Fundamental Activity Constraints Lead to Specific Interpretations of the Connectome

Jannis Schuecker1*, Maximilian Schmidt1*, Sacha J. van Albada1, Markus Diesmann1,2,3, Moritz Helias1,3

1 Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA BRAIN Institute I, Jülich Research Centre, Jülich, Germany, 2 Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen University, Aachen, Germany, 3 Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany

☯ These authors contributed equally to this work.
* j.schuecker@fz-juelich.de

Abstract

The continuous integration of experimental data into coherent models of the brain is an increasing challenge of modern neuroscience. Such models provide a bridge between structure and activity, and identify the mechanisms giving rise to experimental observations. Nevertheless, structurally realistic network models of spiking neurons are necessarily underconstrained even if experimental data on brain connectivity are incorporated to the best of our knowledge. Guided by physiological observations, any model must therefore explore the parameter ranges within the uncertainty of the data. Based on simulation results alone, however, the mechanisms underlying stable and physiologically realistic activity often remain obscure. We here employ a mean-field reduction of the dynamics, which allows us to include activity constraints into the process of model construction. We shape the phase space of a multi-scale network model of the vision-related areas of macaque cortex by systematically refining its connectivity. Fundamental constraints on the activity, i.e., prohibiting quiescence and requiring global stability, prove sufficient to obtain realistic layer- and area-specific activity. Only small adaptations of the structure are required, showing that the network operates close to an instability. The procedure identifies components of the network critical to its collective dynamics and creates hypotheses for structural data and future experiments. The method can be applied to networks involving any neuron model with a known gain function.

Author Summary

The connectome describes the wiring patterns of the neurons in the brain, which form the substrate guiding activity through the network. The influence of its constituents on the dynamics is a central topic in systems neuroscience. We here investigate the critical role of specific structural links between neuronal populations for the global stability of cortex and elucidate the relation between anatomical structure and experimentally observed activity. Our novel framework enables the evaluation of the rapidly growing body of
connectivity data on the basis of fundamental constraints on brain activity and the combination of anatomical and physiological data to form a consistent picture of cortical networks.

Introduction

The neural wiring diagram, the connectome, is gradually being filled through classical techniques combined with innovative quantitative analyses [1, 2] and new technologies [3, 4]. The connectivity between neurons is considered to shape resting-state and task-related collective activity [5, 6]. For simple networks, clear relationships with activity are known analytically, e.g., a dynamic balance between excitatory and inhibitory inputs in inhibition-dominated random networks leads to an asynchronous and irregular state [7–9]. Structures and mechanisms underlying large-scale interactions have been identified by means of dynamical models [10, 11] and graph theory [12, 13]. Furthermore, the impact of local network structure on spike-time correlations is known in some detail [14–16]. Under certain conditions, there is a one-to-one mapping between correlations in neuronal network activity and effective connectivity, a measure that depends on the network structure and on its activity [17]. In conclusion, anatomy does not uniquely determine dynamics, but dynamical observations help constrain the underlying anatomy. Despite advances in understanding simple networks, a complete picture of the relationship between structure and dynamics remains elusive.

Fig 1 visualizes the integrative loop between experiment, model, and theory that guides the investigation of structure and dynamics. In the traditional forward modeling approach, structural data from experimental studies determine the connectivity between model neurons. Combined with the specification of the single-neuron dynamics and synaptic interactions, simulations yield a certain network dynamics. There is a fundamental problem with this approach.

Fig 1. The integrative loop between experiment, model, and theory. Black arrows represent the classical forward modeling approach: Experimental anatomical data are integrated into a model structure, which gives rise to the activity via simulation. The model activity is compared with experimental functional data. The usual case of disagreement leads to the need to change the model definition. By experience and intuition the researcher identifies the parameters to be changed, proceeding in an iterative fashion. Once the model agrees well with a number of experimental findings, it may also generate predictions on the outcome of new experiments. Red arrows symbolize our new approach: informed by functional data, an analytical theory of the model identifies critical connections to be modified, refining the integration of data into a consistent model structure and generating predictions for anatomical studies.

doi:10.1371/journal.pcbi.1005179.g001
Despite the impressive experimental progress in determining network parameters, any neural network model is underdetermined, because of the sheer complexity of brain tissue and inevitable uncertainties in the data. For instance, counting synapses on a large scale presently takes a prohibitive amount of time, and no available technique allows determining precise synaptic weights for entire neural populations. Although it may be possible to quantify the full connectome of an individual in future, inter-individual variability will require modelers to express connectivity as connection probabilities or rules of self-organization if they want to learn about general principles of the brain. In modeling studies, parameters are usually tuned manually to achieve a satisfactory state of activity, which becomes unfeasible for high-dimensional models due to the size of the parameter space. In particular, it is a priori unclear how parameters of the model influence its activity. In consequence, modifications cannot be performed in a targeted fashion, and it is difficult to find a minimal set of modifications necessary for aligning a model with experimental activity data.

Overcoming this problem requires a shift of perspective. Instead of regarding the model exclusively in a forward manner, generating predicted activity from structure, we in addition consider the system in a reverse manner, predicting the structure necessary to explain the observed activity. Our theory, starting from a mean-field description, provides a direct link between structure and activity. In contrast to simulations, the theory is invertible, which we exploit to identify connections critical for the dynamics and to find a minimal set of modifications to the structure yielding a realistic set point of activity. The predicted alterations generate hypotheses on brain structure, thus feeding back to anatomical experiments.

Applying the method to a multi-scale network model of the vision-related areas of macaque cortex, we derive targeted modifications of a set of critical connections, bringing the model closer to experimental observations of cortical activity. Based on the global properties of the bistable phase space of the model, the method refines the model’s construction principles within experimental uncertainties and identifies the connections that decisively shape the dynamics. Preliminary results have been presented in abstract form [18].

Results

Global stability in a simple network

We consider networks with neurons structured into $N$ interconnected populations. A neuron in population $i$ receives $K_{ji}$ incoming connections from neurons in population $j$, each with synaptic efficacy $J_{ij}$. Additionally, each neuron is driven by $K_{ext}$ Poisson sources with rate $\nu_{ext}$ and synaptic efficacy $J_{ext}$. All neurons in one population have identical parameters, so that we can describe the network activity in terms of population-averaged firing rates $\nu_i$.

Our method is applicable if the employed neuron model has a known gain function, either analytically or as a function obtained by interpolating numerical results from simulations. In this study, we model single cells as leaky integrate-and-fire model (LIF) neurons (see “Methods”). The possible stationary states of these networks are characterized by the firing rates that are equilibria of

$$\dot{\nu} = \frac{d\nu}{ds} = \Phi(\nu, A) - \nu,$$

where $s$ is a pseudo-time. The gain function $\Phi$ is known analytically and $A$ indicates its dependence on the model parameters $\{K, J, \nu_{ext}, \ldots\}$ (see “Methods”).

The input-output relationship $\Phi$ typically features a non-linearity which, in combination with feedback connections, can cause multi-stability in the network. In particular excitatory-excitatory loops cause the system defined by Eq (1) to exhibit multi-stable behavior in the
stationary firing rates. A necessary condition for the bistability is that the transfer function has an inflection point. The LIF neuron model can have such an inflection point, originating from the interplay of its leak term and the threshold. Less realistic neuron models, such as the perfect integrate-and-fire model, do not have such an inflection point. To illustrate its origin, we first consider the noiseless case [19] without absolute refractoriness ($\tau_r = 0$). The transfer function initially grows from zero with infinite slope due to the threshold and crosses the identity line (Fig 2C). For larger input the leak term can be neglected and the transfer function approaches a linear function with finite slope $1/\tau_m$ $V_y$ (see, e.g., [20], eq 11), equivalent to a perfect integrator. This is only possible with a negative curvature at intermediate rates, i.e., a reduction in the slope, which makes the transfer function cross the identity line another time, causing the bistability. Network-generated noise only affects the low-rate regime where it smears out the kink causing the transfer function to grow from zero with positive curvature (see, e.g. [21], eq. 22). Importantly, the qualitative picture, i.e., the bistable behavior, is not altered. A finite refractory period only has an effect for very high input values where the transfer function saturates at $1/\tau_r = 500$ spikes/s for the given parameters.

The basic problem is that there is a trade-off between excitation at the fixed points and their stability. In particular, exciting the model to bring a fixed point closer to experimental observations requires a method to preserve its stability. We achieve this by controlling the influence of the excitatory-excitatory loops on the phase space of the network.

As an illustration we first study the mechanism using the simple network architecture depicted in Fig 2A: a population of excitatory neurons is coupled to itself with indegree $K$ and is driven by external Poisson sources with the same indegree $K_{ext} = K$ and rate $\nu_{ext}$. All connections have identical synaptic weights $J$. An increase in the external drive shifts the input-output relationship $\Phi(\nu, \nu_{ext})$ of this one-dimensional system (Fig 2C) to the left. The bifurcation diagram is shown in Fig 2E: for low $\nu_{ext}$ there is only one fixed point with low activity (LA). When increasing $\nu_{ext}$, an additional pair of fixed points of which one is stable and the other is unstable emerges via a saddle-node bifurcation, leading to a bistable system. The second stable fixed point exhibits high firing rates, denoted as the high-activity (HA) state. For even higher values of $\nu_{ext}$, the LA state loses stability.

