Attractor and integrator networks in the brain

Citation: Khona, Mikail and Fiete, Ila R. 2022. "Attractor and integrator networks in the brain." Nature Reviews Neuroscience, 23 (12).

As Published: 10.1038/s41583-022-00642-0

Publisher: Springer Science and Business Media LLC

Persistent URL: https://hdl.handle.net/1721.1/148794

Version: Original manuscript: author's manuscript prior to formal peer review

Attractor and integrator networks in the brain

Mikail Khona[1,2] and Ila R. Fiete *[1]

[1]Department of Brain and Cognitive Sciences & McGovern Institute, MIT
[2]Department of Physics, MIT

Abstract

In this review, we describe the singular success of attractor neural network models in describing how the brain maintains persistent activity states for working memory, error-corrects, and integrates noisy cues. We consider the mechanisms by which simple and forgetful units can organize to collectively generate dynamics on the long time-scales required for such computations. We discuss the myriad potential uses of attractor dynamics for computation in the brain, and showcase notable examples of brain systems in which inherently low-dimensional continuous attractor dynamics have been concretely and rigorously identified. Thus, it is now possible to conclusively state that the brain constructs and uses such systems for computation. Finally, we look ahead by highlighting recent theoretical advances in understanding how the fundamental tradeoffs between robustness and capacity and between structure and flexibility can be overcome by reusing and recombining the same set of modular attractors for multiple functions, so they together produce representations that are structurally constrained and robust but exhibit high capacity and are flexible.

Introduction

One of Biology's grand challenges is to explain how order and complex function spring from inanimate physical systems composed of much simpler parts. The brain creates order in its representations of the world and performs complex functions through the collective interactions of simpler elements. In this review, we will describe and evaluate the hypothesis that attractor dynamics in widespread regions of the central nervous system play a key role in constructing some of these representations, generating long time-scales to support integration and memory functions, and endowing all these functions with robustness. We will review the specific predictions of attractor-based models and the now-extensive body of work testing these predictions. Thus, we will illustrate that the theory and validation of computation with attractor dynamics in the brain is one of the biggest success stories in systems neuroscience.

Some of the first formal circuit-level models of brain function focused on the problem of associative memory and how neural circuits might generate spatially distributed, stable patterns of activity that could function as such a memory [1, 2, 3, 4]. Hopfield networks, with multiple stable states learned from distributed input patterns, were proposed over four decades ago

*Corresponding author

[5, 3, 6]. Network models possessing a continuous set of stable states, that could be used to represent continuous variables, were also first proposed around the same period [7]. Subsequently, many canonical brain circuits for motor control, sensory amplification and memory, motion integration, evidence integration, decision making, and spatial navigation have been modeled using the same general principle – that a set of states stabilized through collective positive feedback can be used for robust representation, memory, and to perform computations that involve memory [8, 9, 10, 11, 12, 13, 14, 15, 16, 17].

Because these are circuit-level models, but were typically inspired by experimental characterization of neurons recorded singly or a few at a time, the patterns of connectivity and cell activity correlations in the models automatically became novel and relatively specific predictions about the population dynamics and architecture of such circuits. As we will discuss below, the combination of these prediction-rich (and often conceptually simple) models, modern experimental breakthroughs in the acquisition of cellular-resolution population activity data, and novel and rigorous analyses of such data based on the model predictions has led to the accumulation of a preponderance of evidence that the brain constructs and exploits attractor networks for performing several essential computations.

We will begin by defining attractors, then describe proposed mechanisms for the construction of attractor network models in neuroscience. We will provide an overview of why attractor networks can be important for computation in the brain and give criteria for determining whether a system has non-trivial attractor dynamics. After this groundwork we will discuss several examples of brain circuits with non-trivial attractor dynamics. Finally, we will end with new directions in our understanding of how these simple circuits could contribute to flexible computation through reuse in multiple contexts.

What are attractors?

To define an attractor, we first define a dynamical system and its states. A *dynamical system* is a set of variables together with all the rules that determine their time-evolution. The instantaneous value of these variables is called the *state* of the system at that moment. The state is a point (vector) in the *state space* of the dynamical system. An *attractor* is the minimal set of states in a state space, to which all nearby states eventually flow [18]. One simple example of an attractor is a stable fixed point: all neighboring states flow to it. Porting these crisp mathematical definitions to the brain involves challenges and simplifications, which revolve around identifying a sufficiently self-contained system and the variables necessary to determine its dynamics.

Defining the state of a neural system: Inherent in the definition of a dynamical system is the assumption that there are no external dynamical inputs to the system (equivalently, the system definition includes all such variables). The first simplification in characterizing the dynamics of a neural circuit is to assume that at least on the time-scale of interest, the system evolves *autonomously*. Given that subcircuits in the brain are interconnected with others, and that the brain itself interacts with the world, it is impossible to completely isolate these circuits into autonomous systems. However, we may define a notion of *effectively autonomous* dynamics over time-scales where inputs are not temporally varying and are untuned in the sense that they do not provide differential drive to subsets of the putative set of attractor states. The second simplification is in defining the states of the system. The time-evolution of a circuit in the brain may depend on the detailed pattern of all the spikes in all neurons, the levels of associated ions, neurotransmitters and modulators, the states of the ion channels. The weights and connections between neurons may be considered as parameters rather than variables on short timescales, but as variables if considering a longer time-scale. One widely used simplification in describing a neural circuit on the timescale of seconds is to use just the spiking outputs of the neurons in the circuit as the states, often further simplified as time-varying spike rates. If such a description is sufficient to predict the evolution of the system at the relevant time-scales, it can be viewed as a reasonable dynamical system model of the circuit. Even though spike or spike rate descriptions ignore sub-cellular and molecular variables to make the grossly simplifying assumption that the relevant circuit dynamics are governed by spikes, the state space of a vertebrate microcircuit described in this way is nevertheless very high-dimensional, comprising the number of neurons in the circuit, or $10^2 - 10^7$ cells. As we will see below, such simplified models can nevertheless yield rich and accurate predictions about neural circuits.

Attractors exist in various flavors: an attractor may consist of a single state or a set of states that trace out a complex shape, such as a curved manifold [1], Fig. 1 (rightmost column). States on an attractor may be stationary, or might flow along the attractor to trace out trajectories that are periodic (limit cycles, Fig. 1f, rightmost column) or chaotic (dynamics that are inherently unpredictable due to high sensitivity to small changes in the state [21]). Various combinations of such attractors, of different dimension, geometry, and topology, may coexist in different regions of the state space of a single dynamical system. Typically, the set of attractors in a dynamical system comprises a small subset of the state space, and attractor manifolds are usually much lower-dimensional than the state space. In cases where a system has multiple attractor states, the initial condition determines the attractor state to which the system flows.

Defining attractors in the presence of noise: Any real physical system unavoidably behaves non-deterministically from the perspective of a model of the system. This is because one cannot observe and describe all variables, and all uncharacterized variables together with true stochastic sources of variation (e.g., synaptic signalling noise from stochastic vesicle release [22], ionic number fluctuations in processes like spike initiation [23], calcium signaling, fluctuations in small copy numbers of proteins [24]) serve as effective sources of noise in the model. Noise can buffet states so they do not strictly localize to the attractor described in a noise-free version of the model, and can drive the system to escape from an attractor over time. However, the general idea of attracting states remains in the following sense: If the system is initialized near such a state, it tends to flow toward it and subsequently remains localized around it, for extended periods. In sum, since attractor states are where systems tend to localize (when not externally driven), they should be observable in the autonomous dynamics of real systems. This basic property is the basis for the most fundamental and robust tests of attractor dynamics in neural systems, as we will discuss below. In a nutshell, the critical signatures of attractors in real systems (discussed in more detail in later sections of this review) can be summarized as: localization of the states of a system to a lower-dimensional subset, flow of the states towards the subset after perturbation, and long-time and (effectively) autonomous stability of states in that subset.

[1] **Attractor manifolds:** If the number of attractor states in a network is large and the points are close to one another, they can behave effectively as a continuous set. If this near-continuous set traces out a surface that is locally Euclidean, it is called a manifold. Nonlinear continuous attractor manifolds can be curved and topologically complex (e.g. rings, torii, etc., Fig. 1c-d, rightmost column [19, 20].

Mechanisms: The construction of neural attractors

The general principle underlying the formation of non-trivial attractor states in neural circuits is strong recurrent positive feedback [2]. Positive feedback fights activity decay to stabilize certain states, and has been conjectured by James, Hebb and others [25, 26, 27, 28] as the basis for the stabilization of memory traces and persistent activity in the brain. Which states become stabilized into attractors depends on how the network sculpts the positive feedback, which according to the synaptic hypothesis is determined by the synaptic weights.[29, 30, 31]

In general, characterizing the relationship between structure and function in a large collection of interacting elements is extremely difficult, as described by Anderson in "More is different" [32]. For instance, a large collection of simple polar 3-atom molecules of hydrogen and oxygen give rise to emergent phenomena like liquidness and wetness and freezing into a solid, which cannot be predicted through intuition or drawing box-and-arrow diagrams. On the flip side, there is also emergent simplicity, in that the transitions and properties of the emergent states can be described with very few key parameters and variables.

One way to characterize the relationship between synaptic weights and attractor dynamics is to ask what attractor states a given set of weights produces (the “forward” problem). With the given weights, one can simulate the circuit and explore the resulting dynamics to find attractors of the system. A more powerful method, the Lyapunov function approach, holds for symmetric weight matrices ($W_{ij} = W_{ji}$) and rate-based neural dynamics. For this class of models, a generalized energy function (the Lyapunov function), which is a function of the weights and neural activation function [6, 5, 2], analytically specifies the network’s dynamics. Stable (unstable) attractor states are the energy minima (maxima) of the derived landscape, and the network’s state flows downhill towards the attractors Fig. 2e in the way a ball rolls down a gravitational potential.

Another way to characterize the relationship between attractors and network structure is to consider the “inverse” problem: given a set of attractors, what network structure could generate it? Neuroscientists want to solve the inverse problem to make predictions about underlying mechanism (and since neural activations are more readily observed than synaptic weights, the inverse problem is more frequently encountered than the forward one), while evolution, the brain, and artificially intelligent systems need to solve the inverse problem to be able to perform computations that require a given type of attractor dynamics (which we will discuss below). Theoretical neuroscience has discovered some solutions to the inverse problem for different types of attractors, as we describe next.

Discrete attractors. A well-known prescription for creating a discrete set of stable attractors at user-defined points in state space is given by the Hopfield network model [5], Fig. 1a: an externally induced and distributed pattern of neural activity is encoded into the weights by a Hebbian-like learning rule that causes co-activated neurons to excite each other and inhibit all the rest. As a result, these patterns become stable attractor states. If a sufficiently small number of patterns are inscribed into the weights, they can be retrieved from partial or corrupted versions of the stored states, thus the network is a content-addressable memory. More generally, the attractors of simple rate-based networks without synaptic delays and with arbitrary symmetric weight matrices[3] consist entirely of fixed points. Some non-symmetric networks can also support point attractors [33], but this is not the generic case and can require additional mechanisms like

[2] Non-trivial attractor states refer to any state other than the null activity state. Positive feedback is not synonymous with excitatory feedback: like mutual excitation, *disinhibition* or inhibition of one’s inhibitor is also a form of positive feedback.

[3] A symmetric weight matrix W satisfies $W^T = W$: it is invariant to reflection of its entries about its diagonal.

homeostatic plasticity [34, 35].

The attractor states in Hopfield-like networks typically have highly mixed and overlapping neural memberships, even when they are well-separated in the state space, Fig. 1a (middle column). In a special case of Hopfield networks, neurons are partitioned into largely disjoint groups with self-excitation within groups and inhibition between groups. In these winner-take-all networks, Fig. 1b, the attractor states consist of non-overlapping active groups of cells, Fig. 1a,b (middle columns).

Continuous attractors. How can one construct networks with a continuum of stationary attractor states? Weights that are invariant to reflection about their diagonal lead to the formation of discrete attractors, as we have seen. If the weights instead exhibit a *continuous* symmetry – for instance, if the weights are invariant to continuous shifts in neural locations – then the set of formed attractors will be related by the same symmetry and could thus form a continuous set.

The general principle for the formation of stationary continuous attractors is pattern formation [36, 37, 38, 39, 40, 41, 42]. Simple and spatially local competitive interactions lead to the emergence of rich stable spatial activity patterns – neurons with excitatory coupling between them become co-active, and suppress the rest of their neighbors through inhibition – a linear (Turning) instability [36].

The following three elements provide a solution to the inverse problem for forming stationary continuous attractors: 1) Nonlinear neurons with saturating responses or inhibition-dominated recurrent interactions with a uniform excitatory drive [7, 16, 43, 44] to keep network activity bounded. 2) Sufficiently strong recurrent weights with competitive dynamics in the form of local excitation or disinhibition with broader inhibition to drive spontaneous pattern formation via the Turing instability [36, 37, 38, 39, 40, 41, 42, 16]; these patterns become the attractor states. 3) Some continuous symmetry in the weights (a continuous weight symmetry is one where, as some variable is varied continuously, the weights remain invariant), such as translational or rotational invariance, Fig. 1c-d, to ensure a continuum of attractor states. These conditions are generally sufficient, but not strictly necessary, for the construction of continuous attractors (see Box on "Correspondences between attractor dynamics and anatomical layout"). If the continuous symmetry of weights is sufficiently corrupted, the continuous attractor will fragment into a discretized set of attractor states. Thus, the existence of stationary continuous attractors is fragile in the sense that it depends on the maintenance of continuous symmetries.

A special set of networks, which do not involve pattern formation to generate continuous attractor dynamics, are those with linear, planar, or hyperplanar attractors generated by neurons with linear or near-linear response functions. In circuits of linear neurons, the network feedback is a linear function of activity ($\mathbf{W}\mathbf{r}$ where $\mathbf{W}$ is the weight matrix and $\mathbf{r}$ are the neural activities), as is the activity decay (given by $-\mathbf{r}$). Such networks can stablize non-zero activity states simply by tuning the strength of the feedback so that positive feedback cancels decay. The feedback matrix W can direct feedback into the different dimensions of the state space; if feedback is directed largely along one dimension, the network can support a line attractor, Fig. 1e. If it is directed equally along two or more dimensions, it can support a plane or hyperplane attractor. To create long-lived attractors requires feedback to precisely cancel decay, thus the strength of network feedback must be finely tuned [9, 45], in contrast with pattern-forming continuous attractor systems.

Non-stationary continuous attractors: Large non-symmetric (and nonlinear) networks with strong connectivity generically exhibit limit cycle attractors or chaotic dynamics [46, 47]. Just as point attractors emerge generically in large networks with strong symmetric weights and bounded state spaces, chaotic attractors emerge generically in large recurrent networks with strong asymmetric weights. Adequate asymmetries are easily achieved if excitatory and inhibtory synapses emerge from distinct sets of neurons neurons [47], as biologically necessitated by Dale's law.

Despite the complexity of chaotic dynamics, these attractors are also highly structured in that they are typically much lower-dimensional than the number of neurons in the network [48]. Non-symmetric networks dominated by inhibition exhibit a single attractor at zero activity, though the flow towards the attractor in responose to perturbations can involve large transients in neural activation that temporarily move the state further away from the attractor [49, 50].

The potential utility of attractors for computation in the brain

Networks with low-dimensional attractor dynamics exhibit myriad properties that can be vital for computation in the brain [4]. These include robust representation, memory, sequence generation, integration, and robust classification and decision making, ideas that have been extensively explored in the literature. In a later section, we will describe how, though attractor dynamics may be rigid and invariant as needed for the roles listed above, recent theoretical and experimental findings are beginning to reveal how these rigid constructions may also be exploited to perform flexible computation through reuse and re-combination across tasks.

Representation and memory

A *representation* of a set of inputs means the assignment of inputs to representational states (a representation need not be injective), with the ability to reproducibly retrieve those states ('labels') when cued. Attractor networks provide a stable internal set of states that can be used for reproducible representation of discrete or analog variables, by mapping states in the world to the attractor states. One way to achieve this mapping is through a feedforward learning process that associates each external state with an internal attractor state, Fig. 2a.

