV) \cdot v_{LFP} v_{spike}^H. \quad (19)$$

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.

$$udv^H = udv^H e^{-i\theta_0} e^{i\theta_0} = e^{-i\theta_0} ud(e^{-i\theta_0} v)^H. \quad (20)$$

We exploit this free parameter to make the GPLA more neuroscientifically interpretable by shifting the phase of 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,

$$\overline{\phi_{LFP}} = \angle \sum_{i=1}^{n_c} (v_{LFP})_i, \quad (21)$$

$$\overline{\phi_{spike}} = \angle \sum_{i=1}^{n_u} (v_{spike})_i. \quad (22)$$

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,

$$\widetilde{v_{spike}} = v_{spike} e^{-i\overline{\phi_{LFP}}} \quad (23)$$

preserves the angular difference of Φ_d of the spikes with respect to the LFP, defined as

$$\Phi_d = \overline{\phi_{LFP}} - \overline{\phi_{spike}}. \quad (24)$$

With this chosen convention, we obtain the final GPLA factorization

$$C \sim (gPLV) \cdot \widetilde{v_{LFP}} \widetilde{v_{spike}}^H. \quad (25)$$

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* ($\widetilde{gPLV} = gPLV e^{-i\Phi_d}$) whose magnitude indicates 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, nevertheless it allows to interpret the GPLA output similarly 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. unwhiten the LFP vector resulting from GPLA. This can be achieved by computing the unwhitening operator W^{-1} and apply it to the LFP vector,

$$v_{LFP}^{unwhiten} = W^{-1} v_{LFP}. \quad (26)$$

In order to find this operator, we first concatenate whitened LFPs of all trials (resulting from Eq 12) into a larger matrix L_w ($n_c^{eff} \times KT$). Then we estimate W^{-1} by using a linear regression with unwhitened and whitened LFPs (W^{-1} is the $n_c \times n_c^{eff}$ matrix of coefficient for regression).

Rescaling: As introduced in Eq 15, the coefficients of the coupling matrix are normalized by the square root of the number of spikes. This choice of normalization is different from the one used in conventional PLV (Eq 13). This will lead to inhomogeneous weighting of spiking units according to their variability of their firing rate. We “revert” this weighting later on by dividing

the spike vector by the square root of number of spikes,

$$\nu_{\text{Spike}}^{\text{rescaled}} = \nu_{\text{Spike}} \oslash \vec{N}, \quad (27)$$

where \oslash is the (entrywise) Hadamard division and $\vec{N} = \{N_m^{\text{tot}}\}_{m=1,\dots,n_s}$, which is a vector consisting of total spike counts (similar to Eq 14) of all the neurons (indexed 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,

$$\nu_{\text{Spike}}^{\text{final}} = \frac{\nu_{\text{Spike}}^{\text{rescaled}}}{\|\nu_{\text{Spike}}^{\text{rescaled}}\|}. \quad (28)$$

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, leading to strong statistical dependencies between the coefficients of the estimated coupling matrix C , which affects the statistics of the singular values (and consequently gPLV). Whitening removes correlations before computing spike-LFP coupling. However, if statistical testing based on surrogate data is intended, it is possible to skip the whitening step and proceed directly with constructing the coupling matrix and low rank estimation (see Fig 3). In that case, entries of the coupling can be filled by conventional PLVs (see Eq 16), or other choices of spike-LFP coupling measures [34–38, 111, 112] (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 entries of coupling matrix, C , to be PLV (LFPs are not whitened and Eq 16 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 the 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 the coupling between spikes and LFP based on gPLV, we develop a surrogate- and a Random Matrix Theory (RMT)-based statistical testing framework exposed in [59]. 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 [45]), but also computationally expensive. This motivates the development of an “analytical” test exploiting minimal computational resources.

Surrogate-based test. In contrast to univariate 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 [129, 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 [45]. 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 appropriate surrogates 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 [45].

In this work, we used an *interval*-jittering rather than a *spike-centered*-jittering (interval- and spike-centered-jittering are also known as hard and soft dithering, respectively), as the former was reported to be more reliable for detecting temporal structures in spike data [130]. We devised the two following spike-jittering-based methods for GPLA. We also verified the appropriateness of our jittering approach 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 [58]. 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 ensemble 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 [45] and considerable increase in the dimensionality of datasets [24, 40, 42, 43], suggest that deriving mathematically (asymptotic) properties of GPLA under the null hypotheses, as is done for univariate testing (e. g. Rayleigh test for PLV [129, Chapter 4]) is an interesting alternative.

In a companion work [59], by using martingale theory [131] we derive an asymptotic distribution for the entries of the coupling matrix in fairly general settings. Furthermore, by exploiting RMT [62] 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.

As mathematical details are described in [59, Theorem 2], 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 15, asymptotically converges to a matrix with i.i.d. complex standard normal coefficients [59 Theorem 3], and the Marchenko-Pastur (MP) law then provides an approximation of the distribution of its squared singular values [59, Theorem 3].

This law [64] has density

$$\frac{d\mu_{MP}}{dx}(x) = \begin{cases} \frac{1}{2\pi\alpha x} \sqrt{(b-x)(x-a)}, & a \leq x \leq b, \\ 0, & \text{otherwise,} \end{cases} \quad (29)$$

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 the these bounds we can define a significance threshold, θ_{DET} , for the largest eigenvalue of hermitian matrix, $\mathbf{S} = \frac{K}{n_u} \mathbf{C} \mathbf{C}^H$:

$$\theta_{DET} = (1 + \sqrt{\alpha})^2. \quad (30)$$

The null hypothesis can be rejected if, the largest eigenvalue of \mathbf{S} (denoted by ℓ_1) is superior to the significance threshold:

$$\ell_1(\mathbf{S}_n) > \theta_{DET}. \quad (31)$$

Therefore, there is a significant coupling between the multi-channel spikes and LFPs, if

$$gPLV > \sqrt{n_u \theta_{DET}}. \quad (32)$$

As mentioned above, to be able to use the result of [59], 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 Eq 17).

For computing α on neural data, the *reduced ranked* $n_c^{eff} < n_c$ entailed by the whitening procedure (see LFP pre-processing for 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 Eq 30 as follows:

$$\theta_{DET} = (1 + \sqrt{\alpha_{eff}})^2, \quad (33)$$

where $\alpha_{eff} = n_c^{eff}/n_u$.

Supporting information

S1 Appendix. Contains method details and analytical developments.
(PDF)

S1 Fig. Use of EPSP as LFP proxy. Difference between phase of excitatory and inhibitory neurons/populations based on GPLA and the excitatory and inhibitory populations in the MassAlpha neural mass model. In this simulation EPSP has been used for the LFP proxy.
(PDF)

S2 Fig. GPLA of CA3 and CA1 activities. For this analysis, CA1 and CA3 data were separately injected into GPLA. (A) Spike vectors represented in polar plots similar to Fig 6E, but for all frequencies (indicated on the left). (B) LFP vectors, similar to Fig 6D, but for all frequencies (indicated in legend in the bottom). (C) gPLV for different frequency ranges of LFPs, similar to Fig 6C. Triangles indicated the significance assessed based on empirical (blue triangles, with significance threshold of 0.05) and theoretical (red triangles) tests. (left) for CA3 and (right) for CA1.
(PDF)

S3 Fig. Joint CA1-CA3 analysis of hippocampal SWRs. For this analysis, CA1 and CA3 data were injected to GPLA together. (A) Spike vectors represented in polar plots similar to Fig 6E, but for all frequencies (indicated on the left). (B) LFP vectors, similar to Fig 6D, but for all frequencies (indicated in legend in the bottom). (C) gPLV for different frequency ranges of LFPs Fig 6C. Triangles indicated the significance assessed based on empirical (blue triangles, with significance threshold of 0.05) and theoretical (red triangles) tests.
(PDF)

S4 Fig. GPLA vs. PLA comparison for hippocampal SWR simulation. Similar to Fig 6D but based on uni-variate phase locking analysis (rather than multivariate GPLA). Each line depicts the phase locking value (PLV) for a fixed spiking units across all LFP channels. Colors indicate the frequency of filtered LFP.
(PDF)

S5 Fig. Phase-modulus relation dependency on level of inhibition. Related to Fig 7D. (A) Same as Fig 7C. for simulations at intermediate levels of recurrent inhibition. (B) Same as Fig 7D. for simulations at intermediate levels of recurrent inhibition. (C) Same as Fig 7E. for simulations at intermediate levels of recurrent inhibition. (D) Magnitude of phase modulus regression coefficient (rescaled by $180/\pi$ to have it in radians) as a function of imaginary part of a derived from Eq 9. (E) Same as (A) for $\text{Im}[a]/\text{Re}[a]$ instead of $\text{Im}[a]$.
(PDF)

S6 Fig. GPLA using IPSP as LFP proxy. To be compared with Fig 7C–7E. (A) gPLV as a function of frequency for both models. (B) Phase of spike vector coefficients as a function of its modulus for the frequency band associated with maximum gPLV for each model (each dot one coefficient, and the continuous lines are plotted based on linear regression). (C) Shift between the averaged phase of spike vector and averaged phase of LFP vector, as a function of frequency.
(PDF)

S7 Fig. Analysis of PFC Utah array data. LFP and spike vectors for frequencies indicated on the right. First column depict the LFP (blue dots) and spike (red dots) in the complex plane. Second column depict fitted von Mises distribution to phase of LFP and spike vectors. Third and forth column respectively representing phase of LFP and spike vectors which remapped to real configuration of electrodes on Utah array (see Fig 8C).
(PDF)

S8 Fig. GPLA vs. PLA comparison for PFC Utah array data for revealing the spatial pattern of coupling. Similar to Utah array maps in Fig 8I but based on uni-variate phase locking analysis (rather than multivariate GPLA). Panels in the first row depict the spatial distribution of phase locking value (PLV) or magnitude of the spike-field coupling on the array (see Fig 8C). Panels in the second row depict the spatial distribution of locking phase on the array. White pixels in all panels indicate the recording channels with insufficient number of spikes

(multiunit activity with a minimum of 5 Hz firing), as it was used in Fig 8I. The colorbars indicate the coupling strength in the first row; and locking phase in the second row. First column, depicts the results based on multivariate GPLA, and second and third column depicts the results based on uni-variate phase locking analysis, but for two different choices of LFP reference channel. The result from ‘Example 2’ is close to what is captured based on GPLA, however result from ‘Example 1’ does not, due to a lack of global coupling.

(PDF)

S9 Fig. GPLA vs. PLA comparison for PFC Utah array data for characterizing the strength of recurrent inhibition in PFC circuits. Similar to Fig 8H but based on uni-variate phase locking analysis (rather than multivariate GPLA). Each row corresponds to analysis in different frequency (the same frequencies used in Fig 8H), i. e., 3–5 Hz, 5–15 Hz, and 15–30 Hz, respectively, first, second and third row. First column indicates the results based on GPLA (notably pairwise coupling measure used here is exactly PLV), and imply the negative slope, similar to Fig 8H. The second and third columns demonstrate a similar analysis based on phase locking analysis, i. e., locking phase plotted versus strength of coupling (PLV) with two example of LFP reference channels (the same used in S8 Fig). Notably, none are compatible with our mean-field analysis (Fig 7). PLA is thus not conclusive about the strength of recurrent inhibition.

(PDF)

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.

Author Contributions

Conceptualization: Shervin Safavi, Theofanis I. Panagiotaropoulos, Michel Besserve.

Data curation: Shervin Safavi, Theofanis I. Panagiotaropoulos, Vishal Kapoor, Michel Besserve.

Formal analysis: Shervin Safavi, Michel Besserve.

Funding acquisition: Nikos K. Logothetis.

Investigation: Shervin Safavi, Theofanis I. Panagiotaropoulos, Vishal Kapoor, Michel Besserve.

Methodology: Shervin Safavi, Juan F. Ramirez-Villegas, Michel Besserve.

Project administration: Theofanis I. Panagiotaropoulos, Michel Besserve.

Resources: Nikos K. Logothetis.

Software: Shervin Safavi, Michel Besserve.

Supervision: Theofanis I. Panagiotaropoulos, Michel Besserve.

Visualization: Shervin Safavi, Michel Besserve.

Writing – original draft: Shervin Safavi, Michel Besserve.

Writing – review & editing: Shervin Safavi, Theofanis I. Panagiotaropoulos, Vishal Kapoor, Juan F. Ramirez-Villegas, Nikos K. Logothetis, Michel Besserve.

References

1. Einevoll GT, Destexhe A, Diesmann M, Grün S, Jirsa V, de Kamps M, et al. The Scientific Case for Brain Simulations. *Neuron*. 2019; 102(4):735–744. <https://doi.org/10.1016/j.neuron.2019.03.027> PMID: 31121126
2. D'Angelo E, Jirsa V. The quest for multiscale brain modeling. *Trends in Neurosciences*. 2022;
3. Safavi S. Brain as a Complex System, harnessing systems neuroscience tools & notions for an empirical approach. PhD Thesis, Universität Tübingen; 2022. Available from: <https://tobias-lib.ub.uni-tuebingen.de/xmlui/handle/10900/128071>
4. Rasch MJ, Gretton A, Murayama Y, Maass W, Logothetis NK. Inferring Spike Trains from Local Field Potentials. *J Neurophysiol*. 2008; 99(3):1461–76. <https://doi.org/10.1152/jn.00919.2007> PMID: 18160425
5. Rasch M, Logothetis NK, Kreiman G. From Neurons to Circuits: Linear Estimation of Local Field Potentials. *J Neurosci*. 2009; 29:13785–96. <https://doi.org/10.1523/JNEUROSCI.2390-09.2009> PMID: 19889990
6. Li CY, Poo MM, Dan Y. Burst Spiking of a Single Cortical Neuron Modifies Global Brain State. *Science*. 2009; 324(5927):643–6. <https://doi.org/10.1126/science.1169957> PMID: 19407203
7. Schwalm M, Schmid F, Wachsmuth L, Backhaus H, Kronfeld A, Aedo Jury F, et al. Cortex-Wide BOLD fMRI Activity Reflects Locally-Recorded Slow Oscillation-Associated Calcium Waves. In: *eLife*. vol. 6; 2017. <https://doi.org/10.7554/eLife.27602> PMID: 28914607
8. Zerbi V, Floriou-Servou A, Markicevic M, Vermeiren Y, Sturman O, Privitera M, et al. Rapid Reconfiguration of the Functional Connectome after Chemogenetic Locus Coeruleus Activation. *Neuron*. 2019; 103(4):702–718.e5. <https://doi.org/10.1016/j.neuron.2019.05.034> PMID: 31227310
9. Sejnowski TJ, Paulsen O. Network oscillations: emerging computational principles. *Journal of Neuroscience*. 2006; 26(6):1673–1676. <https://doi.org/10.1523/JNEUROSCI.3737-05d.2006> PMID: 16467514
10. Chalk M, Gutkin B, Deneve S. Neural oscillations as a signature of efficient coding in the presence of synaptic delays. *Elife*. 2016; 5:e13824. <https://doi.org/10.7554/Elife.13824> PMID: 27383272
11. Muller L, Chavane F, Reynolds J, Sejnowski TJ. Cortical travelling waves: mechanisms and computational principles. *Nature Reviews Neuroscience*. 2018; 19(5):255–268. <https://doi.org/10.1038/nrn.2018.20> PMID: 29563572
12. Peterson EJ, Voytek B. Healthy Oscillatory Coordination Is Bounded by Single-Unit Computation. *bioRxiv [Preprint]*. 2018; p. 309427. <https://doi.org/10.1101/309427>
13. Safavi S, Chalk M, Logothetis N, Levina A. Signatures of criticality in efficient coding networks. *bioRxiv [Preprint]*. 2023; p. 2023–02. <https://doi.org/10.1101/2023.02.14.528465>
14. Engel AK, Fries P, Konig P, Brecht M, Singer W. Temporal binding, binocular rivalry, and consciousness. *Conscious Cogn*. 1999; 8(2):128–51. <https://doi.org/10.1006/ccog.1999.0389> PMID: 10447995
15. Dwarakanath A, Kapoor V, Werner J, Safavi S, Fedorov LA, Logothetis NK, et al. Bistability of prefrontal states gates access to consciousness. *Neuron*. 2023; <https://doi.org/10.1016/j.neuron.2023.02.027> PMID: 36921603
16. Niebur E, Koch C, Rosin C. An oscillation-based model for the neuronal basis of attention. *Vision Res*. 1993; 33(18):2789–802. [https://doi.org/10.1016/0042-6989\(93\)90236-P](https://doi.org/10.1016/0042-6989(93)90236-P) PMID: 8296473
17. Buzsaki G. *Rhythms of the Brain*. Oxford University Press; 2006.
18. Liljenstroem H. Mesoscopic brain dynamics. *Scholarpedia*. 2012; 7(9):4601. <https://doi.org/10.4249/scholarpedia.4601>
19. Buzsaki G, Anastassiou CA, Koch C. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat Rev Neurosci*. 2012; 13(6):407–20. <https://doi.org/10.1038/nrn3241> PMID: 22595786
20. Einevoll GT, Kayser C, Logothetis NK, Panzeri S. Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat Rev Neurosci*. 2013; 14(11):770–85. <https://doi.org/10.1038/nrn3599> PMID: 24135696
21. Hagen E, Dahmen D, Stavrinos ML, Lindén H, Tetzlaff T, Albada V, et al. Hybrid Scheme for Modeling Local Field Potentials from Point-Neuron Networks. *Cereb Cortex*. 2016; 26(12):4461–4496. <https://doi.org/10.1093/cercor/bhw237> PMID: 27797828
22. Herreras O. Local Field Potentials: Myths and Misunderstandings. *Front Neural Circuits*. 2016; 10:101. <https://doi.org/10.3389/fncir.2016.00101> PMID: 28018180
23. Buzsaki G, Logothetis N, Singer W. Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. *Neuron*. 2013; 80(3):751–64. <https://doi.org/10.1016/j.neuron.2013.10.002> PMID: 24183025

24. Pesaran B, Vinck M, Einevoll GT, Sirota A, Fries P, Siegel M, et al. Investigating Large-Scale Brain Dynamics Using Field Potential Recordings: Analysis and Interpretation. *Nature Neuroscience*. 2018; p. 1. <https://doi.org/10.1038/s41593-018-0171-8> PMID: 29942039
25. Fries P. A Mechanism for Cognitive Dynamics: Neuronal Communication through Neuronal Coherence. *Trends Cogn Sci*. 2005; 9(10):474–80. <https://doi.org/10.1016/j.tics.2005.08.011> PMID: 16150631
26. Fries P. Rhythms for Cognition: Communication through Coherence. *Neuron*. 2015; 88:220–35. <https://doi.org/10.1016/j.neuron.2015.09.034> PMID: 26447583
27. Womelsdorf T, Schöffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, et al. Modulation of Neuronal Interactions through Neuronal Synchronization. *Science*. 2007; 316:1609–12. <https://doi.org/10.1126/science.1139597> PMID: 17569862
28. Ito J, Maldonado P, Singer W, Grün S. Saccade-Related Modulations of Neuronal Excitability Support Synchrony of Visually Elicited Spikes. *Cereb Cortex*. 2011; 21(11):2482–2497. <https://doi.org/10.1093/cercor/bhr020> PMID: 21459839
29. Tiesinga PH, Sejnowski TJ. Mechanisms for phase shifting in cortical networks and their role in communication through coherence. *Frontiers in human neuroscience*. 2010; 4:196. <https://doi.org/10.3389/fnhum.2010.00196> PMID: 21103013
30. Besserve M, Lowe SC, Logothetis NK, Schölkopf B, Panzeri S. Shifts of gamma phase across primary visual cortical sites reflect dynamic stimulus-modulated information transfer. *PLoS biology*. 2015; 13(9):e1002257. <https://doi.org/10.1371/journal.pbio.1002257> PMID: 26394205
31. Igarashi KM. Plasticity in oscillatory coupling between hippocampus and cortex. *Current opinion in neurobiology*. 2015; 35:163–168. <https://doi.org/10.1016/j.conb.2015.09.005> PMID: 26425996
32. Zeitler M, Fries P, Gielen S. Assessing Neuronal Coherence with Single-Unit, Multi-Unit, and Local Field Potentials. *Neural Comput*. 2006; 18:2256–81. <https://doi.org/10.1162/neco.2006.18.9.2256> PMID: 16846392
33. Ashida G, Wagner H, Carr CE. Processing of Phase-Locked Spikes and Periodic Signals. In: *Analysis of Parallel Spike Trains*. Springer Series in Computational Neuroscience. Springer, Boston, MA; 2010. p. 59–74.
34. Vinck M, van Wingerden M, Womelsdorf T, Fries P, Pennartz CM. The Pairwise Phase Consistency: A Bias-Free Measure of Rhythmic Neuronal Synchronization. *Neuroimage*. 2010; 51(1):112–22. <https://doi.org/10.1016/j.neuroimage.2010.01.073> PMID: 20114076
35. Vinck M, Battaglia FP, Womelsdorf T, Pennartz C. Improved Measures of Phase-Coupling between Spikes and the Local Field Potential. *J Comput Neurosci*. 2012; 33(1):53–75. <https://doi.org/10.1007/s10827-011-0374-4> PMID: 22187161
36. Jiang H, Bahramisharif A, van Gerven MAJ, Jensen O. Measuring Directionality between Neuronal Oscillations of Different Frequencies. *NeuroImage*. 2015; 118:359–367. <https://doi.org/10.1016/j.neuroimage.2015.05.044> PMID: 26025291
37. Li Z, Cui D, Li X. Unbiased and Robust Quantification of Synchronization between Spikes and Local Field Potential. *J Neurosci Methods*. 2016; 269:33–8. <https://doi.org/10.1016/j.jneumeth.2016.05.004> PMID: 27180930
38. Zarei M, Jahed M, Daliri MR. Introducing a Comprehensive framework to measure Spike-LFP Coupling. *Frontiers in Computational Neuroscience*. 2018; 12. <https://doi.org/10.3389/fncom.2018.00078> PMID: 30374297
39. Dickey AS, Suminski A, Amit Y, Hatsopoulos NG. Single-Unit Stability Using Chronically Implanted Multielectrode Arrays. *Journal of Neurophysiology*. 2009; 102(2):1331–1339. <https://doi.org/10.1152/jn.90920.2008> PMID: 19535480
40. Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, et al. Fully Integrated Silicon Probes for High-Density Recording of Neural Activity. *Nature*. 2017; 551:232–236. <https://doi.org/10.1038/nature24636> PMID: 29120427
41. Juavinett AL, Bekheet G, Churchland AK. Chronically Implanted Neuropixels Probes Enable High-Yield Recordings in Freely Moving Mice. *eLife*. 2019; 8:e47188. <https://doi.org/10.7554/eLife.47188> PMID: 31411559
42. Buzsáki G. Large-Scale Recording of Neuronal Ensembles. *Nature Neuroscience*. 2004; 7(5):446–451. <https://doi.org/10.1038/nn1233> PMID: 15114356
43. Fukushima M, Chao ZC, Fujii N. Studying Brain Functions with Mesoscopic Measurements: Advances in Electrocorticography for Non-Human Primates. *Current Opinion in Neurobiology*. 2015; 32:124–131. <https://doi.org/10.1016/j.conb.2015.03.015> PMID: 25889531

44. Ermentrout GB, Kleinfeld D. Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron*. 2001; 29(1):33–44. [https://doi.org/10.1016/S0896-6273\(01\)00178-7](https://doi.org/10.1016/S0896-6273(01)00178-7) PMID: 11182079
45. Grün S. Data-Driven Significance Estimation for Precise Spike Correlation. *Journal of Neurophysiology*. 2009; 101(3):1126–1140. <https://doi.org/10.1152/jn.00093.2008> PMID: 19129298
46. Salmon WC. *Scientific explanation and the causal structure of the world*. Princeton University Press; 1984.
47. Craver C. Constitutive explanatory relevance. *Journal of Philosophical Research*. 2007; 32:3–20. <https://doi.org/10.5840/jpr20073241>
48. Stevenson IH, Kording KP. How Advances in Neural Recording Affect Data Analysis. *Nature Neuroscience*. 2011; 14(2):139–142. <https://doi.org/10.1038/nn.2731> PMID: 21270781
49. O’Leary T, Sutton AC, Marder E. Computational Models in the Age of Large Datasets. *Current Opinion in Neurobiology*. 2015; 32:87–94. <https://doi.org/10.1016/j.conb.2015.01.006> PMID: 25637959
50. Gao P, Ganguli S. On Simplicity and Complexity in the Brave New World of Large-Scale Neuroscience. *Current Opinion in Neurobiology*. 2015; 32:148–155. <https://doi.org/10.1016/j.conb.2015.04.003> PMID: 25932978
51. Williamson RC, Doiron B, Smith MA, Yu BM. Bridging Large-Scale Neuronal Recordings and Large-Scale Network Models Using Dimensionality Reduction. *Current Opinion in Neurobiology*. 2019; 55:40–47. <https://doi.org/10.1016/j.conb.2018.12.009> PMID: 30677702
52. van der Meij R, Kahana M, Maris E. Phase-Amplitude Coupling in Human Electrocorticography Is Spatially Distributed and Phase Diverse. *J Neurosci*. 2012; 32(1):111–123. <https://doi.org/10.1523/JNEUROSCI.4816-11.2012> PMID: 22219274
53. van der Meij R, Jacobs J, Maris E. Uncovering Phase-Coupled Oscillatory Networks in Electrophysiological Data. *Hum Brain Mapp*. 2015; 36(7):2655–2680. <https://doi.org/10.1002/hbm.22798> PMID: 25864927
54. van der Meij R, Voytek B. Uncovering Neuronal Networks Defined by Consistent Between-Neuron Spike Timing from Neuronal Spike Recordings. *eNeuro*. 2018; 5(3):ENEURO.0379–17.2018. <https://doi.org/10.1523/ENEURO.0379-17.2018> PMID: 29789811
55. Jirsa VK, Haken H. Field theory of electromagnetic brain activity. *Physical Review Letters*. 1996; 77(5):960–963. <https://doi.org/10.1103/PhysRevLett.77.960> PMID: 10062950
56. Karasözen B. Model order reduction in neuroscience. In: *Model order reductionApplications*. Volume 3: Applications. De Gruyter; 2020. p. 237–250.
57. Transtrum MK, Qiu P. Bridging mechanistic and phenomenological models of complex biological systems. *PLoS computational biology*. 2016; 12(5):e1004915. <https://doi.org/10.1371/journal.pcbi.1004915> PMID: 27187545
58. Chavez M, Besserve M, Adam C, Martinerie J. Towards a proper estimation of phase synchronization from time series. *J Neurosci Methods*. 2006; 154(1-2):149–60. <https://doi.org/10.1016/j.jneumeth.2005.12.009> PMID: 16445988
59. Safavi S, Logothetis NK, Besserve M. From Univariate to Multivariate Coupling between Continuous Signals and Point Processes: A Mathematical Framework. *Neural Computation*. 2021; p. 1–67. PMID: 34411270
60. Mukamel R, Fried I. Human Intracranial Recordings and Cognitive Neuroscience. *Annual Review of Psychology*. 2011; 63(1):511–537. <https://doi.org/10.1146/annurev-psych-120709-145401> PMID: 21943170
61. Fried I, Rutishauser U, Cerf M, Kreiman G, editors. *Single Neuron Studies of the Human Brain: Probing Cognition*. The MIT Press; 2014.
62. Anderson GW, Guionnet A, Zeitouni O. *An Introduction to Random Matrices*. Cambridge University Press; 2010.
63. Capitaine M, Donati-Martin C. Spectrum of deformed random matrices and free probability. *arXiv [preprint]* arXiv:160705560. 2016;.
64. Marchenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*. 1967; 114(4):507–536.
65. Nunez PL, Srinivasan R. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press; 2006.
66. Wilson HR, Cowan JD. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*. 1973; 13(2):55–80. <https://doi.org/10.1007/BF00288786> PMID: 4767470

67. Sik A, Penttonen M, Ylinen A, Buzsaki G. Hippocampal CA1 Interneurons: An in Vivo Intracellular Labeling Study. *Journal of Neuroscience*. 1995; 15(10):6651–6665. <https://doi.org/10.1523/JNEUROSCI.15-10-06651.1995> PMID: 7472426
68. Schomburg EW, Anastassiou CA, Buzsáki G, Koch C. The Spiking Component of Oscillatory Extracellular Potentials in the Rat Hippocampus. *Journal of Neuroscience*. 2012; 32(34):11798–11811. <https://doi.org/10.1523/JNEUROSCI.0656-12.2012> PMID: 22915121
69. Taxidis J, Coombes S, Mason R, Owen MR. Modeling Sharp Wave-Ripple Complexes through a CA3-CA1 Network Model with Chemical Synapses. *Hippocampus*. 2012; 22(5):995–1017. <https://doi.org/10.1002/hipo.20930> PMID: 21452258
70. Nö RLD. Action Potential of the Motoneurons of the Hypoglossus Nucleus. *Journal of Cellular and Comparative Physiology*. 1947; 29(3):207–287. <https://doi.org/10.1002/jcp.1030290303>
71. Lindén H, Tetzlaff T, Potjans TC, Pettersen KH, Grün S, Diesmann M, et al. Modeling the Spatial Reach of the LFP. *Neuron*. 2011; 72(5):859–872. <https://doi.org/10.1016/j.neuron.2011.11.006> PMID: 22153380
72. Mazzoni A, Linden H, Cuntz H, Lansner A, Panzeri S, Einevoll GT. Computing the Local Field Potential (LFP) from Integrate-and-Fire Network Models. *PLoS Comput Biol*. 2015; 11(12):e1004584. <https://doi.org/10.1371/journal.pcbi.1004584> PMID: 26657024
73. Moran RJ, Kiebel SJ, Stephan KE, Reilly R, Daunizeau J, Friston KJ. A neural mass model of spectral responses in electrophysiology. *NeuroImage*. 2007; 37(3):706–720. <https://doi.org/10.1016/j.neuroimage.2007.05.032> PMID: 17632015
74. Ledoux E, Brunel N. Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. *Frontiers in Computational Neuroscience*. 2011; 5:25. <https://doi.org/10.3389/fncom.2011.00025> PMID: 21647353
75. Pinotsis DA, Moran RJ, Friston KJ. Dynamic causal modeling with neural fields. *NeuroImage*. 2012; 59(2):1261–1274. <https://doi.org/10.1016/j.neuroimage.2011.08.020> PMID: 21924363
76. Buzsáki G, Horváth Z, Urioste R, Hetke J, Wise K. High-frequency network oscillation in the hippocampus. *Science*. 1992; 256(5059):1025–1027. <https://doi.org/10.1126/science.1589772> PMID: 1589772
77. Ramirez-Villegas JF, Logothetis NK, Besserve M. Diversity of Sharp-Wave-Ripple LFP Signatures Reveals Differentiated Brain-Wide Dynamical Events. *Proc Natl Acad Sci U S A*. 2015; 112:E6379–E87. <https://doi.org/10.1073/pnas.1518257112> PMID: 26540729
78. Ramirez-Villegas JF, Willeke KF, Logothetis NK, Besserve M. Dissecting the Synapse- and Frequency-Dependent Network Mechanisms of In Vivo Hippocampal Sharp Wave-Ripples. *Neuron*. 2018; 100(5):1224–1240.e13. <https://doi.org/10.1016/j.neuron.2018.09.041> PMID: 30482688
79. Csicsvari J, Hirase H, Mamiya A, Buzsáki G. Ensemble Patterns of Hippocampal CA3-CA1 Neurons during Sharp Wave–Associated Population Events. *Neuron*. 2000; 28(2):585–594. [https://doi.org/10.1016/S0896-6273\(00\)00135-5](https://doi.org/10.1016/S0896-6273(00)00135-5) PMID: 11144366
80. Oliva A, Fernandez-Ruiz A, Buzsaki G, Berenyi A. Role of Hippocampal CA2 Region in Triggering Sharp-Wave Ripples. *Neuron*. 2016; 91:1342–55. <https://doi.org/10.1016/j.neuron.2016.08.008> PMID: 27593179
81. Sullivan D, Csicsvari J, Mizuseki K, Montgomery S, Diba K, Buzsáki G. Relationships between hippocampal sharp waves, ripples, and fast gamma oscillation: influence of dentate and entorhinal cortical activity. *Journal of Neuroscience*. 2011; 31(23):8605–8616. <https://doi.org/10.1523/JNEUROSCI.0294-11.2011> PMID: 21653864
82. Davis ZW, Dotson NM, Franken T, Muller L, Reynolds J. Spike-phase coupling patterns reveal laminar identity in primate cortex. *bioRxiv [Preprint]*. 2022; <https://doi.org/10.1101/2022.10.26.513932>
83. Wójcik DK. Current Source Density (CSD) Analysis. In: Jaeger D, Jung R, editors. *Encyclopedia of Computational Neuroscience*. New York, NY: Springer; 2013. p. 1–10.
84. Hindriks R, Arsiwalla XD, Panagiotaropoulos T, Besserve M, Verschure PF, Logothetis NK, et al. Discrepancies between Multi-Electrode LFP and CSD Phase-Patterns: A Forward Modeling Study. *Front Neural Circuits*. 2016; 10:51. <https://doi.org/10.3389/fncir.2016.00051> PMID: 27471451
85. Varga C, Oijala M, Lish J, Szabo GG, Bezaire M, Marchionni I, et al. Functional Fission of Parvalbumin Interneuron Classes during Fast Network Events. *eLife*. 2014; 3. <https://doi.org/10.7554/eLife.04006> PMID: 25375253
86. Whittington MA, Traub RD, Kopell N, Ermentrout B, Buhl EH. Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *International journal of psychophysiology*. 2000; 38(3):315–336. [https://doi.org/10.1016/S0167-8760\(00\)00173-2](https://doi.org/10.1016/S0167-8760(00)00173-2) PMID: 11102670
87. Traub RD, Whittington MA, Buhl EH, Jefferys JG, Faulkner HJ. On the mechanism of the gamma --> beta frequency shift in neuronal oscillations induced in rat hippocampal slices by tetanic stimulation.

- The Journal of Neuroscience. 1999; 19(3):1088–1105. <https://doi.org/10.1523/JNEUROSCI.19-03-01088.1999> PMID: 9920671
88. Börgers C, Kopell N. Synchronization in networks of excitatory and inhibitory neurons with sparse, random connectivity. *Neural computation*. 2003; 15(3):509–538. <https://doi.org/10.1162/089976603321192059> PMID: 12620157
89. Stark E, Roux L, Eichler R, Senzai Y, Royer S, Buzsáki G. Pyramidal cell-interneuron interactions underlie hippocampal ripple oscillations. *Neuron*. 2014; 83(2):467–480. <https://doi.org/10.1016/j.neuron.2014.06.023> PMID: 25033186
90. Brunel N, Wang XJ. What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *Journal of Neurophysiology*. 2003; 90(1):415–430. <https://doi.org/10.1152/jn.01095.2002> PMID: 12611969
91. Teleńczuk B, Dehghani N, Le Van Quyen M, Cash SS, Halgren E, Hatsopoulos NG, et al. Local Field Potentials Primarily Reflect Inhibitory Neuron Activity in Human and Monkey Cortex. *Scientific Reports*. 2017; 7:40211. <https://doi.org/10.1038/srep40211> PMID: 28074856
92. Carr MF, Karlsson MP, Frank LM. Transient Slow Gamma Synchrony Underlies Hippocampal Memory Replay. *Neuron*. 2012; 75(4):700–713. <https://doi.org/10.1016/j.neuron.2012.06.014> PMID: 22920260
93. Zeraati R, Shi YL, Steinmetz NA, Gieselmann MA, Thiele A, Moore T, et al. Intrinsic timescales in the visual cortex change with selective attention and reflect spatial connectivity. *Nat Commun*. 2023; 14(1):1858. <https://doi.org/10.1038/s41467-023-37613-7>
94. Onslow AC, Jones MW, Bogacz R. A canonical circuit for generating phase-amplitude coupling. *PloS one*. 2014; 9(8):e102591. <https://doi.org/10.1371/journal.pone.0102591> PMID: 25136855
95. Logothetis NK, Panzeri S. Local Field Potential, Relationship to BOLD Signal. In: Dieter Jaeger RJ, editor. *Encyclopedia of Computational Neuroscience*. Springer New York; 2014. p. 1–11.
96. Sherfey JS, Ardid S, Hass J, Hasselmo ME, Kopell NJ. Flexible Resonance in Prefrontal Networks with Strong Feedback Inhibition. *PLOS Computational Biology*. 2018; 14(8):e1006357. <https://doi.org/10.1371/journal.pcbi.1006357> PMID: 30091975
97. Zanos TP, Mineault PJ, Pack CC. Removal of Spurious Correlations between Spikes and Local Field Potentials. *J Neurophysiol*. 2011; 105(1):474–86. <https://doi.org/10.1152/jn.00642.2010> PMID: 21068271
98. Ray S, Maunsell JHR. Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biology*. 2011; 9(4):e1000610. <https://doi.org/10.1371/journal.pbio.1000610> PMID: 21532743
99. Elston GN. Pyramidal Cells of the Frontal Lobe: All the More Spinous to Think With. *J Neurosci*. 2000; 20:RC95. <https://doi.org/10.1523/JNEUROSCI.20-18-j0002.2000> PMID: 10974092
100. Elston GN. Cortex, cognition and the cell: new insights into the pyramidal neuron and prefrontal function. *Cerebral Cortex*. 2003; 13(11):1124–1138. <https://doi.org/10.1093/cercor/bhg093> PMID: 14576205
101. Safavi S, Dwarakanath A, Kapoor V, Werner J, Hatsopoulos NG, Logothetis NK, et al. Nonmonotonic Spatial Structure of Interneuronal Correlations in Prefrontal Microcircuits. *Proceedings of the National Academy of Sciences*. 2018; p. 201802356. <https://doi.org/10.1073/pnas.1802356115> PMID: 29588415
102. Sherfey J, Ardid S, Miller EK, Hasselmo ME, Kopell NJ. Prefrontal Oscillations Modulate the Propagation of Neuronal Activity Required for Working Memory. *Neurobiology of Learning and Memory*. 2020; 173:107228. <https://doi.org/10.1016/j.nlm.2020.107228> PMID: 32561459
103. Jin JZ, Weng C, Yeh CI, Gordon JA, Ruthazer ES, Stryker MP, et al. On and off Domains of Geniculocal Afferents in Cat Primary Visual Cortex. *Nature Neuroscience*. 2008; 11(1):88–94. <https://doi.org/10.1038/nn2029> PMID: 18084287
104. Nauhaus I, Busse L, Carandini M, Ringach DL. Stimulus Contrast Modulates Functional Connectivity in Visual Cortex. *Nat Neurosci*. 2009; 12(1):70–6. <https://doi.org/10.1038/nn.2232> PMID: 19029885
105. Ray S, Maunsell JH. Network Rhythms Influence the Relationship between Spike-Triggered Local Field Potential and Functional Connectivity. *J Neurosci*. 2011; 31(35):12674–82. <https://doi.org/10.1523/JNEUROSCI.1856-11.2011> PMID: 21880928
106. Nauhaus I, Busse L, Ringach DL, Carandini M. Robustness of Traveling Waves in Ongoing Activity of Visual Cortex. *J Neurosci*. 2012; 32(9):3088–94. <https://doi.org/10.1523/JNEUROSCI.5827-11.2012> PMID: 22378881

107. Canolty RT, Ganguly K, Kennerley SW, Cadieu CF, Koepsell K, Wallis JD, et al. Oscillatory Phase Coupling Coordinates Anatomically Dispersed Functional Cell Assemblies. *Proc Natl Acad Sci U S A.* 2010; 107:17356–61. <https://doi.org/10.1073/pnas.1008306107> PMID: 20855620
108. Klausberger T, Magill PJ, Marton LF, Roberts JD, Cobden PM, Buzsaki G, et al. Brain-State- and Cell-Type-Specific Firing of Hippocampal Interneurons in Vivo. *Nature.* 2003; 421(6925):844–8. <https://doi.org/10.1038/nature01374> PMID: 12594513
109. Harris KD. Neural Signatures of Cell Assembly Organization. *Nat Rev Neurosci.* 2005; 6(5):399–407. <https://doi.org/10.1038/nrn1669> PMID: 15861182
110. Buzsaki G. Neural Syntax: Cell Assemblies, Synapsembles, and Readers. *Neuron.* 2010; 68(3):362–85. <https://doi.org/10.1016/j.neuron.2010.09.023> PMID: 21040841
111. Grasse DW, Moxon KA. Correcting the Bias of Spike Field Coherence Estimators Due to a Finite Number of Spikes. *J Neurophysiol.* 2010; 104(1):548–58. <https://doi.org/10.1152/jn.00610.2009> PMID: 20484529
112. Lepage KQ, Kramer MA, Eden UT. The Dependence of Spike Field Coherence on Expected Intensity. *Neural Comput.* 2011; 23(9):2209–41. https://doi.org/10.1162/NECO_a_00169 PMID: 21671792
113. Shao K, Ramirez Villegas JF, Logothetis NK, Besserve M. A model of Ponto-Geniculoo-Occipital waves supports bidirectional control of cortical plasticity across sleep-stages. *bioRxiv [Preprint].* 2021. <https://doi.org/10.1101/2021.03.16.432817>
114. Byrne Á, O'Dea RD, Forrester M, Ross J, Coombes S. Next-generation neural mass and field modeling. *Journal of neurophysiology.* 2020; 123(2):726–742. <https://doi.org/10.1152/jn.00406.2019> PMID: 31774370
115. Costa MS, Weigenand A, Ngo HHV, Marshall L, Born J, Martinetz T, et al. A Thalamocortical Neural Mass Model of the EEG during NREM Sleep and Its Response to Auditory Stimulation. *PLOS Computational Biology.* 2016; 12(9):e1005022. <https://doi.org/10.1371/journal.pcbi.1005022>
116. Rosenbaum R, Smith MA, Kohn A, Rubin JE, Doiron B. The Spatial Structure of Correlated Neuronal Variability. *Nat Neurosci.* 2017; 20:107–114. <https://doi.org/10.1038/nn.4433> PMID: 27798630
117. Qubbaj MR, Jirsa VK. Neural Field Dynamics with Heterogeneous Connection Topology. *Phys Rev Lett.* 2007; 98:238102. <https://doi.org/10.1103/PhysRevLett.98.238102> PMID: 17677938
118. Marion M, Temam R. Nonlinear galerkin methods. *SIAM Journal on numerical analysis.* 1989; 26(5):1139–1157. <https://doi.org/10.1137/0726063>
119. Ladenbauer J, McKenzie S, English DF, Hagens O, Ostojic S. Inferring and validating mechanistic models of neural microcircuits based on spike-train data. *Nature communications.* 2019; 10(1):1–17. <https://doi.org/10.1038/s41467-019-12572-0> PMID: 31666513
120. Gonçalves PJ, Lueckmann JM, Deistler M, Nonnenmacher M, Öcal K, Bassetto G, et al. Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife.* 2020; 9:e56261. <https://doi.org/10.7554/elife.56261> PMID: 32940606
121. Liebe S, Hoerzer GM, Logothetis NK, Rainer G. Theta Coupling between V4 and Prefrontal Cortex Predicts Visual Short-Term Memory Performance. *Nature Neuroscience.* 2012; 15(3):456–462, S1–2. <https://doi.org/10.1038/nn.3038> PMID: 22286175
122. Markowitz DA, Curtis CE, Pesaran B. Multiple Component Networks Support Working Memory in Prefrontal Cortex. *Proc Natl Acad Sci U S A.* 2015. <https://doi.org/10.1073/pnas.1504172112> PMID: 26283366
123. Taxidis J, Anastassiou CA, Diba K, Koch C. Local Field Potentials Encode Place Cell Ensemble Activation during Hippocampal Sharp Wave Ripples. *Neuron.* 2015; 87(3):590–604. <https://doi.org/10.1016/j.neuron.2015.07.014> PMID: 26247865
124. Fernández-Ruiz A, Oliva A, Nagy GA, Maurer AP, Berényi A, Buzsáki G. Entorhinal-CA3 Dual-Input Control of Spike Timing in the Hippocampus by Theta-Gamma Coupling. *Neuron.* 2017; 93(5):1213–1226.e5. <https://doi.org/10.1016/j.neuron.2017.02.017> PMID: 28279355
125. Maynard EM, Nordhausen CT, Normann RA. The Utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalography and clinical neurophysiology.* 1997; 102(3):228–239. [https://doi.org/10.1016/S0013-4694\(96\)95176-0](https://doi.org/10.1016/S0013-4694(96)95176-0) PMID: 9129578
126. Denker M, Roux S, Timme M, Riehle A, Grun S. Phase Synchronization between LFP and Spiking Activity in Motor Cortex during Movement Preparation. *Neurocomputing.* 2007; 70(10–12):2096–2101. <https://doi.org/10.1016/j.neucom.2006.10.088>
127. Denker M, Roux S, Lindén H, Diesmann M, Riehle A, Grün S. The Local Field Potential Reflects Surplus Spike Synchrony. *Cereb Cortex.* 2011; 21(12):2681–2695. <https://doi.org/10.1093/cercor/bhr040> PMID: 21508303
128. Datta BN. Numerical Linear Algebra and Applications; 2010.