The equilibria, given by the zeros of the velocity $\dot{\nu}$ in the bistable case, are shown in Fig 2F. An increase of the drive on the one hand increases the firing rate of the LA fixed point (inset) but on the other hand reduces its basin of attraction, indicated by the colored bars in the top left corner. For illustrative purposes, we extend the problem to two dimensions by splitting the excitatory population into two subpopulations of equal size (Fig 2B), mimicking a loop between excitatory populations in the model of the vision-related areas of macaque cortex. The corresponding (symmetric) two-dimensional phase space is shown in Fig 2D. The basin of attraction for the LA fixed point, limited by the separatrix [22], is reduced with increasing external drive.

Since we have a bistable system, there must be at least one unstable fixed point on the separatrix at the intersection of the nullclines, i.e., the subspace for which the velocity $\dot{\nu}_i$ in direction $i$ vanishes. We use the unstable fixed point to preserve the basin of attraction when the external drive $\nu_{ext}$ is increased. For this purpose, we modify the connectivity $K \rightarrow K'$ to reverse the shift of the unstable fixed point due to the parameter change $\nu_{ext} \rightarrow \nu'_{ext}$ (see “Methods” for a detailed derivation). Since the separatrix follows the unstable fixed point, this approximately restores the original basin of attraction.

The resulting velocity of the system $\Phi(\nu'_{ext}, K')$ ($\Phi$ defines the system Eq (1)) is shown in Fig 2F. The firing rate in the LA fixed point is increased as desired (inset), and the unstable fixed point coincides with that obtained in the original system $\Phi(\nu_{ext}, K)$. This pattern of fixed points
is also indicated by the zero vectors of the velocity field $\dot{\nu}$ (Fig 2D). The separatrix follows the unstable fixed point, and the basin of attraction in the system $\Phi(\nu_{ext}, K')$ is restored to that in the original system $\Phi(\nu_{ext}, K)$. Fig 2G shows the behavior of the LA fixed point in more detail. The modification of $K$ does not noticeably change the location of the LA fixed point. In
conclusion, the method allows us to increase the firing rates in the LA fixed point without modifying its basin of attraction.

The purely excitatory network is the simplest model to explain a phase space configuration with a LA and a HA fixed point. Inhibitory feedback is not necessary for this bistability, but it would certainly alter the input-output relationship. For example the classical excitatory-inhibitory network [21] in the balanced regime has an input-output relationship with a negative slope and thus only one fixed point exists. However, if a pair of such balanced E-I networks is coupled by sufficiently strong mutually excitatory connections, these connections cause a bistability in a similar manner. Thus the mechanism shown in the purely excitatory network can also lead to the emergence of a HA attractor in more complex networks with inhibition.

**Bistability in the multi-scale network model**

We investigate a multi-scale network model of cortical areas to understand the structural features essential for a realistic state of baseline activity. The model extends and adapts the microcircuit model presented in [23], which covers $1\,\text{mm}^2$ surface area of early sensory cortex (Fig 3A), to all vision-related areas of macaque cortex (Fig 3B). Based on the microcircuit model, an area is composed of 4 layers (2/3, 4, 5, and 6) each having an excitatory (E) and an inhibitory (I) neural population, except parahippocampal area TH, which consists of only 3 layers (2/3, 5, and 6). A detailed description of the data integration is given in [24].

Simulations of the model (Fig 3C) reveal that, though realistic levels of activity can be achieved for populations in layers 2/3 and 4, populations 5E and 6E of the majority of areas show vanishingly low or zero activity in contrast to empirical data [25, 26]. Inputs from subcortical and non-visual cortical areas are modeled as Poissonian spike trains, whose rate $\nu_{\text{ext}}$ is a free, global parameter. To elevate the firing rates in the excitatory populations in layers 5 and 6, we enhance the external Poisson drive onto these populations parametrized by $\kappa$ (see "Methods"). However, already a perturbation of a few percent leads to a state with unrealistically high rates (Fig 3D), caused by the reduced basin of attraction of the low-activity state similar to Fig 2D. Our aim is to improve the working point of the model such that all populations exhibit spiking activity $\gtrsim 0.05$ spikes/s while preventing the model from entering a state with unrealistically high rates of $\gtrsim 30$ spikes/s (figure 13 of [25], [26]). The employed technique exposes the mechanism giving rise to the observed instability and identifies the circuitry responsible for this dynamical feature.

**Targeted modifications preserve global stability**

We apply the procedure derived in "Methods" and find targeted modifications to the connectivity $K$ that preserve the global stability of the low-activity fixed point for increased values of the external drive $\kappa$.

In the following we choose the inactive state $\mathbf{w}(0) = (0, \ldots, 0)^T$ as the initial condition. The exact choice is not essential since we are only interested in the fixed points of the system. Fig 4B shows the integration of Eq (1) over pseudo-time $s$ for different levels of the external drive to populations 5E and 6E parametrized by $\kappa$. For low values of $\kappa$, the integration converges to the LA fixed point shown in Fig 4D, and is in agreement with the activity emerging in the simulation (Fig 3C). For increased values of $\kappa$, the system settles in the HA fixed point (Fig 4E), again in agreement with the simulation. The population-specific firing rates in the HA state found in the mean-field predictions (Fig 4E) are also close to those in the simulation (Fig 3D), but minor deviations occur due to the violation of the assumptions made in the diffusion approximation. In particular, at these pathologically high rates, the neurons fire regularly and close to the reciprocal of their refractory period, while in the mean-field theory we assume...
Poisson spike statistics. Still, the mean-field theory predicts the bistability found in the simulation. Since the theory yields reliable predictions in both stable fixed points, we assume that also the location of the unstable fixed point in between these two extremes is accurately predicted by the theory.

To control the separatrix we need to find the unstable fixed point of the system. This is non-trivial since the numerical integration of Eq (1) for finding equilibria by construction only converges to stable fixed points. If the unstable fixed point has only one repelling direction (Fig 5A), it constitutes a stable attractor on the N−1 dimensional separatrix. The separatrix is a...
stable manifold [22], and therefore a trajectory originating in its vicinity but not near an unstable fixed point initially stays in the neighborhood. If an initial condition just outside the separatrix is close to the basin of attraction of a particular unstable fixed point, the trajectory initially approaches the latter. Close to the fixed point the velocity is small. Ultimately trajectories diverge from the separatrix in the fixed point’s unstable direction, as illustrated in Fig 4A.

In conclusion, we expect a local minimum in the velocity along the trajectories close to the unstable fixed point. To estimate the location of the unstable fixed point in this manner, we need to find initial conditions close to the separatrix. Naively, we would just fix the value of $\kappa$ and vary the initial condition. However, due to the high dimensionality of our system this is not feasible in practice. Instead, we vary $\kappa$ for a fixed initial condition. Fig 4B shows the firing rate averaged across populations for two trajectories starting close to the separatrix, where the first one converges to the LA fixed point and the second one to the HA state. The trajectories diverge near the unstable fixed point and thus we define the last local minimum of the Euclidean norm of the velocity vector as the critical time $s_c$ at which we assume the system to be close
to the unstable fixed point (Fig 4C). We find four relevant and distinct unstable fixed points, of which two are shown in Fig 6.

To counteract the shift of the separatrix caused by the increase in $\kappa$, we follow the procedure described in “Methods”. We subject the modifications of connectivity to the additional following constraints. In line with the anatomical literature, we do not allow for changes of the connectivity that would lead to cortico-cortical connections originating in the granular layer 4 [27], and we also disallow inhibitory cortico-cortical connections, as the vast majority of long-range connections are known to be excitatory [28, 29]. In addition, we naturally restrict indegrees to positive values. We find that four iterations (numbered by index $j$) corresponding to the four distinct unstable fixed points suffice to preserve the basin of attraction of the LA state with respect to an increase of the external drive up to $\kappa = 1.15$. In the following we concentrate on iterations 1 and 2, where the second one is also representative for iterations 3 and 4, which

![Fig 5. Eigenspectrum analysis of network stability.](image)

A) Eigenvalue spectrum of the effective connectivity matrix $\mathbf{M}$ for the first (blue squares) and second (red dots) iteration. The dashed vertical line marks the edge of stability at a real part of 1. B) Contribution $\eta_l$ (Eq (13)) of an individual eigenprojection $l$ to the shift of the unstable fixed point versus the relative change in indegrees associated with $l$ for the first iteration. The data point corresponding to $\lambda^{(1)}$ is marked in red. The inset shows the relative angles between $\delta \nu^e$ and the eigenvectors $u^e_l$. The red line corresponds to the critical eigendirection. C) Entries of the eigenvector $u^{(1)}$ associated with $\lambda^{(1)}$ in the populations of the model. The affected areas are 46 and FEF. D) Same as C for the second iteration.

doi:10.1371/journal.pcbi.1005179.g005

Fig 6. Unstable fixed points in subsequent iterations. Population firing rates at the unstable fixed point as predicted by the mean-field theory encoded in color for iterations 1 (A) and 2 (B). Same display as in Fig 3.

doi:10.1371/journal.pcbi.1005179.g006
are qualitatively alike. To derive the required modifications of the indegree matrix, we decom-
pose $K$ into its $N$ eigenmodes and quantify the contribution of each eigenmode to the shift of
the unstable fixed point (see “Methods”). This allows us to identify the most effective eigend-
direction: in each iteration $j$ there is exactly one unstable eigendirection with an eigenvalue
$\text{Re}(\lambda_j^{(j)}) > 1$ (Fig 5A). The associated critical eigenvector is approximately anti-parallel to the
shift of the fixed point, $\delta \nu$ (inset of Fig 5B), and of similar length. The critical eigendirection
(red dot in Fig 5B) constitutes the most effective modification, giving the largest contribution
to the desired shift while requiring only a small change of 2.3% in average total indegrees. In
the chosen space of eigenmodes, the modifications are minimal in the sense that only this
most effective eigenmode is changed.

