A Bayesian perspective on the ring attractor for heading-direction tracking in the *Drosophila* central complex

Anna Kutschireiter[1]*, Melanie A Basnak[1], Rachel I Wilson[1], Jan Drugowitsch[1]

[1]Harvard Medical School, Department of Neurobiology, 200 Longwood Avenue, Boston, MA-02115, United States.

*Correspondence: anna_kutschireiter@hms.harvard.edu, jan_drugowitsch@hms.harvard.edu

Summary

Efficient navigation requires animals to track their position, velocity and heading direction (HD). Bayesian inference provides a principled framework for estimating these quantities from unreliable sensory observations, yet little is known about how and where Bayesian algorithms could be implemented in the brain's neural networks. Here, we propose a class of recurrent neural networks that track both a dynamic HD estimate and its associated uncertainty. They do so according to a circular Kalman filter, a statistically optimal algorithm for circular estimation. Our network generalizes standard ring attractor models by encoding uncertainty in the amplitude of a bump of neural activity. More generally, we show that near-Bayesian integration is inherent in ring attractor networks, as long as their connectivity strength allows them to sufficiently deviate from the attractor state. Furthermore, we identified the basic network motifs that are required to implement Bayesian inference, and show that these motifs are present in the *Drosophila* HD system connectome. Overall, our work demonstrates that the *Drosophila* HD system can in principle implement a dynamic Bayesian inference algorithm in a biologically plausible manner, consistent with recent findings that suggest ring-attractor dynamics underlie the *Drosophila* HD system.

Keywords

Drosophila; Central complex; Heading direction neurons; Ring attractor networks; Bayesian inference; Kalman filtering; Population coding; Dynamic inference

Introduction

Navigating a foreign city without a map requires continuous monitoring of one's current heading direction (HD). If we attempt to track our heading solely by 'dead reckoning' (integrating all our turning maneuvers over time), our orientation estimate becomes increasingly less accurate over time. A glimpse at our phone's GPS can help to recalibrate this estimate, yet the GPS signal's

inaccuracy (e.g., in densely built-up areas) renders this calibration imperfect. Thus, efficient navigation requires active use of the involved uncertainties by dynamic Bayesian inference, combining our internal compass estimate with external heading cues, such as a GPS reading, by weighting each in proportion to their reliability.

That humans and other animals make strategic use of their uncertainty, consistent with a Bayesian strategy, has been demonstrated over a wide range of behaviours such as cue combination (Ernst and Banks 2002; Fetsch et al. 2009; Drugowitsch et al. 2014), decision making (Glaze, Kable, and Gold 2015) or visual object tracking (Bill et al. 2020). The homing behavior of desert ants (Merkle and Wehner 2010) is a prime example of an uncertainty-guided strategy in spatial navigation: the radius of their homing search is strategically adjusted to how far the ant has foraged from the nest, reflecting the uncertainty accumulated during path integration. To make strategic use of uncertainties in behavior, the brain needs to represent and use a subjective estimate of uncertainty for the underlying computations. How it does so, however, remains largely unknown (Pouget et al. 2013). In particular, navigation-related neuronal activity has been observed across several brain regions (e.g., (Moser, Kropff, and Moser 2008; Stefanini et al. 2020; Kropff et al. 2015; Taube, Muller, and Ranck 1990)), but whether and how the brain represents *uncertainty* in these navigational variables remains unclear. Recent theoretical effort postulates specific uncertainty representations for dynamic estimation (Beck, Latham, and Pouget 2011; Kutschireiter et al. 2017) and their use in brain computations, but does not clarify whether and how these proposals might map onto specific biological brain networks. By focusing on the sense of orientation as one of the simplest cognitive variables required for spatial navigation, our aim is (i) to understand whether and how neural network dynamics could implement dynamic Bayesian inference for HD tracking, and (ii) whether such a network implementation of probabilistic inference is supported by the connectivity in a specific brain network, namely the *Drosophila* HD system.

The dynamics of orientation-selective neurons, such as head- or heading-direction cells, are traditionally modeled by ring-attractor networks (Zhang 1996; Xie, Hahnloser, and Seung 2002; Skaggs et al. 1995). These networks are recurrent neural networks (RNNs) that feature a continuum of stable fixed points arranged on a topological ring in attractor space, cf. (Eliasmith 2007). Consequently, their dynamics allow them to maintain a unique representation of HD in their population activity even in the absence of inputs. In *Drosophila*, the dynamics of HD neurons in the central complex (CX) qualitatively match that of a ring-attractor network with local excitation and broad inhibition (Kim et al. 2017). Further, CX network topology features network motifs that support the implementation of a ring attractor network (Turner-Evans et al. 2020). However, classic ring attractor networks do not, by design, incorporate a notion of uncertainty and therefore appear unsuited to implement Bayesian inference. Thus, a critical step towards our goal is to identify whether certain signatures of Bayesian inference are nonetheless present in such ring-attractor networks, and which network motifs enable the attractor to do so.

The experimental accessibility and well-understood connectivity of the *Drosophila* CX (Scheffer et al. 2020; Hulse et al. 2021) make it a compelling model network to seek a biological implementation of Bayesian inference. Heading-direction cells in the CX are arranged as a ring, forming the so-called ellipsoid body (EB), and track the fly's HD with a localized increase in neural

activity (`bump') in a compass-like manner (Seelig and Jayaraman 2015; Green et al. 2017; Turner-Evans et al. 2017). Importantly, this internal compass integrates different types of sensory observations such as visual input (Seelig and Jayaraman 2013; Omoto et al. 2017; Y. Sun et al. 2017), wind cues (Okubo et al. 2020), or angular velocity input (Turner-Evans et al. 2017; Green et al. 2017) to update the internal representation of HD. However, whether and to which extent the fly takes into account uncertainty associated with these different cues when updating its own heading estimate remains unclear.

In what follows, we will address how ring-attractor networks, like the one in the *Drosophila* CX, could maintain and update a Bayesian HD estimate, and how they strategically use uncertainty for this computation. Specifically, we start by formulating a Bayesian ideal observer model, which predicts how the HD estimate and its associated uncertainty is updated dynamically in presence of unreliable sensory observations. We show that a specifically tuned ring-attractor network can approximate this ideal observer, which allows us to identify the contribution to the Bayesian computation of different circuit motifs. This network relies on a representation of uncertainty in the amplitude of a neural activity bump. More generally, we show that near-Bayesian integration is inherent in generic ring attractor networks through amplitude dynamics for a wide range of network parameters, as long as their connection strengths allow the network to sufficiently *deviate* from the attractor state. Lastly, we demonstrate that network motifs that support approximate Bayesian inference in ring attractor networks are present in the *Drosophila* CX connectome. Overall, our work provides a principled theoretical foundation for the suggestion that the *Drosophila* CX ring attractor may implement Bayesian HD tracking, and we suggest experiments to further test this hypothesis.

Results

Circular Kalman filtering: an ideal observer model of heading direction tracking

We first formulate how heading direction can be tracked by an ideal Bayesian observer, which results in an algorithm for dynamic circular inference. This algorithm tells us how noisy observations or sensory input ought to be handled in principle, without yet specifying a potential implementation in a neural network.

An animal's heading direction estimate is informed by two qualitatively different types of sensory inputs (Fig. 1a). On the one hand, the animal's internal compass might receive self-turning signals (or **angular velocity observations**), e.g., proprioceptive feedback or visual flow. Integrating only turning signals over time to track HD would result in gradual error accumulation, as angular velocity observations tend to be noisy. In other words: the longer the animal relies on angular velocity integration alone, the more its HD estimate will deviate from the true HD. In addition, an animal may make use of **absolute heading observations**, such as the perceived position of the sun or the wind direction, to calibrate its internal compass. Since these observations are also noisy, they should be combined with the internal HD estimate according to their respective reliabilities.

Here, we formulate the task of dynamically tracking HD with these two types of input as a dynamic Bayesian inference task, ensuring that unreliable sensory input is properly accounted for. We assume the observer has access to both angular velocity observations $v_t \in R$ and absolute heading observations $z_t \in [-\pi, \pi]$. These inform the ideal observer about its angular velocity $\dot{\phi}_t \in R$ and true HD $\phi_t \in [-\pi, \pi]$, respectively, but are corrupted by zero-mean noise. Specifically, we assume angular velocity observations to be corrupted by Gaussian noise that limits the precision of these observations (with precision κ_v, larger κ_v= more reliable), and absolute heading observations by von-Mises noise with precision κ_z, the Gaussian equivalent for circular variables.

The goal of dynamic Bayesian inference is to compute one's **posterior belief** of HD $p(\phi_t | z_{0:t}, v_{0:t})$, and continuously update this belief in light of incoming sensory evidence, $z_{0:t} \equiv z_0, z_1, \ldots$ and $v_{0:t} \equiv v_0, v_1, \ldots$. Importantly, such a belief constitutes a whole probability distribution, rather than a single point estimate, which automatically gives us a measure of uncertainty around the best HD estimate (Knill and Pouget 2004; Dehaene, Coen-Cagli, and Pouget 2021).

The circular nature of heading direction precludes the use of standard inference schemes such as the Kalman filter (Kalman 1960; Kalman and Bucy 1961) to update the posterior belief $p(\phi_t | z_{0:t}, v_{0:t})$ over time. Instead, statistical inference turns out to be analytically intractable (Kurz, Gilitschenski, and Hanebeck 2016) and needs to be approximated (see Methods). We approximate this belief at each point in time by a von Mises distribution, $p(\phi_t | z_{0:t}, v_{0:t}) \approx VM(\mu_t, \kappa_t)$, fully characterized by its mean μ_t, which is the current best HD estimate, and its precision κ_t, which measures the estimate's certainty (Fig. 1b). As these two posterior, or belief, parameters fully specify the HD belief, updates of the belief in light of sensory evidence simplify to updating these to parameters. We derived the parameter update dynamics by a technique called projection filtering (Brigo, Hanzon, and Le Gland 1999; Kutschireiter, Rast, and Drugowitsch 2021), resulting in

$$d\mu_t = v_t dt + \frac{\sqrt{2\kappa_z dt}}{\kappa_t} sin\,(z_t - \mu_t), \qquad (1)$$

$$d\kappa_t = -\frac{f(\kappa_t)}{2\kappa_v}\kappa_t dt + \sqrt{2\kappa_z dt}\, cos\,(z_t - \mu_t). \qquad (2)$$

Here, the factor $\sqrt{2\kappa_z dt}$ ensures proper information scaling of absolute HD observations in the continuous-time limit, and $f(\kappa_t)$ is a monotonically increasing nonlinear function that controls the decay of one's certainty κ_t (see Eq. (11) in Methods). Due to their conceptual similarity to the Kalman filter, we will refer to Eqs. (1) and (2) as the **circular Kalman filter** (circKF). The circular Kalman filter is a general algorithmic method for using noisy data to estimate a circular variable over time.

To provide intuition for the filter's operation, let us represent the belief parameters in its polar coordinate form as a vector on the 2D plane (Fig. 1b inset), whose direction determines the mean HD estimate μ_t, and its length the precision κ_t. Let us now consider how this vector is updated by angular velocity and absolute heading observations in Eqs. (1) and (2).

Angular velocity observations. In the absence of absolute heading observations, i.e., $\kappa_z = 0$, the heading estimate μ_t is fully determined by integrating angular velocity observations v_t over time, i.e., angular path integration (first RHS term in Eq. (1), Fig. 1c). In our vector representation, angular velocity information impacts the belief by a vector pointing in a tangential direction (Fig. 1c), which turns the belief according to the observations (the $v_t dt$ term in Eq. (1)). The ideal Bayesian observer accounts for the error accumulation associated with angular path integration by having the belief's certainty κ_t decay over time, mirroring the estimate's error (first RHS term in Eq. (2)). Accordingly, the belief vector shrinks to account for this reduction in certainty (the $f(\kappa_t)$-related term in Eq. (2)). Interestingly, per Eq. (2), angular velocity observations always decrease certainty. The decrease is more modest for more precise velocity observations (i.e., κ_v large) but nonetheless persists. In other words, if only angular velocity observations are present, the posterior certainty κ_t will inevitably decay towards zero (uniform posterior distribution, i.e., complete lack of knowledge), with decay speed determined by the angular velocity observation's "informativeness" κ_v.

Absolute heading observations. When navigating, absolute heading information, like a GPS reading, helps to calibrate the internal compass. To weigh the reliability of such information against the current HD estimate's certainty, its impact in Eqs. (1) and (2) is scaled by κ_z: if the cue's reliability κ_z is large, the observation z_t will substantially change the mean μ_t towards the direction of the cue. Conversely, if the current certainty κ_t is large compared to the cue's reliability, an absolute heading observation z_t will hardly update the existing estimate.

In our vector representation, including absolute heading information corresponds to adding an absolute heading information vector to the current belief vector (Fig. 1d; see Methods). The direction and length of this heading information vector are determined by the observation's position z_t and reliability κ_z, respectively. Depending on how well the observation is aligned with the current belief (as measured by the cosine in Eq. (2)), the certainty κ_t can either increase or, in the case of a strongly conflicting stimulus, even *decrease* (Fig. 1e). This interesting result is a consequence of the circular nature of the inference task (Murray and Morgenstern 2010), and stands in contrast to the Kalman filter where absolute information *always* increases the estimate's certainty (Wilson and Finkel 2009). It is thus a key distinction between the Kalman filter and the circKF. If more than one heading cue is available, the picture generalizes to adding multiple vectors to the current belief vector (Fig. 1f).

In a dynamic setting, both angular velocity and absolute HD observations are available to the ideal observer as a continual stream. That is, at every point in time, the belief is updated according to Eqs. (1) and (2). In summary, angular velocity observations change the HD estimate and reduce certainty, while absolute heading observations update the HD estimate weighted by their reliability, and either increase certainty (if compatible with one's belief) or reduce certainty (if conflicting with the current belief). These operations are continuously repeated to maintain a belief over time.