129. Fisher NI. *Statistical Analysis of Circular Data*. Cambridge University Press; 1995.
130. Platkiewicz J, Stark E, Amarasingham A. Spike-Centered Jitter Can Mistake Temporal Structure. *Neural Computation*. 2017; 29(3):783–803. https://doi.org/10.1162/NECO_a_00927 PMID: 28095192
131. Aalen OO, Borgan Ø, Gjessing HK. *Survival and event history analysis: a process point of view*. *Statistics for Biology and Health*. New York, NY: Springer; 2008.

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

Michel Besserve^{1,2}, Shervin Safavi¹, Bernhard Schölkopf² and Nikos K. Logothetis^{1,3}

¹Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany.

²Department of Empirical Inference, Max Planck Institute for Intelligent Systems, Spemannstrasse 38, 72076 Tübingen, Germany.

³Centre 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 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. On 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 \mathbf{S} be the (frequency x time) spectrogram matrix, NMF seeks optimal matrices \mathbf{W} and \mathbf{H} having only positive coefficients and such that:

$$\mathbf{S} \approx \mathbf{WH} \quad (1)$$

The number of columns of \mathbf{W} is the same as the number of lines of \mathbf{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 \mathbf{W} correspond to the spectral profile of each of these components, the lines of \mathbf{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-off: 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 (Abeyasuriya et al., 2014a, Abeyasuriya 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 shown 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 (Desgraupe, 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 ($p < .01$, Bonferroni 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 stereotactically 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 (Regierungspräsidium, Tübingen Referat 35, Veterinärwesen) 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 were 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., DEGHANI, 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

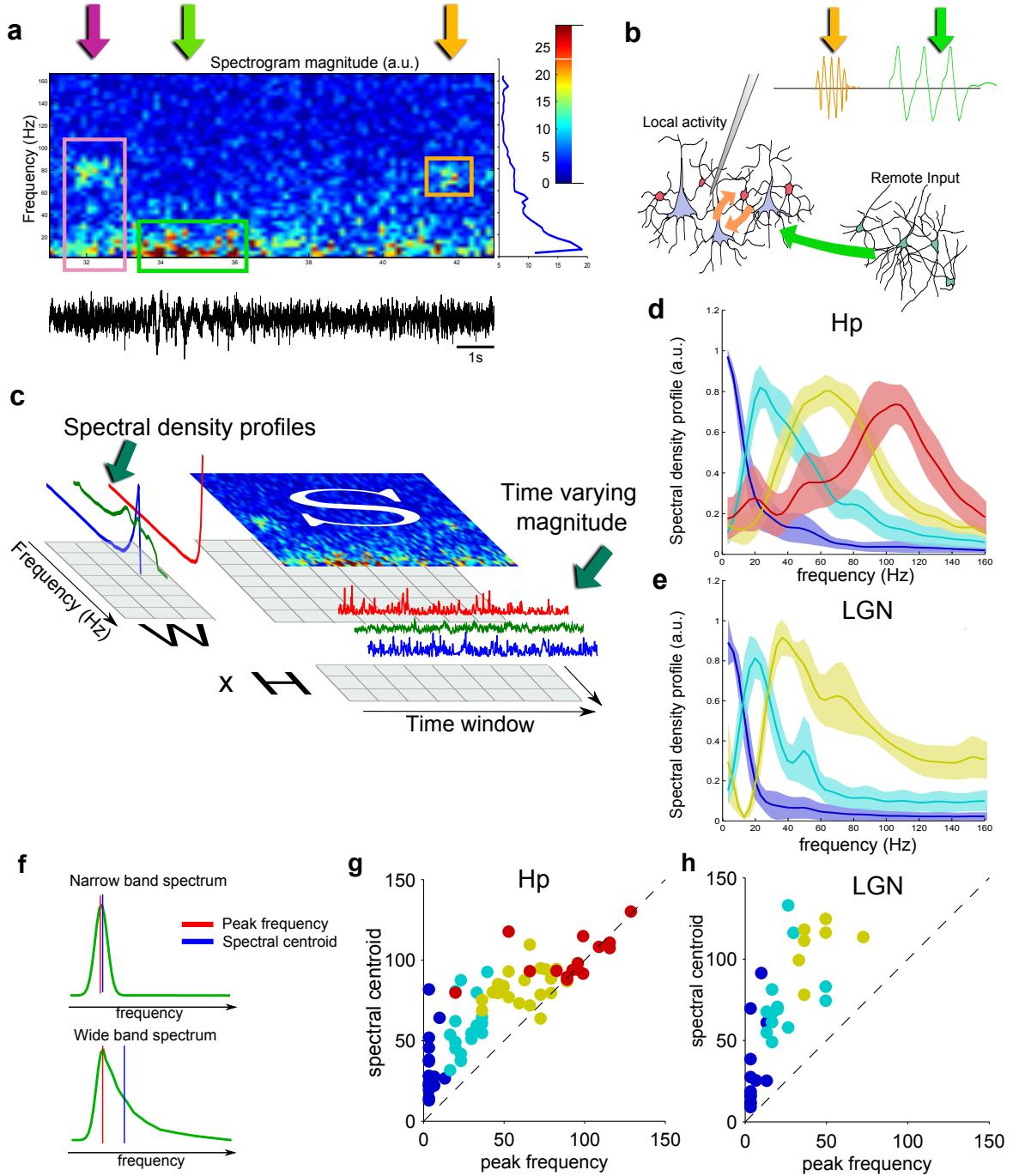


Figure 1: Principle of neural event analysis. (a) Example LFP from the pyramidal layer from monkey CA1 (bottom black trace) and corresponding spectrogram (top) computed with STFT. Rectangles indicate transient spectral power increases, see text. (b) Schematic representation of LFP events (top) and putative 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 \mathbf{S} shown in panel a. \mathbf{S} is approximated by the product of two matrices with non-negative coefficients: \mathbf{W} , gathering the spectral profiles of 3 components, and \mathbf{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.

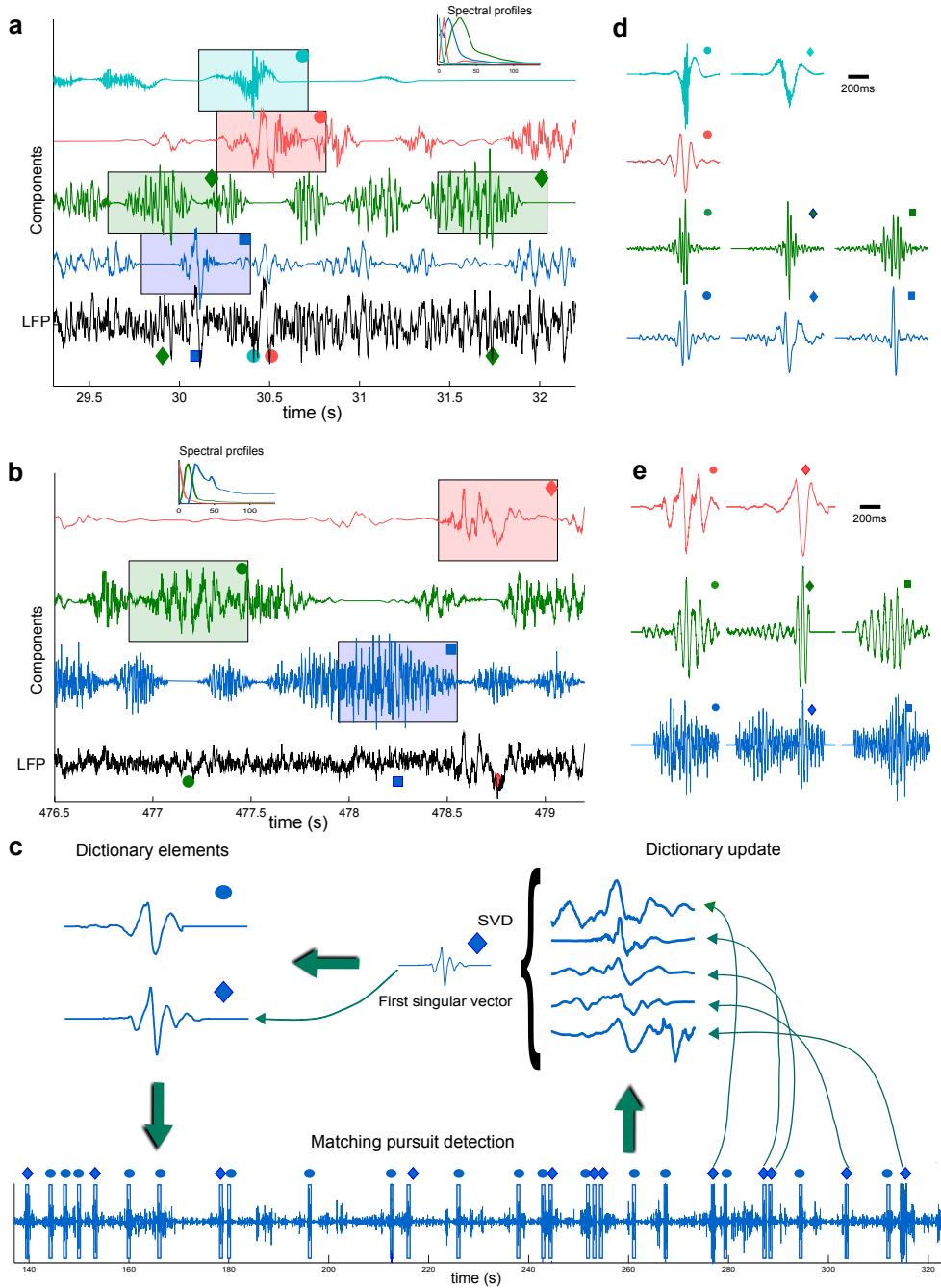


Figure 2: Time course of neural events. (a) Example of dynamical components extracted from CA1 LFP (bottom black trace). Rectangles indicate the location of detected events, the marker in the upper right-hand corner indicates the corresponding dictionary element in panel d. (b) Same as panel a for LFP 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.

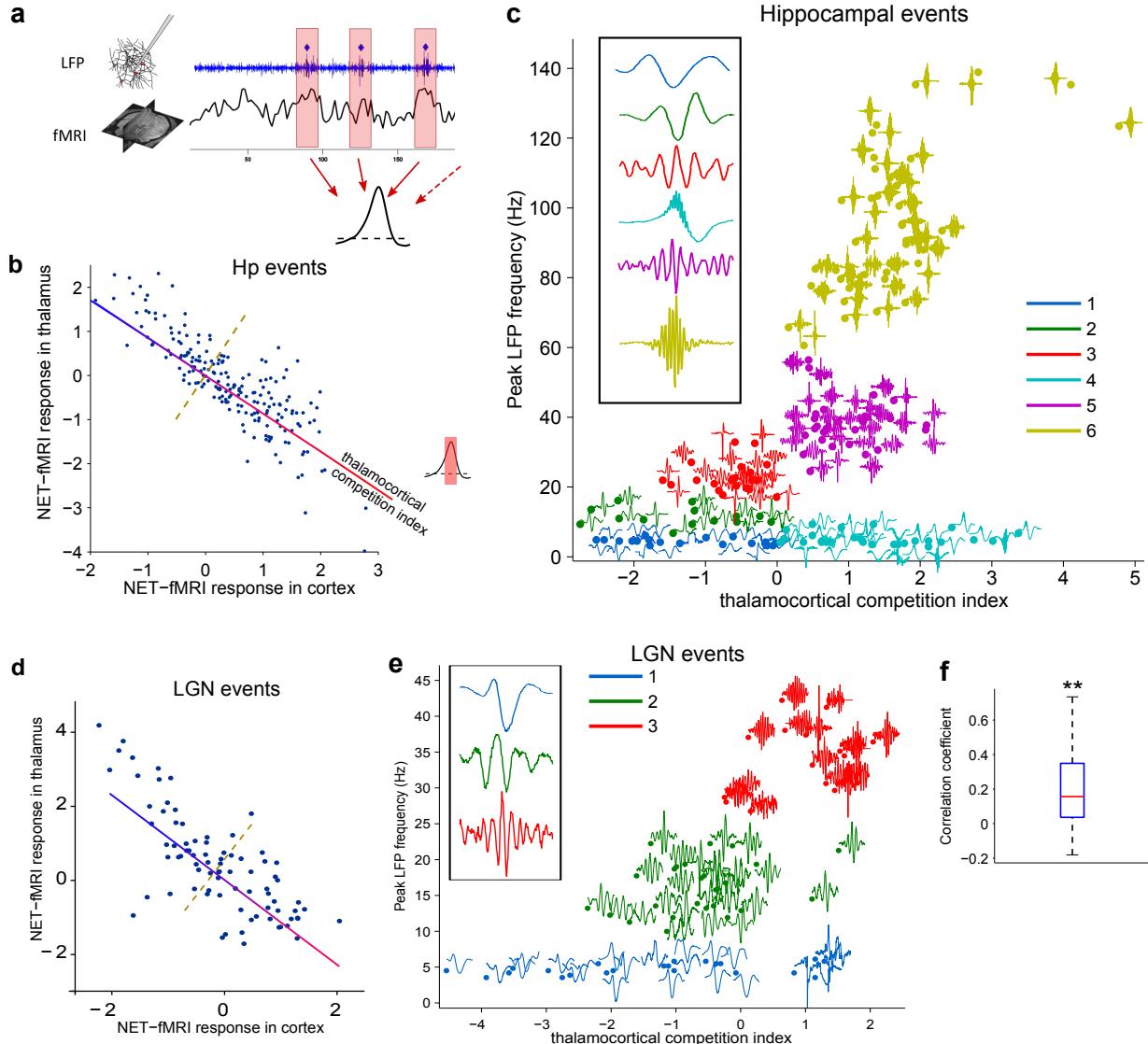


Figure 3: Local versus brain-wide properties of neural events. (a) Schematic representation of the principle of NET-fMRI analysis: the peri-event fMRI traces are averaged across events associated to a given dictionary pattern. (b) Raster plot of the average NET-fMRI response in thalamus and cortex associated 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 identity. (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, $p<.01$, $n=20$)

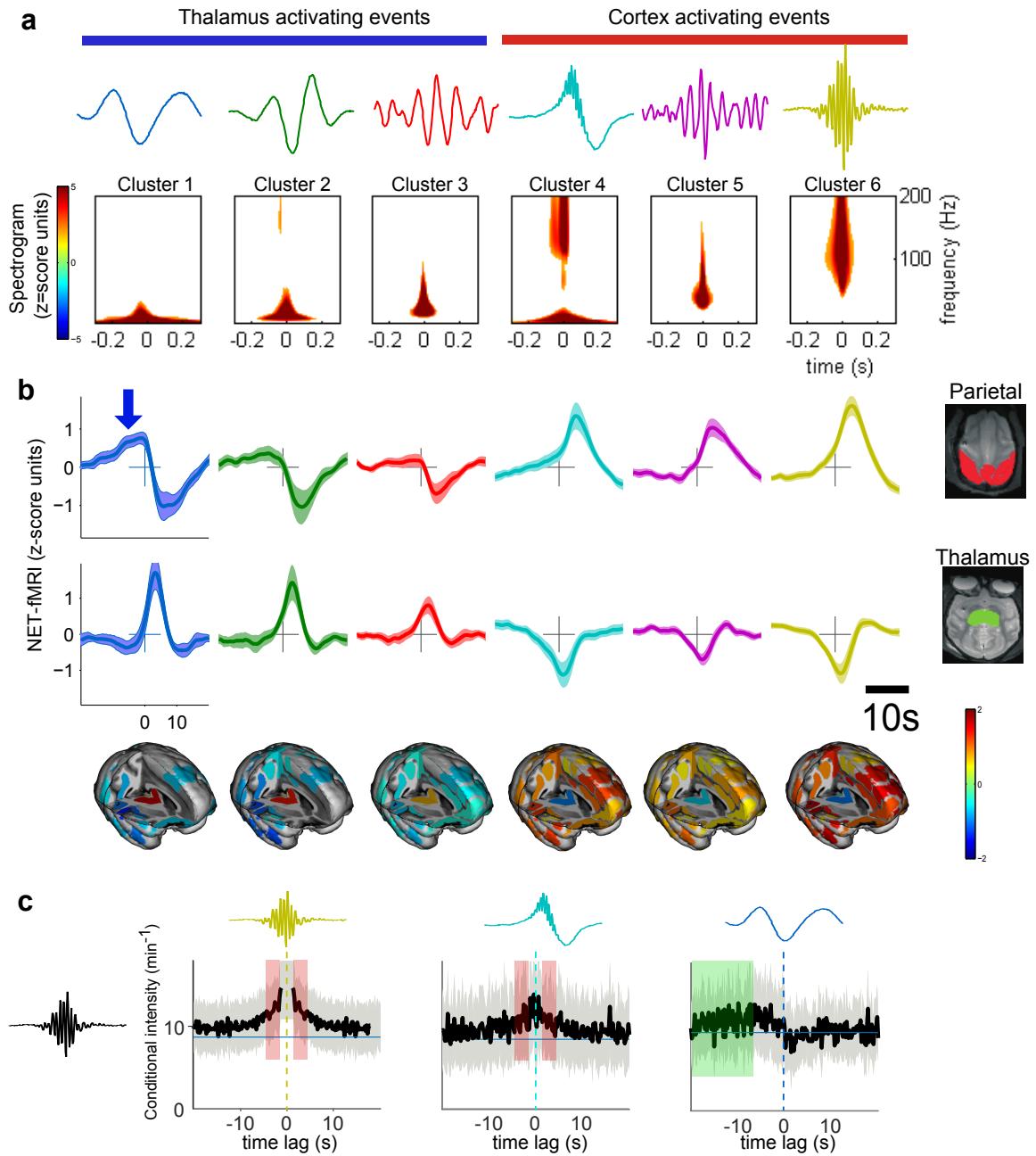


Figure 4: Local and brain-wide physiological properties of hippocampal neural events. (a) Average event triggered time-frequency maps of the 6 clusters of hippocampal events (Z-score with respect to randomized event onsets). Traces at the top represent example dictionary patterns of each cluster. (b) 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 $t=3\text{s}$ overlaid on a template macaque brain.

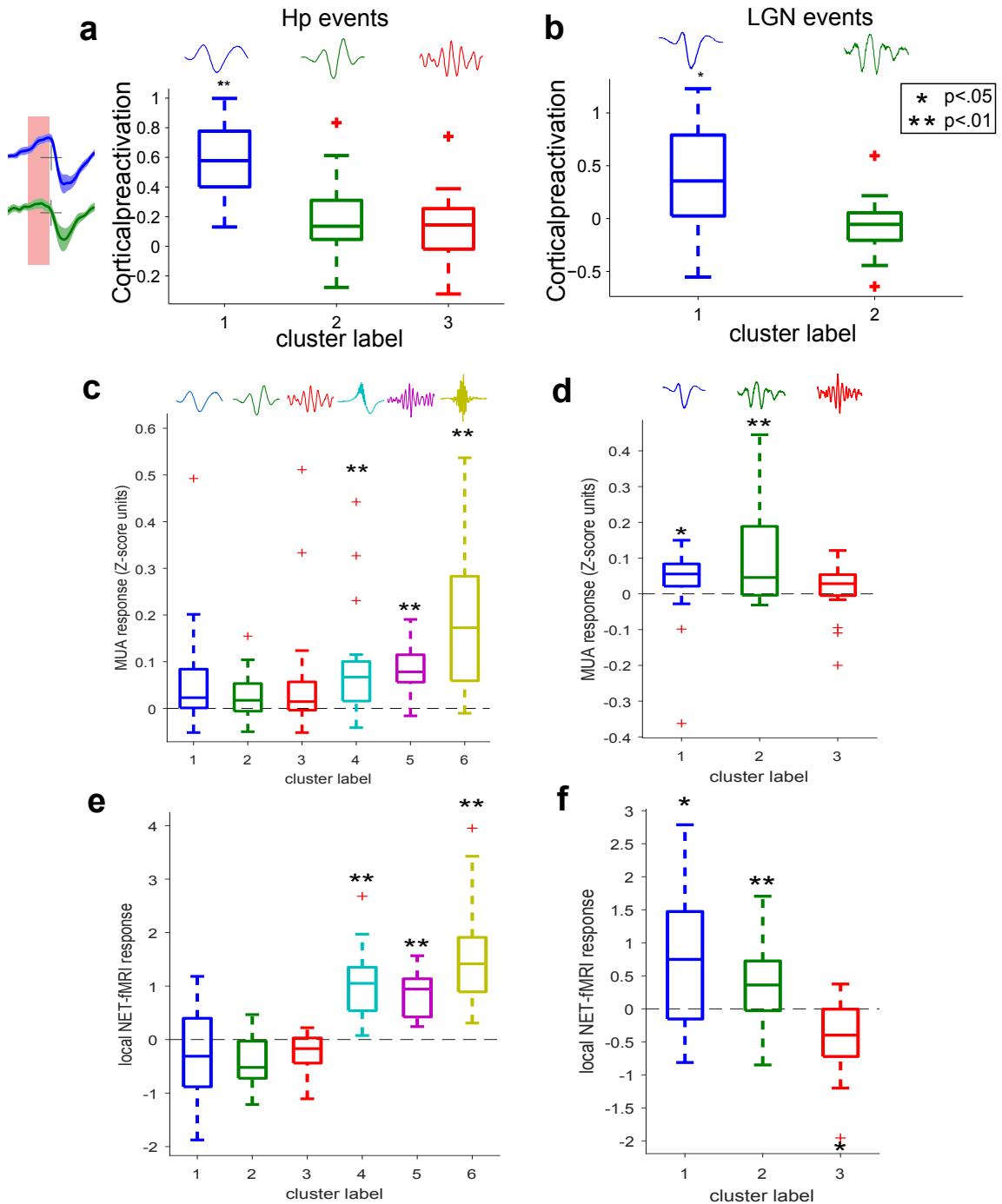


Figure 5: Comparison of LGN and Hp event properties. (a) Distribution of average NET-fMRI response in cortex prior to the event onset (during the pre-activation period, see text) for thalamus activating clusters. Stars indicate significantly positive values (Wilcoxon signed rank test; $p < .01$, Bonferroni 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; $p < .01$, Bonferroni 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; $p < .01$, Bonferroni corrected, from left to right: $n=18, 14, 21, 27, 26, 38$). (d) Same as (c) for LGN events. Stars indicate significantly non-zero median values (Wilcoxon signed rank test; $p < .01$, Bonferroni 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).

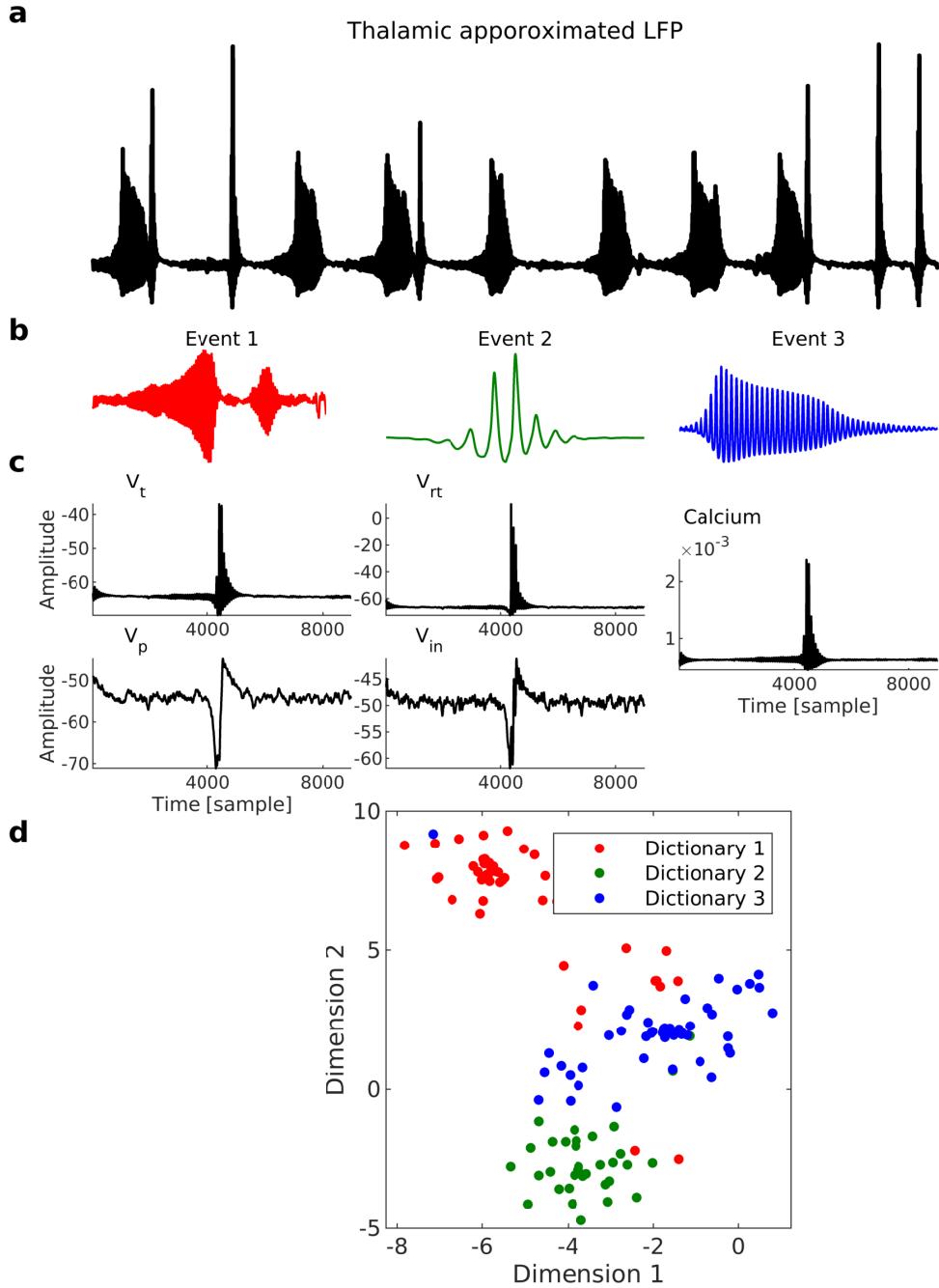


Figure 6: Neural events in a thalamocortical simulation. (a) An exemplary trace for approximated LFP from thalamic module of the thalamocortical simulation. (b) Three Identified events in the approximated LFP. (c) Exemplary trace of cellular variables of the thalamocortical (membrane potential of all 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 Safavi¹, Bernhard Schölkopf² and Nikos K. Logothetis^{1,3}

¹Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany.

²Department of Empirical Inference, Max Planck Institute for Intelligent Systems, Spemannstrasse 38, 72076 Tübingen, Germany.

³Centre 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:

$$\mathbf{x}_n = \sum_{k=1}^K \mathbf{c}_{n,k}, \quad \mathbf{c}_{n,k} = SFT(\{C_k(t_n + m)\}_{m=-T, \dots, T}).$$

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:

$$\mathbf{c}_{k,n} \sim \mathcal{N}_c(0, h_{k,n} \text{diag}(\mathbf{w}_k)).$$

In this expression, the vector $\mathbf{\omega}_k$ 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:

$$\overline{s_{f,n}} = \text{var}(x_{f,n}) = \sum_k h_{k,n} w_{f,k} = (\mathbf{W}\mathbf{H})_{f,n},$$

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 \mathbf{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 $d_{IS}(x|y) \hat{=} \frac{x}{y} - \log \frac{x}{y} - 1$:

$$C_{ML}(\mathbf{W}, \mathbf{H}) =_C \sum_{f,n} d_{IS}(|x_{f,n}|^2 | \sum_k w_{f,k} h_{n,k}) \hat{=} d_{IS}(\mathbf{S} | \mathbf{W}\mathbf{H}).$$

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:

$$d_\beta(x|y) = \begin{cases} \frac{1}{\beta(\beta-1)}(x^\beta + (\beta-1)y^\beta - \beta xy^{\beta-1}) & \beta \in \mathbb{R} \setminus \{0, 1\} \\ x \log\left(\frac{x}{y}\right) + (y - x) & \beta = 1 \\ \frac{x}{y} - \log\left(\frac{x}{y}\right) - 1 & \beta = 0 \end{cases}$$

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 $\nabla_{\mathbf{W}} d_\beta(\mathbf{V}|\mathbf{WH})$, 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 guaranteed 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:

$$X(t) = C_1(t) + \eta(t), \quad \eta(t) = C_2(t) + \dots + C_K(t).$$

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 an 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 learning 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 segments 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 [Abeyseuriya 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 thalamocortical model

1.4.1 Biophysical simulation of thalamocortical 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 have 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 consists 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 have 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 identified 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 thalamocortical (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 dimensional 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 approaches: 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 common center frequency of 5Hz but different bandwidth (.89 and 4.47Hz respectively). Twenty simulations were performed for two cases: either only one or 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 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 through 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

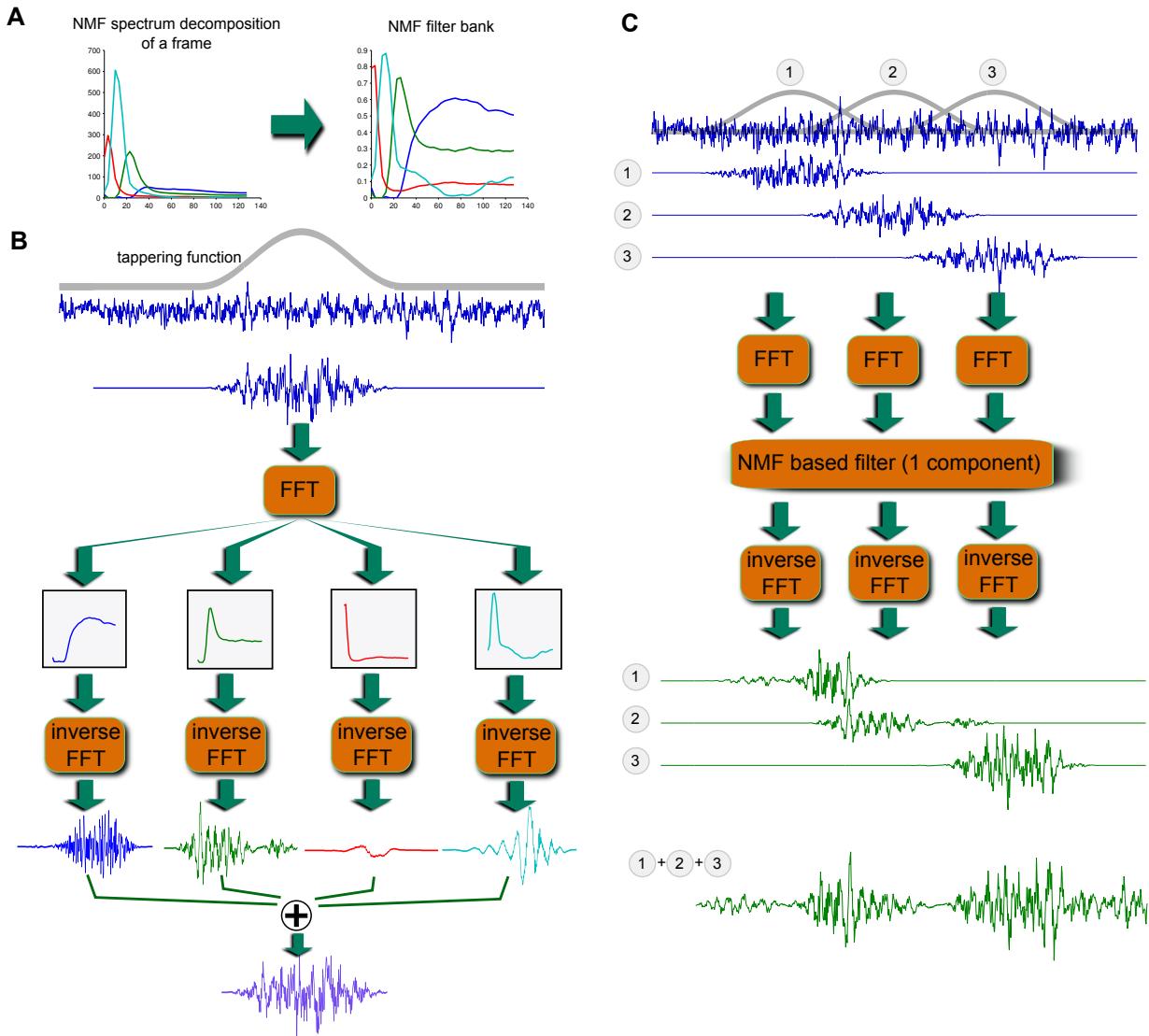
Our approach is a fully unsupervised technique: no additional information (such as behavior or sensory input) is required to assess relevant frequency bands, in contrast to other methods [Magri et al., 2012]. As a consequence our approach is particularly well suited to study ongoing activity in various 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

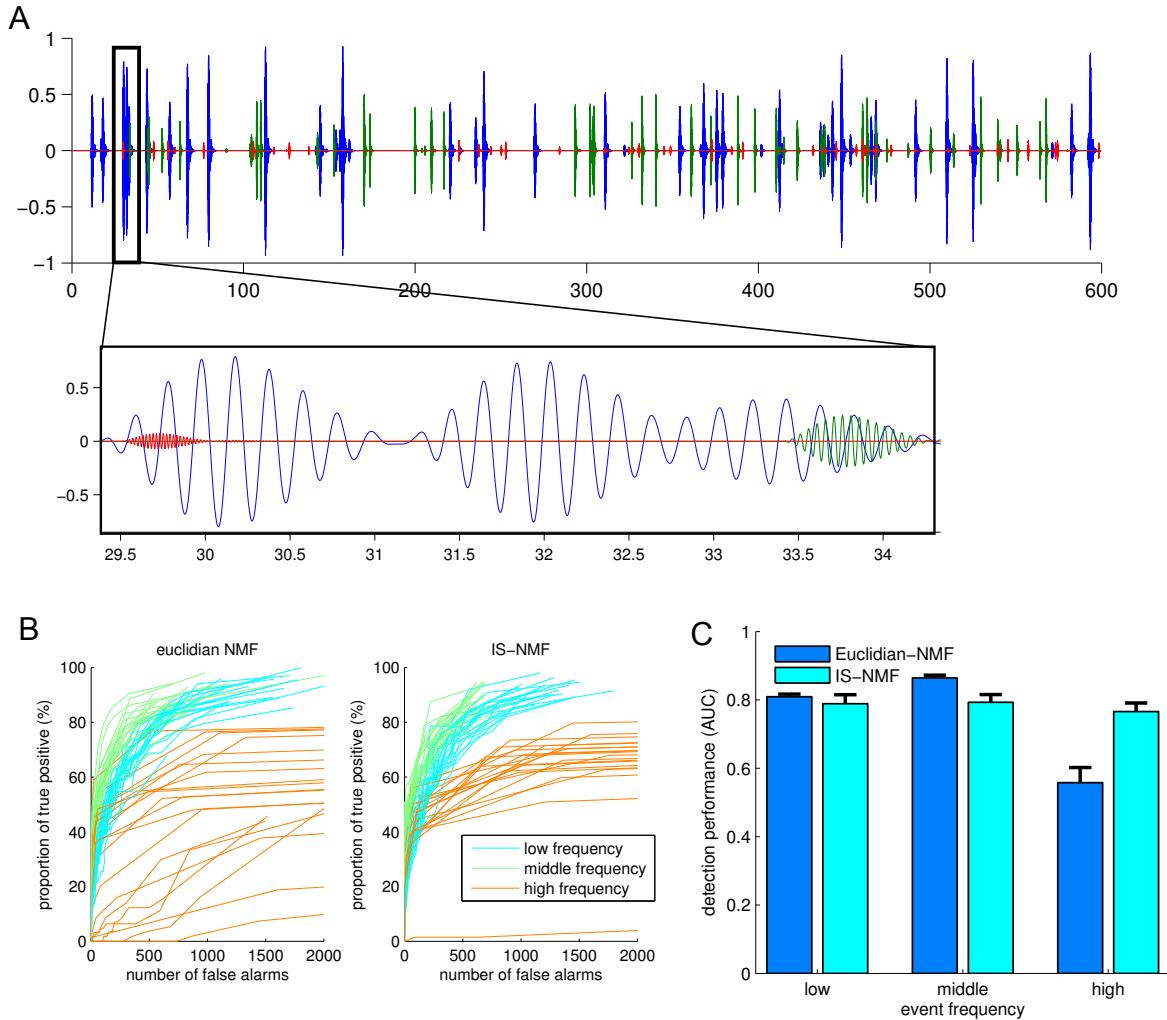
- R. G. Abeysuriya, C. J. Rennie, and P. A. Robinson. Prediction and verification of nonlinear sleep spindle harmonic oscillations. *Journal of Theoretical Biology*, 344:70–77, Mar. 2014. ISSN 0022-5193. doi: 10.1016/j.jtbi.2013.11.013.
- M. Aharon, M. Elad, and A. Bruckstein. -svd: An algorithm for 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.

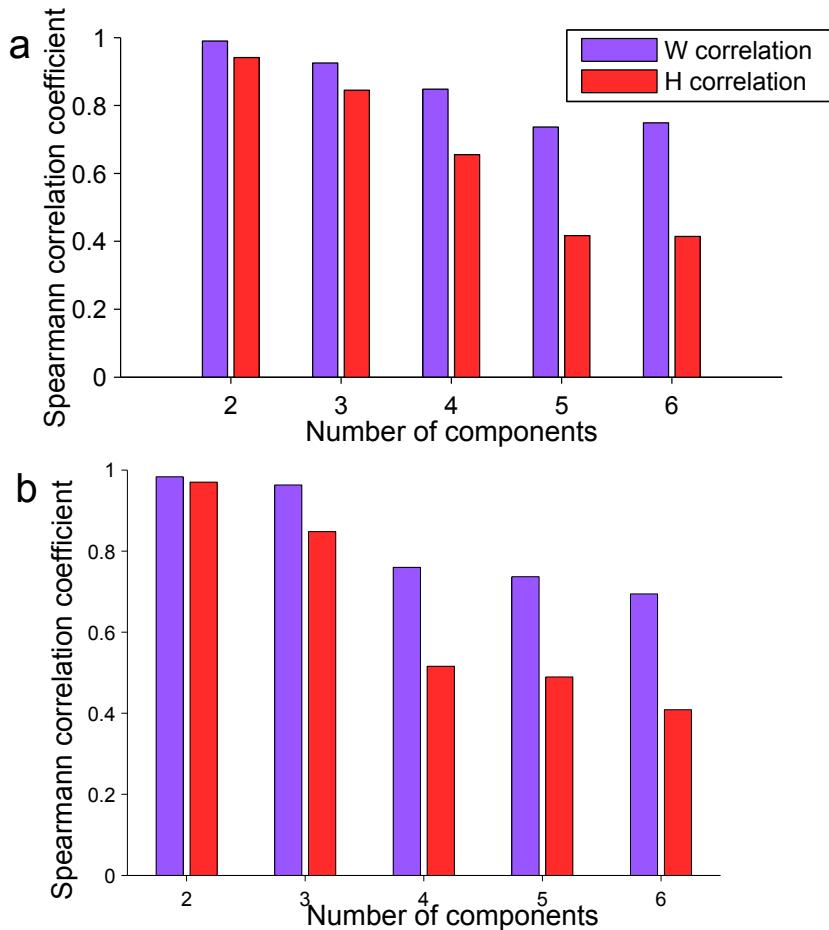
- L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. *J. Mach. Learn. Res.*, 9(86):2579–2605, 2008. ISSN 1533-7928.
- A. Weigenand, M. S. Costa, H.-V. V. Ngo, J. C. Claussen, and T. Martinetz. Characterization of K-Complexes and Slow Wave Activity in a Neural Mass Model. *PLOS Computational 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.



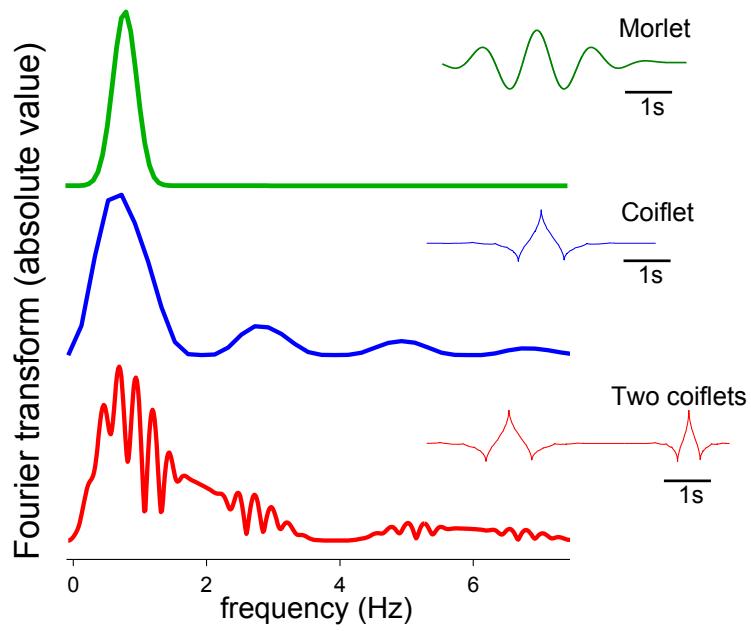
Supplementary Figure 1: Principle of the extraction of the time course of dynamical components based on NMF results. (a) For a given time window, the weighted spectral profiles of all NMF components are normalized for each frequency in order to sum to one. (b) These resulting functions are used as transfer 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.



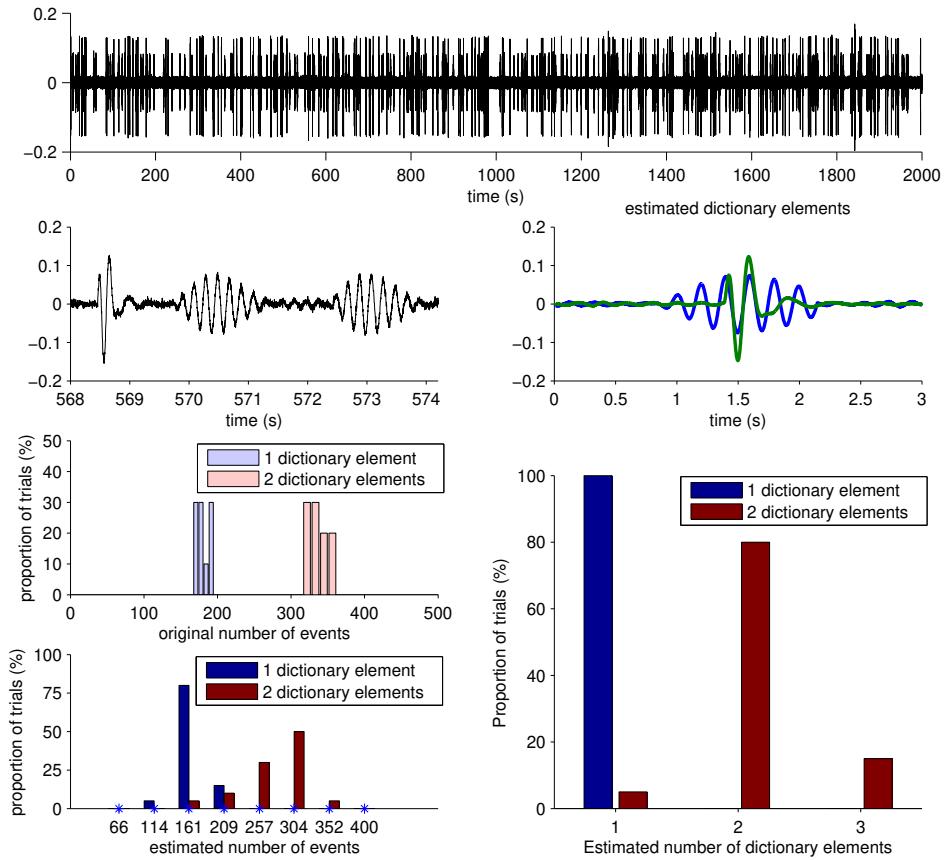
Supplementary Figure 2: **Comparison of Euclidean and Itakura-Saito NMF in simulation.** (a) Example time course of simulated dynamical components, representing low frequency (in blue), middle frequency (in green) and high frequency oscillations (in red). The inset represents a magnified version of the 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).