An attractor network can exhibit two kinds of memory: The first is in the structure of the weights, which specify the set of all attractors. If these weights are specified through an input-driven learning process, this is a form of long-term memory about the inputs. The second is the ability to maintain *persistent activity* in a stationary attractor state: if a system with multiple stationary attractor states is initialized in one of them, it will tend to remain at or near the same state for some time. In other words, the activation levels of the neurons contributing to that state persist while the system remains in the state. This persistent activity response is thus a form of short-term memory of the input that initialized the circuit. If these persistent memory states can be activated without an explicit address, using just the content (or partial content) of the memory, they are content-addressable.

The short-term memory function of attractors depends on the prior formation of stable states through long-term plasticity: For instance, in Hopfield-like networks, states cannot persist if they were not first trained to be attractor states. Even models of short-term memory that are based on synaptic facilitation rather than persistent activity rely implicitly on prior long-term plasticity to construct recurrently stabilized neural ensembles that can be reinstated by random inputs [51]. In

[4] A system could theoretically be perfectly tuned such that every point in state space is a neutrally stable attractor, and thus the system has maximally high-dimensional attractor dynamics. However, because the robustness of attractor networks is related to the low-dimensionality of the attractor states as quantified in the subsections below, the system would lose most of its interesting computational properties: error correction/noise tolerance, nearest-neighbor computation, pattern completion and content-addressable memory. It could perform integration but with no robustness to noise.

other words, these models cannot explain short-term memory for entirely novel inputs; however, combinations of attractors could enable more flexible short-term memory, as we discuss later.

Denoising for fidelity of representation and memory

If the representational states are attractors, then the representations are robust in the sense that they perform denoising: If the input cues or initial conditions reflect noisy or corrupted versions of an attractor state, the dynamics drive the state onto a point on the representational attractor, Fig. 2b (inset). When the attractors form a continuous manifold of dimension $K \ll N$, where N is the number of neurons in the circuit, all noise in $N-K$ dimensions is erased. A noise ball of unit radius in N dimensions (corresponding to random independent noise per neuron) has a projection of size only $\sim \sqrt{K/N} \ll 1$ along K dimensions. If K is low-dimensional, as is often the case, and N ranges from $10^2 - 10^7$ as estimated before for common microcircuits, this constitutes a massive reduction in the sensitivity of the state to internal or input noise, Fig. 2b. Thus, most noise is rendered impotent.

Denoising due to attractor dynamics is especially important for memory maintenance, as otherwise noise-induced deviations would accumulate and grow over time. Discrete attractors continually erase all noise by mapping perturbed states back to the point attractor , resulting in zero drift. With continuous attractors as memory states, all noise orthogonal to the manifold is corrected, thus there is a net reduction of the effects of noise by the factor $\sim \sqrt{K/N} \ll 1$ [52, 53]. However, all states on the attractor manifold are neutrally stable so movements along the attractor are allowed. Thus, components of noise along the K attractor dimensions are not internally corrected and cause an accumulating drift away from the inital state, with variance proportional to KT/N, where T is the elapsed time [52, 16, 53, 54]. Thus, even continuous memory states can be well-stablized in sufficiently large attractor networks.

Although content-addressable long-term memory and error reduction can be instantiated through few-step feedforward computations [55, 56, 57] in place of attractor dynamics, recurrent attractor dynamics are indispensable for the generation of persistent activity states (and thus for short-term memory through persistent activity [58, 59]) and integration, as we discuss below.

Robust classification

When the attractors form a set with a discrete component (e.g. a set of point attractors or a set of continuous attractors), inputs that are not initially on one of the attractors will flow to one of the attractors and thus we may view the identity of the specific attractor to which the input flows as a classification of the input as a class represented by that attractor. The process can perform a pattern-completing nearest-neighbor computation, if the dynamics correctly drives non-attractor states to the nearest attractor states. In other words, the dynamical basins of attraction must align with the Voronoi regions of the attractor states, which is approximately the case for attractor networks operating well below capacity. This property deteriorates when attractor networks are pushed toward their capacity [60].

Integration

Single neurons integrate their inputs, but usually can only do this over the time-scales associated with their membrane capacitances, typically 10-100 milliseconds. Continuous attractor dynamics can allow neural circuits to integrate over much longer time-scales (1-100 seconds).

A non-linear continuous attractor network requires an additional mechanism to gain the functionality of an integrator: A way to shift the internal state along the attractor in response to an input that encodes changes in the external variable, Fig. 2d (left). Conceptually the simplest way to build a shift mechanism is by a *copy-and-offset* construction: construct multiple copies of the attractor network, each with slightly offset (asymmetric) weights in the sense that active neurons center their excitation or point of maximal disinhibition slightly offset from themselves on the neural sheet (e.g Fig. 1g (left) represents a slightly asymmetric version of Fig. 1c). The states in each such network will then form a limit cycle attractor, with patterns flowing in the direction of the asymmetry. If opposing copies are coupled together, the pattern is stabilized through a push-pull balance. A velocity input whose components project differentially to the copies will break the push-pull balance, allowing the more-active population of the moment to drive the pattern along its flow direction (cf Fig. 1g). Thus, the total direction and magnitude of the shift of the pattern, corresponding to movement along the attractor manifold, represents the time-integral of the velocity input to the network. This common principle unifies the mechanisms across diverse integrator models [61, 13, 14, 16, 62].

Decision making

If, instead of a velocity signal, the input to an integrator network consisted of temporally varying positive and negative evidence in support of two options [63], Fig. 2d (right) (or in the case of multiple options, evidence vectors instead of velocity vectors [64]), the network would integrate those inputs and thus perform evidence accumulation.

Decision-making can be viewed as a selection process applied to the integrator, based on a readout that detects when the integrator state has accumulated enough evidence and moved past a decision threshold [65, 54]. The selection process can be external to the integrator in the form of a readout circuit that detects such threshold crossings and outputs the decision; or it can be built into the dynamics of the integrator itself, in the form of a more-complex attractor landscape: the states evolve along a continuous attractor, but at some point the continuous attractor gives way to a pair of discrete attractors toward which the states flow, Fig. 2e. Neural winner-take-all (WTA) models implement such a hybrid analog-discrete computation [63, 17, 66, 67, 68, 64]. The parameters in WTA networks determine the balance between integration dynamics and competitive dynamics, and thus how well the network integrates later evidence (when the network is tuned to be a perfect integrator, its response to inputs is gradual and small amounts of evidence cause (reversible) flow along the continuous attractor manifold. In the case when competition dominates, the response to evidence is a fast flow toward one of the discrete attractors; beyond a point the flow is nearly irreversible, leading to rapid decision making and discounting of later evidence [69]).

Neural winner-take-all networks can accurately and rapidly (in $\sim \log(N)$ time) make the best decision among N alternatives, even if the presented data are noisy (fluctuating over time around their means) [68, 64] and the number of options varies over orders of magnitude [64].

Sequence generation

Attractor dynamics can be important for stabilizing another long time-scale behavior: the generation of sequences. Robust sequences can be constructed as low-dimensional limit cycle attractors, in which high-dimensional perturbations are corrected, while along the attractor there is a systematic, periodic, or quasiperiodic flow of states [70, 71, 72, 73, 74]. The attractor property that affords ongoing de-noising is important for preventing spatial dispersion and temporal dissipation of the activity packet during sequence generation.

As for stationary attractor manifolds, the small components of noise along the limit cycle attractors are not correctable and lead to a gradual accumulation of drift, which for sequence generation is manifest as timing variability: the standard-deviation in the time of reaching the Tth state in the sequence is predicted to grow as $\sqrt{T}$ for unbiased random drift along the attractor [53].

Evidence of attractors in the brain

Criteria for establishing attractor dynamics

The fundamental predictions of attractor models center on the state-space dynamics of the circuit, as first explicitly discussed and tested in a few papers [75, 9, 16, 76]: 1) That the system's states should be found localized at or around a much lower-dimensional set of states corresponding to the attractors in the state space. 2) That perturbations of the system should flow quickly back to the low-dimensional states. 3) That the set of attractor states – quantified by either direct characterization of the full state space or by the relationships between cells – should be invariant, persisting over time and after removal of tuned input, across conditions, across behavioral states, and even when there are induced variations in the mapping from internal states to external inputs [75, 16, 76]. 4) Integrator networks should further exhibit the property of isometry, in which lengths of coding space along a dimension are allocated to equal displacements along a dimension of the external variable. 5) Additional predictions of attractor dynamics models, that are not as fundamental in the sense that they are not theoretically necessary or sufficient but are nevertheless of high significance because they are highly supportive of the mechanisms of attractor dynamics, are anatomical and structural correlates: the existence of low-dimensional structures and symmetries in connectivity between cells.

Because attractor systems are characterized by their internally generated or autonomous dynamics[5], putative attractor networks are best tested in conditions that minimize external cues that are time-varying or tuned to provide localized inputs along the putative attractor.

Innovations in recording methods that made it possible to record multiple neurons simultaneously in animals performing naturalistic behaviors [77, 78, 79, 80], have enabled essential tests of these state-space predictions of attractor models. The newest methods provide activity data from $\sim$ 1000's of neurons within a circuit [81, 82, 83], making it possible to directly characterize the low-dimensional state-space dynamics of whole circuits [84, 85, 19, 86, 87].

When the attractor manifolds are $<$ 3-dimensional, one can directly visualize them by projecting or embedding the high-dimensional state-spaces into ≤ 3 dimensions (using e.g. PCA, multidimensional scaling, tensor factorization, and other linear methods for projection, or Isomap, LLE, tSNE, VAEs, LFADS, nonlinear tensor factorization, and so on, for nonlinear embedding [88, 89, 90, 91]). These methods can also be useful when manifolds have dimension ≥ 3 but are topologically simple [86, 92]. For topologically non-trivial structures (e.g. rings, torii), especially those of dimension ≥ 3, topological data analysis methods become important[93, 94, 95, 19, 96, 97, 87].

[5] Attractor networks dynamics need not be used by the brain in an autonomous setting: inputs that drive attractor networks can be an important part of their function, for instance in integration and evidence accumulation. However, for the purposes of identifying mechanism, it is important that their dynamics are probed in an (effectively) autonomous setting

Testing predictions 1)-3) requires examination of the state-space structure of the population response, rather than the more conventional characterization of relationships (tuning curves) between cell activity and input or output variables. The most direct way to examine the state-space structure is to record enough cells simultaneously that it is possible to characterize the full state-space manifold [19, 96, 20]. However, the existence, stability, and invariance of lower-dimensional state-space structures (predictions 1)-3)) can be inferred from smaller samples of simultaneously recorded cells, by characterizing the invariant structure in pairwise cell-cell relationships, as has been successfully done in a number of studies [75, 98, 99, 76, 100, 101].

Predictions 1) and 2) are necessary but not sufficient for identification of recurrent attractor dynamics in a target network. First, if the behaviors and inputs are themselves low-dimensional, then any observed low dimensionality of the circuit states may be ascribed to the inputs and reveals little about intrinsic constraints imposed by the circuit. Second, even if inputs and behaviors are high-dimensional, a low-dimensional feedforward projection into the target network can generate low-dimensional target states and rapid erasure of high-dimensional perturbations. The *sina qua non* of attractor dynamics is prediction 3), which is that, because the states are internally generated and stabilized by strong recurrent connectivity, the population states and cell-cell relationships should be invariant when probed across time and across a wide and rich variety of input conditions including the removal of tuned input and across waking and sleep. In simple terms, the states observed in 1)-2) should be invariant across a broad range of conditions [16, 76].

Next, to the question of circuit localization: If a circuit exhibits the key signatures of attractor dynamics, does it originate these dynamics or is it a readout of some other region? Localization need not be a primary goal of establishing attractor dynamics: an important first step is to simply characterize whether the brain solves certain problems through attractor dynamics, regardless which circuits create these dynamics. Nevertheless, the persistence of activity states in attractors can lend a helping hand to localization efforts. If a region originates or is upstream of the attractor dynamics, but not downstream of it, then perturbations that succeed in altering its low-dimensional state should persist after the perturbing drive is removed [102].

As we illustrate next, theoretically-motivated analyses of population data have now firmly established that low-dimensional attractor dynamics are ubiquitous in the brain, across levels in the brain's hierarchy and across species.

Discrete attractors

Up and down states

The simplest example of nontrivial discrete attractor dynamics (i.e., beyond a single point attractor) is bistability. Bistable dynamics are a feature of cortical activity in the form of up and down states [103, 104, 105, 106, 4, 107, 108], in which the subthreshold membrane potential of neurons switches between a hyperpolarized state and a relatively depolarized one, with long persistence (100's of milliseconds to seconds) per state, Fig. 3a. The two states are relatively invariant over time, as seen in the relatively sharply peaked histograms (Fig. 3a; predictions 1), 3)), and despite presumed internal noise in the system the peaks are well separated, suggesting a relatively rapid corrective dynamics towards the two states (prediction 2)). There is little evidence of a critical contribution from cellular bistability in supporting these states, suggesting that it is a network-driven phenomenon involving self-excitation and global inhibition [109, 110, 111, 104, 105, 106, 4, 107]. Transitions are believed to be driven through adaptation (from up to down) and stochastic as well as external coordinating events (from down to up) [108]. Though these states and switches can occur in cortex without input from thalamus and striatum, they tend to be synchronous across cortex and striatum. Thus, the origin of up and down states may be highly distributed.

Perceptual bistability

Visual and auditory percepts including binocular rivalry, the Necker cube, and some auditory illusions [112, 113, 114, 115, 116, 117, 118] offer clear examples of bistability in neural processing. In these illusions, the brain (at the level of perceptual reports) selects one possible interpretation of an ambiguous input, often switching between possibilities. Though the phenomenon has long been known and studied, no localized circuit has been identified as the basis of perceptual bistability. Indeed, some percepts may involve top-down activation and modulation of activity across many brain areas [116], suggesting once again a widely distributed circuit for bistability.

Bistability in a premotor area

Recent studies identify and localize discrete attractor dynamics in a mouse premotor area, the anterior lateral motor cortex (ALM) [119, 120, 121, 122]. In a cued 2-alternative delayed response task ALM neurons exhibit persistent activity over a ~ 1s delay period. During the post-cue delay period, activity evolves toward one of two states that guide the response, Fig. 3b (prediction 1)). The delay-period terminal states are similar for cues from different sensory modalities [123] (partial test of prediction 3)). ALM perturbations during the delay are either erased (corrected) by the circuit (Fig. 3b, top) or drive a jump to the opposite state (Fig. 3b, bottom), which results in the animal making the wrong action, suggesting a bistable switching dynamics similar to the mechanisms in either Fig.1b or Fig.2e (prediction 2)).

Given the long training time required for the task and the resulting tailoring of the ALM dynamics to the specific task structure – bistability for a two-choice task – it is likely that this system acquires its dynamics through slow plasticity and thus that the network's recurrent structure is malleable in adult animals. New results showing the existence of small ($\sim 100\mu$m scale) locally recurrent clusters of neurons ALM that can maintain persistent responses to microstimulation [124] may provide experimental evidence of the theoretically posited mixed modular networks (below) hypothesized to support robust and high-capacity memory states [60].

Discrete multistability

Hopfield networks and winner-take-all (WTA) networks are models of multistability beyond bistability[6]

To date, there are somewhat less direct data and exhaustive analyses to establish discrete multistability as a circuit-level brain process, in comparison to the evidence for continuous attractor networks (described next). However, there are many likely candidates systems and brain regions with dynamics suggestive of and consistent with discrete multistability, at least of the special case of WTA attractor dynamics, including in mammalian hippocampus and auditory cortex, and the fly and mamalian olfactory system [133, 134, 135, 136, 137, 138]. In particular, many of these circuits exhibit global inhibition that clearly narrows and refines activity in the circuit (prediction 5)), Fig. 3c top versus middle, and also show evidence of selective recurrent excitation that

[6] WTA networks [125, 126, 127, 67, 128, 129, 130, 131, 132] may be viewed as a special case of Hopfield networks, and bistable switch networks are a special case of WTA networks. As noted earlier, both can express mulitple discrete attractor states, but while the Hopfield network attractors have highly mixed and overlapping neural membership, WTA attractors consist of activity in largely disjoint groups of neurons.

leads to multiple distinct and stably correlated input responses in distinct subpopulations of cells, Fig. 3c, middle versus bottom [133, 134, 135, 136, 137, 138]. In our view, it is likely that these circuits exhibit multiple discrete attractor states but quantitative testing of predictions 1)-3) and direct demonstration of these states as stable and invariant remains an important future direction for characterizing these circuits.

