Continuous Bump Attractor Networks Require Explicit Error Coding for Gain Recalibration

Gorkem Secer[1,2], James J. Knierim[2,3,*], and Noah J. Cowan[1,4,*]

[1]Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA.

[2]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, MD, USA.

[3]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University, Baltimore, MD, USA.

[4]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD, USA.

[*]These authors jointly supervised the work.

February 12, 2024

Abstract

Internal representations of continuous variables are crucial to create internal models of the external world. A prevailing model of how brain maintains these representations is given by continuous bump attractor networks (CBANs). CBANs have been hypothesized as an underlying mechanism in a broad range of brain functions across different areas, such as spatial navigation in hippocampal/entorhinal circuits and working memory in prefrontal cortex. Through recurrent connections, a CBAN maintains a persistent activity bump, whose peak location can vary along a neural space, corresponding to different values of a continuous variable. To track the value of a continuous variable changing over time, a CBAN updates the location of its activity bump based on inputs that encode the changes in the continuous variable (e.g., movement velocity in the case of spatial navigation) – a process akin to mathematical integration. This integration process is not perfect and accumulates error over time. For error correction, CBANs can use additional inputs providing ground-truth information about the continuous variable's correct value (e.g., visual landmarks for spatial navigation). These inputs enable the network dynamics to automatically correct any representation error by shifting the activity bump toward the correct location. Recent experimental work on hippocampal place cells has shown that, beyond correcting errors, ground-truth inputs (e.g., visual landmarks) also fine-tune the gain of the integration process, a crucial factor that links the change in the continuous variable (e.g., movement velocity) to the updating of the activity bump's location. However, existing CBAN models lack this plasticity, offering no insights into the

neural mechanisms and representations involved in the recalibration of the integration gain. In this paper, we explore this gap by using a ring attractor network, a specific type of CBAN, to model the experimental conditions that demonstrated gain recalibration in hippocampal place cells. Our analysis reveals the neural mechanisms behind gain recalibration within a CBAN: Unlike error correction, which occurs through network dynamics based on ground-truth inputs, gain recalibration requires an additional neural signal that explicitly encodes the error in the network's representation. This error signal must be provided by some neurons whose firing rate varies monotonically with respect to one of two signals—either the instantaneous error or the time-integral of the error—for recalibration of the integration gain. Finally, we propose a modified ring attractor network as an example CBAN model that verifies our theoretical findings. Combining an error-rate code with Hebbian synaptic plasticity, this model achieves recalibration of integration gain in a CBAN, ensuring accurate representation for continuous variables.

1 Introduction

The brain's ability to represent and process continuous variables, such as location, time, and sensory information, is fundamental to our understanding and interaction with the external world. A compelling theoretical framework for how the brain constructs these representations is provided by continuous bump attractor networks (CBANs) in a diverse range of brain functions, such as orientation tuning in visual cortex [1], working memory [2,3], evidence accumulation and decision-making [4–7], and spatial navigation [8–11].

The CBAN is a class of recurrently connected neural network, which maintains persistent patterns of population activity through recurrent interactions among its neurons. This persistent activity typically forms the shape of a bell-curve like 'activity bump' when visualized on an appropriate topological arrangement (or manifold [12]) of neurons, such as a plane, circle, or torus. Although the shape of the activity bump is constrained by network dynamics, its center location can vary along this low-dimensional manifold, corresponding to different values of the encoded continuous variable. Neural activity consistent with these key properties of CBANs, namely, the activity bump and the low-dimensional manifold, have been observed in recordings from various regions of the mammalian brain that encode continuous variables [13–15]. More conclusive and direct evidence for CBANs has been found in the fly brain, where a biological CBAN encoding the fly's heading angle, a continuous variable, has been identified based on the connectome and a combination of techniques, such as calcium imaging and optogenetics [16–18]. While these experimental findings support the idea of brain circuits employing CBANs to represent continuous variables, the neural mechanisms that enable CBANs to accurately update their representations in response to changes in continuous variables remains incompletely understood. Understanding these mechanisms is critical to answer one of the big questions in neuroscience – how does the brain create accurate internal models of the external world, which is characterized by a myriad of continuous variables, such as location, time, and sensory information?

