Functional connectivity models for brain state identification: application to decoding of spatial representations from hippocampal CA1 and CA3 recordings

L. Posani[1], S. Cocco[1], K. Jezek[2], and R. Monasson[3]

[1] *Laboratoire de Physique Statistique, Ecole Normale Supérieure and CNRS UMR 8550, PSL Research, Paris Sorbonne UPMC, 24 rue Lhomond, 75005 Paris, France*

[2] *Laboratory of Experimental Neurophysiology, Biomedical Center, Charles University in Prague, Faculty of Medicine in Pilsen, alej Svobody 1655/76, 32300 Pilsen, Czech Republic*

[3] *Laboratoire de Physique Théorique, Ecole Normale Supérieure and CNRS UMR 8549, PSL Research, Paris Sorbonne UPMC, 24 rue Lhomond, 75005 Paris, France*

Abstract

Hippocampus can store spatial representations, or maps, which are recalled each time a subject is placed in the corresponding environment. We consider the problem of decoding the recalled maps as a function of time from multi-cellular recordings. We introduce a functional-connectivity-based decoder, which accounts for the pairwise correlations between the spiking activities of neurons in each map and does not require any positional information, i.e. any knowledge about place fields. We first show, on recordings of hippocampal activity in constant environmental conditions, that our decoder efficiently decodes maps in CA3 and outperfoms existing methods in CA1, where maps are much less orthogonal. Our decoder is then applied to data from teleportation experiments, in which instantaneous switches between environmental conditions trigger the recall of the corresponding maps. We test the sensitivity of our approach on the transition dynamics between the respective memory states (maps). We find that the rate of spontaneous state shifts (flickering) after a teleportation event is increased not only within the first few seconds as already reported, but the network also shows a higher instability level for on much longer (> 1 min) intervals, both in CA3 and in CA1. In addition, we introduce an efficient Bayesian decoder of the rat full trajectory over time, and find that the animal location can be accurately predicted at all times, even during flickering events. Precise information about the animal position is thus always present in the neural activity, irrespectively of the dynamical shifts in the recalled maps.

Significance Statement Hippocampus can store spatial representations (maps), which are retrieved everytime an animal is placed in the corresponding environment. Decoding maps from the instantaneous activity of recorded cells is crucial to track the fast hippocampal dynamics e.g. in response to changes in sensory inputs, but is difficult, especially in CA1 where maps are much less orthogonal than in CA3. We show that functional-connectivity based decoders, which exploit only the pairwise correlations in the recorded cell activities, are accurate both in CA3 and CA1. Our decoder allows us to investigate map retrieval in experiments, where environmental cues are suddenly switched, and could be, more generally, applied to identify brain states in other areas even in the absence of sensory correlates.

1 Introduction

Over the recent decades, multi-cell recording techniques have provided insights into the nature of brain representations and their internal dynamics. While many works have focused on the input-output transfer functions in primary sensory systems (visual, olfactory, etc.), understanding functions corresponding to complex representations in higher cortical circuits is very hard as they are often based on mixed selectivities [1]. In relatively rare cases, such as in the entorhino-hippocampal system, a highly processed neural activity can be reliably correlated with behavior. The so-called 'place cells' in the CA1 and CA3 of hippocampus exhibit sharp spatially tuned and environment specific activity [2], see Fig. 1. Collective activity of the place–cell population coding for the environment defines its neural representation, or *map*. Simultaneous recording of multiple place-cell activity thus allows one to identify a general memory state of the network (specific map), as well as to decode the accurate position of the rodent in the corresponding environment [3].

Recently, Jezek et al. [4] have studied the dynamics of transient change between the spatial maps encoding two different environments in CA3 at high temporal resolution (ca 120 ms time windows). The two environments differed by light cues that could be switched instantaneously ('teleportation procedure'), while the animal hippocampal neural activity was recorded to monitor the course of activation of the proper spatial map. An unstable state generally emerged for some seconds after the light switch, as both maps started to flicker back and forth. This phenomenon, called flickering, was identified through measure of the similarity between the place-cell population activity and its averaged patterns across both environments, recorded earlier in respective reference sessions. Typically, a given 120 ms time window activity of the test data strongly correlated with the average reference activity in one map, and had essentially no correlation with the reference activity in the other map.

Success of such comparison-based decoding methods reflects the strong orthogonality of spatial maps in CA3: across two environments, activity of place cells broadly differ in their mean frequencies and receptive field locations, see Fig. 1(a). Hence, simple map decoders, essentially assuming that cells fire independently of each other, are sufficient to reliably identify the representation expressed by the animal. In contrast, remapping between environments (especially of similar geometry) is less orthogonal in CA1 as it shows higher number of cells firing at corresponding places across rooms, see Fig. 1(b). The population activity vector often correlates well with both concurrent reference templates, which hinders the use of comparison-based methods for map decoding.

Here we address the challenging goal of map decoding in CA1 by introducing a probabilistic graphical model for the neural activity configurations in each map. Graphical models, in which a functional connectivity network accounts for the pairwise correlation structure between neuronal firing events in the recorded population [5], have been applied to various areas so far [6], e.g. to estimate the information conveyed by [7] or the activity of [8,9] retinal ganglion cells in the presence of visual stimuli, to detect learning-related changes in functional connectivity in the prefrontal cortex [10].

We apply our graphical-model decoder to new recordings of the hippocampal

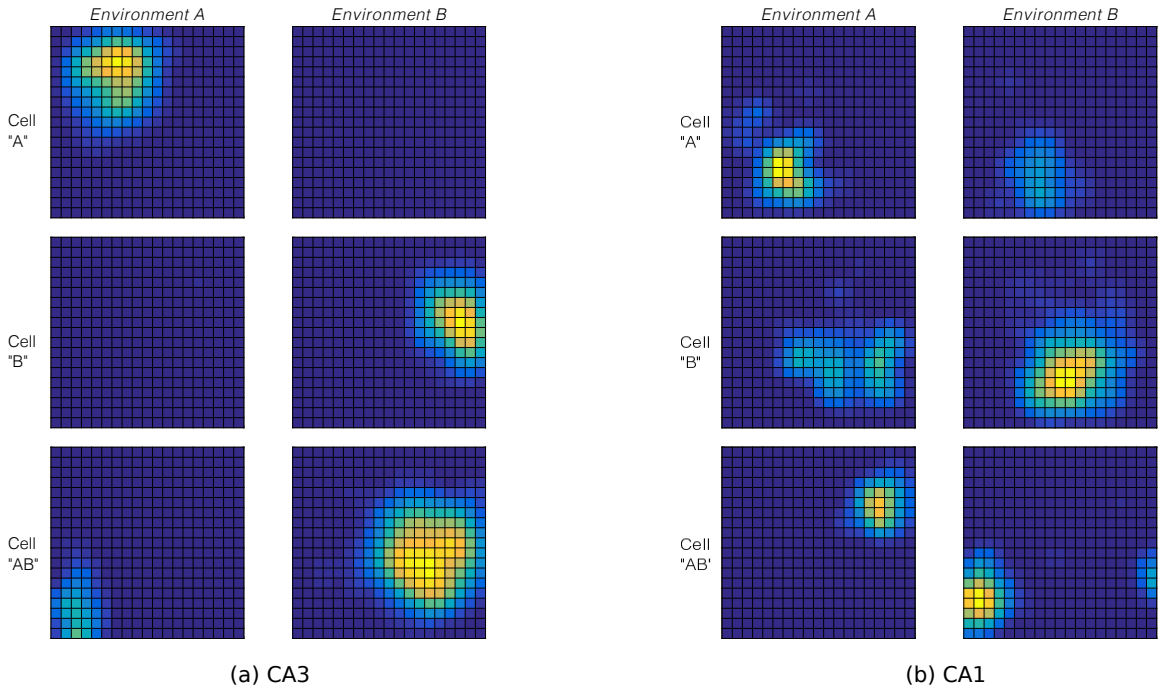


Figure 1: **Hippocampal representations are more orthogonal in CA3 (a) than in CA1 (b).** Each panel shows three rate maps from CA3 (a) and CA1 (b) corresponding to three place cells in the recorded neuronal population, computed from 10 min recordings of the activity during free exploration of environments A and B (same shapes). Whereas CA3 coding is highly sparse and representations are largely orthogonal, CA1 population shows higher amount of cells active in corresponding locations across the two rooms, with peak rates (color scale) changing from one environment to the other. The orthogonality of environment representations in CA3 makes identification of the represented map from neural activity easy compared to the situation in CA1.

activity in CA1, performed within the teleportation setup of Jezek et al. [4]. Our decoder shows very good performances in terms of precision and statistical properties in CA1, and consistently finds back the results for the CA3 recordings already analyzed with simpler approaches [4]. It allows us, in particular, to identify transitions between spatial representations in CA1 (and in CA3) in a statistically robust way. Remarkably, we find that the frequency of these flickering events is increased even minutes after a teleportation switch, both in CA3 and in CA1. In contradistinction with previously used map decoders, ours does not use any position information. It can therefore be applied to decode and study the dynamics of general brain states with unknown input correlates. The only working hypothesis is that we dispose of reference sessions to build statistical models of the corresponding internal states.

In addition, we introduce an efficient decoder of the rat position which maximizes the Bayesian posterior over the full trajectory. We find that the position of the animal is consistently encoded by the neural representation at all times, either during or outside the flickering events identified with our map decoder.

2 Materials and Methods

2.1 Experimental methods

Electrode preparation and surgery. Single unit neuronal activity was recorded in adult Long Evans male rats in hippocampal subfields CA3 and CA1. Rats were implanted with a "hyperdrive" allowing for an independent positioning of 16 tetrodes organized into an ellipsoid bundle. Tetrodes were twisted from 17 um insulated platinum-iridium wire (90% and 10%, respectively, California Fine Wire Company). Impedance of electrode tips was adjusted by platinum plating to 120 – 250 kOhm (at 1 kHz). Anesthesia was introduced by placing the rat into a plexiglass chambre with seal top filled with isoflurane vapour. Then the animnal was shaved and placed into the stereotaxic frame and continued the isoflurane delivery with a face mask. Breathing, heart action and reflexes were monitored continuously. Hyperdrive was then implanted above the right dorsal hippocampus at coordinates AP 3.8 mm and ML 3.2 mm relative to bregma. Stainless steel screws and dental acrylic were used to stabilize the implant on the skull. Two of the screews served as the hyperdrive ground.

Tetrode position. The tetrodes were slowly approached towards CA1 and CA3 within 2-3 weeks after the surgery while the rat was resting in a comfortable pot on a pedestal. To maintain stable recordings, electrodes were not moved at all before and during the experiment on a given day. The recording reference electrode was positioned in corpus callosum. Additional reference for EEG was placed in stratum lacunosum moleculare.

Recording procedures. Neural activity was recorded while the rat was behaving in an apparatus described by Jezek et al. [4]. Signal was recorded differentially against the reference tetrode. Hyperdrive was connected to a multichannel, impedance matching, unity gain headstage and its output conducted through a 82-channel commutator to a Neuralynx digital 64 channel data acquisition system. Signal was band-pass filtered at 600 Hz–6 kHz. Unit waveforms above individually set thresholds (45-70 uV) were time-stamped and digitized at 32 kHz. Position of the light emitting diodes on the headstage was tracked at 50 Hz to assess the animal's position. For the purpose of this study only data from intervals when the rat's movement speed exceeded 5 cm/sec were used. Broadband EEG from each tetrode was recorded continuously at 2000 Hz.

Spike sorting and cell classification. Spikes were sorted manually using 3D graphical cluster-cutting software (SpikeSort, Neuralynx) The feature space consisted of three-dimensional projections of multidimensional waveform amplitudes and energies. Autocorrelation and crosscorrelation functions were used as additional separation tools. Putative pyramidal cells were distinguished from putative interneurons by average rate, spike width and occasional complex spikes.

