The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep

Rishidev Chaudhuri [1,2,3,4]*, Berk Gerçek [1,5,9], Biraj Pandey[1,6,9], Adrien Peyrache [7] and Ila Fiete [1,8]*

Neural circuits construct distributed representations of key variables—external stimuli or internal constructs of quantities relevant for survival, such as an estimate of one's location in the world—as vectors of population activity. Although population activity vectors may have thousands of entries (dimensions), we consider that they trace out a low-dimensional manifold whose dimension and topology match the represented variable. This manifold perspective enables blind discovery and decoding of the represented variable using only neural population activity (without knowledge of the input, output, behavior or topography). We characterize and directly visualize manifold structure in the mammalian head direction circuit, revealing that the states form a topologically nontrivial one-dimensional ring. The ring exhibits isometry and is invariant across waking and rapid eye movement sleep. This result directly demonstrates that there are continuous attractor dynamics and enables powerful inference about mechanism. Finally, external rather than internal noise limits memory fidelity, and the manifold approach reveals new dynamical trajectories during sleep.

It has long been clear that the brain represents sensory, motor and internal variables in distributed codes across large populations of neurons. In turn, theoretical models of neural computation have emphasized that circuit dynamics must be understood in terms of the emergence of simple structures from the collective interactions of large numbers of neurons[1–5], and that robust representation and memory involve the formation of low-dimensional stable states in population dynamics (called 'attractors')[1–5].

Until recently, experimental techniques permitted access to only a few neurons at a time, but simultaneous recordings of multiple neurons are allowing the theoretically suggested approach of characterizing the structure and dynamics of neural responses at the population level. This approach has been illustrated in recent demonstrations of low-dimensional trajectories in sensory and motor circuits[6–8].

Our work proceeds from four central premises. (1) In distributed codes, information representation, computation and dynamics unfold at the level of the neural population, and the collective states across neurons of a circuit are the natural way to understand them. (2) If the primary role of a circuit is to represent a low-dimensional variable of a given dimension and topology, then, by definition, the high-dimensional states of the circuit will be localized to a low-dimensional subspace or 'manifold' of matching dimension and topology. (3) Characterizing the structure of this manifold can enable the unsupervised discovery and decoding of the internally coded (latent) variable. (4) Examining the manifold structure and dynamics on and off the manifold across a range of behavioral states as circuit inputs change can reveal inherently stable states and thus aspects of the circuit mechanism.

We illustrate a method to characterize the manifold structure of data. We use this characterization to discover—in a blind or unsupervised way—low-dimensional internal states, provide blind time-resolved decoding of these states and support the predictions of a classical mechanistic circuit model using the mammalian head-direction (HD) system as our subject. The HD system in mammals and insects[9–15] is a cognitive circuit that uses external and internal cues to estimate the direction that the animal is heading with respect to the external world. It is a proving ground for the manifold-based approach to the unsupervised discovery of encoded variables because it represents an internal cognitive state that need not directly reflect externally measured variables during waking. Moreover, this dissociation between internal and external states holds even truer during sleep (as we will see). Simultaneously, the HD system illustrates how a manifold approach can yield new insights into the structure, dynamics and mechanisms of a long-studied neural circuit, which were impossible to achieve from characterizing the responses of a few neurons at a time.

Two decades ago, theoretical models[4,16,17] of the HD circuit postulated a stable, one-dimensional (1D) ring-shaped manifold in the high-dimensional population activity state space, which is a more abstract and fundamental feature than details about shapes of tuning curves, connectivity profiles or physical placements of neurons. Stability means that perturbations in the high-dimensional space away from the ring should quickly and preferentially flow back to the ring. If the HD circuit is an integrator, then the input to the circuit describes the momentary change in state rather than explicitly specifying the new state; the circuit adds these changes to the existing state to produce the new state. This integration requires that changes in state along the ring for equivalent changes in a represented variable should be equal. HD circuit models have been extended to explain the dynamics of other neurons[5]. The same

[1]Center for Learning and Memory, The University of Texas at Austin, Austin, TX, USA. [2]Center for Neuroscience, University of California, Davis, Davis, CA, USA. [3]Department of Neurobiology, Physiology and Behavior, University of California, Davis, Davis, CA, USA. [4]Department of Mathematics, University of California, Davis, Davis, CA, USA. [5]Department of Basic Neuroscience, University of Geneva, Geneva, Switzerland. [6]Department of Applied Mathematics, University of Washington, Seattle, WA, USA. [7]Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada. [8]Department of Brain and Cognitive Sciences and McGovern Institute, MIT, Cambridge, MA, USA. [9]These authors contributed equally: Berk Gerçek and Biraj Pandey. *e-mail: rchaudhuri@ucdavis.edu; fiete@mit.edu

models can further explain how the brain could form representations in more abstract metric spaces and update them by integrating a signal that encodes changes in the representation[18]. Thus, testing whether these models correctly describe the circuit mechanism is of broad importance.

So far, pairwise correlations between mammalian HD cells[11,19–21] and the discovery of a topographically ordered physical HD-coding ring in flies[13,14] are consistent with hypothesized models. Here, we show that the long-hypothesized low-dimensional state-space ring structure and attractive dynamics can be directly visualized in the population response manifold of the mammalian HD system, whether or not the circuit possesses physical topography that reflects its connectivity and function. The dynamics revealed in this circuit during sleep states provide new evidence about the intrinsic mechanisms that allow these states to be maintained and updated.

Portions of these results have been presented at conferences[22].

Results

The instantaneous (temporally binned) response of N neurons is a point in an N-dimensional state space where each axis represents the activity of one neuron (Fig. 1a). The collection of snapshots of the population activity forms a cloud in state space. If the population encodes some variable of dimension $D_m \ll N$ and a certain topology, the point cloud should trace out a manifold of the same dimension and topology, although the shape may be convoluted. In the following sections, we describe how characterizing the topology and structure of the manifold, then analyzing dynamics on the manifold, can permit us to extract latent encoded variables in an unsupervised way and deduce key aspects of the circuit mechanism.

Spline parameterization for unsupervised decoding (SPUD). To decode the internal state encoded by the manifold, we performed the following steps (details in Methods). (1) Consider binned spiking data as points in a high-dimensional state space (Fig. 1a). (2) Determine the topology of the point cloud using persistent homology[23] (Fig. 1b,c), with the introduction of a neighborhood-thresholded topological data analysis (nt-TDA) method (Supplementary Note 2.3) for increased robustness to noisy data. (3) Estimate the intrinsic manifold dimension using various methods, including correlation dimension[24] (Supplementary Fig. 11). (4) Fit the manifold with a spline of matching topology and dimension (Fig. 1d). (5) Parameterize the spline by a smoothly changing variable of matching dimension and topology (Fig. 1e). Steps 4 and 5 yield a local, on-manifold, minimal-dimensional parameterization of even topologically nontrivial manifolds; the resulting parameterization is interpreted as the values of the encoded latent variable or internal state. (6) Given a population state at any moment, we decode that state by projecting it to the nearest point on the spline; the parameterization value at that point is the unsupervised estimate of the value of the encoded latent variable (Fig. 1f). None of these steps requires, in principle, a global low-dimensional embedding of the data.

To characterize the global topology of the manifold (step 2), we used persistent homology[23]. The method starts by blurring the point cloud of data at different resolutions or scales. Then, at each resolution, we examined the emergence of connected groups of data points called simplicial complexes (Fig. 1b,c). A simplicial complex can contain certain structures, such as a ring or a torus, and so on. Betti numbers form a list of binary structural designations that characterize the complexes (Fig. 1b). In noisy data, if a Betti number for a structure persists over many scales (Fig. 1c), this feature is robust and deemed significant. Topological data analysis uses these Betti numbers, across scales, to characterize the structure of a dataset[23].

Our method fundamentally deals with determining the existence of a nontrivial topological structure in the data manifold, then defining local on-manifold coordinate systems to parameterize it.

Conventional dimensionality reduction methods (including principal component analysis (PCA), Isomap, locally linear embedding and t-distributed stochastic neighbor embedding), by contrast, assume that the manifold is topologically trivial (equivalent to a stretched, folded or crumpled hole-free plane or solid ball of some dimension, or disjoint sets of these; Fig. 1b, first two panels), and find a low-dimensional global space or coordinate system to embed all the data. When the manifold is topologically nontrivial (Fig. 1b, third panel onwards), global dimensionality reduction methods will typically fail to correctly parameterize the latent variable represented on the manifold, thus giving a higher-dimensional embedding and parameterization than the manifold dimension. For instance, the minimum global embedding dimension for a 1D ring is two-dimensional (2D), thus global dimensionality reduction will yield, at best, a 2D parameterization of a 1D circular variable and fail to discover the real 1D latent variable (see the extended discussion and schematic in Supplementary Note 1). Constructing low-dimensional global embeddings is neither sufficient nor actually necessary for spline parameterization for unsupervised decoding (SPUD) (we used an optional initial dimensionality reduction step to reduce the complexity of subsequent operations and possibly for some smoothing of the manifold for undersampled datasets, but this step was not necessary, and applying SPUD directly in higher dimensions led to better fits when there was enough data; Supplementary Fig. 4).