Continuous attractors

The oculomotor integrator

The oculomotor integrator, together with the HD circuit, was one of the first systems in neuroscience to be studied theoretically [8, 139, 9] and experimentally [140] as a continuous attractor network – specifically as a line attractor, Fig. 1e. This network, presynaptic to the motor neurons that control horizontal eye position, is highly conserved across vertebrates, from fish [141, 140] to primates [142, 143]. It integrates pulse-like saccadic eye movement command signals to generate step-like stable muscle tension command signals (Fig.4a) that persist autonomously at various graded activity levels after removal of the movement cue and even in the dark in the absence of visual feedback (Fig.4b; prediction 3)) and thus enable stable gaze fixation at various eccentricities. Saccadic inputs knock the system slightly off the linear response states, but the neural responses rapidly decay back towards the persistent firing states (prediction 2)). Remarkably, the same system also integrates smooth head velocity signals to permit gaze stabilization during head movement. Integration functionality is a network-level rather than single-cell process: single neurons do not generate persistent responses to transient current injections, Fig. 4c (inset), while decreasing network feedback through synaptic blockers reduces the time-constant of integration and results in a leaky integrator [144], Fig.4c. It is possible to induce a reduction or increase in network feedback through training with a virtual surround that generates an artificial retinal slip percept, Fig.4d, showing that the system is capable of error-driven fine-tuning to maintain a high degree of persistence [145]. Finally, a recent EM reconstruction [146, 147] finds recurrent synaptic interconnectivty between integrator neurons, with excitatory connections between ipsilateral neurons and primarily inhibitory contralateral projections, in excellent agreement with line attractor models of the oculomotor circuit [9], Fig.1e (prediction 5)).

Head direction cells

Some of the earliest experiments to suggest the existence of low-dimensional continuous attractor dynamics were done in the rodent head-direction (HD) circuit[98, 75, 148], Fig.5a,b. The HD circuit in mammals maintains an updated internal compass estimate of heading direction (relative to some arbitrary external reference) as animals move around. It does so by integrating internal rotational velocity estimates during navigation and incorporating information from external cues [149, 150, 151, 152, 153]. The HD circuit is modeled by the ring attractor network [10, 61, 11, 154, 13, 14], Fig.1c, g (left). Before large population recordings became available, cell-cell correlations established that the network states remained invariant on a very low-dimensional manifold across environments [98, 75, 148], Fig.5a (predictions 1), 3)). Recently, the complete set of states of the several thousand neuron-sized mammalian HD network was shown to consist solely of a 1-dimensional ring, Fig.5b [19, 96] (prediction 1)), revealing that the brain has completely factorized its navigational representations to dedicate a circuit only to head direction. Further, intervals in the state-space ring manifold map isometrically to intervals of head direction (prediction 4)), as evidenced by a close match between the isometrically parameterized internal ring states and the measured head direction, Fig. 5b (inset, right).

Natural perturbations away from the ring flowed back to it, Fig.5d [19] (prediction 2)), and the ring manifold was invariant across waking and REM sleep, Fig.5e [19, 96] (prediction 3)). These findings explicitly validate the most fundamental predictions (predictions 1)-3)) of ring attractor models and continuous attractor-based integrators (predictions 1)-3) with 5)), providing arguably the most direct and compelling evidence of continuous attractor dynamics in the brain.

In a striking example of convergent evolution [155, 151], *Drosophila* compute HD estimates using apparently very similar dynamics [156, 157, 153, 152]. The fly neural compass circuit is topographically organized such that the neuropil forms a physical ring-shaped structure in the ellipsoid body, with a local moving activity peak that tracks head direction as the fly turns, Fig. 5f. Another notable advantage of the fly circuit in the effort to characterize its mechanisms is that the number of neurons is small and their morphology and connectivity has been fully traced [158], Fig.5g. This detailed view of the circuit permits quantitative, not just qualitative, comparisons with ring attractor models. The combined activity and connectivity data reveal that the fly HD system implements the copy-and-offset double-ring network architecture proposed for velocity integration [159, 14]. The actual dimensionality of the fly HD circuit and its full state-space dynamics remain to be characterized; even though the circuit is organized physically as a ring network, recent evidence suggests that the insect HD circuit may be involved in performing 2-dimensional path integration as well [160, 161], and thus unlike the ADn network in mammals, may not be confined to a 1D ring of attractor states that fully factorizes head direction in its representation of spatial variables.

Finally, the HD system can be re-anchored and reset based on tuned external cues [152, 153], which can change the orientiation tuning curves of cells and moment-by-moment firing rates of cells in a way that remains consistent with prediction 3) for attractor dynamics.

Grid cells

Grid cells encode spatial location though a regular triangular-lattice discharge pattern that tiles explored 2D spaces [162], representing 2D position as a set of spatially periodic 2D phases [163, 129]. They update their states while moving in the light and the dark [162], presumably based on motion cues. Continuous attractor models [15, 164, 16, 165] predict that the population states of a module – a set of grid cells with a common period – should be confined to merely 2 dimensions regardless of environment and behavioral state [16], forming a manifold that is topologically a torus, Fig.1d (rightmost column).

Indeed, grid cells from one module ("co-modular cells") have identical periods and orientations and all possible 2D phases, suggesting a 2D set of states [166, 76] (prediction 1)). The relative firing phase and relative grid parameters of pairs of co-modular cells is tightly conserved even as the spatial tuning and spatial phase of single cells varies across time and across familiar and novel environments (Fig. 6a) [166, 76], across the dimension of the spatial environment (6b) [167], and despite environmental rescaling that leads to large deformations in the spatial tuning of grid cells [76] (prediction 3)). Moreover, the detailed cell-cell relationships (whether a pair of cells is co-active, active in quadrature, or active fully out of phase) that are seen in waking exploration are conserved across overnight REM and non-REM sleep for grid cells but not place cells, Fig. 6c [100, 101], establishing that the low-dimensional states are autonomously generated (prediction 3)). These findings established that the structure of the grid cell response is very low-dimensional on a population level, invariant across environments, time, and behavioral states, and internally stabilized and autonomously generated – validating the fundamental predictions [16] made by continuous attractor grid models. Most recently, these findings were reproduced by a direct visualization of the state-space manifold, Fig6e made possible by large population recordings of grid cells across waking and sleep that confirmed the toroidal state-space structure of co-modular cells [20].

A corollary of these findings is that the grid cell response is *not* derived from upstream place cell inputs since place cells remap across environments and during sleep while grid cells retain their population structure (Fig. 6c); this corollary belies and is inconsistetnt with models in which the place cell response is primary to grid cells [168, 169, 170], as shown in [100].

Given the preserved internal structure and autonomous dynamics of grid cells across states, time, and environments, it follows that various deviations in the spatial tuning curves of grid cells from equilateral grid-like responses in 2- and 3-dimensional spaces [171, 172, 173, 174, 175, 176] likely result from variability in how the invariant internal states are driven by and mapped to external cues and states: for instance through altered velocity estimation [16, 76] or feedforward inputs that shift the phase of the grid cell network [177, 178, 179]. This has been verified in the case of the expansion of grid cells in novel environments [76] and is almost certain to hold – given the preponderance of evidence of internal grid stability [166, 76, 180, 100, 101, 20] – when tested in various other conditions that report grid deformations as well [172, 173, 174, 175, 176, 181].

In sum, the HD and grid cell systems confirm that the same pattern formation principle – based on local excitation or disinhibition, with broader inhibition – that is pivotal for morphogenesis in plants and animals [39] is also fundamental to the genesis of stationary continuous attractor states for computation and representation in the brain.

Graded working memory networks

In monkeys trained to make saccades to previously cued targets (selected from a set arranged in a circle), neurons in PFC and PPC exhibit persistent activity tuned to the direction of the initial cue, across the delay period after cue removal (predictions 1), 3)) [182, 183]. Analysis of delay-period PFC activity [85] in a population of simultaneously recorded shows that the delay period activity bump moves apparently randomly along a 1-dimensional manifold, with the characteristics of a diffusion process, so that the variance in the location of the bump grows linearly with time, as predicted by continuous attractor models [52, 16, 19], but the bump profile remains largely invariant over the duration of the delay (predictions 1) and 2), assuming that the diffusive process is indicative of natural noise-driven perturbations of the system). The bump movement predicts subsequent behavioral errors [85], suggesting that these states are repositories or readouts of the memory.

The need for extensive training on the task and the observed tailoring of the attractor states to this specific multi-cue task suggests that this attractor formed through learning in a flexible system rather by (re)using a genetically pre-specified circuit. Given the apparent malleability of this attractor network, we might therefore also expect a loss of the neural correlation structure if the animal is subsequently trained on other tasks, unlike with the grid and HD cell networks.

Attracting limit cycles and trajectories

The central and peripheral nervous systems contain numerous instances of periodic dynamics, ranging from the spiking of single neurons [184, 185] to circadian rhythms and sleep cycle generation [186], to rhythmic activity in motor circuits. While linear oscillators have amplitudes set by the initial condition, attractive limit cycle oscillators have an intrinsic and invariant amplitude. Thus, not all systems with oscillatory behavior are limit cycle attractors: oscillations that decay or grow over time or whose long-term amplitude or frequency changes after a transient perturbation are not limit cycles. Driven (non-autonomous) systems may exhibit limit cycles because of their inputs rather than intrinsic attractor dynamics [187].

Many of the oscillations noted above maintain a fixed amplitude, and because of their strong functional imperatives for robustness to perturbation are almost certainly generated through attractor dynamics. Particularly well-characterized examples are central pattern generators (CPGs) in motor circuits of the peripheral nervous system, that drive swimming, crawling, walking, breathing, and digestion, and differ in specifics across species but have common principles of mechanism and operation, including high robustness [188, 189]. CPG circuits typically integrate external feedback but have been shown to be able to operate in isolation without external drive [190].

Given the sizeable literature on these topics, we refer the reader to some excellent papers and reviews [191, 192, 193, 194, 195, 186].

Departures from low-dimensional continuous attractor dynamics

Not all circuits hypothesized to exhibit low-dimensional attractor dynamics appear under further experimentation to do so, or currently lack sufficient evidence to establish such dynamics within the circuit. We discuss three potential examples.

Orientation tuning in V1

The circuit of simple cells in V1 satisfies some key properties of ring attractor networks [11]: V1 and V2 cells exhibit strong orientation-tuned responses to real and illusory edges in the visual world [196, 197, 198], and population-wide V1 spontaneous activity during sleep is correlated with these tuned population coding states [199]. While these suggests that illusory edge responses may be driven by self-generated attractor dynamics, they tend to occur after a longer latency than real edge responses, making them more likely to be driven by top-down inputs rather than within-V1 dynamics. Next, moving an attractor state along a continuous attractor manifold requires strong inputs and is slow [200, 201]. These features seem inconsistent with the imperatives of a perceptual system to respond rapidly to changes in input [202], and the dynamics of state fluctuations in sleep appear relatively rapid compared to attractor time-scales [199]. These observations seem to lend more weight to models in which the circuit response is dominated by feedforward drive [196, 203], possibly with recurrently generated but fast non-normal amplification processes [49, 204]. More quantitative characterizations of response speed will be important to draw clear conclusions about competing models for V1 circuit dynamics.

Place cells

Place cells form stable and detailed representations of familiar 1-2 dimensional spatial environments [205], which can persist in the dark [206] and for short intervals after the animal has fallen asleep [207, 208]. In any particular 2-dimensional environment, the population response lies on a low-dimensional manifold in state-space [86]. Accordingly, the place cell circuit has been modeled as a 2-dimensional continuous attractor network [209, 210] or as a superposition (with overlapping neural membership) of a discrete number of such 2-dimensional continuous attractors, each representing a different environment[211]. However, the capacity limitations of generating multiple high-resolution maps within one homogeneous attractor network are severe [212, 213, 163, 214]. And, unlike grid cells, place cells do not, across long sleep bouts, exhibit the spatial correlations measured in awake exploration Fig. 6c [100, 101, 207]. Even during waking, cell-cell correlations are not preserved across environments because of the phenomenon of remapping [215, 216]. Like V1 neurons, place cells might be better described as deriving their tuning by forming conjunctions of multiple feedforward inputs, including from grid cells and cells that encode external cues like borders, landmarks, and reward sites [217, 218, 129, 214]. At the same time, place cells exhibit sequential activation of previous trajectories during activity replay [208, 219, 220, 221], which is hypothesized to be generated by recurrent connections in CA3, suggesting that recurrent and feedforward dynamics collaborate in the generation of place cell states; recent models are beginning to capture this interplay [210]. Closing the book on the question of autonomous low-dimensional dynamics in the far more complex response of place than grid cells requires more detailed experimentation, analysis, and modeling.

Motor cortical trajectories

Finally, recordings of motor cortical activity during stereotyped primate arm movements reveal the existence of stable low-dimensional trajectories [84, 222, 223, 224, 225], similar to the trajectories in state space originally characterized in olfactory circuit responses to odors [226]. Limit cycles and other low-dimensional attractors have been hypothesized to play a key role in cortical movement generation [227, 228]. The behaviors typically performed during neural recording are themselves restricted to be stereotyped and low-dimensional, thus it remains unclear whether activity would remain equally low-dimensional across a richer set of behaviors (e.g. over the set of all possible arm movements). Recent evidence from perturbation experiments [187] suggests that neural trajectories in motor cortex during skilled movements are driven by input from the thalamus, and thus that the circuits for motor pattern generation in the central nervous system might be distributed across multiple brain regions. Characterizing the intrinsic dimensionality of motor cortical activity, and determining whether the command to make more-complex motions involves multiple upstream or distributed primitive attractors, remain important open questions for both clinical brain-machine interfaces and neuroscience.

Flexibility despite rigidity: modern glimpses into the broader potential of attractor networks

Above, the key predictions and experimental validations of attractors in the brain hinged on their invariance, or rigidity, across time and conditions. The identified attractor states were highly structured and low-dimensional. The weight symmetries and asymmetries underlying these states were precisely tailored to the specific tasks performed by the systems. These properties appear to run counter to a key desideratum for representation, memory, and computation in the brain: flexibility.

Recent experimental and theoretical work are beginning to shed light on how the brain might solve the perennial conundrum of stability versus flexibility through attractor networks: the low-dimensional and rigid attractor states might be reused and recombined to create versatile and efficient systems for novel situations.

Exploiting integration for rapid representation: reuse of continuous attractors

Strikingly, all established stationary continuous attractor networks in the brain are also integrators. This seems surprising given that not all continuous variables represented in the brain need be accumulation of evidence or navigation-like variables. Here we discuss how, even for just the problem of representation, the functionality of integration could serve a vital role by enabling rapid construction of new representations.

Building a representation (mapping values of an external variable to the internal states on a continuous attractor) as in Fig. 2a can proceed by painstaking construction of a large set of associative feedforward correspondences: Visit each external state and associate it with an attractor state. Building this lookup table requires full exploration of the space. By contrast, if the attractor is an integrator, only two feedforward correspondences must be built: identify one value of the external variable with one internal state – an anchoring process – then associate the velocity signal with the shift mechanism in the integrator through a learned feedforward projection that is independent of location on the attractor or in the external space, Fig. 2f [229]. The circuit can now automatically generate appropriate and consistent representational states for future and previously unvisited values of the variable based on displacements. This mapping is rapid and does not require exhaustive exploration of the space or reconfiguration of the recurrent attractor circuit, a form of generalizable and rapid learning [230, 231, 229] and *zero-shot memory state construction* [229]. Moreover, an integrator network can correctly infer the current state upon returning to it along a previously untraversed path (novel trajectory), a form of *zero-shot inference* [231, 229, 232].