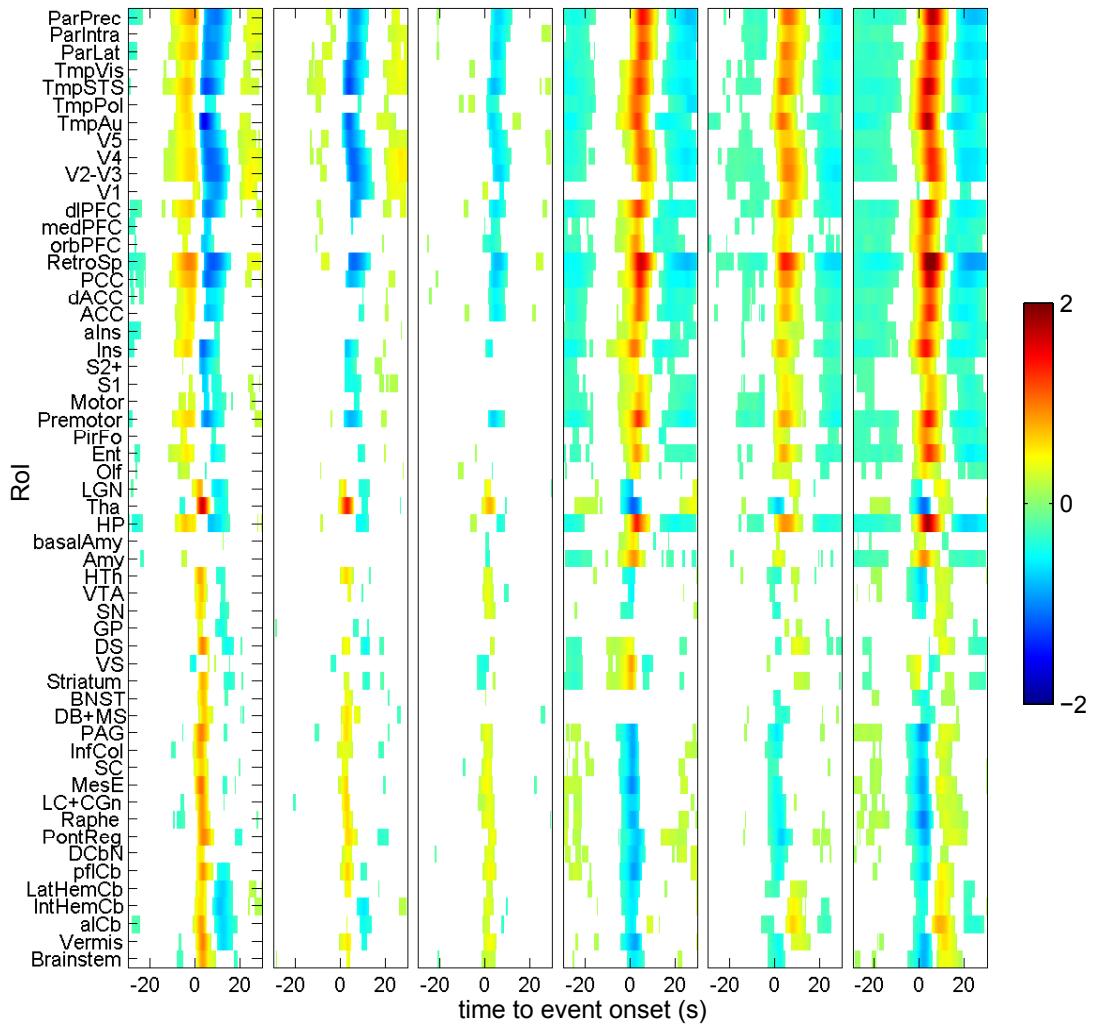
Supplementary Figure 3: **Selection of the number of NMF components.** (a) Average cosine similarity between bootstrapped NMF components averaged across all Hp recording sessions (blue: similarity between spectral profiles stored in the columns of the matrix \mathbf{W} , red similarity between time-varying contribution of profiles to each time window stored in the lines of the matrix \mathbf{H}) (b) Average cosine similarity between bootstrapped NMF components averaged across all LGN recording sessions.



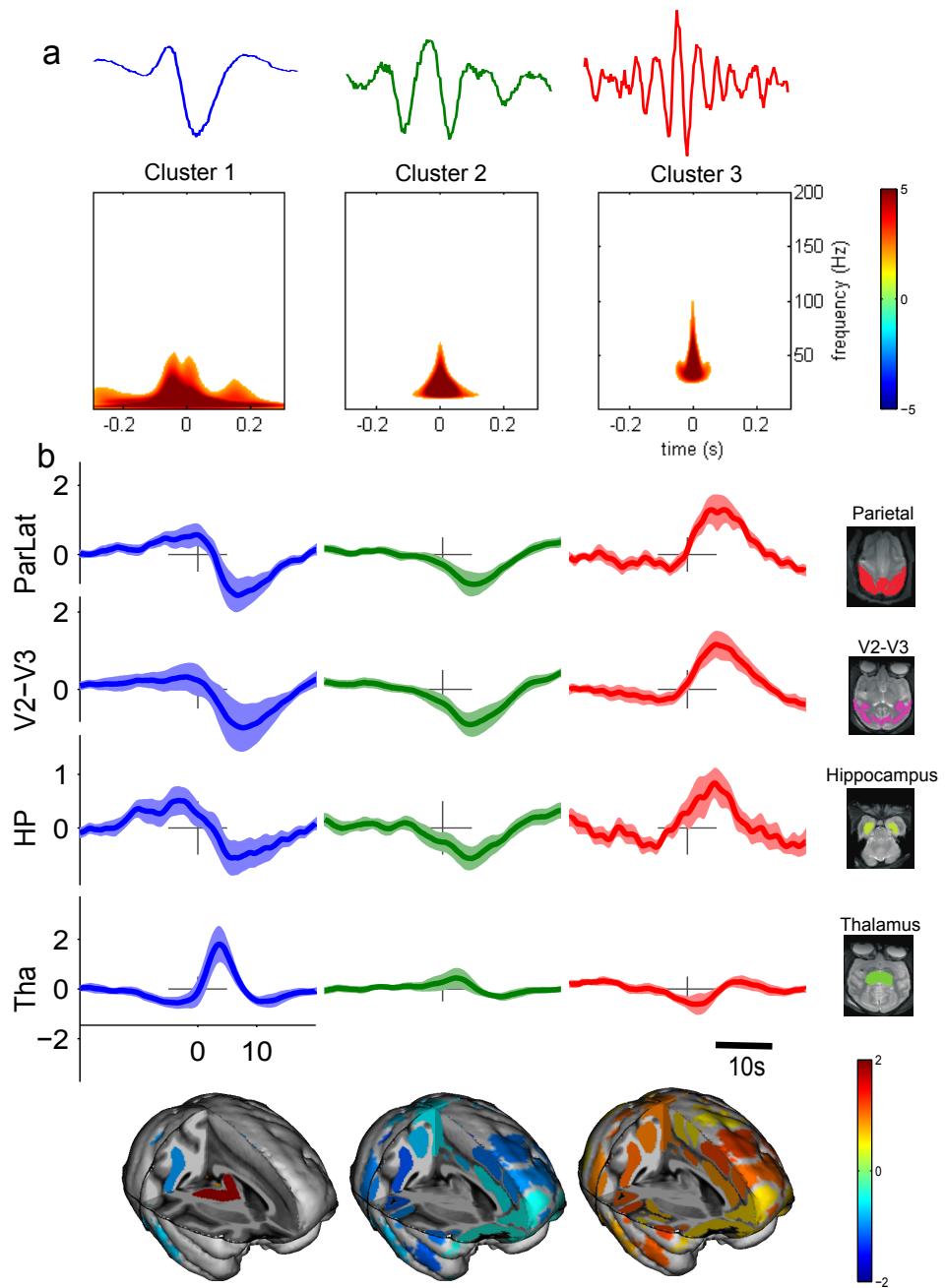
Supplementary Figure 4: **Properties of spectral profiles.** Absolute Fourier transform for different types of temporal patterns (see Supplementary Methods). Insets on the right hand side represent the time course of each pattern.



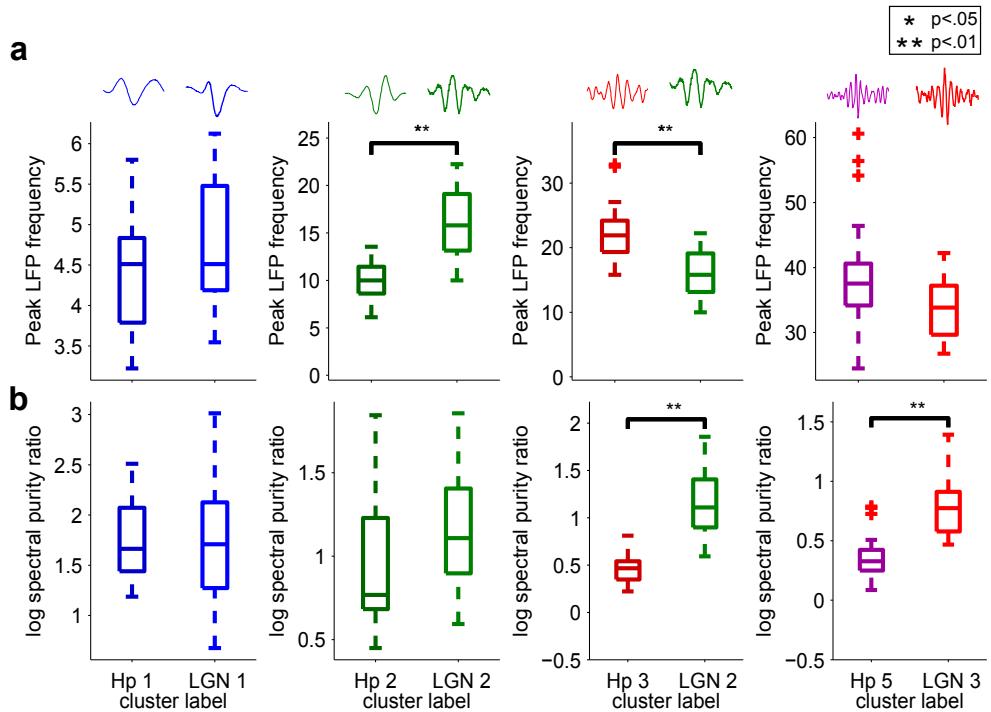
Supplementary Figure 5: **Dictionary learning.** (a) Example time course of a simulated time series with events taken from a dictionary with two patterns. (b) Magnified portion of time-course in (a) showing the occurrence of the first (blue arrow) and the second (green arrow) dictionary pattern. (c) 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.



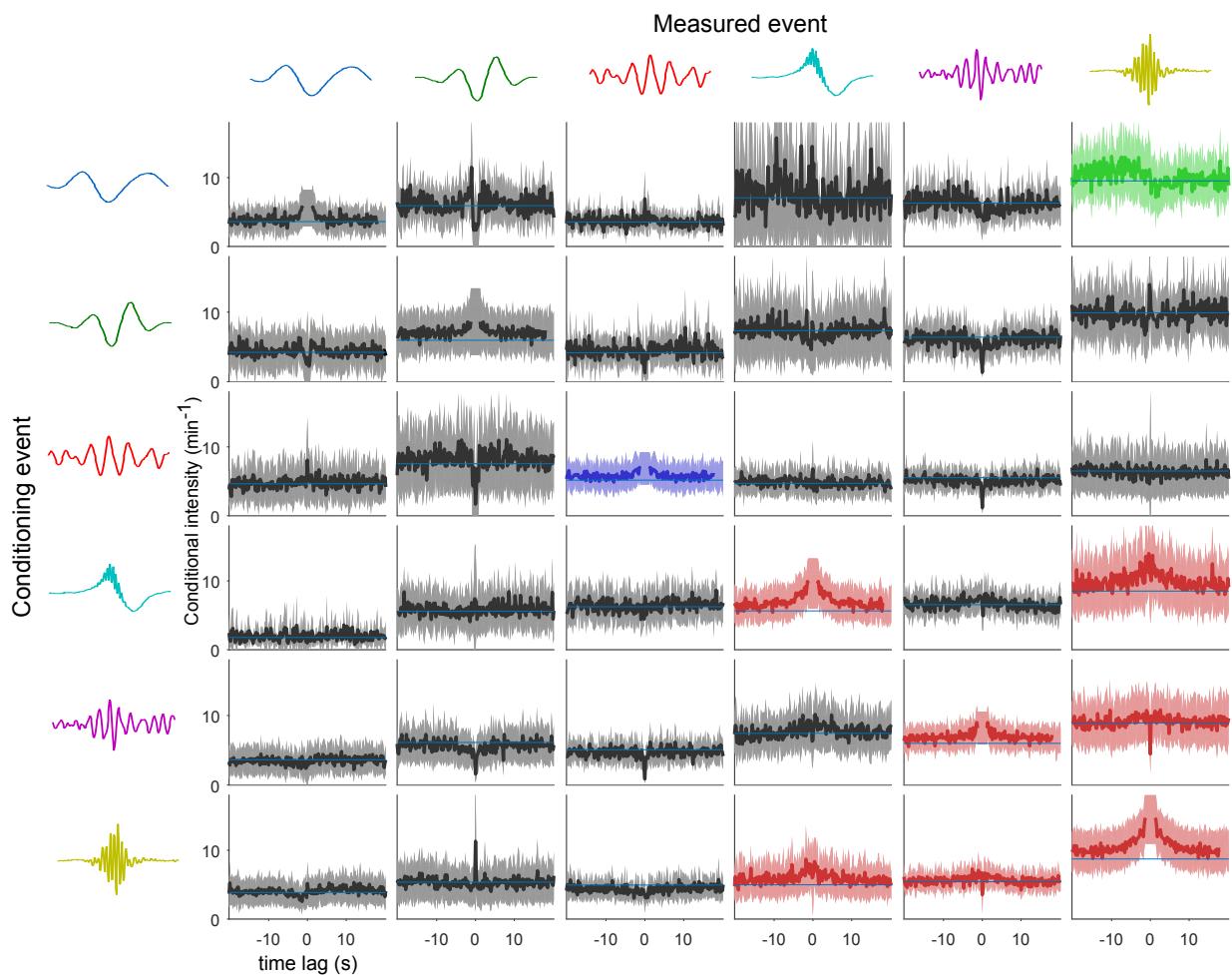
Supplementary Figure 6: **Full NET-fMRI response for hippocampal events.** NET-fMRI response averaged across voxels of each given ROI for each cluster of events. Rows indicate the considered ROI, columns indicate the cluster of events (same order as Figure 4). Values are z-scored with respect to randomized events. 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; $p < .05$, Bonferroni 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; $p < .05$, Bonferroni 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 occurrence 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 rank test; $p < .05$; FDR adjusted): red plots indicate change in the $[-3\text{s}, 3\text{s}]$ time window, excluding a $[-1\text{s}, 1\text{s}]$ interval and tested across all CA event pairs, blue plots indicate the same tested for all TA pairs, green curves indicate change in the $[-20\text{s}, -7\text{s}]$ 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 |
| dIPFC | 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 | Entorhinal Cortex |
| HP | Hippocampus |
| Amy | Amygdala |
| HTh | Hypothalamus |
| Septum | Septum |
| DB | Diagonal Band |
| GP | Globus 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 |
| aICb | Anterior Cerebellar Lobe |
| DCbN | Deep Cerebellar Nuclei |
| pICb | Parafalcinodular 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).

Signatures of criticality in efficient coding networks

Shervin Safavi^{a,b,?}, Matthew Chalk^{c,?}, Nikos Logothetis^{a,d}, and Anna Levina^{a,b,e,?}

^aMax Planck Institute for Biological Cybernetics, Tübingen, Germany; ^bUniversity of Tübingen, Tübingen, Germany; ^cSorbonne Université, Paris, France; ^dInternational Center for Primate Brain Research, Songjiang, Shanghai, China; ^eBernstein Center for Computational Neuroscience Tübingen, Tübingen, Germany

1 The critical brain hypothesis states that the brain can benefit from
 2 operating close to a second-order phase transition. While it has been
 3 shown that several computational aspects of sensory information
 4 processing (e.g., sensitivity to input) are optimal in this regime, it is
 5 still unclear whether these computational benefits of criticality can be
 6 leveraged by neural systems performing behaviorally relevant compu-
 7 tations. To address this question, we investigate signatures of
 8 criticality in networks optimized to perform efficient encoding of stim-
 9uli. We consider a spike-coding network of leaky integrate-and-fire
 10 neurons with synaptic transmission delays and input noise. Previ-
 11ously, it was shown that the performance of such networks varies
 12 non-monotonically with the noise amplitude. Interestingly, we find
 13 that in the vicinity of the optimal noise level for efficient coding, the
 14 network dynamics exhibits signatures of criticality, namely, the dis-
 15 tribution of avalanche sizes follows a power law. When the noise
 16 amplitude is too low or too high for efficient coding, the network
 17 appears either super-critical or sub-critical, respectively. Our work
 18 suggests that two influential, and previously disparate theories of
 19 neural processing optimization – efficient coding, and criticality – may
 20 be intimately related.

Criticality | Efficient coding | Neural dynamics | Neural computation

1 Attempts to understand information processing in the brain
 2 have led to the formulation of various optimality principles.
 3 Two major paths, among others, have been explored to uncover
 4 these principles. On one hand, a large body of studies starts
 5 from the theoretical and experimental finding that neural net-
 6 works operate close to criticality (1, 2). Researchers have thus
 7 sought to investigate what, if any, could be the computational
 8 advantages of a network operating near a critical point (3).
 9 Meanwhile, another line of research presumes that neural net-
 10 works have evolved to efficiently encode natural inputs (given
 11 constraints such as limited energy and noise). Here, the key
 12 question was investigating how neural networks could achieve
 13 such optimal encoding, and what are the resulting dynamics.
 14 In a nutshell: one line of research starts with an observation
 15 about neural dynamics (i.e., that they are near-critical) and
 16 seeks to find the coding advantages; the other starts with a
 17 coding objective (e.g., efficient coding) and seeks to under-
 18 stand the resulting dynamics. However, despite the prevalence
 19 of both approaches, connections between theories based on
 20 closeness to criticality and efficient coding hypothesis remain
 21 elusive.

22 To address this shortcoming, we introduce a complementary
 23 approach. Instead of tuning the network around the critical
 24 point and evaluating its statistical information processing per-
 25 formance, we optimize a network to perform a clearly defined
 26 computation and investigate if signatures of critical dynamics
 27 emerge in the optimized network. We focus on efficiently en-
 28 coding the input, a well-established and functionally relevant

computation, accompanied by a rich body of normative models (4) and neural dynamics (e.g., 5).

We analyzed a network of leaky integrate-and-fire (LIF) neurons that can be optimized (by adjusting the noise) to code a one-dimensional input (6). We evaluate the signatures of criticality, such as scale-freeness of the activity propagation cascades, termed neuronal avalanches, in networks with different noise amplitudes. Interestingly, we only observed scale-free neural avalanches in the vicinity of optimal noise for efficient coding. This result suggests that coding-based optimality co-occur with closeness to criticality.

Results

We investigate a network of LIF neurons consisting of an excitatory and an inhibitory population. The network's dynamics and connectivity is set up such that it can precisely encode a feed-forward input using a minimal number of spikes. In an idealized network with instantaneous synapses (7), recurrent inhibition removes redundancy between neurons. However, the introduction of realistic synaptic delays leads to network synchronization that impairs coding efficiency (for more details, see, e.g., , 6). In the presence of synaptic delays, this network can nonetheless be optimized for efficient coding by adding noise (6, 8, 9) (or increasing the L1 norm which controls the spiking threshold, see 10). As it was shown in previous studies (see, e.g., , 6, 15), the network's performance depends non-monotonically on the noise amplitude, with the optimal performance achieved for an intermediate noise level.

To assess the signatures of criticality in the efficient coding network, we investigate the distribution of neural avalanches in networks with different levels of noise. To begin, we keep the network size fixed, at $N = 100$ (as originally used in, 6). A neuronal avalanche is defined as an uninterrupted cascade of spikes in the network (11). As suggested by (11), the period without spiking signifies the end of the previous avalanche if it is longer than the mean inter-spike interval (ISI) in the compound spike train (obtained by collapsing the spike trains of all neurons). Similar results were obtained using other thresholding choices (see supplementary methods), as in (11).

In a system operating close to criticality, the avalanche size (number of spikes in the cascade) follows a power-law distribution. We demonstrate that the distribution of avalanche

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., M.C. and A.L.; Writing - Review & Editing, S.S., M.C. N.K.L. and A.L.; Visualization, S.S., M.C., A.L.; Supervision and project administration, M.C. and A.L.; Funding acquisition, N.K.L. and A.L.

The authors declare no competing interest.

?To whom correspondence should be addressed. E-mail: research@shervinsafavi.org, anna.levina@uni-tuebingen.de, matthew.chalk@inserm.fr

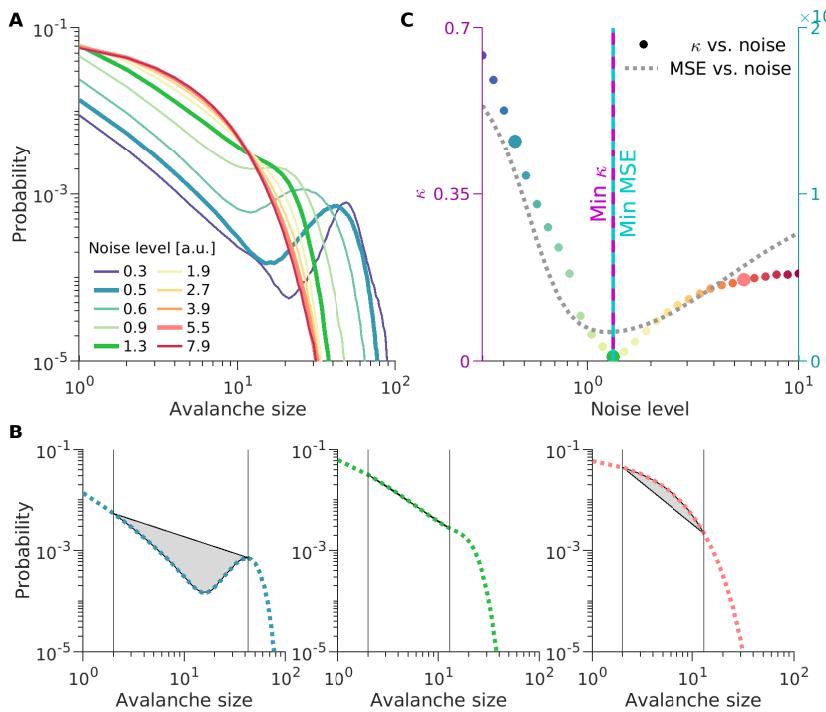


Fig. 1. Co-occurrence of the criticality and optimal settings for efficient coding.

(A) Avalanche-size distributions of efficient coding networks with different noise levels (indicated in the legend and with consistent color code across all panels). (B) Deviation from criticality measure κ for three noise levels. Left: small noise (0.5, network appears supercritical, many large avalanches); middle: medium noise (1.3, network close to criticality); right: strong noise (5.5, network exhibit subcritical behavior with predominantly small avalanches). (thicker lines in A): The area of gray-shaded regions between the actual avalanche size distribution and fitted power-law distribution defines the deviation measure κ (in the middle panel, the filled region is not visible, as the avalanche size distribution is very close to the ideal power-law). Vertical lines indicate the choices of left and right cut-offs (see main text for more details). Distributions for the chosen noise levels are highlighted in bold in panel A (matching colors). (C) Deviation from power-law κ as a function of noise level (left y-axis, color matching the panel A. Gray dotted line indicates mean-square-error (MSE) (right y-axis) as a function of noise. Vertical continuous line (purple) indicates the noise level corresponding to minimal κ (the most scale-free avalanche size distribution), Vertical broken line (cyan) indicates the noise level corresponds to minimum MSE (the best efficient coding performance). These two vertical lines overlap exactly, demonstrating the coincidence of noise levels for scale-free behavior and efficient coding.

70 sizes systematically changes with the strength of added noise
 71 (Figure 1A). In networks with a small amount of noise (e.g.,
 72 noise strength 0.5, thick blue line in Figure 1A, or Figure 1B
 73 left), large avalanches dominate the distribution of avalanche
 74 sizes (a bump in the tail of the distribution signifies a transient
 75 synchronization in the network). On the other hand, for a
 76 large amount of noise (e.g., noise strength 5.5, thick red line
 77 in Figure 1A, or Figure 1B right), the distribution is concen-
 78 trated on the small avalanches (an exponential distribution).
 79 However, for intermediate levels of noise (e.g., noise strength
 80 1.3, thick green line in Figure 1A, or Figure 1B middle), the
 81 avalanche-size distribution resembles a power-law (appears as
 82 a linear function in the log-log coordinates), which is a key
 83 signature of criticality in neural systems (see, e.g., 11).

84 To determine the most scale-free avalanche distribution
 85 (closest to a power-law distribution), we use a deviation mea-
 86 sure κ that quantifies deviation from an ideal power-law dis-
 87 tribution. Our κ measure closely follows the non-parametric
 88 measure introduced by Shew and colleagues (12) but does
 89 not assume a particular scaling exponent (see supplementary
 90 methods). κ is defined as the normalized area between the
 91 empirical and the ideal (fitted power-law for the portion of
 92 data between two cut-offs) distribution (Figure 1B). κ takes
 93 small (close to zero) values for a scale-free distribution (Fig-
 94 ure 1B middle) and deviates from zero otherwise (Figure 1B
 95 left and right panels).

96 We measure how deviations from a power-law, κ , and the
 97 network's reconstruction error depends on the noise strength.
 98 We confirm the previous observation that the performance
 99 of this network depends non-monotonically on the noise am-
 100 plitude (gray dotted curve in Figure 1C), with the optimal
 101 performance achieved for an intermediate noise level (6, 15).
 102 Interestingly, the change in κ with the noise level demon-
 103 strates a similar non-monotonic behavior (colorful circles in
 104 Figure 1C). Remarkably, they both are minimized at the same

105 noise level, resulting in a coincidence of the optimal point for
 106 coding and the most scale-free distribution (vertical purple
 107 and cyan line in Figure 1C). This observation offers additional
 108 support to the criticality hypothesis for the brain, namely that
 109 the various information processing measures are optimized
 110 close to the critical point (3, 13).

111 We next verify the stability of this result to changes in the
 112 network's size by considering networks of various sizes in a
 113 range between $N = 50$ and $N = 400$ neurons. We find that
 114 all networks demonstrate similar non-monotonic behavior for
 115 the dependence of reconstruction error (Figure 2A) and the
 116 scale-freeness deviation measure κ (Figure 2B) on the strength
 117 of noise. This non-monotonic behavior of the reconstruction
 118 error is less pronounced for larger networks. This is expected,
 119 because the recurrent network used in our study is particularly
 120 suitable to code a single dimension of input by a small number
 121 of neurons (6, 7), i.e., hundreds, rather than thousands, of
 122 neurons per input dimension (see 10, as to how this problem
 123 could be alleviated for large networks by encoding higher
 124 dimensional inputs).

125 We observe the co-occurrence of efficient coding optimality
 126 with criticality optimality across all network sizes. The
 127 noise levels where coding error is minimal (x-coordinates in
 128 Figure 2C) and where κ is minimal (y-coordinates in Fig-
 129 ure 2C) are highly correlated across different network sizes.
 130 Furthermore, this observation is robust to variations in the
 131 choice of the right cut-off needed for calculating κ (whiskers
 132 in Figure 2C indicate the standard deviation with respect to
 133 changing the ways κ is computed). Lastly, the location of the
 134 cut-off of the scale-free distribution shifts right (to the larger
 135 values) with the size of the network (Figure 2D), hinting at
 136 the correct finite-size scaling behavior (see, e.g., 11, 14).

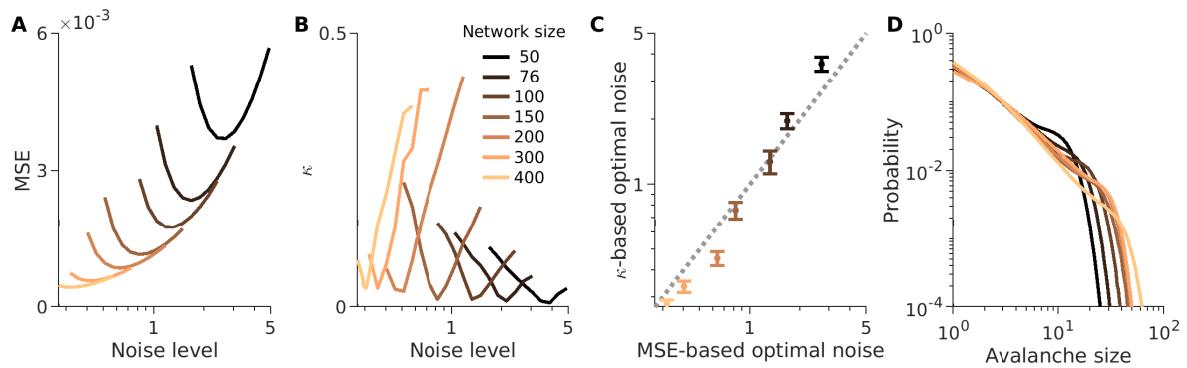


Fig. 2. Co-occurrence of the criticality and efficient coding optimality across networks of different sizes.

(A) Mean-square-error (MSE) of stimulus reconstruction for different injected noise amplitudes (similar to Figure 1C, gray line). Curves with different colors correspond to different network sizes (specified in the legend of panel B). (B) Deviation measure (κ) as function of noise (similar to Figure 1C, colorful dots). As in (A), different curves represent κ -noise relationship for networks of different sizes. (C) Y-coordinate of each point specifies the value of optimal noise chosen based on scale-freeness of the avalanche size distributions (minimum of κ), and the X-coordinate specifies the value of optimal noise chosen based on efficient coding criterion (minimum MSE). Error bars (mean \pm standard deviation) indicate the variability across a wide range of choices of free parameters used for computing the deviation measure κ (see the supplementary text for more detail). (D) The cutoff of the power-law distribution for the most scale-free avalanche size distribution (resulting in smallest κ on panel B) for different network sizes shifts with the network size as expected from finite-size scaling ansatz for critical systems (colors are specified in the legend of panel B).

Discussion

In this study, we probe the connection between the optimality discussed in the context of criticality hypothesis of the brain, and the optimality discussed in theories of neural computations. To this end, we examine an efficient coding network (6, 7) for signatures of criticality. We find signatures of criticality (scale-free dynamics of the neural avalanches) emerging in a network that was *designed based on criticality-agnostic principles* merely by optimizing the coding performance. This suggests criticality and efficient coding are intimately related.

Our approach contrasts with previous work investigating the criticality hypothesis, which used models (e.g., a branching network, a recurrent neural network) that can attain various (critical/non-critical) states depending on a limited number of control parameters (e.g., branching ratio, connection strength) and then quantified how the computational primitives (3, 13), such as sensitivity to an input, depend on these control parameters. These state-generating models aim at reproducing realistic neuronal dynamics. They are typically driven by a slowly delivered noise and have no specific input and no read-out strategy. Therefore, studies based on branching networks are largely agnostic to computational objectives central to theories of neural computation. Our approach thus paves the way for new questions about the relevance of criticality for precisely defined and task-relevant computations.

Future research should address why and how critical dynamics enables optimal efficient coding. For instance, it is not clear what is the exact role of neural avalanches, and why and how their scale-free distribution may be optimal for neural coding. Answering such questions requires going beyond our simulation-based approach, and, similar to (15). Mathematical analysis is needed to understand how the distribution of avalanches depends on different attributes of the network (noise, delay, connection weights, etc) and how those attributes affect the coding optimality. Such approaches will provide a mechanistic insight for our observations, and also, allow us to extend our framework to more sophisticated computations. We believe, our study opens up promising avenues for future investigations to establish the connection between other aspects

of criticality (e.g., 16, 17) and theories of neural computations (e.g., 7).

Materials and Methods

Further details are provided in the Supporting Information.

ACKNOWLEDGMENTS. We thank Roxana Zeraati for sharing her codes for analysis of power-laws and her input on visualization; Victor Buendia for helpful discussions; Joachim Werner and Michael Schnabel for their IT support. This work was supported by the Max Planck Society, International Max Planck Research School for Cognitive and Systems Neuroscience, and the Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation to AL.

1. MA Muñoz, Colloquium: Criticality and dynamical scaling in living systems. *Rev. Mod. Phys.* **90**, 031001 (2018).
2. J O’Byrne, K Jerbi, How critical is brain criticality? *Trends Neurosci.* (2022).
3. WL Shew, D Plenz, The functional benefits of criticality in the cortex. *The Neurosci. : a review journal bringing neurobiology, neurology psychiatry* **19**, 88–100 (2013).
4. M Chalk, O Marre, G Tkacik, Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 186–191 (2018).
5. J Kadmon, J Timcheck, S Ganguli, Predictive coding in balanced neural networks with noise, chaos and delays. *ArXiv200614178 Cond-Mat Q-Bio Stat* (2020).
6. M Chalk, B Gutkin, S Deneve, Neural oscillations as a signature of efficient coding in the presence of synaptic delays. *eLife* **5** (2016).
7. M Boerlin, CK Machens, S Deneve, Predictive coding of dynamical variables in balanced spiking networks. *PLoS computational biology* **9**, e1003258 (2013).
8. V Koren, S Deneve, Computational Account of Spontaneous Activity as a Signature of Predictive Coding. *PLoS computational biology* **13**, e1005355 (2017).
9. CER Buxó, JW Pillow, Poisson balanced spiking networks. *PLOS Comput. Biol.* **16**, e1008261 (2020).
10. N Calaim, FA Dehmelt, PJ Gonçalves, CK Machens, The geometry of robustness in spiking neural networks. *eLife* **11**, e73276 (2022).
11. JM Beggs, D Plenz, Neuronal avalanches in neocortical circuits. *The J. neuroscience : official journal Soc. for Neurosci.* **23**, 11167–77 (2003).
12. WL Shew, H Yang, T Petermann, R Roy, D Plenz, Neuronal Avalanches Imply Maximum Dynamic Range in Cortical Networks at Criticality. *J. Neurosci.* **29**, 15595–15600 (2009).
13. N Tomen, JM Herrmann, U Ernst, eds., *The Functional Role of Critical Dynamics in Neural Systems*, Springer Series on Bio- and Neurosystems. (Springer International Publishing, Cham) Vol. 11, (2019).
14. A Levina, JM Herrmann, T Geisel, Dynamical synapses causing self-organized criticality in neural networks. *Nat Phys* **3**, 857–860 (2007).
15. J Timcheck, J Kadmon, K Boahen, S Ganguli, Optimal noise level for coding with tightly balanced networks of spiking neurons in the presence of transmission delays. *PLOS Comput. Biol.* **18**, e1010593 (2022).
16. S Khajehabdollahi, J Prosi, E Giannakakis, G Martius, A Levina, When to Be Critical? Performance and Evolvability in Different Regimes of Neural Ising Agents. *Artif. Life* pp. 1–21 (2022).

222 17. R Zeraati, et al., Attentional modulation of intrinsic timescales in visual cortex and spatial
223 networks (2021).
224

¹ **Supplementary Information for**

² **Signatures of criticality in efficient coding networks**

³ **Shervin Safavi, Matthew Chalk, Nikos Logothetis, Anna Levina**

⁴ **Corresponding Author name.**

⁵ **E-mail: research@shervinsafavi.org, anna.levina@uni-tuebingen.de, matthew.chalk@inserm.fr**

⁶ **This PDF file includes:**

⁷ Supplementary text

⁸ SI References

9 **Supporting Information Text**

10 **Materials and methods**

11 **Efficient coding network.** The neuronal network model used in this study was introduced and described extensively in the
 12 previous studies (1, 2), thus we restrict ourselves to a brief explanation of the key aspect of the model. Our network can be
 13 optimized to encode a sensory input efficiently (i.e., with a minimal number of spikes) and accurately (i.e., with minimal
 14 reconstruction error). Network optimization objective is incorporated in the loss function $E(t)$,

$$15 \quad E(t) = (x(t) - \hat{x}(t))^2 + \alpha \sum_i r_i(t) + \beta \sum_i r_i(t)^2, \quad [1]$$

16 where $x(t)$ is a given one-dimensional sensory input (similar to 2, 3), $\hat{x}(t)$ is the reconstructed sensory input, $r_i(t)$ is the firing
 17 rate of the neuron i , and α and β are the weights of the $L1$ and $L2$ penalties on the firing rate.

18 It is assumed that the input can be reconstructed by performing a linear readout of the spike trains, more precisely, by a
 19 weighted leaky integration of output spike trains,

$$20 \quad \tau \frac{d\hat{x}(t)}{dt} = -\hat{x}(t) + \sum_i w_i o_i(t), \quad [2]$$

21 where o_i indicates the output spike trains for the neuron i ,

$$22 \quad o_i(t) = \sum_k \delta(t - t_i^k), \quad [3]$$

23 and τ is the read-out time constant*, and w_i is a constant read-out weight associated to the neuron i .

24 Given an idealized network with instantaneous synapses, the optimal network could be derived from first principles. Boerlin
 25 et al. (1) demonstrated that the dynamics of each leaky-integrate and fire (LIF) neuron can be expressed by conventional
 26 differential equation governing the dynamics of the membrane potentials,

$$27 \quad \tau \frac{dV_i(t)}{dt} = -V_i(t) + w_i c(t) - w_i \sum_k w_k o_k(t) - \beta o_i(t) + \sigma \nu_i(t), \quad [4]$$

28 where V_i is the membrane potential of the neuron i , w_i is the constant readout which was introduced in Equation 2, $c(t)$ is the
 29 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
 30 white noise with unit variance that was manually added in the original derivation of (1) for biological realism. Notably, in this
 31 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 input
 32 is the result of a fully connected network. In this network, neurons that receive a common input, decorrelate their activity to
 33 avoid communicating redundant information via instantaneous recurrent inhibition.

34 Chalk et al. (2) introduced a more biologically plausible variants of (1)'s network by incorporating synaptic delays and
 35 introducing a balance network of inhibitory and excitatory population of neurons. They incorporated realistic synaptic delays
 36 by assuming that each spike generates a continuous current input to other neurons, with a dynamic that is described by the
 37 conventional alpha function,

$$38 \quad h(t) = \begin{cases} \frac{1}{\tau_d - \tau_r} \left[e^{\frac{-(t-\tau_{tr})}{\tau_d}} - e^{\frac{-(t-\tau_{tr})}{\tau_r}} \right] & \text{if } t > \tau_{tr} \\ 0 & \text{if } t < \tau_{tr} \end{cases} \quad [5]$$

39 where τ_r and τ_d are respectively synaptic rise and decay times. Adding realistic synaptic delays, led to network synchronization,
 40 which impairs coding efficiency. Chalk et al. (2) demonstrated that, in the presence of synaptic delays, this network of LIF
 41 neurons can nonetheless be optimized for efficient coding by adding noise to the network. In this study, we implement the
 42 additional noise, as white noise added to the membrane potentials. However, (2) also demonstrated similar dependency of
 43 network's performance to noise by using other ways of incorporating noise, for instance, by inducing unreliability in spike
 44 elicitation (also see, 4–6, for other approaches).

45 The original network introduced by (1) was a pure inhibitory network. (2) introduced a variant of this network that respects
 46 the Dale's law. In their network, they introduce a population of inhibitory neurons that tracks the estimate encoded by the
 47 excitatory neurons, and provides recurrent feedback to the excitatory population (for further detail see, 1, 2)

48 **Avalanche detection.** To investigate the scale-free characteristic of the spiking activity (as a potential signature of networks
 49 operating close to criticality), similar to previous studies (7), we probe the distribution of neural avalanches. A neuronal
 50 avalanche is defined as an uninterrupted cascade of spikes in the network (7). In a system operating close to criticality, the
 51 distribution of avalanche sizes (number of spikes in a cascade) and avalanche life-time follows a power-law (in this study we
 52 have only investigated the distribution of avalanche sizes).

*In the efficient coding network used in this study (as in 2), 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 (1) they are not necessarily the same for more general computations.

53 For detecting the avalanches, we followed the procedure used in previous studies (e.g., 7). The period of no spiking activity
54 signifies the end of the previous avalanche if it is larger than a threshold Δ . We mainly used as a threshold the mean inter-spike
55 interval ($\Delta = \langle ISI \rangle$) in the compound spike train (obtained by collapsing the spike trains of all neurons onto a single time-line).
56 Thus, when the compound spike train is interrupted for an interval larger than Δ , we consider that the current avalanche is
57 over, and a next spike will be an onset of a new avalanche. The size of the avalanche is the number of spikes between these two
58 silent time-points. A slightly different procedure has also been used for avalanche detection. In the alternative approach, for
59 computing the Δ , one counts the synchronous spike only once, i.e., excluding zero $ISIs$. Notably, similar results were obtained
60 using the alternative method.

61 This choice of threshold Δ potentially can be made separately for individual network with different noise levels. However, to
62 avoid introducing an additional variability across different levels of noise, we fixed the threshold for all the noise levels. For a
63 given network size, we took the threshold from the network with the noise level corresponding to minimal mean-square-error
64 (MSE), and use that for all noise levels. We also checked that taking slightly different thresholds would not change the results
65 of our study.

66 **Closeness to criticality assessment.** We consider the scale-free distribution of neural avalanche as signature of criticality in
67 the network (8, 9). Thus, to determine the most scale-free avalanche distribution, we introduce a deviation measure κ , which
68 quantifies deviation from an ideal power-law distribution. Our κ measure closely follows the non-parametric measure introduced
69 by Shew *et al.* (10), but does not assume a particular scaling exponent, which might be important, because we do not know
70 *a priori* what is the relevant universality class for neuronal avalanches (see, e.g., 11). We define κ as the area between the
71 empirical and the ideal (fitted power-law) distribution, normalized by the number of data-points (in the empirical distribution)
72 between left and right cut-offs. Larger values of κ correspond to larger deviations from the power-law distribution. Power-laws
73 were fitted between two cut-offs. The left cut-off was always chosen to be 2 (i.e., avalanches with at least two spikes). The right
74 cut-off is typically chosen subjectively based on the problem at hand (12), here we swept over a wide range of choices to be as
75 objective as possible. We choose it between two possible options: it was either a certain percentile of number of avalanches
76 (within the range of 50–95%), or a fraction of the network size (within the range of 10–25%). Between the mentioned choices
77 above, the one led to inclusion of more data, i.e., larger proportion of avalanches was selected. The ideal power-law distribution
78 were also determined based on a linear fit between final choices of left and right cut-offs. Notably, the results were robust to
79 variations (in the ranges noted above) in the choice of cut-offs (see Figure 2C of the main text).

80 References

- 81 1. M Boerlin, CK Machens, S Deneve, Predictive coding of dynamical variables in balanced spiking networks. *PLoS computational biology* **9**, e1003258 (2013).
- 82 2. M Chalk, B Gutkin, S Deneve, Neural oscillations as a signature of efficient coding in the presence of synaptic delays. *eLife* **5** (2016).
- 83 3. J Timcheck, J Kadmon, K Boahen, S Ganguli, Optimal noise level for coding with tightly balanced networks of spiking
84 neurons in the presence of transmission delays. *PLOS Comput. Biol.* **18**, e1010593 (2022).
- 85 4. V Koren, S Deneve, Computational Account of Spontaneous Activity as a Signature of Predictive Coding. *PLoS computational biology* **13**, e1005355 (2017).
- 86 5. CER Buxó, JW Pillow, Poisson balanced spiking networks. *PLOS Comput. Biol.* **16**, e1008261 (2020).
- 87 6. N Calaim, FA Dehmelt, PJ Gonçalves, CK Machens, The geometry of robustness in spiking neural networks. *eLife* **11**,
88 e73276 (2022).
- 89 7. JM Beggs, D Plenz, Neuronal avalanches in neocortical circuits. *The J. neuroscience : official journal Soc. for Neurosci.*
90 **23**, 11167–77 (2003).
- 91 8. D Plenz, et al., Self-Organized Criticality in the Brain. *Front. Phys.* **0** (2021).
- 92 9. GF Grosu, et al., The fractal brain: Scale-invariance in structure and dynamics. *Cereb. Cortex* p. bhac363 (2022).
- 93 10. WL Shew, H Yang, T Petermann, R Roy, D Plenz, Neuronal Avalanches Imply Maximum Dynamic Range in Cortical
94 Networks at Criticality. *J. Neurosci.* **29**, 15595–15600 (2009).
- 95 11. MA Muñoz, R Dickman, A Vespignani, S Zapperi, Avalanche and spreading exponents in systems with absorbing states.
96 *Phys. Rev. E* **59**, 6175–6179 (1999).
- 97 12. J Alstott, E Bullmore, D Plenz, Powerlaw: A Python package for analysis of heavy-tailed distributions. *PLoS ONE* **9**,
98 e85777 (2014).