Ring manifold and unsupervised decoding. We applied SPUD to activity recorded from the anterodorsal thalamic nucleus (ADn) of mice that were awake and foraging in an open 2D environment along variable paths, as well as intervening rapid eye movement (REM) and non-REM (nREM) periods[21]. The behavior, even after reduction to abstract coordinates on a 2D plane, was at least five-dimensional (location, orientation, linear speed and angular speed); the actual behaviors and inputs across sensory modalities are much higher-dimensional.

We included all recorded thalamic cells, without subselection based on tuning or other criteria, using binned spike counts throughout (~100-ms resolution). With larger, simultaneously recorded populations, it will become possible to perform higher temporal resolution decoding; doing so with 5–10 ms of precision would allow us to probe fast dynamics and resolve information that may be encoded in shorter-timescale spike patterns[25].

To determine whether the data exhibit a low-dimensional manifold structure in state space, we used both direct visualization of nonlinear low-dimensional embedding from the high-dimensional state space and topological data analysis, in particular the persistent homology of simplicial complexes[23] (topological methods are more general because they permit the characterization of topologically nontrivial and higher-dimensional manifolds, even when direct visualization is not possible). Both methods revealed that network states during waking exploration lie on a strikingly low-dimensional, albeit highly nonlinear, manifold in the form of a convoluted ring (Fig. 2a,b; see Supplementary Fig. 1 for data for all seven animals and Supplementary Video 1 for a 3D view). Moreover persistent homology revealed no evidence of a toroidal or more complex topological structure (Fig. 2b, H2 plot; contrast with Supplementary Fig. 3). The 1D structure is of much lower dimensionality than the behavior or the sensory inputs.

With the confirmation of a ring topology, we fitted a nonlinear spline with the same topology to the manifold (Fig. 2c) and isometrically parameterized the spline along its length with a circular variable α, whose values are indicated by the color of the spline (Fig. 2d). Now, α is the unsupervised or SPUD estimate of the latent variable encoded by the population manifold. Points on the manifold are colored according to the nearest value of the latent variable estimate (LVE), α (Fig. 2e).
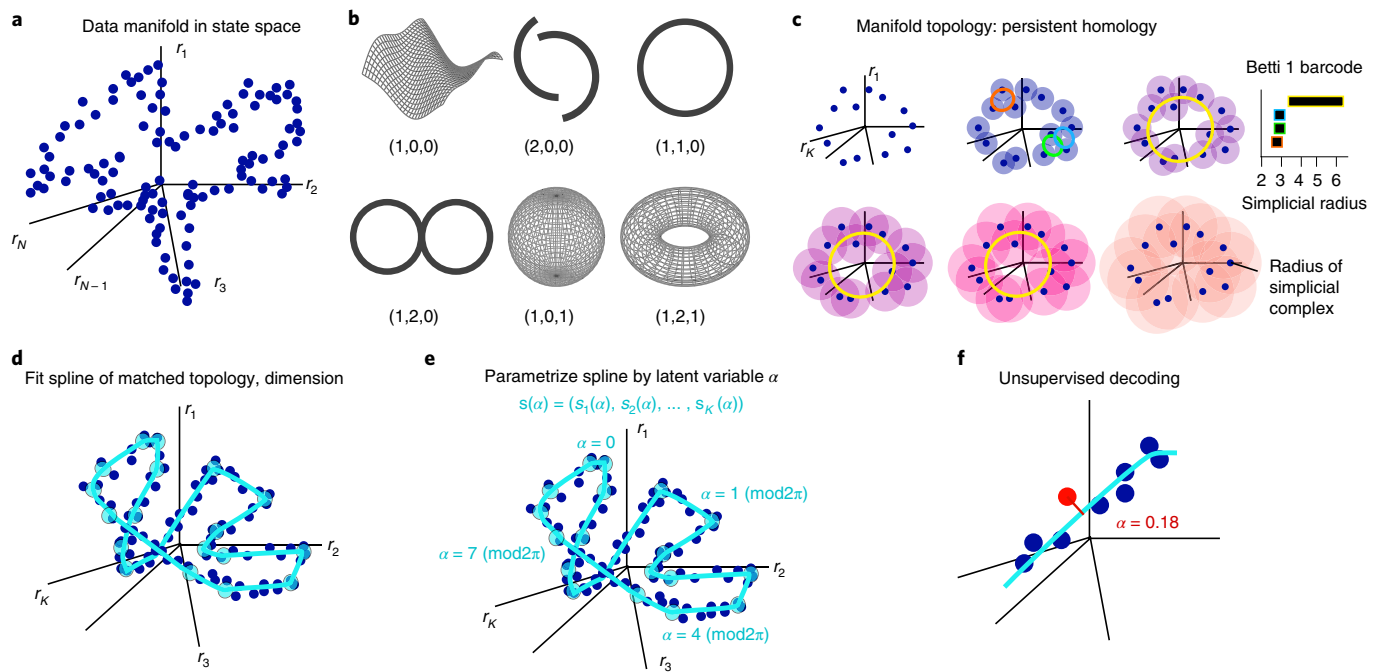
Fig. 1 | Population activity as a manifold and a method for manifold characterization. a, Each blue point is a vector representing instantaneous activity in the neural population: the vector component (r_i) is the activity of the ith neuron. The points form a manifold. **b**, Six sample manifolds, with their Betti 0, 1 and 2 numbers listed underneath. Betti 0 indicates the number of connected components that form the manifold (contrast the first and second manifolds), Betti 1 indicates the number of 1D holes in the manifold (contrast the third and fourth manifolds), and Betti 2 indicates the number of 2D holes or voids (the sphere and torus both enclose a single 2D void). **c**, Persistent homology was used to determine the topology of the manifold underlying the noisy point cloud data. Balls of different sizes centered on the data points represent different simplicial radii (r) or scales. At a given radius (scale), sets of connected points form simplicial complexes (see Supplementary Note 2.2 for details). Each complex is characterized by a set of features, with accompanying Betti numbers (illustrated in **b**). Persistent homology tracks these across scales. Depicted here are the ring features of the simplicial complexes (colored rings). The upper middle panel shows rings appearing, which then disappear (are filled in, upper right panel). The upper right panel shows that another ring (yellow) appears, which persists until the last (lower middle panel); it is deemed significant because of its persistence. The inset shows the Betti 1 barcode, or the range of radii over which each ring feature persists (different ring in each row). **d**, The data manifold is fit by a spline (cyan line) of dimension and topology as determined in **a–c**. The spline uses a few anchor points (cyan circles) determined by clustering methods, with connecting polynomial curves. **e**, The spline is parameterized by assigning coordinates along its length. The coordinates represent the values of an internal (latent) state that the circuit is assumed to encode. **f**, Moment-by-moment decoding of the internal state is done by reading out the parameterization value at the point on the spline closest to the data point (red shows the sample decoded point).

The LVE very closely matches (up to an arbitrary choice of origin and direction) the directly measured head angle (Fig. 2f,g; see Supplementary Figs. 4–6 for data for the other animals). Moreover, regressing the firing rates of individual cells onto the LVE recovered neural tuning curves in a fully blind way (Fig. 2h). The match is a direct validation of the hypothesis that the topology of neural representations should match the topology of the represented variables.

Isometric parameterization along the neural manifold produces excellent decoding. This was not clear a priori, and implies that equal amounts of neural code length or population activity variation are devoted to equal changes in the head angle. This isometry property is exactly consistent with expectations for an accurate head velocity integrator, in which coding states must be equivalently changeable so that a unit velocity input produces a unit change in the represented angle, regardless of the starting state.

The unsupervised LVE better matches an internal state estimate constructed from a supervised (tuning-curve-based) decoder (Fig. 2g; Supplementary Fig. 5) and explains more of the variance of neural spiking (cross-validated; Fig. 2i and Supplementary Fig. 6) than the measured HD. Thus, the LVE more accurately tracks the internal representation of an animal than the measured HD. The internal representation may differ from the measured HD for various reasons, including the possibility that the animal is representing

an inaccurate HD estimate, or past or future HD states, or because of errors in the experimental HD measurement.