This rapid integration-based mapping process permits another use: A single attractor can be easily *reused* to represent mutiple variables, Fig. 2f: By simply adding another anchor point for another variable Z and driving the shift mechanism with velocities related to changes in Z, the network can switch from representing variable X to de-novo representing Z, or can alternate between representing X and Z without any reconfiguration of the attractor itself [229]. Consistent with this idea, it appears that the brain (re)uses grid and place cells when navigating through both the spatial environment and through non-spatial cognitive domains [233, 234, 235].

Multiple modular attractors for high-capacity representation

In general, a fully connected symmetric attractor network of N neurons permits the construction of $\sim N$ chosen attractor states (where the notation $\sim$ refers to the functional scaling with N with potential prefactors that do not depend on N), a result established by a large body of work from statistical physics [236, 237, 238] and information theory [239]. The result is independent of learning rules [239, 236] and not qualitatively altered by adding hidden neurons [60]. It also applies to (near-)continuous attractors, for which the total number of distinguishable states (resolution) on an attractor manifold also scales in the same way.

A D-dimensional attractor with resolution $\sim P$ per dimension would thus require a number of neurons that grows exponentially with dimension, $N \sim P^D$ (this is one aspect of the *curse of dimensionality*). For 2-dimensional space at a resolution of 10 cm per dimension, the full population of $\sim 10^6$ rodent hippocampal cells – which have been hypothesized to function as a Hopfield-like associative memory – could at most represent a combined 100m^2 area [163]. Similarly, the number of cells in hippocampus is vastly smaller than the combinatorially large number of sparse coding states of cortical neurons in rodents and humans, which might set the scale for the number of items to be stored in memory. These arguments point to the need for much higher-capacity representations and memory than possible with fully-connected Hopfield-like networks.

Modular attractor networks permit efficient construction of a large number of representational and memory states in neural networks [163, 129, 240, 241, 242, 243, 60, 214]. If M modular subnetworks have $\sim N$ discrete attractor states each, and these can update independently (uncoupled modules), the combined system expresses a set of $\sim N^M$ states, exponential in the number of modules. Though these states are not attractors, it is possible to couple together these subnetworks to generate exponentially many attractor states so they each have reasonable-sized basins and are thus robust [244, 241, 242, 60, 57], Fig. 2. (By contrast, randomly connected Hopfield networks also typically have exponentially many fixed points, but the basins are often small.) Similar ideas have been applied in the temporal domain to show how networks might support exponentially long activity sequences [245].

If each module expresses a continous attractor of dimension K and the subnetworks are independent, their combined states define an $MK-$dimensional manifold (with no error correction between states on the manifold), solving the curse of dimensionality for representing higher-dimensional variables while maintaining a structured representation for the variable that goes well beyond random combinatorial codes [163, 129, 246]. These subnetworks can be coupled together through their shift mechanisms, forcing a certain fixed relationship between modules in how their relative states update [243], in which case the coupled system exhibits one K-dimensional attractor but with large capacity, containing $\sim e^M$ states per dimension.

In short, modular subnetworks can work together to greatly expand representational capacity in terms of number of attractor states while also maintaining a large denoising capability [244, 241, 242, 60, 57, 163, 129, 246, 243]. However, the vast set of attractor states made possible by these coupled-module constructions are a rather structured set pre-defined as combinations of the states in the individual modules, rather than being arbitrarily specified as the states are in a standard Hopfield network. They do not directly encode user-defined patterns as the attractor states, and thus do not violate the capacity limitations of neural networks [236, 237, 238, 239]. A critical question is how high-capacity and robust but structured sets of attractor states could be leveraged for general memory and computation. Three recent works have begun to address this question, showing that structured attractor states can be leveraged for robust labeling and action selection [247], robust classification [248], and as a component of a heterogeneously structured general associative memory that exhibits smooth degradation instead of catastrophic memory loss as more patterns beyond capacity are added to the network sharma2022content.

Mixed modular codes for flexible representation

Finally, M modular subnetworks that are each integrators in K dimensions can be (re)used without any rewiring of recurrent weights to represent and store inputs of any input dimension $\leq MK$, Fig. 2h, using a *mixed-modular coding scheme* [229]. This scheme combines five concepts: Rapid representation learning with the integration mechanism, the reuse of the same attractors for different variables, the capacity of modular attractors, the fact that a high-dimensional variable can be represented unambiguously by multiple independent lower-dimensional projections, and the fact that random projections tend to be independent.

In mixed modular coding, movement along each dimension in the external space is randomly projected to the shift mechanisms of all modular integrators. Every module is thus involved in representing every input dimension – a form of holographic representation [249, 250]. If the number of input dimensions D is smaller than MK, all input dimensions are represented without information loss, and excess module capacity ($MK - D$ excess dimensions) is automatically convertible into extending the coding range or resolution for each of the D input dimensions, instantly trading off the number of represented dimensions and the dynamic range of each represented dimension without recurrent plasticity.

Looking ahead

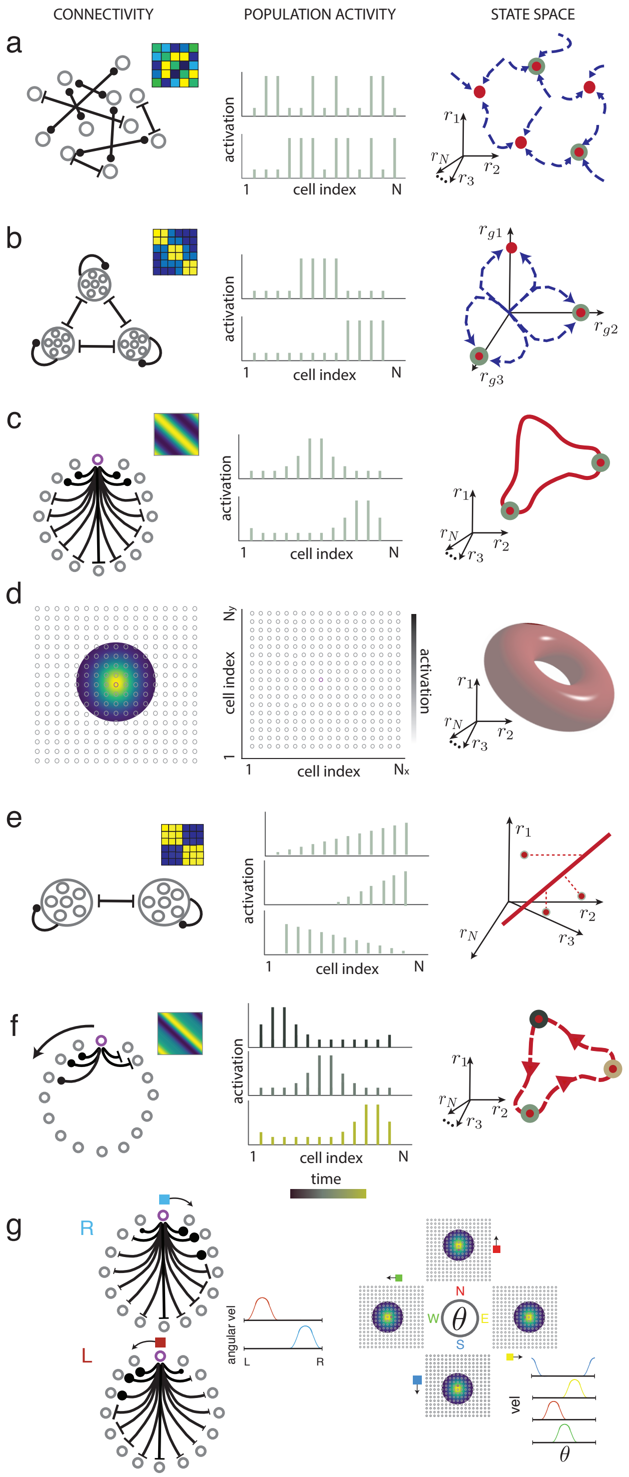
The theory of attractor dynamics in the brain has provided a powerful and unifying conceptual framework for understanding integration, representation, memory, error-correction, and efficient learning and inference in the brain. The experimental effort to study candidate attractor circuits and test their predictions has been a fertile field of research, and population-wide physiology techniques have led to breathtaking direct visualizations of attractor dynamics at work in the brain.

The theory is also proving to be a powerful tool in interpreting how artifical neural networks (ANNs) solve complex tasks. ANNs trained to robustly solve memory, integration, and decision-making tasks in domains as diverse as spatial navigation, vision, and language, develop attractor dynamics [251, 252, 253, 254, 255], suggesting that not only are attractor networks able to solve such problems but might be necessary when the computing elements are memoryless neurons. Further, equipping networks with preconfigured attractor networks can help to produce faster, more data-efficient and generalizable learning [230, 229]. Because ANNs can be trained on complex tasks and then fully examined after learning, they will potentially more readily contribute to the next chapter in our understanding of how continuous attractor networks can interact and combine with other mechanisms to allow the brain to solve richer problems associated with intelligence.

Notable mechanistic questions about attractor networks also remain open, including: Moving away from the high-firing-rate asynchronous spiking regime [256, 257] to better understand whether low-firing-rate synchronous spiking networks might support attractor dynamics – and thus permit a combination of fast time-scale dynamics like spike synchronization and oscillatory phase dynamics [256, 258, 259]. For continuous attractors, understanding how the brain deals with the problem of fine-tuning in linear networks or the imposition and maintenance of a continuous symmetry across neurons remains unknown and ripe for resolution [34, 260].

A few continuous attractor development models show how they could emerge simply through unsupervised associative plasticity [10, 211, 178]; others are based on combining feedback of known or plausible error signals with neural activity in relatively simple learning rules [261, 10, 262]; the rest train networks on a high-level goal with error backpropagation, combined with other constraints on architecture or the form the solutions should take [263, 264, 252, 253, 230, 265, 254]. These models are incomplete for different reasons: the unsupervised models require uniform exploration of the input variable space and suppression of recurrent weights during their training; the backpropagation models do not offer an account of how the loss functions, learning, and additional constraints might be generated in biological systems.

There is much left to do in the field and an exciting vista ahead. On the experimental side, tools for high-resolution population-level neural recrdings and perturbation across multiple brain areas [266, 82, 83] let us peer further and deeper than ever. On the theory side, future developments will help us conceptualize how such circuits could help underwrite intellegent computation through the formation, interaction, and reuse of multiple low-dimensional structures.


CONNECTIVITY
POPULATION ACTIVITY
STATE SPACE
a
activation
1
cell index
N
r_1
r_N
r_2
r_3
b
r_{g1}
r_{g2}
r_{g3}
c
d
N_y
1
activation
N_x
e
f
time
g
R
L
angular vel
N
W
E
S
θ
vel

Correspondences between attractor dynamics and anatomical layout, and weight symmetries in models and biology

Anatomical topography, in which functionally similar neurons are near one another, is neither a necessary nor sufficient condition for the existence of an attractor: Any low-dimensional attractor network is mathematically unchanged if all weights are preserved but neuron locations are scrambled. However, if the network is merely a spatially scrambled version of the idealized model, then the symmetries of the weight matrix can be revealed after an appropriate reordering of the neurons. An advantage of anatomical topography from a biological perspective is that it can reduce the complexity of development, in that wiring decisions can be guided by spatial proximity rather than depending entirely on activity or other target cell signalling mechanisms; for instance, the locally competitive interactions of grid and HD circuit models could be largely constructed through local arborization. It also reduces overall wiring length in the mature circuit [275]. However, a circuit with ≥ 3-dimensional dynamics that are represented in an unfactorizable form cannot be embedded topographically in a 2-dimensional cell layout, which limits the utility of topographic layouts for circuits representing higher-dimensional manifolds.

Second, the posited weight symmetries in simple models of attractors need not exist in a biological instance of the circuit with the same dynamics: If low-dimensional attractor dynamics is only needed downstream of the size-N recurrent network that generates the dynamics, in a set of $M < N$ neurons, then the symmetries required for continuous attractor dynamics can be spread across the recurrent and readout weights and the recurrent weights alone without taking into account the readout weights will not reflect the relevant symmetries [253]. These considerations suggest a hypothesis for circuits with ≤ 2-D continuous attractors: Evolutionarily conserved circuits that do not require extensive early experience [276, 277] should be topographically organized. We might thus predict that the circuit that originates HD signals in mammals should be topographically organized. By contrast, if the low-dimensional dynamics only emerges on the basis of activity-dependent plasticity with repetitive training, we may not expect the circuit to be topographically organized (or even localized to single brain regions).

Remarkably, despite these caveats and in a beautiful example of the predictive power of simple theories in neuroscience, the recent empirical evidence from the anatomy of the zebrafish oculomotor integrator and the fly HD circuit show that nature has used precisely the hypothesized constructions proposed in simple circuit models to build some integrator networks.

Figure 1 *(preceding page)*: **Mechanisms of attractor formation.** In all plots, open gray neurons represent neurons, connections between them are excitatory (black lines ending in bars) or inhibitory (black lines ending in circles). Left column: layout of neurons and connections; connectivity matries shown as inset, with blue to yellow colors indicating strongly inhibitory to excitatory interactions. Middle column: examples of stable population activity patterns. Right column: state-space views of population states and dynamics. Red circles with gray-green ring indicate the activity states shown in middle column. (a) A network with dense symmetric connections determined by associative Hebbian learning on a set of input patterns (middle) stores them as stable attractor states. This defines a Hopfield network. (b) Disjoint groups of neurons interacting through within-group excitation and across-group mutual inhibition leads to group winner-take-all dynamics. Stable states are any patterns with only one winning group (the state-space plot collapses all activities of neurons in group gi along the axis r_{gi}). (c) Neurons arranged in a ring with global inhibition and either local excitation or a lack of local inhibition combined with uniform excitatory input to all neurons produces localized activity bumps (middle) as the stable states. Bumps may be centered anywhere on the neural ring, defining a near-continuum of attractor states that form a ring in state-space (right). (d) Neurons arranged on a two-dimensional sheet, interacting through local inhibition and either center-excitation or a lack of inhibition near the center together with uniform excitatory input to all neurons results in a pattern of multiple, periodically spaced activity bumps (middle). Any two-dimensional phase shift of the periodic pattern upto the lattice periodicity are distinct but equivalent stable states, then the states repeat; thus these are predicted to form a torus in the state-space. (e) Two neuron groups with in-group excitation and across-group inhibition, precisely tuned interaction strengths, and quasi-linear neural fI responses can counteract activity decay in the network and produce persistent activity over a continuum of activity levels in the two populations, defining ramp-like population activity states and a line of attractor states. (f) Neurons arranged on a ring with asymmetric connections that bias neural activity to flow in a particular direction (middle) The network forms localized activity bumps that sequentially move around the ring in that direction (right) The state space contains a limit cycle. (g) The copy-and-offset mechanism for constructing integrators, illustrated for the ring (left) and grid (right) attractor circuits. Each network copy receives velocity inputs tuned to the corresponding shift direction.