Histology. After the experiment was finished, the rat was overdosed with Equithesin and was perfused intracardially with saline followed by 4 % formaldehyde. Brain coronal sections (30 μm) were stained with cresyl violet. Traces of all 14 tetrode locations were identified. Each tip location was considered as the place

in the section before the tissue damage became negligible. Only recordings from tetrodes with their tips in CA3 and CA1 were used in this study.

Behavioral procedure. Animals were first pre-trained according to the procedure described in [4]. Briefly speaking, the apparatus consisted of two identical black plastic boxes (60 × 60 cm, 50 cm in height). The two environments differed only by sets of light cues, one placed on the upper rim of the box, the second was positioned under the semi-transparent floor with an additional cue on one wall, respectively. There were no other visual cues present as the experiment was otherwise carried in darkness provided by surrounding light-proof curtains. The training consisted of four phases. Initially, the two boxes were connected with an alley so the rat could freely explore both of them within three 20 min. sessions for 3 days. In the second phase, after the first 20 min. session, the alley was removed and the animal was placed into box A or B, respectively, in a quasi-random manner so that it received two 10 min. sessions in each of them, respectively. The next day the rat received two 10 min. sessions in each environment as the day before. Then we removed the double maze and replaced it with a single box equipped with both sets of lights that was presented at the original locations with just one cue set switched on at the given session. The rat was given another two 10 min. sessions in each environment that day. Finally, the next day, after two sessions in the original locations, the box was presented in a central location. Again, the animal was presented another two 10 min sessions in each environment, respectively, in a quasi-random order. In all stages, the running sessions were separated by a 20 min. break in the resting pot. On the test day, both environments were presented in two "reference" recording sessions (10 min each). After a 20 minutes break, the test session begun. The animal was inserted to the box with one set of lights on, and the lights were switched between the both sets either one or multiple times in a row.

2.2 Data structure

Cross validation of environment decoding methods. For the validation of environment decoding methods (Section 3.1) a total amount of four recordings were used for each brain region (CA1 and CA3). Two of these sessions, one in the environment A and one in the environment B, were used to infer activity models and reference statistics. The other two (again one in environment A and one in environment B) have then been used to test inferred models in the task of inferring which environment is internally-represented by the rodent.

Teleportation sessions. In the post-teleportation analysis shown in section 3.2 we used recordings from two different experiments. In CA1 the light switch is performed once, and the activity is recorded for some minutes before and after the teleportation. CA1 data comes from two recording sessions, each of which includes a reference session for both environments and two teleportation sessions, one from A to B and one from B to A. The CA3 data (same as published in [4]) contains multiple switches performed in a time-sequence, and short-term effects become more evident. It includes in total two reference recordings and one multiple teleportation session.

2.3 Map decoding methods

We consider two classes of decoders: *Rate-map based decoders*, which expressely use the knowledge of place fields and the rat trajectory as an input, and *Activity-only decoders* that do not rely on any information about the correspondence between position and neural firing. Throughout this section neural activities are binned with time resolution Δt; we define the number of spikes of neuron i in time bin t, $n_{i,t}$, and the binary activity, $s_{i,t} = \min(n_{i,t}, 1)$. Little information is lost when considering s instead of n as long as Δt is smaller than the typical inter-spike interval of the cells.

2.3.1 Activity-only decoders

Bayesian approach to map decoding. We introduce probabilistic models for the distribution of activities $\{s_i\}$ in a time bin, $P(\{s_i\}, \Theta)$. Those models are parametrized by a set of variables, Θ, which are fit to maximize the likelihood of the data in reference sessions. Two sets of parameters $\Theta^{(m)}$ are fitted, one for each reference session $m = A, B$. We then define the difference in log–likelihood, for each time bin t,

$$\mathcal{E}(\{s_{i,t}\}) = \log P(\{s_{i,t}\}|\Theta^{(A)}) - \log P(\{s_{i,t}\}|\Theta^{(B)}) \ . \tag{1}$$

The sign of the quantity may be used to decode the map at time t. Significance levels, based on the percentiles of the distribution of $\mathcal{E}$ can be imposed, see Results, Section 3.2.

Independent-cell model. We here assume that the neural activities s_i are independent from cell to cell. The probability distribution P is parametrized by a set $\Theta = \{h_i\}$ of N 'inputs' h_i: $P(\{s_i\}|\Theta) = \prod_i e^{h_i s_{i,t}}/(1 + e^{h_i})$. h_i is fit such that the average value of s_i with P matches $\mu_i^{(m)}$, the mean value of $s_{i,t}$ across the time bins t in reference session relative to map m, yielding $h_i^{(m)} = \log\left(\mu_i^{(m)}/(1-\mu_i^{(m)})\right)$.

Graphical Ising model. We now take into account pairwise correlations between neural activities s_i in a time bin. We introduce couplings J_{ij} to express the conditional probability that cell i is active given the activity of cell j. The probability distribution P is now parametrized by the set $\Theta = \{h_i, J_{ij}\}$ of N inputs h_i and $\frac{1}{2}N(N-1)$ couplings J_{ij}:

$$P\big(\{s_i\}|\Theta\big) = \frac{1}{\mathcal{Z}[\{h_i, J_{ij}\}]} \exp\left(\sum_i h_i\, s_i + \sum_{i<j} J_{ij}\, s_i\, s_j\right) \tag{2}$$

where $\mathcal{Z}$ is the normalization constant. Parameters $h_i^{(m)}$ and $J_{ij}^{(m)}$ are computed such that the average values of s_i and $s_i s_j$ with P match the mean values of, respectively, $s_{i,t}$ and $s_{i,t} s_{j,t}$ across the time bins t in reference session relative to map m. This hard computational problem can be approximately solved with the Adaptive Cluster Expansion (ACE) algorithm [11, 12], which provides estimates of $\{h_i^{(m)}, J_{ij}^{(m)}\}$ and $\mathcal{Z}^{(m)}$ in Eq. (2).

2.3.2 Rate-map based decoders

Computation of rate maps. The squared box is partitioned into a 20×20 grid of 3×3 cm^2 bins, and the rat position during the two reference sessions is discretized with respect to this grid. The coordinates (x_t, y_t) associated to time bin t correspond to the first spatial bin visited by the rat in the time interval $[t - \Delta t; t]$. We define the average firing rate $r_i^{(m)}(x, y)$ as the total number of spikes emitted by neuron i in the reference session m when the rat is at position (x, y), divided by the total time $T^{(m)}(x, y)$ spent by the animal in this spatial bin. These rate maps are then smoothed to fill missing bins through discrete cosine transform [13].

Pearson decoder. The observed firing pattern $\{n_{i,t}\}$ is compared to the average firing rates $\left\{r_i^{(m)}(x_t, y_t)\right\}$ in the position occupied by the animal at the time the neural activity is recorded [4]. This comparison is made through the Pearson correlation

$$\mathcal{C}^{(m)}(\{n_{i,t}\}) = \frac{\langle n\, r^{(m)}(x_t, y_t)\rangle_t - \langle n\rangle_t \, \langle r^{(m)}(x_t, y_t)\rangle_t}{\sqrt{\left(\langle n^2\rangle_t - \langle n\rangle_t^2\right)\left(\langle r^{(m)}(x_t, y_t)^2\rangle_t - \langle r^{(m)}(x_t, y_t)\rangle_t^2\right)}} \,, \tag{3}$$

where the notation $\langle f\rangle_t := \frac{1}{N}\sum_{i=1}^{N} f_{i,t}$ denotes the average of the quantity $f_{i,t}$ over the N neurons i in time bin t. The decoding of the map is done according to the sign of $\mathcal{E}(\{n_{i,t}\}) = \mathcal{C}^{m=A}(\{n_{i,t}\}) - \mathcal{C}^{(m=B)}(\{n_{i,t}\})$.

Dot-product decoder. The second method used in [4] compares directly the activity to the firing rates at the rat position. The decoding of the map m is done according to the sign of $\mathcal{E}(\{n_{i,t}\}) = \langle n\, r^{(m=A)}(x_t, y_t)\rangle_t - \langle n\, r^{(m=B)}(x_t, y_t)\rangle_t$.

Bayesian Poisson rate model. This model assumes that each neuron fires independently following a Poisson statistics, with a position-dependent firing rate $r_i^{(m)}(x, y)$ in map m. The likelihood of the number of spikes $\{n_i\}$ emitted by the cells in a time bin when the rat is at position (x, y) reads

$$P^{(m)}\big(\{n_i\}|\ (x, y)\big) = \prod_i \frac{\left(r_i^m(x, y)\,\Delta t\right)^{n_i}}{n_i!}\, e^{-r_i^{(m)}(x,y)\,\Delta t} \tag{4}$$

The prior probability over positions is $P^{(m)}(x, y) = T^{(m)}(x, y)/\sum_{x',y'} T^{(m)}(x', y')$, and we assume that both maps m are *a priori* equally likely. We obtain the likelihood of the activity conditioned to the map m by marginalizing over positions, $P(\{n_i\}|m) = \sum_{x,y} P\big(\{n_i\}|(x, y)\big) \times P^{(m)}(x, y)$. We then define the difference in log-likelihoods, $\mathcal{E}(\{n_i\}) = \log P(\{n_i\}|m = A) - \log P(\{n_i\}|m = B)$.

2.4 Performance measure of a binary decoder

To quantitatively assess decoding performance of map-decoding methods we refer to binary classifier theory [14–17].

Receiver Operating Characteristic (ROC) diagrams. A standard framework to assess the performance of binary decoders is the so-called ROC diagram [14]. For each time bin t the decoder outputs either map A or map B. To match the vocables used in the ROC framework we will arbitrarily say that the output is

positive if the map is decoded to be A, and *negative* if the map is predicted to be B. If the output of the decoder matches the environment defined by the light cues at the same time t, the prediction is said to be *True*, otherwise it is said to be *False*. For instance, a time bin such that the decoder predicts A, in agreement with the cues, corresponds to a True Positive event. The 2×2 possible events are shown in Table 1. Two important quantities are: the True Positive Rate (TPR, also called Recall), that is, the number of true positive predictions divided by the total number of positive events, and the False Positive Rate (FPR), that is, the number of false positive predictions, divided by the total number of negative events. In other words, the TPR measures the fraction of time bins with A–cues that are correctly decoded as A, while the FPR is the fraction of time bins with B–cues that are incorrectly predicted to be A.

decoder output	A	B	A	B
cue	A	A	B	B
denomination	True Positive	False Negative	False Positive	True Negative

Table 1: Denomination used for the four possible events, depending on the output of the decoder and on the environment-defining cue.

Our binary decoders are all based on thresholding the *estimator* variable $\mathcal{E}$. Within the Bayesian framework, for instance, we compute $\mathcal{E}$ as the difference between the logarithms of the posterior probabilities of A and B, and output Positive if the difference is larger than $\theta = 0$, Negative otherwise. The value of the significance threshold θ can be arbitrarily changed, with the consequence of modifying the TPR and FPR values. A ROC curve shows the parametric plot of TPR vs. FPR as the threshold varies, and describes a curve in the unit square, see Results, Section 3.1. The two extreme points of the ROC curves have coordinates (0,0), and (1,1); (0,0) is obtained for a very large significance threshold θ, the decoder never outputs Positive and both TPR and FPR vanish; (1,1) is obtained when the significance threshold is very low, the decoder always outputs Positive and both TPR and FPR are equal to unity. Very good decoders are such that the TPR is close to unit, while maintaining a very low value for the FPR. A random-guessing decoder would give equal values for the TPR and FPR, and the ROC curve would coincide with the diagonal of the unit square.