The associated eigenvector $u_{\nu}^{(1)}$ predominately points into the direction of populations 4E
and 5E of areas FEF and 46 (Fig 5C), while $u_{\nu}^{(2)}$ has large entries in the 5E populations of two
areas (Fig 5D). The high rates of these populations at the unstable fixed points (cf. Figs 6A and
6B with 5C and 5D) reflect that the instability is caused by increased rates in excitatory popu-
lations, particularly in population 5E. Each iteration shifts the transition to the HA state (the
value of $\kappa$ for which the separatrix crosses the initial condition) to higher values of $\kappa$ and
increases the attainable rates of populations 5E and 6E in the LA state (Fig 7A). After all four
iterations, the average total indegrees (summed over source populations) of the system are
changed by 11.3%. The first iteration mainly affects connections within and between areas 46
and FEF (Fig 7B). In particular, the excitatory loops between the two areas are reduced in
strength, especially those involving layer 5 (Fig 7C). We thus identify two areas forming a criti-
cal loop. In the remaining iterations, the changes are spread across areas and especially con-
nections originating in layer 5 are weakened (Fig 7D). In conclusion, the method identifies
critical structures in the model both on the level of areas and on the level of layers and popu-
lations, and leads to a small but specific structural change of the model.

Fig 7. Altered phase space and modified connections. A Firing rates average $d$ across populations 5E and
6E and across areas for different stages from the original model (black) to iteration 4 (light gray) as a function
of $\kappa$, predicted by the mean-field theory. B Relative changes in the indegree $\delta K_{AB}/\sum K_{AB}$ between areas $A, B$ in the first iteration. C Layer-specific relative changes $\delta K_{ln}/\sum K_{ln}$ in the connections within and between areas FEF and 46, for the first iteration. Populations are ordered from 2/3E (left) to 6I (right) on the horizontal axis and from 6I (bottom) to 2/3E (top) on the horizontal axis as in panel D. D Relative changes in population-specific indegrees summed over target populations, $\sum \delta K_{ln}/\sum K_{ln}$, combined for iterations two, three and four.

doi:10.1371/journal.pcbi.1005179.g007
Analysis of the modifications

In the following we analyze the modifications of the connectivity with respect to the internal and inter-area connections in detail. The intrinsic circuits of the areas are modified in different directions, as shown for two exemplary areas V4 and CITv in Fig 8A. Despite this heterogeneity, significant changes affect mostly excitatory-excitatory connections (Fig 8A, bottom panel) with connections from population 5E experiencing the most significant changes (top panel of Fig 8A). In fact, the anatomical data [30] underlying the microcircuit model [23] contain only two reconstructed excitatory cells from layer 5, but considerably more for other cell types, indicating a higher uncertainty for layer 5 connections. Fig 8B shows the correlation between intrinsic connectivity changes for all pairs of areas, with areas ordered according to a hierarchical clustering using a farthest point algorithm [31] on the correlation matrix. We find four clusters each indicating a group of areas which undergo changes with similar patterns. The groups are displayed in different colors in the histogram in Fig 8B. The areas of the model are categorized into architectural types based on cell densities and laminar thicknesses (see “Methods”). Areas with architectural type 4, 5 and 6 are distributed over several clusters. We can

---

**Fig 8. Analysis of changes in connectivity.** A Top panel: relative changes in population-specific intrinsic indegrees summed over target populations and averaged across areas, \(\delta K_{ij}^{\text{target}} / \sum K_{ij}^{\text{target}}\). Bottom panel: changes in the indegrees within and between exemplary areas V4 and CITv relative to the total indegrees of the target populations, i.e., \(\delta K_{ij}^{\text{total}} / \sum K_{ij}^{\text{total}}\). Populations ordered as in Fig 7C. B Pearson correlation coefficient of the changes of the internal indegrees \(\delta K_{ij}^{\text{target}}\) between all pairs of the 32 areas. Areas ordered according to hierarchical clustering using a farthest point algorithm [31]. The heights of the bars on top of the matrix indicate the architectural types of the areas (types 1 and 3 do not appear in the model) with color representing the respective clusters. C FLN of the modified connectivity after 4 iterations versus the original FLN of the model. Only FLN > 10^{-3} are shown for a better overview. The overlapping red dots represent the connections between areas 46 and FEF. Unity line shown in gray. D Histogram of the cumulative changes in SLN over all four iterations (\(\delta SLN = SLN^{(4)} - SLN^{(0)}\)).

---

doi:10.1371/journal.pcbi.1005179.g008
interpret this as a differentiation of these types into further subtypes. The resulting changes of the intra-areal connectivity are small (Fig 8A), but still significant for network stability.

Connections between areas can be characterized by their FLN and SLN (see “Methods”). The FLN reflects the overall strength of an inter-areal connection and is only weakly affected across connections (Fig 8C), with a correlation between original and modified logarithms of FLN of $r_{\text{Pearson}} = 0.79$. Significant variations in the FLN occur mostly for very weak connections that are likely to have substantial relative uncertainties in the experimental data. The two overlapping red dots in Fig 8C represent the connections between areas 46 and FEF, which are modified in the first iteration (Fig 7C). The SLN determines the laminar pattern of the location of source neurons for cortico-cortical connections. Overall, data are available for 24% of the inter-areal connections in the parcellation of Felleman & van Essen [27], while the SLN for the rest are derived from the sigmoidal law. The majority of connections undergo small changes in their laminar source pattern (Fig 8D) and connections with large modifications ($|\delta_{\text{SLN}}| > 0.5$) are weak (average $\text{FLN} = 6 \cdot 10^{-4}$ compared to $\text{FLN} = 10^{-2}$ in the model as a whole). Because weak connections are represented by low counts of labeled neurons, they have a relatively large uncertainty in their laminar patterns, justifying larger adjustments. Spearman’s rank correlation between the SLN of the original model that were directly taken from experiments and the logarithmic ratios of cell densities is $\rho = -0.63$ ($p = 3 \cdot 10^{-11}$, p-value of a two-sided test for uncorrelated data). For the modified model, we take the SLN of all connections into account and obtain $\rho = -0.40$ ($p = 6 \cdot 10^{-20}$), indicating a reduced, but still significant, monotonic dependence between SLN and the logarithmic ratios of cell densities.

To judge the size of the modification to the connectivity, we compare it to the variability of measured cortico-cortical connection densities [1]. We quantify the latter as the average inter-individual standard deviation of the logarithmic FLN, i.e., $\sigma = \sqrt{\text{var}(\text{logFLN})}$, where the overbar denotes the average over injections and $\langle \cdot \rangle$ the average over connections. This variability equals 2.17 while the average modification of the logarithmic FLN is 1.34. The main experimental connection probabilities used to construct the intra-areal connectivity of the model have an average relative standard deviation of 30% across electrophysiological experiments (cf. Table 1 of [23]) while the intra-areal connection probabilities of the model are modified by 9% on average. The authors of [32] report even greater variability in their review on cortico-cortical and thalamocortical connectivity. These considerations show that on average, the changes applied to the connectivity are well within the uncertainties of the data. Overall, 35 out of 603 connections were removed from the network. In the CoCoMac database, 83% of these are indicated by only a single tracer injection, while the overall proportion of connections measured by a single injection is 59%.

For the modified connectivity and $\kappa = 1.125$, which we choose to avoid being too close to the transition (Fig 7A), the theory predicts average rates in populations 5E and 6E of 1.3 and 0.18 spikes/s, which is closer to experimentally observed rates compared to the original model. Furthermore we find that the modified connectivity allows us to decrease the inhibition in the network to $g = -11$. Simulating the full spiking network model then results in reasonable rates across populations and areas (Fig 9B and 9D). The average rates in populations 5E and 6E are increased compared to a simulation of the original model from 0.09 and $2 \cdot 10^{-5}$ spikes/s to 3.0 and 0.4 spikes/s, respectively. All populations exhibit firing rates within a reasonable range of 0.05 to 30 spikes/s (Fig 9D), as opposed to the original state in which a considerable fraction of excitatory neurons is silent (Fig 3E). The theoretical prediction is in excellent agreement with the rates obtained in the simulation (Fig 9A and 9C). Small discrepancies are caused by violations of the employed assumptions, i.e., Poissonian spiking statistics [33]. Differences between theory and simulation are small, and negligible for the central
aim of the study: the integration of activity constraints into the data-driven construction of multi-scale neuronal networks.

