The attractor states of the functional brain connectome

Robert Englert
University Medicine Essen

Tamas Spisak[1]
University Medicine Essen

Tuesday 1st August, 2023

Abstract

Abstract:
todo

Keywords

Key Points:

- We present a simple yet powerful computational model large-scale brain dynamics
- The model computes "activity flow" across brain regions using a continuous Hopfield artificial neural network.
- Instead of training the network weights to solve specific tasks, they are initialized with empirical functional brain connectivity.
- It defines an energy level for any arbitrary brain activation patterns and a trajectory towards one of the finite number of stable patterns (attractor states) that minimize this energy
- The model captures the dynamic repertoire of the brain in resting conditions
- It conceptualizes both task-induced and pathological changes in brain activity as a shift in these dynamics.
- We validate the model through eight studies involving approximately 2000 participants.

1 Introduction

Brain function is characterized by the continuous activation and deactivation of anatomically distributed neuronal populations. While the focus of related research is often on the direct mapping between changes in the activity of a single brain area and a specific task or condition, in reality, regional activation never seems to occur in isolation [Bassett and Sporns, 2017].

Irrespective of the presence or absence of explicit stimuli, brain regions appear to work in concert, giving rise to a rich and complex spatiotemporal fluctuation [Gutierrez-Barragan et al., 2019]. This fluctuation is neither random, nor stationary over time [Liu and Duyn, 2013, Zalesky et al., 2014]. It exhibits quasi-periodic properties [Thompson et al., 2014], with a limited number of recurring patterns known as "brain states" [Greene et al., 2023, Vidaurre et al., 2017, Liu and Duyn, 2013, Richiardi et al., 2011].

From hidden Markov models, to point-process analyses, a wide variety of descriptive techniques have been previously employed to characterize whole-brain dynamics. [Smith et al., 2012, Vidaurre et al., 2017, Liu and Duyn, 2013, Chen et al., 2018], providing accumulating evidence not only for the existence of dynamic

[1]Correspondence to: tamas.spisak@uk-essen.de

brain states but also for their clinical significance. [Hutchison et al., 2013, Barttfeld et al., 2015, Meer et al., 2020]. However, the underlying driving forces remain elusive due to the descriptive nature of such studies.

Brain state dynamics can be assessed with multiple techniques, such as dynamic connectivity analysis (), independent component analysis [Smith et al., 2012], hidden markov models [Vidaurre et al., 2017], clustering [Chen et al., 2018] or point-process analyses to capture co-activation patterns (CAPs, [Liu and Duyn, 2013, Chen et al., 2015, Liu et al., 2013, Meer et al., 2020].

Questions regarding the mechanisms, that cause these remarkable dynamics, can be addressed through approaches that are based on computational models, which have the potential to shift our understanding from mere associations to causal explanations. Conventional computational approaches attempt to solve this puzzle by going all the way to the biophysical properties of single neurons, and aim to construct a model of larger neural populations, or even the entire brain [Breakspear, 2017]. Although these approaches have shown numerous successful applications [Kriegeskorte and Douglas, 2018, Heinz et al., 2019], the estimation of all the free parameters in such models presents a grand challenge. This hampers the ability of these techniques to effectively bridge the gap between explanations at the level of single neurons and the complexity of behavior [Breakspear, 2017].

An alternative approach, known as "neuroconnectomism" [Doerig et al., 2023] shifts the emphasis from "biophysical fidelity" of models to "cognitive/behavioral fidelity" [Kriegeskorte and Douglas, 2018], by using artificial neural networks (ANNs) that are trained to perform various tasks, as brain models. While this novel paradigm has already made significant contributions to expanding our understanding of the general computational principles of the brain (see [Doerig et al., 2023], the need to train ANNs for specific tasks inherently limits their ability to explain the spontaneous, and largely task-independent, macro-scale dynamics of neural activity [Richards et al., 2019].

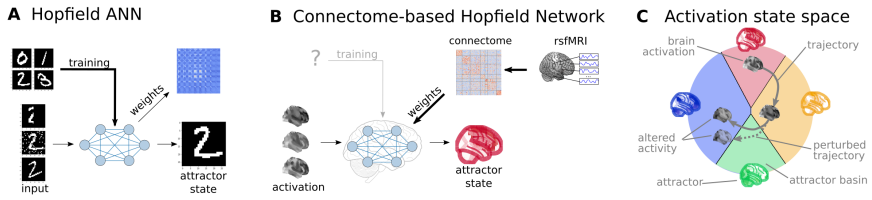
In this work, we adopt a middle ground between traditional computational modeling and neuroconnectionism to investigate the phenomenon of brain dynamics. Similar to neuroconnectionism, our objective is not to attain a comprehensive bottom-up understanding of neural mechanisms. Therefore, we utilize an artificial neural network (ANN) as a high-level computational model of the brain (Figure 1A). However, we do not train our ANN for a specific task, instead we set its weights empirically, with data based on the "activity flow" [Cole et al., 2016, Ito et al., 2017] across regions within the functional brain connectome, as measured with functional magnetic resonance imaging (fMRI, Figure 1B).

We employ this neurobiologically motivated ANN architecture, based on the established architecture of a continuous-space Hopfield network [Hopfield, 1982, Krotov, 2023]. Within this architecture, the topology of the functional connectome naturally establishes an "energy" level for any arbitrary activation patterns and determines a trajectory towards one of the finite number of stable patterns, known as attractor states, that minimize this energy. In the presence of weak noise, the system does not reach equilibrium (i.e. it does not converge to an attractor state). Instead, it traverses extensive regions of the state space, with dynamics influenced by multiple attractor states, arising form the topology of the functional brain connectome (Figure 1C). Through this walk across the state space, our model also offers a natural explanation for brain state dynamics.

In this simplistic yet powerful framework, both spontaneous and task-induced brain dynamics can be conceptualized as a winding, high-dimensional path that meanders on the energy landscape, restricted by the "gravitational pull" of the attractors states. The framework provides a generative model for both resting state and task-related brain dynamics, offering novel perspectives on the mechanistic origins of resting state brain states and task-based activation maps.

In the present work, we first explore the attractor states of the functional brain connectome and construct a low-dimensional representation of the energy landscape. Subsequently, we rigorously test the proposed model through a series of experiments, conducted on data obtained from 8 experimental and clincial studies, encompassing a total of n≈2000 individuals. These analyses encompass the evaluation of robustness and replicability, testing the model's ability to reconstruct various characteristics of resting state brain dynamics, as well as its capacity to detect and explain changes induced by experimental tasks or alterations characteristic to brain disorders.

These experiments provide converging evidence for the validity of connectome-based Hopfield networks (CBH) as models of brain dynamics, and highlight their potential to provide a fresh perspective on a wide range of research questions in basic and translational neuroscience.



A Hopfield artificial neural networks (ANNs) are a form of recurrent ANNs that serve as content-addressable ("associative") memory systems. Hopfield networks can be trained to store a finite number of patterns (e.g. via Hebbian learning). During the training procedure, the weights of the Hopfield ANN are trained so that the stored patterns become stable attractor states of the network. Thus, when the trained network is presented partial or noisy variations of the stored patterns, it can effectively reconstruct the original pattern via an iterative relaxation procedure that converges to the attractor states. **B** We consider regions of the brain as nodes of a Hopfield network. Instead of training the Hopfield network to specific tasks, we use the set its weights empirically, with the interregional activity flow estimated via functional brain connectivity. Following form the strong analogies between the relaxation rule of Hopfield networks and the activity flow principle that links activity to connectivity in brain networks, we propose the constructed connectome-based Hopfield (CBH) network as a computational model for macro-scale brain dynamics.
C The proposed computational framework assigns an energy level, an attractor state and a position in a low-dimensional embedding to brain activation patterns. Additionally, it models how the entire state-space of viable activation patterns is restricted by the dynamics of the system and how alterations in activity and/or connectivity modify these dynamics.

A Hopfield artificial neural networks (ANNs) are a form of recurrent ANNs that serve as content-addressable ("associative") memory systems. Hopfield networks can be trained to store a finite number of patterns (e.g. via Hebbian learning). During the training procedure, the weights of the Hopfield ANN are trained so that the stored patterns become stable attractor states of the network. Thus, when the trained network is presented partial or noisy variations of the stored patterns, it can effectively reconstruct the original pattern via an iterative relaxation procedure that converges to the attractor states. **B** We consider regions of the brain as nodes of a Hopfield network. Instead of training the Hopfield network to specific tasks, we use the set its weights empirically, with the interregional activity flow estimated via functional brain connectivity. Following form the strong analogies between the relaxation rule of Hopfield networks and the activity flow principle that links activity to connectivity in brain networks, we propose the constructed connectome-based Hopfield (CBH) network as a computational model for macro-scale brain dynamics.
C The proposed computational framework assigns an energy level, an attractor state and a position in a low-dimensional embedding to brain activation patterns. Additionally, it models how the entire state-space of viable activation patterns is restricted by the dynamics of the system and how alterations in activity and/or connectivity modify these dynamics.