A complementary measure of decoding performances is the Precision versus Recall (or TPR) curve, obtained by scanning the values of the significance threshold θ, see Results, Section 3.1. The Precision is defined as the number of true positive events, divided by the total number of positive predictions. When lowering the significance threshold the Precision decreases from 1 to 0, while the Recall increases from 0 to 1.

Area Under the Curve (AUC). A quantitative measure of the decoding performances is the Area Under the (ROC) Curve [14]. According to this measure, the ideal decoder has AUC = 1, while random guessing would give AUC = 0.5. Note that this measure is invariant with respect to the arbitrary choice of assigning *positive* value to environment A: if we assign positive to B and negative to A instead of the previous choice, ROC curves will undergo a symmetry transformation

with respect to the top-left/bottom-right diagonal, resulting in an identical area under the curve. This is granted by the fact that positive and negative values are mutually exclusive and complementarily cover the whole data set: for each Θ value the fraction of False Positive (B decoded as A) equals one minus the fraction of True Negative (B decoded as B) events.

2.5 Continuity prior for map decoding

A continuity prior can be included in map inference in order to reduce noise in the decoding and highlight clusters of contiguous transited time bins. To do so, we consider the output $\{\mathcal{E}_t\}$ of the map decoder (see Section 2.3); for Bayesian decoders $\mathcal{E}_t$ is the difference between the log-likelihoods of the two maps $m_t = +1$ and -1 in time bin t. We then introduce a prior, controlled by a strength parameter K, which favors persistence between decoded maps in nearby time bins. Informally speaking, K is the cost (in log-likelihood) we are willing to pay for flipping the map index in time bin t predicted by the sign of $\mathcal{E}_t$ to its opposite value, if it then matches the map indices of the neighboring time bins, $t-1$ or $t+1$. The prior may thus be effective in changing the map prediction m_t if the differences between $\mathcal{E}_{t-1}$, $\mathcal{E}_t$, $\mathcal{E}_{t+1}$, ... are of the order of K (in absolute value). Two situations are encountered: (1) for some decoders, *e.g.* Pearson, $\mathcal{E}_t$ takes value in $[-2; 2]$, and the variations of $\mathcal{E}$ over successive time bins is bounded; (2) for other decoders, *e.g.* Independent-Cell, Poisson and Ising, the difference between $\mathcal{E}_t$ and $\mathcal{E}_{t+1}$ can take arbitrarily large values and show wide fluctuations as t varies across the recording. In the latter case, a uniform prior K is unadequate in large portions of the recording. To circumvent this difficuly we introduce a scale factor $\beta < 1$, and multiply all outcomes $\mathcal{E}_t$ by this factor. As a result we get a smoother time course of $\mathcal{E}_t$ over the time index t, on which a uniform prior can now be applied.

The joint probability of the time sequence of map predictions $\{m_t\}$ reads

$$P(m_1, m_2, \ldots, m_T) = \frac{1}{\mathcal{Z}} \exp\left(\frac{\beta}{2}\sum_{t=1}^{T} \mathcal{E}_t\, m_t + K \sum_{t=1}^{T-1} m_t\, m_{t+1}\right) \tag{5}$$

where $\mathcal{Z}$ is a normalization coefficient. To decode the map in time bin t we compute the marginal probability P_t over m_t from the joint distribution P. Exploiting the analogy with the one-dimensional model of statistical physics, this computation can be done with the transfer matrix method, also called dynamic programming in a time scaling linearly with the total number of time bins. Then the outcome of our combined decoder+prior is

$$\mathcal{E}_t^{\text{decoder+prior}} = \frac{1}{\beta}\left(\log P_t(m_t = +1) - \log P_t(m_t = -1)\right) \tag{6}$$

The presences of the $\frac{1}{2}$ and $\frac{1}{\beta}$ factors in, respectively, Eq. (5) and Eq. (6) ensure that, for $K = 0$, $\mathcal{E}_t^{\text{decoder+prior}}$ and $\mathcal{E}_t$ coincide. In practice we choose $\beta = \frac{1}{|\mathcal{E}_0|}$, where $\mathcal{E}_0 := \max_t \{|\mathcal{E}_t|\}$.

Induced correlation as a function of K. The transfer matrix technique allows us to compute also the correlation between the maps decoded τ bins apart, defined

as

$$C(\tau) := \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \left(\langle m_t m_{t+\tau} \rangle - \langle m_t \rangle \langle m_{t+\tau} \rangle \right) , \tag{7}$$

where the angular-bracket notation denotes the average over the probability distribution in Eq. (5). $C(\tau)$ decays exponentially with τ, over a characteristic 'time' monotically growing with K in Eq. (5), see Results, Section 3.1.

2.6 Position decoding

Decoding of rat trajectory in a given map. The square box in which the rat is moving is discretized into 60×60 squared cells of width 1 cm, with integer coordinates x, y. For each reference environment m (= A or B) we construct the binary rate map $p_i^{(m)}(x, y)$, equal to the average value of the activity $(= 0, 1)$ of neuron i (in time bins of width Δt) when the rat is at position x, y. Assuming that the cell activities $\{s_{i,t}\}$ at time t and at *fixed* position x, y are independent their likelihood reads

$$P^{(m)}(\{s_{i,t}\}|x, y) = \prod_i \left[p_i^{(m)}(x, y) \cdot s_{i,t} + \left(1 - p_i^{(m)}(x, y)\right) \cdot (1 - s_{i,t}) \right] . \tag{8}$$

To account for the continuity of the rat trajectory over time we introduce a Gaussian prior between successive positions,

$$P(x_t, y_t \mid x_{t-1}, y_{t-1}) \propto \exp\left(-\frac{1}{2\,(v\,\Delta t)^2} \left((x_t - x_{t-1})^2 + (y_t - y_{t-1})^2 \right) \right) \tag{9}$$

where v has the dimension of velocity, and is optimized over the reference sessions to minimize the position decoding error. The two-step procedure of [3] consists of maximizing $P^{(m)}(\{s_i\}|x_t, y_t) \times P(x_t, y_t \mid x_{t-1}, y_{t-1})$ over x_t, y_t, recursively over the bin index t. This greedy procedure does not provide the most likely trajectory of the rat over the whole time horizon $t = 1, 2, \ldots, T$. To do so we resort to the Viterbi algorithm (see Section 2.7 below).

Combining map and position inferences. While the map m is reliably known in reference sessions, it may change due to light cue switch or competing attractor dynamics [18] in test sessions. To overcome this ambiguity we first use a map decoder (Methods, Section 2.3) to infer the temporal sequence of maps $\{m_t\}$, and then plug it in Eq. (8) to infer the rat trajectory.

2.7 Inference of temporal sequences of maps or positions with the Viterbi algorithm

The inference of the temporal sequence of maps $\{m_t\}$ or of positions $\{x_t, y_t\}$ may be framed in the context of hidden Markov models [19]. In this framework the internal *state* of the system undergo *stochastic transitions* across time. The binary, discretized neural activity vector $\{s_{i,t}\}$ is the stochastic observable signal, whose *emission probability* depends on the state of the system at time t. We now detail the sets of states, of transition and emission probabilities considered in our analysis:

Map decoding. States are maps $m \in \{A, B\}$. The transition probability is

$T(m_t \to m_{t+1})$ is proportional to e^K if $m_t = m_{t+1}$ and to e^{-K} if $m_t \neq m_{t+1}$, see Eq. (5). The emission probability is $P(\{s_{i,t}\}|\Theta^{(m_t)})$, where $\Theta^{(m)}$ is the set of parameters defining the activity distribution model for map m (Methods, Section 2.3), see Eq. (2).

Position decoding. States are the 60×60 positions (x, y) in the box. The transition probability $T\big((x_{t-1}, y_{t-1}) \to (x_t, y_t)\big)$ is the Gaussian prior in Eq. (9). The emission probability is $P^{(m)}\big(\{s_{i,t}\}|(x_y, y_t)\big)$ given by Eq. (8). The map m is known, either constant across time, or inferred before the trajectory is computed.

The Viterbi algorithm consists of two major steps: (1) a *forward* procedure, which recursively computes the most probable temporal sequences of states for each possible final state, and (2) a *backward* propagation, which starts from the most likely final state and traces back the optimal temporal sequence of states [19].

3 Results

3.1 Cross validation of map-decoding methods

Decoding methods and number of parameters used

We start by presenting map-decoding methods and their performances. For each environment, A and B, we have two recorded sessions with constant light cues: the first one, called reference session, is used to infer the decoder parameters. The second one, called test session, is used for cross-validation, *i.e.* to assess the performances of the decoder. We compare the performances of five different decoders, described in Methods, Section 2.3. Our decoders mainly differ by the fact that they may use or not knowledge of the rat positions and of the spatial rate maps (place fields). They are also based on simple comparison methods or on more sophisticated probabilistic frameworks.

- **Rate-map based decoders** require the computation of the rate maps during the reference session. Knowledge of the position $\vec{x}(t)$ of the rats and of the neural firing rates $r_i(t)$ as a function of time t allows one to build the rate maps, that is, the average firing rate of each cell i as a function of the rat position $\vec{x}$, $r_i^{(E)}(\vec{x})$ for environment $E = A, B$. The similarities between those reference population activities and the activity measured during the test sessions may then be used as a simple estimator of the map retrieved by the rodent. We consider two such comparison-based approaches, called *Dot Product* and *Pearson* [4]. A more sophisticated decoder, called *Poisson*, consists in assuming that each place cell i fires with a Poisson process, with average rate $r_i^{(E)}(\vec{x})$ when the rodent is at position $\vec{x}$, and in estimating the likelihood of the test spiking activity with this multiple Poisson process and for maps $E = A$ and B. The posterior distribution for the (binary) map variable E can then be computed, and we decode the map as the one with larger posterior probability. Poisson is based on a more solid probabilistic framework than Dot Product and Pearson, while making use of the same rate maps estimated from the reference sessions.
- **Activity-only decoders** do not need any information about rat position and place fields. Those models provide approximate expressions for the probability distribution of population neural activity over short time bins, *i.e.* of binary (silent or active neuron in the time bin) strings of length N (the number of recorded neurons). The *Independent-cell* model is the simplest maximum-entropy model [20] ; it reproduces the N average activities of the neuron only. The second model, called *Ising* in statistical physics, is a graphical model that, in addition, reproduces the pairwise correlations between the neural activities in a time bin [8, 11, 20]. The Ising model requires the inference of pairwise effective couplings between every two cells, which we have performed with the Adaptive Cluster Expansion method [11, 12]. Similarly to Poisson, the Independent-cell and Ising models provide estimates of the likelihood of the population activity in a time bin, and can be used to compute the posterior distribution for the map variable, E, and to decode the retrieved map through maximization over E.

As a consequence the numbers of parameters to be learned from the reference sessions vary a lot with the decoders. For N recorded neurons (38 in one of the data sets studied here, see Materials) and a discretization of the environment into S (=20×20 in the present analysis) spatial bins, the numbers of parameters to be extracted from the reference sessions are, respectively $N = 38$ for the Independent-cell decoder, $\frac{1}{2}N(N+1) = 741$ for the Ising decoder, and $N \times S = 15,200$ for the Poisson, Pearson, and Dot Product decoders.

Inferred Ising couplings are fingerprints of environment representation in CA1

As a result of global remapping taking place in CA3 (Fig. 1) the populations of active cells in the two environments are essentially orthogonal. This property can be seen from the comparison of the inputs $\{h_i\}$ in the Ising models inferred in the reference sessions of the two environments, see Fig. 2 (a). Inputs are increasing functions of the average firing rates (Methods, Section 2.3.1). As expected the few cells strongly active in one environment, *i.e.* having large inputs h_i, show very little activity (if any) in the other environment, *i.e.* have very low inputs. In CA1, the presence of rate remapping and the higher firing frequencies of cells make the input h_i larger and much more similar across the environments, see Fig. 2 (b).