Discussion

This study investigates the link between experimentally measured structural connectivity and neuronal activity in a multi-scale spiking network model of the vision-related areas of macaque cortex [24]. We devise a theoretical method that systematically combines anatomical connectivity with physiological activity constraints. Already weak constraints, demanding the activity to neither vanish nor be pathologically high, yield a set of specific but small structural modifications necessary to increase the model’s excitation to a realistic level. We do not fit the model parameters to experimental activity data in the sense of minimizing an error function, since the sparse relevant experimental data do not allow for defining such a function in a meaningful way. Nevertheless, we considerably improve the network activity with a change in only a small set of parameters. The procedure constrains the experimentally obtained connectivity maps to a realization that is compatible with physiological experiments. This establishes a path from experimentally observed activity to specific hypotheses about the anatomy and demands for further experiments.

Connections are modified both inside and between cortical areas, on average well within the known uncertainties of the underlying data. The model areas are based on the early sensory cortex model presented in [23]. This circuit is adapted to individual areas by taking into account neuronal densities and laminar thicknesses. The model definition renders areas with equal architectural type similar in their internal connectivity, a drastic but inevitable simplification due to the lack of more detailed experimental data. The proposed method softens this assumption: small adaptations of internal connectivity distinguish the architectural types into further subtypes. These modifications are significant for the global stability of the network. Thus, our approach enables purely anatomy-based area categorizations to be refined with dynamical information.
Connections between areas are changed in terms of total strength and laminar patterns. Overall, the changes are small, but significant for specific connections. The loop formed by areas 46 and FEF is critical to the global stability of the network. Both areas have been investigated in [1], albeit in a different parcellation scheme than the scheme used here [27]. Our method suggests a weaker coupling of these two areas than found in the anatomical data set. Uncertainties, partly due to the mapping between parcellations, leave room for this interpretation. Areas 46 and FEF belong to prefrontal cortex and are multimodal, indicating that the influence of other parts of cortex could stabilize them, a mechanism outside the scope of the present model of vision-related areas. Both explanations can be tested either in an experimental study or with an extended model.

A few weak connections undergo large changes in their laminar patterns. With the present activity constraints, the method hereby weakens hierarchical relations in the model structure, indicating that preserving these relations requires additional dynamical constraints such as layer-specific coherence between areas [34, 35]. Conversely, it is possible to achieve satisfactory population rates with a less pronounced hierarchical structure.

Our analysis reveals that layer 5 excitatory cells play a critical role in the model’s dynamics, in line with the observed ongoing activity in mammalian neocortex [36, 37]. This critical role is often attributed to single-neuron properties, with a subset of layer 5 neurons displaying pacemaker activity [38–40]. In addition, we here find that the network architecture itself already explains the strong impact of layer 5 on the phase space of the network, suggesting that single-neuron properties and network structure jointly enable layer 5 to exert its dominant influence.

The cortico-cortical connectivity of the model is compiled from the extensive dataset of [1] combined with the CoCoMac database [41, 42], which collects data from hundreds of tracing studies. One could consider alternative methods for combining this information into one connectivity graph, for instance taking into account how consistently a given connection is reported across studies [43], and compare different methods by analyzing the resulting network dynamics. The presented mean-field theory could then be used to estimate the firing rates of each network instance without performing time-consuming simulations. However, here we first choose the integrative approach that accumulates all available experimental evidence into a single model and afterwards modify the resulting connectivity with our analytical procedure, thereby effectively discarding uncertain connections.

We restrict this study to networks of leaky integrate-and-fire model neurons, consistent with the key concept of the models we consider: individual cells are modeled in a simple manner to expose the impact of structural connectivity on the network dynamics. Moreover, the current-based leaky integrate-and-fire neuron can reproduce in-vivo like activity [44, 45] and is analytically tractable, which enables the identification of mechanisms underlying specific network effects. More complex neuron models can be incorporated into the method by replacing the gain function of each neuronal population with an analytical expression or an interpolated function obtained from spiking single-neuron simulations. For example, one could use a conductance-based point-neuron model for which the network dynamics can be described by population rate models [46], featuring a non-monotonic gain function: the gain is reduced if excitatory and inhibitory inputs are increased in a balanced manner [47]. Generally, this renders a system more stable. However, the bistability considered in our work is caused by excitatory inputs. Since conductance-based models also have a monotonically increasing gain function in dependence of the excitatory conductance alone, we expect the bistability to occur for such models as well.

Importantly, our method only requires a description of the system’s fixed points and their dependence on model parameters. The employed theory uses the diffusion approximation to derive a self-consistency equation for the stationary population rates valid for high indegrees.
and small synaptic efficacies. These requirements are fulfilled in the multi-area model and therefore the theoretical prediction agrees with the stationary activity of the simulation. Moreover, the theory predicts the bistability of the model, which is non-trivial, as the mean-field assumption of Poisson statistics of the activity is generally violated in the high-activity state. Nevertheless, since the activity in this state is mostly driven by strong mean inputs and the theory converges to the noiseless solution in the mean-driven limit, its predictions still provide viable approximations. The firing rates in the unstable fixed points are predominantly low, while the exceptions with very high rates are again mean-dominated. Presumably this explains the accuracy by which the theory predicts the locations of these fixed points and the resulting global stability properties.

Since the high-activity attractor of the model under consideration is unrealistic, we aim to prevent a transition to this state. However, in the high-dimensional system it is not a priori clear in which direction the separatrix has to be shifted to ensure stable dynamics. We therefore choose a pragmatic approach and shift the separatrix back to its initial location, inverting the shift which reduced the global stability of the low-activity fixed point. We achieve this to a good approximation by preserving the location of unstable fixed points on the separatrix. To this end, we use a linearization around these locations and an eigenmode decomposition to identify the set of connections to adjust. In the multi-area model, this linearization is justified because the system operates close to an instability so that only minor modifications are required. The method can be generalized to larger modifications by changing parameters iteratively in small steps.

Biological networks have various stabilization mechanisms not considered here, which render them less critical. For instance, during growth, homeostatic mechanisms guide the system toward the right structure. Furthermore, short-term synaptic plasticity (reviewed in [48]), homeostatic synaptic scaling [49] and spike-frequency adaptation (e.g. summarized in [50]) may prevent the system from entering the high-activity state. However, introducing these self-organizing mechanisms increases model complexity, causing a more intricate relation between structure and activity. Therefore, we start from a mean-field description on the level of neuronal populations, ignoring details of synaptic dynamics. Mild constraints on the activity lead to a network structure within the anatomical range of parameters. This network yields globally stable activity, suggesting that additional stabilization mechanisms are not required to achieve this. Nonetheless, they can potentially render the network more robust against external stimulation.

The observed inter-individual variability may reflect that mechanisms of self-organization and homeostasis find structurally different implementations of the same function. Thus, in studies across individuals we cannot expect that progress in experimental technology narrows down the variability of parameters indefinitely. The combined ranges of parameters rather specify the solution space, and our method provides a way to find a particular solution.

In principle, the method applies to any model parameter. It would be possible to modify, for example, synaptic weights. Since experimental data on synaptic weights are sparse, this is another natural choice. Moreover, such an analysis may provide hints about suitable synaptic plasticity rules that dynamically stabilize the model. The method can be applied to networks with more complex sets of attractors compared to the bistable case considered here. Though in high-dimensional systems such as our multi-area model, a larger number of attractors would make it more challenging to find all relevant unstable fixed points, the underlying idea of preserving the location of a separatrix is general. In contrast to the model considered here, transitions between fixed points can have a functional meaning in certain neuronal networks with multiple attractors. The specific location of the separatrix is then functionally relevant. Our method exposes the sensitivity of the location of the separatrix to certain model parameters and allows controlling its location in a specific manner.
In this work, we analyzed the global stability properties of the neuronal network on the population level. In contrast, Ostojic [51] performs a local stability analysis on the level of single neurons of an initially stable fixed point in a system with only one attractor. The author investigates the point at which the real parts of the eigenvalues of the Jacobian matrix evaluated at this fixed point become positive, i.e., the fixed point turns spectrally unstable and the system undergoes a transition to a heterogeneous asynchronous state. Analyzing the spectral stability on the single-neuron level does not reveal the global stability properties required in the current work: While a local stability analysis only considers infinitesimal perturbations, studying the basin of attraction gives information about the size of fluctuations against which the fixed point is stable. We expect both attractors to be spectrally stable because they do not show strong rate fluctuations and the mean-field theory predicts the activity accurately in both cases. Nevertheless, a heterogeneous state could occur if the synaptic weights were increased or the external drive was stronger. However, [52] show that the transition in stochastic systems that quantitatively resemble the spiking network [53], does not coincide with the loss of spectral stability.

One striking feature of the heterogeneous state is bursty spiking behavior of individual cells. Bursty spiking is also observed in the multi-area model for increased synaptic weights of inter-area connections [24]. The fixed points cannot be accurately described in this case because the fluctuations need to be taken into account [54]. Simulations show (figures 4 and 5 of [24]) that the modifications obtained in this study are still able to prevent the system from a transition to a HA attractor also in the presence of bursting neurons. This indicates that the phase space does not change qualitatively and our results are robust against such bursting behavior.