CBAN updates its representation of a continuous variable based on two distinct types of inputs. The first type provides 'absolute' information, namely, the true value of a continuous variable, such as spatial location relative to visual landmarks or the item to be held in the working memory. When this absolute

information is available, it provides localized input on the low-dimensional manifold of the CBAN about a location that is associated with the true value of the continuous variable. In response to this localized input, internal dynamics of the CBAN creates a 'sink' on its low-dimensional manifold toward which the activity bump gravitates, resulting in nearly perfect encoding of the continuous variable [19–21]. This phenomenon has been robustly observed in decades of research on spatial representations in the rodent brain. When visual landmarks are rotated around an open arena, the activity bump of the hippocampal place cell system and thalamic head direction system also rotated by a similar amount [22–26]. In contrast to the first type of inputs providing absolute information to the CBAN, the second type provides 'relative' information, namely, the changes in the continuous variable. Sources of such inputs may be, for instance, self-generated movements in the context of spatial navigation or sensory cues serving as pieces of evidence to be accumulated in the context of decision-making. In response to these inputs, the internal dynamics of the CBAN shifts the activity bump along the low-dimensional manifold. This process, akin to mathematical integration, ensures the bump's location reflects the value of the continuous variable. However, compared to the absolute information, the encoding accuracy of this 'integration' process based on the relative information depends critically on an additional factor, namely, the integration gain of the network that relates the cumulative change in the continuous variable to the updating of the bump location in a proportional manner [27, 28]. If this gain factor is miscalibrated, the result of the CBAN's integration begins drifting away from the true value of the continuous variable; that is, it accumulates error. In the simultaneous presence of absolute information sources, such as visual landmarks for spatial localization, these encoding errors are minimal as the errors are corrected by the absolute information through the aforementioned 'sink' mechanism. However, in the absence of absolute information sources, error accumulation does not stop, which may cause, for example, a CBAN integrating evidence to reach a decision threshold too soon (assuming low integration gain) or a CBAN integrating an animal's angular head velocity to veer off from the correct head direction. Thus, a finely tuned integration gain is crucial for a CBAN to accurately encode a continuous variable based on inputs with only relative information.

Present CBAN models of path integration treat the integration gain as a constant that is perfectly set via carefully chosen, hard-wired model parameters (e.g., synaptic weights) [11, 29–31] (but see [32]). However, recent data from time cells and place cells of the rodent hippocampal formation, hypothesized to rely on CBANs [33], showed that the integration gain is actually a plastic variable whose value is adjusted based on the feedback from absolute information sources. In the first study that demonstrated this phenomenon on place cells [34], the virtual visual landmarks, which provided the absolute information, were moved as a function of the animal's movement. This movement created persistent error between the encoded location and the actual location relative to the landmarks. Prolonged exposure to this conflict led to recalibration of the system's integration gain, altering it in a direction and by an amount that reduced the positional encoding error in the pure integration based on movement information. This recalibration was most evident after the landmarks were extinguished (i.e., when the absolute information to the putative CBAN was abolished); the space encoded by hippocampal cells during pure integration of movement information either expanded or contracted, depending on the direction of the preceding landmark manipulation. A more recent study demonstrated a similar recalibration effect for

time encoding [35]: extending or contracting the time intervals unexpectedly in a temporal bisection task led to rescaling of the time encoded by hippocampal neurons, analogous to rescaling of space in [34]. Therefore, an open question is how CBANs adjust their integration gain based on error feedback from absolute information sources.

In the present paper, we aim to address this open question and generate testable physiological predictions about the neural mechanisms of the gain recalibration in brain circuits that encode continuous variables. As a representative problem, we focus on hippocampal place coding and theoretically examine how visual landmarks, being the absolute information source, recalibrate the integration gain of a CBAN that encodes the animal's position on a circular track, as demonstrated experimentally in [34]. To this end, we first derive analytical expressions of the integration gain, along with a simple dynamical model of the activity bump's location in the ring attractor network, a specific type of CBAN for circular position encoding. In contrast to the previous work implicitly assuming the network's integration gain as a constant, global parameter independent of the bump location in the low-dimensional manifold, our analysis reveals that the integrator gain of a CBAN is a spatially distributed, possibly inhomogeneous, parameter. We then employed control theory techniques to dissect the necessary algorithmic conditions for accurate recalibration of this spatially distributed integration gain via feedback from the absolute information sources, i.e., the visual landmarks in our representative problem. Mapping these necessary conditions from the algorithmic level to the mechanistic level uncovered a key mechanistic requirement for gain recalibration. We found that, unlike correction of encoding errors that happens automatically through network dynamics when feedback from an absolute information source is available, without the need for an explicit representation of the error, the process of learning/recalibrating the integration gain requires an additional neural signal that *explicitly encodes the error* in the network's representation. This error signal must be provided by some neurons whose firing rate varies monotonically with respect to one of two signals—either the instantaneous error or the time-integral of the error—for recalibration of the integration gain. Finally, we propose a modified ring attractor network as an example CBAN model that verifies our theoretical findings. Combining an error-rate code with Hebbian synaptic plasticity, this model achieves recalibration of integration gain in a CBAN, ensuring accurate representation for continuous variables.