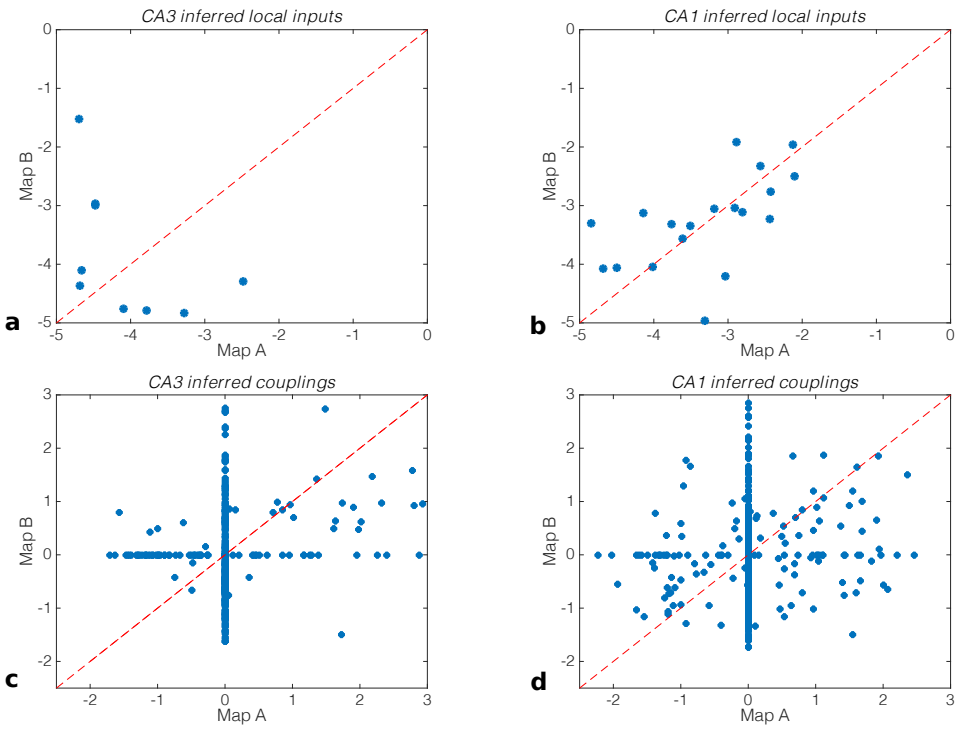


Figure 2: **Comparison of inferred Ising parameters across the two maps**. Top: Inputs h_i of the Ising models inferred from CA3 **(a)** and CA1 **(b)** reference sessions. Only values greater than -5, corresponding to a firing rate of c.a. 0.05 Hz in the independent-cell model, are shown. Bottom: Couplings J_{ij} of the Ising models inferred from CA3 **(c)** and CA1 **(d)** reference sessions. Note the presence of many zero couplings in both environments. Analysis performed with discretization time bin $\Delta t = 120$ ms.

Distinction between the neural representations of the environments in CA1 can,

however, be drawn from the correlational structure of firing events in the place-cell population. Place cells with overlapping firing fields in one environment are indeed more likely to be simultaneously active during the animal's exploration, and their activity is thus correlated. Due to remapping the amplitudes of these correlations are specific to each environment. The inferred Ising couplings $\{J_{ij}\}$, which captures the direct correlation between cells i, j not mediated by other recorded cells (Methods, Section 2.3), are vastly different from one environment to the other, as shown in Fig. 2 (d). The set of effective couplings $\{J_{ij}\}$ is therefore a *fingerprint* of the environment [21], which we can exploit to distinguish between maps, *i.e.* to decode the neural representation. Note that these effective, functional couplings are not directly related to the physiological synaptic interactions, which are not accessible from the data.

Focusing on effective couplings rather than the simpler pairwise correlations allows us to *score* any configuration of the population activity, that is, to quantitatively assess its similarity with typical activities in each environment, as shown below. This score is, in practice, given by the Ising likelihood, see Methods, Eq. (2), and heavily relies on the inferred couplings. Use of simpler scoring functions, based on inputs $\{h_i\}$ only, see Independent-cell model in Methods, is sufficient to decode neural representations in CA3 but not in CA1.

Comparison of performances of map-decoding methods

We present a systematic study of the performances of map-decoding methods within the framework of binary-decoder theory (Methods, Section 2.4). Results for CA1 and CA3 are reported, respectively, in Figs. 3 and 4.

We plot in Fig. 3 (a) the Receiver Operating Characteristic (ROC) curve for CA1 for the Ising and Pearson decoders. Briefly speaking, ROC curve shows the value of the True Positive Rate (fractions of time bins in reference session for environment A for which the decoder rightly decodes map A) as a function of the False Positive Rate (fractions of time bins in reference session for environment B for which the decoder erroneously recognizes map A). A random decoder would have equal values for TPR and FPR, and lies on the diagonal line of the unit square in Fig. 3 (a). A perfect decoder would always recognize map A in environment A and never in environment B, and would thus correspond to TPR=1, FPR=0. Varying the threshold for significance of the decoder change concomitantly the values of TPR and FPR, with the resulting ROC shown in Fig. 3 (a). We observe that the Ising decoder shows much better performances than the Pearson decoder. An alternative representation of the decoder performances is given by the Precision-Recall curve, shown in Fig. 3 (b), see Methods, Section 2.4 for definition.

A measure of the accuracy of the decoder is given by the integral of the ROC curve, called Area Under the Curve (AUC), which ranges from 0.5 for a random decoder to 1 for a perfect decoder. To compare the five decoders we plot in Fig. 3 (c) their AUC values as a function of the elementary time bin Δt, ranging from 10 ms to 1s. The Ising model, which takes into account the correlational structure of the population activity, has higher decoding precision and retrieval capacity than other decoders in CA1 recordings (Fig. 3 (a,b)). As a consequence, in terms of AUC (Fig. 3 (c)), *Ising* is generally the most performant model, followed by

Poisson and lastly by equally-performant *Pearson* and *independent-cell* decoders. *Dot Product* method is the best performant on very short time scales (< 20 ms), but its performance increases very slowly with the time bin width, and as a consequence it has the worst performance for $\Delta t > 100$ ms.

This behavior has an explanation in terms of sensitivity of the different models to the *average number of active neurons per time bin*. Bayesian models, whose predictions do not depend on the specific position of the rat at each time, rely on information conveyed by activity alone. As a consequence, when the number of simultaneously active neurons for each time bin is very small, bayesian models may be less accurate than decoders that take into account spatial information, like Dot Product.

Maps are remarkably well predicted by all decoders from CA3 recordings, see Fig. 4. The AUC takes values higher than 0.97 for all decoders. The character of *global remapping* taking place in CA3, contrary to CA1, ensures indeed that the firing rates of the neurons largely vary from one environment to the other (Fig. 1). As a consequence, independent-cell models and correlation methods carry enough information to identify the internally-represented map.

As a general feature we observe that the performances of all decoders improve on larger discretization time scales. Statistical fluctuations in the activity are indeed smaller for large time bins, resulting in smaller errors in the decoding task. A similar reduction of statistical fluctuations can be obtained through the introduction of a continuity prior, which prevents switching back and forth between spatial maps.

Map decoding with continuity prior

Map decoding can be combined with a continuity prior that enhances persistence in the decoded maps over consecutive time bins, see Methods, Section 2.5. The motivation for the continuity prior is two-fold. First, in situations where the latency between a delivery of external stimulus and the network state change is the main parameter to be measured (e.g. after pharmacology treatment, etc.), one needs to search for a single time point of the state transition. This can be achieved by imposing a strong continuity prior, allowing for the presence of a single transition between maps along the whole recording session.

Secondly, with moderate continuity prior, dynamical events (such as state transitions) can be detected with more precision, at the price of discarding events that happen on a time-scale shorter than a given temporal resolution. When the prior is applied, the resulting correlation between decoded maps in two time-bins that are t bins apart decays exponentially with t, see Fig. 5 (a) and Methods, Eq. (7). A persistence 'time' t_0 can be computed through an exponential fit of the correlation: t_0 is the characteristic number of bins over which decoded maps are persistent. Its value can be chosen at our convenience by tuning the prior strength parameter K, see Fig. 5 (b) and Methods, Section 2.5. Hence, we can choose a temporal resolution t_0 and exploit the noise-cancelling property of the continuity prior over larger time scales.

Unless otherwise specified we set in the following the characteristic persistence time to the small value $t_0 = 2$ time bins. As shown in Fig. 3, the use of this weak

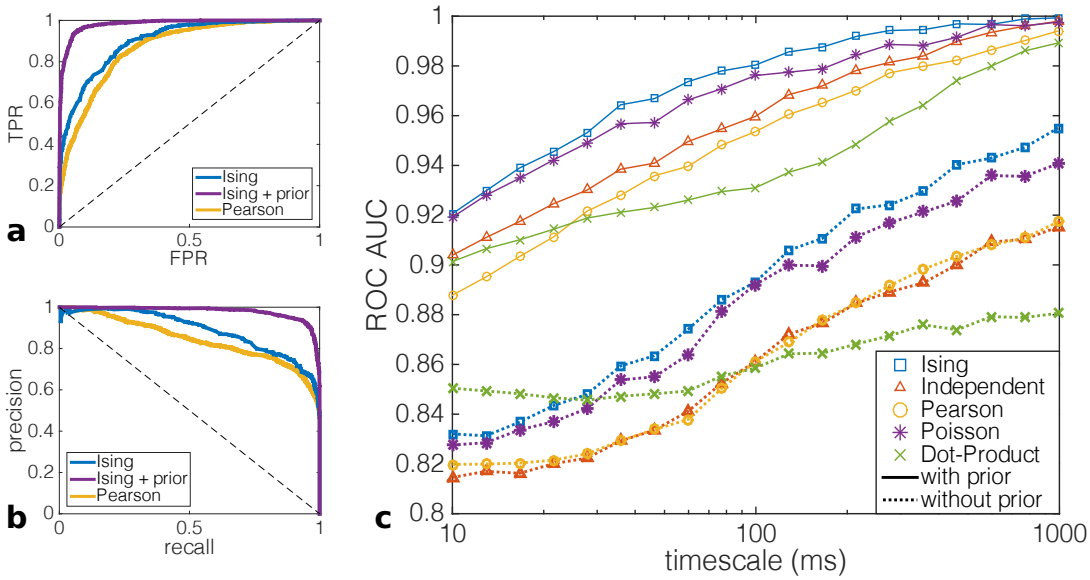


Figure 3: **Decoding performances in CA1.** ROC **(a)** and Precision-Recall **(b)** curves computed at fixed time scale $\Delta t = 120$ ms for a combination of two test sessions in environments A and B, recorded in CA1. Maps A and B correspond, respectively, to positive and negative predictions, see Table 1. The True Positive Rate, also called Recall, is the number of true positive predictions divided by the total number of positive events. The False Positive rate is the number of false positive predictions, divided by the total number of negative events. Precision is defined as the fraction of identified positive events that are true positives. **(c)** performances of *Ising*, *Independent-cell*, *Poisson*, *Pearson*, and *Dot Product* decoders (with and without the addition of a continuity prior) as a function of the discretization time scale, applied to CA1 neural recordings. Full and dashed curves correspond to predictions, respectively, without and with continuity prior; in the latter case the correlation C in Eq. (7) decays over $t_0 = 2$ time bins (Methods, Section 2.5 and Fig. 5 (a)).

continuity prior sensibly improves decoding performances with the Ising method. The AUC increases by about 10%, see Fig. 3 (c). For direct comparison, if one instead increases the time-bin resolution Δt by a factor 2, the increase in AUC is much lower (Fig. 3 (c)): for instance, Ising AUC is equal to 0.90 for $\Delta t = 120$ ms and to 0.92 for $\Delta t = 240$ ms, while it reaches 0.98 for $\Delta t = 120$ ms with a continuity prior such that $t_0 = 2$ time bins. This results shows the statistical efficiency of the continuity prior.