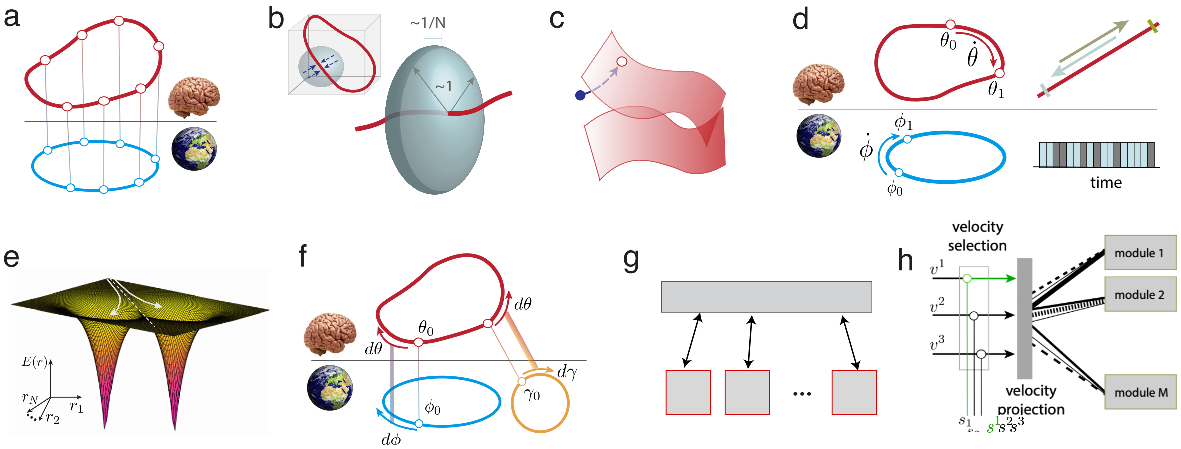


Figure 2: **Utility of low-dimensional attractor networks.** (a) Persistent and stable states generated by attractor networks (red) can be used to represent and remember external variables (blue) by constructing an appropriate mapping between them (vertical lines). (b) Noise-robustness: attractor networks error-correct by mapping noisy states to the nearest attractor state [267]. Main: N-dimensional noise drawn from the unit sphere centered on a 1D attractor has a projection strength of only $1/N$ along the attractor: in this counter-intuitive high-dimensional geometry, a ball is more like a pancake with the attractor orthogonal to the large dimensions [19]. (c) Flow to the nearest (continuous or discrete) attractor can perform a nearest-neighbor computation and thus perform classification: e.g. the two attractors may represent "cat" and "dog" perceptual manifolds, and the blue dot a specific input data point. (d) Left: Continuous attractors can become integrators if velocities or movements in the external space are inputs to the network and induce proportional shifts in the internal attractor state: The current state on the attractor is then the integral of past velocity inputs relative to the starting state. Right: if the input to an integrating attractor consists of temporally varying evidence pulses (bottom, evidence about one option in blue and evidence about the opposing option in khaki), these will move the state on the attractor (top) so its current state reflects the integral of the total evidence.(e) The energy landscape of a combined integration and decision making network: inputs push the state left or right, and as it integrates, the network state also moves toward one of two discrete attractors (left and right; white arrows: two sample trajectories). Arrival to the neighborhood of one of the discrete attractors is a decision point [63, 64]. (f) An integrator can be quickly re-purposed to represent multiple different and new external variables simply by yoking its velocity shift mechanism to different external velocities cues by feedforward learning. It also does zero-shot learning and inference: Given an initial state and an input velocity trajectory, it will generate a self-consistent representation for the current state even if the trajectory if different and new each time [229, 231, 232]. (g) A set of (continuous or discrete) attractor subnetworks (red boxes at bottom) can interact bidirectionally with a shared network to form a high-capacity attractor network [241, 242, 268, 60]. (h) Mixed modular representations can enable representation of inputs of different dimensions by resuing the same attractors of fixed dimension each. Velocities (v_i) from external spaces of potentially different dimension are selected by a set of selection signals (s_i). The selected velocity (green) is routed through random projections to a set of M modular integrator networks of dimension K each. This kind of mixed modular circuit can interchangeably represent a variety of input spaces of dimension $D \leq MK$ while smoothly trading off resolution for dimension [229].

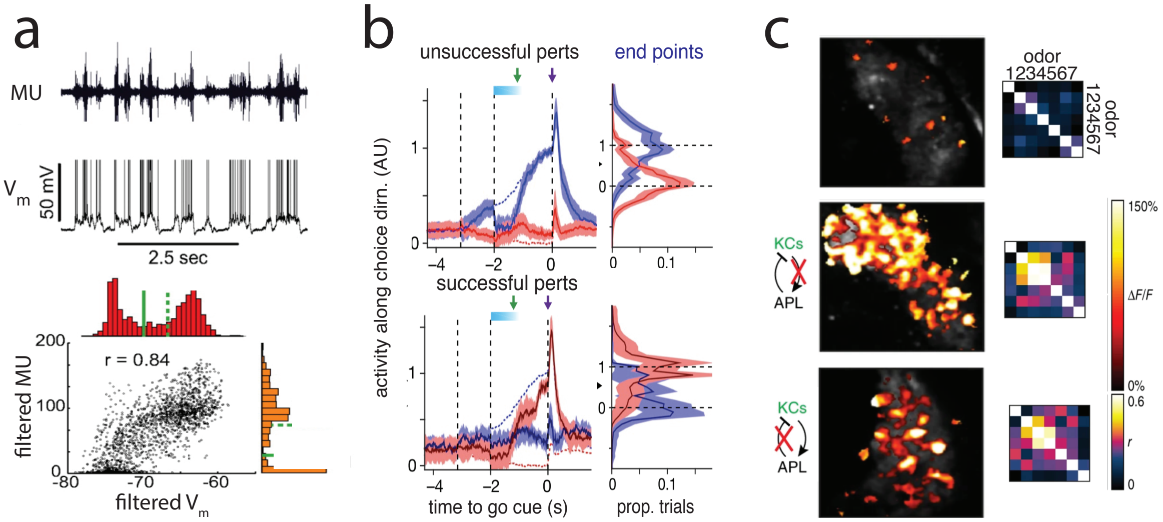


Figure 3: **Evidence of discrete attractor dynamics in the brain.** (a) Multi-unit (above) and single-unit (middle) activity during cortical up- and down-states show signatures of bistability (clusters and histograms at bottom). Reproduced with permission from [269]. (b) Delay-period dynamics in rodent premotor area ALM during a binary decision task before the animal can make a motor report of its decision appears to converge to one of two discrete end points (blue and red curves and histograms, top). Perturbations are either robustly ignored (top), or flip the dynamics so that the end points are reversed (bottom). Reproduced with permission from [123, 119]. (c) Evidence of all-to-all inhibition and competitive winner-take-all recurrent dynamics in the fly olfactory system: Kenyon cell (KC) responses to odors, with input from the globally projecting APL inhibitory neuron, are sparse (top left, Ca fluorescence response to odor IA) and decorrelated across odors (top right); blockage of KC drive to APL or APL inhibition to KCs results in dense and correlated odor responses (middle, bottom). Reproduced with permission from [137].

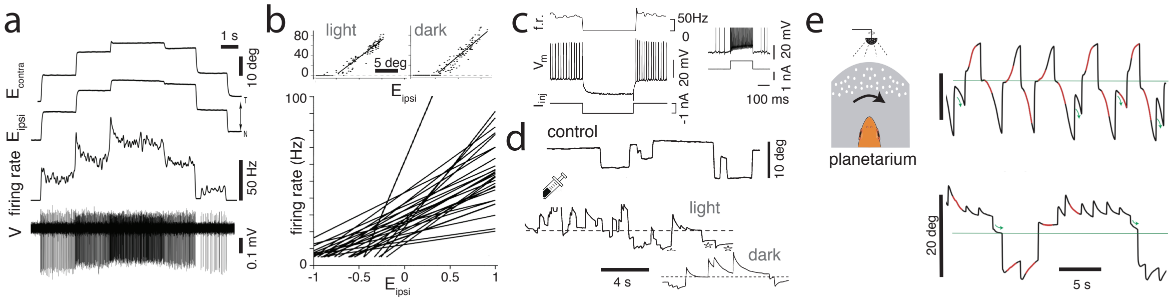


Figure 4: **Linear attractor dynamics generated by nework feedback in the oculomotor integrator.** (a) Stable horizontal gaze during fixation at different angular positions (top two traces) is supported by stable steps in firing rate by oculomotor integrator neurons (bottom two traces) that integrate transient ($\sim$ 100 ms) saccadic command bursts. Reproduced with permission from [140]. (b) Oculomotor neurons maintain eye position through linearly ramping tuning curves (bottom); responses are the same in the light and dark (top) and thus do not depend on visual input for gaze stabilization on the time-scale of seconds. Reproduced with permission from [270]. (c) Transient current injection into single oculomotor neurons reveals a transient, not persistent, decrease or elevation (inset) in firing rate, consistent with lack of a cellular origin for persistent intersaccadic firing. Reproduced with permission from [140]. (d) Injection of kainic acid into the oculomotor integrator produces leaky dynamics even in the light (inset, faster leak in the dark), consistent with network models. Adapted with permission from [271]. (e) Visual feedback mimicking leaky or unstable eye positions in goldfish can mistune the oculomotor integrator, making it unstable or leaky, respectively. Adapted with permission from [272, 145].

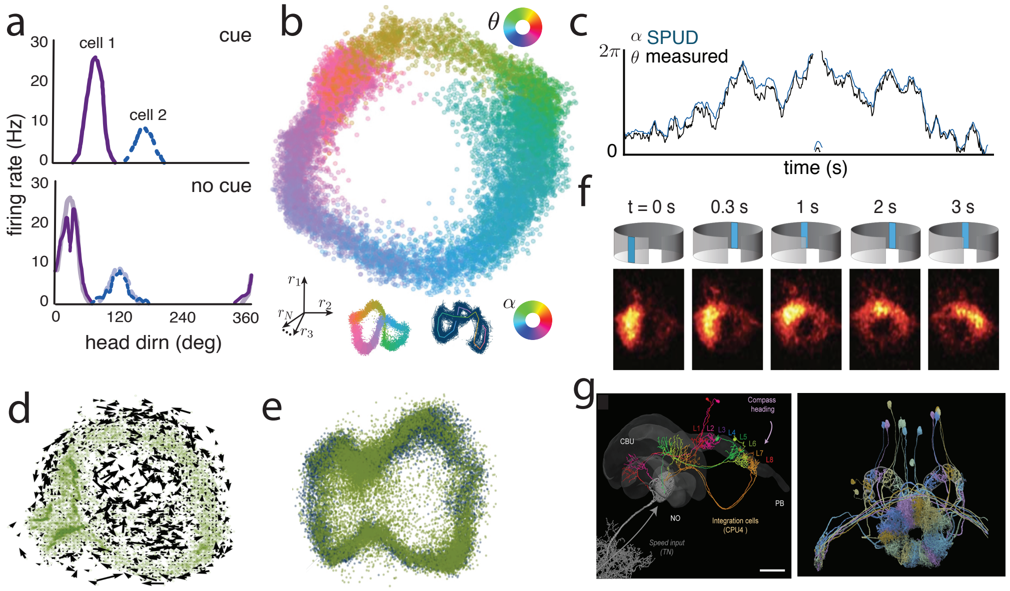


Figure 5: **The head direction circuit: a ring attractor in the brain.** (a) Activity of two cells in the rat HD circuit during free foraging in a 2-dimensional circular arena with a globally orienting cue (top). When the cue is removed (bottom), the fields rotate but the cells maintain their tuning shapes and relative tuning angles (pale curves: top plot, globally rotated. Adapted with permission from [75, 98]. (b) A nonlinear 2-dimensional embedding of the population-level states of the thousands-of-neurons sized mammalian thalamic area ADn recorded during 2-dimensional free-foraging: the states are confined to a 1-dimensional ring (cf. Fig. 1c); here, colors encode the measured head direction of the rodent. Inset: Non-linear embedding of states from a different rodent (left), with coloring obtained by an isometric parametrization along the ring by SPUD [19]. Reproduced with permission from [19]. (c) A close match between unsupervised isometric parametrization of the manifold from (b, inset) and the externally measured head direction of the rodent. Reproduced with permission from [19]. (d-e) The same cells as in (b, inset), recorded during REM sleep (green): the states during REM remain confined to a 1-dimensional ring that precisely overlays the ring of waking states (blue, in (e)), and exhibit large flows back toward the ring (in (d)) Reproduced with permission from [19]. (f) Calcium imaging of activity in the physically ring-shaped *Drosophila* ellipsoid body reveals a localized bump of excitation that follows the movement of a cue in the fly's visual field. Reproduced with permission from [156]. (g) Combination of electrophysiology and EM imaging of the central complex in bees (left) and flies (right) provides detailed layout and connectivity data for comparison with predicted connectivity in ring attractor models. Reproduced with permission from [161, 273].

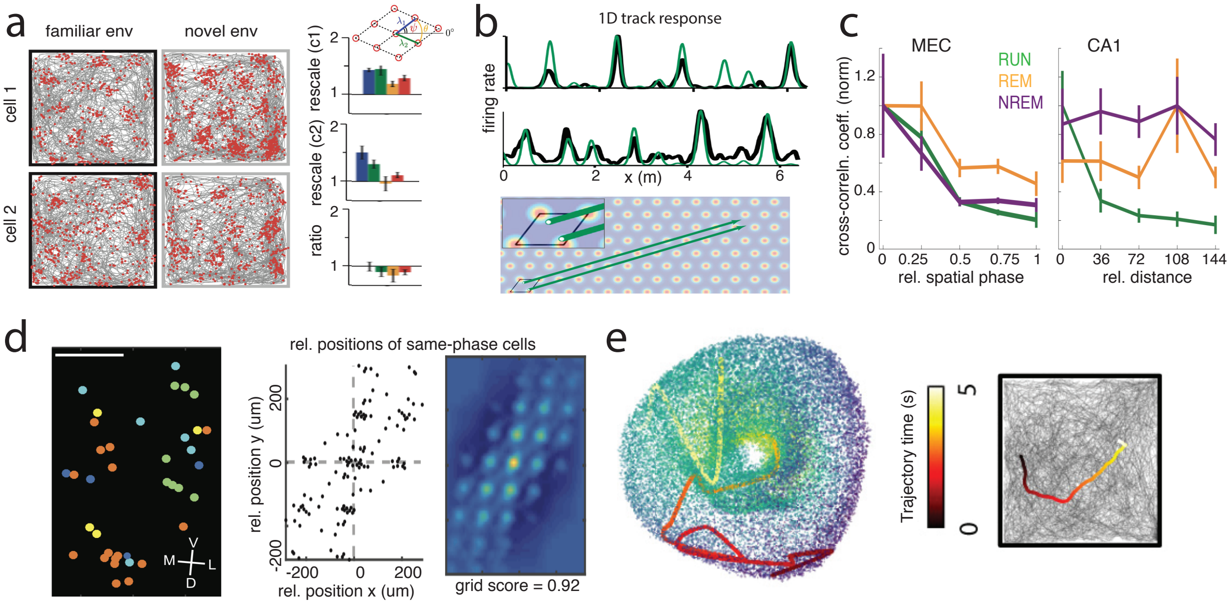


Figure 6: **A set of 2-dimensional toroidal attractors in the grid cell system.** (a) An example pair of grid cells (left column) whose spatial tuning periods and orientations reconfigure in novel environments (middle column), but the changes are tightly yoked to preserve cell-cell relationships (right column). Each color corresponds to a variable describing the lattice of the spatial tuning curve of the cell as shown in the schematic. Adapted with permission from [76]. (b) Responses of co-modular cells on 1D linear tracks can explained by parallel slices through a 2-dimensional grid, suggesting preserved 2-dimensional circuit dynamics across diverse environments. Reproduced with permission from [167]. (c) Pairwise correlations between grid cells measured during navigation are preserved across overnight sleep, while those of place cells are not. Reproduced with permission from [100, 101]. (d) Grid cells are anatomically arranged according to their relative spatial firing phases. (left) Cell positions are colored according to the phase of their spatial tuning curves. The relative cortical positions of same-phase cells make a triangular lattice pattern (middle), with a grid-like autocorrelation pattern (right). Reproduced with permission from [274] (e) Non-linear dimensionality reduction and topological data analysis directly reveal that the states of individual grid modules lie on a torus (left); as the animal follows a spatial trajectory (right), the state moves along the manifold (left). Reproduced with permission from [20].

Acknowledgements

I.R.F. acknowledges funding from the Simons Foundation, ONR, HHMI through the Faculty Scholars Program, the Department of Brain and Cognitive Sciences, MIT, and the McGovern Institute, MIT. M.K. is supported by a Friends of the McGovern Institute Fellowship, a MathWorks Fellowship, and the Department of Physics, MIT.
We are grateful to Kayvon Daie, the anonymous reviewers, Natasha Bray, and Sarthak Chandra and other members of the Fiete lab for helpful comments on the manuscript.

References

[1] S-I Amari. Neural theory of association and concept-formation. *Biological cybernetics*, 26(3):175–185, 1977.

[2] John J Hopfield. Neural networks and physical systems with emergent collective computationalabilities. *Proc Natl Acad Sci U S A*, 79:2554–2558, April 1982.

[3] William A Little. The existence of persistent states in the brain. In *From High-Temperature Superconductivity to Microminiature Refrigeration*, pages 145–164. Springer, 1974.