Experimental data on stationary activity in cortex are sparse. We therefore restrict ourselves to fundamental constraints and increase the drive to the model in an area-unspecific way to fulfill them. The resulting heterogeneity of the firing rates across areas is thus not imposed by the method, but rather arises from the connectivity that remains strongly informed by anatomical data. Alternatively, one could predefine a desired state and investigate the parameter changes necessary to achieve it.

The presented analytical method that combines anatomy and activity data into a consistent model is restricted to stationary firing rates. In future studies, also higher-order statistical measures of activity can be used as constraints. Resting-state fMRI, for example, provides information on the functional connectivity between areas as a second-order measure. When combined with analytical predictions of functional connectivity, the method may shed light on the anatomical connection patterns underlying inter-area communication.

**Methods**

In this study, we model single cells as leaky integrate-and-fire model neurons with exponentially decaying postsynaptic currents. Table 1 specifies the model parameters.

In the diffusion approximation, the dynamics of the membrane potential \( V \) and synaptic current \( I_s \) is [33]

\[
\frac{\tau_m}{dt} V = -V + I_s(t) \\
\frac{\tau_s}{dt} I_s = -I_s + \mu + \sigma \sqrt{\frac{\tau_m}{\tau_s}} \xi(t),
\]

where \( \tau_m \) is the membrane time constant and \( \tau_s \) the synaptic time constant, respectively. The membrane resistance \( \tau_m/C_m \) has been absorbed into the definition of the current. The input spike trains are approximated by a white noise current with fluctuations \( \propto \sigma^2 \) and mean value \( \mu \). Here \( \xi \) is a centered Gaussian white process satisfying \( \langle \xi(t) \rangle = 0 \) and \( \langle \xi(t) \xi(t') \rangle = \delta(t - t') \).
Whenever the membrane potential \( V \) crosses the threshold \( \theta \), the neuron emits a spike and \( V \) is reset to the potential \( V_r \), where it is clamped during \( \tau_r \). All neurons in one population have identical parameters, so that we can describe the network activity in terms of population-averaged firing rates \( \nu_i \) that depend on population-dependent input \( \mu_i, \sigma_i \) determined by the connectivity. Using the Fokker-Planck formalism, the stationary firing rates for each population \( i \) are given by [33]

\[
\frac{1}{\nu_i} = \tau_r + \tau_m \sqrt{\pi \int_{-\infty}^{\nu_i} \frac{d\nu}{\sigma_i^2}} \frac{1}{\sqrt{\tau_m}} e^{x^2/(2\tau_m)} (1 + \text{erf}(x)) dx
\]

\[
= 1/\Phi_i(\nu, A)
\]

(3)

\[
\mu_i(A) = \tau_m \sum_j K_{ij} J_{ij} \nu_j + \tau_m K_{ext} J_{ext} \nu_{ext}
\]

(4)

\[
\sigma_i^2(A) = \tau_m \sum_j K_{ij} J_{ij}^2 \nu_j + \tau_m K_{ext} J_{ext}^2 \nu_{ext}
\]

(5)

which is correct up to linear order in \( \sqrt{\tau_r/\tau_m} \) and where \( \gamma = |\zeta(1/2)|/\sqrt{2} \), with \( \zeta \) denoting the Riemann zeta function [55]. Here, \( A \) is chosen from the set of model parameters \( \{K, J, \nu_{ext}, \ldots\} \). If \( A \) is a matrix, we vectorize it by concatenating its rows and indicate this by lower case, i.e., \( a = (a_{00}, a_{01}, \ldots, a_{0N}, a_{10}, \ldots, a_{1N}, a_{N0}, \ldots, a_{NN}) = \text{vec}(A^T)^T \) following [56]. If the chosen parameter is a scalar we denote it with \( a \).

We find the fixed points of Eq (3) by solving the first-order differential equation Eq (1) [57]

\[
\dot{\nu} = \frac{d\nu}{ds} = \Phi(\nu, A) - \nu,
\]

using the classical fourth-order Runge-Kutta method (RK4) with step size \( h = 0.01 \), where \( s \) denotes a dimensionless pseudo-time. The same approach can be used to solve the activity on

| Table 1. Specification of the neuron and synapse parameters. |
|-----------------|----------------|---------------------------------|
| **Synapse parameters** | **Name** | **Value** | **Description** |
| \( J \pm \delta J \) | 87.8 \( \pm \) 8.8 pA | excitatory synaptic strength |
| \( g \) | -16 (Fig 3) -11 (Fig 9) | relative inhibitory synaptic strength |
| \( d_x \pm \delta d_x \) | 1.5 \( \pm \) 0.75 ms | local excitatory transmission delay |
| \( d_y \pm \delta d_y \) | 0.75 \( \pm \) 0.375 ms | local inhibitory transmission delay |
| \( d \pm \delta d \) | \( d = s/\nu_t \pm \frac{1}{2} s/\nu_t \) | inter-areal transmission delay, with \( s \) the distance between areas |
| \( \nu_t \) | 3.5 m/s | transmission speed |

| **Neuron model** | **Name** | **Value** | **Description** |
|-----------------|----------------|---------------------------------|
| \( \tau_m \) | 10 ms | membrane time constant |
| \( \tau_r \) | 2 ms | absolute refractory period |
| \( \tau_s \) | 0.5 ms | postsynaptic current time constant |
| \( C_m \) | 250 pF | membrane capacity |
| \( V_r \) | -65 mV | reset potential |
| \( \theta \) | -50 mV | fixed firing threshold |
| \( E_L \) | -65 mV | leak potential |

doi:10.1371/journal.pcbi.1005179.t001

PLOS Computational Biology | DOI:10.1371/journal.pcbi.1005179 February 1, 2017 17 / 25
a single neuron level [58]. Note that Eq (1) does not reflect the real time evolution of the population rates, but rather is a mathematical method to obtain the system’s fixed points. In contrast to [57] we do not only search for stable fixed points, but also use Eq (1) to obtain unstable attractors (cf. “Results”), an idea originating from the study of simple attractor networks ([59] esp. their figure 2 and eq 12).

In a bistable situation, the initial condition of Eq (1) determines which fixed point the system settles in. However, studying the behavior for a particular initial condition is of minor interest, since the actual spiking network is a stochastic system which fluctuates around the fixed points of the deterministic system defined by Eq (1). Even if we knew that Eq (1) would relax to the LA fixed point for one particular initial condition, this would not necessarily imply that this state is indefinitely stable. Global stability is determined by the size of the basin of attraction of the LA fixed point.

In the following, we derive the equations leading us to targeted modifications of a parameter $b$ necessary to compensate for the changes in the global stability induced by the change of another parameter $a$. To this end, we study the behavior of the fixed points with respect to an infinitesimal change $\delta a = a' - a$ in the chosen model parameter. Let $v'(a)$ and $v'(a')$ be the corresponding locations of the fixed points and $\delta v = v'(a') - v'(a)$ their separation. We can then expand $v'(a')$ into a Taylor series up to first order in $\delta a$ and obtain

$$
\begin{align*}
v'(a') &= v'(a) + \Delta_a \delta a \\
\Leftrightarrow \delta v &= \Delta_a \delta a,
\end{align*}
$$

with

$$
\begin{align*}
\Delta_a &= \frac{dv(\mu, \sigma)}{da} \\
&= \frac{d\Phi(\mu, \sigma)}{da} \\
&= \frac{\partial \Phi}{\partial \mu} \frac{d\mu}{da} + \frac{\partial \Phi}{\partial \sigma} \frac{1}{\tau_i} \frac{d\sigma^2}{da},
\end{align*}
$$

where we notice that $S_i$ and $T_i$ only depend on the target population $i$. We accordingly define two diagonal matrices $S$ and $T$ with $S_a = S_i$ and $T_a = T_i$. We further define the connectivity matrix $W = K \odot J$, where $\odot$ denotes element-wise multiplication, also called the Hadamard product (see [60], for a consistent set of symbols for operations on matrices). The derivatives with respect to $a_j$ have the compact expressions

$$
\begin{align*}
\frac{d\mu_n}{da_j} &= \frac{\partial \mu_n}{\partial a_j} + \sum_n \frac{\partial \mu_n}{\partial v_n} \frac{dv_n}{da_j} \\
&= \left( D_a \mu \right)_j + \tau_m \sum_n \left( K_{in} J_{nm} \right) \Delta_{a,n},
\end{align*}
$$

with the Jacobian $\left( D_a f \right)_j = \frac{\partial f}{\partial a_j}$ of some vector-valued function $f$ and

$$
\frac{d\sigma^2}{da_j} = \left( D_a \sigma^2 \right)_j + \tau_m \sum_n \left( K_{in} J_{nm} \right) \Delta_{a,n},
$$

where we use the Hadamard product again to define the matrix $W_2 := K \odot J \odot J$. Inserting the
total derivatives into Eq (7), we derive the final expression for $\Delta_a$, reading

$$
\Delta_a = S[D_\mu + \tau_m W \Delta_a] + T[D_\sigma + \tau_m W_2 \Delta_a]
$$

$$
\Rightarrow \Delta_a = \left[1 - \tau_m(SW + TW_2)\right]^{-1} \left[S D_\mu + TD_\sigma\right].
$$