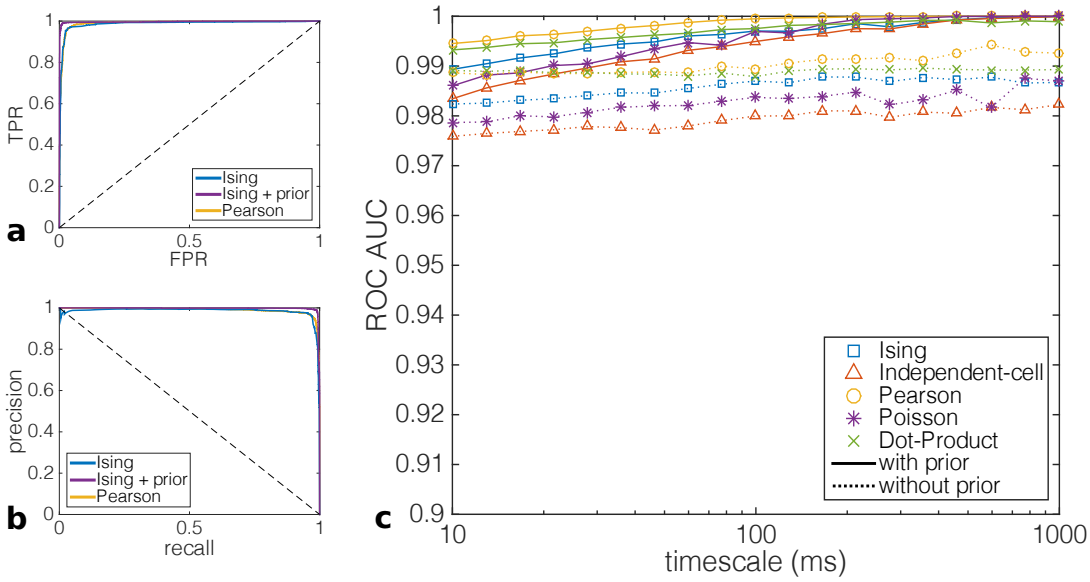


Figure 4: **Decoding performances in CA3**. ROC **(a)** and Precision-Recall **(b)** curves computed at fixed time scale $\Delta t = 120$ ms for a combination of two test sessions in environments A and B, recorded in CA3. Maps A and B correspond, respectively, to positive and negative predictions, see Table 1. The True Positive Rate, also called Recall, is the number of true positive predictions divided by the total number of positive events. The False Positive rate is the number of false positive predictions, divided by the total number of negative events. Precision is defined as the fraction of identified positive events which are true positives. **(c)** performances of *Ising*, *Independent-cell*, *Poisson*, *Pearson*, and *Dot Product* decoders (with and without the addition of a continuity prior) as a function of the discretization time scale, applied to CA3 neural recordings. Full and dashed curves correspond to predictions, respectively, without and with continuity prior; in the latter case the correlation C in Eq. (7) decays over $t_0 = 2$ time bins (Methods, Section 2.5 and Fig. 5 (a)).

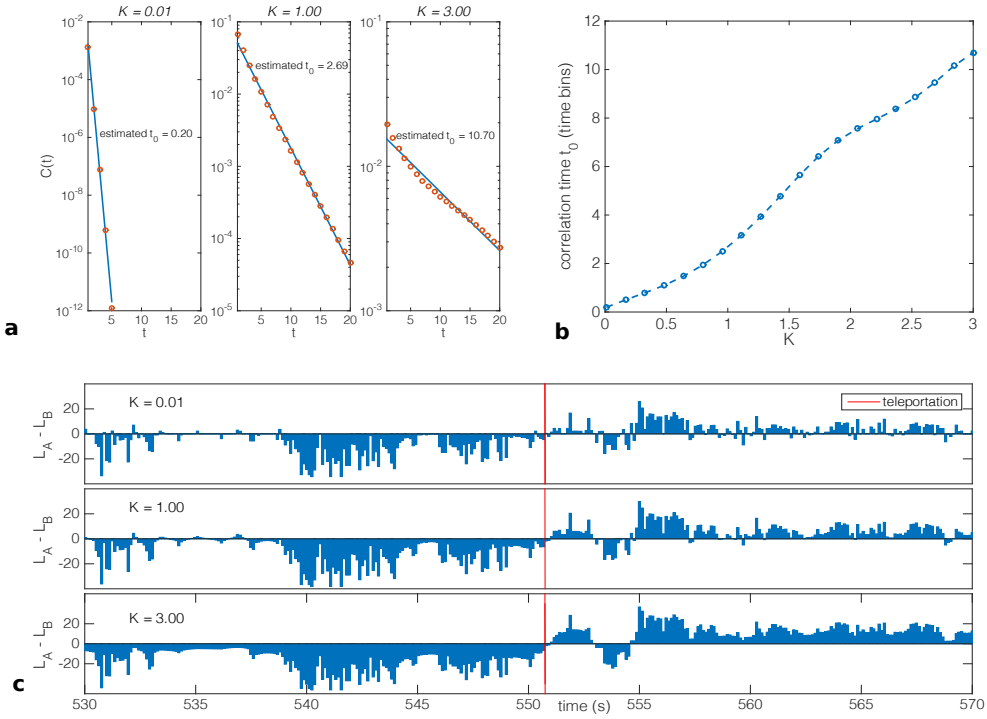


Figure 5: **Continuity prior for map decoding. (a)** Correlation (Methods, Eq. (7)) between maps decoded in two time bins as a function of their separation t (measured in units of time bins), for three values of the prior strength K. Correlations are well fitted by exponential decaying functions, over a characteristic number of bins t_0. **(b)** Value of t_0 as a function of the prior strength K. **(c)** Application of the prior on CA1 teleportation session for different values of prior strength parameter K. Ising decoder, with a discretization time bin $\Delta t = 120$ ms.

.

3.2 Transitions between maps in "teleportation" experiment

The Ising decoder presented above can be used to identify transitions between different activity states of dynamic systems. Brain hippocampal memory circuitry is a dynamic system expressing distinct states of activity - neural representations of surrounding space - with attractor properties [4, 22, 23]. We used data recorded in a 'teleportation' experiment [4], in which the appearance of recording box is abruptly changed by switching between two familiar light cue settings (A and B, respectively) while the laboratory rat continuously explores it (Methods, Section 2.2). This procedure was shown to induce a rapid exchange of corresponding hippocampal representations, including periods of instability with spontaneous fast flickering between them. Transitions between the maps were identified based on activity models of representations A and B, respectively, inferred from reference recordings in both environments under stable conditions preceding the 'teleportation test session'.

Teleportation procedure induces long-term network instability

We use our Ising decoder to assess the effect of teleportation on subsequent network state expression. Jezek et al. [4] described a period of state-instability in CA3 lasting for several seconds after the change of environment appearance. Representative evolution of the log–likelihood (with the Ising model inferred from the reference sessions) of the identified network state is shown before and after the teleportation event for CA1 and CA3 in, respectively, Figs. 6 (a) and (b). The criterion for accepting given bin as corresponding to representation of environment A or B, respectively, was set to match 1% error derived from stable reference sessions, see Fig. 6 (c,d). This ensures that a time bin is identified as A only if there is 99% confidence that its signal value does not come from state B (and vice-versa).

To characterize the kinetics of network state development, we identify the amount of time bins expressing a neural representation that was incongruent (non-corresponding) with the present environment, *i.e.* coding for the environment presented before the teleportation. We estimate the short term effect within interval of first 10 seconds, and an eventual long term effect within subsequent 20-40 seconds after teleportation, both in CA3 and in CA1, respectively. The rates of incongruent bins are shown in Fig. 7 (a,b) for CA1 and CA3. The amount of non–corresponding events were larger during the short-term period (less than 10 sec) after the light switch than those identified in the stable conditions before the teleportation test, both in CA3 as found in [4] and in CA1 (CA3: 0.011 ± 0.001 S.E.M. stable; 0.26 ± 0.04 S.E.M. after switch, $F = 34.18$, $p < 3 \times 10^{-6}$; CA1: 0.017 ± 0.003 S.E.M. stable; 0.101 ± 0.025 S.E.M. after switch, $F = 11.24$, $p < 0.015$). Across both areas, the short-term increase was larger in CA3 than in CA1, despite the initial incongruency rates were generally lower in CA3 than in CA1 prior to the teleportation, see Fig. 7 (a,b).

An interesting finding comes from the analysis of long-term effect shown in Fig. 7 (a,b). Compared to stable environment conditions, the rate of flickering is generally increased both in CA3 and CA1, respectively (CA3: 0.036 ± 0.007 S.E.M. long-term post switch, $F = 12.6$, $p < 0.014$; CA1: 0.048 ± 0.006 S.E.M. long-term post switch, $F = 19.21$, $p < 0.005$). In CA1 the difference between the long-term

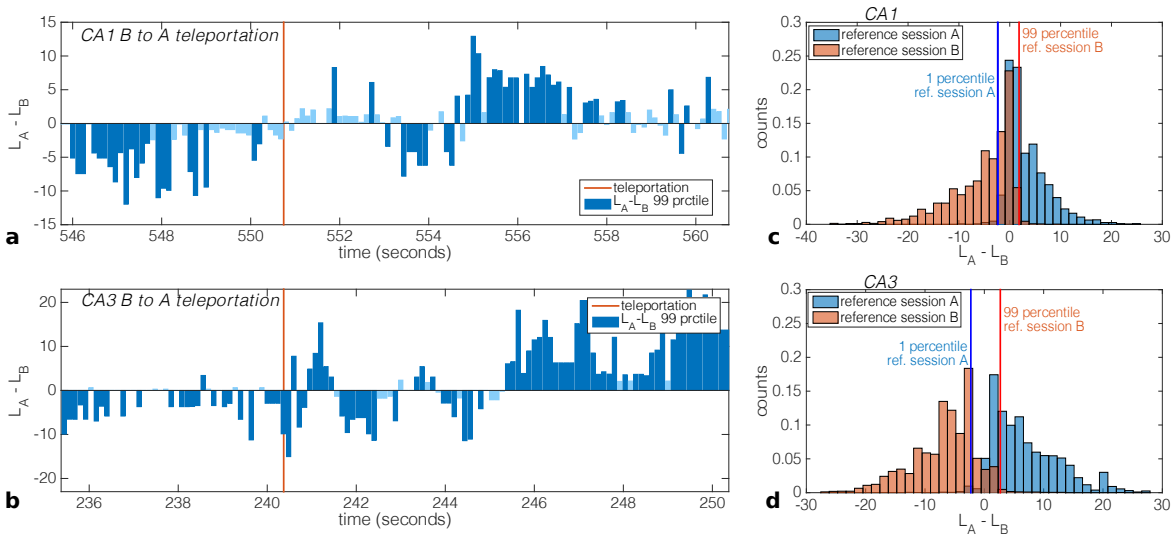
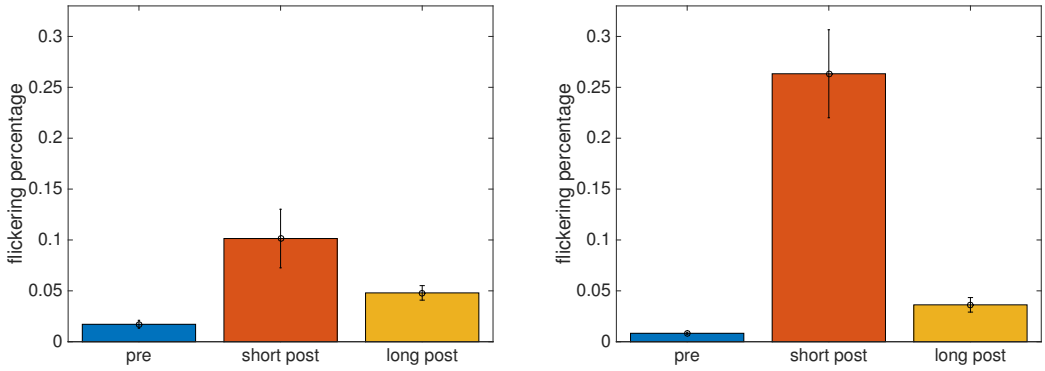