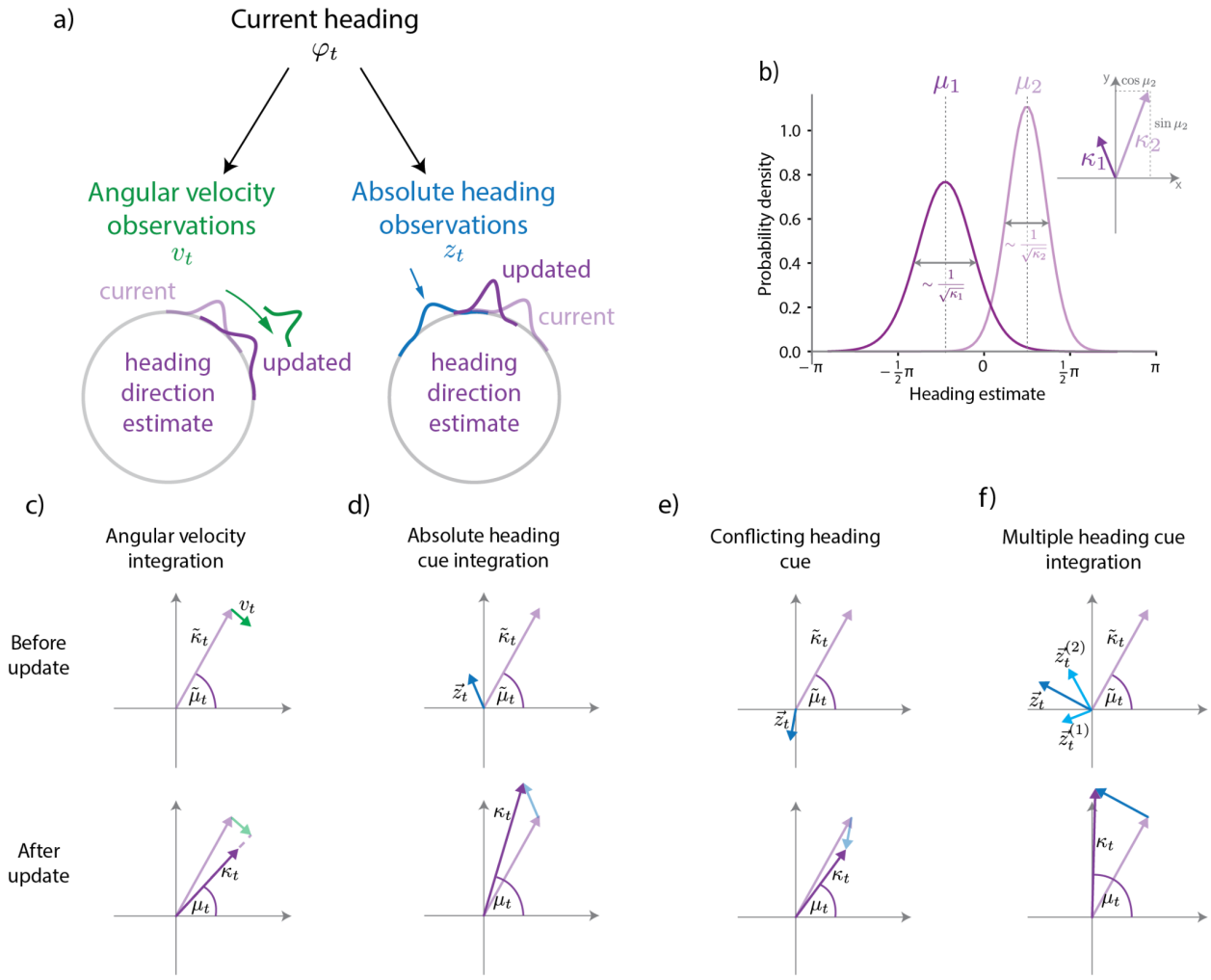

Figure 1. The circular Kalman filter as an ideal observer model for heading direction tracking.

a) Two different types of observations inform about one's current heading direction ϕ_t. Angular velocity observations v_t are assumed to be drawn from a Gaussian distribution centered around the true angular velocity $\dot{\phi}_t$, with precision κ_v. Absolute heading observations z_t are drawn from a von Mises distribution centered around the true heading ϕ_t, with precision κ_z.

b) At every point in time, the belief $p(\phi_t|v_{0:t}, z_{0:t})$ is approximated by the unimodal von Mises distribution, the "Gaussian on the circle". It is fully characterized by its mean parameter μ_t, which determines the position of the mode, and its precision parameter κ_t. Interpreted as the polar coordinates in the 2D plane, these parameters constitute a convenient vector representation of the belief (inset).

c) Angular velocity observations v_t turn the current belief vector in the direction of the observations (angular path integration). The vector length reduction (top vs. bottom), and associated reduction of certainty, accounts for the error accumulation resulting from angular path integration.

d) Integrating absolute heading observations corresponds to adding the absolute heading observation vector (cyan) to the current belief vector (purple).

e) Absolute heading observations that are in conflict with the current belief result in a shortening of the belief vector (top vs. bottom) and an associated reduction of the belief's certainty.

f) Integration of multiple absolute heading cues, such as wind and vision, can be considered as a sum of a single observation vector.

Neural encoding of heading direction estimates

The derived circKF algorithm updates the parameters of one's belief about heading direction in light of incoming sensory information without a reference to neural activities or neural dynamics. If the brain performs this inference, it presumably encodes these parameters by neuronal activities, and updates them through neuronal dynamics. Downstream neurons need to be able to use these encoded inference parameters in further computations, ideally with fairly simple, for example linear, computations. What we therefore need as the next step is a model that encodes the belief parameters μ and κ into neural activity. For this, we will assume a **linear probabilistic population code** (lPPC).

The central idea behind an lPPC (Ma et al. 2006; Pouget et al. 2013) is that neuronal activity encodes a probability distribution, rather than a single point estimate (e.g., the most likely HD without its uncertainty), such that the natural parameters can be retrieved through linear operations, i.e., a weighted sum of neural activity. For the von Mises distribution used here, the natural parameters are the Cartesian end point coordinates $(\kappa\, cos\,(\mu), \kappa\, sin\,(\mu))$ of its vector representation (Fig. 1b) that uniquely reconstruct the distribution's shape. Here, we assume that the neural population consists of independent Poisson spike generators with uniformly distributed HD (ϕ) tuning. For an lPPC, this assumption implies that the tuning curve of neuron *i* is von-Mises-shaped around its preferred heading direction ϕ_i, and that the tuning curve's gain is proportional to the encoded certainty κ ((Ma et al. 2006), see SI).

The brain's task is to estimate HD from noisy data. Each neuron's tuning to the mean of the HD estimate (μ) will be the average of its tuning with respect to each individual HD (ϕ), weighted by how likely each value of ϕ is for a given μ (Eq. (13) in Methods). Low certainty (i.e., a small κ) results in bump-shaped tuning curves that are slightly wider and flatter than the von Mises shaped ϕ-tuning curves (Fig. 2a). For high certainty (i.e., large κ), they will approach a von Mises shape (Fig 2b). The encoded distribution (Fig. 2c) thus results in bump-shaped population activity profiles that follow the same pattern: activity reflecting low certainty is close to a cosine (Fig. 2d), whereas high certainty makes it approach a von Mises shape (Fig. 2e). The corresponding height of the profile scales approximately linearly with certainty κ, while the width of the profile is mostly constant and only depends weakly on κ for low certainties (Fig 2f). Therefore, the population profile is not just a rescaled version of the encoded probability distribution (Fig 2c), because an increase in certainty does not cause the bump to sharpen indefinitely.

This encoding model has several useful properties. First, the resulting population activity profile mimics the unimodal bump of neural activity observed in the *Drosophila* EB (Seelig and

Jayaraman 2015). Second, it encodes both mean estimate μ and associated uncertainty κ, as required for implementing the Bayesian update rules (Eqs. (1) and (2)). Third, consistent with IPPCs, the natural parameters of the von Mises belief, $\theta_1 = \kappa_t \cos(\mu_t)$ and $\theta_2 = \kappa_t \cos(\mu_t)$, can be recovered by taking a weighted sum of the neural population's activity. This makes these parameters fully accessible to downstream neurons via simple (linear) neural operations. Fourth, $\theta_1$and θ_2 represent the von Mises distribution in terms of Cartesian vector coordinates in the 2D plane (where μ and κ are its polar coordinates, cf. Fig. 1b). Such a representation is similar to the phasor representation of neural activity (Lyu, Abbott, and Maimon 2020; Hulse et al. 2021), which translates bump position and bump amplitude to polar coordinates in the 2D plane. Since in our model the activity bump is scaled by certainty, the phasor representation of neural activity approximately corresponds to the vector representation of the von Mises distribution (cf. Fig 2c,d,e). Thus, in what follows, we will make use of this duality and exploit our knowledge on how to manipulate those vectors in order to implement the operations in the optimal observer model (cf. Fig 1 c, d) through neural dynamics.

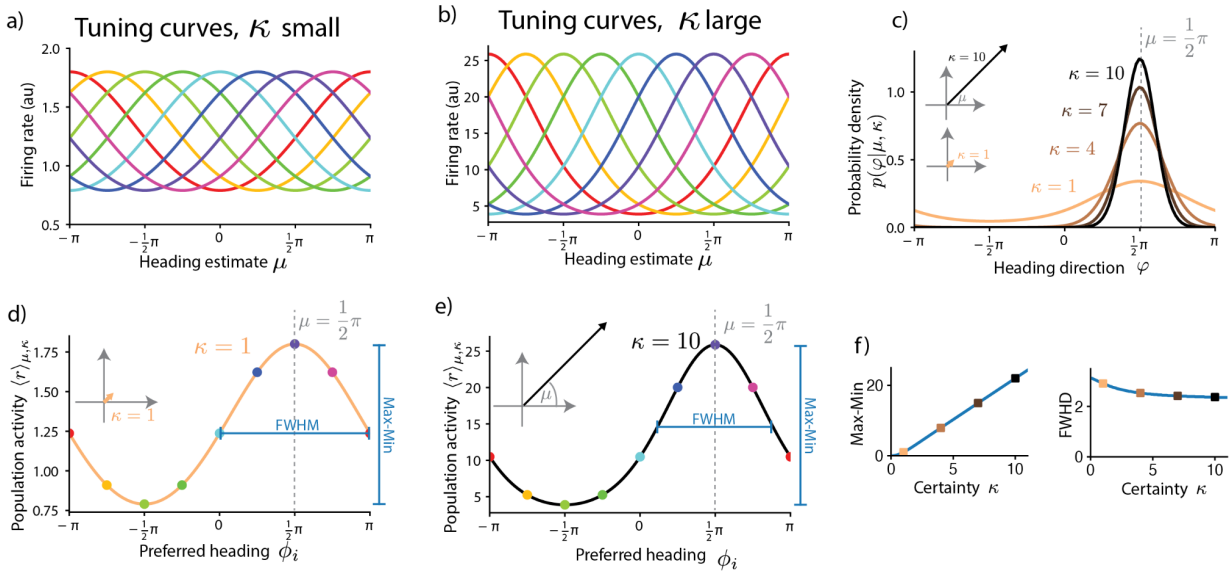

Figure 2. Encoding the heading direction estimate in neural population activity.

- **a)** Tuning curves to mean HD for small values of certainty κ are cosine-shaped. Here, we show the tuning curves of 8 example neurons with $\kappa = 1$ (colors = different preferred μ's).
- **b)** Tuning curves to mean HD for large values of certainty κ are von-Mises shaped (same 8 example neurons as in **a**, but for $\kappa = 10$).
- **c)** Von Mises probability densities for different values of certainty κ and fixed mean $\mu = \pi/2$. Note that the density sharpens around the mean with increasing certainty. Inset shows vector representation of a von Mises distribution with mean $\mu = \pi/2$, and, respectively, $\kappa = 10$ and $\kappa = 1$.
- **d)** Population activity profile (average neural firing rate conditioned on heading estimate μ and certainty κ) encoding the von Mises densities with mean $\mu = \pi/2$ and certainty $\kappa = 1$. Neurons are sorted by preferred heading directions ϕ_i. Colored dots correspond to activity of neurons with tuning curves in **a**. The phasor representation of the neural

activity (inset) matches the vector representation of the encoded von Mises distribution in **c**.

e) Population activity profile encoding the von Mises densities with mean $\mu = \pi/2$ and certainty $\kappa = 10$.

f) Left: The amplitude (Max-Min) of the activity profile scales (approximately) linearly with certainty κ, except for very small values of κ. Right: The population activity bump's width (full width at half maximum, FWHM) is mostly unaffected by uncertainty κ, and saturates at a finite value for large κ, unlike the von Mises distribution it encodes (e.g., **b**), whose FWHM approaches zero for large values of κ.

Recurrent neural networks can track Bayesian HD estimates

Equipped with a model for how neural population activity can encode HD belief parameters, we now consider how an RNN with such a belief parameter encoding can continuously update these parameters to implement the circular Kalman filter (Eqs. (1) and (2)). We will start with an idealized network that only assumes a single neural population, similar to many generic ring-attractor networks (e.g., (Zhang 1996; Hansel and Sompolinsky 1998)). Later, we will use this idealized network as a basis for the more distributed CX-like network that matches the *Drosophila* connectome (Turner-Evans et al. 2020; Kim et al. 2017; Hulse et al. 2021).

A simple and analytically accessible ansatz for network dynamics that implement the circular Kalman filter (Eqs. (1) and (2)) is of the form

$$\dot{\mathbf{r}}_t = -\frac{1}{\tau}\mathbf{r}_t - g(\mathbf{r}_t)\mathbf{r}_t + W \cdot \mathbf{r}_t + \mathbf{I}_t^{\mathrm{ext}}, \qquad (3)$$

where r_t denotes a vector of neural activity levels, with neurons ordered by their preferred headings $\varphi^{(i)}$, τ is the leak time constant, W is the recurrent connectivity matrix, and ${I_t}^{ext}$ is a vector of external inputs to the network. We will specify the details of the synaptic inhibition $g(r_t)$ later, and for now only mention that its output increases with bump amplitude, thus implementing nonlinear global inhibition.

