Integration and multiplexing of positional and contextual information by the hippocampal network

Lorenzo Posani · Simona Cocco · Rémi Monasson

Abstract The hippocampus is known to store cognitive representations, or maps, that encode both positional and contextual information, critical for episodic memories and functional behavior. How path integration and contextual cues are dynamically combined and processed by the hippocampus to maintain these representations accurate over time remains unclear. To answer this question, we propose a two-way data analysis and modeling approach to CA3 multi-electrode recordings of a moving rat submitted to rapid changes of contextual (light) cues, triggering back-and-forth instabitilies between two cognitive representations (Jezek et al, Nature 478, p 246 (2011)). We develop a dual neural activity decoder, capable of independently identifying the recalled cognitive map at high temporal resolution (comparable to theta cycle) and the position of the rodent given a map. Remarkably, position can be reconstructed at any time with an accuracy comparable to fixed-context periods, even during highly unstable periods. These findings provide evidence for the capability of the hippocampal neural activity to maintain an accurate encoding of spatial and contextual variables, while one of these variables undergoes rapid changes independently of the other. To explain this result we introduce an attractor neural network model for the hippocampal activity that process inputs from external cues and the path integrator. Our model allows us to make predictions on the frequency of the cognitive map instability, its duration, and the detailed nature of the place-cell population activity, which are validated by a further analysis of the data. Our work therefore sheds light on the mechanisms by which the hippocampal network achieves and updates multi-dimensional neural representations from various input streams.

Lorenzo Posani · Simona Cocco
Laboratoire de Physique Statistique, Ecole Normale Supérieure and CNRS UMR 8550, PSL Research, Paris Sorbonne UPMC, 24 rue Lhomond, 75005 Paris, France

Rémi Monasson
Laboratoire de Physique Théorique, Ecole Normale Supérieure and CNRS UMR 8549, PSL Research, Paris Sorbonne UPMC, 24 rue Lhomond, 75005 Paris, France

1 Introduction

Following the discovery of place cells, which specifically fire at determined positions in space [1], the hippocampus was recognized as an essential brain area for spatial representations and memories. These cognitive representations, or maps, actually code for more than position in physical space, and are also strongly informative about context [2], including physical features of the background, such as visual landmarks, light, odors, auditory stimuli, as well as more abstract conditions, such as the emotional state or the task to be performed [3–8].

A fundamental property of the hippocampus is its capacity to memorize multiple cognitive maps [1,9–11]. This property may result from specific recurrent synaptic connectivity in the hippocampal CA3 region [12, 13], and can be theoretically understood in the framework of continuous attractor neural networks (CANN) [14, 15]. Thanks to the remapping properties of place cells, multiple maps can be memorized in the same connectivity matrix with almost no interference between them [16–19]. Cognitive maps may be retrieved when the animal explores again the corresponding environments, or be quickly and intermittently recalled depending on the most relevant behavorial information at that moment [20].

Different sources of inputs to the hippocampus concur to form, recall, and dynamically maintain cognitive maps. Changes in visual cues and landmarks may substantially affect place field shape and positioning [5]. The Path Integrator (PI), capable of integrating proprioceptive, vestibular and visual flow inputs and possibly supported by the grid-cell network in the medial-enthorinal cortex (mEC) [21], allows the animal to update the neural representation during navigation [22]. The path integrator is itself sensitive to other sources of inputs, and undergoes reset in case of large disagreement with external landmarks or sensory information [23].

Insights about how these different inputs contribute to hippocampal representations were recently obtained by studying the effects of mismatches between path-integration and visual sensory information, in particular in virtual reality settings [24,25]. In another study Jezek et al showed how abrupt changes in visual context (light conditions) during active exploration by a rodent resulted in fast flickering between context-associated maps in CA3 on the theta time scale [26] (Fig. 1A). Though they are largely artificial, these conditions offer a rare window on the fast dynamics of the place-cell population, and on how this dynamics is shaped by the various inputs.

Despite these studies, how contextual and PI inputs are combined by the hippocampal network to produce cognitive maps and accurate positional encoding remains unclear. In this work, we show, by reanalyzing the data of Jezek et al, that hippocampal activity always encodes the correct location in the retrieved map, even during the fast, unstable dynamics of the cognitive maps. This result is obtained thanks to the introduction of a dual inference method capable of extracting reliably and independently the encoded map [27] and the encoded position [28,29] from the recorded spiking activity alone (Fig. 1B). To explain this finding, we propose a CANN model of the hippocampal circuitry, fed by visual-cue and path-integration inputs projecting on both cognitive maps. The path integrator is, in turn, influenced by the hippocampal activity, closing an interaction loop between the hippocampus and the mEC [30]. Beside reproducing the flickering phenomenology, the model makes several predictions on the dynamics of cognitive maps, the relative strength of inputs, and the intricate activation of place-cell populations supporting the two maps, which are corroborated by a detailed analysis of the data.

Our work therefore proposes explicit mechanisms by which the hippocampus could be capable of encoding various contextual and self-locomotion information in multi-dimensional representations, and of updating them accurately on fast time scales.

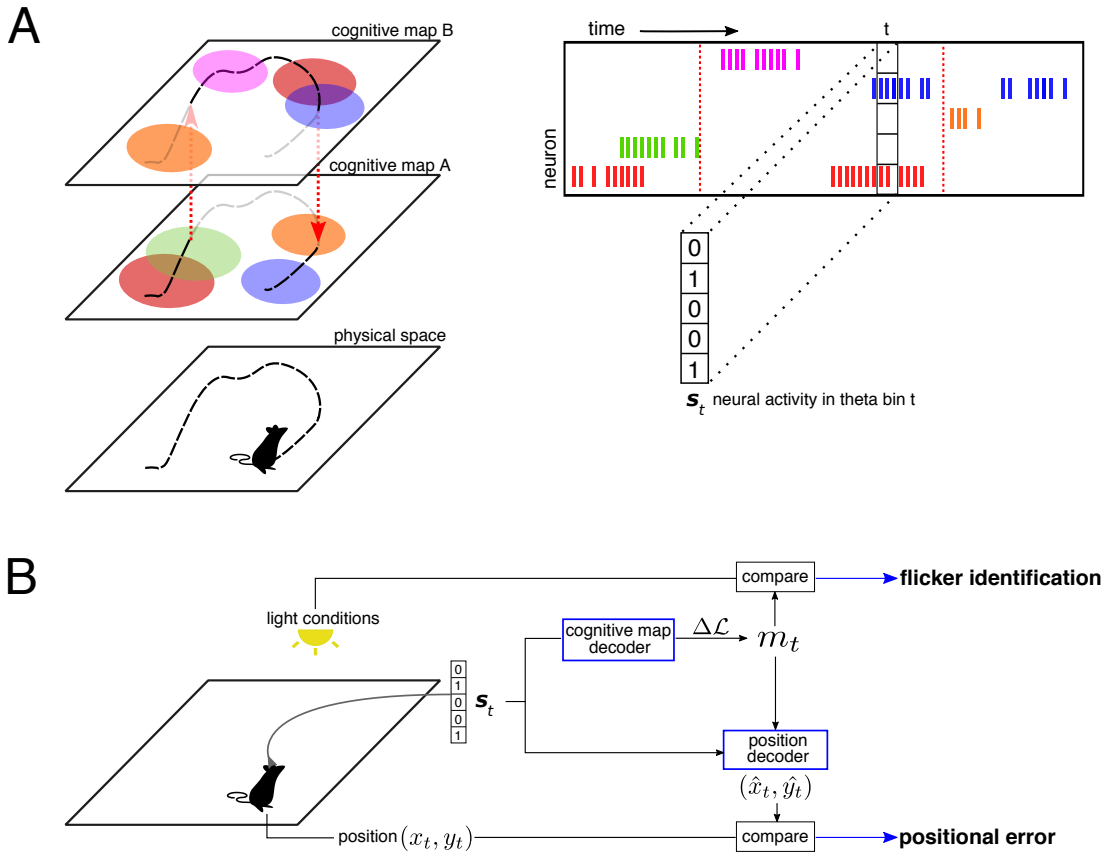

Fig. 1: **Neural encoding and decoding of cognitive maps and position. A. Schematic description of Jezek et al's experiment.** As a rodent is moving in an environment, its position $\mathbf{r} = (x, y)$ is tracked over time, and a population of place cells is recorded (see raster plot). The activity of each cell is then binarized (0: silent cell, 1: active cell) to define the activity pattern $\mathbf{s}_t$ in theta cycle t. One out of two cognitive maps, established during the training sessions, is recalled at any time; change of maps are located by vertical dashed arrows. Place cells may have place fields in both cognitive maps (red, orange, and blue cells) or in one map only (green and purple cells). In addition, pair of neurons may be active simultaneously or not depending on the map; for instance the red and blue neurons have overlapping place fields in map B, but not in A. Hence, pairwise correlations are a fingerprint of the map. **B. Sketch of the dual decoder.** The neural activity alone is used to decode the retrieved map m_t as a function of time, and then to infer the position of the animal based on the place fields in the decoded cognitive map. Mismatches between the decoded maps and the external light cues define flickers over time. The distance between the predicted and real positions defines the positional error ϵ_t.

2 Results

Jezek et al trained a rodent in two environments (square boxes), equal in size and shape, but differing by their light conditions [26]. A population of 34 CA3 place cells was recorded during reference sessions with fixed light conditions, and shown to define environment-specific maps, denoted by A and B. In a subsequent test session, taking place in a single box, instantaneous switches between environmental light conditions triggered the instability of the recalled cognitive map, which flickered back and forth between the two corresponding environments.

The neural activity $\mathbf{s}_t$ of the population in any theta cycle t encodes both syntactic (the set of rules that connects position to activity, i.e. the place fields defining the cognitive map) and semantic (the position of the rodent) information (Fig. 1A). We first introduce a dual decoder, able to independently infer the cognitive map and the position, at high temporal resolution. By comparing the inferred position to the true animal location, we then assess how precisely the position is represented in the population activity, irrespectively of the cognitive map in which it is neurally encoded (Fig. 1B).

Functional network-based decoding of the cognitive map dynamics

Due to the global remapping properties of CA3, the intensities and mutual superpositions of place fields are specific to each environment (Fig. 1A). Consequently, the average firing rates and pairwise correlations of the place-cell population define a fingerprint of the corresponding cognitive map [31, 32].

We use the reference session recordings in each environment m (A or B) to compute this fingerprint statistics. We then build a model $P_m(\mathbf{s})$ that approximates the probability of observing the neural activity $\mathbf{s}$ when the cognitive map m is recalled. This model relies on the inference of a functional network of couplings between the place cells, reproducing the fingerprint statistics of map m [32, 33] (Methods).

Given the activity $\mathbf{s}_t$ recorded in theta bin t during the test session, we then compare the two probabilistic models P_A and P_B to estimate which map m is more likely to have generated $\mathbf{s}_t$. The log-ratio

$$\Delta\mathcal{L}(\mathbf{s}_t) = \log\left[\frac{P_A(\mathbf{s}_t)}{P_B(\mathbf{s}_t)}\right] \tag{1}$$

indicates whether the neural activity $\mathbf{s}_t$ is more similar to the neural patterns encountered in map A than to the ones of map B (large and positive $\Delta\mathcal{L}$), or typical of B and not of A (large in absolute value and negative $\Delta\mathcal{L}$). Comparing $\Delta\mathcal{L}$ to a statistical significance threshold allows us to infer the map m_t (Methods). If the decoded map m_t is discordant with the imposed light conditions the theta bin is identified as a *flicker*.