2 Model Setup: Ring Attractor Network

A continuous bump attractor network is a recurrently connected neural network in which neighboring neurons excite one another and inhibit distant neurons according to a connectivity pattern known as local excitation and global inhibition [1,38]. This connectivity gives rise to a persistent bump of activity as a stable, equilibrium state of the system. Invariance of the connectivity across the network leads to a continuum of such equilibrium states, called attractor states. Arrangement of neurons and the exact pattern of the recurrent connectivity determine the topology of this attractor. In the case of a ring attractor, neurons are arranged conceptually as a topological ring [11]. By sustaining an activity bump whose location can be shifted along the ring based on external inputs (i.e., relative and absolute information sources), a ring attractor network is well-suited to represent a variable on a closed curve

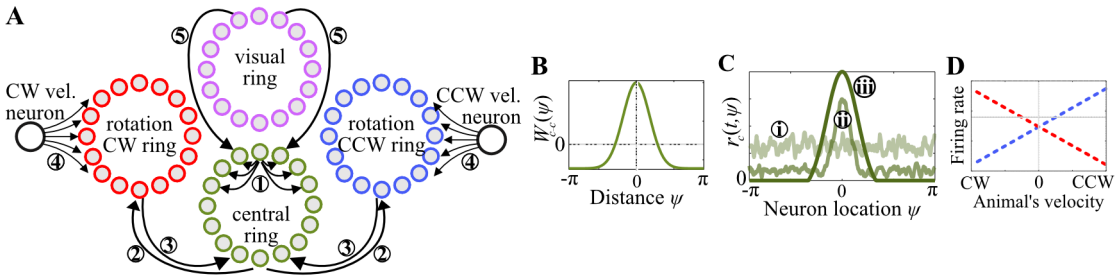

Figure 1: Ring attractor network model [10, 31, 36, 37]. (A) Schematic representation of the model. The central ring forms the main body of the model based on its recurrent connections (labeled with ①). The central ring is also reciprocally connected with the rotation CW and CCW rings (labeled with ② and ③), together creating a push-pull mechanism that modulates the intrinsically controlled neural activity of the central ring based on external inputs from the CW and CCW velocity neurons (labeled with ④). An additional external modulatory input is provided to the central ring from the visual ring (labeled with ⑤), corresponding to a set of sensory neurons that are tuned to visual landmarks. (B) Synaptic weight function $W_{\text{c-c}} : S^1 \to \mathbb{R}$ that describes the recurrent connections within the central ring according to the well-known local excitation and global inhibition pattern. (C) Numerical demonstration of how recurrent connectivity within the central ring can autonomously maintain a persistent activity bump. Simulation of the central ring neurons was started with initial conditions that are assigned pseudo-randomly (light green line labeled with ⓘ). Quickly within about 100 milliseconds, a bump of activity emerges (medium green line labeled with ⓘⓘ). Eventually, the firing rates converge to an equilibrium, forming a persistent bump of activity (dark green line labeled with ⓘⓘⓘ). (D) Tuning curves of CCW and CW velocity neurons shown with blue-dashed and red-dashed lines, respectively.

(e.g., angular location of an animal on a circular track). Augmenting the arrangement of neurons to a two dimensional plane results in a plane attractor whose activity bump is well-suited to represent two variables, for example, the x and y coordinates of location in a room [37].

The activities of place and grid cells in 2D environments have been traditionally modeled using a plane attractor [12, 31, 37]. However, in the present investigation of gain recalibration based on location encoding, originally demonstrated in place cells from rats running laps on a 1D circular track, we choose the ring attractor as the basis of our model because of its analytical tractability.