Careful tuning of the network parameters yields a class of networks that, in the limit of infinitely many neurons, exactly implements the circular Kalman filter (see Methods for details, and the SI for the formal analysis). In this class, probabilistic angular velocity integration is implemented by an interplay between recurrent connectivity (W), leak ($1/\tau$), and synaptic inhibition ($g(r_t)$). Modulating the strength of the recurrent connectivity's sine-shaped (or odd) component (Fig. 3a) by angular velocity observations v_t causes the heading estimate to rotate (Fig. 3b). Meanwhile, leak and global inhibition together cause certainty to decay over time (Fig. 3c). Absolute heading information (e.g., information about the position of a visual cue) enters the network via the external input vector ${I_t}^{ext}$, in form of a cosine-shaped bump with amplitude modulated by perceptual reliability κ_z (Fig. 3d). Adding this input activity effectively implements the vector addition required for proper absolute heading information integration. Then, the external information's weight is determined by the ratio between input amplitude and bump amplitude, in line with the weighting between cue reliability and own certainty required by the circKF. Bump position and amplitude

dynamics derived from such a network well-approximate - and for certain choices of parameters are identical to - the parameter dynamics of the circular Kalman filter (Eqs. (1) and (2); Fig. 3e; see Eqs. (15) and (16) in Methods for the dynamics of the bump parameters).

Importantly, the network that exactly implements the circular Kalman filter does *not* constitute a ring-attractor network. In fact, in the absence of external input, $I_t^{ext} = 0$, the network activity decays towards its fixed-point at zero (Fig 3e). This stands in contrast to ring attractor networks that, without input, commonly maintain a constant, non-zero activity. In our network, activity decay is by design, and reflects the continuously decreasing certainty arising from pure angular path integration in the absence of absolute HD observations.

We can regain some similarity to previously proposed ring attractor networks by approximating the Bayesian inference. We do so by adding even (cosine-shaped) recurrent connectivity, and by choosing $g(r_t)$ such that the second term in Eq. (3) becomes a quadratic function of the bump amplitude. The use of a quadratic function has the added advantage of making this network fully analytically accessible, as we see in the Methods. This allows us to tune the network parameters such that this network exactly implements **a quadratic approximation** of the circKF filter. The circKF implementation becomes non-approximate (i.e., precise) in the limit of large posterior certainty κ_t. With these modifications, the network's activity ceases to decay to zero in absence of absolute HD observations, and instead settles on a non-zero stable fixed point. In other words, the bump's amplitude dynamics will correctly reflect the posterior's certainty for large bump amplitudes but will deviate from it in the small-certainty/small-amplitude limit. Thus, the network has two operating regimes: a regime close to the attractor state with constant amplitude, encoding approximate constant certainty, and a dynamic regime away from the attractor state, where it correctly implements dynamic Bayesian inference. Numerical simulations confirm the existence of these two regimes: the network tracks the HD estimate and its associated certainty just like the circular Kalman filter in the dynamic regime, but features a slightly lower HD tracking precision, and overestimates its confidence, in the attractor regime (Fig. 3f, orange vs. blue).

In summary, three network motifs support Bayesian inference in an RNN (Fig 4b-d): (i) odd recurrent connectivity with strength modulated by angular velocity observations v_t, (ii) global inhibition that is approximately quadratic in bump amplitude, and (iii) a cosine-shaped external input at the position of the absolute HD observation, whose strengths is modulated by the reliability of this observation κ_z. Motifs (i) and (ii) implement probabilistic angular path integration, whereas motif (iii) updates the network's current HD estimate in light of uncertain absolute HD observations.

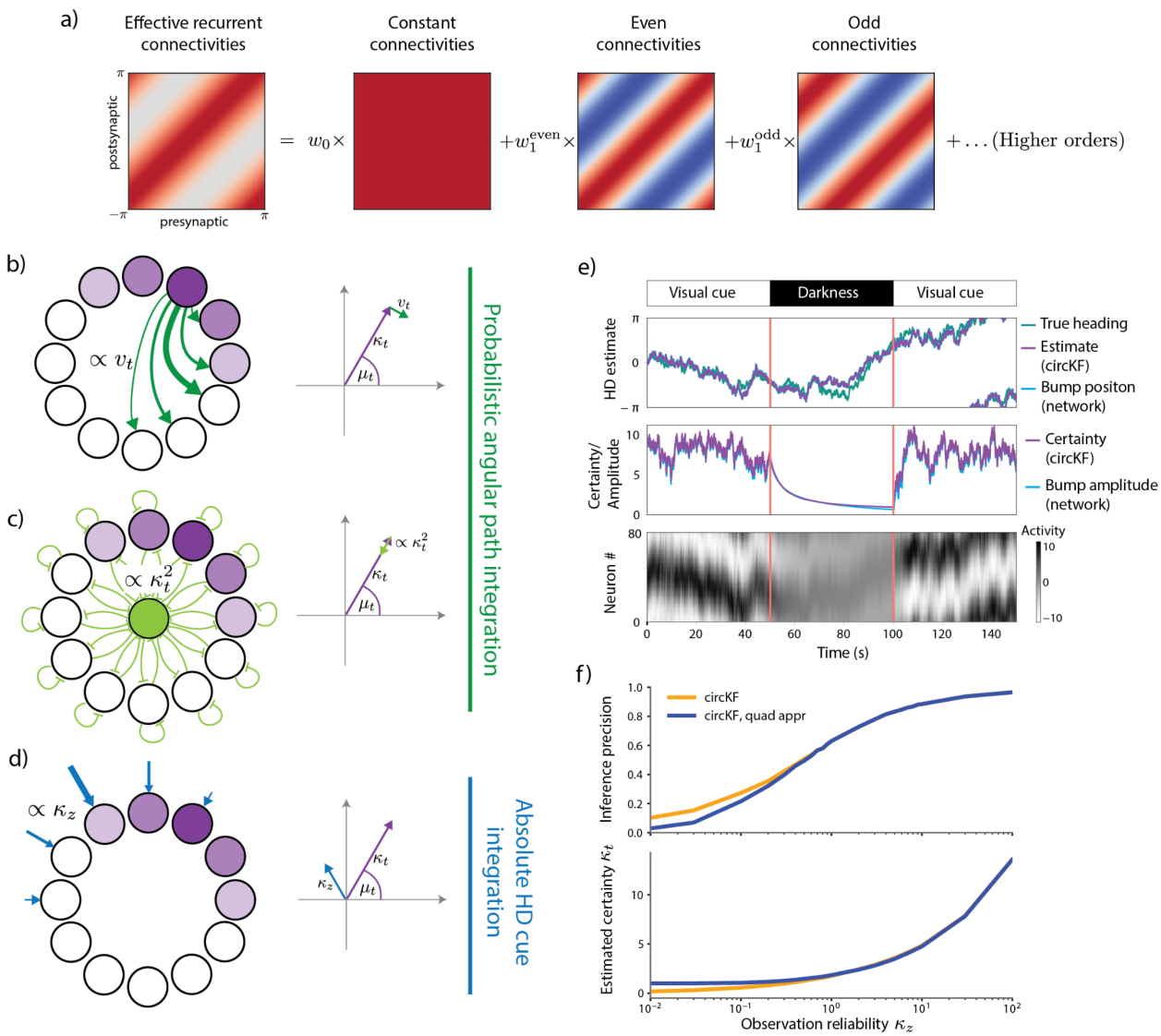

Figure 3. A network implementation of the circular Kalman filter.

a) Rotation-symmetric recurrent connectivities (here: neurons are sorted according to their preferred HD) can be decomposed into constant, cosine-shaped (even), sine-shaped (odd) and higher-order frequency components (basis function). Red and blue denote excitatory and inhibitory components, respectively.

b) Network motifs that implement the circular Kalman filter (**b-d**). Rotations of the HD estimate are mediated by sine-shaped (or odd) recurrent connectivities, whose strength is modulated by angular velocity observations.

c) Decay in amplitude arises from leak and global inhibition.

d) A cosine-shaped input to the network provides external absolute HD cue input. The strengths of this input is modulated by observation reliability κ_z.

e) The dynamics of the network implement the dynamics of the ideal observer's belief, as shown in a simulation of a network with 80 neurons. In the 'visual cue' period both absolute heading observations and angular velocity observations were available. During 'darkness', only angular velocity observations were available.

f) The network implementation with quadratic leak approximation (circKF, quadratic approx) tracks the HD estimate with the same precision (top; higher = lower circular distance to true HD) as the circular Kalman filter (circKF, Eqs. (1) and (2)) if absolute heading observations are reliable (large κ_z), but with slightly lower precision once they become less reliable (small κ_z). This drop comes with an overestimate in the estimate's confidence κ_t (bottom). Plots are averages over 10000 simulations (see Methods for simulation details).

Ring attractors approximate Bayesian inference for HD tracking

As just shown, a ring-attractor network with quadratic inhibition and carefully tuned network parameters can implement a quadratic approximation of the circKF. Our network differs in two ways from classical ring-attractor networks that qualitatively model neural dynamics in HD systems (Zhang 1996; Skaggs et al. 1995; Xie, Hahnloser, and Seung 2002; Kim et al. 2017). First, by design, classical networks do not feature a notion of uncertainty, and therefore assign no interpretation to their bump's amplitude. Second, they are designed to operate close to their attractor states, whereas our network performs close-to-optimal Bayesian inference only in a regime away from its attractor states. Here, we investigate whether generic ring-attractor networks may nonetheless approximate Bayesian computations through their amplitude dynamics. To see which features of the amplitude dynamics in our ring-attractor network enable such computations, we asked more generally how much of the (ideal) Bayesian computation can be captured by ring-attractor dynamics that are not perfectly tuned to the ideal observer.

Our RNN with quadratic inhibition is a ring-attractor network that exhibits attracting states with a finite amplitude. In networks of such structure, bump position changes upon absolute heading inputs in proportion to the ratio between input amplitude (≙ reliability of absolute heading cue) and bump amplitude (≙ own certainty) (see Eq. (15) in Methods). Carefully tuning the network parameters, in particular recurrent weights, network time constant, and inhibitory weights, allows us to change the bump amplitude dynamics (Eq. (17) in Methods) to explore network regimes away from approximate Bayesian inference. Specifically, we tuned the network parameters to modulate two amplitude characteristics. The first is the attractive fixed point $\kappa *$ in the amplitude of the population activity bump (specified by the parameter $\kappa *$, which defines both the amplitude and the width of the bump)[1]. The second is the inverse time constant β, which describes how fast the amplitude approaches this fixed point (larger β implies faster dynamics, Fig. 4a).

In the limit of infinitely fast dynamics, $\beta \to \infty$, the network becomes a `strict attractor' with a bump amplitude that never moves away from its fixed point. As the bump amplitude encodes the HD estimate's uncertainty, such a strict attractor never updates its uncertainty, and consequently lacks proper reliability weighting of absolute HD observations. Nonetheless, strict attractors with a numerically optimized fixed point amplitude $\kappa *$ (Fig. 4b inset) can still perform competitive HD

[1] To facilitate comparison, we adhere to a slight abuse of notation in this section, and refer to the bump amplitude as \kappa, because it enters the phase dynamics in exactly the same way as the precision dynamics enters the mean dynamics in the circKF.

estimation for moderately reliable observations (Fig. 4b). Adjusting this fixed point for each level of information reliability *individually* further boosted the network's estimation performance, and effectively re-established proper Bayesian reliability weighting of incoming absolute HD observations, similar to the circKF (Fig. 4c). Even then, strict attractors lack temporal adjustment of their internal certainty estimates, as required by Bayes-optimal path integration.

We next asked whether we could increase performance by loosening the attractor. Such a relaxation would permit the network to operate farther away from its attracting state, to which it will decay in the absence of absolute heading observations (the dynamics that implement the quadratic approximation of the circKF are a special case of this network, see Fig. 4a). Indeed, slower dynamics (smaller inverse time constant β) boosted overall performance (Fig. 4d). In fact, with tuned network parameters, HD estimation performance becomes practically indistinguishable from that of the ideal Bayesian observer. In this regime, the HD estimate and bump amplitude/certainty dynamics become almost identical to the circKF, whereas those resulting from the strict attractor clearly deviate, despite optimized fixed-point value (Fig. 4e). Interestingly, the optimal network parameters do not necessarily coincide with the quadratic approximation of the circKF, which we found by analytically matching the certainty dynamics in the large-certainty limit rather than by numerical optimization. In fact, a wide range of network parameters lead to a relatively small performance loss (<10%, Fig. 6d). Therefore, accurate parameter tuning might be unnecessary, as long as the network dynamics remain sufficiently slow.

Overall, this demonstrates that proper HD estimation relies on weighting absolute heading observations both globally (Figs. 4 b,c), i.e. according to the average level of reliability, and dynamically (Figs. 4d,e), according to the dynamics of one's own certainty. Nonetheless, decent performance can be achieved for a wide range of network parameters. This may indicate a "built-in" implicit reliability weighting in attractor networks through their amplitude dynamics, as previously suggested for static absolute HD observations (Esnaola-Acebes, Roxin, and Wimmer 2021), as long as the attractor dynamics around the fixed point are reasonably slow and allow for deviation from the attractor state. Such a finding may explain why ring-attractor networks perform reasonably well in practice, even though they lack an explicit representation of uncertainty.