As a control, we check that $\Delta\mathcal{L}$ is mostly positive in reference sessions for environment A and negative for B, see Fig. 2A. Applying the decoder to the test session, we observe the presence of flickers, see Fig. 2B (yellow bins); flickers were first found in [26] with correlation-based methods requiring knowledge of the true position of the animal. An analysis of the temporal correlation of these flickers reveals that they typically persist over ~ 6 theta bins (Methods and Fig. 2C); hence, cognitive maps show some inertia extending beyond the theta scale.

Position is accurately encoded even during flickering instabilities of the cognitive map

To assess if the fast dynamics of cognitive maps affects the quality of positional encoding we next re-use the neural activity pattern $\mathbf{s}_t$ in theta bin t, this time to infer the position of the animal. A naive Bayesian decoder [28, 29] takes as an input the above-decoded map m_t (Fig. 1B) and uses its place fields to estimate the position. The distance between the inferred position, $\hat{\mathbf{r}}_t$, and the true position, $\mathbf{r}_t$, defines the positional error ϵ_t. As shown in Fig. 3A, the positional error ϵ_t (blue line) is independent of the

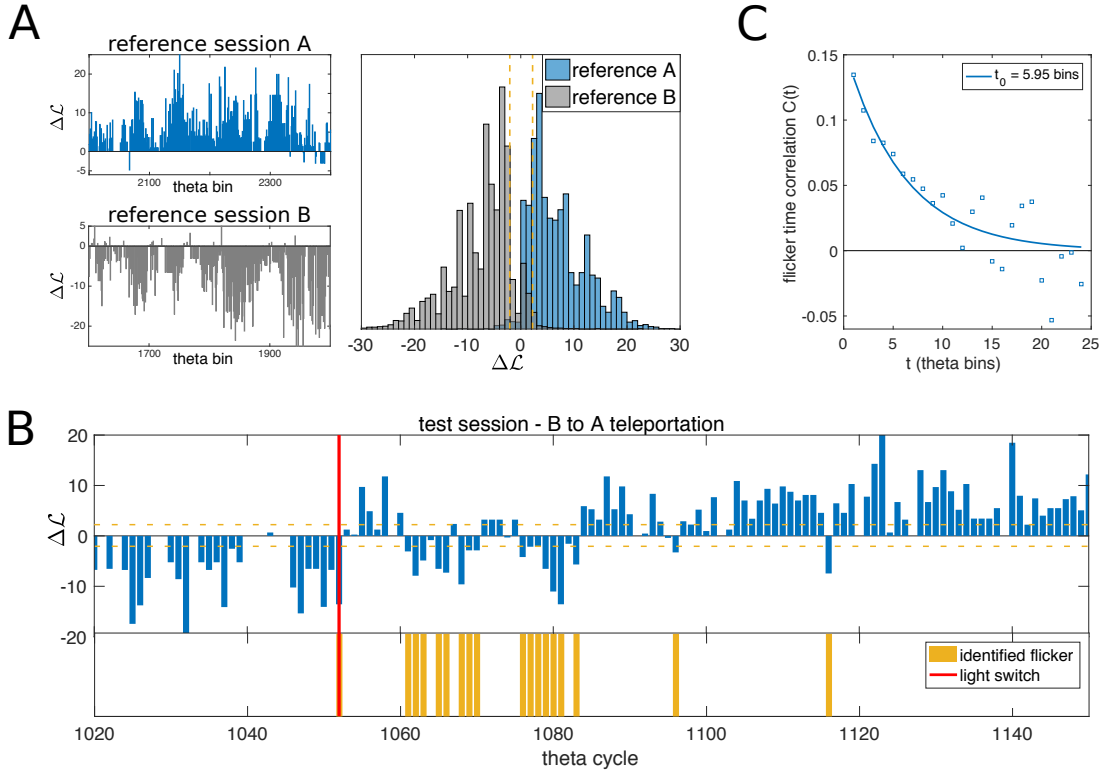

Fig. 2: **Decoding the cognitive map from neural activity. A.** Map-decoding procedure applied to constant-environment reference sessions. The log-ratio $\Delta\mathcal{L}_t$, Eqn. (1), which is the difference of likelihoods of map A and B given the recorded neural pattern in theta bin t (Methods), is mostly positive in the reference session with constant light-cue evoking A (blue) and mostly negative in the reference session B (grey). **B.** Time course of $\Delta\mathcal{L}$ in a portion of the test session around one light switch (red vertical line). The emergence of flickers (disagreement between the recalled map and the light cue) is clearly visible after the light switch. Yellow horizontal dashed lines show the statistical threshold applied in map decoding and flicker identification ($|\Delta\mathcal{L}| > L_0 = \log 10$, Methods). Yellow bars represent the identified flicker instabilities, i.e. significative discordances ($\Delta\mathcal{L} < -L_0$) between the decoded cognitive map (here, B) and the post-switch light conditions (here, A). **C.** time correlation of flickers, computed with significance threshold $L_0 = \log 10$. The exponential fit shows that correlations extend over ~ 6 theta bins, highlighting the tendency of the cognitive map to persist beyond the theta cycle.

time elapsed after the light switch, and has a value comparable to the one obtained in fixed-environment conditions (blue dashed line).

This result crucially depends on the fact that position is estimated according to the decoded map m_t, which varies with time t. For comparison, in Fig. 3B we show the error if we decode the position according to the new, post-switch map (green line) or to the old, pre-switch one (red line) at all times. Both procedures result in similar, higher errors right after the light-switch, where flickers are frequent. The error with the post-switch map eventually decrease to fixed-environment value after few seconds, due to the rarity of flickers long after the light switch.

In summary, the output of our map decoder, m_t, can be interpreted as the correct cognitive state (syntax) to read the positional code (semantics), see Fig. 2A. Even in the presence of fast dynamical flickers of the cognitive map, the location of the animal is robustly and coherently represented at all times. Our findings show that the hippocampus representation encodes both positional and contextual information in an independent and accurate way. Interestingly, the positional error computed with the

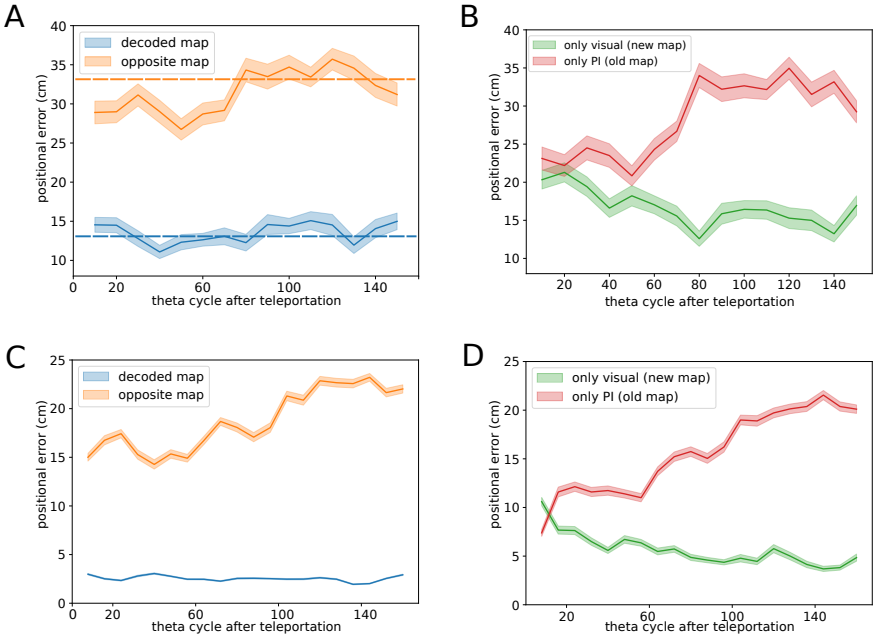

Fig. 3: **Positional error as a function of time from the teleportation light switch. A. Recorded data.** Positional error computed as a function of time from the teleportation (in units of theta cycle). In each theta bin t, the map decoder is used to select which map m_t (set of place fields) to use to infer the position from the neural activity (Fig. 1A). When using the decoded map m_t, the positional error (blue line) is comparable to the one found in constant-environment conditions (blue dashed line). The orange line shows the positional error if the opposite map (alternative to m_t) is used for inferring position. The positional error is significantly reduced with respect to constant-environment conditions (orange dashed line) in the vicinity of the teleportation, and reaches comparable values after ~ 80 theta bins. Shaded areas represent the standard error computed over 15 teleportation events. **B. Recorded data.** Same as in panel A but using a fixed map of reference for position inference, irrespectively from the decoded map. Red and green lines show results with, respectively, the 'old' (pre-teleportation) and the 'new' (post-teleportation) maps. **C & D. Simulations of CANN model**: same analysis as in panels A & B, computed over 15 simulated teleportation events, with the same trajectory of the rodent as in the experimental data. Model parameters: $N = 400$ neurons, $\gamma_J = 0.0025$, $\gamma_V = \gamma_{PI} = 0.4$, $\gamma_W = 6.25$, $\beta = 15$, see Methods, Supplementary Text, and Supplementary Information Fig. 2&3 for detailed discussion of the choice of parameters.

map opposed to the decoded one (orange line in Fig. 3A) shows a significant reduction in the first seconds after the light switch; this non-trivial effect will be explained in detail in the next sections.

Continuous attractor neural network model for the interplay between path integrator, visual cues, and memory

The findings above suggest that the stream of positional information to the hippocampus is maintained despite the presence of rapid changes of cognitive representations following the abrupt modification of visual cue (V) after the light switch. A natural hypothesis is that the path-integrator (PI) sends to the hippocampus information relative to the position in the 'old' map [11], competing with the visual cue input associated to the 'new' map. To formalize this assumption, we introduce a continuous attractor neural network (CANN) model that contains the minimal ingredients to understand the effect of conflicting PI and V stimuli onto the hippocampal activity. In the CANN paradigm for memory storage and retrieval of cognitive maps [15–17, 34], the animal location at a certain time is represented as a self-sustained bump of neural activity. The bump is localized in the current position within a two-dimensional

manifold, where place cells are embedded according to the positions of their place field centers in the real environment. We generalize this classical model by including two informational inputs on the memory network, from allothetic (visual cues) and idiothetic (path integrator) stimuli. The proposed interaction model is composed of the following four ingredients, see Fig. 4A and Methods:

(a) A CANN *memory*, including excitatory recurrent connections of strength γ_J and global inhibition, designed to store and support two cognitive maps m, denominated A and B, which mimics the status of the CA3 place-cell network after learning of the two 'environments'. This model of stochastic neurons was described and studied in [19, 34, 35].
(b) A *visual-cue input* of amplitude γ_V onto place cells whose place fields match the current position of the rat, $\mathbf{r}_t$, in the cognitive map corresponding to the external light cue. The latter is denoted by $V = A$ or B.
(c) A *path-integrator input* of amplitude γ_{PI} that projects onto place cells whose place fields match the current position in the cognitive map corresponding to its own internal cognitive state [11], denoted by the variable $PI = A$ or B.
(d) An *effective feedback* from the hippocampal network to the path integrator, which stochastically maintains coherence between the recalled cognitive map and the path-integrator state.