Figure 1: **Connectome-based Hopfield networks as models of macro-scale brain dynamics.**

A Hopfield artificial neural networks (ANNs) are a form of recurrent ANNs that serve as content-addressable ("associative") memory systems. Hopfield networks can be trained to store a finite number of patterns (e.g. via Hebbian learning). During the training procedure, the weights of the Hopfield ANN are trained so that the stored patterns become stable attractor states of the network. Thus, when the trained network is presented partial or noisy variations of the stored patterns, it can effectively reconstruct the original pattern via an iterative relaxation procedure that converges to the attractor states. **B** We consider regions of the brain as nodes of a Hopfield network. Instead of training the Hopfield network to specific tasks, we use the set its weights empirically, with the interregional activity flow estimated via functional brain connectivity. Following form the strong analogies between the relaxation rule of Hopfield networks and the activity flow principle that links activity to connectivity in brain networks, we propose the constructed connectome-based Hopfield (CBH) network as a computational model for macro-scale brain dynamics.
C The proposed computational framework assigns an energy level, an attractor state and a position in a low-dimensional embedding to brain activation patterns. Additionally, it models how the entire state-space of viable activation patterns is restricted by the dynamics of the system and how alterations in activity and/or connectivity modify these dynamics.

2 Results

2.1 Connectome-based Hopfield network as a model of brain dynamics

First, we explored the attractor states of the functional brain connectome in a sample of n=41 healthy young participants (Table **??**). We estimated interregional activity flow [Cole et al., 2016, Ito et al., 2017] as the study-level average of regularized partial correlations among the resting state fMRI timeseries of m = 122 functionally defined brain regions (BASC brain atlas, see Methods for details). We then used the standardized functional connectome as the w_{ij} weights of a continuous-state Hopfield network [Hopfield, 1982, Koiran, 1994] consisting of m neural units, each having an activity $a_i \in [-1, 1]$. Hopfield networks can be initialized by an arbitrary activation pattern (consisting of m activation values) and iteratively updated until convergence is reached (i.e. "relaxed"), according to the following equation:

$$\dot{a}_i = S(\beta \sum_{j=1}^{m} w_{ij} a_j - b_i) \tag{1}$$

where $\dot{a}_i$ is the activity of neural unit i in the next iteration and $S(a_j)$ is the sigmoidal activation function $S(a) = tanh(a)$ and b_i is the bias of unit i and β is the so-called temperature parameter. For the sake of simplicity, we set $b_i = 0$ in all our experiments. We refer to this architecture as a connectome-based Hopfield network (CBH network). Importantly, the relaxation of a CBH network can be conceptualized as the repeated application of the activity flow principle [Cole et al., 2016, Ito et al., 2017] , simultaneously for all regions: $\dot{a}_i = \sum_{j=1}^{m} w_{ij} a_j$. The update rule also exhibits strong analogies with the inner workings of neural mass models [Breakspear, 2017] as applied e.g. in dynamic causal modeling(see Discussion for further details).

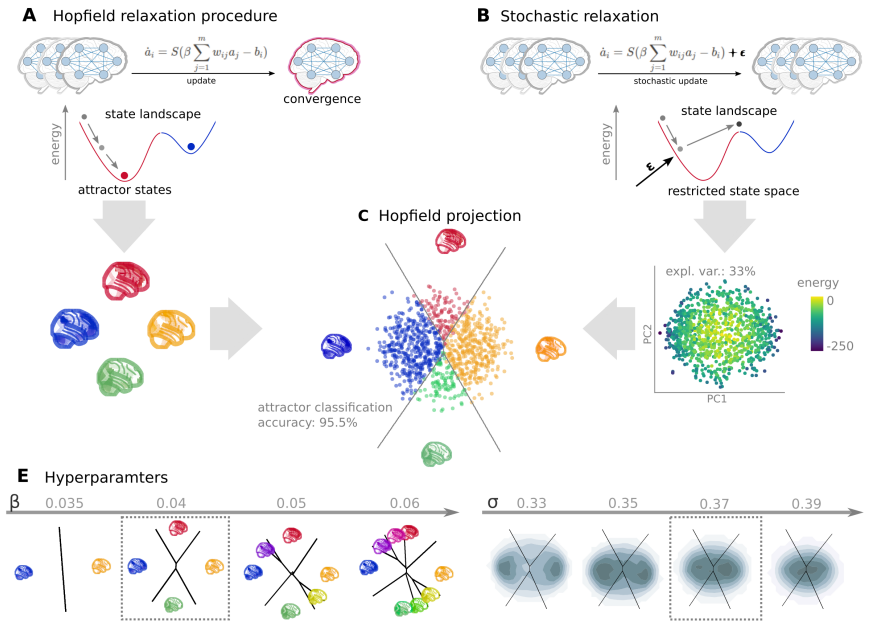
Hopfield networks assign an energy value to each possible activity configuration (see Methods), which decreases during the relaxation procedure until reaching an equilibrium state with minimal energy (Figure 2.1A, top panel, [Hopfield, 1982, Koiran, 1994]. We used a large number of random initializations to obtain all possible attractor states of the connectome-based Hopfield network in study 1 (Figure 2.1A, bottom panel).

Consistent with theoretical expectations, we observed that increasing the temperature parameter β led to an increasing number of attractor states ((Figure 2.1E, left), appearing in symmetric pairs (i.e. $a_i^{(1)} = -a_i^{(2)}$).For simplicity, we set the temperature parameter for the rest of the paper to a value resulting in 4 distinct attractor states ($\beta = 0.4$).

Connectome-based Hopfield networks, without any modifications, always converge to an equilibrium state. To incorporate stochastic fluctuations in neuronal activity [Robinson et al., 2005], we introduce weak Gaussian noise to the CBH relaxation procedure. This procedure, referred to as stochastic relaxation, prevents the system from reaching equilibrium and, somewhat similarly to Stochastic DCM [Daunizeau et al., 2012], induces complex CBH system dynamics (equivalent to brain activity fluctuations in our framework) that may traverse extensive regions of the state space, determined by the "gravity field" (basins) of multiple attractor states (Figure 2.1B).

We hypothesise that the resulting dynamics capture essential characteristics of spontaneous activity fluctuations in the brain and can serve as a valuable generative computational model for large-scale brain dynamics. To sample the resulting state space, we obtained 100,000 iterations of the stochastic relaxation procedure with a Hopfield network initialized with the mean functional connectome in study 1. Next, in order to enhance interpretability, we conducted a principal component analysis (PCA) on the resulting state space sample and obtained the first two principal components. These components were used to construct a low-dimensional embedding. Largely independent on the free parameter σ (variance of the noise), the first two principal components (PCs) explained around 15% of the variance in the state space, with attractor states (minimal energy) located at the extremes of the PCs (Figure 2.1B, bottom plot).

The PCA embedding exhibited high consistency across different values of β and σ (Figure 2.1E). For all subsequent analyses, we set $\sigma = 0.37$, which was determined through a coarse optimization procedure aimed at reconstructing the bimodal distribution of empirical data in the same projection (Figure 2.1E, see Methods for details). On the low-dimensional embedding, which we refer to as the *Hopfield projection*, we observed a clear separation of the attractor states (Figure 2.1C), with the two symmetric pairs of attractor states located at the extremes of the first and second PC. To map the attractor basins on the space spanned by the first two PCs (Figure 2.1C), we obtained the attractor state of each point visited during the stochastic



A Top: During so-called relaxation procedure, activities in the nodes of a connectome-based Hopfield (CBH) network are iteratively updated based on the activity of all other regions and the connectivity between them. The energy of a connectome-based Hopfield network decreases during the relaxation procedure until reaching an equilibrium state with minimal energy, i.e. an attractor state. Bottom: Four attractor states obtained by a CBH, initialized with a group-level functional connectivity matrix from Table **??** (n=44). **B** Top: Similarly to stochastic dynamic causal modeling, in presence of weak noise (stochastic update), the system does not converge to an equilibrium anymore. Instead, it the system transverses on the state landscape in a way restricted by the topology of the connectome and the "gravitational pull" of the attractor states. Bottom: We sample the state space by running the stochastic relaxation procedure for an extended amount of time (e.g. 100.000 consecutive stochatsic updates), each point representing a possible activation configuration (state). To construct a low-dimensional representation of the state space, we take the first principal components of the simulated activity patterns. The first two principal components explain approximately 55-85% of the variance of state energy (depending on the noise parameter σ, see Supplementary Material **X**). **C** We map all states of the state space sample to their corresponding attractor state, with the conventional Hopfield relaxation procedure (A). The four attractor states are also visualized in their corresponding position on the PCA-based projection. The first two principal components yield a clear separation of the attractive state basins (cross-validated classification accuracy: 95.5%, Supplementary Material **X**). We refer to the resulting visualization as the Hopfiled projection and use it to visualize CBH-derived and empirical brain dynamics throughout the rest of the manuscript. **E** At its simpliest form, the CBH framework entails only two free hyperparamters: the temperature parameter β (left) that controls the number of attractor states and the noise parameter of the stochastic relaxation σ. To avoid overfitting these parameters to the empirical data, we set $\beta = 0.04$ and $\sigma = 0.37$ for the rest of the paper.