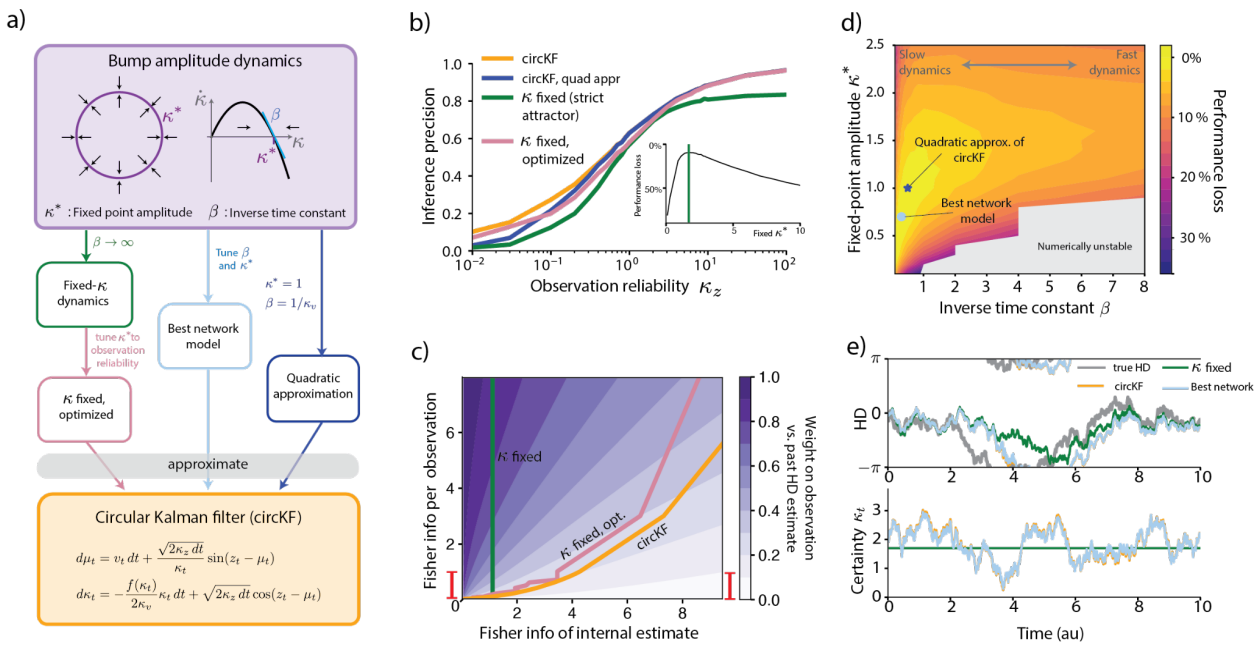

Figure 4. Attractor models with slow dynamics approximate Bayesian inference.

a) A linear RNN with quadratic inhibition can operate in different regimes. Its bump amplitude dynamics can be characterized by their fixed point amplitude $\kappa *$ and inverse time constant β. Note that the bump position dynamics is described by the same equation across all compared regimes (Eq. (15) in Methods). However, the position dynamics depend on bump amplitude, whose dynamics differ across regimes. This causes HD tracking behavior to differ across network regimes.

b) HD estimation performance as measured by inference precision (as defined by $1 - circVar$, see Methods). For the strict attractor, we chose $\kappa *$ to numerically minimize performance loss (compared to the performance-maximizing particle filter) across all levels of observation reliability, weighted by a prior $p(\kappa_z)$ on this reliability (see Methods). For the optimized network, we adapted $\kappa *$ individually for each level of observation reliability to maximize performance.

c) The weight with which a single observation contributes to the HD estimate varies with informativeness of both the absolute HD observations and the current HD estimate. We here illustrate this for an absolute heading observation that is orthogonal to the current HD estimate, resulting in the largest possible estimate change ($|z_t - \mu_t| = 90$deg in Eq. (1). The weight itself quantifies how much the observation impacts the HD estimate as a function of how informative this observation is (vertical axis, measured by Fisher information of a 10ms observation) and our certainty in the HD estimate (horizontal axis, also measured by Fisher information) before this observation. A weight of one implies that the observation replaces the previous HD estimate, whereas a weight of zero implies that the observation does not impact this estimate. The close-to-optimal update weight of the circKF (yellow) forms a nonlinear curve through this parameter space. Fisher information per observation is directly related to the observation reliability κ_z, and the vertical red bar shows the equivalent range of observation reliabilities, $\kappa_z \in [10^{-2}, 10^{2}]$, shown in panel b.

d) Overall inference performance loss (compared to a particle filter; performance measured by avg. inference precision, as in **b**, 0%: same average inference precision as particle filter, 100%: average. inference precision = 0), averaged across all levels of observation reliability (weighted by prior $p(\kappa_z)$, see Methods) as a function of the bump amplitude parameters $\kappa *$ and β. For too small fixed point amplitudes and too fast dynamics, numerical simulations become unstable (grey area).

e) Simulated example trajectories of HD estimate/bump positions of HD estimate/bump positions (top) and certainties/bump amplitudes (bottom). To avoid cluttering, we are not showing the quadratic approximation of the circKF (visually indistinguishable from circKF and best network).

Drosophila CX can implement a circular Kalman filter

To show that fruit flies can in principle perform Bayesian HD tracking, we need to establish that the network motifs that support Bayesian inference in our simple RNN are also present in the *Drosophila*'s CX. To do so, we first revisit what is known about the CX's network anatomy as well as past work suggesting that this network implements a ring attractor network for (non-Bayesian) HD tracking (Kim et al. 2017; Hulse and Jayaraman 2020). Then, we use the *Drosophila* connectome (Scheffer et al. 2020) as a template to build a distributed, CX-like Bayesian ring-attractor network. This reveals that the network motifs that implement the ring attractor in the CX are sufficient to implement a Bayesian inference algorithm, allowing for a probabilistic interpretation of these motifs.

The ring attractor in the *Drosophila* HD system is composed primarily of three core cell types, called EPG, $\Delta 7$, and PEN1 (Turner-Evans et al. 2020; Scheffer et al. 2020; Hulse et al. 2021), cf. Fig. 5a. First, EPG neurons are tuned to HD, and this depends in large part on the sensory inputs that EPG neurons receive from so-called ER neurons. External sensory cues rotating around the head rotate the pattern of neural activity in the ER population, and thus also the pattern of activity in the EPG population (Seelig and Jayaraman 2013; 2015; Omoto et al. 2017; Fisher et al. 2019; Kim et al. 2019). Next, EPG neurons are recurrently connected with inhibitory $\Delta 7$ neurons. Each $\Delta 7$ neuron is excited by a set of EPG neurons with similar heading tuning, and inhibits EPG neurons with opposite heading tuning (Turner-Evans et al., 2020). This creates network stability and ensures a single EPG activity bump at any given time. Finally, EPG neurons are also recurrently connected with excitatory PEN1 neurons. The function of PEN1 neurons is to bring angular velocity information to the EPG ensemble (Green et al. 2017; Turner-Evans et al. 2017). Each half of the PEN1 population also contains a topographic map of heading direction, which is inherited from EPG neurons. The pattern of PEN1 projections back to EPG neurons is shifted to drive the bump of EPG activity around the ring whenever the fly rotates, and thus effectively establishes the odd recurrent connectivity required to initiate turns in traditional ring-attractor models (Zhang 1996; Skaggs et al. 1995). This makes the rotational velocity of the bump roughly proportional to the fly's own rotational velocity, when ER neurons are inactive, i.e., in absence of external sensory cues that would inform about the fly's HD.

We used the CX connectome, determined from the Hemibrain dataset (Scheffer et al. 2020) as the number of synapses between the different neuron types (Fig. 5b), as a template for the

connectivities of an idealized RNN that can perform Bayesian inference (Fig. 5c). To avoid confusion with actual neurons in the Drosophila connectome, we refer to the neuronal populations in this idealized RNN as heading-direction (HD), angular velocity (AV+ and AV-, in reference to the two hemispheres), inhibitory (INH) and external input (EXT) populations. The relative connection strengths between (rather than within) these populations as well as decay time constant were determined analytically such that the dynamics of the bump parameters implement the circular Kalman filter (see SI for derivation; Methods for network dynamics). By design, the RNN builds its ring attractor dynamics on the motifs available in the *Drosophila* connectome. Differential activation in the AV+ and AV- population in combination with a shifted connectivity pattern implements the effective odd connectivity needed to turn the HD bump by angular velocity observations (Fig. 5d), which is equivalent to how turning signals are thought to be integrated in the ring attractor in the CX (Green et al. 2017; Turner-Evans et al. 2017), as described above. Absolute heading information, such as from visual or wind sensors, is provided via inhibitory inputs from the EXT population. As for the ER neurons in the CX, the EXT population is connected to the HD neurons such that they inhibit those HD neurons opposite of the current input. By scaling the size of the EXT input with the reliability of absolute HD cues, the RNN achieves Bayes-optimal reliability weighting between external cue and internal reliability (Fig. 5e). In absence of absolute HD observations, the HD population's bump amplitude should decay over time to reflect the certainty decay during probabilistic angular path integration. We implemented this decay through a quadratic interaction with the INH population. More precisely, the INH population is not only responsible for network stability, as was hypothesized for the *Drosophila* CX (Turner-Evans et al. 2020), but additionally implements the certainty decay resulting from probabilistic angular path integration (Fig. 5f).

In summary, we have built an RNN that uses the *Drosophila* connectome to implement Bayesian HD tracking. The resulting ring-attractor network is thus fully consistent with previously proposed implementations of ring attractors in the CX (Kim et al. 2017; Kakaria and de Bivort 2017; Hulse and Jayaraman 2020), but offers a different interpretation of the underlying computation that the network putatively performs. Specifically, we identified all the motifs necessary to implement a Bayesian computation in the *Drosophila* connectome. Overall, this indicates that the *Drosophila* HD system could, at least in principle, implement such an algorithm.

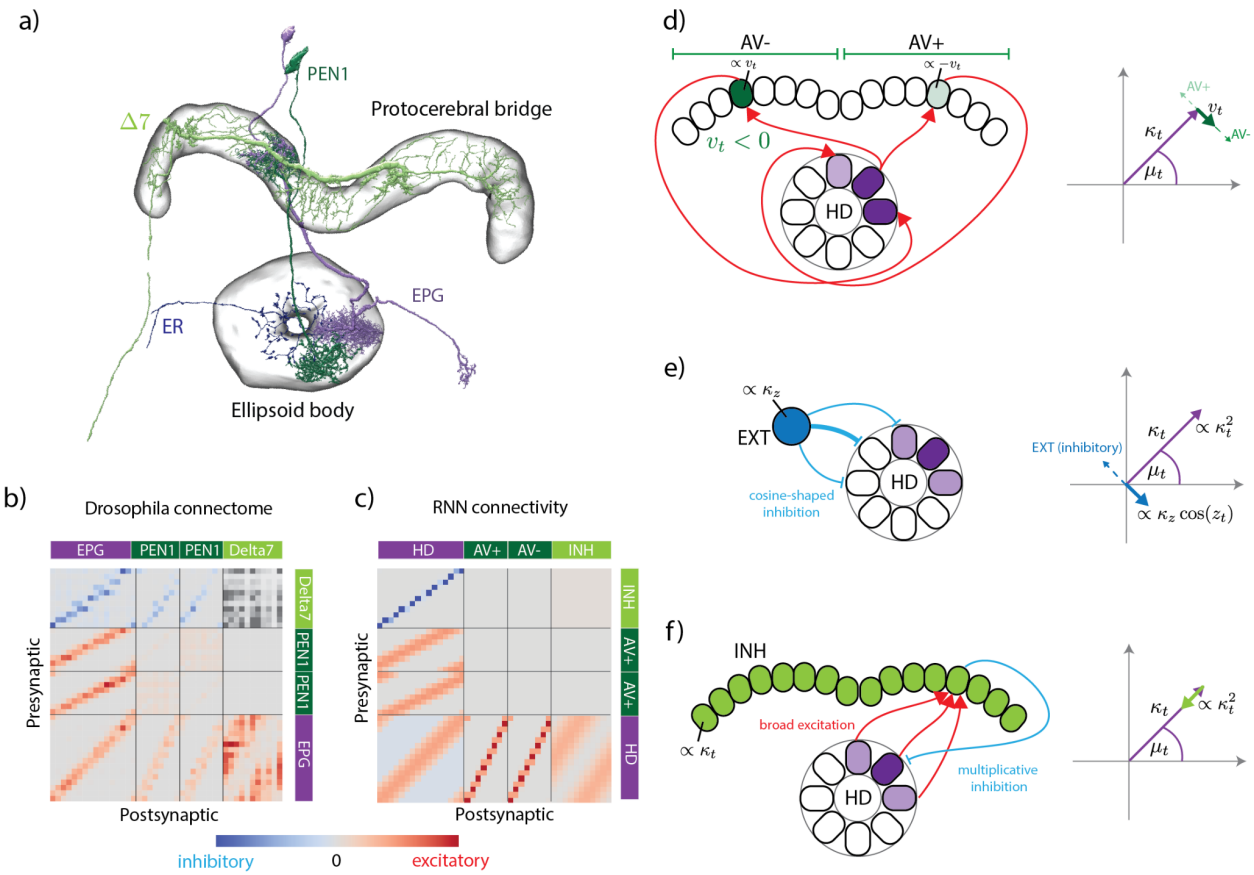

Figure 5. A Drosophila-like network implementing the circular Kalman filter.

a) Cell types in the *Drosophila* brain that could potentially contribute to implementing a circular filtering algorithm for HD tracking.

b) Connectivity between EPG, *Δ*7 and PEN1 neurons, plotted from the hemibrain:v1.2.1 database (Scheffer et al. 2020). ER neurons were omitted because they are not essential elements of the ring attractor in the *Drosophila* CX. Here, neurons were grouped according to anatomical region as a proxy for preferred HD. *Δ*7 to *Δ*7 connectivities are greyed out, as the polarity of these connections (inhibitory or excitatory) remains unknown.

c) RNN connectivity profile that implements an approximation of the circular filtering algorithm is strikingly similar to the connectivity of neurons in the *Drosophila* CX.

d) Differential activation of AV populations (left/right: high/low) in the two hemispheres as well as a shifted feedback connectivity from AV to HD populations effectively implement the odd (or shifted) connectivity needed to turn the bump position (here: clockwise shift for anti-clockwise turn).

e) Broad excitation of the INH population by the HD population, together with a one-to-one multiplicative interaction between INH and HD population, implement the quadratic decay of the bump amplitude needed for probabilistic path integration.

f) External input is mediated by inhibiting HD neurons with preferred direction *opposite* the location of the absolute HD observation, effectively implementing a vector sum of belief with absolute heading input.

Signatures of Bayesian inference in neural population activity

Our network model makes specific predictions for neuronal activities. These predictions come in two types. The first type arises solely from the combination of circKF (Eqs. (1) and (2)) and IPPC encoding (Eq. (3)), but is agnostic to the exact network topology implementing this filter. The first two predictions are of this type. We nonetheless illustrate them using our CX-like network model (Fig. 5c). The second type, which becomes relevant for the last prediction, arises from the specific choices we made to implement this filter in a CX-like network.

Strongly conflicting absolute HD cues transiently decreases certainty of estimate. The circular Kalman filter predicts that strongly conflicting heading direction cues, such as HD observations opposite of the current belief, should decrease the certainty of the estimate (Fig. 1e). Thus, we predict that a strongly conflicting HD observation, such as a sudden rotation of the visual input, should result in a transient reduction of bump amplitude (see Fig. 6a, b for example). The circKF furthermore predicts how the mean estimate changes after such a jump of absolute HD observation ought to depend on the observation's reliability right after the jump. Absolute HD observations with low reliability should cause the HD estimate to slowly drift to the observation's position. Highly reliable observations instead should result in a sudden, albeit slightly delayed, HD estimate correction. These circKF predictions are mirrored in the activity dynamics of our network model (Fig. 6a, b).