From a functional point of view, the CANN model mostly behaves, for a fixed position $\mathbf{r}$ of the rodent, as an effective two-state model for the hippocampal activity, as sketched in Fig. 4B. These two states correspond to the activity localized in map A or B; their probabilities are controlled by the intensity of, respectively, the path-integrator and visual-cue inputs. Note that the emergence of two well separated collective states from the microscopic CANN model is intrinsically due to the presence of recurrent connections, see Fig. 4B; A characterization of the effective barrier between the states is reported in SI, see Supplementary Fig. 1. The height of the barrier, controlled by the parameter γ_J, and the amount of stochasticity in the individual neural dynamics are crucial ingredients to determine the dynamics of the model. In particular, these variables control the the time-correlation of flickers (Methods); For the chosen simulation parameters, the time correlation decays over ~ 7 theta bins (Fig. 4C) in accordance with data (Fig. 2C).

The typical outcome of a simulated experiment is shown in Fig. 4D. During the exploration phase preceding the light switch, the visual (b) and path-integrator (c) inputs jointly contribute to the stability of the internal representation of the position. A localized bump of activity, sustained by the recurrent connections (a), can be observed in the pre-switch map (Fig. 4A, left), say, $m = PI = V = A$. Right after the switch, the hippocampal network receives conflicting streams of information: $PI = A$ differs from $V = B$. The path integrator is still activating place cells coding for the current position of the animal in the 'old' map, while the visual stream points to neurons coding for the same position in the 'new' map (Fig. 4A, center). This results in a *conflict* between the two bump representations, which are mutually incompatible due to the orthogonality of global remapping. Flickering is produced as an alternance between these two possible states, $m = PI = A$ and $m = V = B$. During this conflicting phase (Fig. 4D, red region), the feedback (d) from the memory network to the path integrator tries to achieve coherence between the hippocampal and path-integrator states. When the bump is in the visually-driven, post-switch map ($m = B$), incoherence is strong, and the path integrator is more likely to be reset. Realigning the path integrator state with the external cue, $PI = V = m = B$, brings the conflict phase to an end, and the hippocampal state reaches stability (Fig. 4A, right).

Despite its conceptual simplicity, the model shows a rich phenomenology and reproduces in a strikingly-accurate manner the results of the analysis of the CA3 teleportation recordings. In Fig. 4D we show a representative time trace of the log-ratio $\Delta\mathcal{L}$ (Eqn. (1)) in a simulated teleportation session. Alternate intervals of positive and negative $\Delta\mathcal{L}$ signal the presence of map instability, as in [26], following the light switch (red vertical lines) and the path-integrator realignment (green vertical lines). Applying to the simulated data the same positional-error analysis as for the recorded data (Fig. 3A&B), we observe the same qualitative picture, see Fig. 3C&D. In particular, position is coherently encoded in the recalled cognitive map at all times.

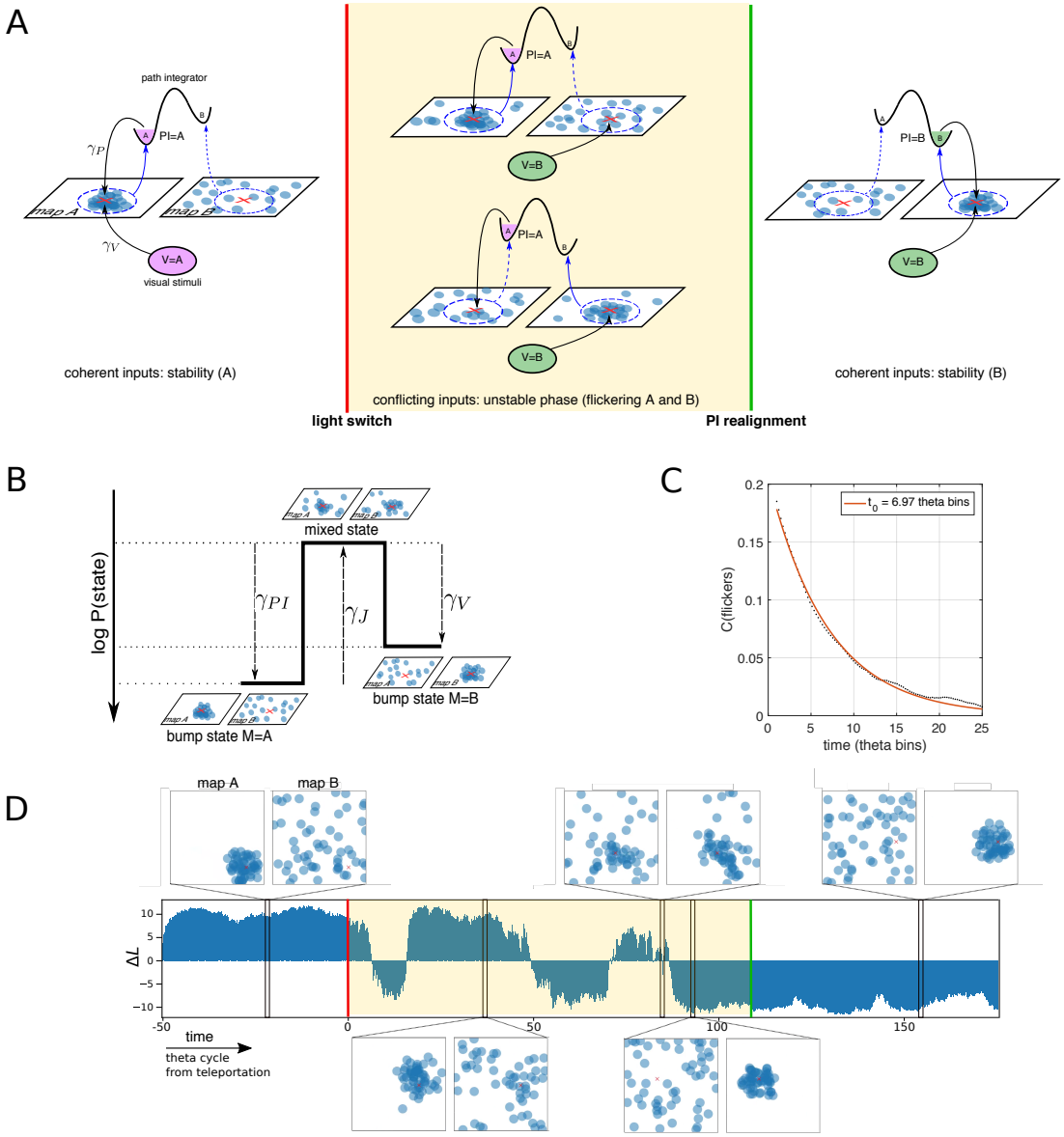

Fig. 4: **CANN model for interplay between path integrator, external stimuli, and memory. A. Phenomenology of the CANN model.** The model is composed of a recurrent hippocampal network that has memorized two cognitive maps (place fields dispositions) denominated A and B, a path integrator input (PI), and a visual input (V). In the left panel both PI and V are activating place cells whose place-field centers (shown by blue dots) correspond to the position of the rodent (red cross X) in the cognitive map A. The activity is said to be localized in a bump around X in t map A, while it appears as sparse and uninformative if interpreted with respect to the place-field locations in map B. A feedback projection (blue arrows) from the hippocampal state (bump) to the path-integrator state (purple) maintains the stability of the system by enforcing that the retrieved hippocampal map and the PI state agree. After the light conditions have been switched (teleportation, red line), V projects on place cells encoding position X in map B. The two hippocampal cognitive maps are therefore in conflict, and the bump of activity is alternatively localized in A (center-top) or in B (center- bottom). When the hippocampal activity is localized in the cognitive map B, the feedback projection tries to realign the internal state of PI along the corresponding map. Once the reaglinment has succeeded (green line), both inputs are back to a coherent state, and stability is reached in the cognitive map relative to the post-teleportation external light conditions. **B.** Representation of the effective model for the activity bump and effects of parameters. The input strengths, γ_{PI} and γ_V, contribute to push the hippocampal activity towards the corresponding cognitive states. Increasing the strength of recurrent connections, γ_J, results in an effective barrier separating the two collective hippocampal states, giving rise to well formed bumps in either map A (left) or in map B (right). Due to this effective trap the bump state remains localized in either map for more than a single theta bin. **C.** Temporal correlation of flickering events (theta bins with cognitive state opposite to external light conditions) decay over c.a. 7 theta bins in simulated data. Same model parameters as in Fig. 3. **D.** Time trace of the log-likelihood difference $\Delta\mathcal{L}$ (Eqn. (15) in Methods) in a simulated teleportation session; flickers can be observed during the conflicting phase following teleportation (shaded region). Prior to teleportation, and after the PI is realigned, inputs are coherent, and the system is stable: the sign of $\Delta\mathcal{L}$ is constant, mirroring the localization of the bump in one map. Screenshots of the activity projected on the two cognitive maps are shown for five different times. From left to right: bump localized in map A, bump localized in map A during the conflicting phase, mixed state during the conflicting phase, bump localized in map B during the conflicting phase, bump localized in map B. The video of the simulation can be found in SI.

Flickering frequency is constant throughout the conflicting phase, whose duration is exponentially distributed

Our model predicts that (1) the duration of the conflict phase, i.e. the time elapsed from a light switch to the subsequent PI realignment, is exponentially distributed (Fig. 5A, top panel); (2) during the conflict phase, the flickering frequency i.e. the percentage of theta bins identified as flickers, is constant and independent of time (Fig. 5B, top panel).

In order to test these two predictions on CA3 recordings, we introduce a method to disentangle the flickering dynamics of the cognitive map and the realignment of the PI, the latter bringing an end to the former. We first infer the most likely PI-realignment time for each light-switch event, given the sequence of identified flickers (Methods). The outcomes are shown as green lines in Fig. 5D, and correctly separate conflicting phases (rich in flickering events) from coherent periods (during which the hippocampal representation is much more stable). The distribution of conflicting phase durations is approximately exponential in agreement with model prediction (1), with decay time $\tau = 53$ theta bins (Fig. 5A, bottom panel). Dividing the test session into conflicting and coherent phases, we compute the frequency of flickers in the conflicting phase only. Consistently with the model prediction (2), the frequency of flickers is independent of the delay after the switch, with about 60% of theta bins in the conflicting phase carrying flickers (Fig. 5B).

Similar frequencies of flickers, close to one half, are obtained in the model when the two inputs have comparable strengths ($\gamma_{PI} \simeq \gamma_V$ in Fig. 4B, see also Supplementary Fig. 3). A testable consequence of this balance is that the distributions of the sojourn times (durations of the periods in which the neural activity persists in a cognitive map, see Methods) in map A and in map B are similar. This prediction is confirmed by a further analysis of the CA3 recordings: the two distributions of the sojourn times are both exponential, with roughly the same decay times (Supplementary Fig. 5). This common time scale is related to the correlation time of the flickers (Fig. 2C & 4C), see Methods.