A natural question that arises is whether the neurons encode additional undiscovered variables. The manifold is clearly primarily 1D, but we can ask whether there is additional structure, for example, in the thickness of the ring. With a finite signal-to-noise ratio (SNR) in the dataset, it is impossible to exclude structures that are significantly smaller than the noise. We thus searched for additional coding structures down to the noise floor by asking whether the data exhibit either a spread or structure that is not explained by the 1D ring structure with independent spiking noise. First, we generated synthetic data based only on tuning curves for the 1D LVE, with spikes generated using an independent point process per cell with data-matched dispersion (Fano factor). The resulting point cloud closely matched the data (Fig. 2j). Second, shared angular coding around the ring manifold accounted for 94% of the covariation between neurons. Third, the residual covariance after removing tuning to the LVE exhibited little discernible structure (Fig. 2k). By contrast, there were additional coding dimensions in postsubicular HD cells (which code for head velocity and behavioral state; data not shown) and in the ADn during nREM sleep (shown later). Larger numbers of simultaneously recorded cells will improve the SNR, thus allowing the discovery of a finer additional structure or further downgrading the possibility that it exists.
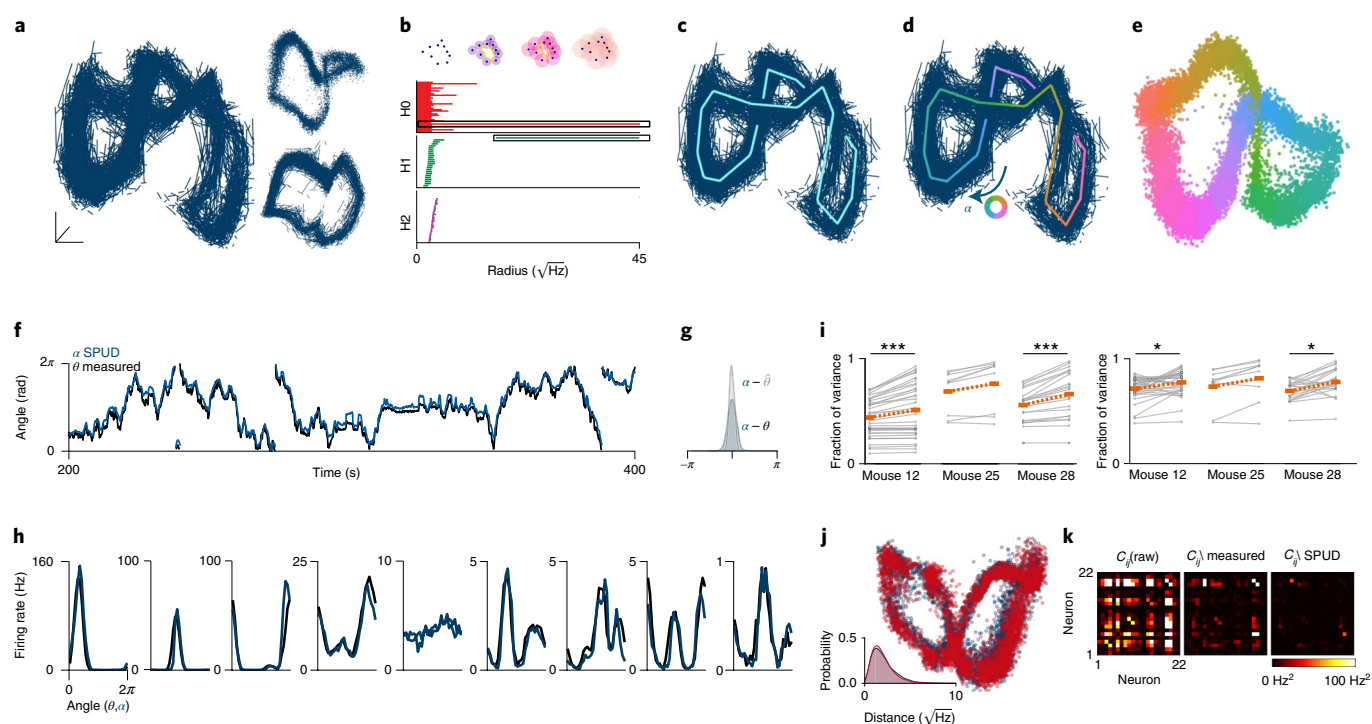
Fig. 2 | Unsupervised discovery and time-resolved decoding of encoded variables through manifold characterization. Throughout this figure, θ represents the direct measurement of the head orientation of the animal from light-emitting diode tracking, $\hat{\theta}$ represents the supervised decoder's estimate of the brain's representation (using tuning curves), and α represents the unsupervised latent variable estimate. This figure shows ADn data from a single animal for a full waking episode (31-min interval). **a**, Visualization of the manifold (using Isomap[43]), with every alternate pair of temporally adjacent points connected. The inset shows a point cloud (upper) and an alternative view of the manifold (lower). Note that the manifold does not occupy a very low-dimensional linear subspace, and PCA often fails to extract a ring structure (Supplementary Fig. 2). **b**, Betti 0, 1 and 2 barcodes. Each row shows a different feature. Simplicial complex radii are shown on the abscissa (schematic is shown at the top of the chart: complexes were constructed from data at different radii). The start and end of a horizontal bar in the middle plot signals the appearance and disappearance of some ring (a non-zero Betti 1 feature) in the data at the corresponding radii. The long bar represents a ring that appears at ~16$\sqrt{Hz}$ and persists until ~43$\sqrt{Hz}$ ($\sqrt{Hz}$ because of variance stabilization; see Methods). **c**, Spline fit to the point cloud. **d**, Parameterization of the spline by coordinate α (arbitrary origin). **e**, Coloring of neural states via the unsupervised latent variable estimate (that is, α). **f**, Comparison of α and θ. The origin and direction around the ring for the measured head angle and for unsupervised decoding, both arbitrary choices, are matched to facilitate comparison, only after unsupervised decoding is complete. **g**, Histogram of differences between α and θ and between α and $\hat{\theta}$. **h**, Fully unsupervised tuning-curve estimate (blue) versus supervised tuning-curve estimate (black). Unsupervised tuning curves capture 71±2.8% of the variance of tuning curves constructed using the traditional, supervised way (Supplementary Fig. 6). **i**, Left: fraction of variance explained by θ (left) and α (right) under a Poisson-spiking model. Means are shown in orange. Significance is calculated from two-sided binomial tests. Right: as in the left panel, but using an overdispersed model. P values left to right: $P < 10^{-4}$, $P = 0.11$, $P < 10^{-4}$, $P = 0.047$, $P = 0.11$ and $P = 0.017$ (*$P < 0.05$, ***$P < 10^{-3}$). $n = 37$, 10 and 22 cells from mouse 12, mouse 25 and mouse 28, respectively. **j**, Manifold from data (blue) and from an overdispersed spiking model (red), with overdispersion (Fano factor) estimated from the data and applied as uncorrelated across neurons. The inset shows the distribution of distances from the manifold fit for data and model. **k**, Covariance of firing rates (left) and covariance conditioned on either θ (center) or α (right). The covariance matrix is only 6% of the raw covariance matrix when conditioning on α, which suggests that α captures ~94% of data covariance; the ratio after conditioning on θ is 25%. All panels show data from mouse 28, session 140313.

In summary, down to the noise floor (SNR) of the present data, and if the recorded cells are representative samples, the population manifold reveals that the several thousand neuron-sized ADn population collectively encodes a single 1D variable, and no other, during waking.

The manifold is autonomously generated and attractive. We next show how the manifold perspective directly reveals the collective intrinsic dynamics of the circuit. These analyses test the key predictions[1,3–5] of continuous attractor models (properties 1, 3–5) and models of neural integrators for continuous variables (properties 1–6; Fig. 3a; see Supplementary Fig. 10 for a network model). (1) The high-dimensional network response occupies a low-dimensional continuum of states with a dimension and topology matching the encoded variable (or variables). (2) There is isometry of encoded state intervals so that equal velocity inputs produce equal changes in the encoded state, regardless of the starting state.

(3) States are autonomously generated and stabilized, and capable of self-sustained activation when sensory inputs are removed. (4) The manifold is an attractor, whereby states initialized away from the manifold rapidly flow back. (5) Manifold states are energetically equal, with no net flow along the manifold. (6) A velocity input, encoding the time-derivative of the represented variable, drives the circuit in a special direction in the high-dimensional state space, specifically along the low-dimensional manifold. These predictions are fundamentally applied in terms of the population manifold and hence most naturally tested at that level.

The results presented above directly support properties 1 and 2, which alone are not sufficient for establishing continuous attractor dynamics. To study autonomous dynamics, we examined the circuit during sleep, in the absence of spatial or directional input from the world.