(8)

where we use $\mathbb{1}$ for the identity matrix and define the effective connectivity matrix $M$ and the matrix $\Delta_a$. The latter has dimensionality $N \times P$, where $P$ is the dimension of $a$ (for example, $P = N^2$ for $a = k$ the vector of indegrees). With the aid of Eq (8), evaluating Eq (6) at the unstable fixed point predicts the shift of the separatrix (Fig 2D) to linear order. We now consider an additional parameter $b$ which is modified to counteract the shift of the unstable fixed point caused by the change in parameter $a$, i.e.,

$$
\Delta_a \delta a = -\Delta_a \delta b
$$

(9)

where we used that the inverse of $\mathbb{1} - M$ appears on both sides of the equation and hence drops out. Note that the tuple $(a, b)$ may represent any combination of model parameters, for example external input, indegrees, synaptic weights, etc. For a particular choice of $(a, b)$ we solve Eq (9) for $\delta b$. For the illustrative example shown in Fig 2, where $a = v_{\text{ext}}$ and $b = K$, Eq (9) simplifies to

$$
\delta K = \frac{\tilde{\Delta}_a \delta v_{\text{ext}}}{\tilde{\Delta}_K} = \frac{K_{\text{ext}} \delta v_{\text{ext}}}{v},
$$

since $S$ and $T$ appearing in the respective $\tilde{\Delta}$'s cancel each other.

To determine critical connections in a more complex model, we choose $b = k$, i.e. the vector of indegrees, and solve Eq (9) with a decomposition into eigenmodes. We can write the right-hand side as

$$
-(\tilde{\Delta}_a \delta k)_i = -\sum_j \tau_m \left(SJ_{ij} + T(J^\top)^2\right)_{ij} v_{ij} \delta K_{ij}.
$$

(10)

The equation holds because $\mu_i, \sigma_i$ are only affected by connections to population $i$, and therefore their derivatives $\partial \mu_i/\partial k_l, \partial \sigma_i/\partial k_l$ and hence $\Delta_{k,l}$, vanish for $l \notin [(i-1)N+1, iN]$. We now make the ansatz

$$
\delta K_{ij} = \sum_l \epsilon_l \frac{(u^l v^l)^T}{\tau_m(SJ + T(J^\top)\delta k_l)}
$$

(11)

which decomposes the changes $\delta K$ into eigenmodes of the effective connectivity. The $u^l$ and $v^l$ are the $l$-th right and left eigenvectors of $M$ as defined in Eq (8), fulfilling the bi-orthogonality $v^\top u = \delta_{ln}$ and the completeness relation $\sum_l u^l v^l = \mathbb{1}$. Inserting Eq (10) with Eq (11) into Eq (9) yields

$$
\tilde{\Delta}_a \delta a = -\sum_l \epsilon_l u^l v^l v.
$$

(12)
Thus we can solve for \( \epsilon_l \) by multiplying from the left with \( v^T \)

\[
\begin{align*}
\nu^T & \Delta \delta a = - \sum_l \epsilon_l \nu^T u_l v^T v \\
\epsilon_n & = - \tilde{\alpha}^n \tilde{\gamma}^n.
\end{align*}
\]

Our goal is to find a set of connections which dominate the size of the basin of attraction of the LA fixed point. Inserting Eq (12) into Eq (6) leads to

\[
\delta \nu^* = \sum_l (1 - M)^{-1} \epsilon_l u_l v^T v \\
= \sum_l \frac{\epsilon_l}{1 - \lambda_l} u_l v^T v \\
= \sum_l \frac{-\tilde{\alpha}^l}{1 - \lambda_l} u_l,
\]

where \( \lambda_l \) are the eigenvalues of \( M \), which are either real or complex conjugate pairs since \( M \in \mathbb{R}^{N \times N} \). To determine the influence of each eigenmode on the shift of the fixed point, we project the eigenvectors \( u_l \) onto the fixed-point shift \( \delta \nu \) by multiplying each side with \( \delta \nu \delta \nu^T \) and solve again for \( \delta \nu^* \) to obtain

\[
\delta \nu^* = \sum_l \frac{-\tilde{\alpha}^l}{1 - \lambda_l} \delta \nu^T u_l \delta \nu^T \delta \nu^*, \tag{13}
\]

where we define the (possibly complex-valued) coefficients \( \tilde{\eta}_l \). We aim at a decomposition of \( \delta \nu^* \) into real components. If \( \text{Im}(\lambda_l) = 0 \), \( \tilde{\eta}_l \) is real, so we can work directly with \( \eta_l \). Complex eigenvalues \( \text{Im}(\lambda_l) \neq 0 \) and corresponding eigenvectors come in conjugate pairs, so in this case we combine the corresponding coefficients \( \eta_l = \tilde{\eta}_l + \tilde{\eta}_l^* \), to have all contributions \( \eta_l \in \mathbb{R} \) and \( \sum_l \eta_l = 1 \) by construction Eq (13). Each \( \eta_l \) quantifies how much of the total fixed-point shift can be attributed to the \( l \)-th eigenmode, which allows identification of the most effective eigendirection (see “Results”), where we apply the ansatz Eq (11) to a multi-area, multi-layer model of the vision-related areas of macaque cortex.

The spiking simulations of the network model were carried out on the JUQUEEN supercomputer [61]. All simulations were performed with NEST version 2.8.0 [62] with optimizations for the use on the supercomputer which will be included in a future release. The simulations use a time step of 0.1 ms and exact integration for the subthreshold dynamics of the leaky integrate-and-fire neuron model (reviewed in [63]).

**Multi-area model**

The multi-area model of the vision-related areas of macaque cortex uses the microcircuit model of [23] as a prototype for all 32 areas in the FV91 parcellation [27] and customizes it based on experimental findings on cortical structure. From anatomical studies, it is known that cortical areas in the macaque monkey are heterogeneous in their laminar structure and can be roughly categorized into 8 different architectural types based on cell densities and laminar thicknesses. This distinction was originally developed for prefrontal areas [64], and then extended to the entire cortex [65]. The visual cortex, and thus the model, comprises areas of categories 2, 4, 5, 6, 7 and 8. Precise layer-specific neuron densities are available for a number
of areas, while for other areas, the neuron density is estimated based on their architectural type (see [24] for details).

The inter-areal connectivity is based on binary data from the CoCoMac database [27, 41, 42, 66, 67] indicating the existence of connections, and quantitative data from [1]. The latter are retrograde tracing data where connection strengths are quantified by the fraction of labeled neurons (FLN) in each source area. The original analysis of the experimental data was performed in the M132 atlas [1]. Both the FV91 and the M132 parcellations have been registered to F99 space [68], a standard macaque cortical surface included with the software tool Caret [69]. This enables mapping between the two parcellations.

On the target side, we use the exact coordinates of the injections to identify the equivalent area in the FV91 parcellation. To map the data on the source side from the M132 atlas to the FV91 parcellation, we count the number of overlapping triangles on the F99 surface between any given pair of regions and distribute data proportionally to the amount of overlap using the F99 region overlap tool on http://cocomac.g-node.org. In the model, this FLN is mapped to the indegree \( K_{AB} \) the target area \( A \) receives from source area \( B \) divided by its total indegree, i.e., \( FLN_{AB} = K_{AB} / \sum^B K_{AB} \). Here, \( K_{AB} \) is defined as the total number of synapses between \( A \) and \( B \) divided by the total number of neurons in \( A \). On the source side, laminar connection patterns are based on CoCoMac [27, 67, 70–73] and on fractions of labeled neurons in the supragranular layers (SLN) [2]. Gaps in these data are bridged exploiting a sigmoidal relation between SLN and the logarithmized ratio of overall cell densities of the two areas, similar to [74]. We map the SLN to the ratio between the number of synapses originating in layer 2/3 and the total number of synapses between the two areas, assuming that only excitatory populations send inter-area connections, i.e., \( SLN_{AB} = \sum^i K'_{2/3i}N_i / \sum^j K'_{3j}N_j \), where the indices \( i \) and \( j \) go over the different populations within area \( A \) and \( B \), respectively. In the context of the model, we use the terms FLN and SLN to refer to the relevant relative indegrees given here. On the target side, the CoCoMac database provides data from anterograde tracing studies [27, 66, 75, 76].

Missing inputs in the model, i.e., from subcortical and non-visual cortical areas, are replaced by Poissonian spike trains, whose rate \( v_{ext} \) is a free, global parameter. In the original model all populations of a particular area receive the same indegree of external inputs \( K_{ext} \). The only exception to this rule is area TH1 where the absence of granular layer 4 is compensated by an increase of the external input to populations 2/3E and 5E by 20%. To elevate the firing rates in the excitatory populations in layers 5 and 6, we increase the external drive onto these populations. The possibility of a higher drive onto these populations is left open by the sparseness of the corresponding experimental data. We enhance the external Poisson drive of the 5E population, parametrized by the \( K_{5E,ext} \) incoming connections per target neuron (indegree), in all areas by increasing \( \kappa = K_{5E,ext}/K_{ext} \). The simultaneous increase in the drive of 6E needs to be stronger, since the firing rates in population 6E of the original model (Fig 3C) are even lower than the rates of 5E (averaged across areas: 0.09 spikes/s for 5E compared to \( 2 \cdot 10^{-5} \) spikes/s for 6E). We thus scale up \( K_{6E,ext} \) linearly with \( \kappa \) such that \( \kappa = 1.15 \) results in \( K_{6E,ext}/K_{ext} = 1.5 \).