A Top: During so-called relaxation procedure, activities in the nodes of a connectome-based Hopfield (CBH) network are iteratively updated based on the activity of all other regions and the connectivity between them. The energy of a connectome-based Hopfield network decreases during the relaxation procedure until reaching an equilibrium state with minimal energy, i.e. an attractor state. Bottom: Four attractor states obtained by a CBH, initialized with a group-level functional connectivity matrix from Table **??** (n=44). **B** Top: Similarly to stochastic dynamic causal modeling, in presence of weak noise (stochastic update), the system does not converge to an equilibrium anymore. Instead, it the system transverses on the state landscape in a way restricted by the topology of the connectome and the "gravitational pull" of the attractor states. Bottom: We sample the state space by running the stochastic relaxation procedure for an extended amount of time (e.g. 100.000 consecutive stochatsic updates), each point representing a possible activation configuration (state). To construct a low-dimensional representation of the state space, we take the first principal components of the simulated activity patterns. The first two principal components explain approximately 55-85% of the variance of state energy (depending on the noise parameter σ, see Supplementary Material **X**). **C** We map all states of the state space sample to their corresponding attractor state, with the conventional Hopfield relaxation procedure (A). The four attractor states are also visualized in their corresponding position on the PCA-based projection. The first two principal components yield a clear separation of the attractive state basins (cross-validated classification accuracy: 95.5%, Supplementary Material **X**). We refer to the resulting visualization as the Hopfiled projection and use it to visualize CBH-derived and empirical brain dynamics throughout the rest of the manuscript. **E** At its simpliest form, the CBH framework entails only two free hyperparamters: the temperature parameter β (left) that controls the number of attractor states and the

relaxation and fit a multinomial logistic regression model to predict the attractor state from the first two PCs. The resulting model demonstrated high prediction accuracy, achieving an out-of-sample accuracy of 96.5%. The attractor basins were visualized by using the decision boundaries obtained from this model. (Figure 2.1C). We propose the Hopfield projection depicted on (Figure 2.1C) as a simplified representation of brain dynamics, and use it as a basis for all subsequent analyses in this work.

2.2 Reconstruction of resting state brain dynamics

The obtained attractor states closely resemble frequently described brain patterns. ({numref}rest-validityA). The first pair of attractors (mapped on PC1) resemble the two complementary "macro" systems described by and as well as the two primary brain states previously described by . This state-pair has previously been described as an "extrinsic" system which exhibits a stronger direct connection to the immediate sensory environment and an "intrinsic" system, whose activity is primarily associated with dynamic changes in higher-level internal context and closely linked to the default mode network. The other pair of attractors spans an orthogonal axis connecting regions that are commonly associated with perception–action cycles

Importantly, the discovered attractor states demonstrate a remarkable level of replicability (mean Pearson's correlation 0.93) across the discovery datasets (study 1) and two independent replication datasets (Table **??**, Figure 2.2C). This is in line with the previously described, high robustness of Hopfield networks to noisy input [Hopfield, 1982] as well as their ability to tolerate corrupted weights. (**ref**, Supplementary Material **X**) and renders CBH networks highly promising for computational analyses of brain dynamics.

Further analysis in study 1 showed that connectome-based Hopfield models accurately reconstructed multiple characteristics of true resting-state data. First, the Hopfield projection accounted for a substantial amount of variance in the real resting-state fMRI data in study 1 (mean $R^2 = 0.15$) and generalized well to study 2 (mean $R^2 = 0.13$) and study 3 (mean $R^2 = 0.12$) (Figure 2.2E). Remarkably, the explained variance of the Hopfield projection significantly exceeded that of a PCA performed directly on the real resting-state fMRI data itself (Figure 2.2E).

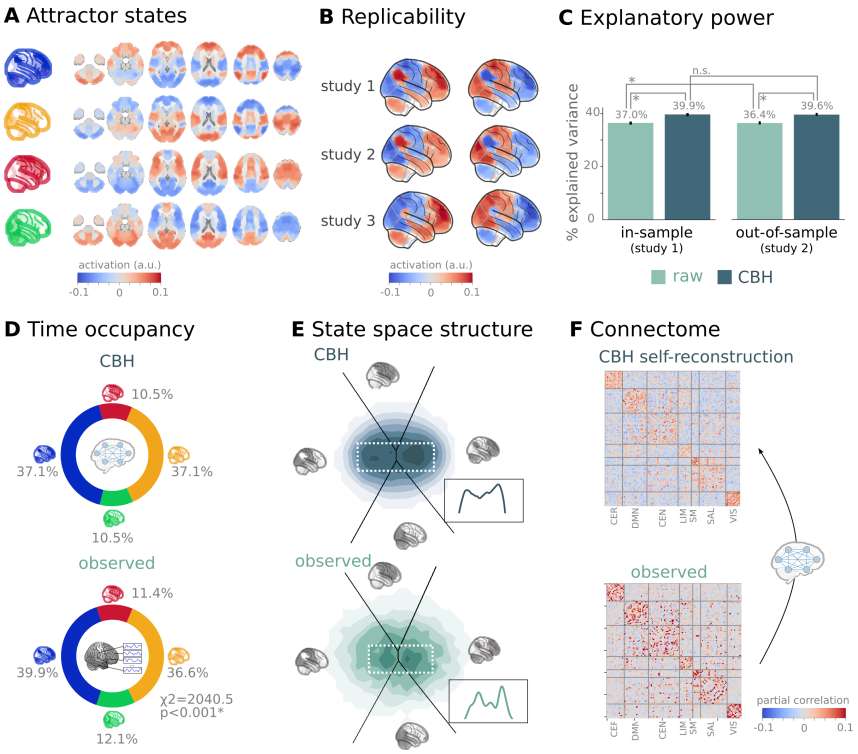
Second, CBH analyses accurately reconstruts true resting state brain state dynamics. During stochastic relaxation, the CBH network spends approximately three-quarters of the time on the basis of the first two attractor states, with an equal distribution between them. The remaining one-quarter of the time is spent on the basis of the second pair of attractor states, also equally distributed. This ratios match the properties of real resting state data very closely. We obtained normalized and cleaned mean timeseries in $m = 122$ regions from all participants in study 1 and calculated the attractor state of each time-frame via the CBH network. We observed highly similar temporal occupancies to those predicted by the model (^{2}-test with the null hypothesis of uniform occupancies: p<0.00001, Figure 2.2B).

The validity of the CBH model extends beyond the accurate reconstruction of brain state temporal occupancies. It successfully reproduces the bimodal distribution observed in the real resting-state fMRI data when projected onto the Hopfield projection (Figure 2.2F and Figure 2.1E). These findings suggest that brain dynamics are indeed governed by a limited number of attractor states that emerge from the flow of activity across functional connectivity networks.

Finally, during the stochastic relaxation procedure, CBH models were found to generate regional time series that preserve the partial correlation structure of the real functional connectome used for network initialization. This important result indicates that a dynamic system in which activity flows across nodes of a complex network inevitably "leaks" its underlying structure into the activity time series, providing a high level of construct validity for the proposed approach (Figure 2.2D).

The ability of the connectome-based Hopfield model to reconstruct all these characteristics of real data is remarkable, especially considering that the model was neither trained to reconstruct nor informed about any spatial (bi-modal distribution, explanatory performance) or temporal patterns (temporal state occupancy) of the brain.

The CBH model was solely provided with the functional connectome as initialization and constraint during stochastic relaxation. The model's ability to accurately reconstruct characteristics of resting-state brain dynamics, which are not explicitly encoded in the connectome, strongly suggests that it captures essential relationships between the topology of the brain's connectome and the dynamics of its activation.