The ring attractor that we analyzed is a network model consisting of three groups of neurons ordered in a ring arrangement: a central ring, a clockwise (CW) rotation ring, and a counter-clockwise (CCW) rotation ring [8, 11], as depicted in Fig. 1A. Neurons in this network receive synaptic input from other neurons in the network via intrinsic connections between the neurons of the central and rotation rings and from upstream neurons carrying relative and absolute information via extrinsic connections. In the specific case of location encoding, the relative information conveys the animal's velocity and the absolute information conveys the positional feedback from visual landmarks.

We can mathematically model the dynamics of the ring attractor network using a set of equations that model the firing rate of a continuum of neurons in response to their synaptic inputs. By parameterizing a neuron based on its angle $\psi \in S^1$ in the circular neural space, we can model dynamics of the central ring neurons via

$$\tau_{\text{c}} \frac{\partial r_{\text{c}}(t,\psi)}{\partial t} = -r_{\text{c}}(t,\psi) + \sigma\left[W_{\text{c-c}}(\psi) \circledast r_{\text{c}}(t,\psi) + I_{\text{ext}}(t,\psi)\right], \tag{1}$$

where $r_{\text{c}}(t,\psi)$ denotes the firing rate of the central ring neuron ψ at time t, τ_{c} denotes the synaptic time constant of central ring neurons, $\circledast$ denotes the circular convolution operation, σ denotes an activation function (chosen as rectified linear unit (RELU) in our current study), $I_{\text{ext}}(t,\psi)$ denotes external synaptic inputs to the central ring, and $W_{\text{c-c}} : S^1 \to \mathbb{R}$ denotes a rotationally invariant synaptic weight function that describes the recurrent connections (① in Fig. 1A) according to the pattern known as local excitation

and global inhibition (Fig. 1B). This recurrent connectivity pattern is a hallmark of the continuous bump attractors, including the ring attractor, leading to stabilization of a persistent "bump" of activity within the network [38–40] (see Fig. 1C for a numerical illustration). While the shape of the emergent activity bump is determined by the shape of the recurrent connectivity pattern, the location of the bump can be controlled by external synaptic inputs to the central ring.

These external inputs to the central ring are provided by neurons of the rotation rings and the visual ring (providing the absolute, ground-truth information in the specific case of location encoding). The rotation rings conjoin the relative movement information with the positional information by receiving inputs from two different afferent neurons: CW and CCW 'velocity' neurons, encoding the animal's velocity differentially as shown in Fig. 1D, provide the movement information to their respective rotation rings through synaptic weight functions $W_{\text{v-cw}}$, $W_{\text{v-ccw}} : S^1 \to \mathbb{R}$ (④ in Fig. 1A), whereas the central ring provides the positional information to both rotation rings through synaptic weight functions $W_{\text{c-cw}}$, $W_{\text{c-ccw}} : S^1 \to \mathbb{R}$ (② in Fig. 1A). The resulting conjunctive codes of velocity and position within the rotation rings are then transmitted back to the central ring through *offset* synaptic weight functions $W_{\text{cw-c}}$, $W_{\text{ccw-c}} : S^1 \to \mathbb{R}$ (③ in Fig. 1A), leading to a *shift* of the persistent activity bump within the central ring proportional to the animal's velocity. In contrast to the rotation rings, neurons of the visual ring do not explicitly receive any inputs; instead, they are presumed to autonomously fire at specific locations of the animal relative to landmarks, capturing the absolute positional information received from visual landmarks available at each position (modeling how visual processes can calculate position from landmarks is beyond the scope of this paper). Through connections between visual and central rings (⑤, in Fig. 1A), this firing of the visual rings provides a bump-like synaptic input to the central ringencoding the animal's "true" position relative to landmarks. This bump-like input from the visual ring pulls the activity bump of the central ring, hence correcting positional errors in the ring attractor's representation.