Figure 6: **Log-likelihoods of neural activities in CA1 and CA3 and teleportation event**. Ising decoder applied to neural activity in a teleportation sessions in CA1 **(a)** and CA3 **(b)**. The light switch is marked with a red line, predictions higher than 99 percentile value of reference sessions are colored in dark blue, weaker prediction are colored in light blue. Panels **(c)** and **(d)** show the distributions of log-likelihoods in, respectively, CA1 and CA3 across reference sessions. A percentile value θ in $[0, 100]$ (normally in the interval $[90, 100]$) is defined. We consider a test time bin as significantly decoded as A only if the log-likelihood of the activity configuration in the time bin is higher than the θ percentile value of reference session B, and as B only if its value is lower than the $100 - \theta$ percentile value of reference session A. The underlying reasoning is to decode a test time bin as A only if it is very unlikely that it comes from reference population B, and vice-versa.

and short-term enhancements in network instabilities is less evident, considering the statistical uncertainty (CA3: $F = 26.9$, $p < 2 \times 10^{-5}$; CA1: $F = 4.33$, $p = 0.08 > 0.05$).

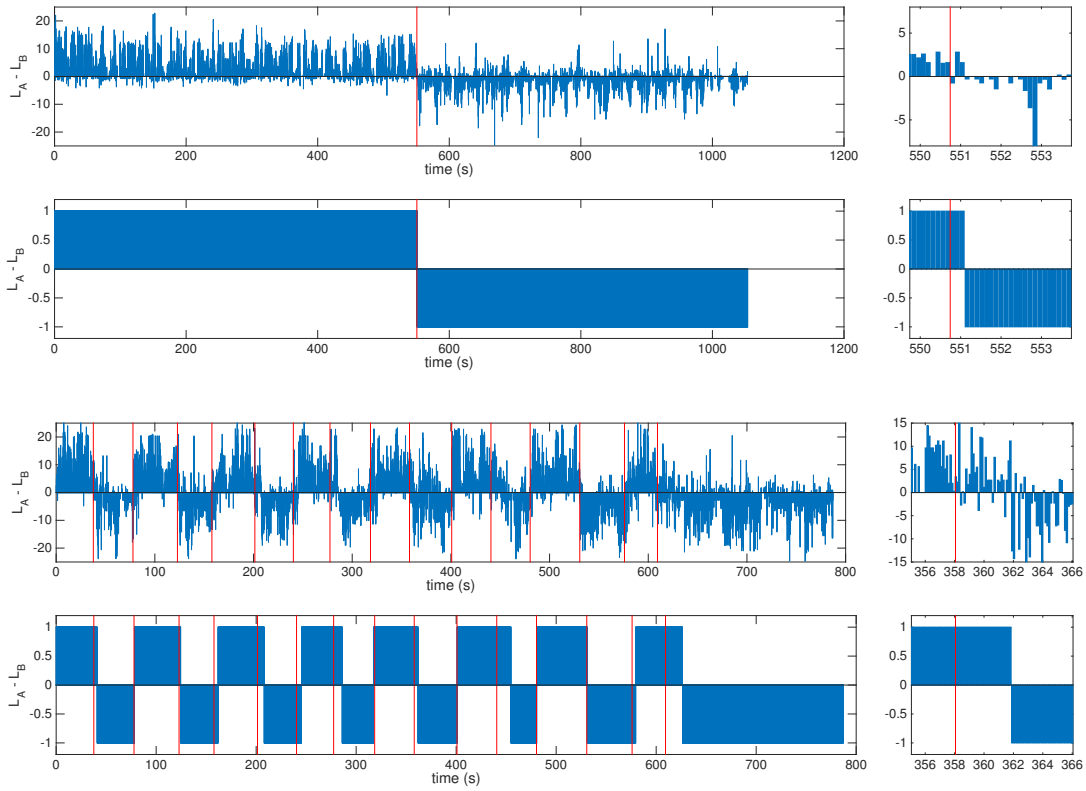
Identification of transitions with strong continuity prior

The network state-decoding procedure with implemented continuity prior can be used to detect internal state-shifts under predefined criteria. For instance, when the prior strength is brought to extreme values the decoding procedure discards the fast instability-driven dynamics and, instead, returns a single state transition time point that reflects the evolution of log-likelihood values across the continuum of temporal bins. Taking as an illustration the CA1 teleportation session in Fig. 8(a), we see that the response of network activity state to the teleportation event is identified with high accuracy. This is avaluable tool to measure the most probable moment of network remapping even under widely fluctuating dynamics. A further illustration is provided by decoding of maps from CA3 recordings, in which the strong post-teleportation instability makes identification of a state transition rather difficult. Figure 8(b) shows a multi-teleportation session with identified transitions between the CA3 representations, each corresponding to a single teleportation event.



(a) CA1 (b) CA3

Figure 7: **Teleportation enhances network instability over both short and long term periods.** Percentage of temporal bins expressing the environment-incongruent coding computed in rest conditions (pre), during the first 10 seconds after a light switch (short post), and in long-term period after the teleportation event (long post, 20-40 seconds after light switch). Only bins expressing likelihood values higher than 99 percentile of reference sessions have been taken into account. Values were averaged over 4 sessions in CA1 and over 12 teleportations within the same session in CA3. Analysis performed with Ising environment decoder with discretization time bin $\Delta t = 120$ ms.



(a)

(b)

Figure 8: **Network state transitions identified by implementation of continuity prior.** Ising decoder and Viterbi algorithm with strong continuity constraint applied to neural activity in a CA1 **(a)** and CA3 **(b)** teleportation session with enlarged examples. Light switches are marked with red lines. Analysis performed with time bin $\Delta t = 120$ ms.

3.3 Combining map and position decoders allows for reliable identification of transitions in CA3

An important issue is to distinguish real flickering events, in which the internally-represented map is incongruent to the external physical environment, from erroneous map predictions produced by the decoder, *i.e.* False Positive events due to noise and imperfect orthogonality of representations. One option is to use very strong statistical thresholds, as done in previous sections, to identify statistically-reliable transitions. However, the thresholding procedure inevitably discards a large amount of data, leaving many data points unresolved.

Another option is to exploit the recorded path of the rat, which has not been used in our analysis so far. Place-cell activities 'follow' the trajectory of the animal in the environment. Therefore we expect that a strong mismatch between the physical location of the rodent and the reconstructed position in the decoded map might signal an erroneously predicted map [4]. Therefore we hypothesize the correctly decoded map will correspond to a match between the animal's real and reconstructed coordinates. In this section we use a combined map & position approach as a novel way to increase the reliability of decoded results.

Position decoding

We first introduce our position decoding procedure. Several methods of position decoding, exploiting the spatial firing specificity of place cells, have been proposed in the literature [3, 24, 25]. Among the best performing approaches is a two-step algorithm, which computes the most likely position of the rat given the activity recorded under the condition that this position does not vary much from the identified location at previous times [3]. Here we extend this procedure in a full Bayesian setting, as we compute the *full trajectory* of the rat given the whole recording of the activity and a continuity prior requiring that the rat apparent velocity (computed from the change in the reconstructed positions between two successive time bins) does not exceed much its average velocity (measured during the experiment). The full-trajectory posterior maximization is performed through the so-called Viterbi algorithm [19], see Methods, Section 2.7.

We first apply our spatial decoding procedure to CA1 and CA3 recordings. Figure 9 (c) shows the comparison between true position and inferred position in the constant-environment (no switch) CA1 test session. Use of the Viterbi algorithm to maximize probability over the full path yields an average error, defined as the distance between those two positions $|\vec{x}_{\text{infer}} - \vec{x}_{\text{true}}|$, of 5.5 cm on the test session, with 40 ms time bins and 36 recorded cells. The average error over position for different discretization time scales is shown in Fig. 9 (a). The error stays roughly constant over all time scales less than $\sim$ 100 ms, and increases for longer scales. This is coherent with the value of the average running speed of the rodent, ca 10 cm/s: over long time scales more than one spatial bin (of 1 cm width) may be explored in each time bin. In addition, in Fig. 9, we show that our Viterbi-based full-trajectory reconstruction approach outperforms the two-step procedure of Zhang et al. [3] on our data set.

The same analysis conducted in the CA3 region (34 cells) produces higher average errors, as expected, see Fig. 9 (b). The lower number of simultaneously

active cells, compared to CA1, results in a less precise position decoding outcome for an equal number of simultaneously-recorded neurons.

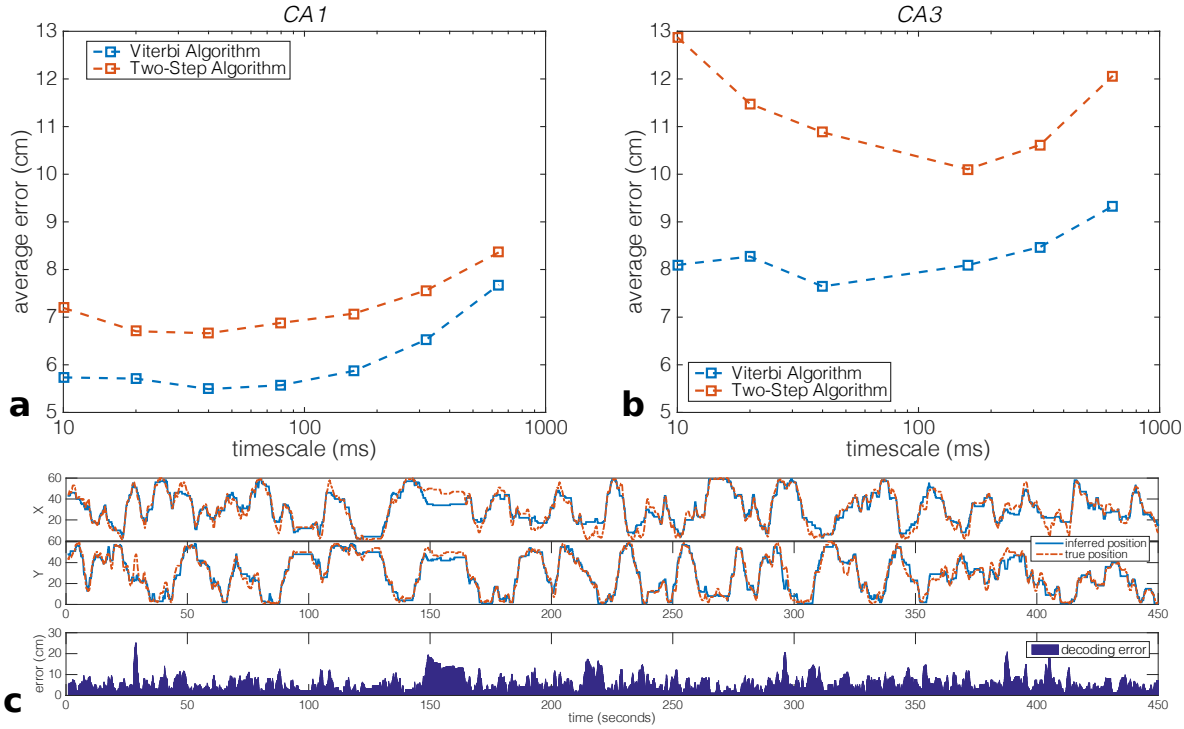


Figure 9: **Position decoding from activity of place-cell population.** Average position errors (in cm) as a function of the time–bin width Δt, with our position decoder (blue curve) and the two-step procedure of [3] (red curve) for CA1 **(a)** and CA3 **(b)** constant-environment test sessions. **(c)** Prediction of position with the Viterbi-based decoder (blue curve), compared to the true location of the rat (red curve) from CA1 recording. The top two panels refer to the X and Y coordinates as a function of time, for a representative 450 second long interval. The bottom panel shows the error over position defined as the distance between predicted and real positions. Analysis performed on CA1 constant-environment test session with time bin $\Delta t =$ 40 ms, continuity prior strength $\sigma = 1.53$ cm. Average positional error $= 5.49$ cm.