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 Einhäuser, 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 LPFC 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 Einhäuser, 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
- Scheinberg, 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.

Copyright © 2014 Safavi, Kapoor, Logothetis and Panagiotaropoulos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



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}

^aDepartment Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany; ^bInternational Max Planck Research School for Cognitive and Systems Neuroscience, University of Tübingen, 72074 Tübingen, Germany; ^cDepartment of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637; ^dDivision of Imaging Science and Biomedical Engineering, University of Manchester, 72074 Manchester, United Kingdom; and ^eCognitive Neuroimaging Unit, Commissariat à l'Energie 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.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹S.S. and A.D. contributed equally to this work.

²To 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 $400 \mu\text{m}$, electrode length 1 mm; Fig. 1*A*) to record spiking activity from the inferior convexity of the ventrolateral PFC (vIPFC) 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 \pm SEM) single units and $5,305 \pm 1,681$ pairs during anesthesia (Fig. S1*A*)

and 107 ± 14 single units and $5,758 \pm 1,675$ pairs during wakefulness (Fig. S1B).

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 vIPFC.

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

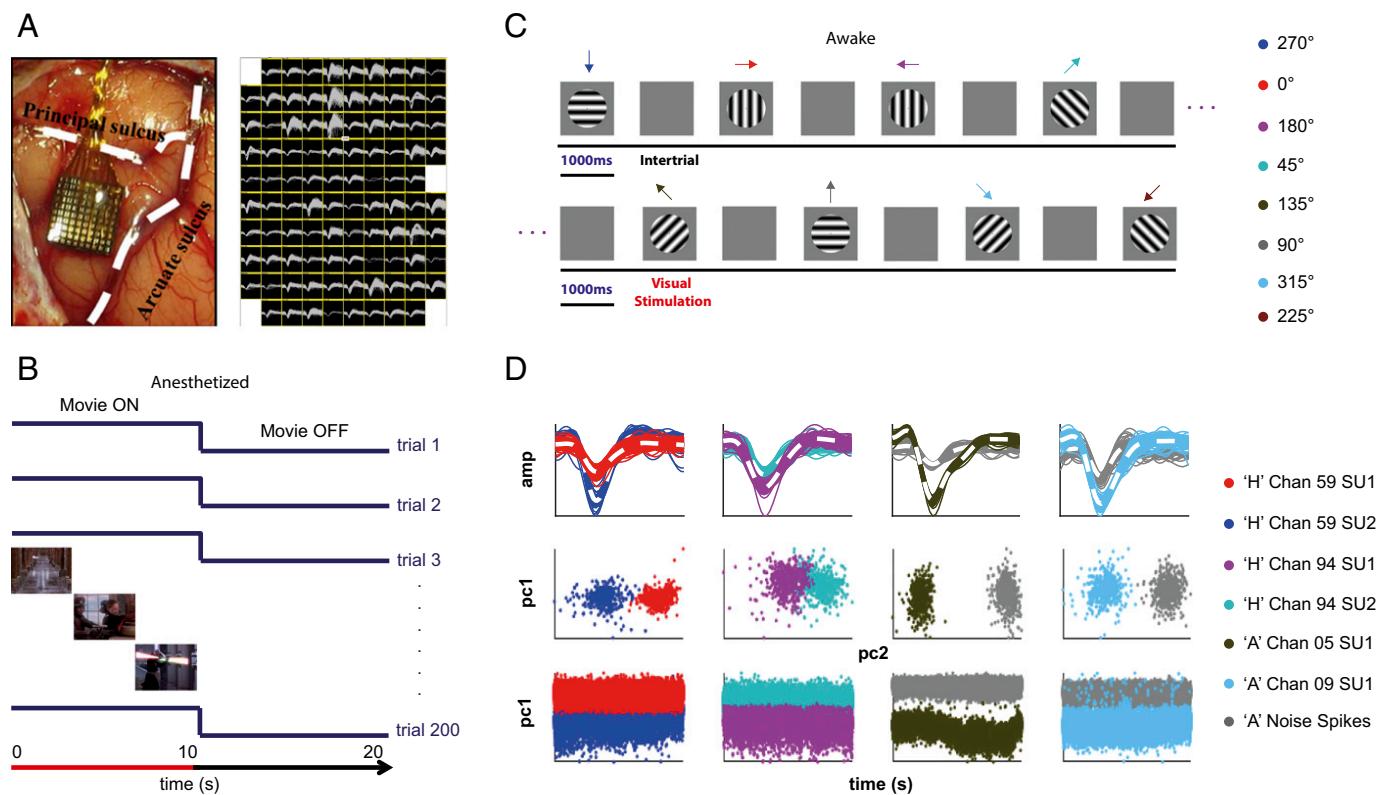


Fig. 1. Implantation, visual stimulation, and quality of single unit isolation. (A) Location of the implanted array with respect to arcuate and principal sulci and an example of typical waveforms acquired across the implanted cortical patch during a typical recording session in an awake animal. (B) Anesthetized 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 (pc_1 and pc_2) 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 $\sim 3.5\text{--}4$ h. amp, amplitude.

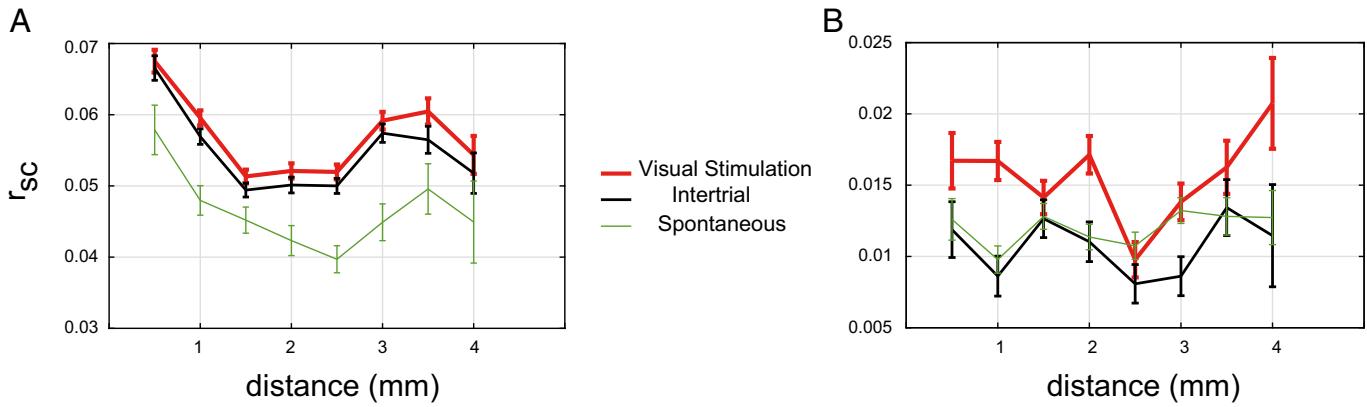


Fig. 2. Spatial structure of correlated variability. (A) Spike count correlations (r_{sc}) during visual stimulation (red), intertrial (black), and spontaneous activity (green) as a function of lateral spatial distance (millimeters) between cell pairs for anesthetized-state recordings (error bars represent mean \pm 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 ($\bar{r}_{sc}^{0.5 \text{ mm}} = 0.0675 \pm 0.0016$ vs. $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0520 \pm 0.0010$; $P < 10^{-10}$; Fig. 2A, red curve, and Fig. S2A) 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 ($\bar{r}_{sc}^{1 \text{ mm}} = 0.0167 \pm 0.0013$ vs. $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0098 \pm 0.0012$; $P = 0.0038$; Fig. 2B, red curve, and Fig. S2D). Following this minimum, correlations during anesthesia significantly increased from 2.5 to 3 mm ($\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0520 \pm 0.0010$ vs. $\bar{r}_{sc}^{3 \text{ mm}} = 0.0592 \pm 0.0012$; $P = 0.012$) and 3.5 mm ($\bar{r}_{sc}^{3.5 \text{ mm}} = 0.0605 \pm 0.0018$; $P = 0.0040$). 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 ($\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0098 \pm 0.0012$ vs. $\bar{r}_{sc}^{3.5 \text{ mm}} = 0.0163 \pm 0.0019$; $P = 0.0076$) and 4 mm ($\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0098 \pm 0.0012$ vs. $\bar{r}_{sc}^{4 \text{ mm}} = 0.0207 \pm 0.0032$; $P = 0.0079$).

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 ($\bar{r}_{sc}^{1 \text{ mm}} = 0.0167 \pm 0.0013$ vs. $\bar{r}_{sc}^{3.5 \text{ mm}} = 0.0163 \pm 0.0019$, $P = 0.7$; and $\bar{r}_{sc}^{4 \text{ mm}} = 0.0207 \pm 0.0032$, $P = 0.3$; Fig. 2B, red curve, and Fig. S2D). 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 ($\bar{r}_{sc}^{0.5 \text{ mm}} = 0.075 \pm 0.0016$ vs. $\bar{r}_{sc}^{3.5 \text{ mm}} = 0.0605 \pm 0.0018$, $P = 0.0056$; and $\bar{r}_{sc}^{4 \text{ mm}} = 0.0544 \pm 0.0027$, $P = 0.0008$; Fig. 2A, red curve, and Fig. S2A), 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. 2B and Fig. S2D). 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 (\bar{r}_{sc}^{anesth}) and awake (\bar{r}_{sc}^{awake}) states were observed despite significant differences in the average magnitude of correlations [$\bar{r}_{sc}^{anesth} = 0.0574 \pm 3 \times 10^{-4}$ (mean \pm SEM) vs. $\bar{r}_{sc}^{awake} = 0.0153 \pm 3 \times 10^{-4}$, $P = 0$; Fig. 3A]. 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 ($\bar{r}_{sc}^{\text{visual}} = 0.0574 \pm 3 \times 10^{-4}$ vs. $\bar{r}_{sc}^{\text{intertrial}} = 0.0554 \pm 3 \times 10^{-4}$, $P < 10^{-3}$; Fig. 3B) and awake recordings ($\bar{r}_{sc}^{\text{visual}} = 0.0153 \pm 3 \times 10^{-4}$ vs. $\bar{r}_{sc}^{\text{intertrial}} = 0.011 \pm 4 \times 10^{-4}$, $P = 6.7 \times 10^{-5}$, Fig. 3C).

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 ($\bar{r}_{sc}^{0.5 \text{ mm}} = 0.0119 \pm 0.0020$ vs. $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0081 \pm 0.0014$, $P > 0.3$; Fig. 2B, black curve, and Fig. S2B) or intermediate and distant populations ($\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0081 \pm 0.0014$ vs. $\bar{r}_{sc}^{3.5 \text{ mm}} = 0.0134 \pm 0.0020$, $P = 0.08$; and vs. $\bar{r}_{sc}^{4 \text{ mm}} = 0.0115 \pm 0.0036$, $P > 0.25$; Fig. 2B, black curve, and Fig. S2B). Spatially homogeneous correlations were also observed during periods without any structured visual input or task engagement, in data collected during spontaneous activity (Fig. 2B, green curve). In these epochs, we also found no difference between local and intermediate correlations ($\bar{r}_{sc}^{0.5 \text{ mm}} = 0.0126 \pm 0.0015$ vs. $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0107 \pm 9.5 \times 10^{-4}$,

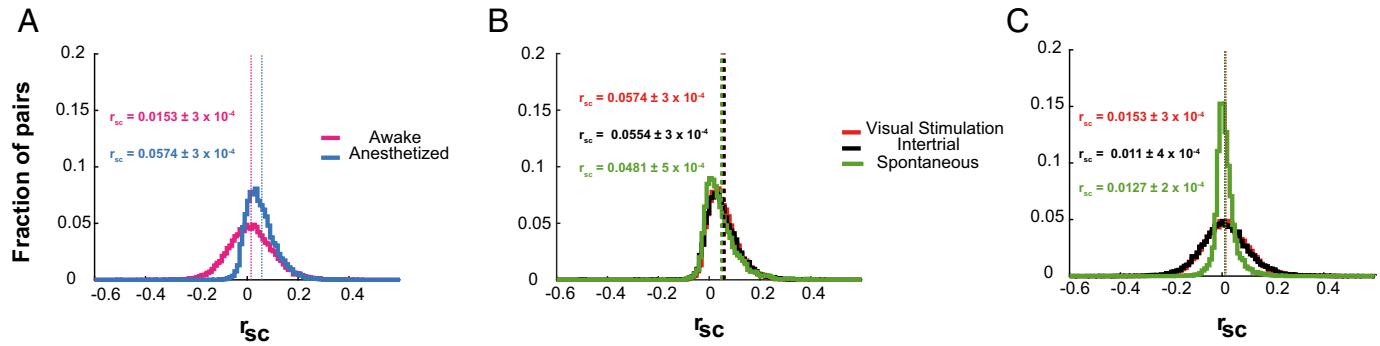


Fig. 3. Distributions of correlated variability across different states and conditions. (A) Distribution of pairwise correlated variability (fraction of pairs) and mean values (dotted lines) during visual stimulation for anesthetized (blue) and awake (pink) recordings. Correlated variability was significantly 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 ($\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0107 \pm 9.5 \times 10^{-4}$ vs. $\bar{r}_{sc}^{3.5 \text{ mm}} = 0.0128 \pm 0.0013$, $P > 0.9$; vs. $\bar{r}_{sc}^{4 \text{ mm}} = 0.0127 \pm 0.0019$, $P > 0.17$; Fig. 2B, 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. 4A), 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. 4B); 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. 6A) 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 ($\bar{r}_{sc}^{0.5 \text{ mm}} = 0.0266 \pm 0.0032$ vs. $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0165 \pm 0.0021$, $P = 0.07$; and $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0165 \pm 0.0021$ vs. $\bar{r}_{sc}^{4 \text{ mm}} = 0.0266 \pm 0.0054$, $P = 0.1$; green curve in Fig. 6B). 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. 7A). 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 ($\sim 2.5 \text{ mm}$) 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 ($\bar{r}_{sc}^{0.5 \text{ mm}} = 0.015 \pm 0.0033$ vs. $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0111 \pm 0.0022$, $P = 0.3$; and $\bar{r}_{sc}^{4 \text{ mm}} = 0.008 \pm 0.0023$ vs. $\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0111 \pm 0.0022$, $P = 0.9$;

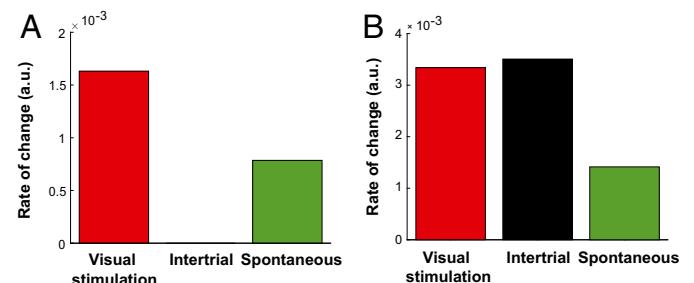


Fig. 4. Quantification of spatial inhomogeneity in the structure of correlated variability. (A) Spatial inhomogeneity in the structure of correlations across different conditions during awake-state recordings. Spatial inhomogeneity was quantified by computing the average of the absolute rate of change 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.

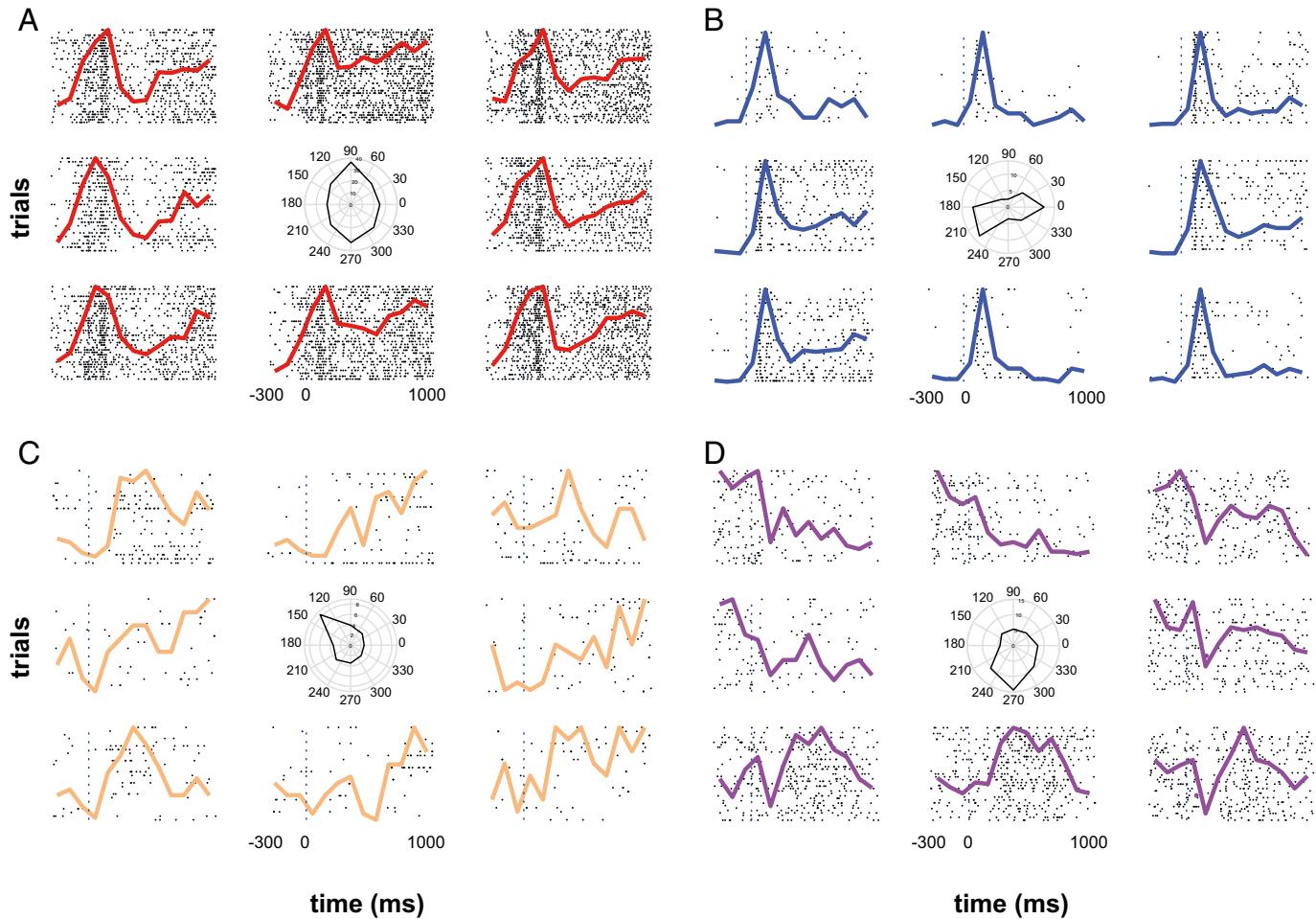


Fig. 5. Visual modulation of single unit activity during wakefulness. Spike raster plots and overlaid peristimulus time histograms (PSTHs) for four single units across all eight orientations ($0\text{--}315^\circ$) are shown. Polar plots for each unit show the preferred direction(s) of motion. The green 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 ($\bar{r}_{sc}^{2.5 \text{ mm}} = 0.0111 \pm 0.0022$ vs. $\bar{r}_{sc}^{3.5 \text{ mm}} = 0.0077 \pm 0.0034, P = 0.8$; and vs. $\bar{r}_{sc}^{4 \text{ mm}} = 0.0057 \pm 0.0068, P = 0.9$). 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 ($0.5\text{--}1 \text{ mm}$) and distant ($3.5\text{--}4 \text{ mm}$) 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 ($P_{VisStim \text{ vs. } Intertial}^{0.5\text{--}1 \text{ mm}} = 0.00024; P_{VisStim \text{ vs. } Intertial}^{3.5\text{--}4 \text{ mm}} = 0.017; P_{VisStim \text{ vs. } Intertial}^{2.5 \text{ mm}} = 0.14$). 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 ($P_{VisStim \text{ vs. } Intertial}^{0.5\text{--}1 \text{ mm}} = 0.64; P_{VisStim \text{ vs. } Intertial}^{3.5\text{--}4 \text{ mm}} = 0.7$). 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

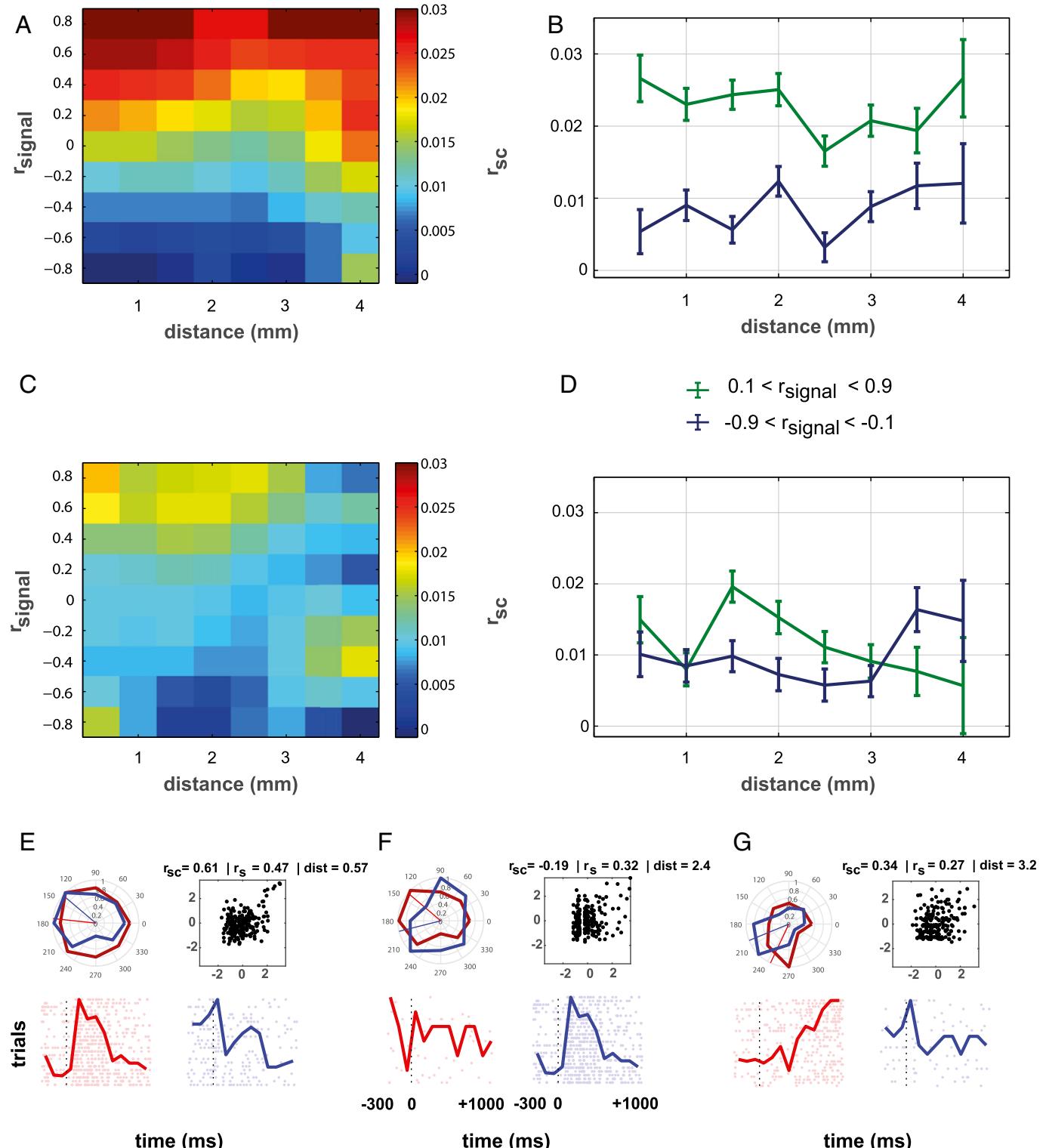


Fig. 6. Effect of functional similarity on the spatial structure of correlated variability. (A) Correlated variability as a function of distance and signal correlation for the pooled data recorded during visual stimulation in the awake state. The color of each pixel indicates the average correlated variability 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 \pm 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).

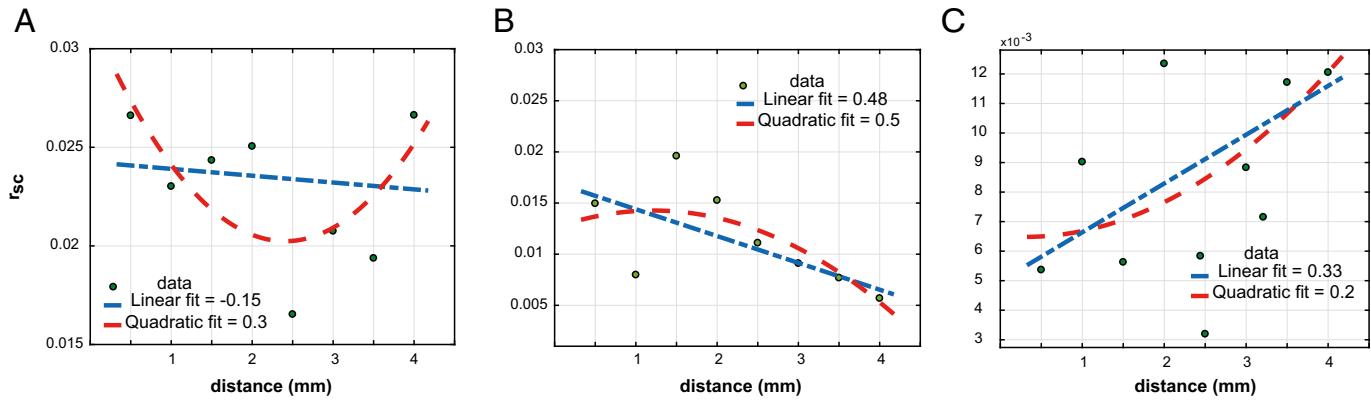


Fig. 7. Fitting trends to the relationship between noise correlations and distance for functionally similar populations. (A) Linear ($y = ax + b$) and quadratic ($y = ax^2 + bx + c$) trends fit to noise correlations as a function of distance for functionally similar neurons ($0.1 < r_{signal} < 0.9$) 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. 2B), 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., vIPFC, since none of these studies involved recordings in the vIPFC, 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 vIPFC 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 vIPFC is fundamentally different compared with early sensory cortices such as V1 or V4. Correlated variability in the vIPFC is spatially nonmonotonic, and a major source of non-monotonicity 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 $400 \mu\text{m}$) 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 vIPFC (Fig. 1A). 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 47/40; 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 $1.5\text{--}2^\circ$ 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. $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$, and 315° (Fig. 1B), 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 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 8 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, Nadasy 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, Buzsaki 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, Buzsaki 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.

ARTICLE



<https://doi.org/10.1038/s41467-022-28897-2>

OPEN

Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports

Vishal Kapoor^{1,2,7✉}, Abhilash Dwarakanath^{1,7}, Shervin Safavi^{1,3}, Joachim Werner¹, Michel Besserve^{1,4}, Theofanis I. Panagiotaropoulos^{1,5,8✉} & Nikos K. Logothetis^{1,2,6,8}

A major debate about the neural correlates of conscious perception concerns its cortical organization, namely, whether it includes the prefrontal cortex (PFC), which mediates executive functions, or it is constrained within posterior cortices. It has been suggested that PFC activity during paradigms investigating conscious perception is conflated with post-perceptual processes associated with reporting the contents of consciousness or feedforward signals originating from exogenous stimulus manipulations and relayed via posterior cortical areas. We addressed this debate by simultaneously probing neuronal populations in the rhesus macaque (*Macaca mulatta*) PFC during a no-report paradigm, capable of instigating internally generated transitions in conscious perception, without changes in visual stimulation. We find that feature-selective prefrontal neurons are modulated concomitantly with subjective perception and perceptual suppression of their preferred stimulus during both externally induced and internally generated changes in conscious perception. Importantly, this enables reliable single-trial, population decoding of conscious contents. Control experiments confirm significant decoding of stimulus contents, even when oculomotor responses, used for inferring perception, are suppressed. These findings suggest that internally generated changes in the contents of conscious visual perception are reliably reflected within the activity of prefrontal populations in the absence of volitional reports or changes in sensory input.

¹Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany. ²International Center for Primate Brain Research, Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Institute of Neuroscience (ION), Chinese Academy of Sciences, Shanghai, China. ³International Max Planck Research School, Tübingen 72076, Germany. ⁴Department of Empirical Inference, Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany. ⁵Cognitive Neuroimaging Unit, CEA, DSV/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France. ⁶Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK. ⁷These authors contributed equally: Vishal Kapoor, Abhilash Dwarakanath. ⁸These authors jointly supervised this work: Theofanis I. Panagiotaropoulos, Nikos K. Logothetis. ✉email: vishal.kapoor@tuebingen.mpg.de; theofanis.panagiotaropoulos@tuebingen.mpg.de

One of the most elusive problems in science is to understand the biological basis of consciousness¹. 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 within the reach of neuroscience². Since then, several theoretical treatises^{3–5}, including the frontal lobe hypothesis³, higher order (HOT)⁵ and global neuronal workspace (GNW) theories^{4,6}, postulate the prefrontal cortex (PFC) as a critical node in mediating conscious perception. Evidence supporting its involvement comes from functional magnetic resonance imaging (fMRI)^{6–8}, experience of visual hallucinations upon its electrical stimulation^{9,10}, impaired conscious processing following PFC lesions^{11–15} and intracortical recordings^{16–20}.

Alternative approaches such as the integrated information (IIT) or the recurrent processing (RPT) theories diverge on the contribution of PFC to conscious perception^{21,22}. Their proponents suggest that activity in posterior cortical areas reflects conscious contents, and PFC is primarily responsible for its executive role of processing the behavioral and cognitive consequences of conscious perception like task demands, introspection about the perceived stimuli and motor reports of perception^{23–27}. This is corroborated with functional magnetic resonance imaging (fMRI) studies utilizing paradigms where either no volitional report is required or perceptual changes are indiscernible and therefore unreportable^{23,26,28}. These studies suggest that consequences of perception such as decision making or motor action drive frontal activity, thus casting doubt if PFC represents conscious content^{8,11,25,27}. However, the univariate contrastive fMRI analysis, as well as the indirect nature and limited spatial resolution of BOLD fMRI signal compared to direct neuronal recordings²⁹ leaves open the possibility that prefrontal ensembles could reflect the content of consciousness even without report requirements or other conflating variables. Importantly, this cannot be firmly established from previous electrophysiological studies probing the PFC during conscious perception since they either leveraged an external manipulation of stimulus presentation, which affected the feedforward sensory drive¹⁷ or utilized motor reports^{16,18–20}. These variables could conflate the signals intrinsically related to conscious perception (also see Supplementary Table 1). For example, it is possible that prefrontal activity during conscious perception paradigms utilizing external changes in sensory input reflects feedforward signals relayed via the posterior cortical areas¹⁷. In addition, if motor reports are required to indicate conscious perception, it is plausible that prefrontal signals could reflect the decision about the specific motor report that is associated with a particular visual stimulus^{16,18–20}. Therefore, ascertaining the presence of conscious content related signals in the PFC, in the absence of these variables, is necessary for a resolution of the aforementioned debate and crucial for understanding whether PFC is a part of the cortical network underlying conscious perception.

We examined this by simultaneously probing the activity of neuronal populations in the macaque ventrolateral PFC with multi-electrode arrays during a no-report binocular rivalry (BR) paradigm. BR is a phenomenon, wherein presentation of incongruent, dichoptic visual input to corresponding retinal locations results in stochastic, endogenously driven alternations in subjective perception of the presented stimuli. It therefore allows a dissociation of conscious perception from sensory input, since the contents of consciousness fluctuate without a change in sensory stimulation^{30,31}. Typical BR experiments require perceptual reports from subjects, thus conflating the neural activity related to consciousness with signals related to its consequences such as volitional reports, decision making and introspection^{23,32–34}. Objective indicators of perception like the optokinetic nystagmus

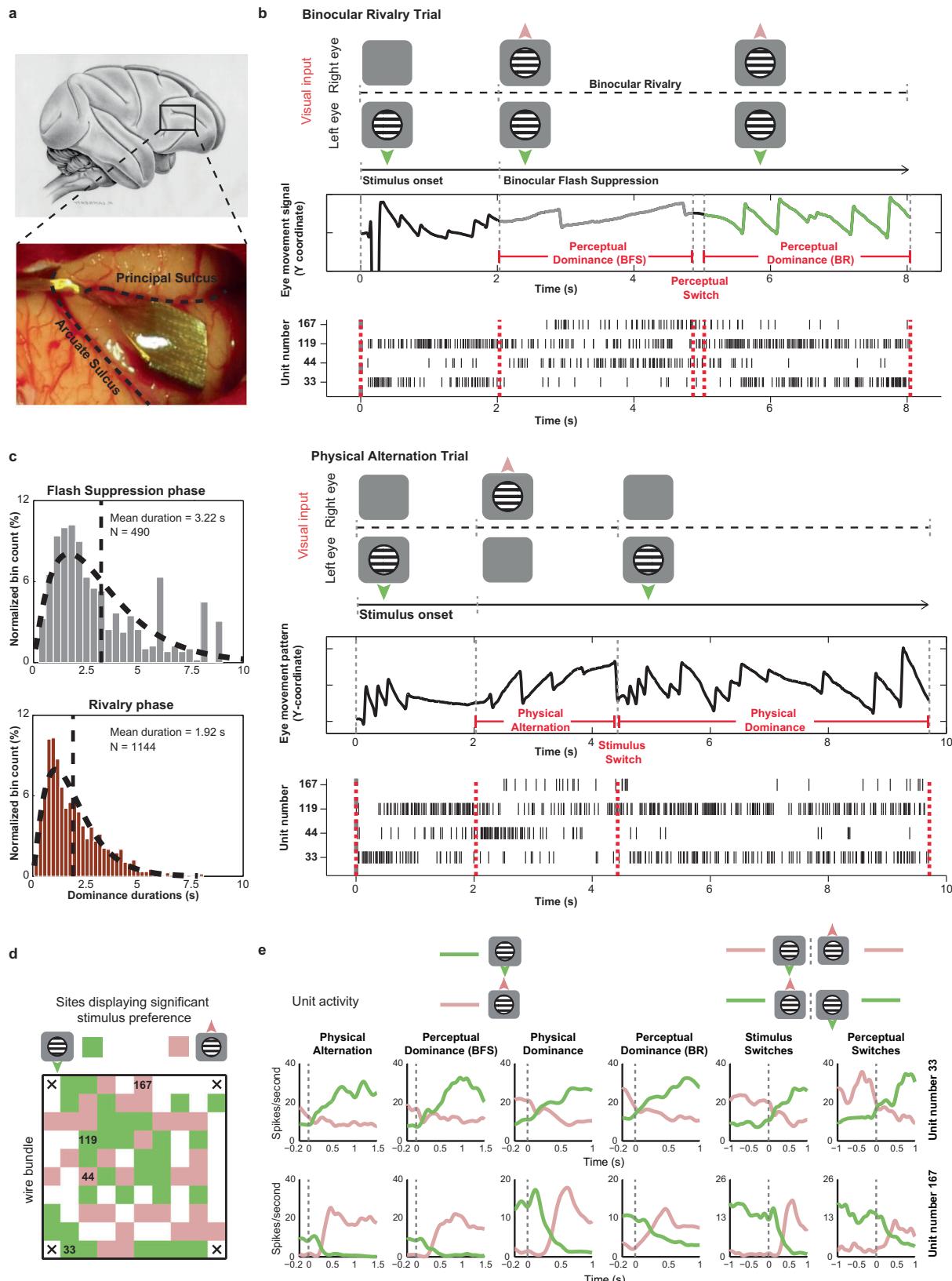
(OKN) reflex, a combination of smooth pursuit and fast saccadic eye movements, provide a solution to this problem. OKN polarity correlates with the perceptual reports during BR elicited with stimuli containing opposing motion content^{35,36}. Thus, exploiting OKN as a surrogate measure to infer conscious perception can address confounds in the neural activity originating from motor reports. Importantly, utilizing such a stimulus set allows scrutinizing, if PFC represents simple low-level features such as direction of motion, when they are consciously perceived. To this end, we probed prefrontal ensembles in rhesus macaques (which were not trained for reporting their perception during a behavioral task) while they experienced no-report BR to investigate representations of conscious content in the PFC. Therefore, a central aim of the present work was to unravel the report-free prefrontal correlates of conscious visual perception during internally generated changes in the contents of consciousness.

Here we show that stimulus selective, prefrontal activity is modulated in concordance with subjective perception, during both externally induced and internally generated changes in conscious perception. Importantly, we reliably decoded the contents of conscious perception from prefrontal activity in the absence of subjective reports. These findings suggest a decodable representation of the ongoing conscious percept within an executive brain region such as the PFC.

Results

No-report BR paradigm and the temporal dynamics of perception. The no-report BR paradigm consisted of two trial types, physical alternation (PA, wherein perceptual changes were induced externally without visual competition) and binocular rivalry (BR, where perceptual changes could also occur endogenously because of visual competition, see Fig. 1b and Supplementary Fig. 1 and methods). Each trial started with a fixation spot cueing the animal to initiate fixation, lasting ~300 milliseconds, followed by monocular presentation of a stimulus (upward or downward drifting motion) for 1 or 2 sec. After this, an oppositely drifting stimulus (grating or random dots) was presented to the contralateral eye during BR trials, typically inducing perceptual suppression of the first stimulus, a phenomenon termed as binocular flash suppression (BFS)^{37,38}. BFS, typically resulted in a switch of the OKN polarity (Fig. 1b), indicating perception of the newly presented direction of motion (Fig. 1b, marked in grey). Following this, visual competition ensued, resulting in spontaneous perceptual switches between the two stimuli as inferred from the OKN polarity. We evaluated the onset and offset of perceptual dominance periods during every trial based on the stability of the OKN pattern (see methods) during BFS and BR. Perceptual dominance durations displayed a gamma distribution (Fig. 1c), a key characteristic of multistable perception dynamics³⁹, which has been previously established for both human and monkey observers that reported their perception⁴⁰.

Feature selective PFC activity during changes in conscious perception. We targeted the inferior convexity of the PFC (Fig. 1a and methods), where neurons display selective responses to complex visual stimuli and direction of motion^{41–43}. Neural representations of conscious content are assessed with such feature selectivity. If neurons or units (see methods and see below) during BR reliably increase and decrease their firing rate contemporaneously with the perceptual dominance and suppression of their preferred stimulus respectively, then they are considered perceptually modulated and provide an explicit representation of conscious content⁴⁰. We spike sorted and identified putative single units in our recordings. The spiking waveforms, recorded from a given electrode, which could not be sorted to a given single



unit, were collected and denoted as a multi-unit. The valid spiking activity recorded from a given electrode together is referred to as a site. For a majority of analyses presented in this study (unless stated otherwise), we utilized the sorted single units and multi-units and they are collectively referred to as units. After

sorting, we identified 987 units across two animals (734 units from H07, and 253 units from A11). We would like to add that the chronic nature of recordings might lead to an overlap of units across different recording sessions. At the same time, our approach is similar to other recent studies, wherein recording

Fig. 1 Recorded area, BR paradigm and example unit activity. **a**, Utah array location displayed over the vIPFC (schematic macaque brain and post implantation location in one animal). **b**, Visual input, OKN and spiking activity during an example BR and PA trial. A drifting sinusoidal grating (downward) was monocularly presented first during both trials. During BR, an upward drifting grating was presented to the contralateral eye 2000 ms later, resulting in perceptual suppression (BFS) of downward motion, as inferred from the OKN (grey curve). Externally induced perceptual suppression lasted ~3000 ms, following which a spontaneous switch (~5000 ms) reinstated the perception of downward motion (green curve). Units 33 and 119 display strong spiking activity when downward drifting grating is perceptually dominant, while units 44 and 167 respond more when upward drifting grating is perceived. Below is a PA trial. Following initial monocular presentation of downward drifting grating, an upward drifting grating was presented to the contralateral eye. The stimulus was switched later (~4500 ms) resulting in a change in OKN polarity. Individual units were preferentially modulated by similar direction of motion as during BR. **c**, Perceptual dominance distributions during flash suppression and rivalry phases (derived from the OKN traces) are approximated well by a gamma distribution. **d**, Sites with significant stimulus preference (Wilcoxon rank-sum test, two-sided, $p \leq 0.05$) during PA trials (physical alternation phase) projected on the array for one recording session (selectivity here was computed using spiking activity recorded from a given electrode). Location of units (identified after spike sorting) presented in b are marked. Green and pink pixels reflect sites, where activity was stronger for downward or upward drifting gratings, respectively. **e**, Average spike density functions of two simultaneously recorded PFC units, 33 (preferred downward motion) and 167 (preferred upward motion), during PA and BR trials. Pink and green colors in the first four columns depict responses pertaining to downward and upward drifting grating respectively. The last two columns display the activity during a stimulus or perceptual switch from downward to an upward drifting grating (pink) and vice versa (green). Source data are provided as a Source Data file.

sessions (with chronically implanted arrays) on separate days are treated as independent^{44–46}. We further note that we also carried out the decoding analysis (presented later) on individual recording sessions, which alleviates the aforementioned concern.

Figure 1b displays discharges of four such, simultaneously recorded prefrontal units and OKN during one PA and one BR trial. While two units fired more when the downward drifting grating was presented, another two displayed stronger modulation for the opposite motion direction during a PA trial (Fig. 1b). Spiking activity of these units was similarly modulated with subjective changes in conscious perception during a BR trial, for both externally induced perceptual suppression (BFS) and a subsequent spontaneous switch in conscious content (BR) (Fig. 1b).