During REM sleep, states again lie on a 1D ring (Fig. 3b; Supplementary Video 2; see also Supplementary Note 2.3 for a new
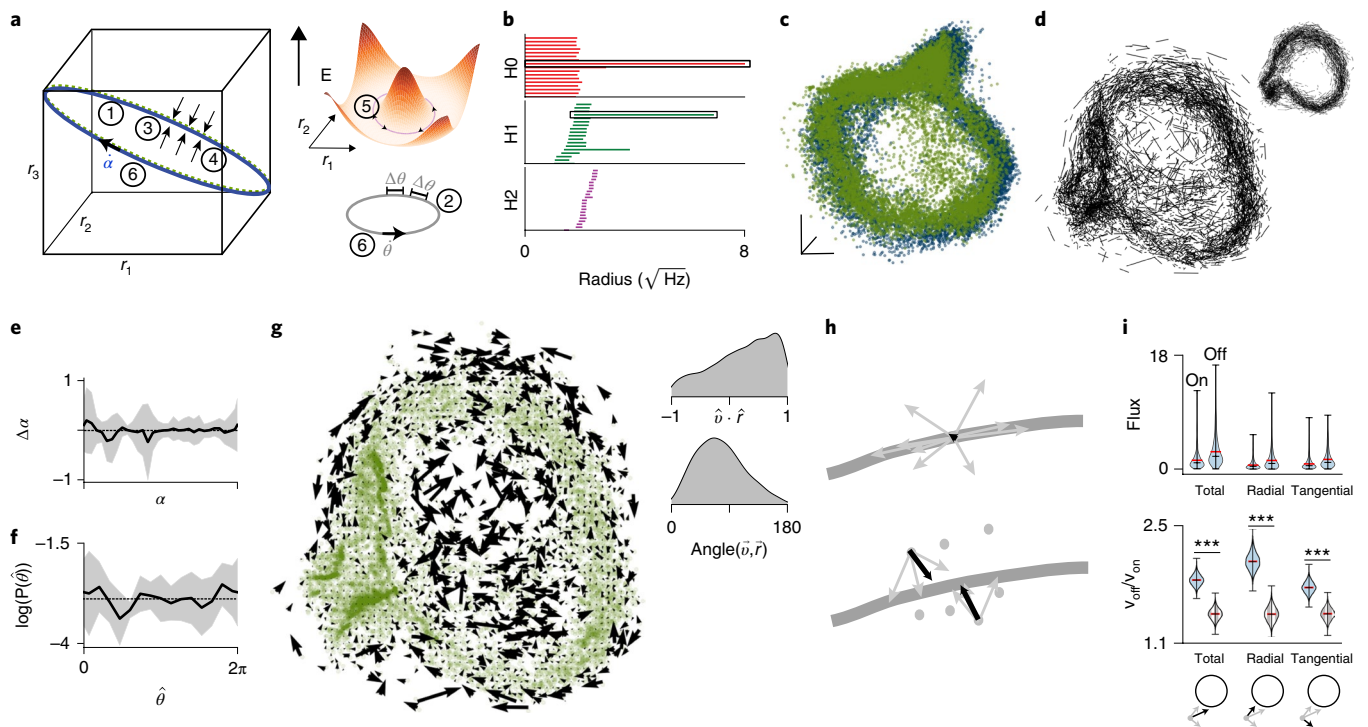
Fig. 3 | REM sleep states, fluxes and dynamics suggest that the manifold is internally generated and attractive. a, Schematic of attractor model predictions (see main text for list). **b,** Betti 1 barcode reveals a 1D ring structure preserved during REM sleep. **c,** Joint visualization of REM (green) and waking data (dark blue) using Isomap[43]. **d,** Flows along the REM manifold, with each bar representing a velocity vector at a moment in time. The inset shows the wake manifold. **e,** Single session mean and standard deviation of change in the decoded angle as a function of angle (14,679 time points). **f,** Mean and standard deviation of angle occupancy (from the tuning-curve decoder) across ten sessions. **g,** Flux on and off the manifold. The inset shows the distribution of normalized dot product (upper) and angles (lower) between flux vectors and the vector pointing toward the closest point on the manifold (that is, radial direction; see schematic at bottom of panel **i**). **h,** Schematic showing that flux should be small on the manifold because velocity vectors tend to point in both directions along the manifold and thus average out (upper panel), while flux off the manifold should be large because velocity vectors tend to drive the system back to the manifold (lower panel). **i,** Flux is larger off the manifold. Upper: each pair contrasts the distribution of fluxes on versus off the manifold for the total flux vector (left) and its radial (center) and tangential component (right). $n = 814$ flux vectors for each condition (from different spatial bins). ***$P < 10^{-3}$ for all 3 comparisons, from two-sided permutation tests. Lower: ratio of the mean flux off-manifold to on-manifold. Filled violin plots show data (bootstrapped distribution of the ratio of the means of the distributions in the upper panel), while shaded plots show shuffled controls. A total of 1,000 samples were analyzed for each distribution. In all violin plots, black lines show the minimum, median and maximum values, and red lines show the means. The existence of non-negligible tangential fluxes at off-manifold states suggests that relaxations back to the manifold may not always take the shortest path toward the manifold. Such trajectories might reflect more complex dynamics than in the simplest attractor models, an avenue for future study. All panels show data from mouse 25, session 140130.

method to enhance the SNR of persistent homology in very noisy data) that is essentially identical to the ring during awake exploration (Fig. 3c; see Supplementary Fig. 7 for data for other animals). Thus, the ring manifold is internally generated and autonomous (property 3) to the brain, which is consistent with a conclusion inferred from preserved pairwise correlations during sleep[21]. However, we cannot determine whether the internal dynamics are confined to the anterior thalamus or dependent on longer-range interactions between areas.

Manifold states are equivalent. To test the equivalence of manifold states, we examined the occupancy and dynamics during REM sleep, when state occupancy is not biased by behavior and the external world. First, we plotted instantaneous velocity vectors linking states at adjacent timepoints. If the manifold contained a number of discrete fixed points, there would be fast flow to and high occupancy around those fixed points. These flows would correspond visually to long bars converging near those points, unlike the roughly uniform bars observed (Fig. 3d).

Relatedly, the angular change is independent of the angle value itself (Fig. 3e; see Supplementary Fig. 8 for data for other animals). To gain statistical power from pooling across sessions for each animal, we decoded the angular states on the ring with a supervised

decoder and computed the density of the decoded angles (Fig. 3f). The logarithm of the density of states along the ring—an estimate of the relative energy of states—was flat on the scale of variability across sessions. These results directly support property 5.

Finally, we studied circuit dynamics by examining fluxes of states on and off the manifold. A high-dimensional state space and manifold perspective is critical to this analysis, which cannot be done on the level of single-cell tuning. The flux through a small region is the average over all trajectories that flow into and out of that region (Fig. 3g). For a continuous attractor that is not driven by directional input, we expect roughly zero net fluxes along the manifold because of the isotropic distribution of flow directions along the manifold (property 5) and the omnidirectional nature of random kicks off manifold. However, states not on the manifold should exhibit large net fluxes because of biased flows returning to the manifold (property 4; Fig. 3h). Indeed, net fluxes were larger at off-manifold states (Fig. 3g–i), with an increase in both radial (that is, toward the ring) and tangential (that is, along the ring) components (Fig. 3i). These off-manifold fluxes were directed preferentially toward the ring, even during spontaneous activity, thus showing that population states are attracted toward the manifold, as predicted by attractor models.

Diffusive dynamics along the manifold during REM. We next combined theoretical predictions about dynamical trajectories on continuous attractor manifolds[26] with SPUD decoding (Fig. 4a,b) to gain a quantitative estimate of the nature and influence of noise on the circuit. Noise is an important consideration for integrator, memory and representational circuits because it determines the timescale and fidelity of information stored in the circuit.

First, we used waking data to confirm that SPUD captured the fine-time-scale statistics of trajectories (Fig.4c, left inset). Since HD updates are correlated over short times during waking (Fig. 4d, blue trace; Supplementary Video 3), the squared deflection in estimated angle—if the SPUD estimate is sufficiently time-resolved—should grow quadratically at short times, which is what we found.

By contrast, during REM sleep, estimated angle updates were temporally uncorrelated but local (Fig. 4d, green trace; Supplementary Video 4). The squared angular deflection grew linearly with time (Fig. 4c, green curve). Temporally uncorrelated local updates and a linear growth in squared deflection are characteristic of an unbiased diffusive random walk[26], which is consistent with property 5.

Evidence of input aligned to the manifold. To resolve the nature of the noise driving diffusivity during REM, we made, to our knowledge, the first quantitative comparison between empirically observed diffusion in a neural circuit and theoretical predictions. The diffusion constant of REM dynamics shown in Fig. 4c is $1.1 \pm 0.04 \, \mathrm{rad}^2 \, \mathrm{s}^{-1}$ (0.52 ± 0.03 and 1.3 ± 0.06 for the other two animals; see Supplementary Fig. 9). This diffusivity exceeded, by 20–50 times, the predicted value in a matched neural network model[26] (Fig. 4c; Supplementary Note 4.2; Supplementary Fig. 10), if noise is independent across neurons.

Independent per-neuron noise could arise from Poisson-like spike count variations within the circuit or from high-dimensional input that projects in a spatially uncorrelated way to the neurons. In either case, high-dimensional noise is largely impotent in pushing the network state along the manifold because each unit variance of high-dimensional noise has a variance of only $1/N$ along the manifold[4,26,27] (Fig. 4e; N is the number of neurons in the circuit; over-dispersed noise does not resolve the problem, see Supplementary Note 4.2).

By contrast, a modest amount of low-dimensional noise aligned to the nonlinear manifold (standard deviation of $8.5 \, \mathrm{rad}^2 \, \mathrm{s}^{-1}$ with temporal correlations of 20 ms or less, which is comparable to the head-velocity drive during waking to update the HD estimate, Supplementary Fig. 5) has a much stronger effect[26,27], thus accounting for the measured diffusion (Fig. 4c). In contrast to high-dimensional noise, such manifold-aligned noise tends not to distort the activity states, but keeps them close to the manifold as seen in the REM data. These results suggest that the network receives an input that is aligned to the manifold and is of the right amplitude for moving the state around the ring in response to waking head movements, thus supporting property 6.

Furthermore, the results demonstrate that even in cognitive circuits for memory and integration, as established for low-level sensory circuits and sensorimotor pathways[28,29], information fidelity is primarily limited by input noise or sensory imprecision rather than by internal noise.