A The four attractor states of the connectome-based Hopfiled (CBH) network from study 1 reflect brain activation patterns with a high neurobiologicasl relevance, resembling to sub-systems previously described as being associated for "internal context" (blue), "external context" (yellow), "action/execution" (red) and "perception" (green) [Golland et al., 2008, Cioli et al., 2014, Chen et al., 2018, Fuster, 2004]. **B** The attractor states show excellent replicability in two external datasets (study 2 and 3, mean correlation 0.93). **C** The Hopfield projection (first two PCs of the CBH state space) explains more variance ($p<0.0001$) in the real resting state fMRI data than principal components derived from the real resting state data itself and generalizes better ($p<0.0001$) to out-of-sample data (study 2). Error bars denote 99% bootstrapped confidence intervals. **D** The CBH analysis accurately predicts ($p<0.0001$) the fraction of time spent on the basis of the four attractor states in real restring state fMRI data (study 1) and **E** reconstructs the characteristic bimodal distribution of the real resting state data. **F** CBH networks perform self-reconstruction: the timeseries resulting from the stochastic relaxation procedure mirror the co-variance structure of the functional connectome the CBH network was initialized with.

A The four attractor states of the connectome-based Hopfiled (CBH) network from study 1 reflect brain activation patterns with a high neurobiologicasl relevance, resembling to sub-systems previously described as being associated for "internal context" (blue), "external context" (yellow), "action/execution" (red) and "perception" (green) [Golland et al., 2008, Cioli et al., 2014, Chen et al., 2018, Fuster, 2004]. **B** The attractor states show excellent replicability in two external datasets (study 2 and 3, mean correlation 0.93). **C** The Hopfield projection (first two PCs of the CBH state space) explains more variance ($p<0.0001$) in the real resting state fMRI data than principal components derived from the real resting state data itself and generalizes better ($p<0.0001$) to out-of-sample data (study 2). Error bars denote 99% bootstrapped confidence intervals. **D** The CBH analysis accurately predicts ($p<0.0001$) the fraction of time spent on the basis of the four attractor states in real restring state fMRI data (study 1) and **E** reconstructs the characteristic bimodal distribution of the real resting state data. **F** CBH networks perform self-reconstruction: the timeseries resulting from the stochastic relaxation procedure mirror the co-variance structure of the functional connectome the CBH network was initialized with.

Figure 3: **Connectome-based Hopfield networks reconstruct characteristics of real resting state brain activity.**

A The four attractor states of the connectome-based Hopfiled (CBH) network from study 1 reflect brain activation patterns with a high neurobiologicasl relevance, resembling to sub-systems previously described as being associated for "internal context" (blue), "external context" (yellow), "action/execution" (red) and "perception" (green) [Golland et al., 2008, Cioli et al., 2014, Chen et al., 2018, Fuster, 2004]. **B** The attractor states show excellent replicability in two external datasets (study 2 and 3, mean correlation 0.93). **C** The Hopfield projection (first

2.3 An explanatory framework for task-based brain activity

The proposed framework offers a natural account for how activation patterns in the brain dynamically emerge form the underlying functional connectivity. To illustrate this, we obtained task-based fMRI data from a study by [Woo et al., 2015a] (Table **??**, n=33, see Figure 2.2), investigating the neural correlates of pain and its self-regulation. We found that time-frames obtained from periods with pain stimulation (taking into account hemodynamics, see Methods for details) locate significantly differently on the Hopfield projection than time-frames obtained from periods without pain stimulation (permutation test, $p<0.001$, Figure 2.3A, left). Energies, as defined by the Hopfield model, were also significantly different between the two conditions (permutation test, $p<0.001$), with higher energies during pain stimulation.

When participants were instructed to up- or down-regulate their pain sensation (resulting in increased and decreased pain reports and differential brain activity in the nucleus accumbens, NAc, (see [Woo et al., 2015a] for details) we observed further changes of the location of momentary brain states on the Hopfield-projection (permutation test, $p<0.001$, Figure 2.3A, right). Interestingly, self-regulation did not manifest in significant energy changes (permutation test, $p=0.36$).

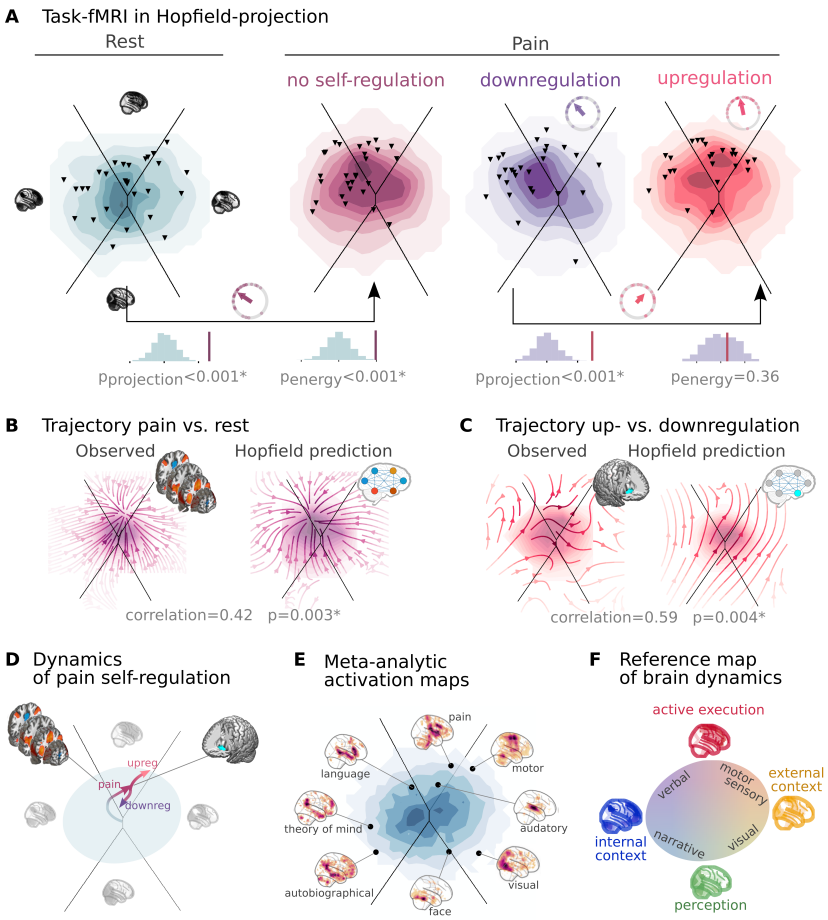
These results provide an intuitive account for how the underlying functional connectivity of the brain can give rise to different activation patterns, depending on the current (extrinsic or intrinsic) input. In the CBH framework, change in input (i.e. a task or stimulation) does not simply switch to the brain into a distinct "mode" of operation but acts as a perturbation of the system's dyanmics, resulting in mean activations changes that are only reliable measurable over an extended period of time, as done by conventional task-based fMRI analyses.

The proposed framework offers much more than visualization and inference of resting state and task based data on the Hopfield projection. It provides a generative model for observed activity changes, enabling the prediction of brain activity under different conditions. To illustrate this, we used the CBHs model to simulate brain activity during pain stimulation and self-regulation. First, we registered the frame-to-frame transitions in the real fMRI data for all four conditions: rest, pain without self-regulation, downregulation, and upregulation. We then transformed these transitions into the Hopfield embedding, resulting in two-dimensional vectors on the Hopfield projection for each transition.

Next, we evaluated the average direction in different segments of the projection on a 6x6 grid. Subsequently, we computed the difference between the mean directions observed during rest and pain (without regulation, Figure 2.3B, left side), as well as between down- and upregulation (Figure 2.3C, left side). This analysis unveiled non-linear trajectory patterns, indicating the most probable direction the brain follows from a specific state (activity pattern) in a particular condition (pain without self-regulation or upregulation), in comparison to the reference state (rest and downregulation, respectively). In the case of pain versus rest, brain activity tends to gravitate towards a "ghost attractor" situated near the Hopfield projection of a typical pain activation map (see e.g. Figure 2.3E). In terms of attractor states, this belongs to the basin of attractor corresponding to action/execution. In case of up vs. downregulation, brain activity is pulled generally towards a similar direction, but with a lack of a clear ghost attractor, potentially resulting in more extreme states.