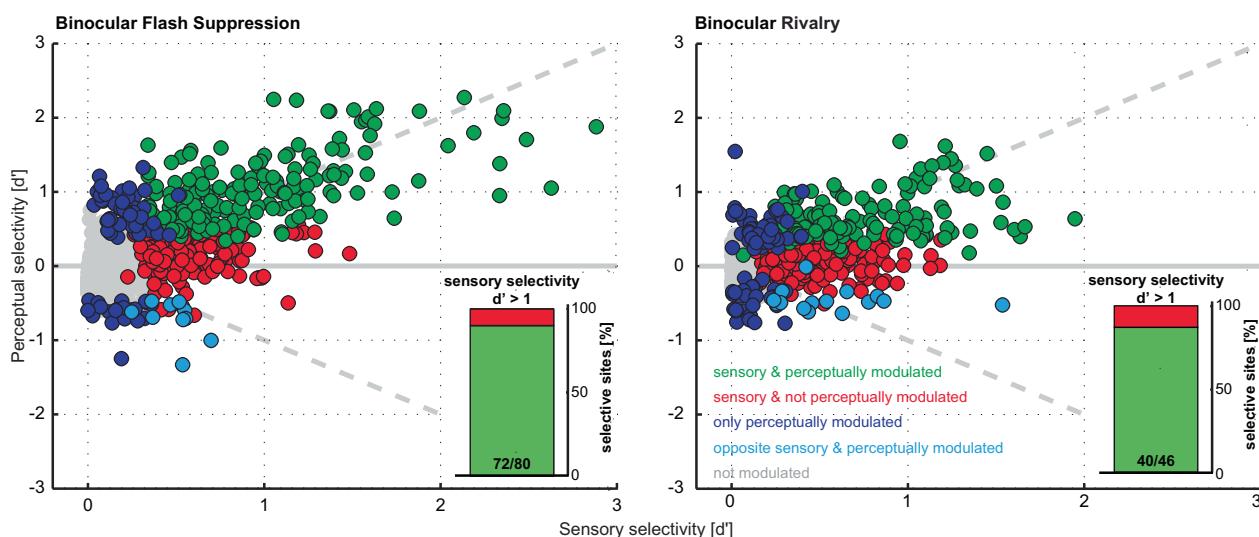
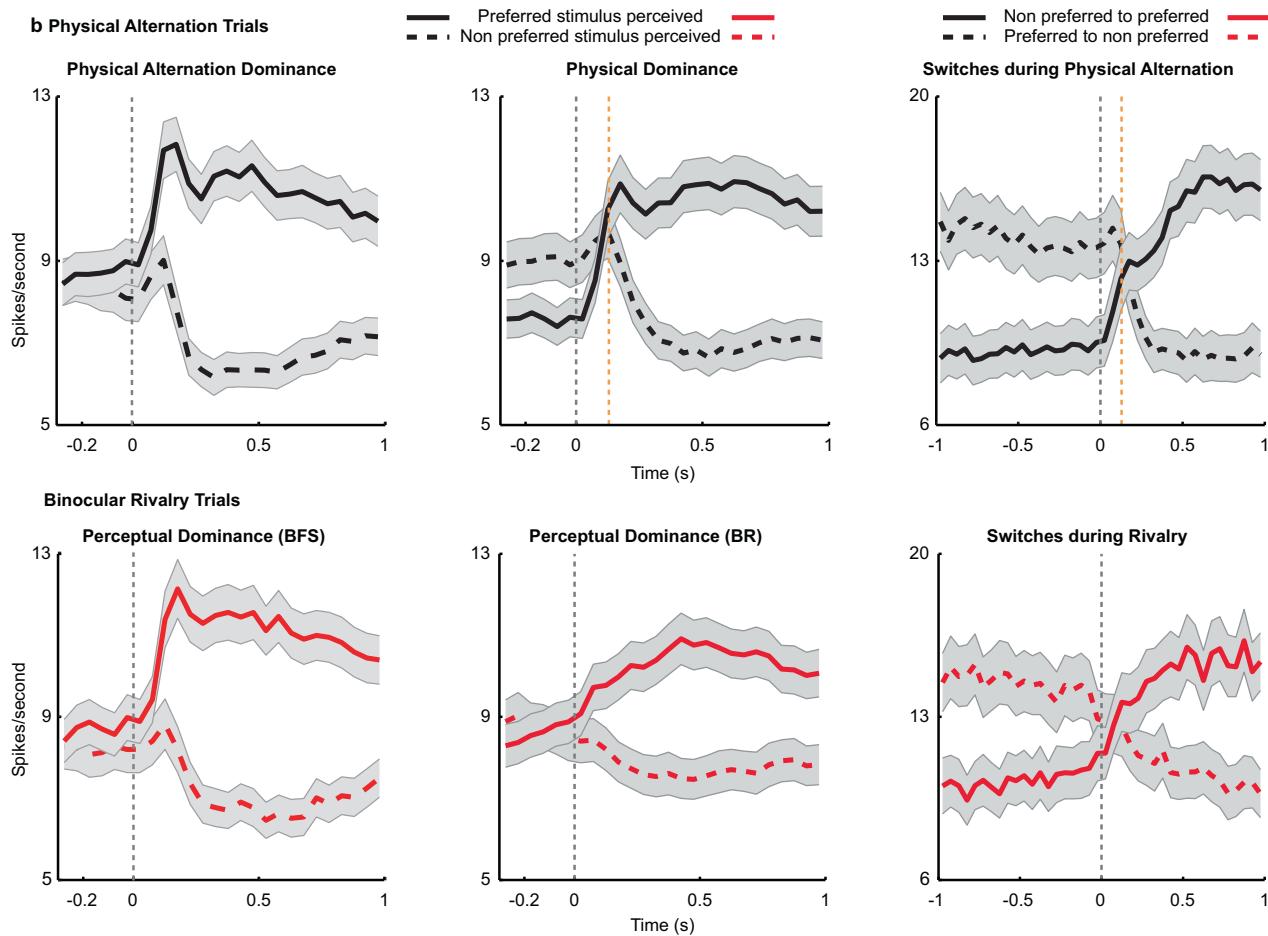
We analyzed the spiking activity of the recorded units separately during perceptual dominance and suppression periods either (i) externally induced during BFS, or (ii) brought about by an endogenous spontaneous switch in BR. Furthermore, selectivity of neural activity was analyzed both before and after such perceptual switches and in all aforementioned cases, compared with selectivity in corresponding temporal phases during PA trials (see methods). We found sites displaying similar as well as opposite stimulus preference (as judged from their underlying spiking activity) distributed in close proximity throughout the electrode array (Fig. 1d and Supplementary Fig. 2a). Individual units recorded from these sites displayed robust perceptual modulation in their average response (examples in Fig. 1e and Supplementary Fig. 2).

Strength of unit selectivity and population responses. We estimated the selectivity of the recorded units (see methods) during sensory, monocular presentations in PA trials and compared it with their selectivity during subjective perception in BFS and BR (see methods). The majority of stimulus selective units (as judged from PA trials) fired on average more when their preferred stimulus was consciously perceived compared to its perceptual suppression (in BR trials). Specifically, during the BFS and BR phases of the paradigm, 84.21% (288/342) (H07: 87.72% (250/285); A11: 66.67% (38/57)) and 76.09% (277/364) (H07: 75.52% (216/286); A11: 78.21% (61/78)) of stimulus selective units, respectively displayed on average stronger activity during the perceptual dominance of their preferred stimulus (with 53.8% (184/342) (H07: 60.35% (172/285); A11: 21.05% (12/57)) and 40.38% (147/364) (H07: 42.31% (121/286); A11: 33.33% (26/78)) being significantly modulated, Wilcoxon rank-sum test, $p \leq 0.05$; also see Supplementary Table 2 and Fig. 2a). This result indicates that conscious content is robustly encoded in PFC. While many units fire more during perceptual suppression of their preferred

stimulus in earlier visual regions (proposed as a part of an inhibitory mechanism independent of the mechanisms of perception)³⁰; such units were a minority in PFC (BFS-2.92% (10/342) (H07: 2.81% (8/285); A11: 3.51% (2/57)) and BR-4.12% (15/364) (H07: 4.89% (14/286); A11: 1.28% (1/78))). Moreover, several units displayed significant preference only during BFS (26.51%, 70/264) (H07: 24.05% (57/237); A11: 48.15% (13/27)) and BR (34.41%, 85/247) (H07: 35.71% (75/210); A11: 27.03% (10/37)), indicating a separate population, whose activity discriminates across the subjective perception of the two visual stimuli reliably, only during visual competition. The pattern of results were largely comparable across animals.

In general, units displayed considerable heterogeneity in stimulus preference strength estimated using a d' index (see methods) during the temporal phases of BFS (d' -physical alternation = 0.3985 ± 0.0131 (Mean \pm SEM), d' -perceptual dominance = 0.4471 ± 0.0133) and BR (d' -physical dominance = 0.3075 ± 0.0098 , d' -perceptual dominance = 0.2719 ± 0.0082). The magnitude of sensory selectivity (d') was a critical factor determining perceptual modulation (Fig. 2a) since almost 90% of units displaying strong sensory selectivity ($d' > 1$), fired more during conscious perception of their preferred stimulus during BR trials (Wilcoxon rank-sum test, $p \leq 0.05$) (90% (72/80) for BFS and 86.96% (40/46) for BR). These proportions are remarkably similar to those observed in the temporal lobe of macaques participating in report based BR⁴⁷, thus indicating that neuronal activity in two cortical regions reflects conscious content. Further, these results suggest that PFC activity correlates with externally and internally driven switches in subjective conscious perception of simple visual features like direction of motion, in addition to the externally induced perceptual suppression of faces and more complex stimuli¹⁷.

Next, we computed the prefrontal population spiking activity by averaging the mean activity of all units that displayed significant modulation (in PA or BR) and similar preference across temporally corresponding phases of PA and BR trials (see methods). Similar to their response during monocular presentations (PA trials), the average population response during both BFS and BR phases displayed stronger activity during conscious perception of a preferred stimulus and a dramatic reduction, when the same stimulus was perceptually suppressed (Fig. 2b, upper and lower row). Interestingly, the strong transient response observed during BFS after the externally induced perceptual dominance was absent in BR, after the spontaneously induced perceptual transitions. Since reliable perceptual modulation was detected in both BFS and BR during subjective perception, this feedforward response component is unlikely to be the main

a Sensory versus Perceptual modulation of spiking activity - [d']**b Physical Alteration Trials**

source of conscious perception representations in the PFC. However, we also note that the lack of a transient response during BR could possibly also result from a jitter of the perceptual changes relative to OKN changes across trials.

Moreover, neural activity switched reliably around both physical stimulus and perceptual switches (Fig. 2b). Similar

results were obtained when the population activity was computed by averaging the responses of the units, selected based upon their significant modulation in either PA or BR trials (Supplementary Fig. 3). Additionally, when neural activity during PA was aligned to OKN changes (see methods), as was done for BR trials, we obtained similar results (Supplementary Fig. 4 and 5 and

Fig. 2 Sensory (PA) versus perceptual (BR) modulation of spiking activity - d' and average population activity. **a**, Scatter plot of sensory vs. perceptual selectivity (d') for all units (dots) during BFS and BR. Displayed with different colors are units; showing no significant modulation in PA or BR trials (grey); those with significant modulation for the same stimulus during both trial types (green); units displaying significant preference only during PA (red) and units displaying significant modulation only during BR trials (blue) and a small percentage of units which fired more when their preferred stimulus was perceptually suppressed (cyan). Proportion of perceptually modulated units for both BFS and BR increased as a function of sensory selectivity strength (insets showing perceptual modulation for $d' > 1$, BFS - 90% and BR - 86%). **b**, Population activity averaged across units, which were significantly modulated during PA (upper row, presented in black) or BR (lower row, presented in red) trials and preferred the same stimulus, is plotted for perceptual dominance brought about by BFS (left) or during BR (middle), as well as switches (right) during BR. Plotted above is population activity during temporally matched phases in PA. Shaded regions depict standard error of the mean. The orange dashed line indicates the average delay between the physical stimulus transition and the OKN derived transition during PA trials (129.4 ± 36.6 ms following the onset of change in the physical stimulus). A remarkable similarity in population activity across trial types indicates robust perceptual modulation. Source data are provided as a Source Data file.

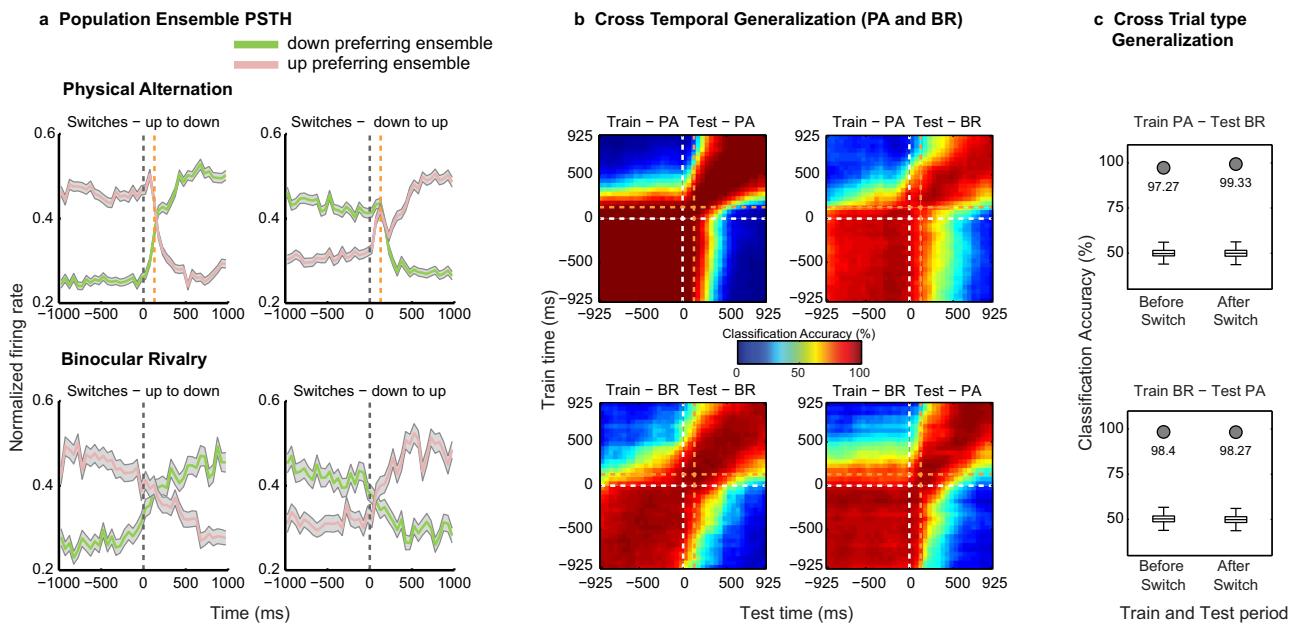


Fig. 3 Decoding the contents of conscious perception from simultaneously recorded prefrontal ensembles. **a**, Normalized population spiking activity of down (green) and up (pink) preferring ensembles during up to down or down to up, PA (upper row) and BR (lower row) switches shows reliable modulation. Data are presented as mean, and shaded regions depict standard error of the mean. **b**, Cross-temporal decoding of conscious contents around switches during PA and BR trials and generalization across the two. Classification accuracy was computed for each pair of train and test time windows (see methods) in steps of 50 ms, using 150 ms bins. In both a and b, the delay between the physical stimulus transition and the OKN derived transition during PA trials is indicated with an orange dashed line (129.4 ± 36.6 ms). **c**, Cross trial-type invariance of the population code assessed by training a classifier on activity during one trial type and testing on the other, before and after a switch for a single 400 ms bin (starting 200 ms pre and post switch). Box plots (for box plot description, see statistical information, methods) depict the distribution of classification results with shuffled labels ($n = 500$), while filled circles denote the highly significant (permutation test, one-sided, estimated p -value: $p = 0.00199$) classification accuracy with real labels. Results suggest invariance of the population code, thus encoding perceptual contents. The presented results were computed with data from two animals pooled together. Similar results were observed for individual animals, which are presented in supplementary figure 6. Source data are provided as a Source Data file.

Supplementary Table 3). Together, these results indicate robust modulation of prefrontal spiking activity during conscious perception.

Decoding the contents of consciousness from prefrontal ensembles. Probing the PFC with multi-electrode arrays allowed us to simultaneously monitor feature specific ensembles, that is, the two groups of units displaying preferential response to each motion direction. Ensemble activity reliably followed both exogenous stimulus changes during PA and endogenous perceptual transitions during BR trials (Fig. 3a). Importantly, this pattern of the population response during the two trial types was observed in prefrontal ensembles recorded in both animals (Supplementary Fig. 6a and d).

Next, we utilized a multivariate decoding approach⁴⁸ to assess the population code's ability to predict contents during single

instances of stimulus and perceptual transitions (see methods). Classifiers trained on neural activity elicited around switches during PA or BR trials discriminated strongly above chance (50%) stimulus or perceptual contents not only within a trial type but crucially also generalized across them with near-perfect accuracy (levels reaching up to 95%, when computed over data collated across both animals, see Fig. 3b and c) revealing a relatively static population code⁴⁹. Such temporal generalization was similarly observed in the population code, when the neural responses recorded in individual animals were subjected to the decoding analysis (Supplementary Fig. 6b, e). Importantly, strong temporal generalization of the classifier trained and tested before and after a switch, across PA and BR when the same stimulus was perceived, indicates an invariance in the population code representing sensory input and its subjective perception (Fig. 3c)⁴⁸. Moreover, such an invariance is crucial to establishing that PFC represents the contents of conscious perception. For

Decoding within individual datasets

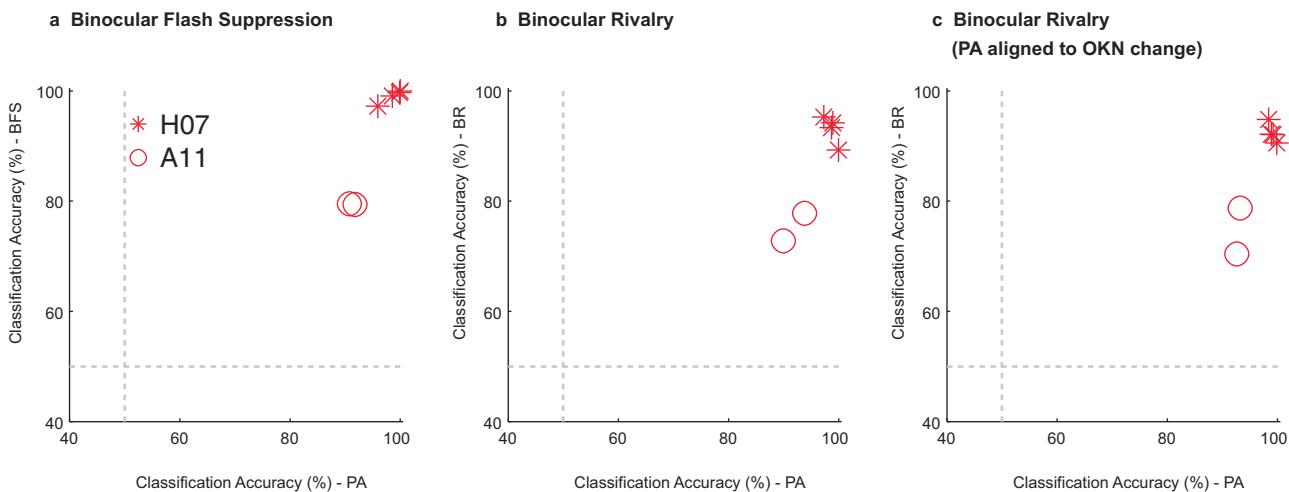


Fig. 4 Assessment of decoding within individual datasets. This figure displays the classification accuracy at the level of individual recording sessions performed across the two animals: four sessions of H07 (red asterisks) and two sessions for A11 (red circles). Each point in the scatter plot corresponds to classification accuracy during PA or BR trials, obtained in a given dataset. The three panels present the results obtained across the different temporal phases of the paradigm: **a**, flash dominance, **b**, rivalry dominance, wherein PA trials are aligned to the physical stimulus change and **c**, rivalry dominance, wherein PA trials were aligned to the change in OKN. The results are computed over an 800 ms window starting 200 ms after the event used for aligning the data (physical stimulus change, or OKN derived change). We note that the classification accuracy for BR in **b** and **c** are similar but not identical, because they were obtained using two separate runs of the decoding analysis. In general, we observed strong classification accuracy even at the level of individual datasets.

instance, if the PFC activity was reflecting the unconscious stimulus, the neural populations would be modulated differently during BR (in comparison to PA), when their preferred stimulus was perceptually suppressed. This would then result in a lack of generalization across the two trial types, PA and BR.

This generalization across trial types during corresponding temporal phases was highly significant (permutation test, $p \leq 0.002$, see methods) when assessed during two temporal windows (400 ms), before (-200 ms to -600 ms) and after (200 ms to 600 ms) a switch (Fig. 3c). Similarly, reliable generalization across the two trial types was observed in both animals (Supplementary Fig. 6c and f). In contrast, training the classifier on activity before and testing after (or training on activity after and testing before) a switch resulted in a dramatic drop in the classification accuracy, given that oppositely drifting stimuli were perceived during the phases compared (Fig. 3b). Therefore, the population activity explicitly encodes the stimulus that enters awareness with minimal representation of the suppressed stimulus. Importantly, a similar overall pattern of results pertaining to the decoding analysis were obtained, when we analyzed the data from the two individual animals separately (Supplementary Fig. 6). Moreover, we observed robust decoding and generalization of stimulus and perceptual content within an individual dataset (see Supplementary Fig. 7 for results pertaining to one example dataset). Figure 4 summarizes the classification accuracy obtained within the two trial types for all individual datasets during the different temporal phases of the paradigm. Reliable classification and generalization accuracy was also obtained, when the PA activity was aligned to OKN change instead of stimulus changes (Supplementary Fig. 7). Together, these results suggest that the prefrontal population code underlying purely sensory perception without visual competition and subjective, conscious perception is similar, reliable and robust.

Further, we also assessed the ability of the population code to generalize across different recording days, by computing the generalization accuracy across recording sessions and compared it

with the classification accuracy obtained within a recording session (see methods). Although strong decoding was observed within individual datasets (mean: Physical Alternation = 94.61, Perceptual Dominance (BFS) = 91.82, Physical Dominance = 95.78, Perceptual Dominance (BR) = 86.35), the generalization accuracy across recording sessions was limited (mean: Physical Alternation = 63.12, Perceptual Dominance (BFS) = 65.09, Physical Dominance = 62.57, Perceptual Dominance (BR) = 57.70) (see Fig. 5 and methods). This is in agreement with recent work⁵⁰ which reported a steady turnover of individual neurons from one recording session to the next. This resulted in reduced prediction accuracy across sessions, when using fixed linear decoders, based directly on the recorded neural activity. Instead, it found low-dimensional latent dynamics were stable across days. The decreased generalization across recording days in our experiments, together with the strong classification performance within a day suggest that a similar turnover may have occurred in our recordings.

Decoding motion content in the presence and during the suppression of eye movements. Finally, given the tight coupling between OKN and perceptual content, we aimed at dissociating neural activity related to oculomotor processes from activity related to visual input. For a majority of units ($n = 581$, H07 ($n = 328$), and A11 ($n = 253$)), we estimated their preference to direction of motion during two control experiments, which utilized the following two paradigms: Fixation Off and Fixation On. During Fixation Off, visual input elicited OKN, similar to the BR paradigm, while during Fixation On, macaques suppressed their eye movements by fixating a centrally presented spot (Fig. 6). Example units displaying similar tuning curves across the two paradigms are displayed in Fig. 6e and f. Our analysis focused on upward and downward motion, which were used for instigating competition during BR. Ensemble population PSTHs (see methods) of significantly modulated units during Fixation Off or Fixation On and preferring the same motion direction across the two conditions displayed stronger activity when the preferred

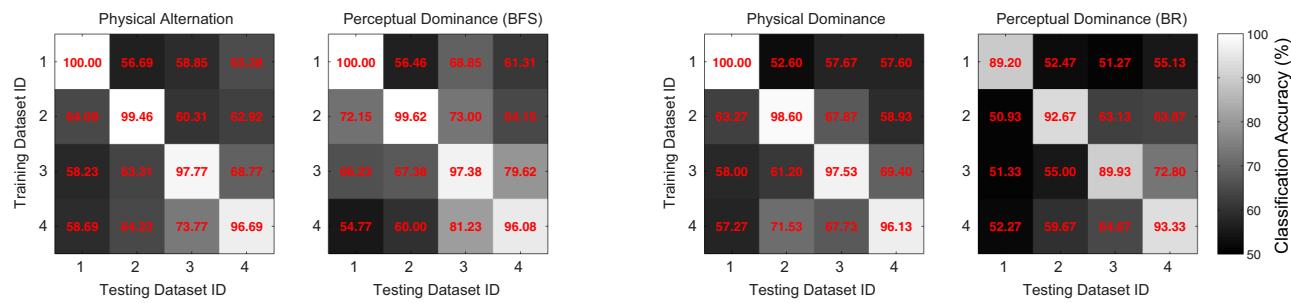
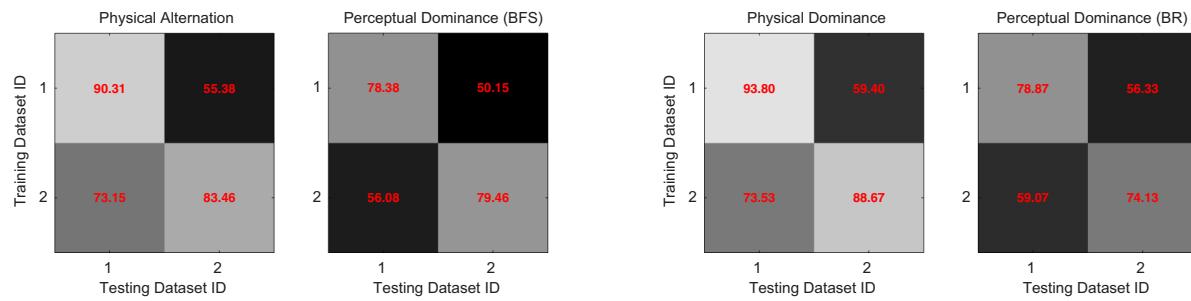
a Generalization across recording sessions (H07)**b Generalization across recording sessions (A11)**

Fig. 5 Decoding within and generalization across recording sessions. Displayed is the classification accuracy, **a**, computed for individual datasets and **b**, generalization accuracy across recording sessions. This analysis requires a correspondence of features across them. The units were identified after spike sorting, and can therefore be different across individual sessions. Thus, for this analysis, the spiking activity recorded from each of the 96 sites recorded with the array served as features, on which we trained and tested the classifier across the recording sessions. Results are presented for both the flash suppression (Perceptual dominance - BFS) and the rivalry phase (Perceptual dominance - BR) of the paradigm computed with neural activity elicited during an 800 ms window starting 200 ms after a stimulus or a perceptual change. The main diagonal (upper left to lower right) represents classification accuracy within a session and lateral diagonals represent generalization accuracy across sessions. While we find on average, high accuracy within individual recording sessions, the generalization across datasets is limited. Moreover, the classification accuracy within datasets was significantly more compared to generalization accuracy across datasets (two-sample t -test, two-sided, Physical Alteration_{within} vs. Physical Alteration_{across}, $p = 2.42 \times 10^{-9}$, Perceptual Dominance (BFS)_{within} vs. Perceptual Dominance (BFS)_{across}, $p = 1.94 \times 10^{-5}$, Physical Dominance_{within} vs. Physical Dominance_{across}, $p = 7.45 \times 10^{-10}$, Perceptual Dominance (BR)_{within} vs. Perceptual Dominance (BR)_{across}, $p = 1.07 \times 10^{-7}$). Potential reasons for such limited generalization could be that units either change or lose their preference, or it is likely, that we sampled from different units across recording sessions.

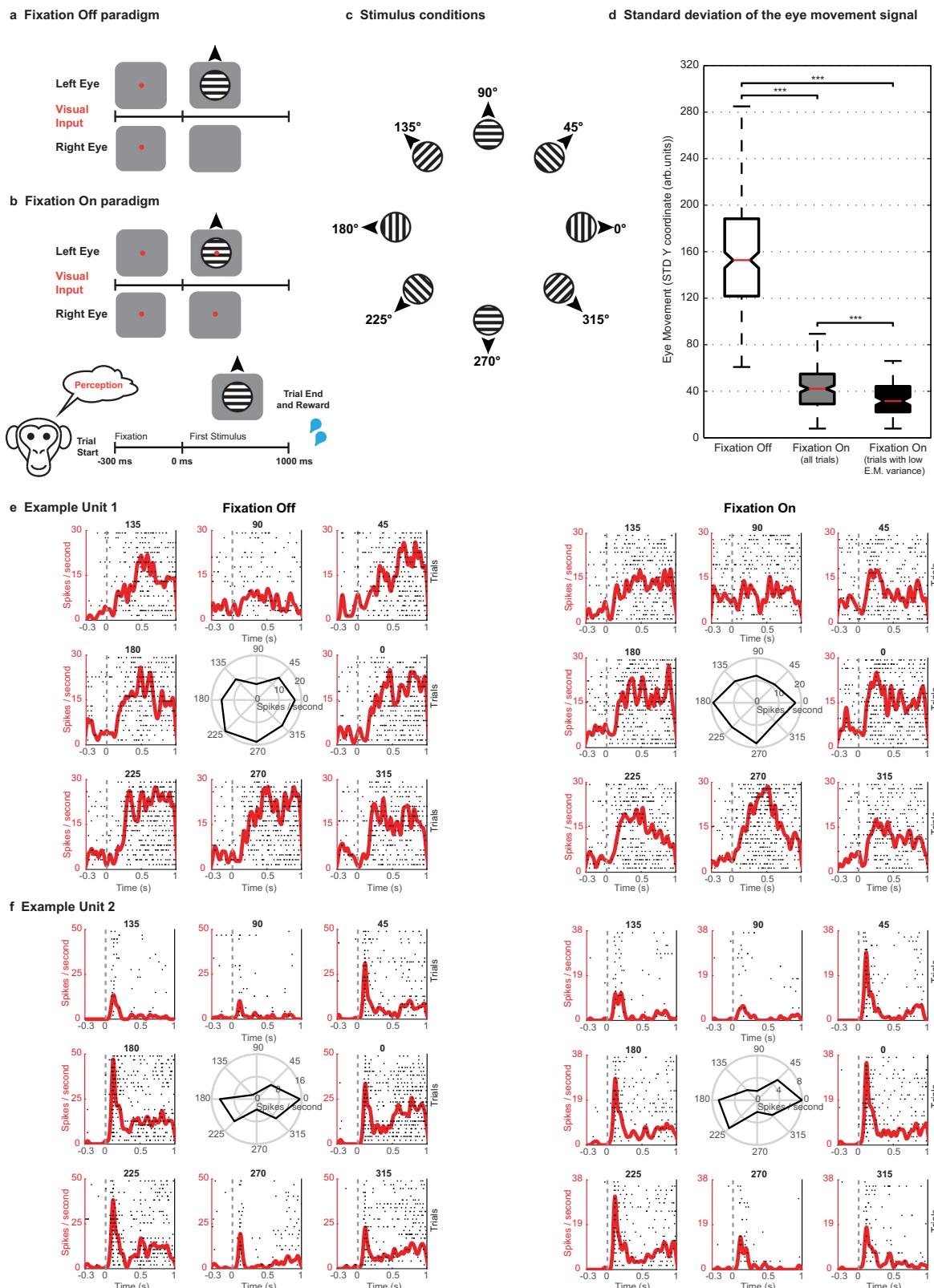
motion direction was presented and a reduction in response to the nonpreferred visual input during both paradigms (Fig. 7a). A similar modulation of population responses was observed in both animals (Supplementary Fig. 8a and d). Further, a classifier trained on neural responses of these units to stimuli that elicited OKN could reliably predict the same stimuli with significant accuracy (permutation test, $p \leq 0.002$, see methods), when macaques viewed them with eye movements suppressed, and vice versa. Importantly, reliable generalization across paradigms was observed, when the classifiers were trained on the data collected from the two individual animals separately (Supplementary Fig. 8c and f). Therefore, this indicates that prefrontal activity contains stimulus information and is not just driven by eye movements (Fig. 7b and c). We also obtained similar results using only those Fixation On trials where eye movements within the fixation window were further controlled (see methods, Fig. 6 and Supplementary Fig. 9). Importantly the cross-paradigm generalization was robust to the unit selection procedure (see methods and Supplementary Fig. 10). These findings corroborate previous work suggesting that the frontal cortex responds to visual motion both in the presence and absence of OKN^{43,51} and indicate that motion content signals contribute to the activity of the tested population.

We further quantified the difference in the modulation strength of the unit population across the two control paradigms

(Supplementary Fig. 11). This was assessed by computing a difference in the d-primes (fix Off – fix On) of individual units. Interestingly, the average of this difference was significantly different from zero for the last 500 ms (mean = 0.27 ± 0.07 , one-sample t -test, $p = 4.85 \times 10^{-4}$), but not the first 500 ms (mean = -0.003 ± 0.05 , one-sample t -test, $p = 0.9552$). A stronger response modulation appearing later in time during Fixation Off paradigm, is possibly related to a visual update, which is instigated by the saccade related to the optokinetic response. However, the similarity of this differential response (Supplementary Fig. 11a) during the first 500 ms across the two paradigms, suggests that prefrontal populations during this period represent direction of motion in the absence of any motor reports or when eye movements are suppressed. In summary, neurons in this prefrontal region reflect a mixture of perceptual and oculomotor signals⁵² and are selective for motion stimuli even when the monkeys fixate⁴³. Oculomotor signals are therefore not solely responsible for prefrontal activity reflecting the contents of visual consciousness.

Discussion

These results suggest that feature selective neural activity in the primate PFC reliably reflects internally generated changes in the content of subjective perception even without voluntary perceptual reports or external changes in visual stimulation, which



would elicit strong feedforward responses. While addressing an ongoing debate among the various theories of consciousness regarding the neural correlates of conscious perception in the PFC^{1,8,11,21,22,25,27,53}, we demonstrate that conscious contents can be reliably decoded from the activity of prefrontal ensembles.

A major debate in contemporary consciousness research focuses on the functional relevance of PFC in representing conscious

perception. Multiple studies indicate a prefrontal involvement during paradigms investigating conscious perception^{6–8,16–20,54}. However, recent theoretical and experimental work argue that previously reported frontal activation instead reflects prerequisites or consequences of perception^{23–27}. In particular, it has been suggested that other cognitive processes mediated by the PFC such as decision-making, introspection or preparation and execution of

Fig. 6 Control paradigms - Fixation Off, Fixation On and example unit activity. Trials started with a fixation spot, cueing the animal to bring and maintain gaze within a fixation window (300 ms), following which a drifting stimulus was presented monocularly. **a**, During Fixation Off, the fixation spot was removed at stimulus onset, thus inducing OKN. **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 $\pm 2^\circ$) until the trial ended, to receive a juice reward. **c**, During each trial, a stimulus drifting in one of eight different directions (pseudorandomized across trials) was presented. **d**, Whisker box plots (for box plot description, see statistical information, methods) displaying the distribution of standard deviations (STD) estimated from the eye movement signal (y-coordinate) elicited on individual trials during stimulus presentation (0–1000 ms). For Fixation On, either all ($n = 187$) or selected trials ($n = 95$), which displayed lower variance in the eye movement (E.M.) signal were analyzed (see methods). The STD was significantly reduced (Wilcoxon rank-sum test, two-sided, *** denotes $p \leq 0.001$, Fixation Off vs. Fixation On (all trials), $p = 7.86 \times 10^{-69}$, Fixation On (all trials) vs. Fixation Off (low E.M. variance), $p = 9.07 \times 10^{-5}$, Fixation Off (all trials) vs. Fixation On (low E.M. variance), $p = 6.09 \times 10^{-46}$) during Fixation On as compared to Fixation Off trials ($n = 239$). The results presented in this figure were computed with data from two animals pooled together. **e** and **f** show spike density functions overlaid on spike raster plots depicting the responses of two units to eight different motion directions during the two paradigms. The middle polar plots display the tuning curves of each unit (average response in Hz to gratings drifting in different directions). Spike rasters are displayed for first ' n ' trials of every motion direction presentation. Here, n is the minimum number of trials presented to the animal across any motion direction during a given paradigm. PSTHs and tuning curves were computed taking all trials (of a given motion direction) into account. **e**, Example Unit 1 displays a stronger response to a stimulus drifting downwards during both paradigms. The unit displayed in **f** responds strongly to two opposite directions of motion, thus displaying orientation preference. Although the firing rate was higher during the Fixation off paradigm, the unit displayed similar preference across both paradigms. Source data are provided as a Source Data file.

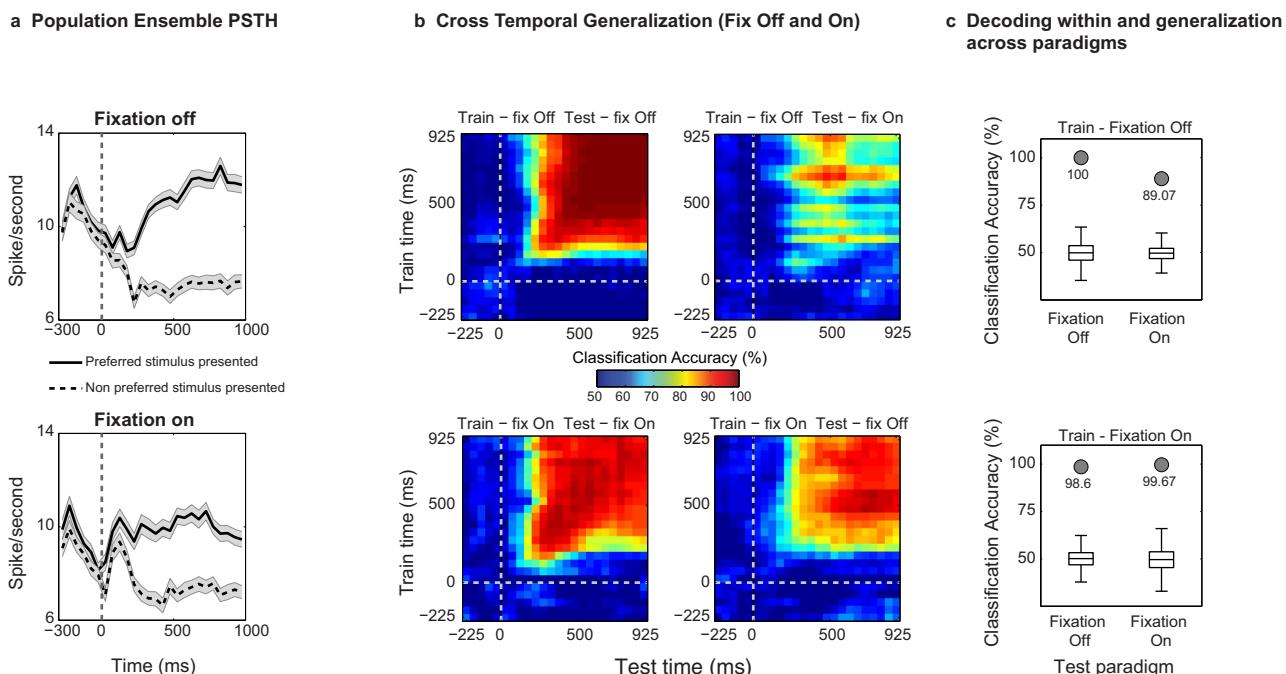


Fig. 7 Decoding of motion content in the presence and during the suppression of OKN eye movements from simultaneously recorded prefrontal ensembles. **a**, Ensemble population spiking activity (see methods) during fixation-Off and fixation-On paradigm for units which were significantly modulated in either paradigms and preferred the same motion direction (also see methods). Black solid and dashed lines depict the response to the preferred and nonpreferred stimulus respectively. Data are presented as mean and shaded regions depict standard error of the mean. **b**, Cross-temporal decoding of stimulus contents (binning parameters similar to Fig. 3b). **c**, Cross-paradigm invariance of the population code was assessed by training a classifier on activity during one paradigm and testing on the other, for a single 400 ms bin (starting 400 ms poststimulus onset) during visual motion presentation. Significant (permutation test, one-sided, estimated p -value: $p = 0.00199$) accuracy (checked by comparing it to accuracy obtained with shuffled labels ($n = 500$), summarized with box plots (for box plot description, see statistical information, methods)) suggests that the population code is invariant to the presence of large OKN, and encodes stimulus motion contents. Classification accuracy for decoding within the paradigm is also presented. The presented results were computed with data from two animals pooled together. We observed similar pattern of results in individual animals, and they are presented in supplementary figure 8. Source data are provided as a Source Data file.

motor reports (typical of these psychophysical paradigms), confound the previously observed correlates of conscious perception^{23,32–34}. The opponents to a prefrontal role therefore posit that the correlates of conscious perception are limited to posterior sensory regions, which has instigated a ‘front vs back’ of the cortex debate regarding the cortical organization underlying conscious perception^{1,11,27,54–57}. This is particularly relevant, since this critical issue separates key theories of consciousness with respect to their predictions about the role of PFC, in comparison

to posterior cortices^{1,55,58}. The present study addresses this debate by utilizing direct recordings of PFC neuronal populations during a no-report BR paradigm, wherein no volitional action was required from the subjects. In particular, this allowed us to carefully control for various cognitive processes associated with reporting, which may confound neural responses related to conscious perception. We find that prefrontal activity during BR reliably reflects changes in subjective perception, in the absence of these confounds.

BR offers a distinct advantage over other paradigms investigating visual consciousness such as BFS or visual masking, since it instigates stochastic, internally driven changes in the subjective perceptual content without a concomitant change in visual stimulation³⁰. Hence, it confers an opportunity to observe the internal neural dynamics contemporaneous with spontaneous changes in the contents of visual consciousness, without the conflation introduced from abrupt changes in the sensory input affecting the feedforward drive, which is typical of other perceptual paradigms, such as binocular flash suppression^{17,37} (also see Supplementary Table 1). Paired with electrophysiological investigations of the primate visual system^{59,60}, 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 to later temporal regions⁴⁰. A recent study reported non-selective modulation of neural activity in human medial frontal cortex (in the anterior cingulate (ACC) and pre-supplementary motor area (pre-SMA)) before spontaneous perceptual transitions during BR, suggesting that some medial frontal areas might reflect the prerequisites of conscious perception¹⁹. In contrast, our results demonstrate conscious content-specific representations in a macaque ventrolateral PFC subregion, where cells are selective for faces, complex visual objects and direction of motion^{41–43} and which is reciprocally connected with the inferotemporal cortex⁶¹. Importantly, previous electrophysiological studies probing the PFC during conscious perception either utilized a motor report^{16,18–20} and were therefore likely conflated by consequences of conscious perception, or investigated perceptual modulation among neurons selective for high-level stimuli, such as faces and complex objects with a no-report BFS paradigm¹⁷. In BFS, perceptual dominance and suppression are externally induced due to an abrupt change in the visual stimulation that results in strong changes in the feedforward input (see Fig. 1 and Supplementary Table 1). In contrast, during BR, perceptual dominance and suppression are endogenously driven and therefore neural activity recorded during this paradigm is more likely to reflect conscious perception¹⁷, uncontaminated from the strong initial feedforward signal associated with a newly presented stimulus. Hence, our results collected during unreported but clearly occurring spontaneous transitions in conscious perception indicate the existence of prefrontal representations of conscious content. Together with results obtained from previous electrophysiological investigations employing BR, present results indicate that the neural correlates of conscious perception are distributed throughout the cortical hierarchy with stronger representations in association cortical areas, including the PFC.

These results constitute a critical advance in our understanding of PFC function. A quintessential node in the perception-action cycle, PFC has long been considered crucial for temporal organization of goal-directed behavior^{62,63}. This is aided by its brain-wide anatomical connectivity, which includes both sensory and motor regions⁶⁴. An important prerequisite suggested for implementing such executive function is a cognitively relevant representation of conscious perception in the PFC⁶³. However, we never trained the animals participating in the current study to categorize or discriminate motion direction and report their perception, thus precluding an expectation that they were mentally generating an action, which was then withheld. Hence, our finding of report-free correlates of conscious content in the frontal cortex suggests that such perceptual information is available within the region, even when no rule-based overt action, such as a learned motor response associated with a specific stimulus, is required. Therefore, these results challenge the notion that post-perceptual processes, in particular, motor reports are the major source of PFC activity during paradigms of conscious

perception^{23,32}. They further challenge a purely task associated representation of perception in the PFC. Additionally, the diverse bi-directional connectivity of prefrontal cortex⁶⁴ endows it with the ability to communicate this disambiguated ongoing perceptual content contemporaneously to multiple regions of the primate brain. This is particularly relevant in cases where top-down influences from the PFC might be required, such as cognitive control, attentional selection or visual processing^{62,65}. Taken together, a holistic understanding of PFC function necessitates an investigation of areal and functional segregation of perceptual and cognitive processes within the region⁶⁶.

In addition to finding conscious perception related signals in the PFC, we could reliably decode conscious content from pre-frontal activity on individual trials with high accuracy. This is noteworthy, given that a multivariate pattern analysis approach, i.e., decoding, is considered a powerful framework for identifying the neural correlates of consciousness^{67–69}, in comparison to univariate contrastive analysis. For example, decoding offers higher sensitivity since it uses the correlation or covariance of signals recorded across multiple units. Further, it confers an ability to judge the consistency with which the population code represents conscious content on a trial by trial basis, as well as across task conditions^{68,69}. Moreover, it can provide insights into how neural representations transform in time by lending itself to a chronological evaluation of the population code⁷⁰. In regards to this, our temporal generalization analysis revealed a population code, which was stable across time and robustly represented perception, within the PFC. This is likely important for mediating perceptual continuity⁷¹.

Crucially, we note that decoding of unambiguous stimuli (PA trials) alone cannot guarantee that neurons also reflect conscious perception of these stimuli. However, our result demonstrating the ability to decode with such high accuracy conscious perception from PFC activity during binocular rivalry, a visual stimulation protocol that dissociates conscious perception from visual input confirms decoding of subjective conscious perception, during visual competition. Specifically, the ability of the classifier to generalize across the unambiguous (PA) and ambiguous visual presentations (BR) suggests an invariance of the prefrontal population code underlying conscious content. In contrast, the absence of invariance between PA and BR could indicate representation of perceptually suppressed stimuli, therefore signaling the unconscious percept, whose representation seems to be minimal in PFC during BR. Interestingly, such a multiplexed representation of perceptually dominant and suppressed stimulus has been suggested recently in the case of temporal lobe neurons⁷².

Such a decoding approach could be used for investigating the population code subserving conscious perception across cortical regions, by for e.g., comparing decoding accuracy. However, the particularly strong classification and generalization accuracies (up to 95% obtained on data combined across both animals, see Fig. 3c) obtained from prefrontal responses, makes it unlikely that stronger decoding of the conscious percept would be observed from activity in other cortical areas.

Our findings are in contrast to the conclusions of recent imaging studies suggesting PFC's reduced involvement in conscious perception^{23,24,26}. However, constraints in the spatio-temporal 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^{29,54}. Such limits are particularly relevant to the frontal cortex, where individual neurons often display a high degree of mixed selectivity^{73,74} or distinct temporal patterns of activity during perceptual paradigms⁶⁶. For example, we found that sites displaying preferential responses to opposite motion directions are

distributed in close proximity (~0.4 mm) (Fig. 1d and Supplementary Fig. 2). Such spatial variability of stimulus selectivity remains difficult to capture with fMRI, especially with univariate methodologies. However, advances in high field imaging together with multivariate pattern analysis hold great promise in providing brain-wide representations of conscious content⁶⁷.