Higher-dimensional manifold and coherent dynamics in nREM sleep. Hippocampal circuits replay waking activity patterns during nREM sleep[30,31], and replays might be important for memory consolidation[25,32]. However, the HD circuit seems to lack replays or even coherent temporal dynamics during nREM when probed using conventional decoding approaches[21,33]. nREM sleep is also described as disrupting the ability of the brain to maintain integrated representations[34], but it is unclear what this disruption means at a more mechanistic level, thus presenting an opportunity
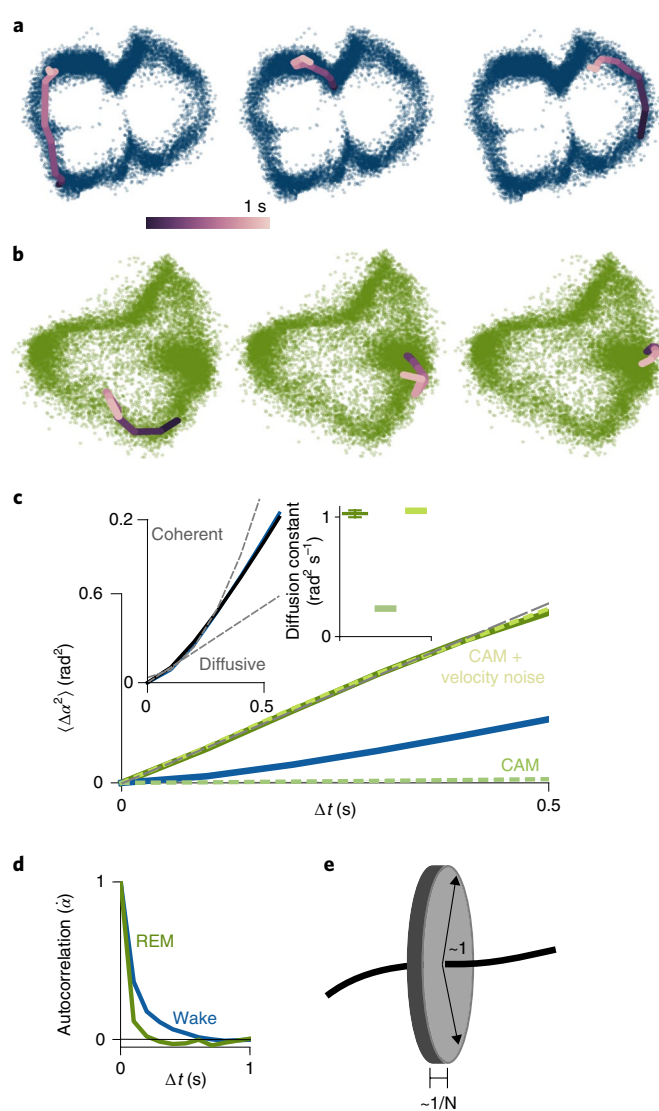

Fig. 4 | Diffusive dynamics during REM sleep. a, The manifold with sample waking trajectories at three different times. Each plot shows several consecutive points over 1s and transitions between them. **b**, As in **a**, but for sample REM trajectories. **c**, Plot of variance of REM angle update over time (unbroken green), with waking shown for comparison in blue. At short timescales, the growth of variance is linear, with the slope given by the diffusion constant. Broken green traces show the continuous attractor model (CAM) with and without noisy velocity input. The left inset shows the waking trace (blue, SPUD; black, the measured head angle) on an expanded scale, highlighting a supralinear increase at small times. Broken lines in the inset show pure diffusion and expected increase for velocity-driven dynamics. Left-most boxplot in the right inset shows the diffusion constant for the decoded angle (5,000 samples; center line, median; box limits, upper and lower quartiles; whiskers, data range). The remaining two values in the right inset show the diffusion constant for the continuous attractor model without noisy velocity input and the continuous attractor model with noisy velocity input. **d**, Autocorrelation of angular velocity. **e**, Schematic showing that in high dimensions, independent noise is almost entirely directed off-manifold and does not move the system much along the manifold. All panels show data from mouse 28, session 140313.

to understand it in the context of a specific integrated representation like HD.

We therefore examined manifold structure and dynamics in the ADn during nREM sleep. We found that the manifold is

Fig. 5 | Higher-dimensional states and coherent dynamics during nREM sleep. a, Absence of a persistent ring in the Betti 1 barcode during nREM sleep (that is, there is no long horizontal line; compare Figs. 2 and 3b). **b**, Joint plot of nREM and waking manifolds using Isomap. nREM points are shown in mustard yellow, while waking points are shown for comparison in dark blue (as before). The inset shows alternative views. **c**, Population firing rate during nREM decoded from the distance to the centroid (actual in light blue, decoded in yellow). The inset shows points colored by the population firing rate. **d**, nREM manifold with three sample trajectories. **e**, Autocorrelation of velocity on the full manifold. Waking, REM traces from Fig. 4d shown for comparison. The timescale over which each type of nREM trajectory persists before switching between types (300 ms) is longer than timescales present in waking (96 ms, blue trace) and REM dynamics (38 ms, green trace), producing the fat tails in the temporal autocorrelation. The inset shows the fraction of time spent in consecutive low-velocity and high-velocity epochs (300 ms of duration each). Box plots show shuffled control (1,000 shuffled samples; center line, median; box limits, upper and lower quartiles; whiskers, data range). ***$P < 10^{-3}$, two-sided permutation test. **f**, Squared change in position on the full manifold for low-velocity and high-velocity epochs. Quadratic and linear fits are shown by broken lines. The inset shows the distribution of angles between successive high (dark) and low (light) velocities on full manifold (100-ms separation). High-velocity epochs are coherent, producing initial quadratic behavior. **g**, Average LFP trace conditioned on small (black) versus large (gray) change in position, along with 95% confidence intervals ($n = 7,450$ samples across time). **h**, Correlation of the total change in position against LFP power in 1-s bins, along with 95% confidence intervals ($n = 500$ bootstrapped samples). All panels show data from mouse 28, session 140313.

higher-dimensional (Fig. 5a,b; Supplementary Video 5; Supplementary Fig. 11), forming a conical surface (Fig. 5b and Supplementary Fig. 11, where the cone is clearer). The manifold only partially overlapped the waking/REM manifold (Fig. 5b), which caps the circular rim of the nREM cone.

The nREM manifold encodes at least two latent variables, which we decoded along the tangential (circular) and radial (distance along spokes emanating from the manifold centroid) dimensions of the manifold using SPUD. The LVE along the tangential direction represents an angular variable, which closely matched (except at low-activity states where the SNR of the neural response is low) the estimates of two wake-trained supervised decoders of head angle that make different assumptions (Supplementary Figs. 12 and 13). This result means that the angular structure of the manifold still represents the HD variable.

The radial LVE encodes population firing rate, capturing the slow, global fluctuations that characterize nREM sleep[35] (Fig. 5c). Unlike in waking and REM, where we inferred that some discrete attractor state in the total drive to the ADn (which will be interesting to identify in future experiments) keeps the manifold radius invariant across divergent behavioral states, this discrete attractor appeared to be lost during nREM, so that the network receives an overall drive of slowly varying amplitude. The highest drive during nREM matched wake and REM, but reduced drive caused the system to visit angle coding states or rings of varying amplitude, down to nearly the zero activity state close to the manifold centroid or cone tip. The nREM responses were well modeled by the same attractor circuit as for waking and REM dynamics, but with the global external input to all neurons undergoing large suppressive amplitude fluctuations of multiplicative amplitude ≤1 (Supplementary Fig. 14).

Dynamics on the higher-dimensional nREM manifold are of two distinct types (Fig. 5d): local diffusive evolution in a confined region of the manifold and larger coherent sweeps (see Methods and Supplementary Video 6 for more detail on the properties of these two types of trajectories). This result is in contrast to the largely
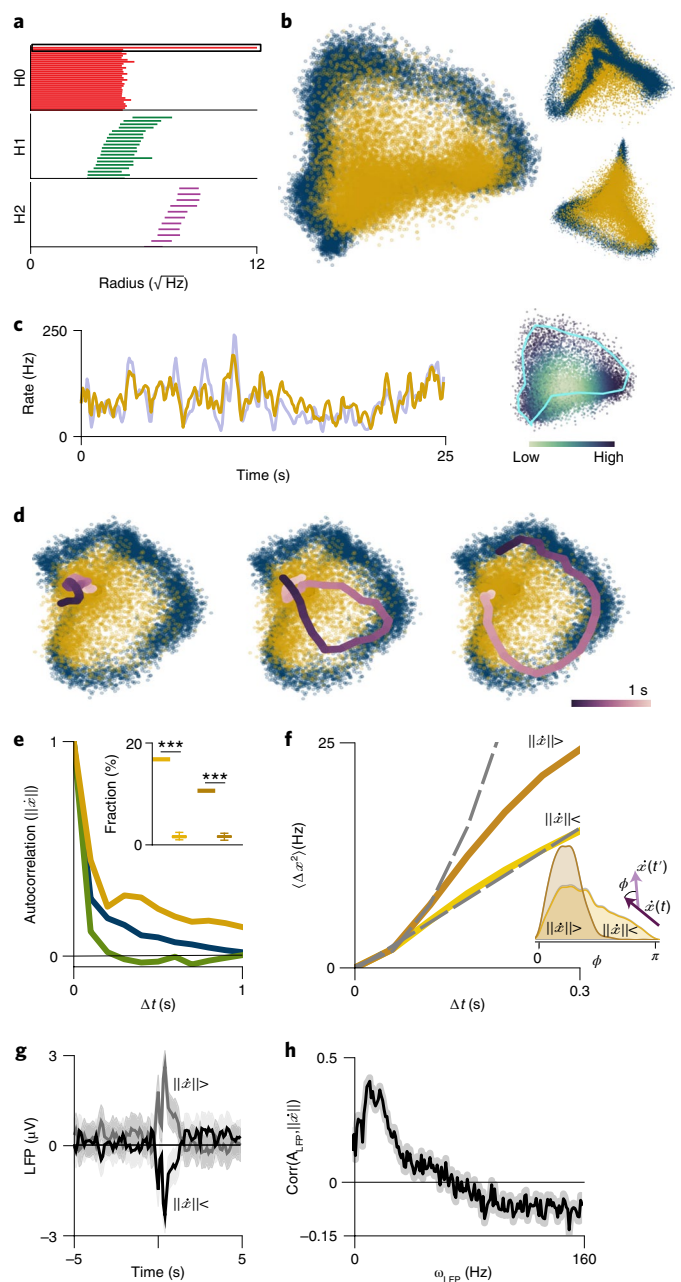
(rapidly) diffusive dynamics identified in nREM by wake-trained supervised decoders[21,36,37] (Supplementary Fig. 12d–f), which effectively project the higher-dimensional manifold states onto the 1D waking ring before estimating temporal dynamics.