Amplitude decreases during noisy angular path integration. In the absence of absolute HD observations, our model predicts that changes in bump positions should correlate with the fly's true angular velocity, but noise in the angular velocity observations should cause the position to progressively deviate from the fly's true heading. This progressive deviation has been observed for the position of the *Drosophila* EB activity bump (Seelig and Jayaraman 2015; Green et al. 2017) and is also present in our network dynamics (Fig. 6c). In parallel, certainty should decrease over time. As certainty is encoded in the bump amplitude, we predict this amplitude to decay over time accordingly (Fig. 6c, lower panel).

Activity of the inhibitory population. In our CX-like network, the inhibitory (INH) population's *overall activity* implements the quadratic decay of the bump amplitude (see SI for formal analysis). This quadratic decay is implemented by multiplicative feedback from the INH to the HD population. As a consequence, our model predicts the feedback from $\Delta 7$ to EPG neurons to be multiplicative, and the overall activity of the inhibitory $\Delta 7$ population to be proportional to encoded HD certainty. As HD certainty decays over time in the absence of reliable absolute HD observations, so should INH activity (Fig. 6d).

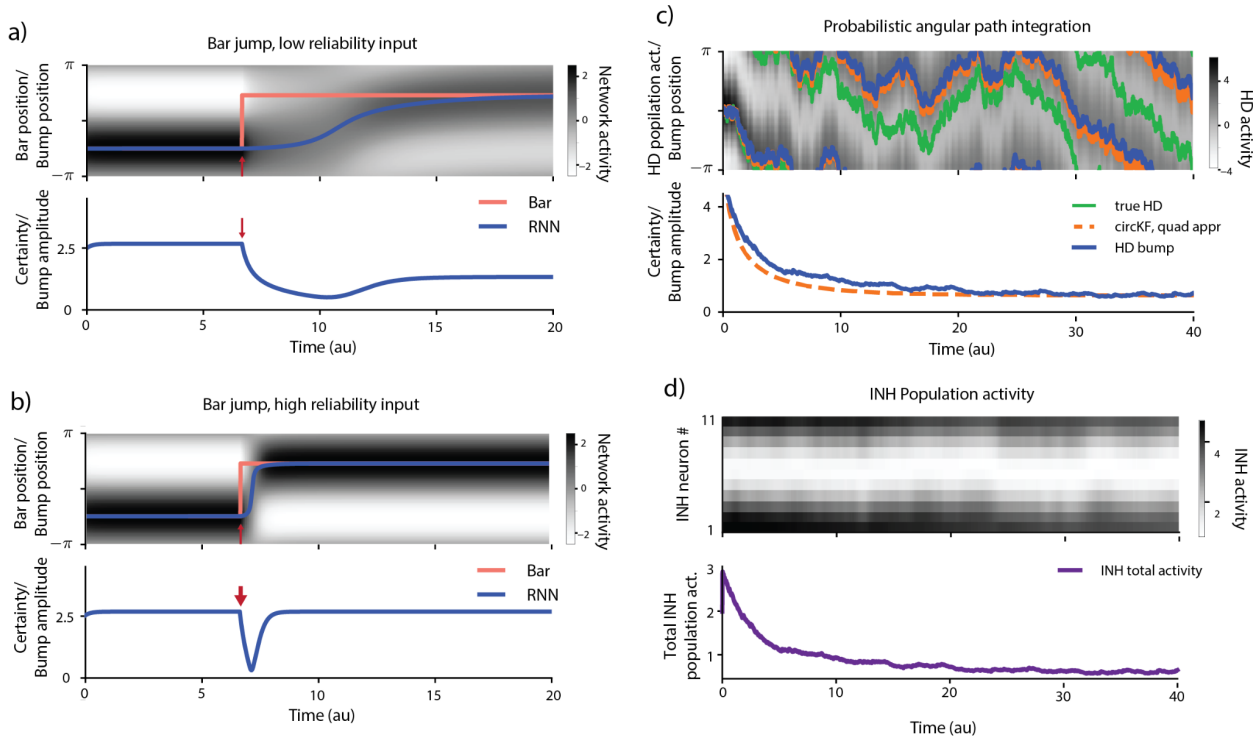

Figure 6. Predictions from encoding ideal observer.

a) Simulation of the bar jump experiments of Kim et al. (2017). We expose our network to a sudden, strongly conflicting absolute HD observation, similarly to a sudden bar jump in a virtual environment (Kim et al.2017). Here, unreliable HD observations result in a slow drift of the HD estimate/bump position towards the new bar position.

b) Highly reliable HD observations, in contrast, result in a rapid jump of the HD estimate/bump position to the new bar position after a brief delay. In both scenarios, the estimate's certainty briefly drops after the bar jump, as reflected by a decreased bump amplitude.

c) Top: simulation showing gradual error accumulation when only angular velocity observations (generated from the true HD time-course, green) are available, capturing features of noisy path integration. Error accumulation occurs in both the circKF mean estimate (orange) and the network's HD population bump position (blue; same network connectivity as in Fig. 5c). Bottom: As certainty decays over time, so does the network's bump amplitude. This amplitude follows the certainty of the quadratic approximation of the circKF, and thus does not decay to zero.

d) The total activity (sum of the activity of all neurons) of the INH population (from the same network simulation as in **a**) decays over time, mirroring the decay in certainty.

Discussion

We have shown how a ring-attractor network with biologically realistic anatomical constraints can perform Bayesian heading direction tracking. The network integrates angular velocity cues and absolute heading cues, and weighs them according to their reliability. To do so, it needs to

represent its own HD estimate and associated certainty, which it does in its neural population activity bump, mirroring the one observed in the *Drosophila* EB. Further, the network motifs required for Bayesian inference can readily be found in the connectome of the *Drosophila* CX, suggesting that, in principle, the fruit fly brain could indeed implement dynamic Bayesian inference to keep track of its heading direction.

One key element of our approach is the representation of uncertainty as the amplitude of a neural activity bump, which sets us apart from recent network models that only consider probabilistic cue integration on the level of the inputs (Sun et al. 2018; Sun et al. 2020). By restricting ourselves to an analytically tractable ring attractor network, we were able to almost exactly map the certainty dynamics of the ideal-observer circKF to the bump amplitude dynamics. Having the network implement the circKF rather than the Kalman filter fully accounts for the circular symmetry of HD estimation. Thus, unlike previous work (Wilson and Finkel 2009), our network does not suffer from imprecise inference once absolute HD observations strongly deviate from the current HD estimate. As a result, it yields fundamentally different predictions for strongly conflicting absolute heading direction stimuli (Fig. 1e). Further, our network performs close-to-optimal Bayesian inference irrespective of the presence or absence of absolute HD observations, unlike (Esnaola-Acebes et al. 2021), who showed that continuous ring-attractors can track orientation and compute the running circular average of an absolute HD stimulus with hand-tuned, rather than self-adjusting, circular weights. Lastly, our network is to our knowledge the first to fully account for the effect of *probabilistic* angular path integration in a principled way.

Even though our Bayesian HD tracking algorithm requires keeping track of the HD estimate's uncertainty, we have shown that imperfectly tuned ring-attractor networks can track heading reasonably well. In fact, even strict attractor networks with a fixed amplitude, and fixed associated uncertainty, can perform close-to-Bayesian cue integration (Figs. 5b; cf. also (Takiyama 2016)). This result raises the question of why neurons should encode uncertainty in the first place. First and foremost, for some animals, uncertainty influences their behavior directly to improve their performance. For instance, the homing behavior of the desert ant (Merkle and Wehner 2010) suggests that the performance gained from tracking one's uncertainty justifies the added complexity for doing so. Further, uncertainty appears to impact the neural encoding of other navigation-related variables. For example, when absolute visual cues are in conflict with path integration cues, grid cells in mouse medial entorhinal cortex are more likely to remap when the visual cues are more reliable (Campbell et al. 2021). Whether such strategic use of uncertainty is also present in *Drosophila* remains to be experimentally tested. Additionally identifying how uncertainty ought to be reflected in their neural activity, as we do here, is required for a comprehensive understanding of the role of uncertainty in the brain's computations.

In the CX-like network (Fig. 5), we only focused on the smallest possible set of neuron types sufficient to implement the Bayesian algorithm and ignored a range of cell types whose role remains to be clarified (Turner-Evans et al. 2020). Further, we assumed that the long-range inhibition required in ring-attractor networks is achieved exclusively by the recurrent connectivity between E-PG and $\Delta 7$ neurons, even though recent work suggests that this long-range inhibition is most likely split across multiple brain regions and classes of neurons (Turner-Evans et al. 2020). As we have demonstrated here, these shortcomings can likely be resolved by further extending

our network, as long as the qualitative ring attractor in the CX is maintained. Lastly, our encoding model assumed uncertainty to be encoded almost exclusively in the bump amplitude, while other bump parameters, such as the width, were not affected by uncertainty. Alternatively, uncertainty might modulate bump width, requiring an adjusted set of network motifs for doing so.

In summary, our work provides a first hypothesis for how the *Drosophila* brain could implement dynamic Bayesian inference. The uncovered network motifs generalize beyond *Drosophila* to other organisms, and might manifest in HD neuron dynamics in other animals, such as mice (Peyrache et al. 2015; Ajabi et al. 2021), monkeys (Robertson et al. 1999), humans (Baumann and Mattingley 2010), or even the dynamics of three-dimensional HD cells found in bats (Finkelstein et al. 2015). More generally, it demonstrates how classic network motifs, like ring attractor networks, can perform close-to-optimal Bayesian inference. Similar lessons could be learned about other networks.

Acknowledgements

We would like to thank Habiba Noamany (Wilson lab, HMS) for assisting us navigating the Neuprint database, and for informed comments on the manuscript. We would further like to thank Johannes Bill & Albert Chen (Drugowitsch lab, HMS) for discussions and feedback on the manuscript, and the entire Drugowitsch lab for valuable discussions.

The work was funded by the NIH (R34NS123819; R.I.W. and J.D.), the James S. McDonnell Foundation (Scholar Award #220020462; J.D.), the Swiss National Science Foundation (grant numbers P2ZHP2 184213 and P400PB 199242; A.K.), and a Grant in the Basic and Social Sciences by the Harvard Medical School Dean's Initiative award program (R.I.W. and J.D.).

Author contributions

Conceptualization, A.K., M.A.B., R.I.W., J.D.; Methodology, A.K., J.D.; Software, A.K.; Formal analysis, A.K., J.D.; Investigation, A.K, M.A.B., R.I.W., J.D.; Resources, R.I.W., J.D; Writing - Original Draft: A.K., J.D.; Writing - Review & Editing: A.K., M.A.B., R.I.W., J.D.; Visualization, A.K.; Supervision, R.I.W, J.D.; Funding Acquisition, A.K., R.I.W., J.D.

Declaration of interests

The authors declare no competing interests.

References

Ajabi, Zaki, Alexandra T. Keinath, Xue-Xin Wei, and Mark P. Brandon. 2021. "Population Dynamics of the Thalamic Head Direction System during Drift and Reorientation." Preprint. Neuroscience. https://doi.org/10.1101/2021.08.30.458266.

Baumann, O., and J. B. Mattingley. 2010. "Medial Parietal Cortex Encodes Perceived Heading Direction in Humans." *Journal of Neuroscience* 30 (39): 12897–901. https://doi.org/10.1523/JNEUROSCI.3077-10.2010.

Beck, J. M., P. E. Latham, and A. Pouget. 2011. "Marginalization in Neural Circuits with Divisive Normalization." *Journal of Neuroscience* 31 (43): 15310–19. https://doi.org/10.1523/JNEUROSCI.1706-11.2011.

Bill, Johannes, Hrag Pailian, Samuel J. Gershman, and Jan Drugowitsch. 2020. "Hierarchical Structure Is Employed by Humans during Visual Motion Perception." *Proceedings of the National Academy of Sciences* 117 (39): 24581–89. https://doi.org/10.1073/pnas.2008961117.

Brigo, D., B. Hanzon, and F. Le Gland. 1999. "Approximate Nonlinear Filtering by Projection on Exponential Manifolds of Densities." *Bernoulli* 5 (3): 495–534. https://doi.org/10.2307/3318714.

Campbell, Malcolm G., Alexander Attinger, Samuel A. Ocko, Surya Ganguli, and Lisa M. Giocomo. 2021. "Distance-Tuned Neurons Drive Specialized Path Integration Calculations in Medial Entorhinal Cortex." *Cell Reports* 36 (10): 109669. https://doi.org/10.1016/j.celrep.2021.109669.

Dehaene, Guillaume P., Ruben Coen-Cagli, and Alexandre Pouget. 2021. "Investigating the Representation of Uncertainty in Neuronal Circuits." *PLOS Computational Biology* 17 (2): e1008138. https://doi.org/10.1371/journal.pcbi.1008138.

Drugowitsch, Jan, Gregory C DeAngelis, Eliana M Klier, Dora E Angelaki, and Alexandre Pouget. 2014. "Optimal Multisensory Decision-Making in a Reaction-Time Task." *ELife* 3 (June): 748–56. https://doi.org/10.7554/eLife.03005.

Eliasmith, Chris. 2007. "Attractor Network." *Scholarpedia* 2 (10): 1380. https://doi.org/10.4249/scholarpedia.1380.

Ernst, Marc O, and Martin S Banks. 2002. "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion." *Nature* 415 (6870): 429–33. https://doi.org/10.1038/415429a.

Esnaola-Acebes, Jose M., Alex Roxin, and Klaus Wimmer. 2021. "Bump Attractor Dynamics Underlying Stimulus Integration in Perceptual Estimation Tasks." Preprint. Neuroscience. https://doi.org/10.1101/2021.03.15.434192.

Fetsch, Christopher R., Amanda H. Turner, Gregory C. DeAngelis, and Dora E. Angelaki. 2009. "Dynamic Reweighting of Visual and Vestibular Cues during Self-Motion Perception." *Journal of Neuroscience* 29 (49): 15601–12. https://doi.org/10.1523/JNEUROSCI.2574-09.2009.