**Author Contributions**

**Conceptualization:** JS MS SJvA MD MH.

**Data curation:** MS JS.

**Formal analysis:** MS JS.

**Funding acquisition:** MH MD.

**Investigation:** JS MS.
Methodology: JS MS MH.

Project administration: SJvA MH MD.

Resources: JS MS SJvA MH MD.

Software: MS JS SJvA MH.

Supervision: SJvA MH MD.

Validation: MS JS.

Visualization: MS JS.

Writing – original draft: JS MS.

Writing – review & editing: JS MS SJvA MH MD.

References

1. Markov NT, Ercsey-Ravasz MM, Ribeiro Gomes AR, Lamy C, Magrou L, Vezoli J, et al. A Weighted and Directed Interareal Connectiv ity Matrix for Macaque Cerebral Cortex. Cereb Cortex. 2014; 24 (1):17–36. Available from: http://cercor.oxfordjournals.org/content/24/1/17.abstract. doi: 10.1093/cercor/bhs270 PMID: 23010748

2. Markov NT, Vezoli J, Chameau P, Falchier A, Quiodran R, Huissoud C, et al. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. J Compar Neurol. 2014; 522(1):225–259. Available from: http://dx.doi.org/10.1002/cne.23458. PMID: 23983048

3. Wedeen VJ, Wang R, Schmahmann JD, Benner T, Tseng W, Dai G, et al. Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. Neuroimage. 2008; 41(4):1267–1277. doi: 10.1016/j.neuroimage.2008.03.036 PMID: 18495497

4. Axer M, Grässel D, Kleiner M, Dammers J, Dickscheid T, Reckfort J, et al. High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging. Front Neuinform. 2011; 5(34). doi: 10.3389/fninf.2011.00034 PMID: 22232597

5. Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE. Intrinsic and task-evoked network architectures of the human brain. Neuron. 2014; 83(1):238–251. doi: 10.1016/j.neuron.2014.05.014 PMID: 24991964

6. Shen K, Hutchison RM, Bezgin G, Everling S, McIntosh AR. Network structure shapes spontaneous functional connectivity dynamics. J Neurosci. 2015; 35(14):5579–5588. doi: 10.1523/JNEUROSCI.4903-14.2015 PMID: 25855174

7. van Vreeswijk C, Sompolinsky H. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. Science. 1996 December, 6; 274:1724–1726. doi: 10.1126/science.274.5293.1724 PMID: 8939866

8. Amit DJ, Brunel N. Dynamics of a recurrent network of spiking neurons and following learning. Network: Comput Neural Systems. 1997; 8:373–404. doi: 10.1088/0954-898X_8_4_003

9. Tetzlaff T, Helias M, Einevoll G, Diesmann M. Decorrelation of neural-network activity by inhibitory feedback. PLoS Comput Biol. 2012; 8(8):e1002596. doi: 10.1371/journal.pcbi.1002596 PMID: 23133368

10. Deco G, Jirsa VK. Ongoing Cortical Activity at Rest: Criticality, Multistability, and Ghost Attractors. J Neurosci. 2012; 32(10):3366–3375. doi: 10.1523/JNEUROSCI.2523-11.2012 PMID: 22399758

11. Cabral J, Klingelbach ML, Deco G. Exploring the network dynamics underlying brain activity during rest. Prog Neurobiol. 2014; 114:102–131. doi: 10.1016/j.pneurobio.2013.12.005 PMID: 24399385

12. Markov NT, Ercsey-Ravasz M, van Essen DC, Knoblauch K, Toroczkai Z, Kennedy H. Cortical High-Density Counterstream Architectures. Science. 2013; 342 (6158). Available from: http://www.sciencemag.org/content/342/6158/1238406.abstract. PMID: 24179228

13. Goulas A, Bastiani M, Bezgin G, Uylings HB, Roebroeck A, Stiers P. Comparative analysis of the macroscale structural connectivity in the macaque and human brain. PLoS Comput Biol. 2014; 10(3): e1003529. doi: 10.1371/journal.pcbi.1003529 PMID: 24676052

14. Ostojic S, Brunel N, Haikim V. How Connectivity, Background Activity, and Synaptic Properties Shape the Cross-Correlation between Spike Trains. J Neurosci. 2009; 29(33):10234–10253. doi: 10.1523/JNEUROSCI.1275-09.2009 PMID: 19692598
15. Pernice V, Staude B, Cardanobile S, Rotter S. How Structure Determines Correlations in Neuronal Networks. PLoS Comput Biol. 2011 May; 7(5):e1002059. doi: 10.1371/journal.pcbi.1002059 PMID: 21625580

16. Trousdale J, Hu Y, Shea-Brown E, Josic K. Impact of network structure and cellular response on spike time correlations. PLoS Comput Biol. 2012; 8(3):e1002408. doi: 10.1371/journal.pcbi.1002408 PMID: 22457608

17. van Albada SJ, Helias M, Diesmann M. Scalability of asynchronous networks is limited by one-to-one mapping between effective connectivity and correlations. PLoS Comput Biol. 2015; 11(9):e1004490. doi: 10.1371/journal.pcbi.1004490 PMID: 26325661

18. Schmidt M, Schuecker J, Diesmann M, Helias M. Shaping phase space of neural networks via connectivity. In: Proceedings of the 11th Göttingen Meeting of the German Neuroscience Society; 2015. p. T26 7C.

19. Dayan P, Abbott LF. Theoretical Neuroscience. Cambridge: MIT Press; 2001.

20. Kriener B, Helias M, Rotter S, Diesmann M, Einevoll G. How pattern formation in ring networks of excitatory and inhibitory spiking neurons depends on the input current regime. Front Comput Neurosci. 2014; 7(187):1–21. doi: 10.3389/fncom.2013.00187 PMID: 24501591

21. Brunel N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J Comput Neurosci. 2000; 8(3):183–208. doi: 10.1023/A:1008925309027 PMID: 10809012

22. Strogatz SH. Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering. Reading, Massachusetts: Perseus Books; 1994. doi: 10.1063/1.4823332

23. Potjans TC, Diesmann M. The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model. Cereb Cortex. 2014; 24(3):785–806. doi: 10.1093/cercor/bhs358 PMID: 23203991

24. Schmidt M, Bakker R, Diesmann M, van Albada S. Full-density multi-scale account of structure and dynamics of macaque visual cortex. arXiv preprint arXiv:151109364v4. 2016;

25. Swadlow HA. Efferent Neurons and Suspected Interneurons in Binocular Visual Cortex of the Awake Rabbit: Receptive Fields and Binocular Properties. J Neurophysiol. 1988; 59(4):1162–1187. PMID: 3373273

26. de Kock CPJ, Sakmann B. Spiking in primary somatosensory cortex during natural whisking in awake head-restrained rats is cell-type specific. Proc Natl Acad Sci USA. 2009; 106(38):16446–16450. doi: 10.1073/pnas.0904143106 PMID: 19805318

27. Fellerman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex. 1991; 1:1–47. doi: 10.1093/cercor/1.1.1 PMID: 1822724

28. Salin PA, Bullier J. Corticocortical connections in the visual system: structure and function. Physiol Rev. 1995; 75(1):107–154. PMID: 7831995

29. Tomioka R, Rockland KS. Long-distance corticocortical GABAergic neurons in the adult monkey white and gray matter. J Comp Neurol. 2007; 505(5):526–538. doi: 10.1002/cne.21504 PMID: 17924571

30. Binzegger T, Douglas R, Martin KAC. A Quantitative Map of the Circuit of Cat Primary Visual Cortex. J Neurosci. 2004; 39(24):8441–8453. doi: 10.1523/JNEUROSCI.1400-04.2004 PMID: 15456817

31. Voorhees EM. Implementing agglomerative hierarchical clustering algorithms for use in document retrieval. Information Processing & Management. 1986; 22(6):465–476. doi: 10.1016/0306-4573(86)90097-X

32. Scannell JW, Grant S, Payne BR, Baddeley R. On variability in the density of corticocortical and thalamocortical connections. Philos Trans R Soc Lond, B. 2000; 355(1393):21–35. doi: 10.1098/rstb.2000.0547 PMID: 10703042

33. Fourcaud N, Brunel N. Dynamics of the firing probability of noisy integrate-and-fire neurons. Neural Comput. 2002; 14:2057–2110. doi: 10.1162/089976602320264015 PMID: 12184844

34. van Kerkorle T, Self MW, Dagnino B, Gariel-Mahdf MA, Poort J, van der Togt C, et al. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. J Neurosci. 2014; 11(40):14332–14341. doi: 10.1073/pnas.1402773111 PMID: 25205911

35. Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, et al. Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron. 2015; 85(2):390–401. doi: 10.1016/j.neuron.2014.12.018 PMID: 25558836

36. Sanchez-Vives MV, McCormick D. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. Nat Neurosci. 2000; 3:1027–1034. doi: 10.1038/79848 PMID: 11017176

37. Beltramo R, D’Urso G, Maschio MD, Farisello P, Bovett S, Clovis Y, et al. Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex. Nat Neurosci. 2013; 16(2):227–234. Available from: http://dx.doi.org/10.1038/nn.3306 PMID: 23313909
38. Le Bon-Jego M, Yuste R. Persistently active, pacemaker-like neurons in neocortex. Front Neurosci. 2007; 1(1):123–129. doi: 10.3389/neuro.01.1.1.009.2007 PMID: 18982123

39. Lörincz ML, Gunner D, Bao Y, Connelly WM, Isaac JT, Hughes SW, et al. A Distinct Class of Slow (0.2–2 Hz) Intrinsically Bursting Layer 5 Pyramidal Neurons Determines UP/DOWN State Dynamics in the Neocortex. J Neurosci. 2015; 35(14):5442–5458. doi: 10.1523/JNEUROSCI.3603-14.2015 PMID: 25855163

40. Neske GT, Patrick SL, Connors BW. Contributions of diverse excitatory and inhibitory neurons to recurrent network activity in cerebral cortex. J Neurosci. 2015 Jan; 35(3):1089–1105. Available from: http://dx.doi.org/10.1523/JNEUROSCI.2279-14.2015. PMID: 25609625

41. Stephan KE, Kamper L, Bozkurt A, Burns GPC, Young MP, Köttter R. Advanced database methodology for the collation of connectivity data on the macaque brain (CoCoMac). Philos Trans R Soc Lond, B. 2001; 356:1159–1186. doi: 10.1098/rstb.2001.0908 PMID: 11545697

42. Bakker R, Thomas W, Diesmann M. CoCoMac 2.0 and the future of tract-tracing databases. Front Neuroinform. 2012; 6(30). doi: 10.3389/fninf.2012.00030 PMID: 23293600

43. Schmitt O, Eipert P, Kettlitz R, Leßmann F, Wree A. The connectome of the basal ganglia. Brain Structure and Function. 2014;p. 1–62.

44. Rauch A, La Camera G, Lüscher H, Senn W, Fusi S. Neocortical Pyramidal Cells Respond as Integrate-and-Fire Neurons to In Vivo Like Input Currents. J Neurophysiol. 2003; 90:1598–1612. doi: 10.1152/jn.00293.2003 PMID: 12750422

45. Jolivet R, Rauch A, Lüscher HR, Gerstner W. Predicting spike timing of neocortical pyramidal neurons by simple threshold models. J Comput Neurosci. 2006 Aug; 21(1):35–49. doi: 10.1007/s10827-006-7074-5 PMID: 16633938

46. Shriki O, Hansel D, Sompolinsky H. Rate models for conductance-based cortical neuronal networks. Neur Comp. 2003; 15:1809–1841. doi: 10.1162/0899766036075053 PMID: 14511514

47. Kuhn A, Aertsen A, Rotter S. Neuronal integration of synaptic input in the fluctuation-driven regime. J Neurosci. 2004; 24(10):2345–2356. doi: 10.1523/JNEUROSCI.3349-03.2004 PMID: 15014109

48. Morrison A, Diesmann M, Gerstner W. Phenomenological models of synaptic plasticity based on spike-timing. Biol Cybern. 2008; 98:459–478. doi: 10.1007/s00422-008-0233-1 PMID: 18491160

49. Turrigiano GG, Leslie KR, Desai NS, Rutherford LC, Nelson SB. Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature. 1998; 391:892–896. doi: 10.1038/36103 PMID: 9495341

50. Benda J, Herz VM. A universal model for spike-frequency adaptation. Neural Comput. 2003; 15:2523–2546. doi: 10.1162/089976603322395653 PMID: 14577853

51. Ostojic S. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. Nat Neurosci. 2014; 17:594–600. doi: 10.1038/nn.3658 PMID: 24561997

52. Goedeke S, Schuecker J, Helias M. Noise dynamically suppresses chaos in neural networks. arXiv. 2016;1603.01880v1 [q-bio.NC].

53. Grytskyy D, Tetzlaff T, Diesmann M, Helias M. A unified view on weakly correlated recurrent networks. Front Comput Neurosci. 2013; 7:131. doi: 10.3389/fncom.2013.00131 PMID: 24151463

54. Mastrogiuseppe F, Ostojic S. Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. arXiv. 2016;p. 1605.04221.

55. Abramowitz M, Stegun IA. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. New York: Dover Publications; 1974.

56. Magnus JR, Neudecker H. Matrix differential calculus with applications in statistics and econometrics. John Wiley & Sons; 1995.

57. Wong KF, Wang XJ. A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. J Neurosci. 2006; 26(4):1314–1328. doi: 10.1523/JNEUROSCI.3733-05.2006 PMID: 16436619

58. Sadeh S, Rotter S. Orientation selectivity in inhibition-dominated networks of spiking neurons: effect of single neuron properties and network dynamics. PLoS Comput Biol. 2015; 11(1):e1004045. doi: 10.1371/journal.pcbi.1004045 PMID: 25569445

59. Amit DJ, Brunel N. Model of Global Spontaneous Activity and Local Structured Activity During Delay periods in the Cerebral Cortex. Cereb Cortex. 1997; 7:237–252. doi: 10.1093/ercpec/7.3.237 PMID: 9143444

60. Cichocki A, Zdunek R, Phan AH, Amari SI. Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation. John Wiley & Sons; 2009. doi: 10.1002/9780470747278

61. Jülich Supercomputing Centre. JUQUEEN: IBM Blue Gene/Q® Supercomputer System at the Jülich Supercomputing Centre. Journal of large-scale research facilities. 2015; 1. Available from: http://dx.doi.org/10.17815/jlsrf-1-18.
62. Eppler JM, Pauli R, Payser A, Ippen T, Morrison A, Senk J, et al. NEST 2.8.0; 2015. Available from: http://dx.doi.org/10.5281/zenodo.32969.

63. Plesser HE, Diesmann M. Simplicity and efficiency of integrate-and-fire neuron models. Neural Comput. 2009; 21:353–359. doi: 10.1162/neco.2008.03-08-731 PMID: 19431263

64. Barbas H, Rempel-Clower N. Cortical structure predicts the pattern of corticocortical connections. Cereb Cortex. 1997; 7(7):635–646. Available from: http://cercor.oxfordjournals.org/content/7/7/635.abstract. doi: 10.1093/cercor/7.7.635 PMID: 9373019

65. Hilgetag CC, Medalla M, Beul S, Barbas H. The primate connectome in context: principles of connections of the cortical visual system. NeuroImage. 2016;p. (in press). doi: 10.1016/j.neuroimage.2016.04.017 PMID: 27083526

66. Rockland KS, Pandya DN. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res. 1979; 179:3–20. doi: 10.1016/0006-8993(79)90485-2 PMID: 116716

67. Barnes CL, Pandya DN. Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. J Compar Neurol. 1992; 318(2):222–244. doi: 10.1002/cne.903180207 PMID: 1583161

68. Van Essen DC. Windows on the brain: the emerging role of atlases and databases in neuroscience. Curr Opin Neurobiol. 2002; 12(5):574–579. doi: 10.1016/S0959-4388(02)00361-6 PMID: 12367638

69. Van Essen DC, Drury HA, Dickson J, Hanwell J, Hanlon D, Anderson CH. An integrated software suite for surface-based analyses of cerebral cortex. Journal of the American Medical Informatics Association. 8(5):443–459. doi: 10.1136/jamia.2001.0080443 PMID: 1152765

70. Suzuki WL, Amaral DG. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. J Compa r Neurol. 1994; 350(4):497–533. doi: 10.1002/cne.903500402 PMID: 7890828

71. Morel A, Bullier J. Anatomical segregation of two cortical visual pathways in the macaque monkey. Visual neuroscience. 1990; 4(06):555–578. doi: 10.1017/S09525238000005769 PMID: 2278934

72. Perkel DJ, Bullier J, Kennedy H. Topography of the afferent connectivity of area 17 in the macaque monkey: A double-labeling study. J Compar Neurol. 1986; 253(3):374–402. doi: 10.1002/cne.902530307 PMID: 3793996

73. Seltzer B, Pandya DN. Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. J Compar Neurol. 1994; 343(3):445–463. doi: 10.1002/cne.903430308 PMID: 8027432

74. Beul SF, Barbas H, Hilgetag CC. A predictive structural model of the primate connectome. arXiv preprint arXiv:151107222. 2015;

75. Jones E, Coulter J, Hendry S. Intracortical connectivity of architectonic fields in the somatic sensory, motor and parietal cortex of monkeys. J Compar Neurol. 1978; 181(2):291–347. doi: 10.1002/cne.901810206 PMID: 99458

76. Seltzer B, Pandya DN. Post-rolandic cortical projections of the superior temporal sulcus in the rhesus monkey. J Compar Neurol. 1991; 312(4):625–640. doi: 10.1002/cne.903120412 PMID: 1761745