The large sweeps are coherent in magnitude (Fig. 5e) and in direction so that motion in a direction tends to continue in that direction (Fig. 5f, inset), thus producing a quadratic (rather than linear diffusive) growth in squared displacement over time (Fig. 5f, dark curve) as seen during waking (Fig. 4c, blue curve); however, the inferred speed of nREM coherent trajectories is eight times the speed of waking trajectories.

To reproduce sweeps in the attractor circuit model requires not only slowly modulating the strength of global inputs to generate matched population firing rate fluctuations but also adding temporally correlated fluctuations (correlation time of 200 ms) projected through the low-dimensional velocity input to the circuit (Supplementary Fig. 14).

The sweeps occur during transient increases in local field potential (LFP) amplitude in the ADn (Fig. 5g), specifically during

upward fluctuations in LFP power in the ~ 12 Hz band, within the 7–15 Hz range for sleep spindles[35] (Fig. 5h), which in turn are correlated with the occurrence of hippocampal sharp waves[38].

To summarize, while nREM dynamics projected to the waking manifold show misleadingly little temporal structure, the dynamics on the nREM manifold switch between a diffusive low-velocity regime and a coherent high-velocity regime. Coherent sweeps are temporally coincident with LFP power in the spindle band and thus with hippocampal sharp waves. Moreover, in continuous attractor models of the circuit, sweeps must be driven by temporally coherent velocity inputs to the circuit. These observations suggest that there are possible connections with replay and memory consolidation events elsewhere in the brain[25,30–32].

Discussion

We obtained a direct glimpse of a clear 1D ring in the activity states of the vertebrate HD circuit, in which neurons may or may not be physically laid out in order of their activity profiles. This visualization provides a compelling mammalian parallel to the beautiful recent results on a topographically ordered HD ring in the invertebrate nervous system[13]. In addition, a purely 1D structure in the state-space manifold would directly imply that the circuit encodes no other information beyond HD, while a physical ring layout of a HD circuit does not rule out the possibility of additional coding dimensions.

A manifold approach reveals the full *N*-point correlations of the circuit, directly uncovering a structure that cannot easily be obtained from individual neural responses and pairwise correlations, including that there is little structure beyond a single 1D ring (down to some noise threshold) in waking and REM sleep. It shows that a presumed discrete attractor holds the circuit's response amplitude fixed across waking and REM, and the constraint of this discrete attractor on response amplitude is lifted in nREM. Thus, our results augment an elegant body of work that inferred intrinsic low-dimensional structure from pairwise correlations in the vertebrate circuits for HD[9,11,19–21], oculomotor control[3], prefrontal evidence accumulation[39] and 2D spatial navigation[36,37,40].

By examining dynamics on and off the manifold across waking and sleep, we showed that the manifold is generated autonomously in the brain and that population dynamics are attractive. In particular, attractive flows onto the manifold from perturbed states are evident at the manifold level, but would be hard to observe from a few neurons at a time. Finally, a manifold approach allows for comparison with theoretical models, whose key predictions are at the level of structured population dynamics.

Unlike many applications of manifold methods[7,8], the animal was not constrained to a specific low-dimensional task or trajectory whose dimensionality determined the manifold dimension. Rather, the manifold during waking behavior (which included variations in location, linear speed, angular speed and orientation, for a minimum of five dimensions of variation) was of much lower dimension than the behavior. Moreover the manifold retained the low-dimensional structure during sleep states, which are not constrained by low-dimensional inputs from the world, and when the circuit likely receives high-dimensional fluctuations. Thus, these analyses have uncovered the intrinsic dimension of the manifold.

With the demonstration of a purely 1D representation in a neural circuit (down to the noise floor), we are now free to wonder, without the usual caveat that these neurons might also be representing other things, why the brain uses thousands of cells for this purpose.

The manifold approach could be particularly useful for discovering unknown variables encoded in high-level brain areas and examining how structured states and dynamics emerge in neural circuits[41], including through development[42], plasticity or learning. Unlike supervised decoding methods[25], we did not force an interpretation of a new ensemble of states (for example, during a different behavioral or brain state or task) by regression onto a previously characterized library of states. Manifold characterization with topological data analysis followed by parameterization in local coordinates as done here rather than by global low-dimensional embedding (as done by, for example, PCA and Isomap[43]) will be important for unsupervised decoding of higher-dimensional manifolds that are topologically nontrivial, for example, toroidal structures produced by simulated grid cells (Supplementary Note 3; we found that ~35 grid cells can be sufficient to reveal 2D toroidal structure, Supplementary Fig. 3).

Persistent homology has been used to analyze neural data in various contexts[44–47], including to recover the topology of the physical environment explored by an animal from place cell activity[45], and to determine the topology of neural activity space[46] or of unknown covariates[47], but not, until now and in a contemporaneous work[48], as part of a procedure for decoding latent variables.

Our innovation is to parameterize the manifold with splines of matching topology and to use this parameterization to decode the represented latent variable, characterize on- and off-manifold dynamics in waking and sleep, and to test theoretical models. By contrast, Rybakken et al.[48] performed blind decoding using a variant of persistent homology called persistent cohomology, which provides a mapping between a Betti 1 feature and a circle, and applied an interesting iterative procedure to find additional coding dimensions (likely reflecting non-thalamic cells excluded from our present analysis and in which we also find additional structure (R.C., B.G., B.P., A.P. and I.F., unpublished observations)). Together, these studies illustrate the broader idea of unsupervised decoding from topologically nontrivial population manifolds.

Persistent homology has several limitations. One is the high sensitivity to outliers, for which we proposed a density-threshold method (Supplementary Note 2.3) that ameliorates this problem. Second, computing persistent features can be computationally slow. Finally, persistent homology cannot distinguish between topologically trivial manifolds with different geometries (for example, a hyperplane versus a filled ball). In general, persistent homology should be used as an initial step in a workflow to detect or rule out nontrivial topological features, followed by use of a manifold parameterization method to suit the topology (see Supplementary Note 1) and geometry.

In summary, manifold-level analyses can enable fully unsupervised discovery and decoding of brain states and dynamics, and quantification of collective dynamics on and off the manifold can give insight into circuit mechanisms. We believe that a manifold perspective and related techniques[41,48–50] will be essential for extracting information from large datasets and during cognitively interesting tasks in which the brain constructs rich latent variables, thus representing the future of neural decoding.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0460-x.

References

1. Amari, S.-I. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**, 77–87 (1977).
2. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
3. Seung, H. S. How the brain keeps the eyes still. *Proc. Natl Acad. Sci. USA* **93**, 13339–13344 (1996).
4. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **15**, 2112–2126 (1996).

5. Burak, Y. & Fiete, I. R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).

6. Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48**, 661–673 (2005).

7. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).

8. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural manifolds for the control of movement. *Neuron* **94**, 978–984 (2017).

9. Ranck, J. B. in *Electrical Activity of Archicortex* (eds Buzsaki, G. & Vanderwolf, C.) 217–220 (Akademiai Kiado, 1985).

10. Taube, J. S., Muller, R. U. & Ranck, J. B. Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).

11. Taube, J. S., Muller, R. U. & Ranck, J. B. Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *J. Neurosci.* **10**, 436–447 (1990).

12. Finkelstein, A. et al. Three-dimensional head-direction coding in the bat brain. *Nature* **517**, 159–164 (2015).

13. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).

14. Green, J. et al. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **546**, 101–106 (2017).

15. Kim, S. S., Rouault, H., Druckmann, S. & Jayaraman, V. Ring attractor dynamics in the *Drosophila* central brain. *Science* **356**, 849–853 (2017).

16. Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. A model of the neural basis of the rat's sense of direction. *Adv. Neural Inf. Process. Syst.* **7**, 173–180 (1995).

17. Sharp, P. E., Blair, H. T. & Brown, M. Neural network modeling of the hippocampal formation spatial signals and their possible role in navigation: a modular approach. *Hippocampus* **6**, 720–734 (1996).

18. Aronov, D., Nevers, R. & Tank, D. W. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature* **543**, 719–722 (2017).

19. Mizumori, S. & Williams, J. Directionally selective mnemonic properties of neurons in the lateral dorsal nucleus of the thalamus of rats. *J. Neurosci.* **13**, 4015–4028 (1993).

20. Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells. *J. Neurophysiol.* **80**, 425–446 (1998).

21. Peyrache, A., Lacroix, M. M., Petersen, P. C. & Buzsáki, G. Internally organized mechanisms of the head direction sense. *Nat. Neurosci.* **18**, 569–575 (2015).