Finkelstein, Arseny, Dori Derdikman, Alon Rubin, Jakob N. Foerster, Liora Las, and Nachum Ulanovsky. 2015. "Three-Dimensional Head-Direction Coding in the Bat Brain." *Nature* 517 (7533): 159–64. https://doi.org/10.1038/nature14031.

Fisher, Yvette E, Jenny Lu, Isabel D'Alessandro, and Rachel I Wilson. 2019. "Sensorimotor Experience Remaps Visual Input to a Heading-Direction Network." *Nature*, no. December 2018 (November). https://doi.org/10.1038/s41586-019-1772-4.

Glaze, Christopher M., Joseph W. Kable, and Joshua I. Gold. 2015. "Normative Evidence Accumulation in Unpredictable Environments." *ELife* 4 (AUGUST2015). https://doi.org/10.7554/eLife.08825.

Green, Jonathan, Atsuko Adachi, Kunal K. Shah, Jonathan D. Hirokawa, Pablo S. Magani, and

Gaby Maimon. 2017. “A Neural Circuit Architecture for Angular Integration in Drosophila.” *Nature* 546 (7656): 101–6. https://doi.org/10.1038/nature22343.

Hansel, David, and Haim Sompolinsky. 1998. “Modeling Feature Selectivity in Local Cortical Circuits.” *Methods in Neuronal Modeling: From Ions to Networks*, 69.

Hulse, Brad K, Hannah Haberkern, Romain Franconville, Daniel B Turner-Evans, Shin-ya Takemura, Tanya Wolff, Marcella Noorman, et al. 2021. “A Connectome of the Drosophila Central Complex Reveals Network Motifs Suitable for Flexible Navigation and Context-Dependent Action Selection.” Edited by Ronald L Calabrese. *ELife* 10 (October): e66039. https://doi.org/10.7554/eLife.66039.

Hulse, Brad K., and Vivek Jayaraman. 2020. “Mechanisms Underlying the Neural Computation of Head Direction.” *Annual Review of Neuroscience* 43 (1): 31–54. https://doi.org/10.1146/annurev-neuro-072116-031516.

Kakaria, Kyobi S., and Benjamin L. de Bivort. 2017. “Ring Attractor Dynamics Emerge from a Spiking Model of the Entire Protocerebral Bridge.” *Frontiers in Behavioral Neuroscience* 11 (February): 1–13. https://doi.org/10.3389/fnbeh.2017.00008.

Kalman, R E. 1960. “A New Approach to Linear Filtering and Prediction Problems.” *Transactions of the ASME Journal of Basic Engineering* 82 (Series D): 35–45. https://doi.org/10.1115/1.3662552.

Kalman, R. E., and R. S. Bucy. 1961. “New Results in Linear Filtering and Prediction Theory.” *Journal of Basic Engineering* 83 (1): 95–108. https://doi.org/10.1115/1.3658902.

Kim, Sung Soo, Ann M Hermundstad, Sandro Romani, L F Abbott, and Vivek Jayaraman. 2019. “Generation of Stable Heading Representations in Diverse Visual Scenes.” *Nature*, no. December 2018: 1–6. https://doi.org/10.1038/s41586-019-1767-1.

Kim, Sung Soo, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. 2017. “Ring Attractor Dynamics in the Drosophila Central Brain.” *Science* 356 (6340): 849–53. https://doi.org/10.1126/science.aal4835.

Kloeden, Peter E., and Eckhard Platen. 2010. *Numerical Solution of Stochastic Differential Equations*. Corr. 3. print. Applications of Mathematics 23. Berlin: Springer.

Knill, David C., and Alexandre Pouget. 2004. “The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation.” *Trends in Neurosciences* 27 (12): 712–19. https://doi.org/10.1016/j.tins.2004.10.007.

Kropff, Emilio, James E. Carmichael, May Britt Moser, and Edvard I. Moser. 2015. “Speed Cells in the Medial Entorhinal Cortex.” *Nature* 523 (7561): 419–24. https://doi.org/10.1038/nature14622.

Kurz, Gerhard, Igor Gilitschenski, and Uwe D. Hanebeck. 2016. “Recursive Bayesian Filtering in Circular State Spaces.” *IEEE Aerospace and Electronic Systems Magazine* 31 (3): 70–87. https://doi.org/10.1109/MAES.2016.150083.

Kutschireiter, Anna, Luke Rast, and Jan Drugowitsch. 2021. “Projection Filtering with Observed State Increments with Applications in Continuous-Time Circular Filtering.” *ArXiv:2102.09650 [Cs, Stat]*, September. http://arxiv.org/abs/2102.09650.

Kutschireiter, Anna, Simone Carlo Surace, Henning Sprekeler, and Jean-Pascal Pfister. 2017. “Nonlinear Bayesian Filtering and Learning: A Neuronal Dynamics for Perception.” *Scientific Reports* 7 (1): 8722. https://doi.org/10.1038/s41598-017-06519-y.

Lyu, Cheng, L.F. Abbott, and Gaby Maimon. 2020. “A Neuronal Circuit for Vector Computation Builds an Allocentric Traveling-Direction Signal in the *Drosophila* Fan-Shaped Body.” Preprint. Neuroscience. https://doi.org/10.1101/2020.12.22.423967.

Ma, Wei Ji, Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. 2006. “Bayesian Inference with Probabilistic Population Codes.” *Nature Neuroscience* 9 (11): 1432–38. https://doi.org/10.1038/nn1790.

Mardia, Kanti V., and Peter E. Jupp. 2000. *Directional Statistics*. John Wiley & Sons. https://doi.org/10.1002/9780470316979.

Merkle, Tobias, and Rüdiger Wehner. 2010. “Desert Ants Use Foraging Distance to Adapt the Nest Search to the Uncertainty of the Path Integrator.” *Behavioral Ecology* 21 (2): 349–55. https://doi.org/10.1093/beheco/arp197.

Moser, Edvard I., Emilio Kropff, and May-Britt Moser. 2008. “Place Cells, Grid Cells, and the Brain’s Spatial Representation System.” *Annual Review of Neuroscience* 31 (1): 69–89. https://doi.org/10.1146/annurev.neuro.31.061307.090723.

Murray, Richard F., and Yaniv Morgenstern. 2010. “Cue Combination on the Circle and the Sphere.” *Journal of Vision* 10 (11): 15–15. https://doi.org/10.1167/10.11.15.

Okubo, Tatsuo S, Paola Patella, Isabel D’Alessandro, and Rachel I Wilson. 2020. “A Neural Network for Wind-Guided Compass Navigation.” *Neuron*, July, 1–17. https://doi.org/10.1016/j.neuron.2020.06.022.

Omoto, Jaison Jiro, Mehmet Fatih Keleş, Bao Chau Minh Nguyen, Cheyenne Bolanos, Jennifer Kelly Lovick, Mark Arthur Frye, and Volker Hartenstein. 2017. “Visual Input to the Drosophila Central Complex by Developmentally and Functionally Distinct Neuronal Populations.” *Current Biology* 27 (8): 1098–1110. https://doi.org/10.1016/j.cub.2017.02.063.

Peyrache, Adrien, Marie M. Lacroix, Peter C. Petersen, and György Buzsáki. 2015. “Internally Organized Mechanisms of the Head Direction Sense.” *Nature Neuroscience* 18 (4): 569–75. https://doi.org/10.1038/nn.3968.

Pouget, Alexandre, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham. 2013. “Probabilistic Brains: Knowns and Unknowns.” *Nature Neuroscience* 16 (9): 1170–78. https://doi.org/10.1038/nn.3495.

Robertson, Robert G., Edmund T. Rolls, Pierre Georges-François, and Stefano Panzeri. 1999. “Head Direction Cells in the Primate Pre-Subiculum.” *Hippocampus* 9 (3): 206–19. https://doi.org/10.1002/(SICI)1098-1063(1999)9:3<206::AID-HIPO2>3.0.CO;2-H.

Scheffer, Louis K, C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J Hayworth, Gary B Huang, et al. 2020. “A Connectome and Analysis of the Adult Drosophila Central Brain.” Edited by Eve Marder, Michael B Eisen, Jason Pipkin, and Chris Q Doe. *ELife* 9 (September): e57443. https://doi.org/10.7554/eLife.57443.

Seelig, Johannes D., and Vivek Jayaraman. 2013. “Feature Detection and Orientation Tuning in the Drosophila Central Complex.” *Nature* 503 (7475): 262–66. https://doi.org/10.1038/nature12601.

———. 2015. “Neural Dynamics for Landmark Orientation and Angular Path Integration.” *Nature* 521 (7551): 186–91. https://doi.org/10.1038/nature14446.

Skaggs, William E, James J Knierim, Hemant S Kudrimoti, and Bruce L McNaughton. 1995. “A Model of the Neural Basis of the Rat’s Sense of Direction.” *Advances in Neural Information Processing Systems*, 10.

Stefanini, Fabio, Lyudmila Kushnir, Jessica C. Jimenez, Joshua H. Jennings, Nicholas I. Woods, Garret D. Stuber, Mazen A. Kheirbek, René Hen, and Stefano Fusi. 2020. “A Distributed Neural Code in the Dentate Gyrus and in CA1.” *Neuron*, 1–22. https://doi.org/10.1016/j.neuron.2020.05.022.

Sun, Xuelong, Michael Mangan, and Shigang Yue. 2018. “An Analysis of a Ring Attractor Model for Cue Integration.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10928 LNAI: 459–70. https://doi.org/10.1007/978-3-319-95972-6_49.

Sun, Xuelong, Shigang Yue, and Michael Mangan. 2020. “A Decentralised Neural Model Explaining Optimal Integration of Navigational Strategies in Insects.” Edited by Mani Ramaswami, Michael B Eisen, and Stanley Heinze. *ELife* 9 (June): e54026. https://doi.org/10.7554/eLife.54026.

Sun, Yi, Aljoscha Nern, Romain Franconville, Hod Dana, Eric R. Schreiter, Loren L. Looger, Karel Svoboda, Douglas S. Kim, Ann M. Hermundstad, and Vivek Jayaraman. 2017.

"Neural Signatures of Dynamic Stimulus Selection in Drosophila." *Nature Neuroscience* 20 (8): 1104–13. https://doi.org/10.1038/nn.4581.

Takiyama, Ken. 2016. "Bayesian Estimation Inherent in a Mexican-Hat-Type Neural Network." *Physical Review E* 93 (5): 052303. https://doi.org/10.1103/PhysRevE.93.052303.

Taube, J S, Robert U Muller, and J B Ranck. 1990. "Head-Direction Cells Recorded from the Postsubiculum in Freely Moving Rats. I. Description and Quantitative Analysis." *The Journal of Neuroscience* 10 (2): 420–35. https://doi.org/10.1212/01.wnl.0000299117.48935.2e.

Turner-Evans, Daniel B., Kristopher T. Jensen, Saba Ali, Tyler Paterson, Arlo Sheridan, Robert P. Ray, Tanya Wolff, et al. 2020. "The Neuroanatomical Ultrastructure and Function of a Biological Ring Attractor." *Neuron*, September, S0896627320306139. https://doi.org/10.1016/j.neuron.2020.08.006.

Turner-Evans, Daniel, Stephanie Wegener, Hervé Rouault, Romain Franconville, Tanya Wolff, Johannes D Seelig, Shaul Druckmann, and Vivek Jayaraman. 2017. "Angular Velocity Integration in a Fly Heading Circuit." *ELife* 6 (May): e23496. https://doi.org/10.7554/eLife.23496.

Wilson, Robert, and Leif Finkel. 2009. "A Neural Implementation of the Kalman Filter." *Advances in Neural Information Processing Systems* 22: 9.

Xie, Xiaohui, Richard H.R. Hahnloser, and H. Sebastian Seung. 2002. "Double-Ring Network Model of the Head-Direction System." *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 66 (4): 9–9. https://doi.org/10.1103/PhysRevE.66.041902.

Zhang, Kechen. 1996. "Representation of Spatial Orientation by the Intrinsic Dynamics of the Head-Direction Cell Ensemble: A Theory." *The Journal of Neuroscience* 16 (6): 2112–26. https://doi.org/10.1523/JNEUROSCI.16-06-02112.1996.

Methods

Ideal observer model

Here, we outline the derivation of the ideal observer model in Eqs. (1) and (2) in discrete time. For a more detailed discrete-time derivation please consult the SI. For a more formal and rigorous derivation based on a continuous-time stochastic differential equation (SDE) formulation please consult (Kutschireiter et al., 2021).

The overall goal is to derive an online estimator for the unobserved true HD $\phi_t \in [-\pi, \pi]$ at each point in time t, conditioned on a continuous stream of noisy angular velocity observations $v_{0:t} = \{v_0, v_{dt}, \dots v_t\}$ with $v_\tau \in \mathbb{R}$ and absolute HD observations $z_{0:t} = \{z_0, z_{dt}, \dots z_t\}$ with $z_\tau \in [-\pi, \pi]$.

We assume that these observations are generated from the (true) angular velocity $\dot{\phi}_t$ and HD ϕ_t that are corrupted by zero-mean noise at each point in time, via

$$v_t | \dot{\phi}_t \sim \mathcal{N}\left(\dot{\phi}_t, \frac{1}{\kappa_v \, dt}\right), \tag{4}$$

$$z_t | \phi_t \sim \mathcal{VM}\left(\phi_t, \sqrt{2\kappa_z \, dt}\right), \tag{5}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian with mean μ and variance σ^2, and κ_v denotes the precision of the angular velocity observations. We denote with $\mathcal{VM}(\mu, \kappa) = \frac{e^{\kappa \cos(\varphi - \mu)}}{2\pi I_0(\kappa)}$ the von Mises distribution of a circular random variable φ with mean μ and precision κ. In Eq. (5), the scaling of absolute observation precision with $\sqrt{2\kappa_z \, dt}$ ensures that the Fisher information of the observations about the true HD scales linearly in time and κ_z in the continuum limit $dt \to 0$ (Kutschireiter et al., 2021, Theorem 2).