In summary, our results lend support to theoretical approaches, which suggest a representation of conscious content in the executive areas of the brain such as the PFC to mediate cognitive functions or motor action^{3,5,6,75}. Future work employing direct activation of such perceptually modulated ensembles could help elucidate the causal mechanisms mediating conscious perception⁷⁶. Together with carefully designed experiments, it could help unravel the similarities and differences in the neural mechanisms underlying conscious perception and other cognitive processes⁵⁶ such as introspection^{23,77}, attention⁷⁸, decision making^{79,80} or cognitive control^{34,81}.

Methods

All experiments were in full compliance with the guidelines of the European community (2010/63/EU) for the care and use of laboratory animals and were approved by the local authorities (Regierungspräsidium, Tübingen, Baden-Württemberg, Germany, protocol KY6/12).

Binocular rivalry, control paradigms, and stimulus presentation. The paradigm consisted of two trial types, namely, the physical alternation (PA) trials and binocular rivalry (BR) trials (Supplementary Fig. 1a). 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 ($\pm 8^\circ$), a drifting sinusoidal grating (size: 8°, speed: 12–13°/sec, spatial frequency: 0.5 cycles/degree, gratings were drifting vertically up or down) was monocularly presented. During one recording session (in A11), we used random dot motion stimulus (field of view 8°, speed 13°/sec, 200 limited lifetime dots, and 100% coherence) since visual inspection of OKN revealed that such a stimulus produced better OKN responses than drifting gratings in this animal. Further, it helped assess if prefrontal responses are correlated with perception of motion, independent of the low-level stimulus properties generating it. After one or two seconds, the first stimulus was removed and a second stimulus drifting in the opposite direction was presented to the contralateral eye in PA trials. During 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 termed as flash suppression^{17,37,38,82} (Supplementary Fig. 1a). After this period, visual input alternated between oppositely drifting stimuli presented monocularly in the PA condition. In comparison, during the BR condition, the percept of the animal switched endogenously between the discordant visual stimuli. The temporal histogram of perceptual dominance during BR could be approximated with a gamma distribution (Fig. 1c). The total duration of a single-trial/observation period was around 8 to 10 s. Note that the perception of the animal (Supplementary Fig. 1a) is similar during the two trial types, even though the underlying visual input is monocular in PA trials, while it is dichoptic during BR. The eye (where the first stimulus was presented), motion direction (which was presented first), and trial types (PA or BR) were pseudorandomized within a single dataset. During the entire period of a trial, animals had been trained to maintain their gaze within a fixation window ($\pm 8^\circ$), which encompassed 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 out of 6 recording sessions, 2 - macaque H'07, 2 - macaque A'11). Both paradigms consisted of trials, where the macaques were presented with a visual stimulus drifting in one of eight different directions (pseudorandomized across trials) for one second (Fig. 6). 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) encompassed the stimulus ($\pm 8^\circ$). In contrast, during the Fixation On paradigm, a fixation spot overlaid on the stimulus indicated that the monkeys must fixate within a smaller fixation window ($\sim \pm 1$ to ± 2 degrees) to complete the trial and receive the 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 1280×1024 on the monitors (running at a 60 Hz refresh rate) using a dedicated graphics workstation. The custom-made stereoscope included two mirrors, slanted to approximately a 45° angle and placed in front of the animal's eyes. The angle of the mirrors could be adjusted to allow precise

binocular stimulation. In addition, the setup included two monitors, each on the left and right side, facing each other and perpendicular to the direction in which the animal faced. This allowed binocular stimulation, wherein images displayed on the left and right monitor could be presented to corresponding retinal locations of the left and right eye respectively, with the aid of slanted mirrors. Prior to the presentation of the BR paradigm, we carried out a previously described calibration procedure⁸², 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 a 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. The animal was trained to make a saccade to the eccentrically 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.

Visual stimulus design and presentation were carried out using OpenGL (version 1.2) and controlled on a windows machine via Tcl/Tk (version 8.0). A QNX real-time operating system (QNX Software Systems, version 4.25, BlackBerry Limited, Waterloo, Canada) managed the precise temporal presentation of the visual stimuli, and sent digital pulses to the Blackrock recording system. An infrared camera captured eye movements (1 kHz sampling rate) with the software iView (SensoriMotoric Instruments GmbH, Germany). Besides monitoring eye movements online, they were also stored for offline analysis in both, the QNX-based acquisition system as well as the Blackrock neural data acquisition system. We used the latter to align the neural data.

Surgical procedures. Two healthy male rhesus monkeys (*Macaca mulatta*), H07 and A11 participated in behavioral and electrophysiological recordings. H07 was ~12 years old and weighed ~11–12 kg at the time of the study. A11 was ~15 years old, and weighed ~8–9 kg, when the experiments were done. 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 (2010/63/EU) for the care and use of laboratory animals. Each animal was implanted with a titanium headpost custom designed to fit the skull based on 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 described⁸³. The MR scan also aided in localizing the inferior convexity of the lateral PFC. Post behavioral training in the task, the animals underwent another surgery, where a Utah microelectrode array (Blackrock Microsystems, Salt Lake City, Utah USA⁸⁴) was implanted in the PFC. The array had a 10 by 10 electrode configuration and was 4 mm by 4 mm in size, with an inter-electrode distance of 400 μ m 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 (Fig. 1a).

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 (sampled at 30 kHz) were recorded with the Neural Signal Processors (Blackrock Microsystems) and band-pass filtered offline between 0.6 – 3 kHz using a second-order Butterworth filter. Spikes were detected with an amplitude threshold set at five times the median absolute deviation⁸⁵. 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 ms) around the peak were extracted for spike sorting. An automatic clustering procedure identified putative single neurons via a Split and Merge Expectation-Maximization algorithm which fits a mixture of Gaussians on the spike feature data consisting of the first three principal components of the spike waveforms⁸⁶. Inspection and manual cluster cutting was carried out in Klusters (Lynn Hazan, Buzsáki lab, Rutgers, Newark, United States)⁸⁷. This way, we sorted and identified putative single units recorded from each of the 96 electrodes in the array. 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 (unless stated otherwise), we combined individual single units and multi-units recorded and they are together referred to as units. After sorting, we identified a total of 734 units from H07 across four recording sessions, and 253 units from A11 across two recording sessions.

Selectivity of unit activity. Eye movements during each BR trial were visually inspected with the aid of a custom written GUI in MATLAB (MathWorks, Natick, United States) and the onset and end of a perceptual dominance (during the rivalry phase) were 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 ($p \leq 0.05$). 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 period of 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 a 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, we considered only those epochs during PA and BR trials for computing selectivity, which consisted of perceptual dominance (BFS and BR) or monocular presentation (PA) of a given stimulus lasting at least 1000 ms. Therefore, only those epochs of flash suppression dominance were included in the analysis, which entailed a successful suppression of the first monocularly presented stimulus for at least 1000 ms (as determined by the OKN). With respect to perceptual switches, we analyzed transitions, which consisted of at least 1000 ms of clear dominance (judged by a stable OKN pattern), before and after an OKN switch. To compare with PA as closely as possible, we analyzed those transitions during BR, which had an interval of less than 250 ms between the end of the preceding dominance, and the onset of the next. Data were 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 ms. 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 ms period. Any relevant figures presented in the main body of the paper were obtained by analyzing the spiking activity elicited during PA trials which was 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 (Fig. 2) was repeated with PA trials aligned according to this new criterion and we obtained very similar results (Supplementary Fig. 4 and 5).

D-prime calculation. For every unit, we computed a preference index denoted as d' , by quantifying the strength of its selectivity during PA trials. It was calculated as follows:

$$d' = (\mu p - \mu np) / \left(\sqrt{\left(\frac{\text{Varp} + \text{Varnp}}{2} \right)} \right) \quad (1)$$

where, μp and μnp is the average spiking response of a given unit during the presentation of its preferred and nonpreferred stimulus, calculated over a duration of 1000 ms after a physical stimulus change during PA trials. The difference between these two quantities is normalized by the square root of the average pooled variance (Varp and Varnp) of the response distributions. The d' was similarly calculated during corresponding periods of BR trials. However, the preferred and nonpreferred stimulus conditions were designated based on PA trials. A unit with a positive d' during both PA and BR trials, signifies similar preference, while a negative d' during BR indicates that the unit had opposite or no preference during PA (that is, it fired on average equal number of spikes for both motion conditions) (Fig. 2a and Supplementary Fig. 4a).

Conventional population PSTHs and ensemble PSTHs. Population PSTHs (Fig. 2b) were computed by averaging the mean neural activity of selective units in response to their preferred and nonpreferred 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 during either 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 trials (Supplementary Fig. 3 and 5) were also computed. Population activity computed for the switches included units, which were significantly modulated both before and after the switch for the same motion direction in PA or BR (Supplementary Fig. 3 and 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. Unit populations, which contributed to ensemble PSTHs, were identified similarly to conventional population PSTHs. Therefore, the population of units contributing to Fig. 2b (switches) and Fig. 3a 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 Fig. 3a. 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 Fig. 7a, Supplementary Fig. 8a, d and 9a were generated similarly as described above, but without the normalization step. Further, the response of the two ensembles has been pooled together in this case. 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 nonpreferred stimulus was presented, were pooled together and averaged for ensemble activity pertaining to the nonpreferred stimulus.

Decoding Analysis. Multivariate pattern analysis was utilized to assess, if the spiking activity of neuronal ensembles in the PFC contained information about the contents of perception on a single transition basis. In this regard, we used a maximum correlation coefficient classifier⁴⁹ implemented as a part of the neural decoding toolbox⁸⁸. All the recorded units ($n = 987$) across the two monkeys were pooled as a pseudopopulation for the decoding analysis pertaining to the BR paradigm (Fig. 3). This is similar to previous studies^{49,89}, where units recorded during independent sessions were compiled together as a pseudopopulation. For testing generalization across sessions (Fig. 5), each of the 96 individual electrodes of the array was considered a unit to maintain correspondence across sessions. The valid spiking activity recorded from this electrode was used for training and testing the classifier across the recording sessions. In either case (independent of how the units were defined), responses of each of these units during 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 left out and put in the test. For decoding during flash suppression dominance (presented in Fig. 4 and Fig. 5), the number of cross-validation splits were 13. This procedure was repeated 50 times (resample runs) to estimate the classification accuracy with a different randomly chosen cross-validation split during each run. All decoding accuracy estimates are zero-one-loss results. Each pixel in cross-temporal generalization plots (Figs. 3b, 7b and Supplementary Fig. 6b, e, 7b, e, 8b, e, 9b) 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 recorded in the frontal and temporal cortex^{49,89}.

Similar steps as described above were used for decoding analysis pertaining to the control paradigms (Fig. 7 and Supplementary Fig. 8 and 9), except the following. Units which were significantly selective (Wilcoxon rank-sum test, $p \leq 0.05$) in either of the two paradigms and preferred the same stimulus, that is stronger response to the same motion direction during both paradigms (Final selection of units used: $n = 103$ (H07: $n = 66$, A11: $n = 37$) for Fig. 7 and $n = 100$ (H07: $n = 65$, A11: $n = 35$) for Supplementary Fig. 9) participated in the decoding procedure. Further, we employed eight cross-validation splits for the decoding analysis pertaining to trials with reduced eye movement variance (Supplementary Fig. 9).

Additionally, we wanted to assess the robustness of the cross-paradigm generalization across the control paradigms to the above mentioned unit selection procedure. To this end, we ran the decoding procedure ten times, and during each run, the units participating in the analysis were selected in the following manner. Similar to as described above, the units were chosen based on their selectivity (Wilcoxon rank-sum test, $p \leq 0.05$) in either of the two paradigms and similar preference (upward or downward) across them. However, this was computed differently: using all of the trials of one paradigm, which was used for subsequent training of the classifier, and only half of the trials of the second paradigm, which were randomly selected during each run. The generalization accuracy of the classifier was subsequently tested by training the classifier (on the activity of the set of units passing the aforementioned criterion) on the first paradigm and testing it on the remaining half of the trials from the second paradigm, which did not participate in the aforementioned unit selection procedure. The number of cross-validation splits employed for this analysis was seven. The rest of the decoding parameters were similar to as described above. We repeated this procedure 10 times, for both comparisons, that is, training the classifier on trials from Fixation Off and testing it on Fixation On, and vice versa (Supplementary Fig. 10). The resulting classification accuracies were strongly significant ($p \leq 0.002$), based on the permutation test.

Selection of trials with reduced eye position variance and decoding. To create a robust dataset for decoding the visual stimulus during passive observation of monocularly presented stimuli, we only picked those trials of the Fixation On paradigm, wherein the eye movements during passive fixation were relatively minimal, that is, there were no strong drifts in eye position. The trial selection was carried out by utilizing the following procedure. Firstly, the Y coordinate of the eye movement signal was detrended. It was then filtered below 20 Hz to remove high-frequency noise. Next, a double differential of this signal 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 computed giving us a distribution of errors. All trials within a condition, whose sum of least-squared error was less than the median of the distribution of these errors obtained from all trials in that condition, were selected for further analysis. This method resulted in a selection of trials with significantly (Wilcoxon rank-sum test, $p \leq 0.001$) reduced variance of the eye movements signal (Fig. 6d).

Statistical information. Significant selectivity of units was analyzed with a Wilcoxon rank-sum test during both PA and BR trials. We compared the spike responses elicited during specific temporal phases, namely the physical alternation, physical dominance, perceptual dominance during BFS or BR, as well as stimulus and perceptual switches as described above in the selectivity of unit activity section. The alpha value was 0.05.

Statistical significance of the classification accuracy was assessed using a permutation test, which involved running the decoding analysis on the data with labels shuffled^{48,88}. 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 resulting 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 ($p \leq 1/500 = 0.002$, or an estimated p -value: $p = 0.00199$). Significance of decoding accuracy was computed using this procedure for the results presented in Figs. 3 and 7 (also see Supplementary Fig. 6, 7, 8, 9 and 10).

With respect to the box plots presented in Figs. 3c, 6d, 7c and Supplementary Fig. 6, 7, 8, 9 and 10, the box denotes the 25th (Q1) and 75th percentiles (Q3) of the data, while the red (Fig. 6d) or black central line (other figures) 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. Outliers are not displayed. In Fig. 6d, the 95% confidence interval around the median is approximated by the notches, whose edges are calculated as $\text{median} \pm 1.57 \times (Q3 - Q1) / (\text{square root of number of samples})$.

Unit related statistics. The average spike density functions of two simultaneously recorded units, 33 (preferring downward motion) and 167 (preferring upward motion), which displayed robust modulation during both PA and BR trials are presented in Fig. 1e. These were plotted with the path function (which utilized a gaussian kernel with a standard deviation of 50 ms) from the Chronux data analysis toolbox (<http://chronux.org/>)⁹⁰. The spiking activity of the units reliably switched for both externally induced and internally driven perceptual switches as assessed with a Wilcoxon rank-sum test. With respect to unit 33, the statistics of significant modulation were as follows: during physical alternation phase of PA trials, (temporally analogous to flash suppression dominance during BR), $p_{PA-33} = 2.82 \times 10^{-14}$ and for binocular flash suppression phase during BR trials, $p_{BFS-33} = 2.39 \times 10^{-6}$. Additionally, it was also significantly modulated during the physical dominance (PD) phase of PA trials and its temporally equivalent rivalry dominance (RD) phase during BR trials. The statistical value was $p_{PD-33} = 8.72 \times 10^{-15}$ for physical dominance and $p_{RD-33} = 7.18 \times 10^{-8}$ for rivalry dominance. For unit 167, the corresponding values were: $p_{PA-167} = 1.13 \times 10^{-17}$, $p_{BFS-167} = 1.20 \times 10^{-9}$, $p_{PD-167} = 1.49 \times 10^{-4}$ and $p_{RD-167} = 7.8 \times 10^{-3}$.

During the eye movement control experiments, a majority of units displaying significant stimulus selectivity (Wilcoxon rank sum test, $p \leq 0.05$) within a given paradigm retained their stimulus preference across the two control paradigms (fix On—69.56 %, 48/69 (H07: 70.83%, 34/48; A11: 66.67%, 14/21); fix Off—55.56 %, 80/144 (H07: 51.55%, 50/97; A11: 63.83%, 30/47)) while a small percentage of units (fix On—14.49 %, 10/69 (H07: 16.67%, 8/48; A11: 9.52%, 2/21); fix Off—6.94 %, 10/144 (H07: 8.25%, 8/97; A11: 4.26%, 2/47); Wilcoxon rank sum test, $p \leq 0.05$) exhibited a significant preference to stimuli with opposing motion content across the two paradigms (see Fig. 6 for tuning curves of example units).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data are provided with this paper. Data with spiking activity, that support the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

We used codes, implemented as part of the neural decoding toolbox for carrying out the various decoding analysis reported within the paper. The toolbox (including the associated codes) is available for download at <http://www.readout.info/downloads/>. The utility of this toolbox has been previously described⁸⁷. Other code is available from the corresponding author upon reasonable request.

Received: 21 October 2020; Accepted: 17 February 2022;

Published online: 22 March 2022

References

- Koch, C. What Is Consciousness? *Nature* **557**, S8–S12 (2018).
- Crick, F. & Koch, C. Towards a neurobiological theory of consciousness. *Seminars Neurosci.* **2**, 263–275 (1990).
- Crick, F. & Koch, C. Consciousness and neuroscience. *Cereb. Cortex* **8**, 97–107 (1998).
- Baars, B. J. Global workspace theory of consciousness toward a cognitive neuroscience of human experience. *Prog. Brain Res.* **150**, 45–53 (2005).
- Lau, H. & Rosenthal, D. Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci. (Regul. Ed.)* **15**, 365–373 (2011).
- Dehaene, S. & Changeux, J.-P. Experimental and theoretical approaches to conscious processing. *Neuron* **70**, 200–227 (2011).
- Rees, G., Kreiman, G. & Koch, C. Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* **3**, 261–270 (2002).
- Brascamp, J., Sterzer, P., Blake, R. & Knapen, T. Multistable perception and the role of the frontoparietal cortex in perceptual inference. *Annu. Rev. Psychol.* **69**, 77–103 (2018).
- Vignal, J. P., Chauvel, P. & Halgren, E. Localised face processing by the human prefrontal cortex: stimulation-evoked hallucinations of faces. *Cogn. Neuropsychol.* **17**, 281–291 (2000).
- Blanke, O., Landis, T. & Seeck, M. Electrical cortical stimulation of the human prefrontal cortex evokes complex visual hallucinations. *Epilepsy Behav.* **1**, 356–361 (2000).
- Odegaard, B., Knight, R. T. & Lau, H. Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* **37**, 9593–9602 (2017).
- Del Cul, A., Dehaene, S., Reyes, P., Bravo, E. & Slachevsky, A. Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* **132**, 2531–2540 (2009).
- Szczepanski, S. M. & Knight, R. T. Insights into human behavior from lesions to the prefrontal cortex. *Neuron* **83**, 1002–1018 (2014).
- Nakamura, R. K. & Mishkin, M. Chronic “blindness” following lesions of nonvisual cortex in the monkey. *Exp. Brain Res.* **63**, 173–184 (1986).
- Colás, I. et al. Conscious perception in patients with prefrontal damage. *Neuropsychologia* **129**, 284–293 (2019).
- van Vugt, B. et al. The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* **360**, 537–542 (2018).
- Panagiotaropoulos, T. I., Deco, G., Kapoor, V. & Logothetis, N. K. Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron* **74**, 924–935 (2012).
- Libedinsky, C. & Livingstone, M. Role of prefrontal cortex in conscious visual perception. *J. Neurosci.* **31**, 64–69 (2011).
- Gelbard-Sagiv, H., Mudrik, L., Hill, M. R., Koch, C. & Fried, I. Human single neuron activity precedes emergence of conscious perception. *Nat. Commun.* **9**, 2057 (2018).
- Thompson, K. G. & Schall, J. D. The detection of visual signals by macaque frontal eye field during masking. *Nat. Neurosci.* **2**, 283–288 (1999).
- Tononi, G., Boly, M., Massimini, M. & Koch, C. Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* **17**, 450–461 (2016).
- Lamme, V. A. F. Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **373**, 20170344 (2018).
- Frässle, S., Sommer, J., Jansen, A., Naber, M. & Einhäuser, W. Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* **34**, 1738–1747 (2014).
- Knapen, T., Brascamp, J., Pearson, J., van Ee, R. & Blake, R. The role of frontal and parietal brain areas in bistable perception. *J. Neurosci.* **31**, 10293–10301 (2011).
- Koch, C., Massimini, M., Boly, M. & Tononi, G. Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* **17**, 307–321 (2016).
- Brascamp, J., Blake, R. & Knapen, T. Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nat. Neurosci.* **18**, 1672–1678 (2015).

27. Boly, M. 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).
28. Safavi, S., Kapoor, V., Logothetis, N. K. & Panagiotaropoulos, T. I. Is the frontal lobe involved in conscious perception? *Front. Psychol.* **5**, 1063 (2014).
29. Logothetis, N. K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).
30. Blake, R. & Logothetis, N. K. Visual competition. *Nat. Rev. Neurosci.* **3**, 13–21 (2002).
31. Wheatstone, C. Contributions to the physiology of vision. part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos. Trans. R. Soc. Lond.* **128**, 371–394 (1838).
32. Tsuchiya, N., Wilke, M., Frässle, S. & Lamme, V. A. F. No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends Cogn. Sci. (Regul. Ed.)* **19**, 757–770 (2015).
33. Pitts, M. A., Metzler, S. & Hillyard, S. A. Isolating neural correlates of conscious perception from neural correlates of reporting one's perception. *Front. Psychol.* **5**, 1078 (2014).
34. Aru, J., Bachmann, T., Singer, W. & Melloni, L. Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* **36**, 737–746 (2012).
35. Fox, R., Todd, S. & Bettinger, L. A. Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vis. Res.* **15**, 849–853 (1975).
36. Logothetis, N. K. & Schall, J. D. Binocular motion rivalry in macaque monkeys: eye dominance and tracking eye movements. *Vis. Res.* **30**, 1409–1419 (1990).
37. Wolfe, J. M. Reversing ocular dominance and suppression in a single flash. *Vis. Res.* **24**, 471–478 (1984).
38. Lansing, R. W. Electroencephalographic correlates of binocular rivalry in man. *Science* **146**, 1325–1327 (1964).
39. Levelt, W. J. Note on the distribution of dominance times in binocular rivalry. *Br. J. Psychol.* **58**, 143–145 (1967).
40. Logothetis, N. K. Single units and conscious vision. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **353**, 1801–1818 (1998).
41. Pigarev, I. N., Rizzolatti, G. & Scandolara, C. Neurons responding to visual stimuli in the frontal lobe of macaque monkeys. *Neurosci. Lett.* **12**, 207–212 (1979).
42. Wilson, F. A., Scalaidhe, S. P. & Goldman-Rakic, P. S. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* **260**, 1955–1958 (1993).
43. Zaksas, D. & Pasternak, T. Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* **26**, 11726–11742 (2006).
44. Nigam, S., Pojoga, S. & Dragoi, V. A distinct population of heterogeneously color-tuned neurons in macaque visual cortex. *Sci. Adv.* **7**, eabc5837 (2021).
45. Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J. & Cohen, M. R. Learning and attention reveal a general relationship between population activity and behavior. *Science* **359**, 463–465 (2018).
46. Milton, R., Shahidi, N. & Dragoi, V. Dynamic states of population activity in prefrontal cortical networks of freely-moving macaque. *Nat. Commun.* **11**, 1948 (2020).
47. Sheinberg, D. L. & Logothetis, N. K. The role of temporal cortical areas in perceptual organization. *Proc. Natl Acad. Sci. USA* **94**, 3408–3413 (1997).
48. Meyers, E. & Kreiman, G. in *Visual Population Codes* (eds. Kriegeskorte, N. & Kreiman, G.) 517–538 (2011).
49. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* **100**, 1407–1419 (2008).
50. Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A. & Miller, L. E. Long-term stability of cortical population dynamics underlying consistent behavior. *Nat. Neurosci.* **23**, 260–270 (2020).
51. Dieterich, M., Bucher, S. F., Seelos, K. C. & Brandst, T. Horizontal or vertical optokinetic stimulation activates visual motion-sensitive, ocular motor and vestibular cortex areas with right hemispheric dominance. An fMRI study. *Brain* **121**, 1479–1495 (1998). Pt 8.
52. Kim, J. N. & Shadlen, M. N. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat. Neurosci.* **2**, 176–185 (1999).
53. Mashour, G. A. The controversial correlates of consciousness. *Science* **360**, 493–494 (2018).
54. Michel, M. & Morales, J. Minority reports: Consciousness and the prefrontal cortex. *Mind Language* **35**, 493–513 (2019).
55. Melloni, L., Mudrik, L., Pitts, M. & Koch, C. Making the hard problem of consciousness easier. *Science* **372**, 911–912 (2021).
56. Block, N. What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends Cogn. Sci. (Regul. Ed.)* **23**, 1003–1013 (2019).
57. Panagiotaropoulos, T. I., Dwarakanath, A. & Kapoor, V. Prefrontal cortex and consciousness: beware of the signals. *Trends Cogn. Sci. (Regul. Ed.)* **24**, 343–344 (2020).
58. Bartels, A. Consciousness: What is the role of prefrontal cortex? *Curr. Biol.* **31**, R853–R856 (2021).
59. Schmid, M. C. & Maier, A. To see or not to see—thalamo-cortical networks during blindsight and perceptual suppression. *Prog. Neurobiol.* **126**, 36–48 (2015).
60. Panagiotaropoulos, T. I., Kapoor, V. & Logothetis, N. K. Subjective visual perception: from local processing to emergent phenomena of brain activity. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **369**, 20130534 (2014).
61. Webster, M. J., Bachevalier, J. & Ungerleider, L. G. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb. Cortex* **4**, 470–483 (1994).
62. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
63. Fuster, J. M. The Prefrontal Cortex—An Update. *Neuron* **30**, 319–333 (2001).
64. Yeterian, E. H., Pandya, D. N., Tomaiuolo, F. & Petrides, M. The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* **48**, 58–81 (2012).
65. Gilbert, C. D. & Li, W. Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14**, 350–363 (2013).
66. Kapoor, V., Besserve, M., Logothetis, N. K. & Panagiotaropoulos, T. I. Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. *Commun. Biol.* **1**, 215 (2018).
67. Haynes, J.-D. Decoding visual consciousness from human brain signals. *Trends Cogn. Sci. (Regul. Ed.)* **13**, 194–202 (2009).
68. Sandberg, K., Andersen, L. M. & Overgaard, M. Using multivariate decoding to go beyond contrastive analyses in consciousness research. *Front. Psychol.* **5**, 1250 (2014).
69. Haynes, J.-D. & Rees, G. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523–534 (2006).
70. King, J. R. & Dehaene, S. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci. (Regul. Ed.)* **18**, 203–210 (2014).
71. Wurtz, R. H. Corollary discharge contributions to perceptual continuity across saccades. *Annu. Rev. Vis. Sci.* **4**, 215–237 (2018).
72. Hesse, J. K. & Tsao, D. Y. A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli. *Elife* **9**, (2020).
73. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
74. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
75. Brown, R., Lau, H. & LeDoux, J. E. Understanding the Higher-Order Approach to Consciousness. *Trends Cogn. Sci. (Regul. Ed.)* **23**, 754–768 (2019).
76. Marshel, J. H. et al. Cortical layer-specific critical dynamics triggering perception. *Science* **365**, eaaw5202 (2019).
77. Goldberg, I. I., Harel, M. & Malach, R. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* **50**, 329–339 (2006).
78. Watanabe, M. et al. Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science* **334**, 829–831 (2011).
79. Shadlen, M. N. & Kiani, R. in *Characterizing Consciousness: From Cognition to the Clinic?* (eds. Dehaene, S. & Christen, Y.) 27–46 (Springer Berlin Heidelberg, 2011).
80. Kang, Y. H. R., Petzschner, F. H., Wolpert, D. M. & Shadlen, M. N. Piercing of Consciousness as a Threshold-Crossing Operation. *Curr. Biol.* **27**, 2285–2295.e6 (2017).
81. Block, N. in *The Future of the Brain* (eds. MARCUS, G. & FREEMAN, J.) 161 (Princeton University Press, 2015).
82. Keliris, G. A., Logothetis, N. K. & Tolias, A. S. The role of the primary visual cortex in perceptual suppression of salient visual stimuli. *J. Neurosci.* **30**, 12353–12365 (2010).
83. Logothetis, N., Merkle, H., Augath, M., Trinath, T. & Ugurbil, K. Ultra high-resolution fMRI in monkeys with implanted RF coils. *Neuron* **35**, 227–242 (2002).
84. Maynard, E. M., Nordhausen, C. T. & Normann, R. A. The Utah intracortical Electrode Array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* **102**, 228–239 (1997).
85. Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* **16**, 1661–1687 (2004).
86. Kadir, S. N., Goodman, D. F. M. & Harris, K. D. High-dimensional cluster analysis with the masked EM algorithm. *Neural Comput.* **26**, 2379–2394 (2014).
87. Hazan, L., Zugaro, M. & Buzsáki, G. Klusters, NeuroScope, NDManager: A free software suite for neurophysiological data processing and visualization. *J. Neurosci Methods.* **155**, 207–216 (2006).
88. Meyers, E. M. The neural decoding toolbox. *Front Neuroinformatics* **7**, 8 (2013).
89. Zhang, Y. et al. Object decoding with attention in inferior temporal cortex. *Proc. Natl Acad. Sci. USA* **108**, 8850–8855 (2011).

90. Mitra, P. & Bokil, H. *Observed Brain Dynamics*. (Oxford University Press, 2007). <https://doi.org/10.1093/acprof:oso/9780195178081.001.0001>

Acknowledgements

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. This study was supported by the Max Planck Society. M.B. was partly supported by German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039B; and by the Machine Learning Cluster of Excellence, EXC number 2064/1 - Project number 390727645.

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 with input from T.I.P. 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 Figs. 6d–f, 7a, Supplementary Fig. 9a, 8a, d, 11. T.I.P. and N.K.L. supervised the study. N.K.L., T.I.P., and J.W. contributed unpublished reagents/analytical tools. N.K.L. provided support to the group. V.K. and T.I.P. wrote the original manuscript draft. All authors participated in the discussion and interpretation of the results and editing the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28897-2>.

Correspondence and requests for materials should be addressed to Vishal Kapoor or Theofanis I. Panagiotaropoulos.

Peer review information *Nature Communications* thanks Chris Klink and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Bistability of prefrontal states gates access to consciousness

Highlights

- vIPFC recordings during visual rivalry reveal frontal dynamics of conscious access
- Update and stability of conscious content in a tug of war between distinct LFP bands
- Changes in the encoding of conscious contents follow an increase in 1–9 Hz activity
- Content-agnostic prefrontal fluctuations gate sensory access to consciousness

Authors

Abhilash Dwarakanath, Vishal Kapoor,
Joachim Werner, Shervin Safavi,
Leonid A. Fedorov,
Nikos K. Logothetis,
Theofanis I. Panagiotaropoulos

Correspondence

abhilash.dwarakanath@tuebingen.mpg.de (A.D.),
theofanis.panagiotaropoulos@tuebingen.mpg.de (T.I.P.)

In brief

Mesoscale ensemble recordings in the vIPFC during no-report binocular rivalry show that stochastic prefrontal state fluctuations are associated with spontaneous changes in the content of consciousness. Transitions from beta band to low-frequency field activity precede both spontaneous changes in perception and encoding of the contents of consciousness by competing neuronal ensembles. Temporal dynamics of prefrontal state fluctuations during awake resting state mimic the temporal dynamics of binocular rivalry, suggesting a causal role in perceptual multistability.

Article

Bistability of prefrontal states gates access to consciousness

Abhilash Dwarakanath,^{1,4,6,*} Vishal Kapoor,^{1,5,6} Joachim Werner,¹ Shervin Safavi,^{1,2} Leonid A. Fedorov,¹ Nikos K. Logothetis,^{1,3,5,7} and Theofanis I. Panagiotaropoulos^{1,4,7,8,*}

¹Department of Physiology of Cognitive Processes, Max Planck Institute for Biological Cybernetics, Tübingen 72076, Germany

²International Max Planck Research School, Tübingen 72076, Germany

³Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, UK

⁴Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale, Commissariat à l'Energie Atomique et aux énergies alternatives, Université Paris-Saclay, NeuroSpin Center, 91191 Gif-sur-Yvette, France

⁵International Center for Primate Brain Research, Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Chinese Academy of Sciences, Shanghai, China

⁶These authors contributed equally

⁷Senior author

⁸Lead contact

*Correspondence: abilash.dwarakanath@tuebingen.mpg.de (A.D.), theofanis.panagiotaropoulos@tuebingen.mpg.de (T.I.P.)

<https://doi.org/10.1016/j.neuron.2023.02.027>

SUMMARY

Access of sensory information to consciousness has been linked to the ignition of content-specific representations in association cortices. How does ignition interact with intrinsic cortical state fluctuations to give rise to conscious perception? We addressed this question in the prefrontal cortex (PFC) by combining multi-electrode recordings with a binocular rivalry (BR) paradigm inducing spontaneously driven changes in the content of consciousness, inferred from the reflexive optokinetic nystagmus (OKN) pattern. We find that fluctuations between low-frequency (LF, 1–9 Hz) and beta (~20–40 Hz) local field potentials (LFPs) reflect competition between spontaneous updates and stability of conscious contents, respectively. Both LF and beta events were locally modulated. The phase of the former locked differentially to the competing populations just before a spontaneous transition while the latter synchronized the neuronal ensemble coding the consciously perceived content. These results suggest that prefrontal state fluctuations gate conscious perception by mediating internal states that facilitate perceptual update and stability.

INTRODUCTION

When the visual system is confronted with ambiguous sensory information, conscious perception spontaneously fluctuates between different possible perceptual interpretations.¹ In a stochastic manner, one of the competing representations temporarily gains access to consciousness while the others become perceptually suppressed, therefore dissociating sensory input from subjective conscious perception. This perceptual multistability is a gateway in understanding the mechanism that mediates the emergence of visual consciousness, due to the spontaneous switching of conscious perception between different co-registered representations.^{2–5}

Multistable visual perception, in particular binocular rivalry (BR), has been combined with direct neuronal recordings in the non-human primate (NHP) brain to study the neuronal correlates of consciousness (NCC). During BR, the content of consciousness spontaneously alternates between two disparate stimuli that are continuously presented to each eye.^{6–8} Neurons that correlate with the content of consciousness in BR should fire

more when their preferred stimulus is perceptually dominant and decrease their firing when it is perceptually suppressed, therefore providing an explicit representation of conscious contents. NHP electrophysiological studies have shown that such explicit representations are more reliably observed in association cortical areas like the prefrontal cortex (PFC) and temporal cortex compared with early visual areas.^{9–14} However, the exact neural mechanisms underlying the spontaneously driven passage of sensory input from non-conscious processing to conscious access, and vice versa, remain largely unknown.¹⁵ Understanding the mechanisms of spontaneous switching between different conscious contents could provide general insights for the emergence of conscious perception in the brain. In particular, the spontaneous nature of perceptual transitions in multistable perception could reveal the interplay between global cortical state fluctuations, which occur spontaneously during wakefulness and reflect fluctuations in internal states like arousal or attention,^{16–22} and the local ignition of content-selective neuronal representations, which has been associated with the access of sensory information to consciousness.^{23–25}

Two mechanisms associated with different cortical areas have been proposed to drive the spontaneous transitions in conscious content during BR: competition between monocular neurons (i.e., eye-specific input) in the primary visual cortex (V1) and a selection process between stimulus representations in higher-order associational cortical areas, in particular the PFC and parietal cortex (PPC).^{26–29} The originally proposed mechanism that involves competition among monocular V1 neurons may not be sufficient to explain rivalrous switches in conscious perception.¹³ This is because BR involves competition among higher-order perceptual representations of the rivaling stimuli that are not bound to eye-specific input,³⁰ and the percentage of neurons that correlate with conscious perception is higher in associational cortical areas.¹¹ By contrast, monocular neurons in V1 continue to fire, even when their preferred stimulus is perceptually suppressed.¹³ Moreover, switches in the activity of ocular dominance columns during BR can be observed in V1 during anesthesia³¹ or for invisible patterns,³² suggesting a dissociation of neural activity and consciousness in V1.

The second candidate mechanism that involves an active selection process among the rivaling stimulus representations in the PFC,^{1,33,34} where the contents of consciousness can be decoded from neuronal population activity,^{9,10,35} has not been systematically explored. This may possibly be implemented through spontaneous fluctuations in cortical states that modulate the ignition of content-specific neuronal representations in the PFC and therefore access of these representations to consciousness.^{36–39}

The concept of ignition is central to the global neuronal workspace theory (GNWT) of consciousness, which posits that access to consciousness is mediated by non-linear amplification (i.e., ignition) of neuronal activity in association cortical areas like the PFC and PPC. Ignition can be thought to result in crossing a threshold⁴⁰ that allows the broadcast of sensory information to widely distributed cortical networks,^{24,27} thus allowing it conscious access. Ignition associated with conscious perception has been classically observed as a sudden and sustained non-linear increase in gross-scale brain activity measured by functional magnetic resonance imaging (fMRI), electroencephalography (EEG), electrocorticography (ECoG), and magnetoencephalography (MEG), around 200–300 ms after the onset of stimuli that are consciously perceived.^{23,24,41–44} At the neuronal level, evidence for ignition comes from studies showing an amplification of activity in the frontal, parietal, somatosensory, and ventral premotor cortices at the single neuronal level, when sensory information enters consciousness.^{34,45}

Interestingly, the propagation of stimulus-driven activity from lower visual areas to the PFC, where (together with the PPC) the ignition of content-selective representations determines conscious access of sensory input, is influenced by the amplitude of global pre-stimulus states of neural activity.³⁴ Intuitively, such pre-stimulus states, known to fluctuate due to variability in internal states like arousal or attention,^{16,19} can be understood as reflecting a gating mechanism: an operation that could potentially control the threshold to conscious perception. For example, gating could be implemented through the bistability of intrinsically generated cortical states suggested to mediate active maintenance and update of information in the cortex.⁴⁶

However, to date, there is scant evidence as to what cortical states are the substrates of prefrontal gating that regulate ignition and enable access to consciousness.

In the present study, we attempted to unravel the mechanisms underlying the emergence of conscious perception in the macaque lateral PFC, using a no-report BR paradigm.⁹ This paradigm allowed us to detect internally driven transitions in the conscious perception of stimuli that moved in opposing directions. We combined this task with multi-electrode recordings of local field potentials (LFPs) and simultaneously sampled direction-of-motion-selective spiking activity of competing neuronal 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, thus identifying signals directly related to spontaneous transitions in the content of consciousness.

Our results suggest that spontaneously occurring, ongoing alternations in the prefrontal state between low-frequency (LF) (1–9 Hz) and beta (20–40 Hz) activity directly relate to the update and stability of prefrontal representations of conscious contents, respectively. We show that the amplitude of these LFP fluctuations in the PFC is content agnostic and precede both spontaneous OKN changes (signaling transitions in the content of consciousness) during BR and spontaneous ignitions of conscious contents, indicating that PFC activity is not a mere consequence of conscious perception. Furthermore, we find that these global fluctuations are also modulated on a local level through interactions with content-selective populations that depend on the activity state of these populations and therefore the perceptual dominance or suppression of their preferred stimulus. These results suggest that waking LFP state fluctuations in the PFC have a regulatory role in the ignition of conscious content representations within the same area, acting as a gate to conscious access.

RESULTS

We performed multi-electrode extracellular recordings of LFPs and direction-of-motion-selective neuronal ensembles in the inferior convexity of the macaque PFC (Figure 1A). The inferior convexity of the macaque ventrolateral PFC (vIPFC) is homologous to the human inferior frontal gyrus and has direct anatomical connections to the PPC,^{47,48} thus forming the “frontoparietal loop,” which is hypothesized to control access to consciousness.²⁷ Two trial conditions were employed: (1) physical alternation (PA) of monocularly alternating gratings with opposing directions of motion and (2) BR, where the initial direction of motion stimulus was not removed but was followed by a grating moving in the opposite direction, presented to the contralateral eye (Figure 1B). This manipulation results in an externally induced period of perceptual suppression for the first stimulus of variable duration (binocular flash suppression [BFS]), which is then followed by spontaneous perceptual transitions, since the two competing representations (upward vs. downward direction of motion) 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, which was previously shown to

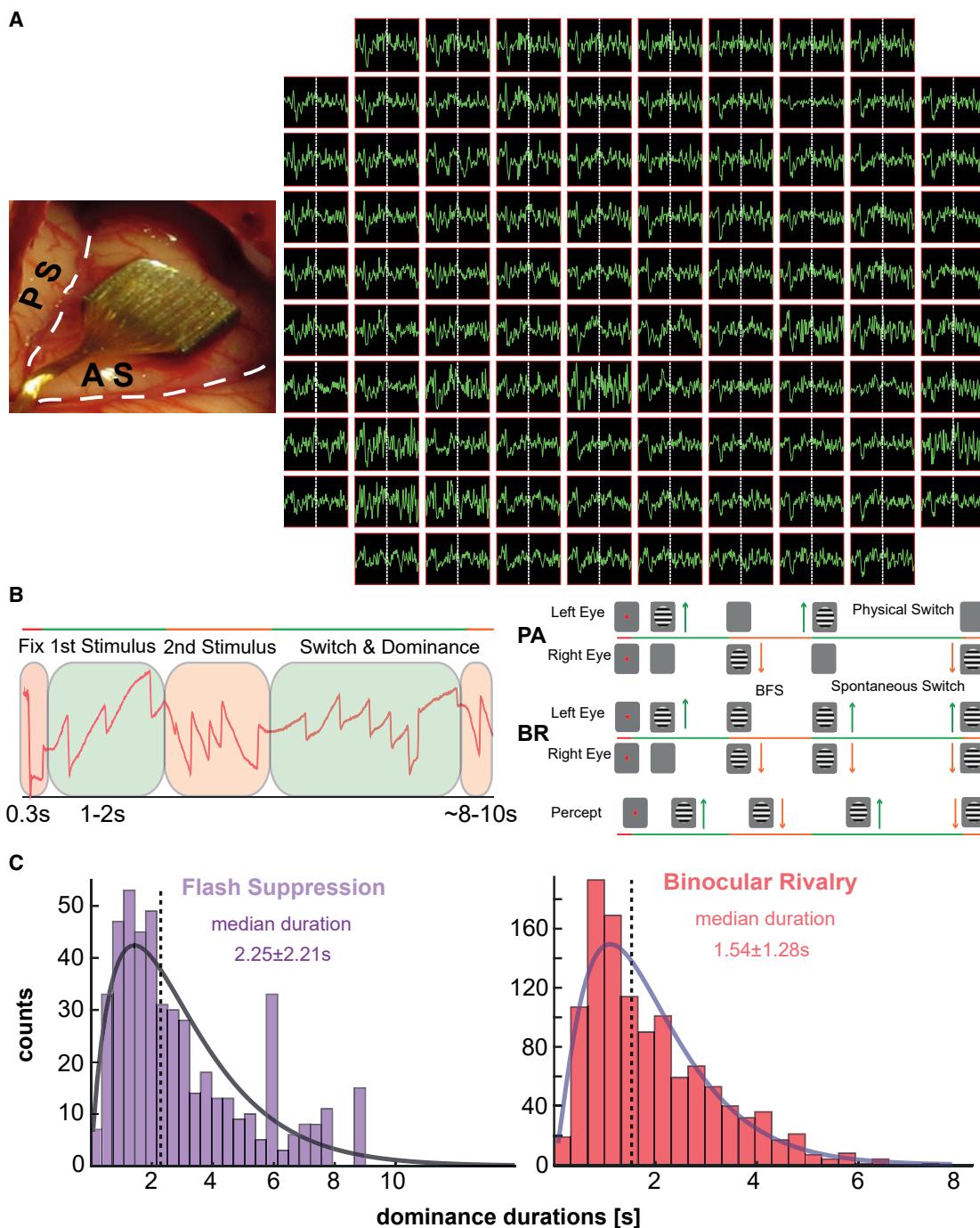


Figure 1. Experiment and typical LFP signals

- (A) Multi-electrode array in the inferior convexity of the PFC (top). AS, arcuate sulcus; PS, principal sulcus. Below, an example spatial map of 0.1–250 Hz LFP signals around a change in the OKN polarity (white line) signaling a spontaneous perceptual transition.
- (B) Task structure: in PA trials, after an initial phase of BFS, the two competing stimuli rivaled for 8–10 s, eliciting different OKN patterns (red trace, highlighted in green and orange). In PA, the stimuli were alternated monocularly.
- (C) Perceptual dominance durations during BFS and BR switches estimated from the OKN follow a classical gamma distribution.

provide an accurate (and faster, compared with manual responses) perceptual state readout in both humans and macaques,^{49–53} was used to infer perceptual dominance periods (**Figure 1B**). Here, by perceptual dominance periods, we mean periods for which perception was stable for one or the other stimulus, based on the OKN readout. These dominance durations followed a gamma distribution, a hallmark of multistable perception,^{1,54,55} with a median dominance duration of 1.54 ± 1.28 s (median \pm SD) for spontaneous transitions in BR and 2.25 ± 2.21 s for transitions involving exogenous perceptual suppression in BFS (**Figure 1C**).