22. Chaudhuri, R., Gercek, B., Pandey, B. & Fiete, I. Unsupervised latent variable extraction from neural data to characterize processing across states. In *Computational and Systems Neuroscience (CoSyNe)* I-56 http://cosyne.org/cosyne17/Cosyne2017_program_book.pdf (2017).

23. Ghrist, R. Barcodes: the persistent topology of data. *Bull. Am. Math. Soc.* **45**, 61–75 (2008).

24. Grassberger, P. & Procaccia, I. Measuring the strangeness of strange attractors. *Phys. D.* **9**, 189–208 (1983).

25. Chen, Z. & Wilson, M. A. Deciphering neural codes of memory during sleep. *Trends Neurosci.* **40**, 260–275 (2017).

26. Burak, Y. & Fiete, I. R. Fundamental limits on persistent activity in networks of noisy neurons. *Proc. Natl Acad. Sci. USA* **109**, 17645–17650 (2012).

27. Moreno-Bote, R. et al. Information-limiting correlations. *Nat. Neurosci.* **17**, 1410–1417 (2014).

28. Bialek, W. Physical limits to sensation and perception. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 455–478 (1987).

29. Osborne, L. C., Lisberger, S. G. & Bialek, W. A sensory source for motor variation. *Nature* **437**, 412–416 (2005).

30. Pavlides, C. & Winson, J. Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *J. Neurosci.* **9**, 2907–2918 (1989).

31. Lee, A. K. & Wilson, M. A. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002).

32. Diekelmann, S. & Born, J. The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010).

33. Brandon, M. P., Bogaard, A. R., Andrews, C. M. & Hasselmo, M. E. Head direction cells in the postsubiculum do not show replay of prior waking sequences during sleep. *Hippocampus* **22**, 604–618 (2012).

34. Massimini, M. et al. Breakdown of cortical effective connectivity during sleep. *Science* **309**, 2228–2232 (2005).

35. Steriade, M., McCormick, D. A. & Sejnowski, T. J. Thalamocortical oscillations in the sleeping and aroused brain. *Science* **262**, 679–685 (1993).

36. Gardner, R. J., Lu, L., Wernle, T., Moser, M.-B. & Moser, E. I. Correlation structure of grid cells is preserved during sleep. *Nat. Neurosci.* **22**, 598–608 (2019).

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

38. Siapas, A. G. & Wilson, M. A. Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* **21**, 1123–1128 (1998).

39. Wimmer, K., Nykamp, D. Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).

40. Yoon, K. et al. Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat. Neurosci.* **16**, 1077–1084 (2013).

41. Low, R. J., Lewallen, S., Aronov, D., Nevers, R. & Tank, D. W. Probing variability in a cognitive map using manifold inference from neural dynamics. Preprint at *biorXiv* https://www.biorxiv.org/content/10.1101/418939v2 (2018).

42. Bassett, J. P., Wills, T. J. & Cacucci, F. Self-organised attractor dynamics in the developing head direction circuit. *Curr. Biol.* **28**, 609–615 (2018).

43. Tenenbaum, J. B., De Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000).

44. Curto, C. & Itskov, V. Cell groups reveal structure of stimulus space. *PLoS Comput. Biol.* **4**, e1000205 (2008).

45. Dabaghian, Y., Mémoli, F., Frank, L. & Carlsson, G. A topological paradigm for hippocampal spatial map formation using persistent homology. *PLoS Comput. Biol.* **8**, e1002581 (2012).

46. Singh, G. et al. Topological analysis of population activity in visual cortex. *J. Vis.* **8**, 1–18 (2008).

47. Spreemann, G., Dunn, B., Botnan, M. B. & Baas, N. A. Using persistent homology to reveal hidden information in neural data. Preprint at *arXiv* https://arxiv.org/abs/1510.06629 (2015).

48. Rybakken, E., Baas, N. & Dunn, B. Decoding of neural data using cohomological feature extraction. *Neural Comput.* **31**, 68–93 (2019).

49. Park, M. et al. Bayesian manifold learning: the locally linear latent variable model. *Adv. Neural Inf. Process. Syst.* **28**, 154–162 (2015).

50. Rubin, A. et al. Revealing neural correlates of behavior without behavioral measurements. Preprint at *biorXiv* https://www.biorxiv.org/content/10.1101/540195v1 (2019).

Author contributions

R.C. and I.F. conceived the decoding method and analyses. R.C., B.G. and B.P. implemented the analyses. A.P. collected the original data and advised on the data and analyses. R.C. and I.F. wrote the paper with input from the other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0460-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.C. or I.F.

Peer review information: *Nature Neuroscience* thanks Vivek Jayaraman, Kate Jeffery, Sung Soo Kim, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Data. We analyzed data from a previous experiment[21]. All experiments were approved by the Institutional Animal Care and Use Committee of New York University Medical Center. Briefly, the dataset contains recordings from the ADn of seven C57BL/6 mice (five male, two female), aged between 3 and 6 months, that were awake and foraging in an open environment along variable paths with variable velocities, as well as during intervening REM and nREM periods, along with measured head angles[21]. For some of the mice, the data also contain recordings from the postsubiculum. Including data from the postsubiculum allows for slightly better manifold decoding of waking HD in some animals. However, this is a separate brain area with a different manifold structure, so we leave those data for a separate study.

We show manifolds (across waking, REM and nREM states) and wake decoding results from all seven mice. We show further analyses of waking dynamics and waking and REM decoding results from the three mice with waking root-mean-square decoding error of <0.5 rad and nREM decoding results from two out of these mice (the manifold for the third mouse was confined to very low activity states; Supplementary Fig. 11). Sessions that allowed good decoding contained 9–50 neurons, and the number of neurons that showed good HD tuning ranged from 8 to 30.

Preprocessing. We first converted spike times into time-varying rates. For analyses except persistent homology, we estimated firing rates by convolving the spike times with a Gaussian kernel of standard deviation 100 ms (50 ms for the plots in Fig. 5f). For the persistent homology analyses, we computed total spike counts in 1-s bins. In all cases, we then replaced the rates by their square root to stabilize the variance.

Before fitting the manifold or applying topological methods, we used Isomap[43] to reduce the large (N-dimensional) ambient dimension by re-embedding the data into a smaller, but still relatively high-dimensional, embedding space of dimension D_e ($D_m \ll D_e \ll N$, where D_m is the intrinsic manifold dimension). We set the number of neighbors to be 5 (higher values also work well) and embedded into 3–20 dimensions (3 for visualization and before decoding; 10 before applying the topological methods below; a range between 3 and 20 for Supplementary Fig. 4). For the joint visualizations of data across states (Figs. 3c and 5d), we concatenated equal amounts of data from the two states and ran Isomap on these combined data.

This preliminary embedding ironed out some of the convolutions in the manifold while preserving its topology and, most importantly, sped up spline fitting. However, the results for decoding are not sensitive to the choice of embedding dimension. Indeed, for two of the three animals for which good decoding was possible, fitting directly in the high-dimensional space yielded a slightly more accurate fit, although the fitting procedure occasionally failed in high dimensions (Supplementary Fig. 4).

Persistent homology. To compute Betti barcodes for the data, we first applied Isomap to reduce it to ten dimensions (or used the full-dimensional state space if the number of neurons was fewer than ten). We then used the package Ripser[51] to generate the Betti 0, 1 and 2 barcodes. For the nt-TDA analysis, we first excluded outliers by considering a neighborhood around each point with the radius defined by the first percentile of the pairwise distance distribution, and then removing all points whose numbers of neighbors lay in the bottom 20th percentile of the distribution of number of neighbors across points (Supplementary Note 2.3).

We plotted features above the 98th, 97th and 30th percentiles of Betti 0, 1 and 2 length distributions, respectively, for Fig. 2. For Fig. 3, the thresholds were 99th, 98th and 90th percentiles, respectively. For Fig. 5, the thresholds were 99.85th, 99.85th and 98th percentiles, respectively. For Supplementary Figs. 1c,d, 7b,c and 11b,c, the thresholds were common across mice and reported in the figure captions. Note that there are a large number of very short-lived features, and no results were sensitive to the threshold for displaying the Betti features.

Spline fit, parameterization and decoding. We fit the manifolds using piecewise linear curves. A curve $L(y)$ is specified by K knots, with locations $\{y_1 \cdots y_K\}$. The knots are ordered, and the ith segment of the curve is a straight line between the ith and $i+$ th knot. Given data points x_i and a number of knots K, we first used k-means to identify K clusters in the data and set the centers of these clusters to be the initial knot locations. We then iteratively updated these knot locations to minimize $\left(\sum_i ||x_i - L(y)||\right)|L(y)|$, where $||x_i - L(y)||$ is the Euclidean distance between the ith data point and the nearest point on the curve $L(y)$, and $|L(y)|$ is the length of the curve. The multiplication by $|L(y)|$ acts as a regularizer that penalizes excessively long or convoluted curves. The improvement from regularization is mild. An alternative cost function of the form $\left(\sum_i ||x_i - L(y)||\right) + \lambda|L(y)|$ (where λ controls the degree of additive regularization) also worked well.

We parameterized points on the manifold by distance along the curve (in embedding space) from some arbitrary origin, with distances rescaled between 0 and 2π for comparison to the actual head angle. We primarily used $K=12$ and embedding dimension $D_e=3$ (see Supplementary Fig. 4 for other values). Points were decoded by mapping them to the nearest point on the manifold, based on the Euclidean norm in the embedding space, and reading off the parameter value there.