The combination of properties (1) and (2) explains the exponential decay in the frequency of flickers with the delay after the switch reported in [26] and [36]. While the frequency of flickers is constant and large in the conflicting phase, and constant and very low in the coherent phase, the duration of the conflicting phase is exponentially distributed. Hence the frequency of flickering theta bins, irrespectively of the phase, shows the same exponential decay, see Fig. 5C (top panel: simulated experiment, bottom panel: analysis of CA3 recordings). A detailed analysis of the data provides overwhelming statistical support to our two-fold explanation compared to a simple exponential decay of the flickering frequency (logarithmic likelihood-ratio test ~ 150, see Methods).

Neural encoding of position reflects the presence of input mismatches

Our model allows us to better understand the subtle differences between the neural encodings of position in the conflicting and the coherent phases. In the latter phase, both path-integrator and visual inputs point to the neurons with place fields overlapping the rodent position $\mathbf{r}$ in map. During the conflicting phase, the two inputs excite the two place-cell populations centered in $\mathbf{r}$ in their respective maps, respectively, $m = PI$ and $m = V$. Hence, while the bump of activity is mostly localized in one of the two maps (varying over time), some dispersion may be expected due to these incoherent inputs.

Mixed activity states, in which two (distinct) populations of neurons encoding the same position in the two maps are active, can be occasionally observed in the snapshots of the simulated activity in Fig. 4D, e.g. around theta bin $t = 80$. The overdispersion present during the conflicting phase has two consequences. First, the accuracy in position encoding is expected to be lower in the conflicting phase that in the coherent phase, see Fig. 6A, left panel. Secondly, the loss in accuracy is not due to some random noise in the neural activity, but to a transient bump-like activity in the 'wrong' map, opposite to the decoded one. This effect is clearly seen when we choose the opposite map to infer the rodent position. While this choice leads to very poor prediction during the coherent phase, the positional error is significantly reduced during the conflicting phase (Fig. 6B, left panel).

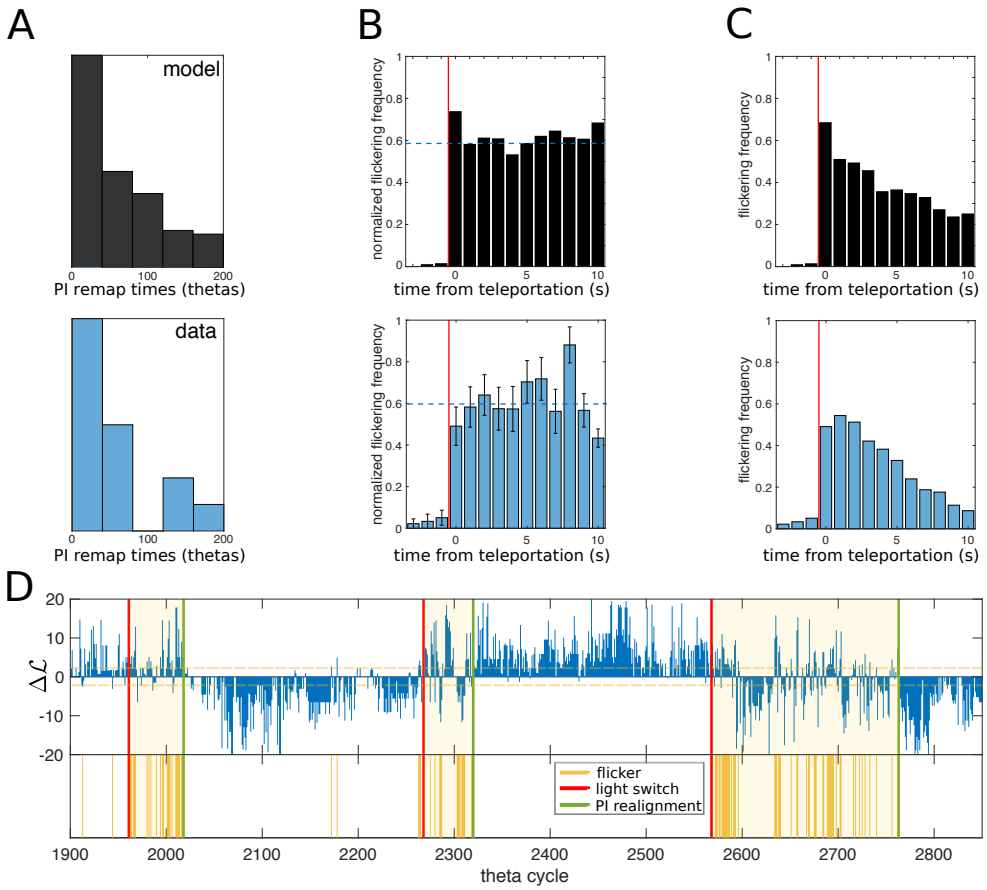

Fig. 5: **Flickering rate is homogenous within each conflicting period. A.** Distribution of the path-integrator realignment times in a simulated session (top) and inferred from the recorded data (bottom). **B.** Mean Frequency of Flickers (MFF) during the conflicting phase, binned over 8 theta bins (~ 1 second) intervals. MFF is constant during the conflicting phase, both in the model (top) and in recordings (bottom). Realignment times inferred from data were obtained with a Bayesian procedure (Methods); histograms show flickering frequency in each time bin t normalized with respect to the fraction of conflicting phases, out of 15 teleportation events, that survived at least up to time t. **C.** Convolution of the two distributions shown in panels A and B, i.e. the MFF computed on the full test session, shows an apparent exponential decay both in the model (top) and in data (bottom, similar to the analysis shown in [26]). **D.** Map decoder output $\Delta\mathcal{L}$ and inferred PI-realignment times for the experimental test session; 3 out of 15 teleportations are shown. Light switches are marked with red lines, inferred PI-realignment times are marked with green lines. Identified conflicting periods are shaded. Simulated data were obtained with the same model parameters as in Fig. 3.

To test these two predictions in CA3 recordings, we combine our positional analysis and our PI-realignment time inference procedure. In Fig. 6A (right), we compare the distributions of positional errors computed with the decoded map during conflicting and coherent phases (blue and red, respectively). Consistently with the model predictions, the positional error is significantly increased during the conflicting phase (ANOVA $p < 0.01$; conflicting: 13.4 ± 0.5 SEM, coherent: 12.1 ± 0.2 SEM). When computed with the opposite map, the positional error is obviously much higher than its counterpart computed with the decoded map, but a substantial decrease is found in the conflicting phase compared to the coherent phase, see Fig. 6B, right panel (ANOVA $p < 10^{-10}$; conflicting: 28.1 ± 0.7 SEM, coherent: 33.0 ± 0.3 SEM), in full agreement with the model prediction. This effect also explains the relatively

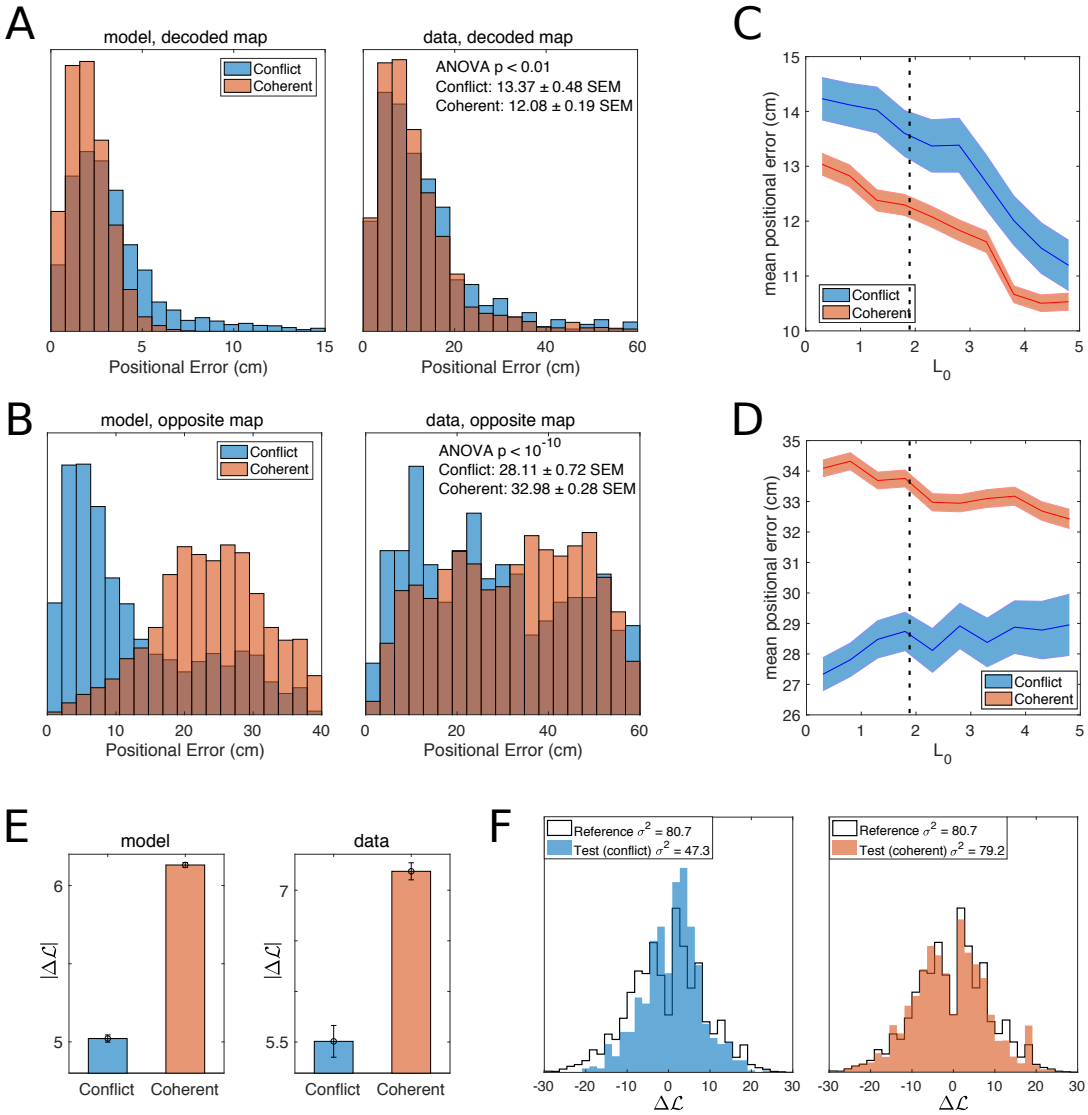

Fig. 6: **Distributions of positional errors in conflicting and coherent phases. A.** Distributions of positional errors in the decoded cognitive map during coherent/stable (red) and conflicting/unstable (blue) phases. The model predicts (top panel) that the mean error is significantly increased during conflicting periods, mirroring the dispersion of the neural representation induced by the conflicting input streams. The same phenomenology is observed in recordings (bottom panel; ANOVA $p < 0.01$; conflicting: 13.4 ± 0.5 SEM, coherent: 12.1 ± 0.2 SEM). **B.** The over-dispersion of the neural code in the 'correct' cognitive state (reported in panel A) is caused by an increased precision of the positional representation in the 'opposite' map, i.e. the one where the bump is not localized. This effect is significant in the model, top panel, and in the recorded data, bottom panel (ANOVA $p < 10^{-10}$; conflicting: 28.1 ± 0.7 SEM, coherent: 33.0 ± 0.3 SEM). **C, D.** Mean positional error (line) and standard error (shaded area) for CA3 data reported in panel A (position inferred according to the decoded map) and B (position inferred according to the opposite map) for a large range of values of the threshold L_0 of the map decoder; higher values of L_0 correspond to fewer theta bins retained in the analysis. Results shown in panels A and B were obtained with $L_0 = 2.3$. **E.** The absolute value of the log-ratio $\Delta\mathcal{L}$, Eqn. (1), is significantly reduced during the conflicting phase, both in the model (top) and in the data (bottom; ANOVA $p < 10^{-15}$; conflicting: 5.51 ± 0.16 SEM, coherent: 7.19 ± 0.08 SEM). $|\Delta\mathcal{L}|$ is a proxy for the completeness and stability of the bump (Methods). **F** Distributions of $\Delta\mathcal{L}$ in the test session during the conflicting (left, blue) and the coherent (right, red) phases, compared to the distribution in the two reference sessions (black contour, same as in Fig. 2A).

low value of the positional error obtained with the opposite map right after the switch, i.e. deep into the conflicting phase, compared to later times, see Fig. 3A&C. We stress that these two results are robust against changes in the statistical threshold L_0 for map decoding (Methods), as shown in Fig. 6C&D.