Position decoding in CA3 teleportation session confirms the presence of flickering events

We now consider inferring the animal position from the neural activity recorded in a CA3 session during which teleportations (light switches) were carried out. We first use the map associated to the light cue for position decoding. Results are shown in Fig. 10 (a), which reports the inferred ($\vec{x}_{\text{infer}}$) and true ($\vec{x}_{\text{true}}$) positions over a 80 second interval with 2 teleportation events. We see that large errors over the predicted positions, largely exceeding the expected error reported in Fig. 9(b), are found, in particular in the periods following teleportation events. More precisely, large errors often coincide with the green impulses in Fig. 10 (a), locating the time bins in which the map identified by the Ising decoder is incongruent with the environmental cues (light conditions). Figure 10 (b) shows that, if the position is inferred based on the decoded map rather than the cue-associated map, the position inference error during incongruent time bins (putative transitions) is strongly

reduced.

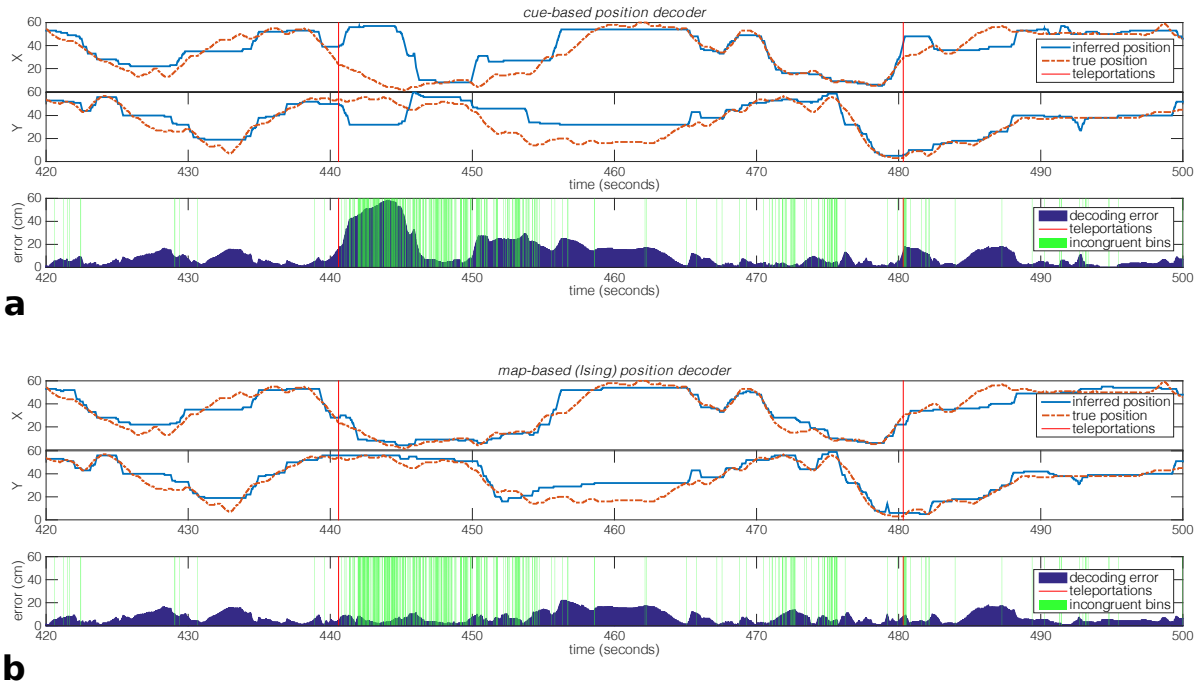


Figure 10: **Cue-based vs. map-based position decoders in teleportation session.** True (orange) and inferred (blue) coordinates X and Y of the rat, and positional error (dark blue) for one representative portion of CA3 recording including two teleportations (light switches, indicated by vertical red bars). Position decoding is done with the maps corresponding to the light cues in panel **(a)** and to the Ising decoder output in panel **(b)**. Time bins predicted by the Ising decoder to carry out neural representations incongruent with the light cues are highlighted with green impulses. Notice the reduction in the error over position when the Ising map is used instead of the cue-associated map. Parameters: time bin $\Delta t = 20$ ms, position–continuity prior strength $\sigma = 1.04$ cm.

We plot the histogram of positional errors restricted to incongruent time bins using the decoded and cue-associated maps in Fig. 11 (b). The strong error reduction obtained using the decoded map (with Ising decoder) rather than cue-associated map (map-based 9.4 ± 0.2 S.E.M, cue-based 17.2 ± 0.5 S.E.M, $F = 211$, $p < 2 \times 10^{-45}$), during incongruent time bins, suggests that our Ising decoder is able to identify real flickering events, during which the neural activity is similar to what would be expected for the same physical position of the rat in the other environment than the one associated to the light cue.

We show in Fig. 11 (d) that the larger the confidence in the decoding prediction (estimated by the difference of log-likelihoods of the position in the two possible maps), the larger is the reduction in the positional errors when using the decoded map. This confirms that incongruent events decoded with large significance threshold are likely to correspond to true internal changes of representation, rather than artifacts of our map-decoding procedure.

Position is steadily encoded in neural activity after teleportation, in spite of map instability

Figure 11 (a) reports the position error computed withing the map decoded by the Ising model, as a function of the time elapsed after the light switch. The error remains roughly independent of time, with a value close to the error computed in the constant-environment test session reported in Fig. 9 (b). This is a remarkable result as, contrary to the constant-environment session, many flickering events arise in the first 10 seconds following teleportation, as shown in Fig. 7 (b). To better characterize the evolution of flickering events we plot the fraction of incongruent bins as a function of time after teleportation in Fig. 11 (c). We observe the strong increase in the amount of bins with non-corresponding representation within the first 10 seconds, followed by a plateau, in agreement with the results of Fig. 7 (b).

The time-dependence of the statistics of flickerings events explains the behaviour of the positional error computed with the maps associated to the environments after and before the teleportation, respectively called cue and non-cue in Fig. 11 (a). At short times both errors are comparable, as roughly half of the time bins are incongruent (Fig. 11 (c)). After about 10 seconds the cue-based positional error decays to its constant-environment level shown in Fig. 9 (b), while the non-cue error increases to a value close to constant-environment error computed with the wrong map. Again we stress that the position of the rat is accurately reconstructed at all times when computed from the Ising–decoded map, either during or outside flickerings.

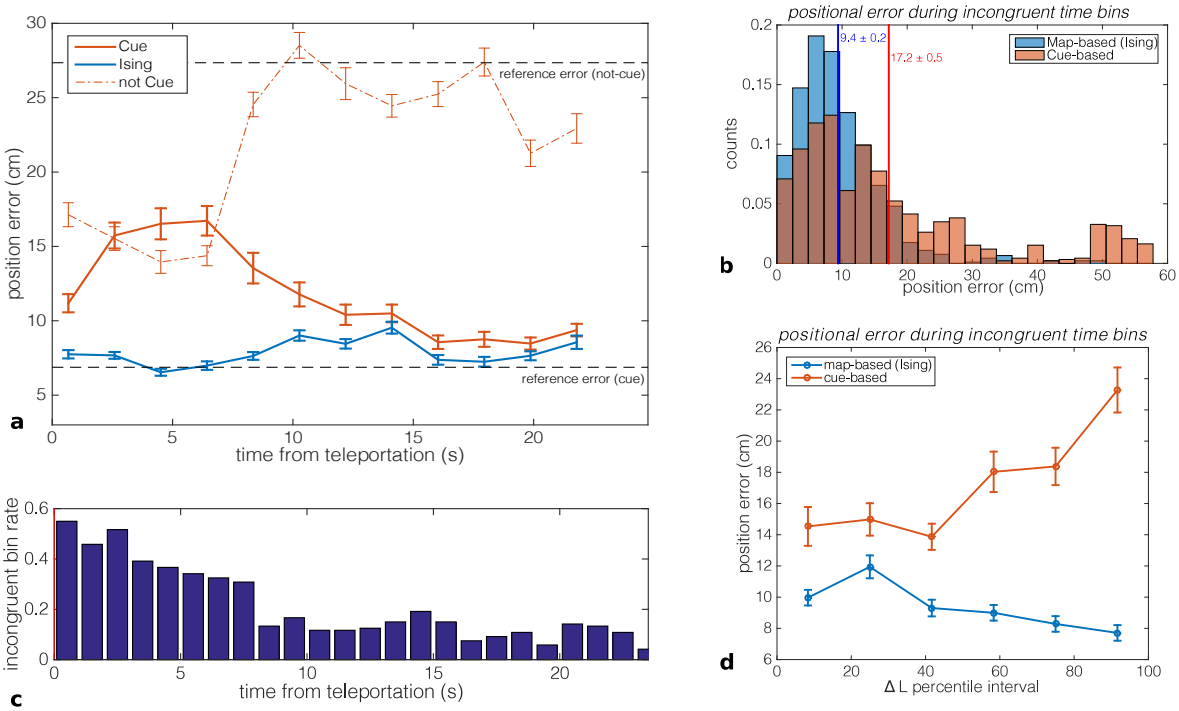


Figure 11: **Positional error in Ising-decoded map is low at all times following teleportation, regardless of flickering events in CA3. (a)** Error over the position of the rat as a function of the time elapsed after teleportation (light switch), computed with the Ising decoder (blue curve), the cue-associated (full red curve) and the non-cue (corresponding to the environment prior to teleportation, dashed red curve) maps. The horinzontal dashed lines show the average errors in reference sessions with no teleportation, computed with the environment-associated (cue) and the other (non-cue) maps. **(b)** Histograms of positional errors computed during time bins identified as incongruent by the Ising decoder. The red and blue bars shows the frequencies of errors computed with, respectively, the cue and non-cue maps. Parameters: $\Delta t = 120$ ms. **(c)** Average rate of flickering events as a function of the delay after teleportation. **(d)** Average positional errors over incongruent time bins identified with the Ising decoder as a function of the percentile θ of the difference between the log-likelihoods assocaited to the two maps, see Results, Section 3.2.

4 Discussion

We first discuss some technical points about our approach, in light of existing literature, before emphasizing the main points of the present work.