In order to properly integrate information over time, we need to specify how the current HD ϕ_t depends on the past HD ϕ_{t-dt}, and how big we expect this change in HD to be over time. Specifically, we assume that, in the absence of further evidence, the HD diffuses on a circle,

$$\phi_t | \phi_{t-dt} \sim \mathcal{N}\left(\phi_{t-dt}, \frac{dt}{\kappa_\phi}\right) \mod 2\pi, \tag{6}$$

with a diffusion coefficient that decreases with κ_ϕ. Setting $\kappa_\phi \to 0$ implies that arbitrary HD changes are possible across consecutive time steps. Combining this dynamic prior with the noisy angular velocity observations in Eq. (4) gives rise to a "modified diffusion" (see SI), conditioned on the angular velocity observations:

$$\phi_t | \phi_{t-dt}, v_t \sim \mathcal{N}\left(\phi_{t-dt} + \frac{\kappa_v}{\kappa_\phi + \kappa_v} v_t \, dt, \frac{dt}{\kappa_\phi + \kappa_v}\right) \mod 2\pi. \tag{7}$$

In the results section, we refrained from specifying a dynamic prior by setting $\kappa_\phi \to 0$, which does not affect our conceptual results. However, Eq. (6) is used in our simulations to draw trajectories from the 'true' HD over time with $\kappa_\phi \neq 0$, and therefore needs to be accounted for in the inference. Thus, we added it here for completeness and reproducibility.

The online estimation problem we would like to solve requires finding the posterior $p(\phi_t|v_{0:t}, z_{0:t})$ and updating it recursively in time as more observations become available. Formally, this corresponds to solving the following:

$$p(\phi_t|v_{0:t}, z_{0:t}) \propto_{\phi_t} p(z_t|\phi_t) \int d\phi_{t-dt}\, p(\phi_t|\phi_{t-dt}, v_t) p(\phi_{t-dt}|v_{0:t-dt}, z_{0:t-dt}). \tag{8}$$

This recursion only relies on the likelihood of the absolute HD observations, the modified diffusion transition density, as well as the posterior at the last time step, $p(\phi_{t-dt}|y_{0:t-dt}, z_{0:t-dt})$. Since such a recursion cannot be solved in closed form for circular state-spaces (Kurz et al., 2016), we employ the following approximation: for each moment in time, we assume that the posterior takes the form of a von Mises-distribution, that is, $p(\phi_t|y_{0:t}, z_{0:t}) \approx \mathcal{VM}(\mu_t, \kappa_t)$. This makes the solution fully characterized by the mean μ_t and precision κ_t. As a consequence, HD tracking reduces to determining proper update equations for these two posterior parameters.

An alternative to specifying the parameters of the von Mises distribution through mean μ_t and precision κ_t is to use its natural parameters, $\theta_t = (\kappa_t \cos\mu_t, \kappa_t \sin\mu_t)^T$, that is, the polar coordinates of (μ_t, κ_t). As we show in the SI, this re-parametrization makes the integration of absolute HD observations, Eq. (5), exact and straightforward. In fact, it becomes a vector addition. In contrast, the convolution in Eq. (8) between the Gaussian transition density and the von Mises posterior relies on the aforementioned approximation of the posterior: using a method called projection filtering (Brigo et al., 1999), we minimize the Kullback-Leibler divergence between the outcome of the convolution and a von Mises distribution, to get closed-form dynamic equations for posterior mean and precision. In the continuum limit, $dt \to 0$, these two steps are taken simultaneously at each point in time, and we arrive at the **Circular Kalman Filter (circKF)**:

$$d\mu_t = \frac{\kappa_v}{\kappa_\phi + \kappa_v} v_t\, dt + \frac{\sqrt{2\kappa_z\, dt}}{\kappa_t} \sin(z_t - \mu_t), \tag{9}$$

$$d\kappa_t = -\frac{f(\kappa_t)}{2(\kappa_\phi + \kappa_v)} \kappa_t dt + \sqrt{2\kappa_z\, dt} \sin(z_t - \mu_t), \tag{10}$$

where $f(\kappa_t)$ is a monotonically increasing nonlinear function,

$$f(\kappa) = \frac{A(\kappa)}{\kappa_t - A(\kappa) - \kappa A(\kappa)^2}, \quad \text{with } A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}, \tag{11}$$

and $I_0(\cdot)$ and $I_1(\cdot)$ denote the modified Bessel functions of the first kind of order 0 and 1, respectively. For the text in the results, we assume for simplicity that $\kappa_\phi \to 0$, that is, the HD can a-priori change arbitrarily across consecutive time steps. This yields Eqs. (1) and (2). Note that conceptually, setting $\kappa_\phi \to 0$ does not change the general vector operations we present in Fig. 1.

If κ is sufficiently large, the nonlinearity $f(\kappa)$ can be approximated by a linear function, $f(\kappa) \approx 2\kappa - 2$. In our **quadratic approximation**, we replace the non-linearity by a quadratic decay:

$$d\kappa_t = -\frac{1}{\kappa_\phi + \kappa_v} \left(\kappa_t^2 - \kappa_t\right)\, dt + \sqrt{2\kappa_z\, dt} \sin(z_t - \mu_t), \tag{12}$$

which thus approximates the circKF for large certainties. Equations (9), (10) and (12) are expressed in form of stochastic differential equations (SDEs), i.e. infinitely small increments, to fully account for the time scaling in the absolute HD observation update.

Encoding and decoding with linear probabilistic population codes

Encoding

Our encoding model is based on a **linear probabilistic population code** (**IPPC**, (Ma et al., 2006; Beck et al., 2011)). The central idea behind these models is that i) neuronal populations encode a probability distribution of the exponential family, rather than a single point estimate, and that ii) the natural parameters of the encoded distribution can be retrieved from the population profile using linear operations. For our purposes this implies that we can linearly decode the natural parameters $\boldsymbol{\theta}_t = (\kappa_t \cos \mu_t, \kappa_t \sin \mu_t)^T$ of the von Mises distribution from population activity.

Here, we assume that the neuronal population consists of independent Poisson neurons, which densely tile the true HD, ϕ_t, stimulus space. We further require that these neurons encode a von Mises distribution with an IPPC, which gives rise to a von Mises-shaped tuning curve of a neuron i with preferred HD $\phi_i \in [-\pi, \pi]$, and gain controlled by certainty κ_t. The resulting population activity conditional on HD will then also be von Mises-shaped (see SI for proof), with a population activity gain that is, again, proportional to certainty.

Note that tuning to heading direction ϕ_t can only be measured if we have access to the encoded HD estimate. To instead find the tuning with respect to μ_t and κ_t that parameterize the distribution of ϕ_t, we need to average the neuron's tuning for a given μ_t and κ_t over all possible realizations of ϕ_t. This results in the following tuning with respect to μ_t and κ_t:

$$f_i(\mu_t, \kappa_t) = \int_{-\pi}^{\pi} d\phi_t \, f_i(\phi_t) \, \mathcal{VM}\left(\phi_t; \mu_t, \kappa_t\right) = \kappa_t \frac{I_0(\tilde{\kappa}_t^{(i)})}{I_0(\kappa_t)}, \tag{13}$$

with $\tilde{\kappa}_t^{(i)} = \sqrt{\xi^2 + \kappa_t^2 + 2\xi\kappa_t \cos(\phi_i - \mu_t)}$, and ξ an additional parameter that controls the width of the tuning curves (to keep the analysis simple, we set $\xi = 1$). This tuning curve is again bump-shaped with peak at μ_t and height modulated by certainty κ_t.

Decoding

While the encoding model, Eq. (13), describes how the parameters of the von Mises distribution change the activity profile of the neurons, it does not directly tell us how to estimate μ_t and κ_t from population activity. The population's qualitative tuning properties (Fig. 2f) suggest that these parameters can be determined from the first Fourier components of the profile, which we confirmed numerically (Supplementary Fig. S1). This is not surprising, as the Fourier transform is a linear operation, and linear decodability of the natural parameters is one of the features of IPPCs for exponential family distributions. The natural parameters can thus be retrieved from a vector of N neural activities $\mathbf{r}_t = (r_{1,t}, \ldots, r_{N,t})^T \in \mathbb{R}^N$ with a decoder matrix $A \propto (\cos(\boldsymbol{\phi}^{(i)}), \sin(\boldsymbol{\phi}^{(i)}))^T$, such that $\boldsymbol{\theta}_t = A \cdot \mathbf{r}_t$.

Network model

Here, we describe a rate-based network model that implements (approximations of) the circKF, by encoding the von Mises posterior in the neuronal population $\mathbf{r}_t$. We focused on the simplest kind of network model to allow for such an approximation, which is of the form:

$$\dot{\mathbf{r}}_t = -\frac{1}{\tau}\mathbf{r}_t - g(\mathbf{r}_t)\mathbf{r}_t + W \cdot \mathbf{r}_t + \mathbf{I}_t^{\text{ext}}, \tag{14}$$

where τ is the time constant, and $g : \mathbb{R}^N \to \mathbb{R}_+$ is a scalar nonlinearity. We decompose the recurrent connectivity matrix into $W = \frac{w_0}{2}W^{\text{const}} + w_1^{\text{even}}W^{\cos} + w_1^{\text{odd}}W^{\sin}$, where W^{const} denotes a matrix with constant entries, and $W^{\cos}$ and $W^{\sin}$ refer to cosine- and sine-shaped connectivity profiles (Fig. 3a). Specifically, due to the circular symmetry, the entries of these matrices only depend on the relative distance in preferred HD, and are given by[1] $W_{ij}^{\text{const}} = \frac{2}{N}$, $W_{ij}^{\cos} = \frac{2}{N}\cos(\phi_i - \phi_j)$, and $W_{ij}^{\sin} = \frac{2}{N}\sin(\phi_i - \phi_j)$. We further consider a cosine-shaped external input of the form $I_t^{\text{ext}}(\phi_i) = I_t\cos(\Phi_t - \phi_i)$ that is peaked around an input location Φ_t.

As we show in the SI, a network of this form features the following dynamics of the cosine-profile parameters μ_t and κ_t:

$$d\mu_t = w_1^{\text{odd}}\,dt + \frac{I_t}{\kappa_t}\sin(\Phi_t - \mu_t), \tag{15}$$

$$d\kappa_t = \left(w_1^{\text{even}} - \frac{1}{\tau}\right)\kappa_t\,dt - g(r_t)\kappa_t\,dt + I_t\cos(\Phi_t - \mu_t). \tag{16}$$

To derive these dynamics, we assumed the following:

1. The network is *rate-based*, i.e. neural activities directly correspond to the average activity profile $r_t^{(i)} = f_i(\mu_t, \kappa_t)$.

2. Our analysis assumed a continuum of neurons, i.e. $N \to \infty$. For numerical simulations, and the network description below, we employed a finite-sized network of size N that corresponds to a discretization of the continuous network. SI Fig. S3 demonstrates only a very weak dependence of our results on the exact number of neurons in the network.

3. Our analysis and simulations focused on the first Fourier mode of the profile in Eq. (13). This is justified, as this mode contains all the information about the posterior (SI Fig. S1), and is independent of any other shape-modifying parameter (e.g., bump width or constant baseline).

Network parameters for Bayesian inference

The network in Eq. (14) can be tuned to exactly match the dynamics of its cosine-shaped bump component, i.e., amplitude and phase, with that of mean and certainty of the circKF (compare Eqs. (9) and (10) with Eqs. (15) and (16)). Specifically, we find for the network parameters:

[1] The scaling factor $\frac{2}{N}$ was chosen to facilitate matching our analytical results from the continuum network to the network structure outlined here.

- Odd recurrent connectivities are modulated by angular velocity observations, $w_1^{\text{odd}} = \frac{\kappa_v}{\kappa_\phi + \kappa_v} v_t$, which shifts the activity profile without changing its amplitude (Zhang, 1996; Skaggs et al., 1995).

- Absolute HD observations z_t are represented as the peak position Φ_t of a cosine-shaped external input, whose input strength is modulated by the reliability of the observation, i.e., $I_t = \sqrt{2\kappa_z \, dt}$. The inputs might contain additional Fourier modes (e.g., a constant baseline), but those do not affect the dynamics in Eqs. (15) and (16).

- The recurrent excitatory input through even connectivities needs to exactly balance the internal activity decay, i.e., $w_1^{\text{even}} - \frac{1}{\tau} = 0$.

- The decay nonlinearity is modulated by the reliability of the angular velocity observations, and is given by $g(r_t) = \frac{1}{2(\kappa_\phi + \kappa_v)} f\big(\kappa(r_t)\big)$, where f is equal the nonlinearity that governs the certainty decay in the circKF (Eq. (11)). This can be achieved, e.g., through interaction with in inhibitory neuron that computes the activity bump's amplitude. This neuron's activation function is then given by the nonlinearity f.

Quadratic approximation

To achieve an analytically more tractable interaction (with the added benefit that the bump amplitude is finite in absence of absolute HD observations, see below), we alternatively consider the recurrent inhibition to be of the form $g(\mathbf{r}_t)\mathbf{r}_t \to w^{\text{quad}}(M \cdot [\mathbf{r}_t]_+) \circ \mathbf{r}_t$, where $[\cdot]_+$ denotes the rectification nonlinearity, M is a constant matrix with entries $M_{ij} = \frac{\pi}{N}$, and $\circ$ denotes the elementwise (Hadamard) product. The resulting recurrent inhibition can be shown to be quadratic in the amplitude κ_t. In the large population limit, this leads to the amplitude dynamics (see SI for derivation)

$$d\kappa_t = \left(w_1^{\text{even}} - \frac{1}{\tau} \right) \kappa_t \, dt - w^{\text{quad}} \kappa_t^2 dt + I_t \cos(\Phi_t - \mu_t). \tag{17}$$

The dynamics of the phase μ_t in Eq. (15) are not affected by the choice of g, and thus remain described by Eq. (15). Such a network thus implements the quadratic approximation of the circKF, given by Eqs. (9) and (12), if we tune the its parameters such that $w^{\text{quad}} = \frac{1}{\kappa_\phi + \kappa_v}$ and $w^{\text{even}} - \frac{1}{\tau} = \frac{1}{\kappa_\phi + \kappa_v}$, while sensory input, i.e. angular velocity v_t and absolute HD observations z_t, enter in the same way as before.