[4] Hugh R Wilson and Jack D Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80, 1973.

[5] John J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci U S A*, 81:3088–3092, May 1984.

[6] Michael A Cohen and Stephen Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE transactions on systems, man, and cybernetics*, (5):815–826, 1983.

[7] S Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern*, 27(2):77–87, Aug 1977.

[8] Stephen C Cannon, David A Robinson, and Shihab Shamma. A proposed neural network for the integrator of the oculomotor system. *Biological cybernetics*, 49(2):127–136, 1983.

[9] H Sebastian Seung. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23):13339–13344, 1996.

[10] K. Zhang. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*, 15:2112–2126, 1996.

[11] R Ben-Yishai, R L Bar-Or, and H Sompolinsky. Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci U S A*, 92(9):3844–8, Apr 1995.

[12] Bard Ermentrout. Neural networks as spatio-temporal pattern-forming systems. *Reports on progress in physics*, 61(4), 4 1998. An optional note.

[13] SM Stringer, TP Trappenberg, ET Rolls, and IETd Araujo. Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Network: Computation in Neural Systems*, 13(2):217–242, 2002.

[14] Xiaohui Xie, Richard H R Hahnloser, and H Sebastian Seung. Double-ring network model of the head-direction system. *Phys Rev E Stat Nonlin Soft Matter Phys*, 66(4 Pt 1):041902, Oct 2002.

[15] Mark C Fuhs and David S Touretzky. A spin glass model of path integration in rat medial entorhinal cortex. *J Neurosci*, 26(16):4266–76, Apr 2006.

[16] Yoram Burak and Ila R Fiete. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput Biol*, 5(2):e1000291, Feb 2009.

[17] Xiao-Jing Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5):955–968, 2002.

[18] J. W. Milnor. Attractor. *Scholarpedia*, 1(11):1815, 2006. revision #186525.

[19] Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature neuroscience*, 22(9):1512–1520, 2019.

[20] Richard J Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A Baas, Benjamin J Dunn, May-Britt Moser, and Edvard I Moser. Toroidal topology of population activity in grid cells. *bioRxiv*, 2021.

[21] Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC Press, 2018.

[22] Christof Koch. *Biophysics of computation: information processing in single neurons*. Oxford university press, 2004.

[23] Michael N Shadlen and William T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of neuroscience*, 18(10):3870–3896, 1998.

[24] Cyril Hanus and Erin M Schuman. Proteostasis in complex dendrites. *Nature Reviews Neuroscience*, 14(9):638, 2013.

[25] William James. *The Principles of Psychology*. Henry Holt Co., 1890.

[26] W. MCDOUGALL. ON THE SEAT OF THE PSYCHO-PHYSICAL PROCESSES. *Brain*, 24(4):579–630, 10 1901.

[27] Donald O. Hebb. *The Organization of Behavior*. John Wiley Sons., 1949.

[28] Richard E Brown, Thaddeus W B Bligh, and Jessica F Garden. The hebb synapse before hebb: Theories of synaptic function in learning and memory before , with a discussion of the long-lost synaptic theory of william mcdougall. *Front Behav Neurosci*, 15:732195, 2021.

[29] SJ Martin, PD Grimwood, and RG Morris. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu Rev Neurosci 2000*, 23:649–711, 2000.

[30] Wickliffe C Abraham, Owen D Jones, and David L Glanzman. Is plasticity of synapses the mechanism of long-term memory storage? *NPJ science of learning*, 4(1):1–10, 2019.

[31] Tomonori Takeuchi, Adrian J Duszkiewicz, and Richard GM Morris. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1633):20130288, 2014.

[32] P. W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.

[33] Huaguang Zhang, Zhanshan Wang, and Derong Liu. A comprehensive review of stability analysis of continuous-time recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7):1229–1262, 2014.

[34] Alfonso Renart, Pengcheng Song, and Xiao-Jing Wang. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, 38(3):473–485, 2003.

[35] Vladimir Itskov, David Hansel, and Misha Tsodyks. Short-term facilitation may stabilize parametric working memory trace. *Frontiers in computational neuroscience*, 5:40, 2011.

[36] Alan. M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72, August 1952.

[37] A. Gierer and H. Meinhardt. A theory of biological pattern formation. *Kybernetik*, 12(1):30–39, dec 1972.

[38] Mark C Cross and Pierre C Hohenberg. Pattern formation outside of equilibrium. *Reviews of modern physics*, 65(3):851, 1993.

[39] A. J. Koch1 and H. Meinhardt. Biological pattern formation : from basic mechanisms to complex structures. *Rev. Modern Physics*, (66):1481–1507, 1994.

[40] François Schweisguth and Francis Corson. Self organization in pattern formation. *Dev. Cell*, (49):659–677, Jun 2019.

[41] Boris Shraiman. Mechanical feedback as a possible regulator of tissue growth. *Proceedings of the National Academy of Sciences*, 102(9):3318–3323, Jun 2005.

[42] Noji S. Ueno N. Maini P.K. Sekimura, T. *Morphogenesis and Pattern Formation in Biological Systems: Experiments and Models*. Springer, 2003.

[43] Christian Boucheny, Nicolas Brunel, and Angelo Arleo. A continuous attractor network model without recurrent excitation: maintenance and integration in the head direction cell system. *Journal of computational neuroscience*, 18(2):205–227, 2005.

[44] Jonathan J Couey, Aree Witoelar, Sheng-Jia Zhang, Kang Zheng, Jing Ye, Benjamin Dunn, Rafal Czajkowski, May-Britt Moser, Edvard I Moser, Yasser Roudi, and Menno P Witter. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci*, 16(3):318–24, Mar 2013.

[45] H Sebastian Seung. Amplification, attenuation, and integration. *The handbook of brain theory and neural networks*, 2:94–97, 2003.

[46] C van Vreeswijk and H Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–6, Dec 1996.

[47] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural networks. *Physical review letters*, 61(3):259, 1988.

[48] Rainer Engelken, Fred Wolf, and LF Abbott. Lyapunov spectra of chaotic recurrent neural networks. *arXiv preprint arXiv:2006.02427*, 2020.

[49] Brendan K Murphy and Kenneth D Miller. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron*, 61(4):635–648, 2009.

[50] Lloyd N Trefethen, Anne E Trefethen, Satish C Reddy, and Tobin A Driscoll. Hydrodynamic stability without eigenvalues. *Science*, 261(5121):578–584, 1993.

[51] Gianluigi Mongillo, Omri Barak, and Misha Tsodyks. Synaptic theory of working memory. *Science*, 319(5869):1543–6, Mar 2008.

[52] Albert Compte, Nicolas Brunel, Patricia S Goldman-Rakic, and Xiao-Jing Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral cortex*, 10(9):910–923, 2000.

[53] Yoram Burak and Ila R Fiete. Fundamental limits on persistent activity in networks of noisy neurons. *Proc Natl Acad Sci U S A*, 109(43):17645–50, Oct 2012.

[54] Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D Cohen. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev*, 113(4):700–65, Oct 2006.

[55] Eric B Baum, John Moody, and Frank Wilczek. Internal representations for associative memory. *Biological Cybernetics*, 59(4-5):217–228, 1988.

[56] Lawrence K Saul and Michael I Jordan. Attractor dynamics in feedforward neural networks. *Neural computation*, 12(6):1313–1335, 2000.

[57] Sugandha Sharma, Sarthak Chandra, and Ila R Fiete. Content addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold. *arXiv preprint arXiv:2202.00159*, 2022.

[58] S Funahashi, C J Bruce, and P S Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol*, 61(2):331–49, Feb 1989.

[59] Clayton E. Curtis and Mark D'Esposito. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci*, 7(9):415–423, Sep 2003.

[60] Rishidev Chaudhuri and Ila Fiete. Bipartite expander hopfield networks as self-decoding high-capacity error correcting codes. In *Advances in Neural Information Processing Systems*, pages 7686–7697, 2019.

[61] David Redish, Adam N Elga, and David S Touretzky. A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems*, 7:671685, 1996.

[62] Pengcheng Song and Xiao-Jing Wang. Angular path integration by moving "hill of activity": a spiking neuron model without recurrent excitation of the head-direction system. *Journal of Neuroscience*, 25(4):1002–1014, 2005.

[63] Xiao-Jing Wang. Decision making in recurrent neuronal circuits. *Neuron*, 60(2):215–34, Oct 2008.

[64] B. Kriener, R. Chaudhuri, and I.R. Fiete. Robust parallel decision-making in neural circuits with nonlinear inhibition. *PNAS*, (to appear), 2020.

[65] Marius Usher and James L McClelland. The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3):550, 2001.

[66] Kong-Fatt Wong and Xiao-Jing Wang. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.

[67] Richard HR Hahnloser, Rahul Sarpeshkar, Misha A Mahowald, Rodney J Douglas, and H Sebastian Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947, 2000.

[68] Rafal Bogacz and Kevin Gurney. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput*, 19(2):442–77, Feb 2007.

[69] Genís Prat-Ortega, Klaus Wimmer, Alex Roxin, and Jaime de la Rocha. Flexible categorization in perceptual decision making. *Nature communications*, 12(1):1–15, 2021.

[70] Brad E Pfeiffer and David J Foster. Autoassociative dynamics in the generation of sequences of hippocampal place cells. *Science*, 349(6244):180–183, July 2015.

[71] Rodrigo Laje and Dean V Buonomano. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nature neuroscience*, 16(7):925–933, 2013.

[72] David Kleinfeld. Sequential state generation by model neural networks. *Proceedings of the National Academy of Sciences*, 83(24):9469–9473, 1986.

[73] Haim Sompolinsky and I Kanter. Temporal association in asymmetric neural networks. *Physical review letters*, 57(22):2861, 1986.

[74] Ila R Fiete, Walter Senn, Claude Z H Wang, and Richard H R Hahnloser. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron*, 65(4):563–76, Feb 2010.

[75] J S Taube, R U Muller, and J B Ranck, Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. ii. effects of environmental manipulations. *J Neurosci*, 10(2):436–47, Feb 1990.

[76] K.J. Yoon, M.A. Buice, R. Barry, C.and Hayman, N. Burgess, and I.R. Fiete. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat Neurosci*, 16(8):1077–84, Aug 2013.

[77] B L McNaughton, J O'Keefe, and C A Barnes. The stereotrode: a new technique for simultaneous isolation of several single units in the central nervous system from multiple unit records. *J Neurosci Methods*, 8(4):391–7, Aug 1983.

[78] James J Jun, Nicholas A Steinmetz, Joshua H Siegle, Daniel J Denman, Marius Bauza, Brian Barbarits, Albert K Lee, Costas A Anastassiou, Alexandru Andrei, Çağatay Aydın, et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236, 2017.

[79] Misha B Ahrens, Jennifer M Li, Michael B Orger, Drew N Robson, Alexander F Schier, Florian Engert, and Ruben Portugues. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature*, 485(7399):471–477, 2012.

[80] Brian A Wilt, Laurie D Burns, Eric Tatt Wei Ho, Kunal K Ghosh, Eran A Mukamel, and Mark J Schnitzer. Advances in light microscopy for neuroscience. *Annu Rev Neurosci*, 32:435–506, 2009.

[81] Abdulmalik M Obaid, Mina-Elraheb S Hanna, Yu-Wei Wu, Mihaly Kollo, Romeo R Racz, Matthew R Angle, Jan Muller, Nora Brackbill, William Wray, Felix Franke, et al. Massively parallel microwire arrays integrated with cmos chips for neural recording. *bioRxiv*, page 573295, 2019.

[82] Siegfried Weisenburger and Alipasha Vaziri. A guide to emerging technologies for large-scale and whole-brain optical imaging of neuronal activity. *Annu Rev Neurosci*, 41:431–452, 07 2018.

[83] Nicholas A. Steinmetz, Cagatay Aydin, Anna Lebedeva, Michael Okun, Marius Pachitariu, Marius Bauza, Maxime Beau, Jai Bhagat, Claudia Böhm, Martijn Broux, Susu Chen, Jennifer Colonell, Richard J. Gardner, Bill Karsh, Dimitar Kostadinov, Carolina Mora-Lopez, Junchol Park, Jan Putzeys, Britton Sauerbrei, Rik J. J. van Daal, Abraham Z. Vollan, Marleen Welkenhuysen, Zhiwen Ye, Joshua Dudman, Barundeb Dutta, Adam W. Hantman, Kenneth D. Harris, Albert K. Lee, Edvard I. Moser, John O'Keefe, Alfonso Renart, Karel Svoboda, Michael Häusser, Sebastian Haesler, Matteo Carandini, and Timothy D. Harris. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. oct 2020.

[84] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51–56, 2012.

[85] Klaus Wimmer, Duane Q Nykamp, Christos Constantinidis, and Albert Compte. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience*, 17(3):431–439, 2014.

[86] Ryan J Low, Sam Lewallen, Dmitriy Aronov, Rhino Nevers, and David W Tank. Probing variability in a cognitive map using manifold inference from neural dynamics. *bioRxiv*, page 418939, 2018.

[87] Richard J. Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A. Baas, Benjamin A. Dunn, May-Britt Moser, and Edvard I. Moser. Toroidal topology of population activity in grid cells. feb 2021.

[88] Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, Jaimie M Henderson, Krishna V Shenoy, L F Abbott, and David Sussillo. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat Methods*, 15(10):805–815, 10 2018.

[89] J. B. Tenenbaum. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[90] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

[91] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[92] Anqi Wu, Stan Pashkovski, Sandeep R Datta, and Jonathan W Pillow. Learning a latent manifold of odor representations from neural responses in piriform cortex. In *Advances in Neural Information Processing Systems*, pages 5378–5388, 2018.

[93] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discret. Comput. Geom.*, 33(2):249–274, February 2005.

[94] Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[95] Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zornorodian. On the local behavior of spaces of natural images. *Int. J. Comput. Vis.*, 76(1):1–12, January 2008.

[96] Erik Rybakken, Nils Baas, and Benjamin Dunn. Decoding of neural data using cohomological feature extraction. *Neural Comput*, 31(1):68–93, 01 2019.

[97] Gurjeet Singh, Facundo Memoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario L Ringach. Topological analysis of population activity in visual cortex. *Journal of vision*, 8(8):11–11, 2008.

[98] J S Taube, R U Muller, and J B Ranck, Jr. Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *J Neurosci*, 10(2):420–35, Feb 1990.

[99] D Yoganarasimha, Xintian Yu, and James J Knierim. Head direction cell representations maintain internal coherence during conflicting proximal and distal cue rotations: comparison with hippocampal place cells. *J Neurosci*, 26(2):622–31, Jan 2006.

[100] S.G. Trettel, J.B. Trimper, E. Hwaun, I.R. Fiete, and L.L. Colgin. Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors. *Nat Neurosci*, 22(4):609–617, 04 2019.

[101] Richard J Gardner, Li Lu, Tanja Wernle, May-Britt Moser, and Edvard I Moser. Correlation structure of grid cells is preserved during sleep. *Nat Neurosci*, 22(4):598–608, 04 2019.

[102] John Widloski, Michael P Marder, and Ila R Fiete. Inferring circuit mechanisms from sparse neural recording and global perturbation in grid cells. *eLife*, 7:e33503, 2018.

[103] Maria V Sanchez-Vives, Marcello Massimini, and Maurizio Mattia. Shaping the default activity pattern of the cortical network. *Neuron*, 94(5):993–1001, Jun 2017.

[104] Silvia Scarpetta and Antonio de Candia. Alternation of up and down states at a dynamical phase-transition of a neural network with spatiotemporal attractors. *Front Syst Neurosci*, 8:88, 2014.

[105] Rosa Cossart, Dmitriy Aronov, and Rafael Yuste. Attractor dynamics of network up states in the neocortex. *Nature*, 423(6937):283–8, May 2003.

[106] Barthó P Luczak A Compte A de la Rocha J. Jercog D, Roxin A. Up-down cortical dynamics reflect state transitions in a bistable network. *eLife*, (6:e22425), August 2017.

[107] Daniel Jercog, Alex Roxin, Peter Barthó, Artur Luczak, Albert Compte, and Jaime de la Rocha. Up-down cortical dynamics reflect state transitions in a bistable network. *Elife*, 6, 08 2017.