The over-dispersion of the neural bump during the conflicting phase can also be observed from the reduction in (the absolute value of) the log-ratio, $|\Delta\mathcal{L}|$, see Eqn. (1). This quantity can be interpreted as a proxy for the completeness of the bump in one single map (Methods), larger $|\Delta\mathcal{L}|$ corresponding to large bumps in either of the two maps and randomly scattered activity in the other map (Fig. 4A&D). We find that the absolute value of $\Delta\mathcal{L}$ is significantly reduced during the conflicting phase in CA3 data, see Fig. 6E (left panel, ANOVA $p < 10^{-15}$; conflicting: 5.51 ± 0.16 SEM, coherent: 7.19 ± 0.08 SEM). Figure 6F shows the bimodal nature of the distributions of $\Delta\mathcal{L}$ in the conflicting and coherent phases. While the reference and coherent-phase distributions coincide, the conflicting-phase distribution is more narrow, due to the overdispersion of the bump (reference $\sigma^2 = 80.7$, coherent $\sigma^2 = 79.2$, conflict $\sigma^2 = 47.3$). This result provides further evidence for the predictive power of the CANN model.

3 Discussion

Our statistical inference-based data analysis allows us to quantify how well the CA3 neural activity encode various cognitve maps, and the position therein. Correlation-based procedures, e.g. used in [26], decode the cognitive state by comparing the instantaneous population activity to the average activity recorded in reference sessions at the same position of the rodent. Our functional-network based map decoder, instead, relies on the fact that the joint pairwise spiking activity of neurons is a fingerprint of the cognitive map [27,32]. It does not need any knowledge of the sensory correlate (here, position), and could be used to decode generic brain states in other areas.

The fast dynamics of cognitive maps studied in [26] and here results from an unrealistic sensory situation. Imposing artificial conflicts between inputs and studying their consequences is a standard approach to unveil the circuitry underlying the processing of multimodal sensory information in the hippocampus [24, 25] as well as in other brains areas, see for instance [37] for an illustration in the primary visual cortex where mismatches involve sensory and motor inputs. However, fast retrieval of functionally relevant maps, characterized by grouping and cognitive control, has also been observed in realistic settings, in which a behaving animal is required to maintain representations of two distinct spatial frames [20].

The position of the animal was accurately inferred at all times from the spiking activity using the place fields of the retrieved cognitive map (Fig. 1B). As a main finding, we show that the hippocampus maintains high-quality encoding of the position even if the contextual variable undergoes fast dynamical changes. This is explained in the model by the fact that inputs point to place cells coding for the physical position in both competing maps (Fig. 4A), and that the bump of activity is most often localized around these place cells in one map, and scattered all over the other map. The emergence of unambiguous, non-mixed representations was also underlined in [26], and shown to take place in the second half of the theta cycle. However, the detailed analysis of the CA3 recordings and of the model data shows a loss of quality of the bump state (reduction in absolute value of log-ratio $|\Delta\mathcal{L}|$) and an increased quality of position decoding in the opposite map (Fig. 6), providing evidence for the presence of partially mixed states.

Our model for the retrieval of hippocampal cognitive maps in the presence of inputs from the path integrator and visual cues is based on CANN theory [30, 38]. Two-dimensional CANN attractors, were previously applied to networks of place [15, 17] and grid [39, 40] cells. Indirect experimental evidence supporting CANN is now accumulating in various animals and brain areas. Evidence of a ring-shaped attractor region associated to head direction representation was recently reported in the drosophila central brain [41]. Attractor dynamics has also been associated to behavioral observation in a study on the monkey prefrontal cortex [42].

As for space representation, experimental support for attractor behaviour has been found in hippocampal CA1 [43] as well in grid cell [44] recordings. Further indirect evidence is provided by the pattern of connectivity in CA3, compatible with its functional role as an auto-associative attractor network [13], as suggested long ago based on anatomical and computational considerations [12, 45], and by the active nature of dendrites of mEC neurons, which enhances the robustness of attractors under environmental changes [46].

The detailed analysis of the CA3 recordings done here provides another indirect support for CANN theoretical framework, when multiple (two) cognitive maps are memorized. Memorization of the two attractors is obtained, in the model, by adding the corresponding connectivity matrices into the unique CANN connectivity matrix [16–19]. A detailed theoretical study of the mechanisms for transition from map to map was obtained in the absence of inputs, i.e. for spontaneous transitions induced by neural noise only [34]. A similar picture is found here in the presence of visual-cue inputs pointing to the 'new' map, while the path-integrator inputs point to the 'old' map in the conflicting phase. As inputs are of comparable magnitude, no single map is favored. The stochastic fluctuations resulting from the noise of the individual neurons are sufficient for the system to cross the activation barrier between the two memory states (maps) of the network, see Fig. 4B and see Supplementary Fig. 1. The hippocampal network jumps intermittently from one cognitive map to the other, reproducing the flickering events experimentally identified and described in [26]. Transition rates between the two maps increase with the neural noise, modeled here by the parameter β, see Eqn. (14) in Methods, Supplementary Fig. 4, and Supplementary text. Neural noise relative to the population activity could also be effectively increased through the introduction of periodic modulations of the activity (theta rhythm) [35, 47]. Transitions could also be facilitated by particular 'confounding' landmarks or positions in space, where the maps happen to be locally similar [16, 34].

The present model reproduces accurately all the observed flickering properties, without any need for a post-learning short-term plasticity of the CA3 network hypothetized in [36]. In particular, our model predicts that the flickering frequency is independent from the time spent after the teleportation event in the conflicting phase (Fig. 5B). Accurate data analysis show that this prediction, when combined with the exponentially distributed realignment time of the path integrator (Fig. 5A), is much more likely than the hypothesis of (exponentially) decreasing flickering frequency put forward in [26] and in [36].

Our model is based on the existence of two streams of inputs conveying, respectively, external landmark and self-navigation information. Recent studies have pointed to the grid cells network in mEC as the possible region that supports path integration, as their firing patterns are maintained in the dark [48], and the relative phases of grid cells seem to be largely unaffected by global remapping between environments of similar shapes [11, 49]. CANN-based approaches have been proposed to model grid-cell networks [39, 40], differing from hippocampal CANN mostly by the short-range nature of the inhibitory couplings. In much the same way the microscopic hippocampal CANN proposed here can effectively be reduced to a 2-state model (Fig. 4B), we expect CANN models for the grid-cell networks to be approximately described by a 2-state model, corresponding to the PI aligned with map A or B [11, 50, 51]. This motivates the simple model for the PI we have considered here.

In addition to sending projections towards the hippocampus, our model PI receives a feedback from the CANN, greatly increasing the probability of transition to the state agreeing with the instantaneous cognitive map [30, 52, 53]. Eventually, the state of the PI is realigned along the visual cue inputs, which stops the conflicting phase. Our model effectively implements a ratchet mechanism, locking the system into the coherent phase after a conflicting transient. Realignment of the path integrator based on visual landmarks is an important functional property, intended to limit the accumulation of errors in position estimation [54], and observed for large mismatch between external and internal inputs [23]. From a physiological point of view, projections exist from CA1 to mEC [30], and have been shown to be important for the formation of grid cells [53]. Hence, the feedback from the CANN, thought here to model CA3 activity, to the PI should be understood as effective.

As recently reported in [49], the impairement of the mEC grid firing resulted in a loss of path integrator in behaving rodents. As in our model, the recall of the pre-teleportation map, and, therefore,

the whole flickering phenomenology are driven by the input stream from the path integrator to the CA3 network, we conjecture that flickering instabilities would disappear upon grid-cell impairement. Simultaneous recordings of mEC and CA3, as in [11], would be extremely useful to test our predictions and better describe the effect of the path integrator on the cognitive status of CA3.

Acknowledgments: We are grateful to K. Jezek for providing us with the data of [26] and for fruitful discussions. We thank C. Schmidt-Hieber for insightful discussions and comments. This work was partly supported by the HFSP project RGP0057–2016.