Graphical models for neural representation identification. Methods for decoding considered in this work can be divided in two classes, depending on whether they make use of positional information or not. Remarkably, the latter methods do not show worse performances than the former approaches. The Ising decoder, which uses no information about the rat position and firing fields, yields the highest performance on all time scales in CA1 (Fig. 3). The higher performance of the pairwise Ising model, combined with the lower number of parameters involved in the inference process, suggests that the correlational structure of neural firing activities conveys essential information about the internal representation of memorized environments.

While we have here considered the activity vectors as discretized in regular-spaced time windows of duration Δt it would easy to extend our analysis to process activity in elementary windows in correspondence to Theta cycles. Such an analysis would be interesting to assess the plausibility of transition scenarios put forward by theoretical studies [26] and deepen our understanding of the role of Theta oscillations for the dynamics of transitions [4]. In this regard repeating the present study with probabilistic models capable of capturing some aspects of the activation dynamics in recorded spiking sequences, such as Generalized-Linear Models [27] could be potentially interesting. Contrary to their Ising model counterparts effective couplings in the GLM approach are not necessarily symmetric, and may reflect specific ordering in neuron activations. However, some basic assumptions underlying GLM, such as the Poissonian nature of firing events are questionable for hippocampal place cell activity [28].

Continuity prior for identification of transitions. We introduced a continuity prior for map decoding for two main purposes:

Weak continuity priors, obtained for low values of the prior strength parameter K in Eq. (5), help to smooth out spurious transitions. Clusters of contiguous time bins with the same decoded map reinforce their coherent signal, highlighting transitions of internal representation that linger on longer time scales. As a result spatial maps transitions taking place on long time scales (set by the prior strength, see Fig. 5 (b)) are easier to detect at the expense of faster transitions.

Strong continuity priors, with large strength K allow for the search of map transitions that meet preset criteria for the network stability. The Ising decoder with continuity prior allowed us to identify time points with highest probability of expected state shift following teleportations, see Figs. 8 (a,b) for applications to CA1 and CA3 recordings. Such an implementation makes possible to perform hypothesis-driven measurements. This can be beneficial whenever the latency between the sensory input and the global network response is of importance, *e.g.* in treatments with expected effect on attention, to characterize feedback dynamics of the neural system with respect to experimental conditions. The same procedure could also be used to identify attractor dynamics in experiments where the input signal (environmental light cues, in our data) is changed smoothly or gradually

over time [29].

Decoding of position. Though the inference of position from neural activity was not our primary goal in the present work, we have introduced a position decoder based on the Viterbi algorithm, which maximizes the likelihood of the full rodent trajectory. Our decoding procedure is intrinsically *off-line*, since the information contained in the whole temporal sequence of neural activities is taken into account to decode the animal position in each single time bin.

We use data from hippocampal CA1 to estimate the accuracy of position reconstruction; CA1 activity is most frequently analyzed in the literature for this purpose, partly due to its higher firing rate with respect to CA3. In our analysis, with 36 recorded neurons, we obtained an average error of 5.5 cm with a time definition of 40 ms. Note that our algorithm for position decoding bears strong similarities with the two-step maximization of Zhang et al. [3], which includes a continuity prior over position similar to ours. The main difference is that Zhang et al's procedure maximizes the posterior distribution over the position at time t, and uses the outcome for inference at later times. Contrary to our Viterbi-based decoder this greedy procedure does not generally find the trajectory with maximal posterior probability, and leads to higher errors on position compared to our Viterbi decoder, see Fig. 9 (a,b).

Though it is not possible to compare directly the performances of position decoders over different data sets (due to differences in the numbers of recorded cells, in animals, in the maze shapes and sizes, in the purpose for which cells were selected, ...) let us mention that the errors reported here are comparable to results reported in the literature [3, 24, 25].

Perspectives: Hippocampal CA3 and CA1 network state decoding. We believe our technique can be successfully used for analyzing various aspects of activity states in brain neural networks. Originating in early works of David Marr [30], memories have been conceptually framed as discrete states of collateral neural networks with attractor properties. Hippocampal representation of space is considered as a physiological substrate of spatial memory. Its systematic research over the last almost 50 years has brought a substantial knowledge about organization of memory in the brain in general. However, numerous aspects of memory-related processes are still poorly explored. One of such fields is a physiology of recall of stored memory patterns in neural networks – probably because of its obvious volatility – under normal conditions memory recollection takes place usually within a single second after stimulus delivery while its brain network correlate is indeed even faster. In this light, methods allowing for the identification of a memory state present in the network at high temporal resolution are of particular importance. A recent analysis of memory replay in the prefrontal cortex of rats with Ising models can be found in [10].

In hippocampus, patterns of place cell activity across different environments behave as uncorrelated network states with attractor properties [22]. In this paper we used multiunit recordings from hippocampal areas CA3 and CA1. Both are parts of the entorhino-hippocampal loop, an essential circuit for spatial memory and navigation in mammalian brain. Despite being directly connected in series (CA3 signaling into CA1), they very much differ in their architecture – while CA3

is organized as a recurrent network with attractor properties, CA1 has a feed forward structure – and in their connections with other brain areas involved into space representation [31]. Our recordings were taken during a free exploration in two environments in an experimental paradigm shown to induce rapid switches between hippocampal activity states. We showed that the Ising model is both sensitive and robust to decode a memory state expressed in the network, with temporal resolution high enough to reflect natural time patterning of activity provided by local theta oscillation (ca. 6-11 Hz). Whereas state identification is, due to high level of orthogonality in environment coding, rather straightforward with numerous approaches in CA3, to decode network states in CA1 is more of a challenge. The degree of orthogonality is smaller in CA1 as a consequence of the fact that firing fields of certain subpopulation of place cells tend to overlap across environments, especially those with similar shapes. Firing fields of these cells occupy corresponding coordinates and differ mostly by firing rates. This obviously does not affect position decoding within each environment, but weakens the sensitivity in state identification across environments.

In comparison with other tested decoders the Ising algorithm showed strongest performance in analyzing CA1 data and proved to be a perspective analytical tool for experiments focusing on identification of activity states in brain neural networks. It allowed us to identify network state kinetics following the sensory input switch in both CA3 and CA1. In agreement with previous report we detected a higher degree of flickering in CA3 over CA1, reflecting probably the absence of recurrent connections in CA1. Importantly, we found so far unseen flickering increase in long-term scale after the teleportation event (20-40 seconds) in both CA3 and CA1 compared with the stable environment conditions, see Fig. 7. This effect is rather surprising as the network usually achieves relative stability within a couple of seconds after the cue switch. Whereas the short-term instability can be explained by reverberatory activity in CA3 and by competing visual and self-motion inputs, the long-term flickering will probably reflect inner, sensory-independent processing with so far unknown origin.

Interestingly, we find that the position of the animal can be reconstructed with accuracy from the CA3 hippocampal neural activity, when choosing the map decoded by the Ising model, whether it agrees with the environmental cue or not (flickering), see Fig. 11. The stability of the path integrator is a non trivial phenomenon, as flickerings may occur at high rate especially in the few seconds following teleportation. Understanding the underlying mechanims, deeply related to the kinetics of phase and orientiation shift in the medio-enthorinal cortex associated to global remappings in CA3 [32] would be extremely interesting.

Acknowledgments. We are indebted to S. Rosay, who contributed to the early stage of the CA3 data analysis. We are grateful to J. Tubiana and A. Treves for useful discussions and suggestions. This study benefited from partial fundings from the CNRS-InphyNiTi INFERNEUR project and from GACR 15-20008S, PRVOUK P-36 and NPU I LO1503 of the Czech Republic.

References

[1] Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585–590, 2013.

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

[3] Kechen Zhang, Iris Ginzburg, Bruce L McNaughton, and Terrence J Sejnowski. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *Journal of neurophysiology*, 79(2):1017–1044, 1998.

[4] Karel Jezek, Espen J Henriksen, Alessandro Treves, Edvard I Moser, and May-Britt Moser. Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, 478(7368):246–249, 2011.

[5] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[6] Ian H Stevenson, James M Rebesco, Lee E Miller, and Konrad P Körding. Inferring functional connections between neurons. *Current opinion in neurobiology*, 18(6):582–588, 2008.

[7] Gašper Tkačik, Jason S Prentice, Vijay Balasubramanian, and Elad Schneidman. Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences*, 107(32):14419–14424, 2010.

[8] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[9] Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[10] Gaia Tavoni, Ulisse Ferrari, Francesco Paolo Battaglia, Simona Cocco, and Rémi Monasson. Inferred model of the prefrontal cortex activity unveils cell assemblies and memory replay. *bioRxiv*, page 028316, 2015.

[11] Simona Cocco and Rémi Monasson. Adaptive cluster expansion for inferring boltzmann machines with noisy data. *Physical review letters*, 106(9):090601, 2011.

[12] John Barton, Eleonora De Leonardis, Alice Coucke, and Simona Cocco. Ace: adaptive cluster expansion for maximum entropy graphical model inference. *Bioinformatics*, 2016.

[13] Damien Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational statistics & data analysis*, 54(4):1167–1178, 2010.

[14] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[15] Charles E Metz. Basic principles of roc analysis. In *Seminars in nuclear medicine*, volume 8, pages 283–298. Elsevier, 1978.

[16] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

[17] Michael J Pencina, Ralph B D'Agostino, Ralph B D'Agostino, and Ramachandran S Vasan. Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. *Statistics in medicine*, 27(2):157, 2008.

[18] Rémi Monasson and Sophie Rosay. Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Collective motion of the activity. *Physical Review E*, 89(3):032803, 2014.

[19] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[20] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

[21] Murat Okatan, Matthew A Wilson, and Emery N Brown. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural computation*, 17(9):1927–1961, 2005.

[22] Tom J Wills, Colin Lever, Francesca Cacucci, Neil Burgess, and John O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–876, 2005.

[23] Laura L Colgin, Stefan Leutgeb, Karel Jezek, Jill K Leutgeb, Edvard I Moser, Bruce L McNaughton, and May-Britt Moser. Attractor-map versus autoassociation based attractor dynamics in the hippocampal network. *Journal of neurophysiology*, 104(1):35–50, 2010.

[24] Emery N Brown, Loren M Frank, Dengda Tang, Michael C Quirk, and Matthew A Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *The Journal of Neuroscience*, 18(18):7411–7425, 1998.

[25] Riccardo Barbieri, Matthew A Wilson, Loren M Frank, and Emery N Brown. An analysis of hippocampal spatio-temporal representations using a bayesian algorithm for neural spike train decoding. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 13(2):131–136, 2005.

[26] Rémi Monasson and Sophie Rosay. Transitions between spatial attractors in place-cell models. *Physical Review Letters*, 2015.

[27] W. Truccolo, U.T. Eden, M.R. Fellows, J.P. Donoghue, and E.P. Brown. A point process framework for relating neural spiking activity to spiking history,

neural ensemble,and extrinsic covariate effects. *J. Neurophysiol.*, 93:1071–89, 2005.

[28] A.A. Fenton and R.U. Muller. Place cell discharge is extremely vairiable during individual passes of the rat through the firing field. *Proc. Natl. Acad. Sci. USA*, 95:3182–3187, 1998.

[29] Jill K Leutgeb, Stefan Leutgeb, Alessandro Treves, Retsina Meyer, Carol A Barnes, Bruce L McNaughton, May-Britt Moser, and Edvard I Moser. Progressive transformation of hippocampal neuronal representations in "morphed" environments. *Neuron*, 48(2):345–358, 2005.

[30] David Marr, David Willshaw, and Bruce McNaughton. Simple memory: a theory for archicortex. In *From the Retina to the Neocortex*, pages 59–128. Springer, 1991.

[31] James J Knierim. Neural representations of location outside the hippocampus. *Learning & Memory*, 13(4):405–415, 2006.

[32] 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.