In the waking state, we shifted the global origin and chose the orientation around the curve to match the measured head angle, but made no other

modifications (for example, we did not rescale the coordinate differently in different parts of the ring). During sleep, when comparing to a tuning curve decoder, we performed a similar shift and choice of orientation dictated by the tuning-curve decoded angle.

Supervised tuning-curve decoder. We computed the supervised tuning curve of the ith cell, $f_i(\theta)$, as its mean response to the measured head angle for each of the 30 angular bins as follows:

$$f_i(\theta) = \frac{\text{Number of spikes fired by cell } i \text{ around angle } \theta}{\text{Time spent by animal around angle } \theta}.$$

We then decoded the head angle using maximum likelihood estimation under the model that at angle θ, neuron i responds independently with C_i spikes drawn from a Poisson distribution with rate $f_i(\theta)$ as follows:

$$\hat{\theta}_t = \underset{\theta}{\text{argmax}} P(\{C_{it}\}_{i=1,\cdots,N}|\theta) = \underset{\theta}{\text{argmax}} \prod_{i=1}^{N} \text{Poiss}(C_{it}; f_i(\theta)\Delta t)$$

Variance explained and excluding other encoded variables. To compute the variance explained, we considered the spike counts extracted in 100-ms bins. If the spike counts of the ith neuron are C_i, then the variance explained by X is as follows:

$$\text{Var}_{\text{exp},X} = \text{Var}[\mathbb{E}(C_i|X)] + \mathbb{E}[\phi\mathbb{E}(C_i|X)] \qquad (1)$$

Here, X is the measured head angle or decoded head angles (binned in 30 bins between 0 and 2π). For the Poisson model, $\phi = 1$. For the overdispersed model, we estimated ϕ as $\min_X \text{Var}(C_i|X)/\mathbb{E}(C_i|X)$. For a true overdispersed process, taking the minimum is likely to underestimate the overdispersion; thus this estimate is conservative.

Throughout (except for the measured head angle), we used a training set (80% of the data) to fit the manifold or to construct tuning curves, and a test set (remaining 20% of the data) to evaluate the model. As shown in Fig. 2i, we evaluated significance by computing the number of cells that were better explained by the unsupervised LVE than by the measured angle and comparing this to a null model in which both explained the data equally well (that is, two-sided binomial test).

For Fig. 2j, we generated synthetic data using the tuning curves to the unsupervised LVE. Given a decoded latent variable α, we generated a spike count for neuron i from a normal distribution with mean $\mathbb{E}(C_i|\alpha)$ and variance $\text{Var}(C_i|\alpha)$. Generating counts this way assumes that neural firing is overdispersed, but that neurons are independent given α, thus explicitly removing additional structure in the population. To compare the covariance explained (Fig. 2k), we compared the ratio of the Frobenius norm of the residual covariance matrix (that is, after conditioning on either the measured or the SPUD angle) to the norm of the raw covariance matrix.

Diffusion curves. The diffusion curve at time shift τ is $D(\tau) = \langle \alpha(t+\tau) - \alpha(t) \rangle_t$, where the average value is taken over time (that is, all pairs of time points separated by τ). To compute diffusion constants, we fitted a straight line to the first 200 ms of the squared change in decoded angle against time. To obtain a bootstrapped estimate of error, we resampled 200-ms epochs from the data with replacement (number of samples chosen to match the length of data) and recomputed the diffusion constant. We repeated this resampling procedure 1,000 times.

Fluxes. To estimate the flow fields (Fig. 3g), we considered a 2D Isomap embedding of the manifold and averaged together velocity vectors of points in each of the 900 spatial bins (30 equal bins per dimension).

For the distributions shown in Fig. 3g, inset, and Fig. 3i, we carried out the same analysis in 3D (to limit distortions from the low-dimensional embedding) using 20 bins per dimension. We defined a radial vector from the center of each bin to the nearest point on the manifold and projected onto this vector to construct radial and tangential flux components. For Fig. 3i, upper panel, we rescaled the tangential components by $1/\sqrt{2}$ to reflect that there are two tangential dimensions and only one radial dimension (the ratio plots in the lower panel of Fig. 3i and the significance tests are unaffected by this rescaling).

To quantify the difference between on- and off-manifold points (Fig. 3i), we divided the bins into on- and off-manifold based on the 50th percentile of the distance to the fitted spline, and show respective flux distributions. We then compared the ratio of the average norm of the velocity vector on and off the manifold along with the equivalents for the radial and tangential components. The distributions of these ratios were computed by resampling the data with replacement 1,000 times.

To test for significance, we shuffled the assignment of velocity vectors to points, repeated the spatial binning and averaging, and recomputed the ratios (1,000 times). We then computed a P value by calculating the probability of getting a value as or more extreme than the observed data under the null distribution (two-sided permutation test).

nREM dynamics. For the SPUD-based firing rate decoding shown in Fig. 5c, we computed the distance of points on the nREM manifold to the manifold centroid, and plotted the best linear predictor of population firing rate as a function of this distance.

To classify trajectories on the full manifold as sweeps, we looked for 300-ms epochs (here, six 50-ms time bins; results were similar for three 100-ms time bins) when the speed is above the 60th percentile of the speed distribution. For confined trajectories, we did the same but extracted trajectories that remained below the 40th percentile of speed. To compute a null distribution for the fraction of times during which the circuit is in a sweep or confined trajectory, we repeatedly shuffled the velocities over time (1,000 times) and recomputed the fractions. We computed the significance from these null distributions using a two-sided permutation test.

We estimated LFP at a shank as the median of the LFP recorded on each channel, and then averaged these estimates across shanks (results were similar across shanks). For Fig. 5g, we considered changes in the nREM manifold position over 200 ms, and plotted the mean LFP for the 5 s before and after a large (>50th percentile) or small (<50th percentile) change, along with a 95% confidence interval computed as 1.96 times the standard deviation across time samples. For Fig. 5h, we converted the LFP to a spectrogram using a sliding Fourier transform, calculated the total power at each frequency in 1-s windows and correlated this with the summed absolute change in manifold position over ten 100-ms bins (that is, $\sum_{i=1}^{10} \|x(t_0 + 0.1 \times i) - x(t_0 + 0.1 \times (i-1))\|$, where t_0 is the time at which the signals are being compared). We plotted these correlations along with a 95% bootstrapped confidence interval, where we repeatedly resampled the LFP–change in manifold position pairs and recomputed the correlation (1,000 times).

Attractor model. We used a slightly modified version of the continuous attractor model from a previous study[26] and reproduced the waking, REM and nREM data by only changing the inputs to the model. Further details on model construction are provided in Supplementary Note 4 and in Supplementary Figs. 10 and 14.

Statistics. No statistical methods were used to predetermine sample sizes. We analyzed data from all animals reported in a previous publication[21], and the number of animals and recorded cells were similar to previous studies[33,52]. There was no randomization or division into experimental groups. Data collection and analyses were not performed blinded to the conditions of the experiments. For significance, nonparametric permutation tests or binomial tests were used. All statistical tests used were two-sided, and data distributions were not assumed to be normal. Correlations are reported using Pearson's correlation coefficient. Further details are available in the Nature Research Reporting Summary.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Data have been previously reported[21] and are available on the CRCNS website at http://crcns.org/data-sets/thalamus/th-1.

Code availability
The code is available at https://fietelab.mit.edu/code/.

References
51. Bauer, U., Tralie, C. & Saul, N. Ripser. *Github* https://github.com/ctralie/ripser (2017).
52. Knierim, J. J., Kudrimoti, H. S. & McNaughton, B. L. Place cells, head direction cells, and the learning of landmark stability. *J. Neurosci.* **15**, 1648–1659 (1995).

natureresearch

Corresponding author(s):   Ila Fiete, Rishidev Chaudhuri

Last updated by author(s):   Jun 5, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data in the original study was collected using the Amplipex recording software. Spike sorting was performed using KlustaKwik and Klusters.
Data analysis	Data was analyzed using scripts written in Python 2.7, along with NumPy and Scipy. Plots were made using Matplotlib. Analysis code is publicly available on the group repository.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data have been previously reported (Peyrache, 2015) and are available on CRCNS: http://crcns.org/data-sets/thalamus/th-1. Code is available at https://fietelab.mit.edu/code/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined by the study of Peyrache et al. (2015) and is similar to previous studies. We analyzed all 7 animals from this study.
Data exclusions	We show topological analysis and manifolds from all 7 animals across all 3 behavioral states (waking, REM, nREM). We also performed and show waking decoding on all animals. We show waking dynamics and REM dynamics results from the 3 animals where there were enough cells for good waking decoding (RMSE <0.5 rad) and non-REM decoding/dynamics for 2 of these animals (manifold was confined to low activity states for third animal).
Replication	Manifolds are reproducible across all 7 animals and 3 states (see Supplementary Figures). Decoding/dynamics results reproduced for animals for which good decoding was possible (i.e., enough cells), and are shown in Supplementary Figures.
Randomization	No group allocation.
Blinding	No group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	7 C57/BL6 mice (5 male, 2 female), aged between 3 and 6 months.
Wild animals	No wild animals.
Field-collected samples	No field-collected samples.
Ethics oversight	All experiments were approved by the Institutional Animal Care and Use Committee of New York University Medical Center.

Note that full information on the approval of the study protocol must also be provided in the manuscript.