Next, our objective was to evaluate the extent to which the proposed framework can reconstruct these non-linear dynamics. To simulate the alterations in brain dynamics during pain stimulation, we acquired a meta-analytic pain activation map [Zunhammer et al., 2021] and incorporated it as an additional signal, along with Gaussian noise, during the stochastic relaxation procedure. While incorporating such a signal naturally induces a minor linear shift on the Hopfield projection for each state generated during the stochastic relaxation procedure, this alone could only marginally explain the observed nonlinear dynamics in the real data (Supplementary material X). After conducting a coarse-grained optimization across five different signal-to-noise (SNR) values (logarithmically spaced between 0.001 and 0.1), we found that by adding a minimal amount of signal (SNR = 0.01), the CbH model achieved a remarkably precise reconstruction of the observed non-linear disparities in brain dynamics between the pain and rest conditions, encompassing the characteristic pain related "ghost attractor". (Spearman's $\rho = 0.42$, $p=0.003$, Figure 2.3B, right side).

The same model was also able to reconstruct the observed non-linear differences in brain dynamics between the up- and downregulation conditions (Spearman's $\rho = 0.59$, $p=0.004$) without any further optimization (SNR=0.01, Figure 2.3C, right side). The only change we made to the model was the addition (downregulation) or subtraction (upregulation) of activation in the NAc (the region in which [Woo et al., 2015a] observed significant changes between up- and downregulation).



A Functional MRI time-frames during pain stimulation from Table **??** (second Hopfiled projection plot) and self-regulation (third and fourth) locate significantly differently on the Hopfield projection than brain states during rest (first projection, permutation test, p<0.001 for all). Energies, as defined by the Hopfield model, are also significantly different between rest and the pain conditions (permutation test, p<0.001), with higher energies during pain stimulation. Triangles denote participant-level mean activations in the various blocks (corrected for hemodynamics). Circle plots show the directions of the change for each individual (points) as well as the mean direction across participants (arrow), as compared to the reference state (downregulation for the last circle plot, rest for all other circle plots). **B** The average difference between the characteristic directions of the single time-frames on the Hopfield projection reveal a non-linear flow difference between pain the brain dynamics during pain and rest (left). When introducing weak pain-related signal in the CBH network during stochastic relaxation, it accurately reproduces these non-linear flow differences (right). **C** Similarly simulating activity in the nucleaus accumbens (NAc) reconstructs a non-linear flow difference between up- and downregulation (left). When introducing weak self-regulation-related signal similar to the observed dynamics (characterized by NAc activation differences, as observed by [Woo et al., 2015a]. **D** Schematic representation of brain dynamics during pain and its up- and downregulation, visualized on the Hopfield projection. Pain shifts spontanous brain dynamics towards the "action" subsystem, converging to a putative "ghost attractor of pain". Up-regulation by NAc de-activation exerts force towards a similar dircetionm while down-regulation by NAc activation exhibist an opposite effect on brain dynamics, leading to the brain less frequent "visiting" pain-associated states. **E** Visualizing meta-analyitic activation maps on the Hopfiled projection informs our theoretical interpretative framework **F** for spontaneous and task-based brain dynamics. In the proposed framework, taks-based activity is not a mere response to external stimuli in certain brain locations but a perturbation of the brain's charcteristic dynamic trajectories. In this framework, conventional task-based fMRI analyses capture mean differences of the whole brain dyanmics, resulting in the widely reported focal "activation maps" thought to be specific to various taks and stimuli. In the CBH framework, the brain's characteristic trajectories are constrained by the underlying functional connectivity and only perturbed by external inpout, rather than predestinated.

A Functional MRI time-frames during pain stimulation from Table **??** (second Hopfiled projection plot) and self-regulation (third and fourth) locate significantly differently on the Hopfield projection than brain states during rest (first projection, permutation test, p<0.001 for all). Energies, as defined by the Hopfield model, are also significantly different between rest and the pain conditions (permutation test, p<0.001), with higher energies during pain stimulation. Triangles denote participant-level mean activations in the various blocks (corrected for hemodynamics). Circle plots show the directions of the change for each individual (points) as

These findings offer a novel insight into the neural mechanisms underlying pain and its self-regulation, providing a mechanistic explanation for the involvement of both nociception-related regions and the NAc (nucleus accumbens) in pain regulation. (Figure 2.3D). Additionally, these findings emphasize that the conceptual differentiation between resting and task states may, to a considerable extent, be an artificial dichotomy. Instead, the brain remains in a continuous state of flux, which is not radically altered by task states, even in the presence of highly salient stimuli such as pain.

To provide a comprehensive picture on how other tasks map onto the Hopfield projection, we obtained various task-based meta-analytic activation maps from Neurosynth (see Supplementary material X for details) and plotted them on the Hopfield projection (Figure 2.3E). This analysis demonstrated that the Hopfield projection can effectively visualize and quantify the dynamics of various cognitive processes, encompassing sensory, motor, cognitive, and social domains. Furthermore, the analysis revealed that the two primary axes of the projection correspond well to the differentiation between internal and external context, as well as the perception-action axis, respectively.

In this coordinate system, visual processing is labeled "external-perception", sensory-motor processes "external-active", language, verbal cognition and working memory is labelled "internal-active" and long-term memory as well as social and autobiographic narrative fall into the "internal-perception" regime (Figure 2.3F).

These results highlight a very powerful feature of the proposed generative framework, namely that it can be used to simulate and predict brain activity under different conditions. Predicting the effect of lower or higher level of activity in certain regions, or lower or higher connectivity among them, on global brain dynamics and responses to various tasks provides unprecedented opportunities for forecasting the effect of interventions, such as pharmacological or non-invasive brain stimulation, on brain function.

2.4 Clinical relevance

In our final analysis, we provide a brief outlook towards the potential clinical applications of CBH analysis. We analyzed three large public clinical databases as provided by the Autism Brain Imaging Data Exchange (Table **??**: ABIDE, [Di Martino et al., 2014], the Centers of Biomedical Research Excellence (Table **??**: COBRE, [Aine et al., 2017]) and the Alzheimer's Disease Neuroimaging Initiative (Table **??**: ADNI, [Petersen et al., 2010]). Resting state fMRI data of patients with autism spectrum disorder (ASD), schizophrenia (SCZ) and Alzheimer's disease (AD) was contrasted to their respective control groups (typically developing controls for ASD, healthy control participants for SCH and individuals with mild cognitive impairment (MCI), respectively). In all three datasets, we used the CBH model from study 1 and projected the fMRI timeseries of all involved participants onto the Hopfield projection. For each participant, we obtained the average activation of all time-frames belonging to the same attractor state (4 maps per participant) and compared these across groups with permutation tests, Bonferroni corrected across brain regions and attractor states (122*4 comparisons).

We found several significant differences the mean attractor activation of patients as compared to the respective controls. In ASD, all four attractor activation maps showed significant differences (Figure 5A, **table**), characterized by altered activation in the *precuneus, posterior congulate, sensory-motor system, posterior insula, and cerebellum.*

In SCZ, the most prominent differences were found in the subsystem for internal context, with elevated activity of regions that are not typically active in this state, including the *thalamus, the striatum and several cortical regions* (Figure 5B, **table**). Additional activation increases in *visual and motor* areas were observed in the active inference subsystem.

In the AD vs. MCI comparison, we found significant differences in two of the four attractor activation maps (Figure 5C, **table**), indicating changes in the resting state activity of subsystems for passive inference and internal context (both of which together host long-term memory processes, see Figure 2.3F). At the regional level, differences are characterized by altered activation in the *dorsolateral prefrontal cortex (DLPFC) and the cerebellum.*

3 Discussion

Regions of the brain engage in an ongoing exchange of information, leading to co-activations that are commonly known as functional connectivity. The quantity of information exchanged between brain regions is not uniform, but rather exhibits substantial variation among different pairs of regions, encompassing the intricate network referred to as the functional connectome. In this study, we have introduced a simple yet robust

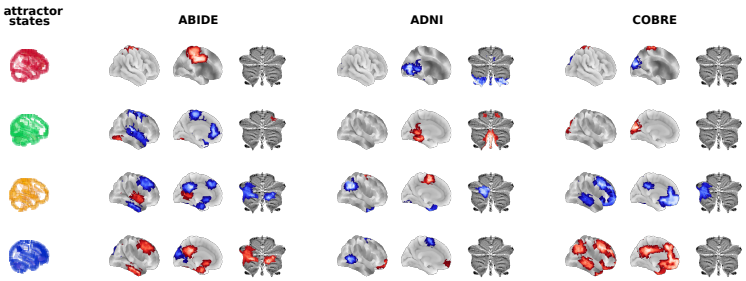


Figure 5: **Connectome-based Hopfield analysis as a sensitive tool for the study of clinical disorders.**
<\br>We quantified attractor state activations in three clincal datasets ((Table **??**) as the individual-level mean activation of all time-frames belonging to the same attractor state. CBH analysis of attractor state activations revealed significant differences in all three datasets. **A** Comparison of individuals with autism spectrum disorder (ASD) and typically developing controls (TD) is characterized by **todo**. **B** The most prominent Schizophrenia (SCZ)-related differences (as compared to healthy controls (HC) are related to the activity of the internalalization-related subsystem. **todo C** Alzheimer's disease (AD) is characterized by altered activation in **todo** the subsystems for passive inference and internal context (both of which together host long-term memory processes, see Figure 2.3F). All results are corrected for multiple comparisons across brain regions and attractor states (122*4 comparisons) with Bonferroni-correction. See Table **X** for detailed results.

model that elucidates how activity propagates through the intricate network topology of the brain, thereby constraining the system's dynamics and giving rise to distinct brain states along with characteristic dynamic responses to perturbations. Through a series of experiments, we have demonstrated that the proposed model can effectively reconstruct and predict large-scale brain activity across diverse conditions. These findings offer unprecedented possibilities for forecasting the impact of interventions, including pharmacological treatments or non-invasive brain stimulation, on brain function.