Ring-attractor networks

In absence of absolute HD observations ($I_t = 0$), the amplitude dynamics in Eq. (17) has a stable **fixed point** at $\kappa^* = \frac{w^{\text{even}} - 1/\tau}{w^{\text{quad}}}$, while any value of the phase is equally favoured, hence constituting a ring-attractor network. The bump amplitude approaches this fixed point with **inverse effective time constant** $\beta = w^{\text{even}} - \frac{1}{\tau}$, which was determined by linearizing the dynamics around its fixed point. A large value of β denotes faster dynamics and thus indicates more rigid attractor dynamics. For the quadratic approximation network, we find $\kappa^* = 1$ and $\beta = \frac{1}{\kappa_\phi + \kappa_v}$. Further, in our simulations in Fig. 4, we explored network dynamics with a range of κ^* and β values by adjusting network parameters accordingly.

Multiple population network

We extended the single population network dynamics, Eq. (14), to encompass five populations: a HD population, which we designed to track HD estimate and certainty with its bump parameter dynamics, two angular velocity populations (AV+ and AV-), which are tuned to HD and are differentially modulated by angular velocity input, an inhibitory population (INH), and a population that mediates external input (EXT), corresponding to absolute HD observations. The resulting network dynamics become:

$$\dot{\mathbf{r}}_t^{HD} = -\frac{1}{\tau_{HD}}\mathbf{r}_t^{HD} + W_{HD\leftarrow HD}\cdot\mathbf{r}_t^{HD} + W_{HD\leftarrow AV+}\cdot\mathbf{r}_t^{AV+} + W_{HD\leftarrow AV-}\cdot\mathbf{r}_t^{AV-} \tag{18}$$

$$+ (W_{HD\leftarrow INH}\mathbf{r}_t^{INH}) \circ \mathbf{r}_t^{HD} + \mathbf{I}_t^{ext}, \tag{19}$$

$$\tau_{AV^+}\dot{\mathbf{r}}_t^{AV^+} = -\mathbf{r}_t^{AV^+} + (o^{AV} + v_t)W_{AV^+\leftarrow HD}\cdot\mathbf{r}_t^{HD}, \tag{20}$$

$$\tau_{AV^-}\dot{\mathbf{r}}_t^{AV^-} = -\mathbf{r}_t^{AV^-} + (o^{AV} - v_t)W_{AV^-\leftarrow HD}\cdot\mathbf{r}_t^{HD}, \tag{21}$$

$$\tau_{INH}\dot{\mathbf{r}}_t^{INH} = -\mathbf{r}_t^{INH} + W_{INH\leftarrow HD}\cdot[\mathbf{r}_t^{HD}]_+ + W_{INH\leftarrow INH}\cdot\mathbf{r}_t^{INH}. \tag{22}$$

Here, the $Wto \leftarrow from$ denote connectivities within and between populations, and o^{AV} is a constant baseline in the AV populations.

The network parameters are tuned such that the activity profile in the HD population tracks the dynamics of the circKF quadratic approximation, in the same way as for the single-population network, Eq. (14). To limit the degrees of freedom, we further constrained the connectivity structure between HD and AV+/- and INH populations by the known connectome of the *Drosophila* HD system (hemibrain dataset, Scheffer et al. 2020). Specifically, we focused on the connectivities between EPG, PEN1 and $\Delta 7$ neurons (which in our model corresponds to HD, AV+/- and INH neurons), sorted according to anatomical regions within the ellipsoid body and the protocerebral bridge (Fig. 5b). Thereby, we used total number of synaptic connections between two regions as a proxy for connection strength. We further assumed that interactions within AV+/- populations and between AV+/- and INH populations were negligible. The resulting connectivity profile in Fig. 5c was determined by matching the *Drosophila* connectome as closely as possible, while allowing for modulation of the connection strengths c_0^{HD},c_1^{HD}, $c^{AV^\pm\leftarrow HD}$, $c^{HD\leftarrow AV^\pm}$., $c_0^{INH\leftarrow HD}$, $c_1^{INH\leftarrow HD}$, c_0^{INH}, c_1^{INH}, and $c^{HD\leftarrow INH}$.

As shown in the SI, the following connectivity matrices implement the quadratic approximation of the circKF in the HD population.

Recurrent excitation within HD population:

$$\begin{aligned}(W_{HD\leftarrow HD})_{ij} &= \frac{2}{N_{HD}}\left(c_0^{HD} + c_1^{HD}\left[\cos\left(\phi_i^{HD} - \phi_j^{HD}\right)\right]_+\right),\\ \text{with } c_1^{HD} &= 2\left(\frac{1}{\kappa_\phi+\kappa_v} + \frac{1}{\tau_{HD}} - o^{AV}\frac{\kappa_v}{\kappa_\phi+\kappa_v}\right),\ \ c_0^{HD} < \frac{1}{2\tau} - \frac{c_1^{HD}}{\pi}.\end{aligned} \tag{23}$$

Recurrent excitation between HD and AV+ and AV- populations:

$$(W_{AV^{\pm}\leftarrow HD})_{ij} = c^{AV^{\pm}\leftarrow HD}\delta_{ij}, \tag{24}$$

$$(W_{HD\leftarrow AV^{\pm}})_{ij} = \frac{2}{N_{AV^{\pm}}} c^{HD\leftarrow AV^{\pm}} \left[\sin\left(\phi_i^{HD} - \phi_j^{AV^{\pm}} \pm \frac{\pi}{4}\right)\right]_+, \quad \text{with } c^{HD\leftarrow AV^{\pm}} = \frac{\sqrt{2}}{c^{AV^{\pm}\leftarrow HD}} \frac{\kappa_v}{\kappa_\phi + \kappa_v}. \tag{25}$$

Recurrent inhibition between HD and INH populations:

$$(W_{INH\leftarrow HD})_{ij} = \frac{2}{N_{HD}} \left(\frac{c_0^{INH\leftarrow HD}}{2} + c_1^{INH\leftarrow HD} \cos\left(\phi_i^{INH} - \phi_j^{HD}\right)\right), \tag{26}$$

$$(W_{INH\leftarrow INH})_{ij} = \frac{2}{N_{INH}} \left(\frac{c_0^{INH}}{2} + c_1^{INH} \cos\left(\phi_i^{INH} - \phi_j^{HD}\right)\right), \quad \text{with } |c_1^{INH\leftarrow HD}| << |1 - c_1^{INH}| \tag{27}$$

$$(W_{HD\leftarrow INH})_{ij} = c^{HD\leftarrow INH}\delta_{ij}, \quad \text{with } c^{HD\leftarrow INH} = -\frac{1}{\kappa_\phi + \kappa_v} \frac{\pi(1 - c_0^{INH})}{c_0^{INH\leftarrow HD}}. \tag{28}$$

This network still has a considerable number of degrees of freedom. That is, the baseline o^{AV}, network connectivity strengths $c^{AV^{\pm}\leftarrow HD}$, $c_0^{INH\leftarrow HD}$, $c_1^{INH\leftarrow HD}$, c_0^{INH}, c_1^{INH}, and time scales τ_{HD}, τ_{AV^+}, τ_{AV^-} and τ_{INH} can essentially be chosen freely. If the number of neurons N differs between populations, the δ_{ij}'s can be replaced by a normalized, Gaussian-shaped kernel with a finite width. For our analytical results to hold, we require $\tau_{HD} \gg \tau_{AV^+}, \tau_{AV^-}, \tau_{INH}$ (see SI). We further constrained the network by choosing $c_0^{INH\leftarrow HD} > 0$, $c_1^{INH\leftarrow HD} \leq 0$ and $|c_0^{INH\leftarrow HD}| > |c_1^{INH\leftarrow HD}|$, which leads to the broad excitatory input into the INH population, and the formation of an 'antibump', similarly to the one observed in $\Delta 7$ neurons (Turner-Evans et al., 2020).

Activities of the EXT population were assumed to give rise to a bump-shaped inhibitory input opposite the absolute HD observation, loosely related to how ring neurons mediate such input to the EPG neurons (Fisher et al., 2019; Kim et al., 2019). We thus modeled this bump-shaped input to the HD population directly without explicitly representing a dynamics of the EXT population:

$$I_{i,t}^{ext} = -2\sqrt{2\kappa_z dt}\left[\cos(\phi_i^{HD} - z_t + \pi)\right]_+. \tag{29}$$

Simulation details

Numerical integration

Our simulations in Figs. 3, 4 and 6 used artificial data that matched the assumptions underlying our models. In particular, the 'true' HD ϕ_t followed a diffusion on the circle, Eq. (6), and observations were drawn at each point in time from Eqs. (4) and (5). To simulate trajectories and observations, we used the Euler-Maruyama scheme (Kloeden and Platen, 2010), which supports the numerical integration of stochastic differential equations. Specifically, for a chosen discretization time step Δt, this scheme is equivalent to drawing trajectories and observations from Eqs. (6), (4) and (5)

directly while substituting $dt \to \Delta t$. The same time-discretization scheme was used to numerically integrate the SDEs of the circKF, Eqs (9) and (10), its quadratic approximation, Eq. (12), and the network dynamics, Eqs. (14) and (18)-(22).

Performance measures

To measure performance, in Figs. 3f, 4b and 4d we compute the circular average distance (Mardia and Jupp, 2000) of the estimate μ_T from the true HD ϕ_T at the end of a simulation of length $T = 20$ from $P = 10'000$ simulated trajectories by $m_1 = \frac{1}{P}\sum_{k=1}^{P} \exp\left(i\left(\mu_T^{(k)} - \phi_T^{(k)}\right)\right)$. The absolute value of the imaginary-valued circular average, $0 \leq |m_1| \leq 1$ denotes an empirical precision (or 'inference precision'), and thus measures how well the estimate μ_t matches the true HD ϕ_T. Here, a value of 1 denotes an exact match. The inference precision is related to the circular variance via Var$_{circ} = 1 - |m_1|$. In SI Fig. 2, we provide histograms with samples $\mu_T - \phi_T$ with different numerical values of $|m_1|$, to provide some intuition for the spread of estimates for a given value of the performance measure.

We estimated performance through such averages for all absolute HD observation reliabilities κ_z in Figs. 3f and 4b. For the inset of Fig. 4b, and for Fig. 4d, we additionally performed a grid search over the fixed-point amplitude κ^* (inset of Fig. 4b), or both the fixed-point amplitude κ^* and of the inverse time constant β (Fig. 4d). For each setting of κ^* and β we assessed the performance by computing an average over this performance for a range of observation reliability κ_z, weighted by how likely each observation reliability is a-priori assumed to be. The latter was specified by a log-normal prior,$p(\kappa_z) =$ Lognormal$(\mu_{\kappa_z}, \sigma^2_{\kappa_z})$, favouring intermediate reliabilitiy levels. We chose $\mu_{\kappa_z} = 0.5$ and $\sigma^2_{\kappa_z} = 1$ for the prior parameters, but our results did not strongly depend on this parameter choice. The performance loss shown in Fig. 4d also relied on such a weighted average across κ_z's for a particle filter benchmark (PF, see SI for details). The loss itself was then defined as $1 - \frac{\text{Performance}}{\text{Performance PF}}$.

Update weights

In Fig. 4c, we computed the weight with which a single observation with $|z_t - \mu_t| = 90°$ changes the estimate. We defined this weight as the change in HD estimate, normalized by the value of the maximum possible change, $w = \frac{\Delta\mu_t}{\pi} = \frac{1}{\pi}\tan^{-1}\frac{\alpha(\kappa_z\, dt)}{\kappa_t}$. Here, $\alpha(\kappa_z\, dt)$ denotes a function that ensures a linear scaling of the Fisher information with sampling time step (see. Kutschireiter et al. (2021), Theorem 2, for details about this function). Thus, by design of the observation model, the Fisher information of a single observation with reliability κ_z during a time interval Δt is given by $I_{z_t}(\phi_t) = \kappa_z\, \Delta t$. We plot the weight as a function of the Fisher information of a single update (how informative is the observation?) and the Fisher information of the current HD estimate (how informative is the current estimate?), which is given by

$$I_{\mu_t,\kappa_t}(\phi_t) = \mathbb{E}\left[\left(\frac{\partial}{\partial\phi}\log\mathcal{VM}(\phi,\mu_t,\kappa_t)\right)^2\right] = \kappa_t\frac{I_1(\kappa_t)}{I_0(\kappa_t)}. \tag{30}$$

Simulation Parameters and additional details

Times are measured in units of inverse diffusion time constant κ_ϕ. For all our simulations we set $\kappa_\phi = 1s$ for convenience. Figures were generated based on simulations with the following parameters:

- Figure 3e: $\kappa_v = 2$, $\kappa_z = 10$ (during 'Visual cue' bout), $\kappa_z = 0$ (during 'Darkness' bout), $\Delta t = 0.01$.
- Figure 3f, 4b, 4d: $\kappa_v = 1$, $T = 20$, $\Delta t = 0.01$, $P = 10000$.
- Figure 4e: $\kappa_v = 1$, $\kappa_z = 1$, $T = 10$, $\Delta t = 0.01$.
- Figure 6a: $\kappa_v = 0$, $T = 20$, $\kappa_z = 0.1$ (before cue jump), $\kappa_z = 0.001$ (after cue jump), $\Delta t = 0.01$. Network architecture according to single-population network in Eq. (14) with $N = 80$ units, time scale $\tau = 1$, connection strengths $w_0 = 0$ (which for this network is the only free parameter).
- Figure 6b: same as 6a, with $\kappa_z = 0.1$ after cue jump.
- Figure 6c, d: $\kappa_v = 5$, $T = 20$, $\kappa_z = \Delta t = 0.001$. Network architecture according to the full network in Eqs. (18)-(22), with baseline $o^{AV} = 0$, time scales $\tau_{HD} = 0.1$, $\tau_{AV^+} = \tau_{AV^-} = 0.01$, $\tau_{INH} = 0.001$, connection strengths $c_0^{HD} = -0.2$, $c_1^{HD} = 0$, $c^{AV^\pm \leftarrow HD} = 1$, $c_0^{INH \leftarrow HD} = 0.5$, $c_1^{INH \leftarrow HD} = -0.5$, $c_0^{INH} = 0.1$, $c_1^{INH} = 0$.

Trajectory simulations and general analyses were performed on a MacBook Pro (Mid 2019) running 2.3 GHz 8-core Intel Core i9. Parameter scans were run on the HMS O_2 HPC cluster. For all our simulations, we used Python 3.9.1 with NumPy 1.19.2. Jupyter notebooks, Python scripts, and data to reproduce the figures will be made available upon acceptance of the manuscript.