[108] Maria V Sanchez-Vives, Marcello Massimini, and Maurizio Mattia. Shaping the default activity pattern of the cortical network. *Neuron*, 94(5):993–1001, Jun 2017.

[109] Peter E Latham, BJ Richmond, PG Nelson, and S Nirenberg. Intrinsic dynamics in neuronal networks. i. theory. *Journal of neurophysiology*, 83(2):808–827, 2000.

[110] Albert Compte, Maria V Sanchez-Vives, David A McCormick, and Xiao-Jing Wang. Cellular and network mechanisms of slow oscillatory activity (< 1 hz) and wave propagations in a cortical network model. *Journal of neurophysiology*, 89(5):2707–2725, 2003.

[111] Fernando Kasanetz, Luis A Riquelme, Patricio O'Donnell, and M Gustavo Murer. Turning off cortical ensembles stops striatal up states and elicits phase perturbations in cortical and striatal slow oscillations in rat in vivo. *The Journal of physiology*, 577(1):97–113, 2006.

[112] Diana Deutsch. An auditory illusion. *Nature*, 251(5473):307–309, 1974.

[113] Emily J Ward and Brian J Scholl. Stochastic or systematic? seemingly random perceptual switching in bistable events triggered by transient unconscious cues. *Journal of Experimental Psychology: Human Perception and Performance*, 41(4):929, 2015.

[114] Randolph Blake and Nikos K Logothetis. Visual competition. *Nature Reviews Neuroscience*, 3(1):13–21, 2002.

[115] Richard McWalter and Josh H McDermott. Illusory sound texture reveals multi-second statistical completion in auditory scene analysis. *Nature communications*, 10(1):1–18, 2019.

[116] Megan Wang, Daniel Arteaga, and Biyu J He. Brain mechanisms for simple perception and bistable perception. *Proceedings of the National Academy of Sciences*, 110(35):E3350–E3359, 2013.

[117] Shashaank Vattikuti, Phyllis Thangaraj, Hua W Xie, Stephen J Gotts, Alex Martin, and Carson C Chow. Canonical cortical circuit model explains rivalry, intermittent rivalry, and rivalry memory. *PLoS Comput Biol*, 12(5):e1004903, 05 2016.

[118] Rubén Moreno-Bote, John Rinzel, and Nava Rubin. Noise-induced alternations in an attractor network model of perceptual bistability. *J Neurophysiol*, 98(3):1125–39, Sep 2007.

[119] Hidehiko K Inagaki, Lorenzo Fontolan, Sandro Romani, and Karel Svoboda. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature*, 566(7743):212–217, 02 2019.

[120] Nuo Li, Kayvon Daie, Karel Svoboda, and Shaul Druckmann. Robust neuronal dynamics in premotor cortex during motor planning. *Nature*, 532(7600):459–464, 2016.

[121] Alex T Piet, Jeffrey C Erlich, Charles D Kopec, and Carlos D Brody. Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics. *Neural computation*, 29(11):2861–2886, 2017.

[122] Jeffrey C Erlich, Bingni W Brunton, Chunyu A Duan, Timothy D Hanks, and Carlos D Brody. Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *Elife*, 4:e05457, 2015.

[123] Hidehiko K Inagaki, Miho Inagaki, Sandro Romani, and Karel Svoboda. Low-dimensional and monotonic preparatory activity in mouse anterior lateral motor cortex. *Journal of Neuroscience*, 38(17):4163–4185, 2018.

[124] Kayvon Daie, Karel Svoboda, and Shaul Druckmann. Targeted photostimulation uncovers circuit motifs supporting short-term memory. *Nat Neurosci*, 24(2):259–265, 02 2021.

[125] John Lazzaro, Sylvie Ryckebusch, Misha Anne Mahowald, and Caver A Mead. Winner-take-all networks of o (n) complexity. In *Advances in neural information processing systems*, pages 703–711, 1989.

[126] Xiaohui Xie, Richard HR Hahnloser, and H Sebastian Seung. Selectively grouping neurons in recurrent networks of lateral inhibition. *Neural computation*, 14(11):2627–2646, 2002.

[127] E Majani, Ruth Erlanson, and Yaser S Abu-Mostafa. On the k-winners-take-all network. In *Advances in neural information processing systems*, pages 634–642, 1989.

[128] Kevin A Bolding and Kevin M Franks. Recurrent cortical circuits implement concentration-invariant odor coding. *Science*, 361(6407):eaat6904, 2018.

[129] Sameet Sreenivasan and Ila Fiete. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat Neurosci*, 14(10):1330–7, Sep 2011.

[130] Licurgo de Almeida, Marco Idiart, and John E Lisman. The input-output transformation of the hippocampal granule cells: from grid cells to place fields. *J Neurosci*, 29(23):7504–12, Jun 2009.

[131] Claudia Espinoza, Segundo Jose Guzman, Xiaomin Zhang, and Peter Jonas. Parvalbumin+ interneurons obey unique connectivity rules and establish a powerful lateral-inhibition microcircuit in dentate gyrus. *Nat Commun*, 9(1):4605, 11 2018.

[132] Simone Kurt, Anke Deutscher, John M Crook, Frank W Ohl, Eike Budinger, Christoph K Moeller, Henning Scheich, and Holger Schulze. Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions. *PLoS One*, 3(3):e1735, Mar 2008.

[133] Licurgo de Almeida, Marco Idiart, and John E Lisman. The input-output transformation of the hippocampal granule cells: from grid cells to place fields. *J Neurosci*, 29(23):7504–12, Jun 2009.

[134] Claudia Espinoza, Segundo Jose Guzman, Xiaomin Zhang, and Peter Jonas. Parvalbumin+ interneurons obey unique connectivity rules and establish a powerful lateral-inhibition microcircuit in dentate gyrus. *Nat Commun*, 9(1):4605, 11 2018.

[135] Simone Kurt, Anke Deutscher, John M Crook, Frank W Ohl, Eike Budinger, Christoph K Moeller, Henning Scheich, and Holger Schulze. Auditory cortical contrast enhancing by global winner-take-all inhibitory interactions. *PLoS One*, 3(3):e1735, Mar 2008.

[136] Sheena A Josselyn and Susumu Tonegawa. Memory engrams: Recalling the past and imagining the future. *Science*, 367(6473), 01 2020.

[137] Andrew C Lin, Alexei M Bygrave, Alix de Calignon, Tzumin Lee, and Gero Miesenböck. Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination. *Nat Neurosci*, 17(4):559–68, Apr 2014.

[138] Charles F Stevens. What the fly's nose tells the fly's brain. *Proc Natl Acad Sci U S A*, 112(30):9460–5, Jul 2015.

[139] DB Arnold and DA Robinson. The oculomotor integrator: testing of a neural network model. *Experimental brain research*, 113(1):57–74, 1997.

[140] Emre Aksay, G Gamkrelidze, H Sebastian Seung, Robert Baker, and David W Tank. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nature neuroscience*, 4(2):184–193, 2001.

[141] Am Pastor, La De Rr Cruz, and R Baker. Eye position and eye velocity integrators reside in separate brainstem nuclei. *Proc Natl Acad Sci U S A.*, 91(2):807–11., Jan 18 1994.

[142] C. Cannon and D.A. Robinson. Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey. *J Neurophys*, 57(5):1383–1409, May 1987.

[143] P. Mettens, E. Godaux, G. Cheron, and H.L. Galiana. Effect of muscimol microinjections into the prepositus hypoglossi and the medial vestibular nuclei on cat eye movements. *J Neurophysiol.*, 72(2):785–802, Aug 1994.

[144] Chris RS Kaneko. Eye movement deficits after ibotenic acid lesions of the nucleus prepositus hypoglossi in monkeys. i. saccades and fixation. *Journal of neurophysiology*, 78(4):1753–1768, 1997.

[145] Guy Major, Robert Baker, Emre Aksay, Brett Mensh, H Sebastian Seung, and David W Tank. Plasticity and tuning by visual feedback of the stability of a neural integrator. *Proc Natl Acad Sci U S A*, 101(20):7739–44, May 2004.

[146] Alexandro D.Ramirez Jeff W.Lichtman Emre R.F.Aksay H. SebastianSeung Ashwin Vishwanathan, Kayvon Daie. Electron microscopic reconstruction of functionally identified cells in a neural integrator. *Current Biology*, 27(14):2137–2147, July 2017.

[147] Ashwin Vishwanathan, Alexandro Ramirez, Jingpeng Wu, Alex Sood, Runzhe Yang, Nico Kemnitz, Dodam Ih, Nicholas Turner, Kisuk Lee, Ignacio Tartavull, et al. Predicting modular functions and neural coding of behavior from a synaptic wiring diagram. *bioRxiv*, pages 2020–10, 2021.

[148] J S Taube. Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *J Neurosci*, 15(1 Pt 1):70–86, Jan 1995.

[149] Ryan M Yoder and Jeffrey S Taube. The vestibular contribution to the head direction signal and navigation. *Front Integr Neurosci*, 8:32, 2014.

[150] Ryan M Yoder, James R Peck, and Jeffrey S Taube. Visual landmark information gains control of the head direction signal at the lateral mammillary nuclei. *J Neurosci*, 35(4):1354–67, Jan 2015.

[151] Brad K Hulse and Vivek Jayaraman. Mechanisms underlying the neural computation of head direction. *Annu Rev Neurosci*, 43:31–54, 07 2020.

[152] Yvette E Fisher, Jenny Lu, Isabel D'Alessandro, and Rachel I Wilson. Sensorimotor experience remaps visual input to a heading-direction network. *Nature*, 576(7785):121–125, 12 2019.

[153] Sung Soo Kim, Ann M Hermundstad, Sandro Romani, L F Abbott, and Vivek Jayaraman. Generation of stable heading representations in diverse visual scenes. *Nature*, 576(7785):126–131, 12 2019.

[154] Alexei Samsonovich and Bruce L. McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci*, 17:5900–5920, August 1997.

[155] Dora E Angelaki and Jean Laurens. The head direction cell network: attractor dynamics, integration within the navigation system, and three-dimensional properties. *Current Opinion in Neurobiology*, 60:136–144, 2020.

[156] Sung Soo Kim, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849–853, 05 2017.

[157] Jonathan Green, Atsuko Adachi, Kunal K Shah, Jonathan D Hirokawa, Pablo S Magani, and Gaby Maimon. A neural circuit architecture for angular integration in drosophila. *Nature*, 546(7656):101, 2017.

[158] Daniel B Turner-Evans, Kristopher T Jensen, Saba Ali, Tyler Paterson, Arlo Sheridan, Robert P Ray, Tanya Wolff, J Scott Lauritzen, Gerald M Rubin, Davi D Bock, and Vivek Jayaraman. The neuroanatomical ultrastructure and function of a biological ring attractor. *Neuron*, 108(1):145–163.e10, 10 2020.

[159] W E Skaggs, J J Knierim, H S Kudrimoti, and B L McNaughton. A model of the neural basis of the rat's sense of direction. *Adv Neural Inf Process Syst*, 7:173–80, 1995.

[160] Cheng Lyu, LF Abbott, and Gaby Maimon. A neuronal circuit for vector computation builds an allocentric traveling-direction signal in the drosophila fan-shaped body. *bioRxiv*, 2020.

[161] Thomas Stone, Barbara Webb, Andrea Adden, Nicolai Ben Weddig, Anna Honkanen, Rachel Templin, William Wcislo, Luca Scimeca, Eric Warrant, and Stanley Heinze. An anatomically constrained model for path integration in the bee brain. *Current Biology*, 27(20):3069–3085, 2017.

[162] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E.I. Moser. Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806, 2005.

[163] Ila R Fiete, Yoram Burak, and Ted Brookings. What grid cells convey about rat location. *J Neurosci*, 28(27):6858–71, Jul 2008.

[164] Yoram Burak and Ila Fiete. Do we understand the emergent dynamics of grid cell activity? *J Neurosci*, 26(37):9352–9354, 2006.

[165] Alexis Guanella, Daniel Kiper, and Paul Verschure. A model of grid cells based on a twisted torus topology. *Int. J. Neural Syst.*, 17(4):231–240, August 2007.

[166] Marianne Fyhn, Torkel Hafting, Alessandro Treves, May-Britt Moser, and Edvard I Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–194, 2007.

[167] KiJung Yoon, Sam Lewallen, Amina A Kinkhabwala, David W Tank, and Ila R Fiete. Grid cell responses in 1d environments assessed as slices through a 2d lattice. *Neuron*, 89(5):1086–99, Mar 2016.

[168] E. Kropff and A. Treves. The emergence of grid cells: intelligent design or just adaptation? *Hippocampus*, 18(12), 2008.

[169] Yedidyah Dordek, Daniel Soudry, Ron Meir, and Dori Derdikman. Extracting grid cell characteristics from place cell inputs using non-negative principal component analysis. *Elife*, 5:e10094, Mar 2016.

[170] Kimberly L Stachenfeld, Matthew M Botvinick, and Samuel J Gershman. The hippocampus as a predictive map. *Nat Neurosci*, 20(11):1643–1653, Nov 2017.

[171] Caswell Barry, Robin Hayman, Neil Burgess, and Kathryn J Jeffery. Experience-dependent rescaling of entorhinal grids. *Nat Neurosci*, 10(6):682–684, 2007.

[172] Julija Krupic, Marius Bauza, Stephen Burton, Caswell Barry, and John O'Keefe. Grid cell symmetry is shaped by environmental geometry. *Nature*, 518(7538):232–235, Feb 2015.

[173] Robin M A Hayman, Giulio Casali, Jonathan J Wilson, and Kate J Jeffery. Grid cells on steeply sloping terrain: evidence for planar rather than volumetric encoding. *Front Psychol*, 6:925, 2015.

[174] Charlotte N Boccara, Michele Nardin, Federico Stella, Joseph O'Neill, and Jozsef Csicsvari. The entorhinal cognitive map is attracted to goals. *Science*, 363(6434):1443–1447, 2019.

[175] Gily Ginosar, Johnatan Aljadeff, Yoram Burak, Haim Sompolinsky, Liora Las, and Nachum Ulanovsky. Locally ordered representation of 3d space in the entorhinal cortex. *Nature*, 596(7872):404–409, Aug 2021.

[176] Roddy M Grieves, Selim Jedidi-Ayoub, Karyna Mishchanchuk, Anyi Liu, Sophie Renaudineau, Éléonore Duvelle, and Kate J Jeffery. Irregular distribution of grid cell firing fields in rats exploring a 3d volumetric space. *Nat Neurosci*, Aug 2021.

[177] Peter E Welinder, Yoram Burak, and Ila R Fiete. Grid cells: the position code, neural network models of activity, and the problem of learning. *Hippocampus*, 18(12):1283–1300, 2008.

[178] John Widloski and Ila R Fiete. A model of grid cell development through spatial exploration and spike time-dependent plasticity. *Neuron*, 83(2):481–495, Jul 2014.

[179] Kiah Hardcastle, Surya Ganguli, and Lisa M Giocomo. Environmental boundaries as an error correction mechanism for grid cells. *Neuron*, 86(3):827–39, May 2015.

[180] Yi Gu, Sam Lewallen, Amina A Kinkhabwala, Cristina Domnisoru, Kijung Yoon, Jeffrey L Gauthier, Ila R Fiete, and David W Tank. A map-like micro-organization of grid cells in the medial entorhinal cortex. *Cell*, 175(3):736–750.e30, 10 2018.

[181] William N Butler, Kiah Hardcastle, and Lisa M Giocomo. Remembered reward locations restructure entorhinal spatial maps. *Science*, 363(6434):1447–1452, 2019.

[182] J W Gnadt and R A Andersen. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp Brain Res*, 70(1):216–20, 1988.

[183] C Constantinidis, M N Franowicz, and P S Goldman-Rakic. Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *J Neurosci*, 21(10):3646–55, May 2001.

[184] Eugene M Izhikevich. *Dynamical systems in neuroscience*. MIT press, 2007.

[185] Peter Ashwin, Stephen Coombes, and Rachel Nicks. Mathematical frameworks for oscillatory network dynamics in neuroscience. *The Journal of Mathematical Neuroscience*, 6(1):1–92, 2016.