The construct validity of our model is rooted in the activity flow principle, first introduced by [Cole et al., 2016]. The activity flow principle states that functional connectivity between regions A and B can be conceptualized as the degree to which activity is transferred form A to B. This principle has been shown to successfully predict held out brain activations by a weighted sum of the activations of all the regions where the weights are set to the functional connectivity of those regions to the held-out region [Cole et al., 2016, Ito et al., 2017, Mill et al., 2022, Hearne et al., 2021, Chen et al., 2018].

ToDo: latent FC-based modelling: McCormick et al. [2022]

Our model was born from the intuition that the repeated, iterative application of the activity flow equation results in a system showing close analogies with a type of recurrent artificial neural network, know as Hopfield networks [Hopfield, 1982].

Hopfield networks have previously been shown to exhibit a series of characteristics that are also highly relevant for brain function, including the ability to store and recall memories [Hopfield, 1982], self-repair (**ref**), a staggering robustness to noisy or corrupted inputs [Hertz et al., 1991] and the tendency to produce multistable dynamics organized by the "gravitational pull" of a finite number of attractor states [Khona and Fiete, 2022].

The proposed link between activity flow and Hopfield networks has an important implication: network weights must be initialized with functional connectivity values, (specifically, partial correlations, as recommend by [Cole et al., 2016], instead of applying an explicit training procedure (common in the "neuroconnectomist" approach [Doerig et al., 2023]) or using the structural connectome (a standard practice of conventional computational neuroscience [Cabral et al., 2017].

Using functional conncetome-based Hopfield (CBH) model provides a simple yet powerful framework for the mechanistic understanding of brian dynamics. Its simplicity comes with important advantages.

Figure 6: **this is the title**

First, increasing model complexity results in an exponential explosion of the parameter space. Although complex, fine-grained computation models hold promise a full-blown understanding, they very easily overfit real data (**ref**). The basic CBH approach has only two hyperparameters (temperature and noise) and produce fairly consistent behavior on a wide range of parameter values. To demonstrate the power of simplicity, in the present work, we deliberately minimized fine-tuning of any free parameters. We fixed the temperature parameter at a value that robustly provides 4 attractor states and used a single noise level for all experiments (selected with a coarse optimization procedure to approximately mimic the distribution of real data).

Second, increasing complexity means increasing burden in terms of interpretability. The CBH model establishes a simple and direct link between two most popular measures of brain function: functional connectivity and brain activity. This link is not only conceptual, but also mathematical, and allows us to investigate and forecast changes of the system's dyanmics in response to perturbations of both activity and connectivity.

In this initial investigation, we further reduced complexity by restricting the analysis to a simplified 2-dimensional embedding of the state-space generated by the CBH approach, which we refer to as the Hopfield projection. This projection is a powerful tool for the visualization of the CBH model's dynamics, and allows for a direct comparison with the dynamics of the original brain activity.

However, the Hopfield projection only conveys a small proportion of the richness of the full state-space dynamics reconstructed by the CBH model. Investigating higher-dimensional dynamics, fine-tuned hyperparameters, the effect of different initializations and perturbations is an important direction for future work, with the potential to further improve the model's accuracy and usefulness.

Given these intentional simplifications, it is remarkable, if not surprising, how accurately the CBH model is able to reconstruct and predict brain dynamics under a wide range of conditions. Next to accurately reconstructing the distribution of, and the time spent in, different brain states during resting state, its superiority in explaining, and generalizing to, resting state brain activation patterns over principal components derived from the same data is particularly striking. The question arises, how can a relatively simple model, which is informed about empirical brain dynamics only through the functional connectome, be so powerful? A possible answer is that, while empirical data (and its principal components) are corrupted by noise and low sampling rate, the highly noise tolerant nature and the self-repair properties of the CBH architecture allow it to capture and reconstruct the basic principles of the underlying dynamics.

The noise-tolerance of the proposed architecture also explains the high replicability of CBH attractors across different datasets (study 2 and 3). The observed level of replicability allowed us to re-use the CBH model constructed with the connectome of study 1 for all subsequent studies, without any further fine-tuning or study-specific parameter optimization.

The connectome obtained form study 1 was also used to evaluate the model's ability to capture and forecast task-induced brain dynamics in study 4 and 5. In these analyses, the CBH model was not only able to capture participant-level activity changes induced by pain and self-regulation (showing significant differences on the Hopfield projection and in terms of state energy), but also accurately predicted the non-linear changes in activity flow induced by characteristic activity changes.

Brain dynamics can not only be perturbed by task or other types of experimental or naturalistic interventions, but also by pathological alterations. In our analysis of clinical samples study 6-8 we found that mean attractor activations show characteristic alteration in autism spectrum disorder (ASD), Schizophrenia (SCH) and Alzheimer's disease (AD). These changes were also detectable on the Hopfield projection, and were accompanied by significant changes in the state energies. The Hopfiled projection also allowed us to visualize the effect of different types of perturbations on the brain's attractor landscape, providing a novel perspective on the pathophysiology of these disorders.

> ToDo: more details on clinical outlook
>
> ToDo: discuss: what are attractor states at all? Platonic idealizations of brain states, that are continously approximated by the brain?
>
> Todo: for spontaneous and task-based brain dynamics. In the proposed framework, taks-based activity is not a mere response to external stimuli in certain brain locations but a perturbation of the brain's charcteristic dynamic trajectories. In this framework, conventional task-based fMRI analyses capture mean differences of the whole brain dyanmics, resulting in the widely reported focal "activation maps" thought to be specific to various taks and stimuli. In the CBH framework, the brain's characteristic trajectories are constrained by the underlying functional connectivity and only perturbed by external inpout, rather than predestinated.
>
> ToDo: discuss: the CBH model is not a model of brain function, but a model of brain dynamics. It does not strice to explain various brain regions ability to perform certain computations, but the brain's characteristic trajectories, which are perturbed by tasks and other types of interventions.

Together, these results open up a series of exciting opportunities for the mechanistic understanding of brain function. By its generative nature, the CBH model could foster analyses that aim at disentangling causal relationships, which are extremely difficult to infer in case of systems as complex as the brain. It could, for instance, aid the differentiation of primary causes and secondary effects of particular activity or connectivity changes in various clinical conditions.

Moreover, the CBH approach might provide testable predictions about the effects of interventions on brain functions, like pharmacological or non-invasive brain stimulation (e.g. transcranial magnetic or direct current stimulation, focused ultrasound) or neurofeedback. For instance, in the context of pain, the CBH model might

be used to predict the effect of various analgesic drugs (or other treatment strategies with known neural correlates) on the individual level (e.g. based on the individual functional connectome). Aiding the design of personalized medicine approaches is a particularly promising field of application for the proposed framework.

The generative nature of the proposed framework may be also used to generate synthetic brain activity data, which can be used to train and test machine learning algorithms, such as deep neural networks, for the prediction of brain activity from functional connectivity. This approach may be particularly useful in the context of clinical applications, where the amount of available data is often limited.

4 Conclusion

To conclude, here we have proposed a novel computational framework that accurately captures and predicts brain dynamics under a wide range of conditions. The framework models large-scale activity flow in the brain with a recurrent artificial neural network architecture that, instead of being trained to solve specific tasks or mimic certain dynamics, is simply initialized with the empirical functional connectome. The framework identifies biologically meaningful attractor states and provides a model for how these restrict brain dynamics. The proposed framework, referred to as the connectome-based Hopfield (CBH) model, can accurately reconstruct and predict brain dynamics under a wide range of conditions, including resting state, task-induced activity changes, and pathological alterations. CBH analyses provide a simple, robust, and highly interpretable computational alternative to the conventional descriptive approach to investigating brain function and establish a link between connectivity and activity. The generative nature of the proposed model opens up a series of exciting opportunities for future research, including novel ways of assessing causality and mechanistic understanding, and the possibility to predict the effects of various interventions, thereby paving the way for novel personalized medical approaches.