References

1. J. O'Keefe, "Place units in the hippocampus of the freely moving rat," Experimental neurology, vol. 51, no. 1, pp. 78–109, 1976.
2. J. R. Manns and H. Eichenbaum, "A cognitive map for object memory in the hippocampus," Learning & Memory, vol. 16, no. 10, pp. 616–624, 2009.
3. J. O'keefe and D. Conway, "Hippocampal place units in the freely moving rat: why they fire where they fire," Experimental brain research, vol. 31, no. 4, pp. 573–590, 1978.
4. J. O'Keefe and A. Speakman, "Single unit activity in the rat hippocampus during a spatial memory task," Experimental brain research, vol. 68, no. 1, pp. 1–27, 1987.
5. R. U. Muller and J. L. Kubie, "The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells," Journal of Neuroscience, vol. 7, no. 7, pp. 1951–1968, 1987.
6. M. I. Anderson and K. J. Jeffery, "Heterogeneous modulation of place cell firing by changes in context," Journal of Neuroscience, vol. 23, no. 26, pp. 8827–8835, 2003.
7. D. M. Smith and S. J. Mizumori, "Hippocampal place cells, context, and episodic memory," Hippocampus, vol. 16, no. 9, pp. 716–729, 2006.
8. T. A. Allen, D. M. Salz, S. McKenzie, and N. J. Fortin, "Nonspatial sequence coding in ca1 neurons," Journal of Neuroscience, vol. 36, no. 5, pp. 1547–1563, 2016.
9. E. Bostock, R. U. Muller, and J. L. Kubie, "Experience-dependent modifications of hippocampal place cell firing," Hippocampus, vol. 1, no. 2, pp. 193–205, 1991.
10. S. Leutgeb, J. K. Leutgeb, C. A. Barnes, E. I. Moser, B. L. McNaughton, and M.-B. Moser, "Independent codes for spatial and episodic memory in hippocampal neuronal ensembles," Science, vol. 309, no. 5734, pp. 619–623, 2005.
11. M. Fyhn, T. Hafting, A. Treves, M.-B. Moser, and E. I. Moser, "Hippocampal remapping and grid realignment in entorhinal cortex," Nature, vol. 446, no. 7132, pp. 190–194, 2007.
12. E. T. Rolls, "An attractor network in the hippocampus: theory and neurophysiology," Learning & Memory, vol. 14, no. 11, pp. 714–731, 2007.
13. S. J. Guzman, A. Schlögl, M. Frotscher, and P. Jonas, "Synaptic mechanisms of pattern completion in the hippocampal ca3 network," Science, vol. 353, pp. 1117–1123, 2016.
14. S.-i. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," Biological cybernetics, vol. 27, no. 2, pp. 77–87, 1977.
15. M. Tsodyks and T. Sejnowski, "Associative memory and hippocampal place cells," International journal of neural systems, vol. 6, pp. 81–86, 1995.
16. A. Samsonovich and B. L. McNaughton, "Path integration and cognitive mapping in a continuous attractor neural network model," Journal of Neuroscience, vol. 17, no. 15, pp. 5900–5920, 1997.
17. F. P. Battaglia and A. Treves, "Attractor neural networks storing multiple space representations: a model for hippocampal place fields," Physical Review E, vol. 58, no. 6, p. 7738, 1998.
18. J. J. Hopfield, "Neurodynamics of mental exploration," Proceedings of the National Academy of Sciences, vol. 107, no. 4, pp. 1648–1653, 2010.
19. R. Monasson and S. Rosay, "Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Phase diagram," Physical review E, vol. 87, no. 6, p. 062813, 2013.
20. E. Kelemen and A. A. Fenton, "Dynamic grouping of hippocampal neural activity during cognitive control of two spatial frames," PLoS biology, vol. 8, no. 6, p. e1000403, 2010.
21. B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser, "Path integration and the neural basis of the'cognitive map'," Nature Reviews Neuroscience, vol. 7, no. 8, pp. 663–678, 2006.
22. M.-L. Mittelstaedt and H. Mittelstaedt, "Homing by path integration in a mammal," Naturwissenschaften, vol. 67, no. 11, pp. 566–567, 1980.
23. K. M. Gothard, W. E. Skaggs, and B. L. McNaughton, "Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues," Journal of Neuroscience, vol. 16, no. 24, pp. 8027–8040, 1996.
24. G. Chen, J. A. King, N. Burgess, and J. O'Keefe, "How vision and movement combine in the hippocampal place code," Proceedings of the National Academy of Sciences, vol. 110, no. 1, pp. 378–383, 2013.
25. O. Haas, J. Henke, C. Lebold, and K. Thurley, "Distinct subpopulations of locomotion- and vision-induced place fields in the hippocampus," Submitted, vol. 16, no. 24, pp. 8027–8040, 2019.

26. K. Jezek, E. J. Henriksen, A. Treves, E. I. Moser, and M.-B. Moser, "Theta-paced flickering between place-cell maps in the hippocampus," Nature, vol. 478, no. 7368, pp. 246–249, 2011.
27. L. Posani, S. Cocco, K. Ježek, and R. Monasson, "Functional connectivity models for decoding of spatial representations from hippocampal ca1 recordings," Journal of Computational Neuroscience, pp. 1–17, 2017.
28. K. Zhang, I. Ginzburg, B. L. McNaughton, and T. J. Sejnowski, "Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells," Journal of neurophysiology, vol. 79, no. 2, pp. 1017–1044, 1998.
29. E. N. Brown, L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson, "A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells," Journal of Neuroscience, vol. 18, no. 18, pp. 7411–7425, 1998.
30. C. Barry and N. Burgess, "Neural mechanisms of self-location," Current Biology, vol. 24, no. 8, pp. R330–R339, 2014.
31. S. Cocco, R. Monasson, L. Posani, S. Rosay, and J. Tubiana, "Statistical physics and representations in real and artificial neural networks," Physica A: Statistical Mechanics and its Applications, p. doi/10.1016/j.physa.2017.11.153, 2017.
32. S. Cocco, R. Monasson, L. Posani, and G. Tavoni, "Functional networks from inverse modeling of neural population activity," Current Opinion in Systems Biology, 2017.
33. E. Schneidmann, M. J. Berry, R. Segev, and W. Bialek, "Weak pairwise correlations imply strongly correlated network states in a population," Nature, vol. 440, pp. 1007–1012, 2006.
34. R. Monasson and S. Rosay, "Transitions between spatial attractors in place-cell models," Physical review letters, vol. 115, no. 9, p. 098101, 2015.
35. R. Monasson and S. 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, vol. 89, no. 3, p. 032803, 2014.
36. S. Mark, S. Romani, K. Jezek, and M. Tsodyks, "Theta-paced flickering between place-cell maps in the hippocampus: A model based on short-term synaptic plasticity," Hippocampus, 2017.
37. G. B. Keller, T. Bonhoeffer, and M. Hübener, "Sensorimotor mismatch signals in primary visual cortex of the behaving mouse," Neuron, vol. 74, no. 5, pp. 809–815, 2012.
38. D. Bush, C. Barry, and N. Burgess, "What do grid cells contribute to place cell firing?," Trends in neurosciences, vol. 37, no. 3, pp. 136–145, 2014.
39. M. C. Fuhs and D. S. Touretzky, "A spin glass model of path integration in rat medial entorhinal cortex," Journal of Neuroscience, vol. 26, no. 16, pp. 4266–4276, 2006.
40. Y. Burak and I. R. Fiete, "Accurate path integration in continuous attractor network models of grid cells," PLoS computational biology, vol. 5, no. 2, p. e1000291, 2009.
41. S. S. Kim, H. Rouault, S. Druckmann, and V. Jayaraman, "Ring attractor dynamics in the drosophila central brain," Science, vol. 356, no. 6340, pp. 849–853, 2017.
42. K. Wimmer, D. Q. Nykamp, C. Constantinidis, and A. Compte, "Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory," Nature neuroscience, vol. 17, no. 3, pp. 431–439, 2014.
43. T. J. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'keefe, "Attractor dynamics in the hippocampal representation of the local environment," Science, vol. 308, no. 5723, pp. 873–876, 2005.
44. K. Yoon, M. A. Buice, C. Barry, R. Hayman, N. Burgess, and I. R. Fiete, "Specific evidence of low-dimensional continuous attractor dynamics in grid cells," Nature neuroscience, vol. 16, no. 8, pp. 1077–1084, 2013.
45. A. Treves and E. T. Rolls, "Computational analysis of the role of the hippocampus in memory," Hippocampus, vol. 4, no. 3, pp. 374–391, 1994.
46. C. Schmidt-Hieber, G. Toleikyte, L. Aitchison, A. Roth, B. A. Clark, T. Branco, and M. Häusser, "Active dendritic integration as a mechanism for robust and precise grid cell firing," Nature neuroscience, vol. 20, no. 8, p. 1114, 2017.
47. F. Stella, E. Cerasti, B. Si, K. Jezek, and A. Treves, "Self-organization of multiple spatial and context memories in the hippocampus," Neuroscience & Biobehavioral Reviews, vol. 36, no. 7, pp. 1609–1625, 2012.
48. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," Nature, vol. 436, no. 7052, pp. 801–806, 2005.
49. M. Gil, M. Ancau, M. I. Schlesiger, A. Neitz, K. Allen, R. J. De Marco, and H. Monyer, "Impaired path integration in mice with disrupted grid cell firing," Nature neuroscience, vol. 21, no. 1, p. 81, 2018.
50. F. Savelli, J. Luck, and J. J. Knierim, "Framing of grid cells within and beyond navigation boundaries," eLife, vol. 6, 2017.
51. J. D. Monaco and L. F. Abbott, "Modular realignment of entorhinal grid cell activity as a basis for hippocampal remapping," Journal of Neuroscience, vol. 31, no. 25, pp. 9414–9425, 2011.
52. N. M. Van Strien, N. Cappaert, and M. P. Witter, "The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network," Nature Reviews Neuroscience, vol. 10, no. 4, pp. 272–282, 2009.
53. T. Bonnevie, B. Dunn, M. Fyhn, T. Hafting, D. Derdikman, J. L. Kubie, Y. Roudi, E. I. Moser, and M.-B. Moser, "Grid cells require excitatory drive from the hippocampus," Nature neuroscience, vol. 16, no. 3, pp. 309–317, 2013.
54. K. Hardcastle, S. Ganguli, and L. M. Giocomo, "Environmental boundaries as an error correction mechanism for grid cells," Neuron, vol. 86, no. 3, pp. 827–839, 2015.

Methods

Cognitive map decoding from neural activity

Theta bins are identified with the Hilbert transform procedure of [1]. We binarize the activity of the N recorded neurons into each theta bin t: $s_{i,t} = 1$ if neuron i is active in bin t, 0 otherwise. We then infer for each cognitive map $m = A, B$ a probability distribution $P_m(\mathbf{s})$ for the activity configurations $\mathbf{s} = (s_1, s_2, ..., s_N)$. The distribution corresponds to the so-called Ising (graphical) model in the statistical physics literature, and reads

$$P_m(\mathbf{s}) = \frac{1}{\mathcal{Z}_m} \exp\left(\sum_i h_i^{(m)} s_i + \sum_{i<j} J_{ij}^{(m)} s_i s_j\right) , \tag{1}$$

where $\mathcal{Z}_m$ is a normalization constant, ensuring that the sum of $P_m(\mathbf{s})$ over all 2^N activity configurations $\mathbf{s}$ is equal to unity. The functional coupling $J_{ij}^{(m)}$ accounts for the increase in log probability that cell i is active when cell j is firing (in the same theta bin). The parameter $h_i^{(m)}$ is a local input onto neuron i, and is an increasing function of the average activity of the cell during the environment exploration. The N inputs h_i^m and $\frac{1}{2}N(N-1)$ couplings $J_{ij}^{(m)}$ are computed to match the N single-cell activities $\langle s_i \rangle$ and the $\frac{1}{2}N(N-1)$ pairwise activities $\langle s_i s_j \rangle$, where $\langle \cdot \rangle$ denotes the average over the reference session for map m. This was done with the Adaptive Cluster Expansion algorithm, see [2–5].

Once inferred, model (1) can be used to score (compute the probability of) any activity configuration of the test session, not necessarily included in the configurations recorded during the reference sessions. The Ising model was shown, first on retinal ganglion cell recordings, and, more recently, on prefrontal cortex [6] and hippocampal data [7,8], to provide a good approximation for the distribution of population activity configurations. The performance in the decoding task has been shown to be superior to rate-based decoders on CA1 data [7]. Here we applied the same methods to test the decoder on CA3 data, and it showed performances comparable to classical rate-based methods (see Supplementary text and Supplementary Fig. 9).