Prefrontal state fluctuations precede spontaneous changes in conscious perception

To investigate the role of internal cortical states in the update and stability of conscious perception, we first analyzed LFP activity dynamics to identify mesoscopic signal fluctuations occurring around periods of externally induced (PA) and spontaneous (BR) perceptual transitions (**Figures 2A** and **S1**). Transient negative deflections of the channel-averaged, raw LFPs (0.1–250 Hz, blue traces in **Figure 2A**) were observed to disrupt activity 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 after the change in the OKN polarity, induced by external stimulus changes in PA (**Figure 2A**, left), but before the change in the OKN polarity that signals spontaneous perceptual transitions in BR (**Figure 2A**, right). The negative deflections were associated with increases in low-frequency (1–9 Hz, LF) activity (**Figure 2A**, bottom raw; blue vs. purple traces).

In individual BR transitions, a LF-associated beta (20–40 Hz) suppression started clearly before the spontaneous OKN change that signals a perceptual change (**Figure 2B**, right), but it manifested itself after the stimulus and OKN change in PA transitions (**Figure 2B**, left). Pooling all physical stimulus transitions in PA ($n = 1,322$) revealed that an increase in the LF power was concentrated immediately after exogenous stimulus changes and was also accompanied by a temporally transient suppression of the ongoing beta activity (**Figure 2C**, left). The LF-associated beta suppression was also observed for intrinsically generated perceptual transitions in BR ($n = 573$); however, it started well before (~400 ms) the spontaneous OKN change (**Figure 2C**, right). Due to the spontaneous nature of perceptual transitions in BR, the absence of a feedforward response, locked to an external change of the sensory input as in PA, resulted in a temporal jitter of the LF transients across different transitions (data not shown). However, spectrograms aligned to the LF event peaks detected before and after the OKN transitions in BR and PA, respectively, show a similarity in the coupling of LF activity and beta-burst suppression between the two conditions (**Figure S2**). Therefore, these results show that suppression of 20–40 Hz activity during 1–9 Hz transients follows exogenous stimulus changes in PA but precedes spontaneous OKN-inferred perceptual transitions in BR.

To clarify the relationship between the observed broadband transients (negative or positive) and the observed LF bumps in the spectrograms, we first computed the wavelet power spectrum around transitions. The normalized power spectrum (**Figure 3A**) before a spontaneous transition revealed a strong bump in the

beta range (20–40 Hz), indicating periodic, sustained oscillatory activity. The power in the LF band was weaker, yet deviated from a typical 1/f spectrum, thereby casting doubt on its oscillatory vs. aperiodic nature. To resolve this, we also treated the activity in LF as aperiodic and detected events in the same way. We then compared the timings of the detected broadband negative (and positive) transients and the detected LF events, as coincidences per time bin, approaching a transition. We find that these LF events are strongly correlated with negative broadband deflections, as compared with the positive transients (**Figures 3B** and **3C**), both in number and in time approaching a switch. Therefore, in all further analysis, we concentrate on the events and activity in this 1–9 Hz band, and resort to periodic analysis because simple event-based analysis is not amenable to answer complex questions of potential spike-LFP coding mechanisms.

First, we quantified the event rate of LF and beta activity, before and after the time of exogenous (PA) and endogenous (BR) perceptual transitions (described in STAR Methods; only the event rates for LF are reported). LF event rate (events/transition/channel) was significantly higher after the OKN change in PA (0.36 ± 0.0046 , $n = 46,495$ events, post-transition, vs. 0.09 ± 0.0014 , pre-transition, $n = 11,330$ events; $p < 10^{-187}$ mean \pm SEM; unless specified), but before the OKN change in BR (0.17 ± 0.002 , $n = 9,670$ events, pre-transition vs. 0.14 ± 0.002 , $n = 7,734$ events, post-transition; $p < 10^{-43}$ mean \pm SEM) (**Figure 4A**). Furthermore, the LF events occurred significantly more often before a spontaneous perceptual transition in BR than before a physical transition in PA (0.17 ± 0.002 , $n = 9,670$, pre-transition BR vs. 0.09 ± 0.0014 , $n = 11,330$, pre-transition PA; $p < 10^{-150}$). To understand the marginal (in value) yet significant difference between the average LF event time, before and after the spontaneous perceptual transitions in BR, one must consider that a post-switch period in BR could be in many instances the pre-switch period of the subsequent transition. To understand this relationship further, we discarded events occurring 250 ms after the OKN change, as this period encompasses the mean visually evoked potential (VEP) time in PA when aligned to the actual stimulus onset. After this correction, LF events occurred at -114 ± 190 ms (**Figure S4**, top row), thus further enhancing the pre- vs. post-switch difference.

Importantly, LF events occurred on average even before the end of the last OKN/perceptual dominance period preceding a spontaneous BR transition (end of perceptual dominance: -97.4 ± 140 ms, LF events collected up to the beginning of the next transition: -198 ± 133 ms, [median \pm SD] $p < 10^{-67}$, **Figure 4B**; see **Figure 2B** for an example transition). As expected, LF events occurred predominantly and significantly after the OKN change in PA (64 ± 147 ms). When these PA switches were aligned to the visual stimulus onset, the event times were further shifted (190 ± 162 ms, **Figure S3** middle row shows that the average latency between the physical stimulus onset and the change in the OKN is 190 ± 36.5 ms).

LF activation before a spontaneous perceptual switch in BR could also be observed in the evolution of event rate in time (quasi-peristimulus [event] time histogram [PSTH], to obtain an events/second/transition measure; **Figure 4C**). In BR, the peak rate of LF events occurred at -160 ± 237 ms and -28 ± 199 ms before the spontaneous perceptual transitions for the two

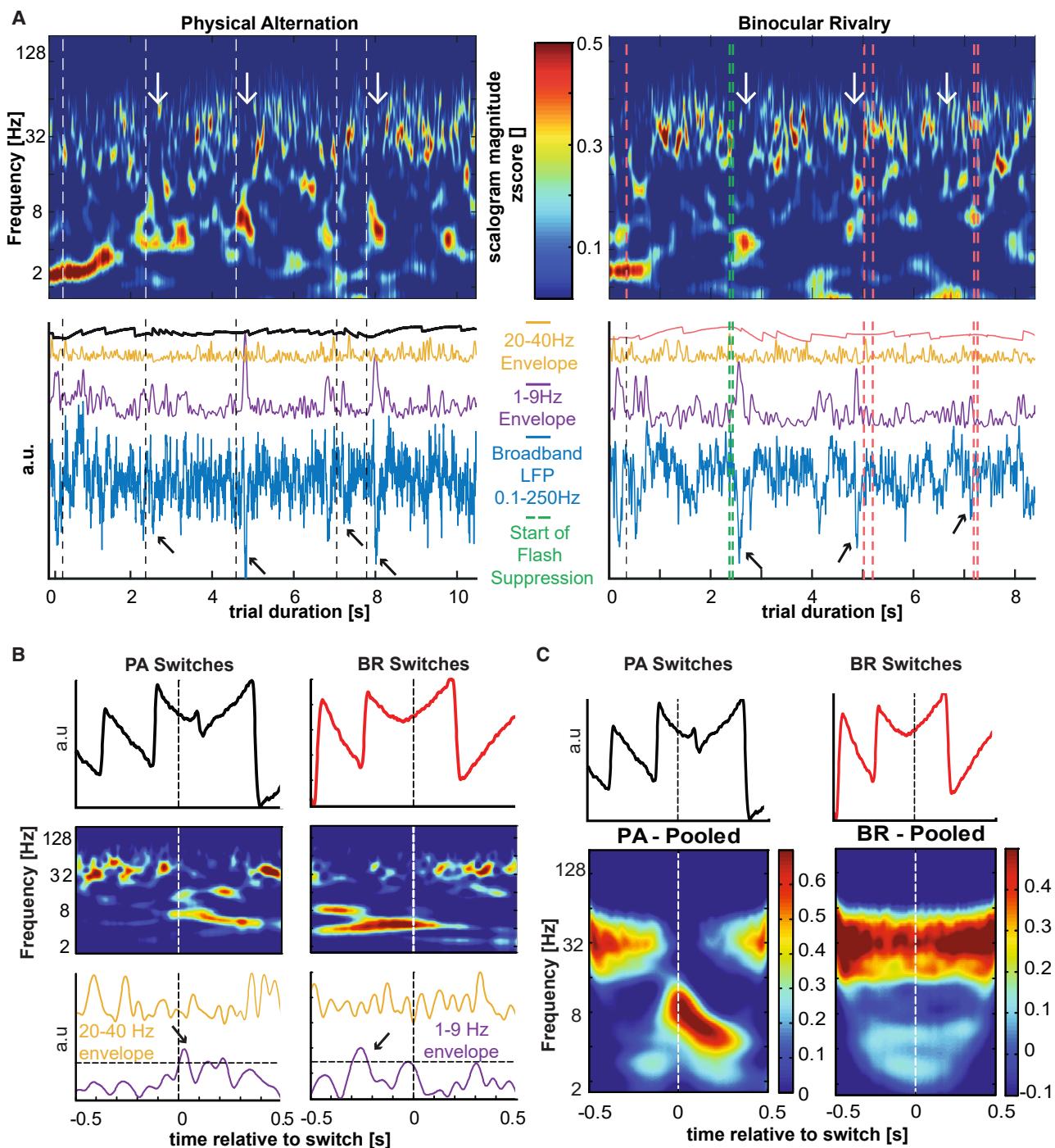


Figure 2. Time-frequency LFP analysis

(A) Z-scored scalograms for a single trial/observation period of PA (left) and BR (right) are shown at the top. White lines in PA reflect the manually marked change in the OKN polarity. Green lines in BR represent the start of the flash suppression phase, and red lines represent the subsequent spontaneous perceptual transitions. LF events occur after a switch in PA but before a switch in BR (white arrows). Bottom panels: broadband LFP (blue), LF amplitude (purple), beta-band amplitude (orange) and corresponding OKN traces. Black arrows point to negative deflections in the broadband trace.

(B) Upper panel: OKN traces around a single physical (black) and spontaneous (red) transition. Middle panel: Z-scored scalograms aligned to the OKN slope change for the two conditions. Lower panel: normalized instantaneous amplitudes of LF (purple) and beta (orange) activity.

(C) Grand average time-frequency analysis of spontaneous transitions (right) and an equal number of randomly sampled physical (left) and transitions. (Colorbar has the same units as A).

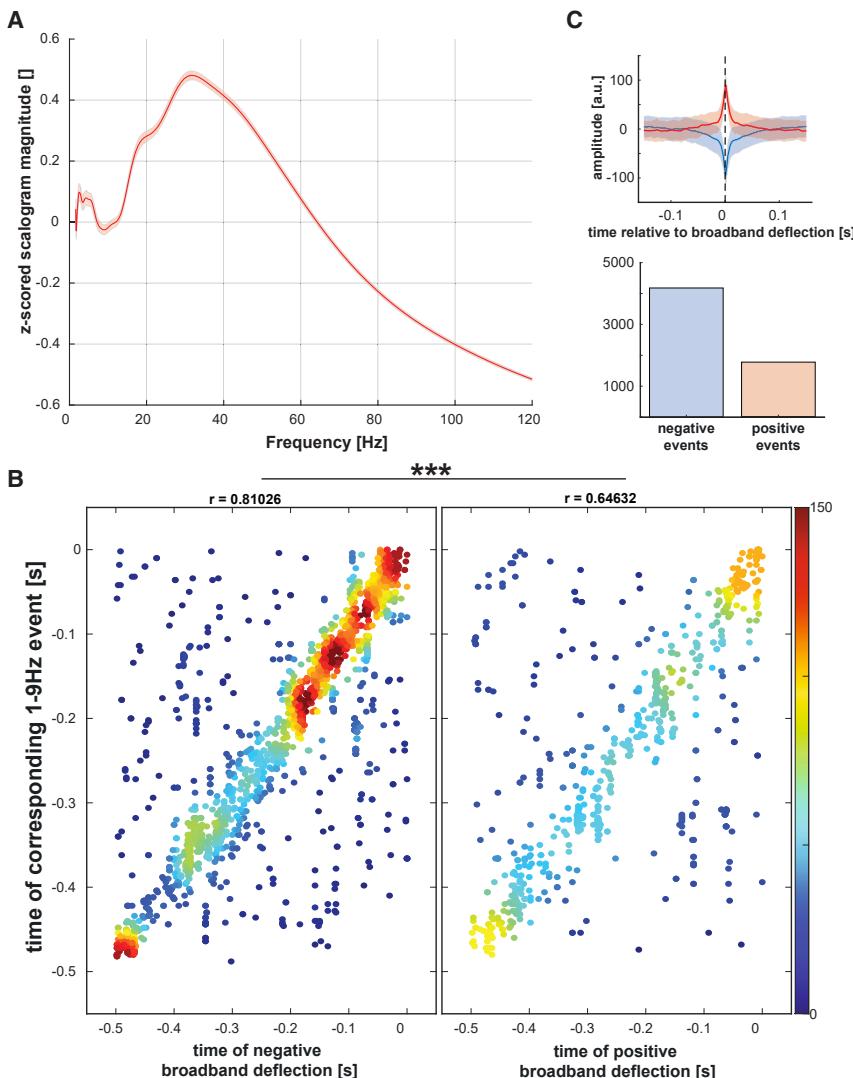


Figure 3. Correlation between 1–9 Hz events and negative broadband deflections

(A) Time-collapsed Z-scored scale-spectrum shows strong oscillatory activity in the beta band (center frequency of ~30 Hz) and weak, quasi-oscillatory/aperiodic activity in LF.

(B) Temporal coincidence density estimate between the detected LF events and negative broadband deviations yields a stronger correlation between LF events and negative transients, than with positive transients. The coincidence per bin for negative transients increases as one approaches a spontaneous switch in time.

(C) Mean positive and negative broadband deflection waveforms and their prevalence.

that in many instances before a spontaneous transition, the last transient LF (1–9 Hz) event was frequently preceded by similar but of lower amplitude events (Figures 5A and S1). Fitting a line to the relationship between the transition-averaged LF event 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 $R^2 = -0.003$) (Figure 5B). Next, we asked if there is a similar increase in the spatial pattern of LF amplitude increase during the period approaching an OKN change/perceptual switch. We counted the number of channels displaying instantaneous power in the LF band higher than the array mean (in 25 ms bins). Indeed, after a very brief non-linear increase, the number of participating sites increases linearly, approaching a switch (Figure 5C). Taken

together, these results indicate the occurrence of a mesoscopic, spatiotemporal spread of LF prefrontal events before spontaneous perceptual reversals operating at a linear scale.

LF activity is stronger before spontaneous transitions toward a clear percept

We further hypothesized that if an increase in LF activity is critical for inducing spontaneous perceptual reversals, then this activity should be significantly weaker when perceptual transitions were not complete but resulted in piecemeal (PM) periods. PM periods were detected in periods that the OKN did not change polarity after the end of a dominance period, therefore indicating that following the OKN change, perception did not unambiguously favor either of the two competing directions of motion. Subtracting the time-frequency decomposition of transitions to a PM percept from that of clean BR perceptual transitions revealed a preponderance of LF activity before a switch (Figure 6A). Additionally, the LF event rate was higher before a clear spontaneous transition compared with the period before transition to a PM

transition types, respectively (Figure 4C, top row), while in PA it occurred at 52 ± 28 ms and 82 ± 64.5 ms following the marked OKN change (Figure 4C, bottom row). Confirming the time-frequency analysis pattern in Figure 2A and suggesting a frequency-specific competitive process (i.e., cortical state fluctuations) in the PFC, the LF and beta event rates were significantly anti-correlated in BR ($r = -0.08$, $p = 0.0071$; pooled across both transition types; Figure 4D). We confirmed the significance of this negative correlation value by shuffling the beta amplitude at LF peak rate 1,000 times. The mean Pearson's correlation coefficient was no different from zero, with a p value of 0.493 (Figure S3, bottom row), suggesting that the small but negative correlation value is biologically relevant.

Spatiotemporal buildup of LF prefrontal activity

Are the LFP events preceding a spontaneous change in the content of consciousness, random large excursions from baseline activity, or do they reflect a gradual buildup process that is critical for inducing a spontaneous transition? Indeed, we noticed

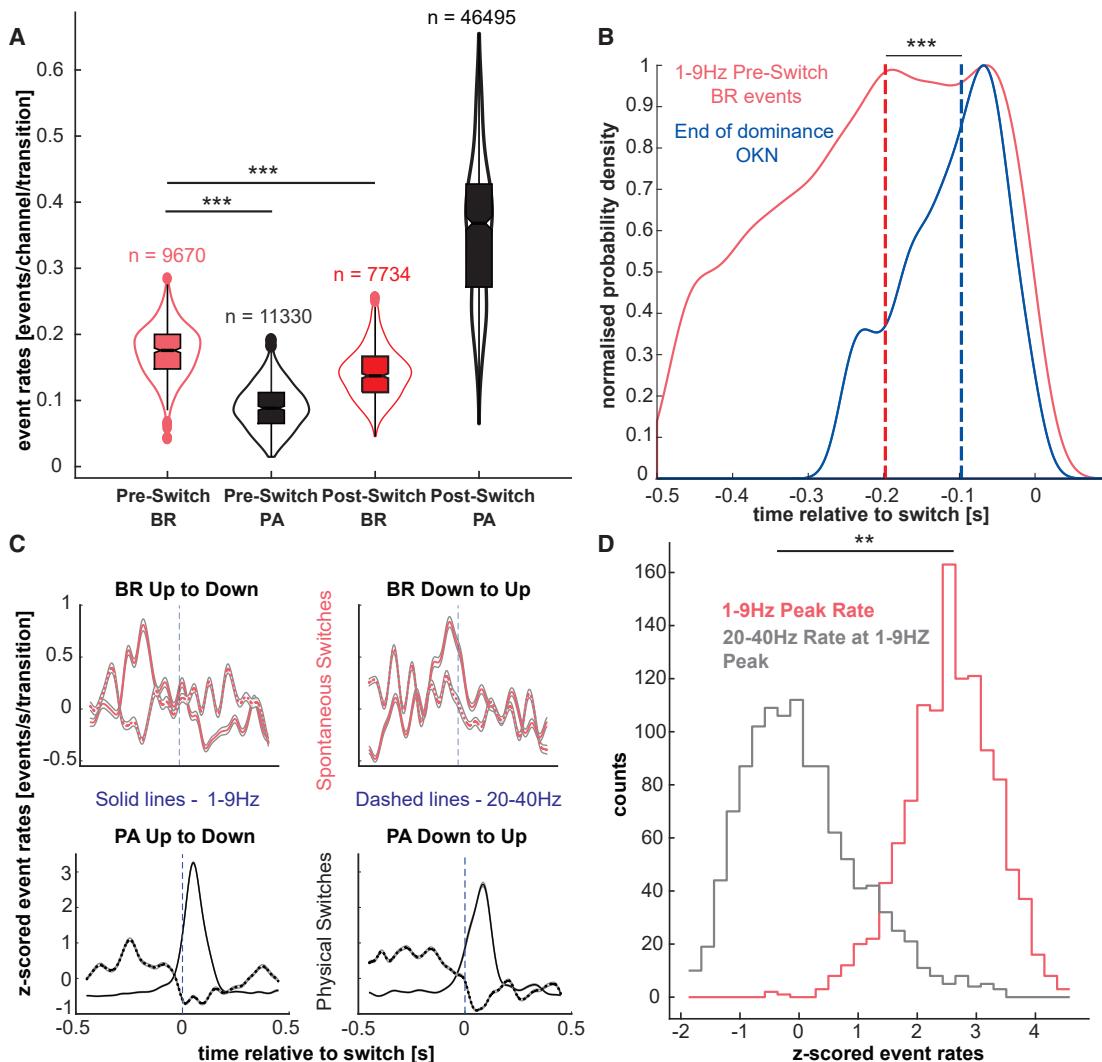


Figure 4. LFP event-rate analysis

(A) Burst rate per transition per channel 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 (gray). Matching color dots represent outliers. More LF events occur before spontaneous but after physical switches. Burst rate before a physical switch is low, suggesting noise levels.

(B) Normalized kernel density estimate of LF event and OKN times marking the end of the previous dominance period. LF events on average occurred before a switch (red dashed line), and significantly before the end of dominance.

(C) Z-scored event rate in time (events/s/transition) during BR (red lines) and PA trials (white lines) for LF (solid lines) and beta activity (dashed lines).

(D) Distribution of beta-band rate (gray) at the peak of LF rate (pink) before spontaneous perceptual transitions in BR showing antagonistic coupling.

$(0.17 \pm 0.0016, n = 9,670, \text{pre-BR vs. } 0.14 \pm 0.004, \text{pre-PM}, n = 2,486; p < 10^{-26})$ (Figure 6C), with the LF peak rate occurring after the transition to PM (Figure 6D). Moreover, the event rate was significantly higher after the transition in PM periods ($0.16 \pm 0.004, \text{PM}, n = 3,531 \text{ vs. } 0.14 \pm 0.004 \text{ pre-PM}, n = 2,486, \text{events/transition/channel}; p < 10^{-5}$), while the anti-correlation between LF and beta was significant but weaker, compared with clear spontaneous transitions ($r = -0.009, p < 10^{-137} \text{ vs. } r = -0.05; p < 10^{-295}, p < 10^{-14}$; Figures 4D and 6F), suggesting that the strength of LF and beta antagonism is critical for completing a perceptual transition to another period of clear dominance.

Next, we sought to ascertain if there is a difference between LF activity leading to a perceptual transition as compared with LF activity when perceptual transitions do not occur, i.e., during periods of sustained dominance. To accomplish this, we collected LFP activity around randomly triggered time points during these periods. Subtracting the time frequency decomposition of these sustained dominance periods from that of BR preserved the pattern observed in the latter (Figure 6B, compare with BR in Figure 2C). This indicates that weak LF activity occurs as baseline noise, which only leads to a perceptual change when it is spatiotemporally ramped up in a structured manner (Figure 6F vs. Figures 2A and 4C). Indeed, during

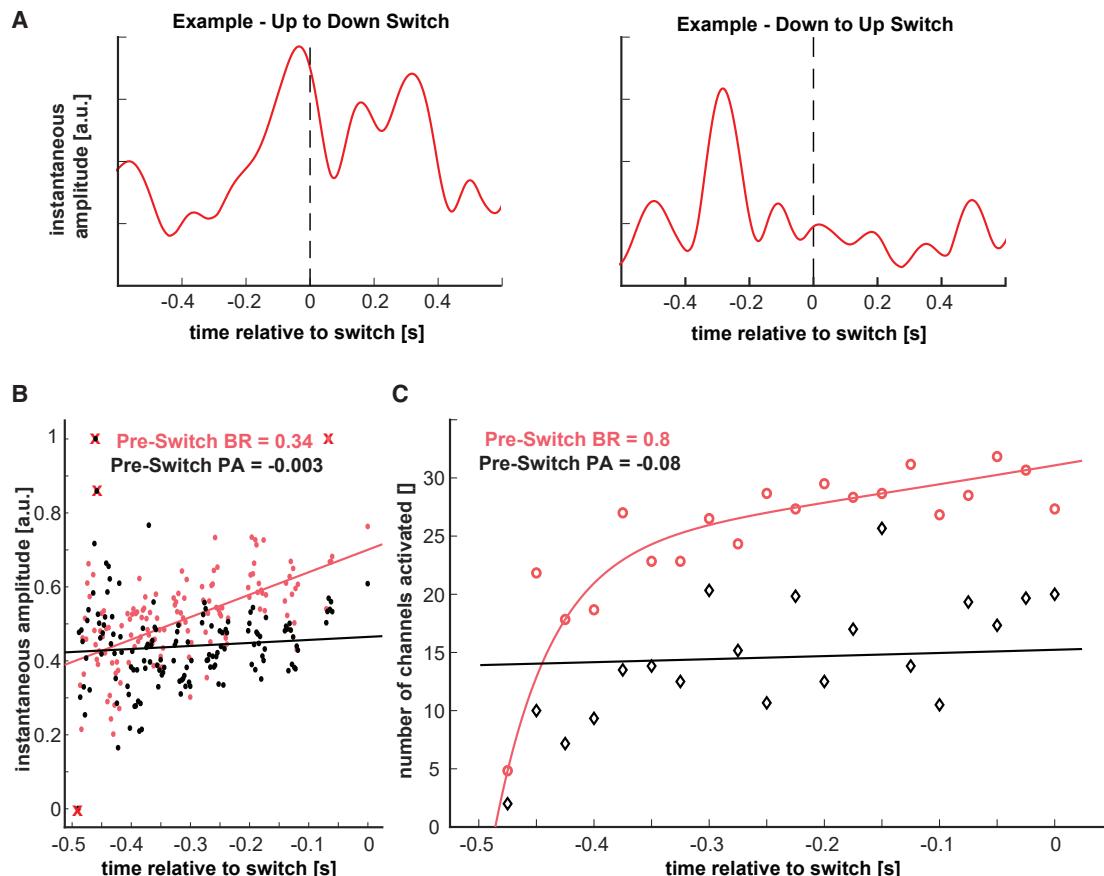


Figure 5. Spatiotemporal buildup preceding spontaneous transitions

(A) Examples of the two types of rivalrous transitions, viz. up to down and down to up, showing a steady increase in LF instantaneous power preceding a switch. (B) Average buildup of LF activity in time. While before BR switch, LF activity ramps up in time (slope = 0.61, $R^2 = 0.34$), before a physical transition, it remains flat (slope = 0.08, $R^2 = -0.003$). X denotes outliers omitted from fitting.

(C) The spatiotemporal spread of LF activity manifests in increasing number of prefrontal sites approaching a BR switch, but not a PA switch.

periods of sustained dominance we computed a mean LF event rate of 0.015 ± 0.0005 ($n = 55,026$ after resampling 100 times, i.e., only 550.26 events per iteration) events per sustained dominance period, an order of magnitude lower than the corresponding periods during BR (0.17 ± 0.0016 , $n = 9,670$, Figure 6C). Furthermore, the proportion of sites that displayed LF events approaching all BR switches was 100%, compared with only 51% during sustained dominance where nearly half the recorded sites were silent. These results further indicate that an increase in LF event rate, buildup, and a larger spatial spread of LF activation characterize spontaneous perceptual transitions. Specifically, LF activity should be significantly up modulated from noise level, potentially crossing both a power and spatial spread threshold, to induce a perceptual transition.

Content-specific modulation of global LFP states

Neuronal ensembles selective for the opposing directions of motion in the PFC show a distinct pattern in their firing during BR, where the firing rate follows the dominant or suppressed

direction of motion, depending on the preference of the ensemble (Figure 7A).⁹ However, LF LFPs collected from the electrodes that detected these ensembles do not show such a divergent pattern (Figure 7A). We confirmed this non-selectivity of the LF LFPs by using a support vector machine to decode representations of conscious contents (Figures 7B and S4). Conscious contents could be decoded from population spiking activity but not from LF or beta LFP activity. Therefore, prefrontal LFPs are content-agnostic signals reflecting fluctuations that are not coding for a particular rivaling stimulus. We also fit a piecewise linear function to the spiking activity of every transition with two degrees of freedom in order to detect change points that could point to the start of the change in encoding. This occurs at around -220 ± 10 ms median \pm SD before 0 and significantly after the rise (significantly after the start of rise in peak-rate and event-rate metric, Figures 4C and 7B; correlation data not shown) in LF activity (Figure 7A bottom row dashed line vs. Figure 7B red and pink distributions), confirming the decoupling of content-specific information from the LFPs.

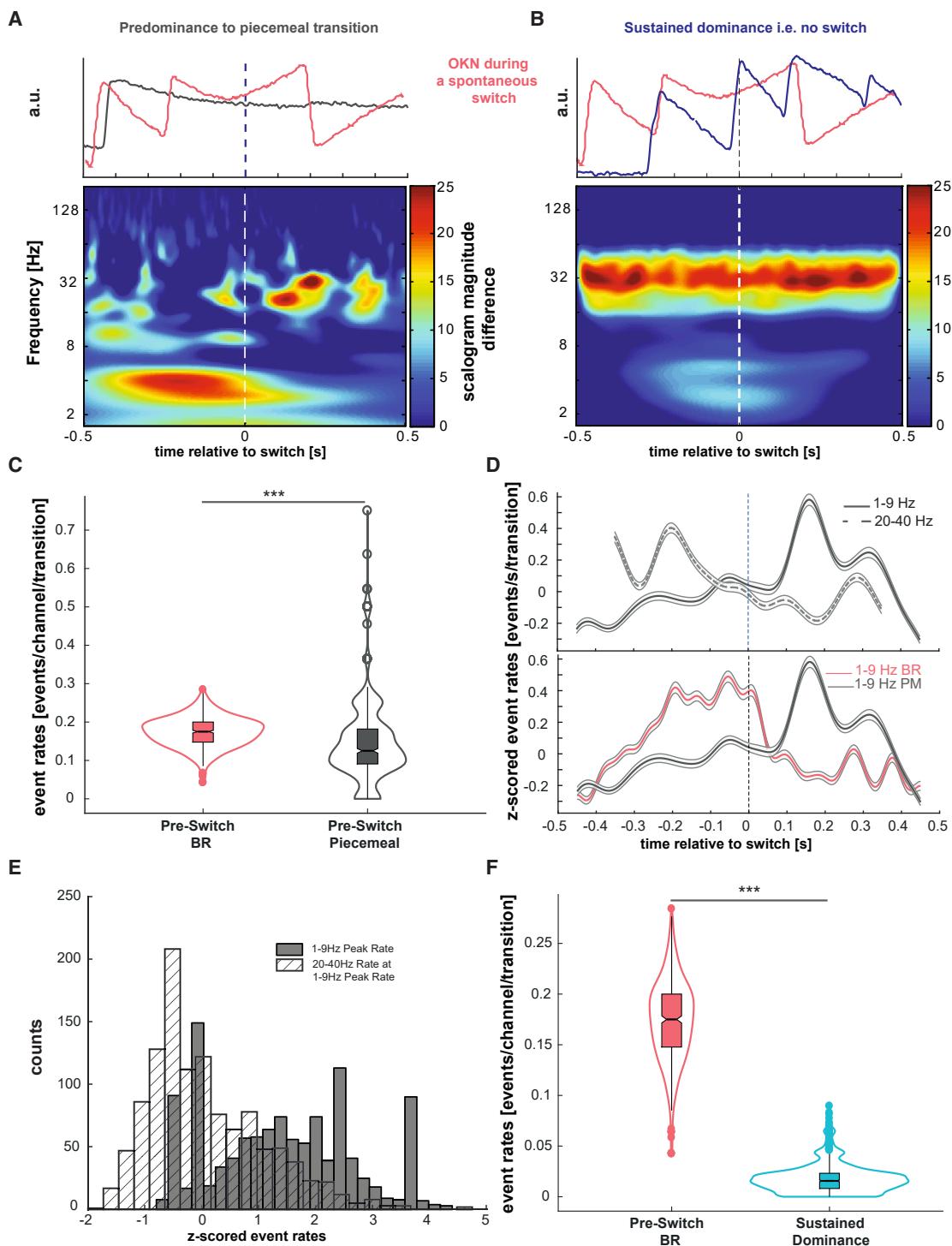


Figure 6. LF amplitude and event rate is critical for inducing clear switches

(A) Top panel: two typical OKN patterns elicited during a spontaneous transition (red) and during a transition to piecemeal (gray). The subtracted spectrogram (BR-PM) below shows a large difference in LF activity before a spontaneous switch, suggesting very weak LF activation before a failed switch.

(B) Top panel shows two typical OKN patterns: spontaneous transition (red) and sustained dominance, i.e., no switches (blue). The subtracted spectrogram preserves the pattern seen before BR, suggesting that LF activity during sustained dominance is minimal.

(C) LF event rate (per transition per channel) before BR is significantly higher than before transition to PM.

(legend continued on next page)

We also investigated whether the non-selective LF LFP activity precedes the change in the encoding of the active conscious percept from the spiking activity of neuronal populations. To understand this temporal relationship, we compared the convergence times of the normalized discharge activity of simultaneously recorded ensembles selective for the rivaling gratings (i.e., the point at which the stimulus-correlated firing rate of the ensembles selective to the dominant and suppressed stimuli flips and reaches saturation, providing a stable ongoing representation of the stimulus) (Figures 7B and S5; see STAR Methods) with the LF event- and peak-rate distributions. Relative to the OKN change signaling a BR switch, we found that discharge activity converged significantly later, compared with both the LF peak rates (-60 ± 222 ms for LFP event/s/transition vs. 209 ± 295 ms, convergence of spiking; $p < 10^{-47}$) and event times (-114 ± 190 ms median event time vs. 209 ± 295 ms, convergence of spiking; $p < 10^{-94}$). In the majority of spontaneous transitions, spiking activity crossovers occurred after the median truncated LF peak rates and event times (in 86.2% of trials compared with peak rates and in 89% of trials compared with event times). These results suggest that LF events, which do not provide an explicit representation of the current conscious content, reflect a pre-conscious process that precedes perceptual update.

To understand how prefrontal state fluctuations relate to spiking-network reorganization and therefore perceptual update and stability, we first plotted the spiking rasters of each transition overlaid with the LF instantaneous power (Figure S6) and the spiking rasters of each negative transient aligned to its trough. Qualitatively, no specific pattern of spike occurrence aligned to the LF power peaks or the negative transients were seen (Figures S6, S7, and S9). However, aligned to negative events detected in both non-preferred to preferred (NP to P) and preferred to non-preferred (P to NP) switch types, a significant increase in population firing rate is observed around the time of the event (Figures S7 and S8), with no significant differences at the peak between the two conditions. Furthermore, in both types of transitions, negative events increased approaching a spontaneous reversal, but the positive events did not (Figure 7C). Therefore, the negative deflections appear to be the most relevant signal that precedes spontaneous perceptual changes, irrespective of the ensemble from which the events are detected, viz. suppressed or dominant.

Because no differences between this activity in competing ensembles could be observed, and because of the quasi-periodic nature of the 1–9 Hz band (and its strong event correlation to observed negative deviations [Figures 2A and 3B]), we hypothesized that differential spike-locking to the ongoing LFP phase could result in spiking network reorganization, leading to crossovers in the observed switch-triggered PSTHs. We therefore performed spike-field coherence⁵⁶ (SFC) and pairwise phase consistency⁵⁷ (PPC) analysis for these competing populations. We computed the SFC in a 1 s window (cut off at 0.9 s to account for edge artifacts) preceding and succeeding a perceptual

switch of the simultaneously recorded, feature-selective ensemble activity and the global broadband LFP across all transitions. After a spontaneous perceptual transition in BR, when the negative LFP deflections and therefore the LF transients were less prevalent, the perceptually dominant ensemble was more coherent in the beta range (~25–37 Hz), compared with the suppressed ensemble (Figures 7D [top and bottom rows] and 7E; $p < 0.03$). However, there were no differences between the suppressed and dominant populations in the period approaching a spontaneous transition when LF transients accompany a suppression of beta bursts. These results suggest that the prefrontal ensemble signaling the current content of consciousness is synchronized in the beta band of the LFP. LF transients dissolve the beta-coherent ensemble, potentially increasing the likelihood for perceptual reorganization. Further evidence for this hypothesis is seen in the mean phase angles of spike-LFP coupling in the LF band (Figure 7F). Before a switch, sites that prefer the suppressed stimulus (i.e., NP to P switch) are locked to the depolarizing phase of the LFP (169.2°), while sites that prefer the dominant stimulus (i.e., P to NP switch) are locked to the hyperpolarizing phase (-147.6°, starting at ~750 ms before the switch, Figure 7F). Although inferring a mechanism from extracellular recordings is challenging, this selective locking could relate to the modulation of the membrane potential, thereby pushing the two ensembles closer to or farther away from the firing threshold, thus increasing or decreasing the firing probability, respectively.

Intrinsic nature of prefrontal state fluctuations

If the competition between LF and beta events that regulate access to visual consciousness is intrinsically generated, reflecting waking state fluctuations, traces of this process should also be observed during resting state—in the absence of any sensory (i.e., visual) input. Indeed, in resting state, LF events suppressed beta activity (Figure 8A). Periods of uninterrupted beta activity, cut off at 10 s to allow for a direct comparison with the psychophysical sample set (Figure 1C), exhibited a gamma distribution with a duration of 1.2 ± 1.44 s, which is close to the psychophysical distribution of stable perceptual dominance durations (1.54 ± 1.28 s) (Figure 8B). However, the full sample set showed a gamma distribution similar to the timescales observed in human bistable studies (data not shown). Therefore, prefrontal state fluctuations appear to reflect an intrinsic process critical to the maintenance and update of conscious information.

DISCUSSION

Our findings suggest that both spontaneous transitions in the content of conscious perception during BR and spontaneous ignitions of content-specific neuronal representations in the PFC are preceded by global (i.e., content-agnostic) LFP state fluctuations. Despite their global nature, these fluctuations are differentially coupled to neuronal populations on a local level. This

(D) Top row: before a transition to PM, beta activity dominates, still signaling the active percept. After the switch, LF inhibits beta, but weakly. Bottom row: LF rate peaks before transition to BR, but after a switch PM.

(E) Distribution of peak LF rates and rate of beta activity at LF peak reveal no significant antagonism before PM, compared with before BR.

(F) LF event rate before a BR switch is significantly higher than during sustained dominance.

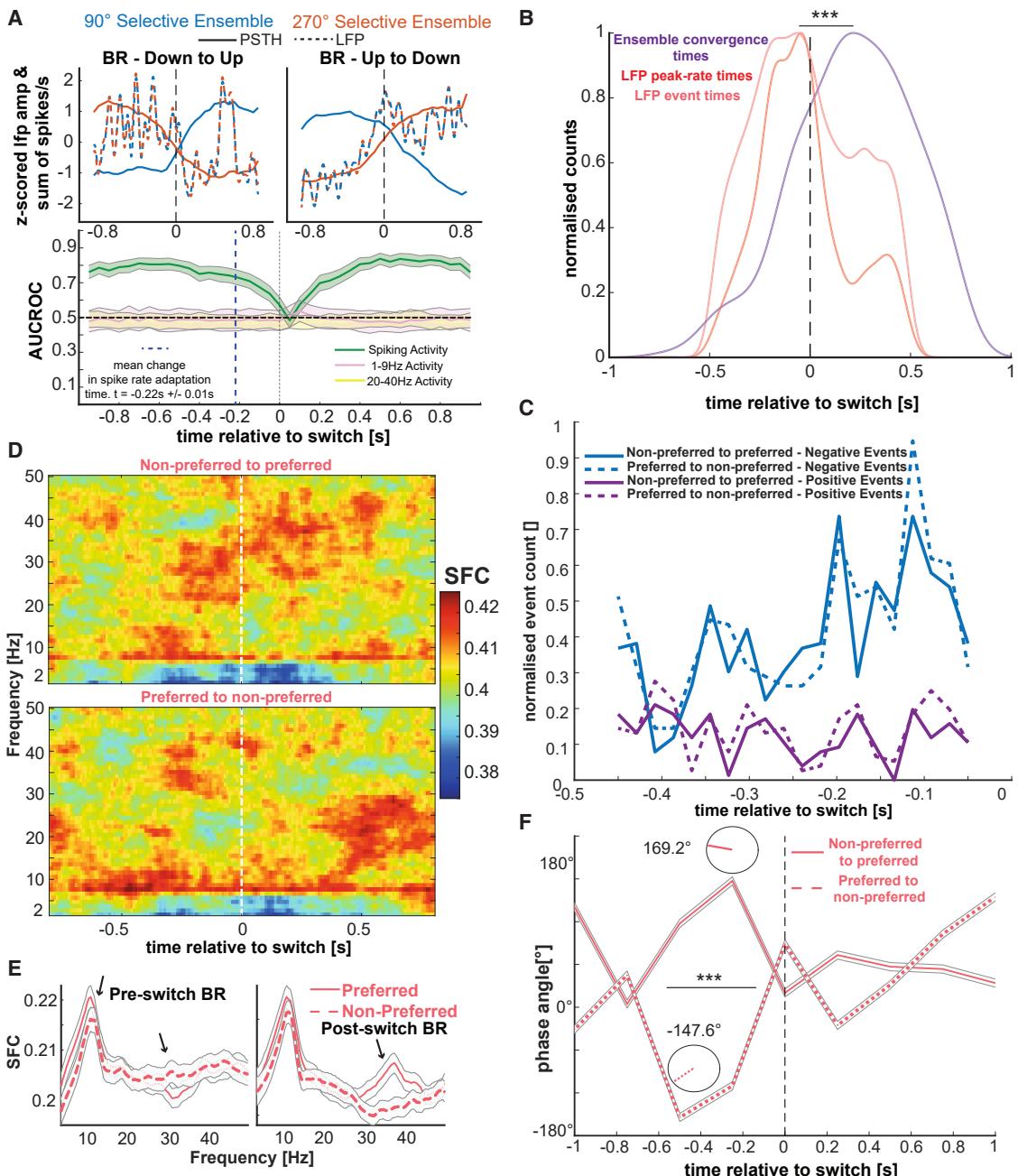


Figure 7. Relationship between the spiking activity of selective ensembles and the LFP

- (A) Top row: average population PSTHs of feature-specific neuronal ensembles during BR, showing divergence due to differential encoding, and LF activity, showing no dichotomy. Bottom row: ensemble selectivity (i.e., the content) can be decoded from population spiking, but not from LF or beta-band activity.
- (B) Distribution of LF event times (pink), peak-rate times (red), and convergence of PSTHs of competing ensembles (black line) demonstrates that the change in encoding of the conscious content follows LF activity.
- (C) Irrespective of the type of switch, i.e., the active ensemble, the number of negative events increases approaching a switch, but the number of positive events remains flat, pointing to the non-specific nature of LF activity.
- (D) Top and bottom rows: time-frequency (top) and time-collapsed representation of the spike-field coherence, for non-preferred to preferred (NP to P) and preferred to non-preferred (P to NP) switches, shows the dissolution of beta coherence and its reformation, depending on the switch type.
- (E) After a spontaneous transition, spiking of the dominant population was significantly more coherent with the beta LFP band, as compared with the suppressed population (black arrows show significant bins). SFC during pre-switch BR periods, when LF transient events are more prevalent, did not exhibit similar

(legend continued on next page)

coupling depends on whether neuronal populations are in an active state, signaling that their preferred stimulus is consciously perceived, or in an inactive state, signaling that their preferred stimulus is perceptually suppressed. Apart from indicating the role of internal state changes in conscious perception, these results suggest that prefrontal activity is not merely a post-perceptual consequence of conscious perception.^{58,59} Rather, prefrontal LFP state fluctuations precede spontaneous perceptual changes and coding of conscious content from neuronal ensembles in the same area. We discuss these findings in the context of the ongoing debate about the role of PFC in conscious perception, the role of internal state fluctuations in perception and consciousness, and the mechanism of spontaneous transitions in perception during BR.

PFC as part of the cortical network mediating conscious perception

The cortical organization of conscious perception, and in particular the role of PFC, is a topic of controversy since it differentiates theories of consciousness like the GNWT and integrated information theory (IIT).⁶⁰ Previous electrophysiological studies in the PFC have revealed representations of the content of consciousness^{10,34,61,62} as well as preparatory spiking activity before spontaneous perceptual changes.^{61,62} However, these activations could not be dissociated from signals related to voluntary motor reports that were used to indicate perceptual changes. Therefore, it has been suggested that conscious perception emerges in posterior cortical areas, and activity found to correlate with the contents of consciousness in the PFC may reflect only the post-perceptual consequences of conscious perception like correlates of reporting the percept or thinking about the stimulus.^{63–69}

However, our previous neuronal recordings in the NHP PFC suggest that conscious content representations can be decoded with very high accuracy from prefrontal populations, even in the absence of subjective reports.^{9,10,35,70} More recently, multivariate searchlight pattern analysis of fMRI recordings has also demonstrated decoding of conscious contents in frontal areas in both report⁷¹ and no-report conditions.⁷² If conscious contents can be decoded in the PFC, suggesting that PFC is part of a cortical network that represents conscious perception, what is the process that gates the ignition of these representations in the same area? Although ignition in the PFC and PPC has been suggested from GNW as the critical factor enabling conscious access,^{24,27,41,73} a gating mechanism that modulates ignition relying on spontaneous activity fluctuations could be the prerequisite, enabling factor for conscious perception.^{16,25,34,74}

Global state fluctuations precede spontaneous transitions in the content of consciousness

Recent work suggests that cortical state fluctuations during wakefulness are not just noise but rather reflect ongoing fluctu-

ations in attention or arousal states,^{16,19} with the PFC being a critical node for both.^{75–77} Here, we hypothesized that such waking state fluctuations detected in the prefrontal LFP activity could be involved in the spontaneous transitions in the contents of consciousness, which are characterized by ignitions of their neuronal representations in the PFC. We used a no-report BR paradigm, where spontaneous changes in the content of visual consciousness are inferred from spontaneous changes in reflexive eye movements (i.e., OKN). In doing so, we removed any confounds that could be the consequence of post-perceptual processes like voluntary motor reports. Our results suggest that dynamic changes in the prefrontal LFP activity precede spontaneous changes in conscious perception in the absence of voluntary report requirements. Specifically, an increase in LF (1–9 Hz) and a decrease in beta (20–40 Hz) LFP activity occur in the PFC before a spontaneous change in the OKN pattern, known to correlate with the perceived direction of motion.^{49–53} Using a decoding approach, we find that neither LF nor beta LFP amplitude represents the conscious contents but rather reflects content-agnostic fluctuations, which occur before a change in the OKN pattern and before the spontaneous ignition of conscious content representations, decoded from the spiking activity of neuronal populations in the PFC. Crucially, these fluctuations in cortical state are more prevalent in spontaneous transitions from one clear percept (i.e., direction of motion) to the other, compared with transitions from one direction of motion to a PM percept (as inferred from the OKN signal).