[186] Antoine R Adamantidis, Carolina Gutierrez Herrera, and Thomas C Gent. Oscillating circuitries in the sleeping brain. *Nature Reviews Neuroscience*, 20(12):746–762, 2019.

[187] Britton A Sauerbrei, Jian-Zhong Guo, Jeremy D Cohen, Matteo Mischiati, Wendy Guo, Mayank Kabra, Nakul Verma, Brett Mensh, Kristin Branson, and Adam W Hantman. Cortical pattern generation during dexterous movement is input-driven. *Nature*, 577(7790):386–391, 2020.

[188] Angela M Bruno, William N Frost, and Mark D Humphries. A spiral attractor network drives rhythmic locomotion. *Elife*, 6, 08 2017.

[189] Annika LA Nichols, Tomáš Eichler, Richard Latham, and Manuel Zimmer. A global brain state underlies c. elegans sleep behavior. *Science*, 356(6344):eaam6851, 2017.

[190] Dirk Bucher, Gal Haspel, Jorge Golowasch, and Farzan Nadim. Central pattern generators. *eLS*, pages 1–12, 2015.

[191] Eve Marder and Dirk Bucher. Central pattern generators and the control of rhythmic movements. *Current biology*, 11(23):R986–R996, 2001.

[192] Eve Marder and Ronald L Calabrese. Principles of rhythmic motor pattern generation. *Physiological reviews*, 76(3):687–717, 1996.

[193] Martyn Goulding. Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Reviews Neuroscience*, 10(7):507–518, 2009.

[194] Ole Kiehn. Decoding the organization of spinal circuits that control locomotion. *Nature Reviews Neuroscience*, 17(4):224, 2016.

[195] Rafael Yuste, Jason N MacLean, Jeffrey Smith, and Anders Lansner. The cortex as a central pattern generator. *Nature Reviews Neuroscience*, 6(6):477–483, 2005.

[196] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.

[197] R von der Heydt, E Peterhans, and G Baumgartner. Illusory contours and cortical neuron responses. *Science*, 224(4654):1260–2, Jun 1984.

[198] D H Grosof, R M Shapley, and M J Hawken. Macaque v1 neurons can signal 'illusory' contours. *Nature*, 365(6446):550–2, Oct 1993.

[199] Amiram Grinvald, Edmund Lieke, Ron D Frostig, Charles D Gilbert, and Torsten N Wiesel. Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature*, 324(6095):361–364, 1986.

[200] Weishun Zhong, Zhiyue Lu, David J Schwab, and Arvind Murugan. Non-equilibrium statistical mechanics of continuous attractors. *Neural Computation, in press*, 2020.

[201] Chi Chung Alan Fung, Tomoki Fukai, et al. Discrete-attractor-like tracking in continuous attractor neural networks. *Physical review letters*, 122(1):018102, 2019.

[202] S Thorpe, D Fize, and C Marlot. Speed of processing in the human visual system. *Nature*, 381(6582):520–2, Jun 6 1996.

[203] David Ferster, Sooyoung Chung, and Heidi Wheat. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380(6571):249–252, 1996.

[204] Guillaume Hennequin, Yashar Ahmadian, Daniel B Rubin, Máté Lengyel, and Kenneth D Miller. The dynamical regime of sensory cortex: Stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*, 98(4):846–860.e5, 05 2018.

[205] J O'Keefe and J Dostrovsky. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Res*, 34(1):171–175, 1971.

[206] G J Quirk, R U Muller, and J L Kubie. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *J Neurosci*, 10(6):2008–2017, 1990.

[207] Matthew A Wilson and Bruce L McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679, 1994.

[208] W E Skaggs and B L McNaughton. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271(5257):1870–3, Mar 1996.

[209] Misha Tsodyks and Terrence Sejnowski. Associative memory and hippocampal place cells. *International journal of neural systems*, 6:81–86, 1995.

[210] Haggai Agmon and Yoram Burak. A theory of joint attractor dynamics in the hippocampus and the entorhinal cortex accounts for artificial remapping and grid cell field-to-field variability. *Elife*, 9:e56894, 2020.

[211] A Samsonovich and B L McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci*, 17(15):5900–5920, 1997.

[212] Alexei Vladimir Samsonovich. *Attractor map theory of the hippocampal representation of space*. PhD thesis, The University of Arizona., 1997.

[213] Aldo Battista and Rémi Monasson. Capacity-resolution trade-off in the optimal learning of multiple low-dimensional manifolds by attractor neural networks. *Physical Review Letters*, 124(4):048302, 2020.

[214] Man Yi Yim, Lorenzo A Sadun, Ila R Fiete, and Thibaud Taillefumier. Place-cell capacity and volatility with grid-like inputs. *Elife*, 10:e62702, 2021.

[215] Laura Lee Colgin, Edvard I Moser, and May-Britt Moser. Understanding memory through hippocampal remapping. *Trends in neurosciences*, 31(9):469–477, 2008.

[216] Charlotte B Alme, Chenglin Miao, Karel Jezek, Alessandro Treves, Edvard I Moser, and May-Britt Moser. Place cells in the hippocampus: eleven maps for eleven rooms. *Proceedings of the National Academy of Sciences*, 111(52):18428–18435, 2014.

[217] Trygve Solstad, Edvard I Moser, and Gaute T Einevoll. From grid cells to place cells: a mathematical model. *Hippocampus*, 16(12):1026–1031, 2006.

[218] Caswell Barry, Colin Lever, Robin Hayman, Tom Hartley, Stephen Burton, John O'Keefe, Kate Jeffery, and N Burgess. The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, 17(1-2):71–98, 2006.

[219] HS Kudrimoti, CA Barnes, and BL McNaughton. Reactivation of hippocampal cell assemblies: effects of behavioral state, experience, and eeg dynamics. *J Neurosci*, 19(10):4090–101, May 15 1999.

[220] Thomas J Davidson, Fabian Kloosterman, and Matthew A Wilson. Hippocampal replay of extended experience. *Neuron*, 63(4):497–507, Aug 2009.

[221] Brad E Pfeiffer and David J Foster. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–9, May 2013.

[222] Daniel W Moran and Andrew B Schwartz. Motor cortical representation of speed and direction during reaching. *Journal of neurophysiology*, 82(5):2676–2692, 1999.

[223] Krishna V Shenoy, Maneesh Sahani, and Mark M Churchland. Cortical control of arm movements: a dynamical systems perspective. *Annual review of neuroscience*, 36:337–359, 2013.

[224] Juan A Gallego, Matthew G Perich, Stephanie N Naufel, Christian Ethier, Sara A Solla, and Lee E Miller. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature communications*, 9(1):1–13, 2018.

[225] Juan A Gallego, Matthew G Perich, Raeed H Chowdhury, Sara A Solla, and Lee E Miller. Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience*, pages 1–11, 2020.

[226] Michael Wehr and Gilles Laurent. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature*, 384(6605):162–166, 1996.

[227] Uri Rokni and Haim Sompolinsky. How the brain generates movement. *Neural computation*, 24(2):289–331, 2012.

[228] Dmitry Kobak, Wieland Brendel, Christos Constantinidis, Claudia E Feierstein, Adam Kepecs, Zachary F Mainen, Xue-Lian Qi, Ranulfo Romo, Naoshige Uchida, and Christian K Machens. Demixed principal component analysis of neural population data. *Elife*, 5:e10989, 2016.

[229] Mirko Klukas, Marcus Lewis, and Ila Fiete. Efficient and flexible representation of higher-dimensional cognitive variables with grid cells. *PLoS Comput Biol*, 16(4):e1007796, 04 2020.

[230] Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J Chadwick, Thomas Degris, Joseph Modayil, et al. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, 2018.

[231] James C R Whittington, Timothy H Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E J Behrens. The tolman-eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263.e23, Nov 2020.

[232] Honi Sanders, Matthew A Wilson, and Samuel J Gershman. Hippocampal remapping as hidden state inference. *Elife*, 9, 06 2020.

[233] Nathaniel J Killian, Michael J Jutras, and Elizabeth A Buffalo. A map of visual space in the primate entorhinal cortex. *Nature*, 491(7426):761–764, 2012.

[234] Dmitriy Aronov, Rhino Nevers, and David W Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722, 2017.

[235] Alexandra O Constantinescu, Jill X O'Reilly, and Timothy EJ Behrens. Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292):1464–1468, 2016.

[236] Elizabeth Gardner. The space of interactions in neural network models. *Journal of physics A: Mathematical and general*, 21(1):257, 1988.

[237] Robert J McEliece, Edward C Posner, Eugene R Rodemich, and Santosh S Venkatesh. The capacity of the Hopfield associative memory. *Information Theory, IEEE Transactions on*, 33(4):461–482, 1987.

[238] Daniel J Amit, Hanoch Gutfreund, and H Sompolinsky. Statistical mechanics of neural networks near saturation. *Annals of Physics*, 173(1):30–67, 1987.

[239] Yaser S Abu-Mostafa and J St Jacques. Information capacity of the hopfield model. *Information Theory, IEEE Transactions on*, 31(4):461–464, 1985.

[240] Hanne Stensola, Tor Stensola, Trygve Solstad, Kristian Frøland, May-Britt Moser, and Edvard I Moser. The entorhinal grid map is discretized. *Nature*, 492(7427):72, 2012.

[241] Christopher J Hillar and Ngoc M Tran. Robust exponential memory in hopfield networks. *The Journal of Mathematical Neuroscience*, 8(1):1, 2018.

[242] IR Fiete, DS Schwab, and Ngoc Mai Tran. A binary hopfield network with information rate and applications to grid cell decoding. In *Proceedings of the 2nd Workshop on Biological Distributed Algorithms*, 2014.

[243] Noga Mosheiff and Yoram Burak. Velocity coupling of grid cell modules enables stable embedding of a low dimensional variable in a high dimensional neural attractor. *eLife*, 8, 2019.

[244] Vincent Gripon and Claude Berrou. Sparse neural networks with large learning diversity. *IEEE transactions on neural networks*, 22(7):1087–1096, 2011.

[245] Samuel P Muscinelli, Wulfram Gerstner, and Johanni Brea. Exponentially long orbits in hopfield neural networks. *Neural computation*, 29(2):458–484, 2017.

[246] A. Mathis, A. Herz, and M. Stemmler. Optimal population codes for space: grid cells outperform place cells. *Neural Comp.*, 24:2280–2317, 2012.

[247] Rishidev Chaudhuri and Ila Fiete. Bipartite expander hopfield networks as self-decoding high-capacity error correcting codes. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32, 2019.

[248] Man Yi Yim, Lorenzo A Sadun, Ila R Fiete, and Thibaud Taillefumier. Place-cell capacity and volatility with grid-like inputs. *Elife*, 10, May 2021.

[249] H. C. LONGUET-HIGGINS. Holographic model of temporal recall. *Nature*, 217(5123):104–104, jan 1968.

[250] A Korpel. Gabor: frequency, time, and memory. *Appl Opt*, 21(20):3624–32, Oct 1982.

[251] Niru Maheswaranathan, Alex H Williams, Matthew D Golub, Surya Ganguli, and David Sussillo. Reverse engineering recurrent networks for sentiment classification reveals line attractor dynamics. *Advances in neural information processing systems*, 32:15696, 2019.

[252] I. Kanitscheider and I. R. Fiete. Training recurrent networks to generate hypotheses about how the brain solves hard navigation problems. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[253] Ingmar Kanitscheider and Ila Fiete. Emergence of dynamically reconfigurable hippocampal responses by learning to perform probabilistic spatial reasoning. *bioRxiv*, page 231159, 2017.

[254] Ben Sorscher, Gabriel Mel, Surya Ganguli, and Samuel Ocko. A unified theory for the origin of grid cells through the lens of pattern formation. In *Advances in Neural Information Processing Systems*, pages 10003–10013, 2019.

[255] Rylan Schaeffer, Mikail Khona, Leenoy Meshulam, Ila Rani Fiete, et al. Reverse-engineering recurrent neural network solutions to a hierarchical inference task for mice. *bioRxiv*, 2020.

[256] Xiao-Jing Wang. Synaptic basis of cortical persistent activity: the importance of nmda receptors to working memory. *Journal of Neuroscience*, 19(21):9587–9603, 1999.

[257] G Bard Ermentrout and Nancy Kopell. Multiple pulse interactions and averaging in systems of coupled neural oscillators. *Journal of Mathematical Biology*, 29(3):195–217, 1991.

[258] Martin Boerlin, Christian K Machens, and Sophie Denève. Predictive coding of dynamical variables in balanced spiking networks. *PLoS computational biology*, 9(11), 2013.

[259] E Paxon Frady and Friedrich T Sommer. Robust computation with rhythmic spike patterns. *Proceedings of the National Academy of Sciences*, 116(36):18050–18059, 2019.

[260] Ran Darshan and Alexander Rivkind. Learning to represent continuous variables in heterogeneous neural networks. *bioRxiv*, 2021.

[261] D. B. Arnold and D. A. Robinson. A learning network model of the neural integrator of the oculomotor system. *Biological Cybernetics*, 64(6):447–454, Apr 1991.

[262] Richard H. R. Hahnloser, H. Sebastian Seung, and Jean-Jacques Slotine. Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Computation*, 15(3):621–638, 2010/12/06 2003/03/01.

[263] H Sebastian Seung. Learning continuous attractors in recurrent networks. In *Advances in neural information processing systems*, pages 654–660, 1998.

[264] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, Nov 2013.

[265] Christopher J. Cueva and Xue-Xin Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. In *International Conference on Learning Representations*, 2018.

[266] Logan Grosenick, James H Marshel, and Karl Deisseroth. Closed-loop and activity-guided optogenetic control. *Neuron*, 86(1):106–39, Apr 2015.

[267] Peter E Latham, Sophie Deneve, and Alexandre Pouget. Optimal computation with attractor networks. *J Physiol Paris*, 97(4-6):683–694, 2003.

[268] Timothy J. Buschman Flora Bouchacourt. A flexible model of working memory. *bioRxiv doi.org/10.1101/407700*, 2018.

[269] Andrea Hasenstaub, Robert N S Sachdev, and David A McCormick. State changes rapidly modulate cortical neuronal responsiveness. *J Neurosci*, 27(36):9607–22, Sep 2007.

[270] E Aksay, R Baker, H S Seung, and D W Tank. Anatomy and discharge properties of premotor neurons in the goldfish medulla that have eye-position signals during fixations. *J Neurophysiol*, 84(2):1035–49, Aug 2000.

[271] Emile Godaux, Philippe Mettens, and Guy Chéron. Differential effect of injections of kainic acid into the prepositus and the vestibular nuclei of the cat. *The Journal of physiology*, 472(1):459–482, 1993.

[272] Guy Major, Robert Baker, Emre Aksay, H Sebastian Seung, and David W Tank. Plasticity and tuning of the time course of analog persistent firing in a neural integrator. *Proc Natl Acad Sci U S A*, 101(20):7745–50, May 2004.

[273] C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth Hayworth, Gary Huang, Kazunori Shinomiya, Jeremy Maitin-Shepard, David Ackerman, Stuart Berg, et al. A connectome of the adult drosophila central brain. *bioRxiv*, 2020.

[274] Yi Gu, Sam Lewallen, Amina A Kinkhabwala, Cristina Domnisoru, Kijung Yoon, Jeffrey L Gauthier, Ila R Fiete, and David W Tank. A map-like micro-organization of grid cells in the medial entorhinal cortex. *Cell*, 175(3):736–750.e30, Oct 2018.

[275] Alexei A Koulakov and Dmitri B Chklovskii. Orientation preference patterns in mammalian visual cortex: a wire length minimization approach. *Neuron*, 29(2):519–527, 2001.

[276] Tom J Wills, Francesca Cacucci, Neil Burgess, and John O'Keefe. Development of the hippocampal cognitive map in preweanling rats. *Science*, 328(5985):1573–1576, Jun 2010.

[277] Rosamund F Langston, James A Ainge, Jonathan J Couey, Cathrin B Canto, Tale L Bjerknes, Menno P Witter, Edvard I Moser, and May-Britt Moser. Development of the spatial representation system in the rat. *Science*, 328(5985):1576–1580, 2010.