Two distributions, $P_A(\mathbf{s})$ and $P_B(\mathbf{s})$, are inferred from, respectively, the reference sessions in environments A and B. Note that, due to the presence of global remapping in CA3, the functional connectivity matrices $J^{(A)}$ and $J^{(B)}$ are largely different, and $\Delta\mathcal{L}$ is substantially different from zero for most activity configurations $\mathbf{s}$. We can then simply compute the log-likelihood difference $\Delta\mathcal{L}$ in Eqn. (1) given the population activity $\mathbf{s}_t$ in the theta bin t, and decode the corresponding map as follows:

$$m_t = \begin{cases} \text{A} & \text{if } \Delta\mathcal{L}(\mathbf{s}_t) > L_0 \ , \\ \text{B} & \text{if } \Delta\mathcal{L}(\mathbf{s}_t) < -L_0 \ , \end{cases} \tag{2}$$

where the threshold L_0 is chosen according to the required statistical confidence. We generally set $L_0 = \log 10 \simeq 2.3$.

Temporal correlation of flickers and sojourn times

After having decoded the map m_t in theta bin t, we define the flicker variable f_t, equal to 1 if m_t does not match the light cue in theta bin t, to 0 otherwise. The time correlation at delay τ is defined as

$$C(\tau) = \frac{1}{T_{tot}} \sum_{i=1}^{S-1} \sum_{t=T_i}^{T_{i+1}-\tau} f_t f_{t+\tau} - \left(\frac{1}{T_{tot}} \sum_{i=1}^{S-1} \sum_{t=T_i}^{T_{i+1}-\tau} f_t\right)^2 \tag{3}$$

where S is the total number of switch events in the recorded data ($S = 16$ in [1]), T_i is the theta bin index of switch i ($< S$), and $T_{tot} = T_S$ is the total number of theta bins in the test session. The time correlation $C(\tau)$ is typically exponentially decaying, with a decay time τ_0, see Fig. 2C.

The correlation time τ_0 is related to the sojourn times of the neural bump in the cognitive maps. We define a sojourn as a sequence of contiguous theta bins that are all decoded in the same map. Theta bins that are not statistically significant, i.e. whose absolute values of $\Delta\mathcal{L}$ are lower than the threshold L_0, are considered as belonging to the same map of the last statistical significant bin. The distribution of sojourn times in each map is shown in Supplementary Fig. 5.

The relationship between correlation and sojourn time can be established by assuming a Markovian dynamics for the 2-state model ($m = A$ or $m = B$) evolving in discrete time. The dynamics is determined by the map transition probabilities from one time bin to the next:

$$\begin{cases} p_{A\to A} = e^{-1/\tau_A} \\ p_{A\to B} = 1 - e^{-1/\tau_A} \\ p_{B\to B} = e^{-1/\tau_B} \\ p_{B\to A} = 1 - e^{-1/\tau_B} \end{cases} \tag{4}$$

where τ_A and τ_B are the mean sojourn times in, respectively, map A and B. A straightforward calculation shows that the time correlation $C(\tau)$ between the map state at times t and $t+\tau$ is decreases exponentially with the delay τ only, with an average time equal to

$$\tau_0 = -\left[\log\left(e^{-1/\tau_A} + e^{-1/\tau_B} - 1\right)\right]^{-1} . \tag{5}$$

Hence, the correlation time τ_0 is approximately given by the smaller mean sojourn time among τ_A and τ_B.

Position decoding from neural activity

Position is inferred by inverting a the simplest binary firing model for neural activity at a given position. The arena is discretized into 60×60 squared bins of 1 cm^2 each, with integer coordinates (x, y) [1]. For each reference environment $m \in$ (A, B) we construct the binary rate map, $p_i^{(m)}(x, y)$, equal to the average of $s_{i,t}$ over all theta bins t in which the rat is at position (x, y). Position is then decoded according to the naive Bayesian framework [9]: the probability of the activity configuration $\mathbf{s}_t = \{s_1, s_2, ..., s_n\}$ in theta bin t and at <u>fixed</u> position (x, y) reads

$$P_m(\mathbf{s}_t|x, y) = \prod_{i=1}^{N} \left[p_i^{(m)}(x, y) \cdot s_i + \left(1 - p_i^{(m)}(x, y)\right) \cdot (1 - s_i)\right] . \tag{6}$$

Once m is known, e.g. using the map decoder, the position of the rodent can be reconstructed from the recorded neural activity through

$$(\hat{x}_t, \hat{y}_t) = \arg\max_{(x,y)} \left[P_m(\mathbf{s}_t|x, y) \times T_m(x, y)\right] , \tag{7}$$

where the maximum is computed over the 60×60 possible positions. $T_m(x, y)$ above is the number of theta bins spent by the rodent at position (x, y) during the reference session of map m; we use it as a prior to favor positions where the rodent is more likely to be, irrespectively of the neural activity.

A crucial role is played by the variable $m = A$ or B in Eqn. (7), i.e. by the set of rate maps used to invert the firing model and retrieve the position from the recorded neural activity in theta bin t. In reference sessions, external conditions are known and stable, and m is constantly set to the corresponding

environment. In the test sessions, light conditions are changed and the internal representation is unstable and alternates between the two maps. We then use the above map decoder to infer the value of m_t, and plug it into Eqn. (7). As our map decoder does not use any information about the rodent position, this procedure allows us to obtain a fair estimate for the position in the retrieved map. Alternatively, we can choose the other map (called opposite map) instead of the decoded one to investigate how position is represented during the conflicting phase (Results).

Continuous Attractor Neural Network model for hippocampal activity

The hippocampal population includes N place cells. For each cell $i = 1...N$ the place-field centers coordinates, $\mathbf{r}_i^A$ and $\mathbf{r}_i^B$, are drawn uniformly and independently at random in the squared environments, respectively, A and B. The linear size of each square is denoted by L.

Neural activities are represented by binary variables: $s_{i,t} = 0$ or 1 if neuron i is, respectively, silent or active in time bin $t = 1, 2, 3, ...$ The duration of a time bin is the theta cycle over K; results reported here were obtained with $K = 4$, which corresponds to approximately 30 ms.

The total input received by neuron i at time t is

$$H_{i,t} = \sum_{j \neq i} J_{ij}\, s_{j,t} + h_i^{(V)}(\mathbf{r}) + h_i^{(PI)}(\mathbf{r}) \ . \tag{8}$$

The three terms on the right hand side of Eqn. (8) represent, in order:

- the input due to recurrent connections in the hippocampal network. The underlying assumption is that connections have emerged from learning during the exploration of the two environments by the rodent. Place cells that turned out to be simultaneously active in either environment have developed positive couplings. Couplings are defined through

$$J_{ij} = J_{ij}^A + J_{ij}^B \quad \text{with} \quad \begin{cases} J_{ij}^A = \gamma_J \times \phi(\mathbf{r}_i^A - \mathbf{r}_j^A) \\ J_{ij}^B = \gamma_J \times \phi(\mathbf{r}_i^B - \mathbf{r}_j^B) \end{cases} , \tag{9}$$

  where γ_J controls the strength of the connections, and

$$\phi(\mathbf{r}) = \frac{L^2}{N \times 2\pi\sigma^2}\, e^{-\frac{|\mathbf{r}|^2}{2\sigma^2}} \ . \tag{10}$$

  Parameter σ in Eqn. (10) is the spatial scale corresponding to the width of the place fields. The prefactor in eqn. (10) ensures that ϕ is dimensionless and that, on average, the sum of the Gaussian factors $\phi(\mathbf{r} - \mathbf{r}_i^m)$ over all neurons i is close to unity for every possible position $\mathbf{r}$ and map m:

$$\sum_{i=1}^{N} \phi(\mathbf{r} - \mathbf{r}_i^m) \sim 1 \qquad \forall \mathbf{r}, \forall m \ . \tag{11}$$

- the visual input

$$h_i^{(V)}(\mathbf{r}) = \gamma_V \times \begin{cases} \phi(\mathbf{r}_i^A - \mathbf{r}) & \text{if } V = A \ , \\ \phi(\mathbf{r}_i^B - \mathbf{r}) & \text{if } V = B \ , \end{cases} \tag{12}$$

  depends on the position $\mathbf{r}$ of the rodent. We again assume that, during the exploration of the environment, visual-cue projections onto place cells have been strenghtened through learning. For simplicity we use the same function ϕ as in the recurrent connections, see Eqn. (9), to characterize the portion of environment in which visual cues project onto a specific place cell i.

– the path-integrator input

$$h_i^{(PI)}(\mathbf{r}) = \gamma_{PI} \times \begin{cases} \phi(\mathbf{r}_i^A - \mathbf{r}) & \text{if } PI = A \ , \\ \phi(\mathbf{r}_i^B - \mathbf{r}) & \text{if } PI = B \ . \end{cases} \tag{13}$$

PI inputs have the same functional dependence over space as visual-cue related inputs. The amplitudes of both input types are tuned by the parameters γ_{PI} and γ_V.

All neurons undergo stochastic updating of their activities from time bin $t \to t+1$ according to their total inputs. The activity of neuron i at time $t+1$ is chosen to be

$$s_{i,t+1} = 0 \text{ with probability } \frac{1}{1+e^{\beta(H_{i,t}-\theta)}} \ , \quad 1 \text{ with probability } \frac{e^{\beta(H_{i,t}-\theta)}}{1+e^{\beta(H_{i,t}-\theta)}} \ . \tag{14}$$

The value of the threshold θ is dynamically adjusted so that an average fraction f of the neurons is active at any time. This is a simple and effective way to enforce global inhibition in the population activity. Parameter β controls the amount of noise in the neural dynamics. For $\beta \to 0$ neuron activities are random and independent of their inputs, while, for $\beta \to \infty$, they deterministically follow the signs of the inputs (after substraction of the threshold θ). The average activity of cell i at time $t+1$ is therefore a monotonously increasing sigmoidal function of its total input $H_{i,t}$ at time t, with maximal slope equal to $\beta/4$ in $H_{i,t} = \theta$.

The properties of this CANN model in the absence of any visual and PI inputs, i.e. for $\gamma_V = \gamma_{PI} = 0$, were analytically studied in [10–12]. A bump of activity can form and diffuse in either map if β is larger than some critical value β_c, and occasionally 'jump' from one map to the other. The mechanism for the jumps was studied in detail. For high values of β, jumps take place between preferred positions in both environments (wormholes in [13]), where maps happen to be locally similar by chance. For lower β (but larger than β_c), the bump may transiently disappear from one map and 'condensate' back in the other map. In the model presented here, due to the presence of inputs related to the rodent position $\mathbf{r}$, jumps take place between bump state coding for $\mathbf{r}$ in both maps.