5 Methods

5.1 Hopfield network

We employ an empirical, connectome-based Hopfield network as a means to model brain activation and dynamics, aiming to bridge the gap between classical computational modeling and neuroconnectionism. The architecture of the Hopfield network [Hopfield, 1982] consists of a single layer of fully connected nodes, the undirected weights connecting all the nodes serve as the system's memory. Instead of training the weights on known patterns, we initialize the weights with a group-level connectivity matrix; each node in the network representing a brain region. Through its associative memory capabilities, the network can retrieve patterns embedded within its memory, when presented with an input similar to a target pattern. During the retrieval process, the network will iterate on the output pattern until the system converges to a stable state, a so-called attractor state. The mathematical energy of all possible states of a trained Hopfield network spans an N dimensional, multi-stable state landscape. This landscape constrains the state configurations, which can be recalled by the Hopfield network across all dimensions. During the memory retrieval process, the network will try to find a state which minimizes the energy function E

$$E = -\frac{1}{2}\mathbf{a}^T\mathbf{W}\mathbf{a} + \mathbf{a}^T\mathbf{b} \tag{2}$$

where W is the weight matrix, a the activation pattern and b the bias, which is set to $b = 0$ for all experiments. The network navigates the state landscape by synchronously updating all regions in the current state, according to the (1). The temperature parameter β scales the estimated activation (activity flow) [Cole et al., 2016]and therefore constrains, how many attractor states can be found, given the convergence criterium of minimal (2). If the value of β is too high however, the network might retrieve spurious states, which meet the convergence criterium, but are composite states which merge multiple states and are not "true" attractor states.

5.2 Hopfield projection

The attractor landscape can be mapped out by stimulating our CBH network with a random input, and adding noise after each iteration of the network relaxation. This prevents the network to reach an energy minimum and the network produces possible state configurations within the landscape, while avoiding the energy minimal attractor basins. We do a principal component analysis (PCA) on the state samples, and

the resulting first two principal components (PC) lay out the coordinate system for our Hopfield projection. Using a Multinomial Logistic Regression, we predict to which attractor state each of the state samples converges to, using the first two PCs as features. We visualize the attractor states position in the projection as well as the decision boundaries between the attractor states, based on the regression model. We set $\beta = 0.04$, which results in 4 attractor states given the connectome of study 1, and do a coarse optimization for the noise level ($\sigma = 0.37$) of the stochastic walk, to reproduce the bimodal distribution of the real fMRI timeseries in the state space (see Figure 2.2).

For all experiments conducted, the connectome of study 1 is used as a base for the CBH and the projection, with the hyperparameters $\beta = 0.04$ and $\sigma = 0.37$, resulting in 4 distinct attractor states.

todo:

- explained variance of energy through state sample
- attractor classification accuracy

5.3 Reconstruction

We use several experiments to investigate the validity of the CBH model and its ability to reconstruct resting state brain dynamics. First, we investigate the reproducibility of the attractor states across studies. The attractor states are highly reproducible across various datasets and scanners, as the attractor states from studies 1,2 and 3 show a 0.93 mean correlation across the first two attractor states. We then compared the explained variance from the first two PCs of our simulated state sample data to the first two PCs of raw fMRI timeseries data (see **ref** for preprocessing), using a linear regression model. For in sample data, the first 2 components of the real data were able to explain 37.0% variance, whereas the simulated counterpart could account for 39.9% variance. For out of sample time series data from study 2, the first two principal components of the hopfield projection of study 1 were able to explain 36.4% and 39.6% variance for real and simulated data respectively. We also investigated the fractional time occupied by each attractor state in real timeseries vs simulated data. For this analysis each timeframe was used as an input to the CBH to generate its corresponding attractor state, a one way $\tilde{\chi}^2$ test was performed on the given frequencies against expected uniform frequencies.

5.4 Task-based activity

The Hopfield projection provides a unique framework in which we can analyze and visualize how activations dynamically change between two conditions. We highlight these properties on the dataset of study 4, which investigated the self-regulation of pain. We preprocess the timeseries data as discussed in **ref**, and divide the samples into task and rest, taking into account the $6s$ delay to adjust for the hemodynamic response function. We group the activations into "rest" and "pain", and transform all single TR activations (density plot) as well was the participant-level mean activations to the Hopfield projection plane. The difference between rest and pain is visualized with a radial plot, showing the participant-level trajectory on the projection plane from rest to pain, denoted with circles, as well as the group level trajectory (arrow). Additionally, we test the significance of the spatial difference of the participant-level mean activation in the projection plane with the L2 norm, as well as the energy difference between the two conditions, both with a permutation test $n_{perms} = 1000$, randomly swapping the conditions. To further highlight the difference between the task and rest conditions, we generate streamplots that visualize the dynamic trajectory of group-level activations. We calculate the direction in the projection plane between each successive TR and calculate a bidimensional binned mean for the x,y position across the direction. We repeat the same for the second condition and visualize the difference in direction between the two conditions, visualized as streamplots. For the simulated data, we introduce a weak signal of SNR=0.01 according to a meta-analytic pain activation map [Zunhammer et al., 2021] to the stochastic walk, aiming to simulate the shift from rest to pain also in the simulated data. We compare the simulated difference to the actual difference through a permutation test ($n_{perm} = 1000$) with the spearman rank-ordered correlation coefficient as the test statistic. The analysis documented in this section is repeated, comparing the pain upregulation and pain downregulation data provided with study 4.

5.5 Clinical data

To assess clinical relevance, we introduce a pipeline that investigates the group differences in raw timeseries activation, during each of the attractor states. We assign each TR a label according to its attractor state, by relaxing the CBH for each TR and then calculate the average participant-level activation for each attractor

study	modality	analysis	n	age (mean±sd)	%female	references
study 1	resting state	discovery	41	26.1±3.9	37%	Spisak et al. [2020]
study 2	resting state	replication	48	24.9±3.5	54%	Spisak et al. [2020]
study 3	resting state	replication	29	24.8±3.1	53%	Spisak et al. [2020]
study 4	task-based	pain self-regulation	33	27.9 ± 9.0	66%	Woo et al. [2015b]
study 5 (Neurosynth)	task-based	coordinate-based meta-analyses	14371 studies in total	•	•	D. [2011]
study 6 (ABIDE, NYU sample)	resting state	Autism Spectrum Disorder	ASD: 98, NC: 74	15.3±6.6	20.9%	[Di Martino et al., 2014]
study 7 (ADNI)	resting state	Alzheimer's Disease vs. Mild Cognitive Impairment	AD: 34, MCI 99:	72.5±7.5	50.4%	[Petersen et al., 2010]
study 8 (COBRE)	resting state	Schizophrenia	SCH: 60, HC: 72	37.0±12.6	29.4 %	[Aine et al., 2017]

state. We implement a permutation test with $n_{perm} = 50000$ to investigate the difference in the average activation during the attractor states between the groups, randomly assigning the group label (preserving the original group stratification). We adjust the significance threshold with a bonferroni correction, accounting for tests across 4 states and 122 regions, resulting in $\alpha = 0.0102$.

5.6 Data preprocessing

The data from studies 1-4 and 6-8 is parcellated with the BASC multiscale atlas into 122 regions [Bellec et al., 2010]. The timeseries is scrubbed with a threshold of 50% and a frame-wise displacement threshold of 0.15, to correct for motion artifacts present in the data. The connectivity matrix used to set the weights of the CBH, is calculated with a partial correlation of the parcellated, scrubbed timeseries, with the diagonal elements set to zero. When initializing the CBH with the connectome, the weights are standard scaled to mean=0 and std=1; the weights w_{ij} are symmetric.

5.7 Data

Todo

Todo

References

Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353–364, 2017.

Daniel Gutierrez-Barragan, M Albert Basson, Stefano Panzeri, and Alessandro Gozzi. Infraslow state fluctuations govern spontaneous fmri network dynamics. *Current Biology*, 29(14):2295–2306, 2019.

Xiao Liu and Jeff H Duyn. Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, 110(11):4392–4397, 2013.