Global state fluctuations are locally modulated during conscious perception

Both LF and beta activity were not conveying information about the content of consciousness and were therefore reflecting global state changes that are not stimulus selective. Although pre-stimulus, non-selective, spontaneous fluctuations have been previously shown to influence visual perception of ambiguous stimuli or stimuli presented at threshold,^{74,78–83} it is unclear if and how such global state changes can influence spontaneous transitions of conscious perception between rivaling stimuli in a selective, content-specific manner. Recent theoretical studies have used a content-agnostic signal that relates to vigilance, which amplifies neuronal responses homogeneously to explain widespread ignition of neural activity relating to conscious perception for individual stimuli presented at threshold.⁸⁴ However, a content-selective modulation may be critical when neuronal representations compete for conscious access like in BR. Recently, there is indeed evidence that global cortical states can be controlled or modulated locally during spatial attention.^{85,86} Our findings point to a similar fine-scale modulation of global LFP state fluctuations during BR. While fluctuations in the prefrontal state were content agnostic since it was not possible to decode the content of consciousness from the amplitude of the LF and beta LFPs, these two states

differences in the beta band. These effects in the beta band were absent in physical transitions where SFC for a dominant preferred stimulus was significantly reduced in a lower-frequency range.

(F) Mean angles of spike-LFP phase locking during an NP to P (dashed line) and P to NP (solid line) transition. The populations selective to the suppressed stimulus before a switch progressively become locked to the depolarizing phase of the LF LFP (169.2°), thereby causing them to increase their firing rate when their preferred stimulus becomes dominant after the perceptual transition.

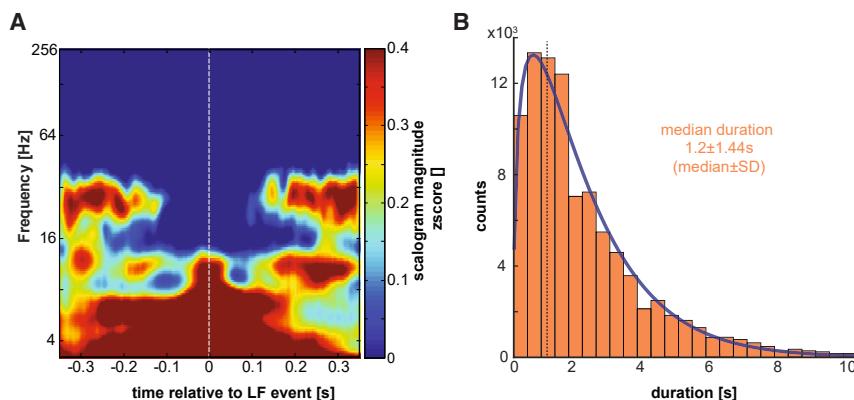


Figure 8. Intrinsic states during resting state show dynamics similar to perceptual transitions

(A) Grand-average ($n = 480$), LF event-triggered spectrogram during resting state showing cortical state fluctuations between LF and beta activity in the absence of structured sensory input.

(B) Periods of sustained beta activation during resting state. The median duration is remarkably similar to the psychophysical gamma distribution (Figure 1C).

interacted differentially with the competing neuronal populations. Specifically, global fluctuations were found to be dynamically modulated on a local level since (1) conscious and non-conscious neuronal representations were locked differentially to the phase of LF activity just before a spontaneous perceptual transition and (2) neuronal populations were coherent with beta activity only when their preferred stimulus was perceptually dominant (i.e., consciously perceived) during BR. These results, pointing to the existence of both global state fluctuations and their local modulation influencing conscious perception during BR in a mesoscopic scale, are in agreement with recent MEG studies showing evidence for both a general pre-stimulus process that is agnostic for the stimulus category and a specific pre-stimulus process that facilitates category-specific recognition.⁷⁴

Beta LFP state reflects perceptual stability during BR

What is the role of beta bursts in the PFC during BR? Beta activity has been suggested to reflect an intrinsic mode of cortical operation that shields ongoing behavioral and cognitive or sensory processing states (status quo) from interference and distractors.^{87–89} As a consequence, transient decreases in cortical beta activity could increase sensory information relay,^{4,11,90–93} providing a mechanism for controlling bottom-up sensory processing through top-down knowledge⁹² that can be content specific.⁹¹

Our findings are in agreement with such a role for beta activity during BR. Beta activity was prominent during periods of perceptual stability, in agreement with scalp EEG and MEG recordings.^{94,95} In particular, neuronal populations were coherently firing in the beta range when their preferred direction of motion was perceived. Dissolution of these content-specific, beta-coherent ensembles from low-frequency activity occurred before spontaneously generated changes in the OKN, signaling perceptual updates. Therefore, beta and LF LFP activity compete in the PFC, reflecting a competition between perceptual stability and updates, respectively. This result is also reminiscent of content (i.e., rule)-selective prefrontal ensembles that are coherent in the beta band with LF activity inhibiting a rule that is about to be deselected.^{96,97} This suggests that conscious perception and cognitive control might recruit similar processes in the PFC.

Prefrontal state fluctuations and gating of access to consciousness

We suggest that these findings could reflect a gating-like mechanism,^{46,98,99} where intrinsically generated fluctuations in the prefrontal cortical state between periods of LF and beta events gate the access of competing perceptual representations to consciousness. Specifically, the disruption of intrinsically generated beta activity by LF activity could reflect the process (perhaps related to fluctuations in arousal or attention) that controls the prefrontal ignition mechanism that has long been hypothesized to gate or control access to consciousness.^{34,73,100–102}

Importantly, we show here that the LF activation occurs earlier compared with both the spontaneous changes in OKN, which are used to infer spontaneous perceptual changes, and the activity of content-selective ensembles. Therefore, it appears that the driver of perceptual reorganization and update might not be the spiking activity of selective neuronal ensembles (which however do show fluctuations before a spontaneous transition); rather, it is a global state signal. Interestingly, during competition between two content representations in BR, these global fluctuations are modulated locally, providing a potential mechanism for content-specific spontaneous changes in conscious perception. This top-down mechanism of perceptual reorganization that has been proposed to drive multistable perception¹³ is fundamentally different from bottom-up mechanisms, proposing that competition between monocular neurons in the V1 resolves BR.^{13,26} Neuronal activity in V1 is indeed only weakly modulated during BR in both monocular and binocular neurons,¹³ while BOLD modulation of V1 is detected in superficial layers, suggesting feedback from higher cortical areas.¹⁰³ Furthermore, ocular dominance columns in V1 rival even during anesthesia,³¹ indicating that V1 activity is not alone sufficient for conscious visual perception.

Limitations of the study

We recorded neural activity only in the PFC and not in other cortical regions. It is therefore unclear whether the prefrontal state change preceding spontaneous transitions is an isolated PFC phenomenon or whether it can be observed or originates in other regions. For example, the anterior insula was recently shown using fMRI to have a gating role in conscious perception by regulating transitions between the default mode and dorsal

attention network.¹⁰⁴ Furthermore, whether the LFP state change before spontaneous perceptual transitions is also observed in the PPC (part of the frontoparietal loop in GNW) is unknown. Multi-site recordings during paradigms of conscious visual perception may help to address the spatial extent and source of these internal state fluctuations.

Another limitation of our study is the difficulty in establishing a firm causal relationship between global changes in the LFP signal and spiking activity, in particular between the LF activity increase and the adaptation in spiking activity that both seem to occur before a spontaneous perceptual transition. The co-occurrence of these processes might indicate that both contribute to a spontaneous perceptual switch. Indeed, theoretical studies have shown that spontaneous changes in BR can be explained as the interplay between deterministic (adaptation and inhibition) and stochastic (noise) forces. In these models, adaptation acts to reduce the inhibition between two pools of neurons that compete for perceptual dominance^{26,105} up to a point where noise drives the transitions.^{106–108} Specifically, adaptation acts to reduce the energy required to transition from one (neuronal) well to another. It is possible that our results reflect these dynamics with both adaptation in encoding and internal noise (i.e., spontaneous LFP state fluctuations) inducing the final transition. We should note, however, that noise (which might relate to internal state fluctuations) is the critical factor in these models since adaptation alone cannot explain the temporal dynamics during BR.

Finally, our study explains the findings based on the GNWT, where conscious perception relates to access consciousness that is characterized by ignition of neural activity and sensory information being broadcasted in a wide cortical network that includes the frontoparietal network.^{24,27} By contrast, IIT suggests that consciousness has a shorter timescale,¹⁰⁹ relies more on posterior cortical areas,⁶⁵ and is characterized by phenomenal experience.¹¹⁰ Despite these theoretical differences, visual stimuli can be decoded from early (~50 ms) neuronal activity in both the PFC and PPC, even under rapid serial visual presentation conditions that might weaken conscious access, suggesting that phenomenal consciousness may also be detected in areas assumed to be involved only in access consciousness.³⁵ Although beyond the scope of this study, the influence of prefrontal state fluctuations in phenomenal vs. access consciousness, and more generally in the predictions of different theories of consciousness, could be a promising field of future investigation.

Conclusions

Spontaneous cortical activity can attain various states during wakefulness that reflect variations in arousal and attention states, replay of sensory-driven activity, or multi-dimensional representations of behaviors and context.^{16,111,112} In the PFC, we observed a suppression of ongoing beta events during LF transients that precedes spontaneous changes in conscious perception. The same pattern was also observed during periods of resting state. This suggests that the source of spontaneous transitions in the content of consciousness may be the interaction of visual input representations with ongoing, waking state fluctuations. Taken together, our results reveal a potential role of prefrontal state fluctuations in a gating process that mediates the emergence of conscious perception.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Electrophysiological recordings
 - Data acquisition, spike sorting and local field potentials
 - Visual stimulation and experimental paradigm
 - Detection of spontaneous transitions
 - Treatment of the LFP data
 - Construction of direction of motion specific neural ensembles
 - Spike-field Coherence
 - Treatment of resting-state activity
 - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.neuron.2023.02.027>.

ACKNOWLEDGMENTS

We thank Dr. Yusuke Murayama and the other technical and animal care staff of the MPI for Biological Cybernetics for excellent technical assistance, Prof. Dr. Nicho Hatsopoulos for help with the implantations of the Utah arrays, Prof. Dr. Stanislas Dehaene for his inputs and insights, and Mr. Akshat Jain for his assistance in the preparation of publishable quality figures. We would also like to thank Dr. Michel Besserve, Prof. Dr. B.S. Dwarakanath, and Dr. Majid Khalili-Ardali for helpful discussions and comments in a previous version of the manuscript. The Max Planck Society supported this study. The Templeton World Charity Foundation, Inc. (TWCF0562) currently supports T.I.P. N.K.L. and V.K. presently receive funding through the Shanghai Municipal Science and Technology Major Project (grant no. 2019SHZDZX02).

AUTHOR CONTRIBUTIONS

Conceptualization, A.D., V.K., N.K.L., and T.I.P. (lead); data curation, A.D. (lead), V.K., and J.W.; formal analysis, A.D. (lead), V.K., J.W., and L.A.F.; funding acquisition, N.K.L.; investigation, A.D. (equal), V.K. (equal), and T.I.P. (supporting); methodology, A.D. (equal), V.K. (equal), J.W. (supporting), S.S. (supporting), and T.I.P. (equal); project administration, T.I.P.; resources, J.W. and N.K.L. (lead); software, A.D. (lead), V.K., J.W., L.A.F., and S.S. (supporting); supervision, T.I.P.; visualization, A.D. (lead) and T.I.P. (supporting); writing – original draft, A.D. and T.I.P. (lead); writing – review & editing, A.D., V.K., L.A.F., S.S., N.K.L., and T.I.P. (lead).

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 10, 2022

Revised: October 24, 2022

Accepted: February 16, 2023

Published: March 14, 2023

REFERENCES

1. Leopold, D.A., and Logothetis, N.K. (1999). Multistable phenomena: changing views in perception. *Trends Cogn.Sci.* 3, 254–264. [https://doi.org/10.1016/s1364-6613\(99\)01332-7](https://doi.org/10.1016/s1364-6613(99)01332-7).
2. Seth, A.K., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 439–452. <https://doi.org/10.1038/s41583-022-00587-4>.
3. Blake, R., Brascamp, J., and Heeger, D.J. (2014). Can binocular rivalry reveal neural correlates of consciousness? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130211. <https://doi.org/10.1098/rstb.2013.0211>.
4. Maier, A., Panagiotaropoulos, T.I., Tsuchiya, N., and Keliris, G.A. (2012). Introduction to research topic - binocular rivalry: a gateway to studying consciousness. *Front. Hum. Neurosci.* 6, 263. <https://doi.org/10.3389/fnhum.2012.00263>.
5. Logothetis, N.K. (1998). Single units and conscious vision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1801–1818. <https://doi.org/10.1098/rstb.1998.0333>.
6. Blake, R., and Logothetis, N.K. (2002). Visual competition. *Nat. Rev. Neurosci.* 3, 13–21. <https://doi.org/10.1038/nrn701>.
7. Clifford, C.W.G. (2009). Binocular rivalry. *Curr. Biol.* 19, R1022–R1023. <https://doi.org/10.1016/j.cub.2009.09.006>.
8. Blake, R. (2001). A primer on binocularrivalry, includingcurrentcontroversies. *BrainMind* 2, 5–38.
9. Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T.I., and Logothetis, N.K. (2022). Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports. *Nat. Commun.* 13, 1535. <https://doi.org/10.1038/s41467-022-28897-2>.
10. 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. <https://doi.org/10.1016/j.neuron.2012.04.013>.
11. 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. <https://doi.org/10.1098/rstb.2013.0534>.
12. Sheinberg, D.L., and Logothetis, N.K. (1997). The role of temporal cortical areas in perceptual organization. *Proc.Natl.Acad.Sci. USA* 94, 3408–3413. <https://doi.org/10.1073/pnas.94.7.3408>.
13. Leopold, D.A., and Logothetis, N.K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549–553. <https://doi.org/10.1038/379549a0>.
14. Logothetis, N.K., and Schall, J.D. (1989). Neuronal correlates of subjective visual perception. *Science* 245, 761–763. <https://doi.org/10.1126/science.2772635>.
15. Sterzer, P., Kleinschmidt, A., and Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn.Sci.* 13, 310–318. <https://doi.org/10.1016/j.tics.2009.04.006>.
16. McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C.R., Tolias, A.S., Cardin, J.A., and McCormick, D.A. (2015). Waking state: rapid variations modulate neural and behavioral responses. *Neuron* 87, 1143–1161. <https://doi.org/10.1016/j.neuron.2015.09.012>.
17. Luczak, A., Barthó, P., and Harris, K.D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* 62, 413–425. <https://doi.org/10.1016/j.neuron.2009.03.014>.
18. Luczak, A., Barthó, P., and Harris, K.D. (2013). Gating of sensory input by spontaneous cortical activity. *J. Neurosci.* 33, 1684–1695. <https://doi.org/10.1523/JNEUROSCI.2928-12.2013>.
19. Harris, K.D., and Thiele, A. (2011). Cortical state and attention. *Nat. Rev. Neurosci.* 12, 509–523. <https://doi.org/10.1038/nrn3084>.
20. Ringach, D.L. (2009). Spontaneous and driven cortical activity: implications for computation. *Curr. Opin. Neurobiol.* 19, 439–444. <https://doi.org/10.1016/j.conb.2009.07.005>.
21. Avitan, L., and Stringer, C. (2022). Not so spontaneous: multi-dimensional representations of behaviors and context in sensory areas. *Neuron* 110, 3064–3075. <https://doi.org/10.1016/j.neuron.2022.06.019>.
22. Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science* 364, 255. <https://doi.org/10.1126/science.aav7893>.
23. Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., Andelman, F., Neufeld, M.Y., Kramer, U., Fried, I., and Malach, R. (2009). Neural “ignition”: enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 64, 562–574. <https://doi.org/10.1016/j.neuron.2009.11.001>.
24. Dehaene, S., and Changeux, J.P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. <https://doi.org/10.1016/j.neuron.2011.03.018>.
25. Moutard, C., Dehaene, S., and Malach, R. (2015). Spontaneous fluctuations and non-linear ignitions: twodynamicfaces of corticalrecurrentloops. *Neuron* 88, 194–206. <https://doi.org/10.1016/j.neuron.2015.09.018>.
26. Blake, R. (1989). A neural theory of binocular rivalry. *Psychol. Rev.* 96, 145–167. <https://doi.org/10.1037/0033-295X.96.1.145>.
27. Mashour, G.A., Roelfsema, P., Changeux, J.P., and Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron* 105, 776–798. <https://doi.org/10.1016/j.neuron.2020.01.026>.
28. Dehaene, S., Changeux, J.-P., and Naccache, L. (2011). The global neuronal workspace model of conscious access: from neuronal architectures to clinical applications. In *Characterizing Consciousness: From Cognition to the Clinic? Research and Perspectives in Neurosciences*, S. Dehaene and Y. Christen, eds. (Springer), pp. 55–84. https://doi.org/10.1007/978-3-642-18015-6_4.
29. Metzger, B.A., Mathewson, K.E., Tapia, E., Fabiani, M., Gratton, G., and Beck, D.M. (2017). Regulating the access to awareness: brainactivityrelated to probe-related and spontaneousreversals in binocularrivalry. *J. Cogn. Neurosci.* 29, 1089–1102. https://doi.org/10.1162/jocn_a_01104.
30. Logothetis, N.K., Leopold, D.A., and Sheinberg, D.L. (1996). What is rivalry during binocular rivalry? *Nature* 380, 621–624. <https://doi.org/10.1038/380621a0>.
31. Xu, H., Han, C., Chen, M., Li, P., Zhu, S., Fang, Y., Hu, J., Ma, H., and Lu, H.D. (2016). Rivalry-like neuralactivity in primaryvisualcortex in anesthetizedmonkeys. *J. Neurosci.* 36, 3231–3242. <https://doi.org/10.1523/JNEUROSCI.3660-15.2016>.
32. Zou, J., He, S., and Zhang, P. (2016). Binocular rivalry from invisible patterns. *Proc.Natl.Acad.Sci. USA* 113, 8408–8413. <https://doi.org/10.1073/pnas.1604816113>.
33. Lumer, E.D., Friston, K.J., and Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science* 280, 1930–1934. <https://doi.org/10.1126/science.280.5371.1930>.
34. van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P.R. (2018). The threshold for conscious report: signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542. <https://doi.org/10.1126/science.aar7186>.
35. Bellet, J., Gay, M., Dwarakanath, A., Jarraya, B., van Kerkoerle, T., Dehaene, S., and Panagiotaropoulos, T.I. (2022). Decoding rapidly presented visual stimuli from prefrontal ensembles without report nor post-perceptual processing. *Neurosci. Conscious.* 2022, niac005. <https://doi.org/10.1093/nc/niac005>.
36. Levinson, M., Podvalny, E., Baete, S.H., and He, B.J. (2021). Cortical and subcortical signatures of conscious object recognition. *Nat. Commun.* 12, 2930. <https://doi.org/10.1038/s41467-021-23266-x>.
37. Weinhammer, V., Fritsch, M., Chikermane, M., Eckert, A.L., Kanthak, K., Stuke, H., Kaminski, J., and Sterzer, P. (2021). An active role of inferior frontal cortex in conscious experience. *Curr. Biol.* 31, 2868–2880.e8. <https://doi.org/10.1016/j.cub.2021.04.043>.

38. Dehaene, S., and Changeux, J.P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLOS Biol.* 3, e141. <https://doi.org/10.1371/journal.pbio.0030141>.
39. 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, e6142. <https://doi.org/10.1371/journal.pone.0006142>.
40. Kang, Y.H.R., Petzschner, F.H., Wolpert, D.M., and Shadlen, M.N. (2017). Piercing of consciousness as a threshold-crossing operation. *Curr. Biol.* 27, 2285–2295.e6. <https://doi.org/10.1016/j.cub.2017.06.047>.
41. Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B., and Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758. <https://doi.org/10.1038/89551>.
42. Noy, N., Bickel, S., Zion-Golumbic, E., Harel, M., Golan, T., Davidesco, I., Schevon, C.A., McKhann, G.M., Goodman, R.R., Schroeder, C.E., et al. (2015). Ignition's glow: ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Conscious. Cogn.* 35, 206–224. <https://doi.org/10.1016/j.concog.2015.03.006>.
43. Del Cul, A., Baillet, S., and Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* 5, e260. <https://doi.org/10.1371/journal.pbio.0050260>.
44. Joglekar, M.R., Mejias, J.F., Yang, G.R., and Wang, X.J. (2018). Interareal balanced amplification enhances signal propagation in a large-scale-circuit model of the primate cortex. *Neuron* 98, 222–234.e8. <https://doi.org/10.1016/j.neuron.2018.02.031>.
45. Noel, J.P., Ishizawa, Y., Patel, S.R., Eskandar, E.N., and Wallace, M.T. (2019). Leveraging nonhuman primate multisensory neurons and circuits in assessing consciousness theory. *J. Neurosci.* 39, 7485–7500. <https://doi.org/10.1523/JNEUROSCI.0934-19.2019>.
46. O'Reilly, R.C. (2006). Biologically based computational models of high-level cognition. *Science* 314, 91–94. <https://doi.org/10.1126/science.1127242>.
47. Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 781–795. <https://doi.org/10.1098/rstb.2005.1631>.
48. Petrides, M., and Pandya, D.N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* 228, 105–116. <https://doi.org/10.1002/cne.902280110>.
49. Fox, R., Todd, S., and Bettinger, L.A. (1975). Optokinetic nystagmus as an objective indicator of binocular rivalry. *Vision Res.* 15, 849–853. [https://doi.org/10.1016/0042-6989\(75\)90265-5](https://doi.org/10.1016/0042-6989(75)90265-5).
50. Logothetis, N.K., and Schall, J.D. (1990). Binocular motion rivalry in macaque monkeys: eye dominance and tracking eye movements. *Vision Res.* 30, 1409–1419. [https://doi.org/10.1016/0042-6989\(90\)90022-d](https://doi.org/10.1016/0042-6989(90)90022-d).
51. Naber, M., Frässle, S., and Einhäuser, W. (2011). Perceptual rivalry: reflexes reveal the gradual nature of visual awareness. *PLoS One* 6, e20910. <https://doi.org/10.1371/journal.pone.0020910>.
52. Fujiwara, M., Ding, C., Kaunitz, L., Stout, J.C., Thyagarajan, D., and Tsuchiya, N. (2017). Optokinetic nystagmus reflects perceptual directions in the onset binocular rivalry in Parkinson's disease. *PLoS One* 12, e0173707. <https://doi.org/10.1371/journal.pone.0173707>.
53. Aleshin, S., Ziman, G., Kovács, I., and Braun, J. (2019). Perceptual reversals in binocular rivalry: improved detection from OKN. *J. Vis.* 19, 5. <https://doi.org/10.1167/19.3.5>.
54. Zhou, Y.H., Gao, J.B., White, K.D., Merk, I., and Yao, K. (2004). Perceptual dominance time distributions in multistable visual perception. *Biol. Cybern.* 90, 256–263. <https://doi.org/10.1007/s00422-004-0472-8>.
55. Brascamp, J.W., van Ee, R., Pestman, W.R., and van den Berg, A.V. (2005). Distributions of alternation rates in various forms of bistable perception. *J. Vis.* 5, 287–298. <https://doi.org/10.1167/5.4.1>.
56. Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5, e1000436. <https://doi.org/10.1371/journal.pcbi.1000436>.
57. van der Meer, M.A.A., and Redish, A.D. (2009). Low and highgamma 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. <https://doi.org/10.3389/neuro.07.009.2009>.
58. Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* 36, 737–746. <https://doi.org/10.1016/j.neubiorev.2011.12.003>.
59. Block, N. (2020). Finessing the bored monkey problem. *Trends Cogn. Sci.* 24, 167–168. <https://doi.org/10.1016/j.tics.2019.12.012>.
60. Melloni, L., Mudrik, L., Pitts, M., and Koch, C. (2021). Making the hard problem of consciousness easier. *Science* 372, 911–912. <https://doi.org/10.1126/science.abj3259>.
61. Gelbard-Sagiv, H., Mudrik, L., Hill, M.R., Koch, C., and Fried, I. (2018). Human single neuron activity precedes emergence of conscious perception. *Nat. Commun.* 9, 2057. <https://doi.org/10.1038/s41467-018-03749-0>.
62. Libedinsky, C., and Livingstone, M. (2011). Role of prefrontal cortex in conscious visual perception. *J. Neurosci.* 31, 64–69. <https://doi.org/10.1523/JNEUROSCI.3620-10.2011>.
63. Boly, M., Massimini, M., Tsuchiya, N., Postle, B.R., Koch, C., and Tononi, G. (2017). 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. <https://doi.org/10.1523/JNEUROSCI.3218-16.2017>.
64. Frässle, S., Sommer, J., Jansen, A., Naber, M., and Einhäuser, W. (2014). Binocular rivalry: frontal activity relates to introspection and action but not to perception. *J. Neurosci.* 34, 1738–1747. <https://doi.org/10.1523/JNEUROSCI.4403-13.2014>.
65. Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* 17, 307–321. <https://doi.org/10.1038/nrn.2016.22>.
66. Tsuchiya, N., Wilke, M., Frässle, S., and Lamme, V.A.F. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770. <https://doi.org/10.1016/j.tics.2015.10.002>.
67. Lamme, V.A.F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. <https://doi.org/10.1016/j.tics.2006.09.001>.
68. Sandberg, K., Frässle, S., and Pitts, M. (2016). Future directions for identifying the neural correlates of consciousness. *Nat. Rev. Neurosci.* 17, 666. <https://doi.org/10.1038/nrn.2016.104>.
69. Boly, M., Seth, A.K., Wilke, M., Ingmundson, P., Baars, B., Laureys, S., Edelman, D.B., and Tsuchiya, N. (2013). Consciousness in humans and non-human animals: recent advances and future directions. *Front. Psychol.* 4, 625. <https://doi.org/10.3389/fpsyg.2013.00625>.
70. Panagiotaropoulos, T.I., Dwarakanath, A., and Kapoor, V. (2020). Prefrontal cortex and consciousness: beware of the signals. *Trends Cogn. Sci.* 24, 343–344. <https://doi.org/10.1016/j.tics.2020.02.005>.
71. Liu, S., Yu, Q., Tse, P.U., and Cavanagh, P. (2019). Neural correlates of the conscious perception of visual location lie outside visual cortex. *Curr. Biol.* 29, 4036–4044.e4. <https://doi.org/10.1016/j.cub.2019.10.033>.
72. Hatamimajoumerd, E., Ratan Murty, N.A., Pitts, M., and Cohen, M.A. (2022). Decoding perceptual awareness across the brain with a no-report fMRI masking paradigm. *Curr. Biol.* 32, 4139–4149.e4. <https://doi.org/10.1016/j.cub.2022.07.068>.
73. Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37. [https://doi.org/10.1016/S0010-0277\(00\)00123-2](https://doi.org/10.1016/S0010-0277(00)00123-2).
74. Podvalny, E., Flounders, M.W., King, L.E., Holroyd, T., and He, B.J. (2019). A dual role of prestimulus spontaneous neural activity in visual object recognition. *Nat. Commun.* 10, 3910. <https://doi.org/10.1038/s41467-019-11877-4>.

75. Martinez-Trujillo, J. (2022). Visual attention in the prefrontal cortex. *Annu. Rev. Vis. Sci.* 8, 407–425. <https://doi.org/10.1146/annurev-vision-100720-031711>.
76. Mashour, G.A., Pal, D., and Brown, E.N. (2022). Prefrontal cortex as a key node in arousal circuitry. *Trends Neurosci.* 45, 722–732. <https://doi.org/10.1016/j.tins.2022.07.002>.
77. Pal, D., and Mashour, G.A. (2022). General anesthesia and the cortical stranglehold on consciousness. *Neuron* 110, 1891–1893. <https://doi.org/10.1016/j.neuron.2022.05.014>.
78. Davis, Z.W., Muller, L., Martinez-Trujillo, J., Sejnowski, T., and Reynolds, J.H. (2020). Spontaneous travelling cortical waves gate perception in behaving primates. *Nature* 587, 432–436. <https://doi.org/10.1038/s41586-020-2802-y>.
79. He, B.J. (2013). Spontaneous and task-evoked brain activity negatively interact. *J. Neurosci.* 33, 4672–4682. <https://doi.org/10.1523/JNEUROSCI.2922-12.2013>.
80. Hesselmann, G., Kell, C.A., and Kleinschmidt, A. (2008). Ongoing activity fluctuations in hMT+ bias the perception of coherent visual motion. *J. Neurosci.* 28, 14481–14485. <https://doi.org/10.1523/JNEUROSCI.4398-08.2008>.
81. Hesselmann, G., Kell, C.A., Eger, E., and Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proc.Natl.Acad.Sci. USA* 105, 10984–10989. <https://doi.org/10.1073/pnas.0712043105>.
82. van Dijk, H., Schoffelen, J.M., Oostenveld, R., and Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *J. Neurosci.* 28, 1816–1823. <https://doi.org/10.1523/JNEUROSCI.1853-07.2008>.
83. Wyart, V., and Tallon-Baudry, C. (2009). How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. *J. Neurosci.* 29, 8715–8725. <https://doi.org/10.1523/JNEUROSCI.0962-09.2009>.
84. Klatzmann, U., Froudast-Walsh, S., Bliss, D.P., Theodoni, P., Mejias, J., Niu, M., Rapan, L., Palomero-Gallagher, N., Sergent, C., Dehaene, S., et al. (2022). A connectome-based model of conscious access in monkey cortex. *bioRxiv*. <https://doi.org/10.1101/2022.02.20.481230>.
85. van Kempen, J., Gieselmann, M.A., Boyd, M., Steinmetz, N.A., Moore, T., Engel, T.A., and Thiele, A. (2021). Top-down coordination of local cortical state during selective attention. *Neuron* 109, 894–904.e8. <https://doi.org/10.1016/j.neuron.2020.12.013>.
86. Engel, T.A., Steinmetz, N.A., Gieselmann, M.A., Thiele, A., Moore, T., and Boahen, K. (2016). Selective modulation of cortical state during spatial attention. *Science* 354, 1140–1144. <https://doi.org/10.1126/science.aag1420>.
87. Ray, W.J., and Cole, H.W. (1985). EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science* 228, 750–752. <https://doi.org/10.1126/science.3992243>.
88. Alayrangues, J., Torrecillas, F., Jahani, A., and Malfait, N. (2019). 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. <https://doi.org/10.1016/j.neuroimage.2018.09.013>.
89. Engel, A.K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. <https://doi.org/10.1016/j.conb.2010.02.015>.
90. David, F., Courtiol, E., Buonviso, N., and Fourcaud-Trocmé, N. (2015). Competing mechanisms of gamma and beta oscillations in the olfactory bulb based on multimodal inhibition of mitral cells over a respiratory cycle. *eNeuro* 2. ENEURO.0018-15.2015. <https://doi.org/10.1523/ENEURO.0018-15.2015>.
91. Spitzer, B., and Haegens, S. (2017). Beyond the status quo: a role for beta oscillations in endogenous content (re)activation. *eNeuro* 4. ENEURO.0170-17.2017. <https://doi.org/10.1523/ENEURO.0170-17.2017>.
92. Miller, E.K., Lundqvist, M., and Bastos, A.M. (2018). Working Memory 2.0. *Neuron* 100, 463–475. <https://doi.org/10.1016/j.neuron.2018.09.023>.
93. Panagiotaropoulos, T.I., Kapoor, V., and Logothetis, N.K. (2013). Desynchronization and rebound of beta oscillations during conscious and unconscious local neuronal processing in the macaque lateral prefrontal cortex. *Front. Psychol.* 4, 603. <https://doi.org/10.3389/fpsyg.2013.00603>.
94. Hardstone, R., Flounders, M.W., Zhu, M., and He, B.J. (2022). Frequency-specific neural signatures of perceptual content and perceptual stability. *eLife* 11, e78108. <https://doi.org/10.7554/eLife.78108>.
95. Zhu, M., Hardstone, R., and He, B.J. (2022). Neural oscillations promoting perceptual stability and perceptual memory during bistable perception. *Sci. Rep.* 12, 2760. <https://doi.org/10.1038/s41598-022-06570-4>.
96. Jensen, O., and Bonnefond, M. (2013). Prefrontal α - and β -band oscillations are involved in rule selection. *Trends Cogn.Sci.* 17, 10–12. <https://doi.org/10.1016/j.tics.2012.11.002>.
97. Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., and Miller, E.K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* 76, 838–846. <https://doi.org/10.1016/j.neuron.2012.09.029>.
98. Hazy, T.E., Frank, M.J., and O'Reilly, R.C. (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1601–1613. <https://doi.org/10.1098/rstb.2007.2055>.
99. Rougier, N.P., and O'Reilly, R.C. (2002). Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cogn. Sci.* 26, 503–520. https://doi.org/10.1207/s15516709cog2604_4.
100. Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends Cogn.Sci.* 15, 365–373. <https://doi.org/10.1016/j.tics.2011.05.009>.
101. Baars, B.J. (1997). In the Theater of Consciousness: the Workspace of the Mind (Oxford University Press). <https://doi.org/10.1093/acprof:oso/9780195102659.001.1>.
102. Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., and Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132, 2531–2540. <https://doi.org/10.1093/brain/awp111>.
103. Qian, C., Liu, C., Zou, J., Zhuo, Y., He, S., and Zhang, P. (2018). BOLD signal modulated with perception in the superficial layer of human V1 during binocular rivalry. *J. Vis.* 18, 955. <https://doi.org/10.1167/18.10.955>.
104. Huang, Z., Tarnal, V., Vlislides, P.E., Janke, E.L., McKinney, A.M., Picton, P., Mashour, G.A., and Hudetz, A.G. (2021). Anterior insula regulates brain network transitions that gate conscious access. *Cell Rep.* 35, 109081. <https://doi.org/10.1016/j.celrep.2021.109081>.
105. Wilson, H.R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proc.Natl.Acad.Sci. USA* 100, 14499–14503. <https://doi.org/10.1073/pnas.2333622100>.
106. Brascamp, J.W., van Ee, R., Noest, A.J., Jacobs, R.H.A.H., and van den Berg, A.V. (2006). The time course of binocular rivalry reveals a fundamental role of noise. *J. Vis.* 6, 1244–1256. <https://doi.org/10.1167/6.11.8>.
107. Huguet, G., Rinzel, J., and Hupé, J.M. (2014). Noise and adaptation in multistable perception: noise drives when to switch, adaptation determines percept choice. *J. Vis.* 14, 19. <https://doi.org/10.1167/14.3.19>.
108. Panagiotaropoulos, T.I., Kapoor, V., Logothetis, N.K., and Deco, G. (2013). A common neurodynamical mechanism could mediate externally induced and intrinsically generated transitions in visual awareness. *PLoS One* 8, e53833. <https://doi.org/10.1371/journal.pone.0053833>.
109. Northoff, G., and Zilio, F. (2022). Temporo-spatial Theory of Consciousness (TTC) - bridging the gap of neuronal activity and phenomenal states. *Behav. Brain Res.* 424, 113788. <https://doi.org/10.1016/j.bbr.2022.113788>.

110. Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Posterior and anterior cortex - where is the difference that makes the difference? *Nat. Rev. Neurosci.* 17, 666. <https://doi.org/10.1038/nrn.2016.105>.
111. Mazzucato, L., Fontanini, A., and La Camera, G. (2015). Dynamics of multistable states during ongoing and evoked cortical activity. *J. Neurosci.* 35, 8214–8231. <https://doi.org/10.1523/JNEUROSCI.4819-14.2015>.
112. Tsodyks, M., Kenet, T., Grinvald, A., and Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286, 1943–1946. <https://doi.org/10.1126/science.286.5446.1943>.
113. Maynard, E.M., Nordhausen, C.T., and Normann, R.A. (1997). The Utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* 102, 228–239. [https://doi.org/10.1016/S0013-4694\(96\)95176-0](https://doi.org/10.1016/S0013-4694(96)95176-0).
114. Hussar, C.R., and Pasternak, T. (2009). Flexibility of sensory representations in prefrontal cortex depends on cell type. *Neuron* 64, 730–743. <https://doi.org/10.1016/j.neuron.2009.11.018>.
115. 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. *Proc.Natl.Acad.Sci. USA* 115, E3539–E3548. <https://doi.org/10.1073/pnas.1802356115>.
116. Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* 16, 1661–1687. <https://doi.org/10.1162/089976604774201631>.
117. Kadir, S.N., Goodman, D.F.M., and Harris, K.D. (2014). High-dimensional cluster analysis with the masked EM algorithm. *Neural Comput.* 26, 2379–2394. https://doi.org/10.1162/NECO_a_00661.
118. Hazan, L., Zugaro, M., and Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. *J. Neurosci. Methods* 155, 207–216. <https://doi.org/10.1016/j.jneumeth.2006.01.017>.
119. 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, 547–553. <https://doi.org/10.1038/nature11618>.
120. Bokil, H., Andrews, P., Kulkarni, J.E., Mehta, S., and Mitra, P.P. (2010). Chronux: a platform for analyzing neural signals. *J. Neurosci. Methods* 192, 146–151. <https://doi.org/10.1016/j.jneumeth.2010.06.020>.
121. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869. <https://doi.org/10.1155/2011/156869>.
122. Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* 6, 461–464. <https://doi.org/10.1214/aos/1176344136>.
123. Hodges, J.L., and Lehmann, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* 34, 598–611. <https://doi.org/10.1214/aoms/1177704172>.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|----------------------------------|---|---|
| Software and algorithms | | |
| FieldTrip | https://www.fieldtriptoolbox.org/ | https://doi.org/10.1155/2011/156869 |
| Chronux Toolbox | https://www.mathworks.com/matlabcentral/fileexchange/68537-chronux-analysis-software | https://doi.org/10.1016/j.jneumeth.2010.06.020 |
| MATLAB 2016b and MATLAB 2022a | Custom scripts for analysis written in MATLAB 2016b and 2022a have been deposited into a repository with a public license. | https://doi.org/10.5281/zenodo.7642910 |

RESOURCE AVAILABILITY

Lead contact

Information and requests for data and code availability should be directed to and will be fulfilled by the lead contact, Theofanis I. Panagiotaropoulos (theofanis.panagiotaropoulos@tuebingen.mpg.de).

Materials availability

Materials availability does not apply here as no wet-lab work was carried out.

Data and code availability

Electrophysiological recordings used in this paper will be used in the future for other studies. Therefore, the data is available upon reasonable request; please contact the [lead contact](#). Custom analysis scripts written in MATLAB are available upon request from the co-first authors.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Two adult, male rhesus macaques (*Macaca mulatta*), weighing between 9.4–9.6kg and 12.2–12.8kg respectively, were employed in the study. The subjects, denoted as H07 and A11, were pair-housed in an enriched animal facility. They were habituated to a primate task training chair of custom design, and the experimental booth.

METHOD DETAILS

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 arrays¹¹³ (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.^{114,115} The arrays were 4×4mm 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 (0.1–30 kHz) were recorded using Neural Signal Processors (NSPs) from Blackrock Microsystems Inc. Signals from the Utah array were digitised, amplified, and then routed to the NSPs for acquisition via CerePlex E headstages and the corresponding pre-amplifiers. 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).¹¹⁶ Spikes were rejected if they occurred within 0.5 ms of each other, (multi-unit refractory period temporal threshold) or if they were larger than 50 times the MAD (to exclude large motion artifacts). All collected spikes were aligned to their minimum. Automatic clustering to detect putative single neurons was performed by a Split and Merge Expectation-Maximisation (SME) algorithm that fits a mixture of Gaussians to the spike feature data which consisted of the first three principal components¹¹⁷ (Klustakwik). The clusters were finalised manually using a cut-and-merge software¹¹⁸ (Klusters). For the analysis of LFP activity, the broadband signal was decimated to 500 Hz sampling rate using a Type I Chebyshev Filter, reliably 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. These 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 60Hz 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 sampling at 1kHz, 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 reliably capture 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 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, in the first five experimental sessions. In the sixth session, 200 random dots at 100% coherence with a limited lifetime of 150ms were presented. This marked the first monocular stimulus epoch in both conditions, viz. Binocular Rivalry (BR) and Physical Alternation (PA). At the end of 1-2s, 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 against each other for a period of 6-10s. 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 2s. Free viewing within the +/-8° window, which included the stimulus, 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 fixation and eye-movement range successfully remained within the specified viewing window during the entire duration of the trial. Every successful trial was followed by a 2-2.5s 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 20Hz 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 at 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 raw decimated LFP signal (0.1-500Hz) was collected around each clean transition. A threshold 3.5 times the SD of the Gaussian noise was set and both positive and negative deviations before and after the switch were detected. Next, the LFP signal 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 in each condition (BR, PA, PM and SD) was first z-scored in the frequency domain to visualise the relative changes in power and then pooled across all channels and averaged. The power spectrum to assess aperiodic and periodic components was computed by collapsing the CWT in time. 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 and the power-spectrum, i.e., the LF (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 obtain the instantaneous amplitude in time, we transformed the signal into the Hilbert space and then computed the absolute value. 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.¹¹⁹ The event-rate in time was computed as a quasi-PSTH by turning the detected events into a binary spike-train and smoothed with a Gaussian kernel of width 25ms, and then averaged across all channels (events/s/transition). The event-rate was computed as the sum of LF events normalised by the number of transitions and channels (events/transition/channel). To compute the build-up in the LF 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 1s post a perceptual switch 2. A preceding perceptual dominance for the competing stimulus must be maintained for 1s, 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 1s 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 1s) or pre-switch (-1 to 0s), 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.

For the decoding analysis, LFP events in both the LF and beta range were rasterised and sorted by switch type, i.e. preferred to non-preferred and non-preferred to preferred, depending on the ensemble they were detected at. These data, along with the binned spiking data were fed into a Support Vector Machine based binary classifier (linear regression), and the decoding accuracy (area under the curve of the receiver operator characteristic – AUCROC) was computed in each 50ms bin, with an overlap of 100ms. Significance testing was achieved by shuffling the labels ($n=100$), with chance decoding at 0.5 (50% accuracy). Finally, to estimate the start of the change in encoding from the spiking activity, a piecewise linear function with 2 degrees of freedom was fit to each transition profile. The earliest detected changepoint was collected as the start time of change in encoding.

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 0.9s window around the time of the marked smooth pursuit OKN change [-0.9s to 0.9s] 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. Finally, all cross-over times either less than -0.75s or more than 0.75s were discarded as these were considered to come from noisy trials.

Spike-field Coherence

The spike-field coherence (SFC) was computed between the spiking activity of selective ensembles for each transition, and the average LFP recorded at these selective electrodes. A rate adjustment and a finite-size correction was applied before computing the SFC via a multi-taper method¹²⁰ (Chronux Toolbox). The mean angle of spike-LFP locking was computed using the pairwise phase consistency method implemented by FieldTrip Toolbox.¹²¹

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 and their respective AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) measures were computed to evaluate the goodness-of-fit.¹²²

Statistical analysis

All statistical comparisons were performed using a Wilcoxon ranksum test¹²³ due to the non-gaussian nature of the underlying distribution from which the data originated (unless specified otherwise). 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. For computing significant phase angles of spike-LFP coupling, the Rayleigh Test for circular data was used.

COLOPHON

This document was typeset using the typographical look-and-feel **classicthesis** developed by André Miede, and modified by Shervin Safavi for the purpose of this thesis. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". **classicthesis** is available for both L^AT_EX and L_YX:

<http://code.google.com/p/classicthesis/>

Temporary page!

`LATEX` was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because `LATEX` now knows how many pages to expect for this document.