There is a deep connection between our CANN model and our functional-network-based decoder. If the rodent position $\mathbf{r}$ is kept fixed, the dynamics of the CANN model will generate stochastically activity configurations $\mathbf{s}$ according to the Ising distribution in Eqn. (1), with couplings given by Eqn. (9) and inputs $h_i = h_i^{(V)}(\mathbf{r}) + h_i^{(PI)}(\mathbf{r})$. When the bump is localized in map m ($= A$ or B), the distribution of stochastically generated activity configurations $\mathbf{s}$ can be approximated by $P^{(m)}(\mathbf{s})$ in Eqn. (1). The log-ratio $\Delta\mathcal{L}$ defined in Eqn. (1) for the decoding of cognitive maps has thus a direct counterpart in our CANN model as the difference between the contributions to the log-probability of an activity configuration $\mathbf{s}$ when the bump is localized in maps A and B,

$$\Delta\mathcal{L}(\mathbf{s}) = \sum_{i<j} (J_{ij}^A - J_{ij}^B)\, s_i\, s_j \ . \tag{15}$$

Mechanism for path-integrator realignment

The path integrator is described as a two-state model, $PI = A$ or B. Its dynamics is stochastic and Markovian: in each time bin t, the state PI can jump into state PI' with transition probabilities $R(PI \to PI')$, independently of the previous states. The feedback from the hippocampal network to the path integrator is expressed in the dependence of R on the hippocampal map M_t at time t. To favor transitions to the state PI' agreeing with the current map M_t, we introduce the following witness function for the presence of the bump in map $m = A, B$:

$$W^{(m)}(\mathbf{s}, \mathbf{r}) = \sum_{i=1}^{N} s_i \ \phi\left(\mathbf{r} - \mathbf{r}_i^m\right) \ , \tag{16}$$

where $\mathbf{s}$ and $\mathbf{r}$ are, respectively, the activity configuration and the position of the rodent at time t. Due to the normalization condition of ϕ in Eqn. (11), we expect $W^{(m)}$ to be close to one for the retrieved map $m = M_t$ and to be much smaller for the opposite map.

We impose the preference for realigning the path-integrator state in accordance with the hippocampal map through the ratio between the two reciprocal transition probabilities,

$$\frac{R(\ PI = B \to PI' = A\ )}{R(\ PI = A \to PI' = B\ )} = e^{\gamma_W \left(W^{(A)}(\mathbf{s},\mathbf{r}) - W^{(B)}(\mathbf{s},\mathbf{r})\right)} \ . \tag{17}$$

Here, γ_W is a positive parameter allowing us to tune the strength of the preference. If the hippocampal bump of activity is localized in, say, map A, the right hand side of Eqn. (17) will be strongly positive, and the probability of realigning the path integrator to $PI' = A$ will be much larger than the probability of the reciprocal transition.

A solution to the constraint expressed by Eqn. (17) is given by

$$\begin{aligned} R(\ PI = B \to PI' = A\ ) &= R_0 \times e^{\gamma_W \left(W^{(A)}(\mathbf{s},\mathbf{r}) - W^{(B)}(\mathbf{s},\mathbf{r})\right)/2} \ , \\ R(\ PI = A \to PI' = B\ ) &= R_0 \times e^{-\gamma_W \left(W^{(A)}(\mathbf{s},\mathbf{r}) - W^{(B)}(\mathbf{s},\mathbf{r})\right)/2} \ , \end{aligned} \tag{18}$$

where R_0 is a positive number. In the absence of bias ($\gamma_W = 0$), the inverse of R_0 may be interpreted as the average time scale between two realignments of the path-integrator state.

The model for the path-integrator dynamics is entirely defined by the transition probabilities in Eqn. (19) and the probability conservation identities:

$$\begin{aligned} R(\ PI = A \to PI' = A\ ) + R(\ PI = A \to PI' = B\ ) = 1 \ , \\ R(\ PI = B \to PI' = A\ ) + R(\ PI = B \to PI' = B\ ) = 1 \ . \end{aligned} \tag{19}$$

Inference of path-integrator realignment times

To identify the realignment times of the PI we first introduce a simple probabilistic model for the hippocampal representation to be incoherent with the light-cue conditions (flickering time bin) as a function of time elapsed after the switch. Defining $t = 0$ as the time bin corresponding to the light switch and T as the time bin corresponding to the next switch (end of analyzed data), the probability $p(t)$ for time bin t to be a flickering event reads:

$$p(t) = \begin{cases} p_0, & \text{if } 1 \le t \le \tau \ , \\ p_e, & \text{if } \tau + 1 \le t \le T \ . \end{cases} \tag{20}$$

Here, p_0 is the constant flickering probability, and p_e is the expected decoding error defined with the percentile used in the flickering identification. Let $f_t = 1$ if time bin t is identified as a flicker and $f_t = 0$ otherwise. We write the log-likelihood of the parameter τ as a function of the identified flickering sequence $\mathbf{f} = \{f_t\}$ as follows:

$$\begin{aligned} \log P(\mathbf{f} \mid \tau, p_0, p_e) = \log p_0 \times \sum_{t=1}^{\tau} f_t + \log(1 - p_0) \times \sum_{t=1}^{\tau} (1 - f_t) \\ + \log p_e \times \sum_{t=\tau+1}^{T} f_t + \log(1 - p_e) \times \sum_{t=\tau+1}^{T} (1 - f_t) \ . \end{aligned} \tag{21}$$

We then maximize this log-likelihood to infer the most likely value τ^* of the realignment time:

$$\tau^* = \arg\max_{\tau} \ \log P(\mathbf{f} \mid \tau, p_0, p_e) \ . \tag{22}$$

The procedure is repeated for all light switches.

The procedure above needs an input value for p_0, the probability of flickering during the conflicting phase, whose consistency can be checked a posteriori by computing the mean flickering frequency in the conflicting phase. In SI, Fig.6 we show that the a-posteriori average flickering frequency remains remarkably stable, around the self-consistent choice ~ 0.6, for any input value of p_0, with a slight dependence on the chosen L_0 threshold. The same value is observed in the model when the strength of PI and V projections are set to similar values ($\gamma_{PI}/\gamma_V \in [0.95, 1.05]$, see SI, Fig. 3.

Independence of frequency of flickers from delay after light switch

We consider two hypothesis:

(a) H_{decay} = the flickering probability depends on time as a decaying function that can be inferred from data. This hypothesis assumes that the average flickering frequency computed on the full test session (15 lght switches) and shown in Fig. 5C, is a good description of the flickering probability in time after every single light switch.

(b) $H_{constant}$ = the flickering probability is constant throughout the conflicting period of varying duration, which can be inferred from data. This hypothesis predicts that the decaying function in Fig. 5C is a consequence of averaging the flickering frequency over many sessions.

To maintain a fair comparison between the two hypothesis, following the Bayesian information criterion [14], we parametrize each model with the same number of variables.

For hypothesis H_{decay}, we estimate the flickering frequency as a function of time from the average frequency computed over the full test session (Fig. 5C, bottom), in bins of one second-width (8 theta cycles), up to 15 seconds after the light switch. For later delays (>15 s) the flickering frequency is set to a baseline error probability, $p_e = 0.01$. We therefore obtain a step-wise decaying flickering probability, p_t, constant over groups of 8 contiguous time bins.

For hypothesis $H_{constant}$, we infer the most likely PI realignment times for each one of the 15 light-switch events (see Section above). The associated flickering probability p_t is then set to $p_0 = 0.55$ until the inferred PI realignment time τ^*, and equal to the baseline probability $p_e = 0.01$ afterwards. Hence, the time course of p_t depends on the light-switch event.

We then compute the likelihoods of both hypothesis given the observed data (identified flickering theta bins $\mathbf{f}$) through

$$\ell(\text{hypothesis}|\text{data}) = \sum_{t=1}^{T} \left[\log p_t \times f_t + \log(1 - p_t) \times (1 - f_t) \right] , \tag{23}$$

where T is the total length of the test session. The above expression is then summed over all light-switch events. We define the difference of the two log-likelihoods as

$$\Delta\ell = \ell(H_{constant}|\text{data}) - \ell(H_{decay}|\text{data}) \ . \tag{24}$$

The constant flickering frequency hypothesis $H_{constant}$ is extremely more likely ($\Delta\ell \sim 150$) than the decaying model H_{decay}. The result is robust against changes in the parameters, see Supplementary Fig. 6. For instance, we obtain $\Delta\ell \simeq 170$ and $\Delta\ell \simeq 60$ when the flickering identification is done based on, respectively, a less ($L_0 = \log 2$) and more ($L_0 = \log 100$) restrictive criterion. Similarly, the log-likelihood difference between the two hypothesis remains very large and positive if we change the constant-rate model p_0 value, e.g. $\Delta\ell \simeq 160$ for $p_0 = 0.4$, $\Delta\ell \simeq 120, 33$ for $p_{inf} = 0.1, 0.001$, or if we extend the definition of H_{decay} up to 30 seconds after the switch (instead of 15 seconds): $\Delta\ell \simeq 125$. In many teleportation events the flickering-dense area is indeed too short (too few flickers) or too long (too many flickers occurring far away the teleportation time) to be explained by the H_{decay} hypothesis, see Supplementary Fig. 7.

References

1. K. Jezek, E. J. Henriksen, A. Treves, E. I. Moser, and M.-B. Moser, "Theta-paced flickering between place-cell maps in the hippocampus," Nature, vol. 478, no. 7368, pp. 246–249, 2011.
2. S. Cocco and R. Monasson, "Adaptive cluster expansion for the inverse ising problem: convergence, algorithm and tests," J. Stat. Phys, vol. 147, no. 2, pp. 252–314, 2012.
3. J. P. Barton and S. Cocco, "Ising models for neural activity inferred via selective cluster expansion: Structural and coding properties," J Stat Mech, p. P03002, 2013.
4. J. Barton, E. De Leonardis, A. Coucke, and S. Cocco, "Ace: adaptive cluster expansion for maximum entropy graphical model inference," Bioinformatics, 2016.
5. S. Cocco and R. Monasson, "Adaptive cluster expansion for inferring boltzmann machines with noisy data," Physical review letters, vol. 106, no. 9, p. 090601, 2011.
6. G. Tavoni, S. Cocco, and R. Monasson, "Neural assemblies revealed by inferred connectivity-based models of prefrontal cortex recordings," Journal of computational neuroscience, vol. 41, no. 3, pp. 269–293, 2016.
7. L. Posani, S. Cocco, K. Ježek, and R. Monasson, "Functional connectivity models for decoding of spatial representations from hippocampal ca1 recordings," Journal of Computational Neuroscience, pp. 1–17, 2017.
8. L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, "Collective behavior of place and non-place neurons in the hippocampal network," Neuron, vol. 96, no. 2, pp. 1178–1191, 2017.
9. D. J. MacKay, Information Theory, Inference and Learning Algorithms. Cambridge University Press, 2003.
10. R. Monasson and S. Rosay, "Crosstalk and transitions between multiple spatial maps in an attractor neural network model of the hippocampus: Phase diagram," Physical review E, vol. 87, no. 6, p. 062813, 2013.
11. R. Monasson and S. 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, vol. 89, no. 3, p. 032803, 2014.
12. R. Monasson and S. Rosay, "Transitions between spatial attractors in place-cell models," Physical review letters, vol. 115, no. 9, p. 098101, 2015.
13. A. Samsonovich and B. L. McNaughton, "Path integration and cognitive mapping in a continuous attractor neural network model," Journal of Neuroscience, vol. 17, no. 15, pp. 5900–5920, 1997.
14. H. Akaike, "A new look at the statistical model identification," IEEE transactions on automatic control, vol. 19, no. 6, pp. 716–723, 1974.