Andrew Zalesky, Alex Fornito, Luca Cocchi, Leonardo L Gollo, and Michael Breakspear. Time-resolved resting-state brain networks. *Proceedings of the National Academy of Sciences*, 111(28):10341–10346, 2014.

Garth John Thompson, Wen-Ju Pan, Matthew Evan Magnuson, Dieter Jaeger, and Shella Dawn Keilholz. Quasi-periodic patterns (qpp): large-scale dynamics in resting state fmri that correlate with local infraslow electrical activity. *Neuroimage*, 84:1018–1031, 2014.

Abigail S Greene, Corey Horien, Daniel Barson, Dustin Scheinost, and R Todd Constable. Why is everyone talking about brain state? *Trends in Neurosciences*, 2023.

Diego Vidaurre, Stephen M Smith, and Mark W Woolrich. Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832, 2017.

Jonas Richiardi, Hamdi Eryilmaz, Sophie Schwartz, Patrik Vuilleumier, and Dimitri Van De Ville. Decoding brain states from fmri connectivity graphs. *Neuroimage*, 56(2):616–626, 2011.

Stephen M Smith, Karla L Miller, Steen Moeller, Junqian Xu, Edward J Auerbach, Mark W Woolrich, Christian F Beckmann, Mark Jenkinson, Jesper Andersson, Matthew F Glasser, et al. Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, 109(8): 3131–3136, 2012.

Richard H Chen, Takuya Ito, Kaustubh R Kulkarni, and Michael W Cole. The human brain traverses a common activation-pattern state space across task and rest. *Brain Connectivity*, 8(7):429–443, 2018.

R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, et al. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378, 2013.

Pablo Barttfeld, Lynn Uhrig, Jacobo D Sitt, Mariano Sigman, Béchir Jarraya, and Stanislas Dehaene. Signature of consciousness in the dynamics of resting-state brain activity. *Proceedings of the National Academy of Sciences*, 112(3):887–892, 2015.

Johan N van der Meer, Michael Breakspear, Luke J Chang, Saurabh Sonkusare, and Luca Cocchi. Movie viewing elicits rich and reliable brain state dynamics. *Nature communications*, 11(1):5004, 2020.

Jingyuan E Chen, Catie Chang, Michael D Greicius, and Gary H Glover. Introducing co-activation pattern metrics to quantify spontaneous brain network dynamics. *Neuroimage*, 111:476–488, 2015.

Xiao Liu, Catie Chang, and Jeff H Duyn. Decomposition of spontaneous brain activity into distinct fmri co-activation patterns. *Frontiers in systems neuroscience*, 7:62295, 2013.

Michael Breakspear. Dynamic models of large-scale brain activity. *Nature neuroscience*, 20(3):340–352, 2017.

Nikolaus Kriegeskorte and Pamela K Douglas. Cognitive computational neuroscience. *Nature neuroscience*, 21(9):1148–1160, 2018.

Andreas Heinz, Graham K Murray, Florian Schlagenhauf, Philipp Sterzer, Anthony A Grace, and James A Waltz. Towards a unifying cognitive, neurophysiological, and computational neuroscience account of schizophrenia. *Schizophrenia bulletin*, 45(5):1092–1100, 2019.

Adrien Doerig, Rowan P Sommers, Katja Seeliger, Blake Richards, Jenann Ismael, Grace W Lindsay, Konrad P Kording, Talia Konkle, Marcel AJ Van Gerven, Nikolaus Kriegeskorte, et al. The neuroconnectionist research programme. *Nature Reviews Neuroscience*, pages 1–20, 2023.

Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. A deep learning framework for neuroscience. *Nature neuroscience*, 22(11):1761–1770, 2019.

Michael W Cole, Takuya Ito, Danielle S Bassett, and Douglas H Schultz. Activity flow over resting-state networks shapes cognitive task activations. *Nature neuroscience*, 19(12):1718–1726, 2016.

Takuya Ito, Kaustubh R Kulkarni, Douglas H Schultz, Ravi D Mill, Richard H Chen, Levi I Solomyak, and Michael W Cole. Cognitive task information is transferred between brain regions via resting-state network topology. *Nature communications*, 8(1):1027, 2017.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

Dmitry Krotov. A new frontier for hopfield networks. *Nature Reviews Physics*, pages 1–2, 2023.

Pascal Koiran. Dynamics of discrete time, continuous state hopfield networks. *Neural Computation*, 6(3): 459–468, 1994.

Peter A Robinson, CJ Rennie, Donald L Rowe, SC O'Connor, Gordon, and E. Multiscale brain modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1043–1050, 2005.

Jean Daunizeau, Klaas Enno Stephan, and Karl J Friston. Stochastic dynamic causal modelling of fmri data: should we care about neural noise? *Neuroimage*, 62(1):464–481, 2012.

Yulia Golland, Polina Golland, Shlomo Bentin, and Rafael Malach. Data-driven clustering reveals a fundamental subdivision of the human cortex into two global systems. *Neuropsychologia*, 46(2):540–553, 2008.

Claudia Cioli, Hervé Abdi, Derek Beaton, Yves Burnod, and Salma Mesmoudi. Differences in human cortical gene expression match the temporal properties of large-scale functional networks. *PloS one*, 9(12):e115913, 2014.

Joaqun M Fuster. Upper processing stages of the perception–action cycle. *Trends in cognitive sciences*, 8 (4):143–145, 2004.

Choong-Wan Woo, Mathieu Roy, Jason T Buhle, and Tor D Wager. Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS biology*, 13(1):e1002036, 2015a.

Matthias Zunhammer, Tamás Spisák, Tor D Wager, and Ulrike Bingel. Meta-analysis of neural systems underlying placebo analgesia from individual participant fmri data. *Nature communications*, 12(1):1391, 2021.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

CJ Aine, Henry Jeremy Bockholt, Juan R Bustillo, José M Cañive, Arvind Caprihan, Charles Gasparovic, Faith M Hanlon, Jon M Houck, Rex E Jung, John Lauriello, et al. Multimodal neuroimaging in schizophrenia: description and dissemination. *Neuroinformatics*, 15:343–364, 2017.

Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, Clifford R Jack, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's disease neuroimaging initiative (adni): clinical characterization. *Neurology*, 74(3):201–209, 2010.

Ravi D Mill, Julia L Hamilton, Emily C Winfield, Nicole Lalta, Richard H Chen, and Michael W Cole. Network modeling of dynamic brain interactions predicts emergence of neural information that supports human cognitive behavior. *PLoS Biology*, 20(8):e3001686, 2022.

Luke J Hearne, Ravi D Mill, Brian P Keane, Grega Repovš, Alan Anticevic, and Michael W Cole. Activity flow underlying abnormalities in brain activations and cognition in schizophrenia. *Science advances*, 7(29): eabf2513, 2021.

Ethan M. McCormick, Katelyn L. Arnemann, Takuya Ito, Stephen José Hanson, and Michael W. Cole. Latent functional connectivity underlying multiple brain states. *Network Neuroscience*, 6(2):570–590, 2022. doi:10.1162/netn_a_00234. URL https://doi.org/10.1162%2Fnetn_a_00234.

John Hertz, Andres Krogh, and Richard G Palmer. Introduction to the theory of neural computation, chapter 7. *Lecture Notes*, 1, 1991.

Mikail Khona and Ila R Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23(12):744–766, 2022.

Joana Cabral, Morten L Kringelbach, and Gustavo Deco. Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *NeuroImage*, 160:84–96, 2017.

Pierre Bellec, Pedro Rosa-Neto, Oliver C Lyttelton, Habib Benali, and Alan C Evans. Multi-level bootstrap analysis of stable clusters in resting-state fmri. *Neuroimage*, 51(3):1126–1139, 2010.

Tamas Spisak, Balint Kincses, Frederik Schlitt, Matthias Zunhammer, Tobias Schmidt-Wilcke, Zsigmond T. Kincses, and Ulrike Bingel. Pain-free resting-state functional brain connectivity predicts individual pain sensitivity. *Nature Communications*, 11(1), jan 2020. doi:10.1038/s41467-019-13785-z. URL https://doi.org/10.1038%2Fs41467-019-13785-z.

Choong-Wan Woo, Mathieu Roy, Jason T. Buhle, and Tor D. Wager. Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biology*, 13(1):e1002036, jan 2015b. doi:10.1371/journal.pbio.1002036. URL https://doi.org/10.1371%2Fjournal.pbio.1002036.

Wager Tor D. NeuroSynth: a new platform for large-scale automated synthesis of human functional neuroimaging data. *Frontiers in Neuroinformatics*, 5, 2011. doi:10.3389/conf.fninf.2011.08.00058. URL https://doi.org/10.3389%2Fconf.fninf.2011.08.00058.