Before concluding this section, we clarify an important distinction between traditional models of the ring attractor network and our model in the present paper. In traditional models, the synaptic weights $W_{\text{v-cw}}$, $W_{\text{v-ccw}}$ and $W_{\text{cw-c}}$,$W_{\text{ccw-c}}$ are treated as constants, each constrained to take a uniform value across the entire neural space. However, in our model, we intentionally relax this constraint. Instead, we treat these weights as functions that can vary throughout the neural space, possibly taking nonuniform values. While this approach may be less mathematically convenient, it becomes necessary when we later explore the possibility of gain recalibration through plasticity of spatially distributed synapses in the ring attractor network. As will be evident in the subsequent sections, our approach has broader implications, especially for the spatial metric of the ring attractor network. The complete mathematical model of the ring attractor, including the functional synaptic weights and the dynamics of the rotation and visual rings, is given in Appendix 6.1.

3 Algorithmic and Mechanistic Requirements for Gain Recalibration

In this section, we analyze the dynamics of our unconstrained ring attractor model to garner insight into how the network's integration gain, hereafter referred to as the path-integration (PI) gain to be

consistent with the spatial navigation terminology, can be recalibrated by visual landmarks. For the same terminological reasons, we will use the term path integration (PI) to refer to the integration computation carried out by the ring attractor. Dynamics of the ring attractor are high-dimensional and mathematically too complex to analyze. Thus, our analysis begins with reduction of these complex dynamics into a simple, one-dimensional differential equation model, including an analytical expression of the PI gain, in Section 3.1. Under certain assumptions, this model can accurately capture how the representation of the animal's location, decoded from the bump location, changes as the animal moves on the circular track. Leveraging the analytical tractability of this simple model, we then identify algorithmic conditions for gain recalibration, i.e., necessary conditions for the convergence of the PI gain to its true value within the simple model, in Section 3.2. Finally, we use the analytical expression of the PI gain to map the algorithmic conditions from the simple model to mechanistic prerequisites for gain recalibration within the high-dimensional network dynamics, in Section 3.3.

3.1 Dimensionality reduction reveals computational principles of the network

To derive approximate models for how the location of the persistent activity bump is controlled by external velocity and visual inputs within our unconstrained ring attractor network, we follow the dimensionality reduction protocol described in [41]. Briefly, the protocol exploits the fact that attractor network dynamics constrain the population activity to form a bump that looks like a 'bell-curve'. Based on this fact, the protocol assumes that the overall shape of the activity bump stays the same, but its center location in the neural space of neurons can change, depending on which neuron fires maximally. Although we do not know the exact solution to the network dynamics that can describe this activity pattern, we can 'guess' a solution form that describes the general properties of a bump without relying on a specific function. This guess, termed an *ansatz* solution, makes the analysis mathematically tractable.

In our case, the ansatz solution reduces the complex network dynamics to a one-dimensional differential equation that tracks the temporal change in the location of the activity bump along the neural space as a function of external inputs. Applying similar dimensionality reduction protocols, prior work derived one-dimensional differential equation models for the traditional ring attractors, where the synaptic weights $W_{\text{v-cw}}$, $W_{\text{v-ccw}}$ and $W_{\text{cw-c}}$,$W_{\text{ccw-c}}$ are constrained to be constant [41,42]. Here, we extend this approach to our unconstrained ring attractor. We develop the differential equation models progressively, starting from the simplest case that the animal is stationary in the absence of landmarks. Next, we add movement inputs and show that the network employs a spatially distributed PI gain. Finally, we extend the model to include landmark-based correction, and show that the network combines integration cues with landmarks in a computation that resembles a Kalman filter [43]. The resulting model forms the basis for our search for the necessary algorithmic conditions of the gain recalibration.

3.1.1 Ansatz solution to network dynamics

When the ring attractor employs a symmetric local excitation and global inhibition in the recurrent weight function $W_{\text{c-c}}$, the firing rates r_{c} of the central ring neurons converge to a persistent activity state

that can be characterized by the following solution

$$r_{\mathrm{c}}(t,\psi) = r_{\mathrm{c}}^{\star}(\psi - \theta(t)). \tag{2}$$

As depicted by the numerical example in Fig. 1C, the function $r_{\mathrm{c}}^{\star}$ underlying this solution takes the form of a symmetric bump, featuring a single peak at $\psi = \theta$ corresponding to the internal representation of the animal's position. We also assume that, at a given time t, only a subset of central ring neurons are active, restricting this bump function $r_{\mathrm{c}}^{\star}$ to take nonzero values in a limited range. Functions such as thresholded Gaussians (the von Mises function in the circular case of the ring attractor) or sinusoids possess these characteristics and are often employed to explain neurophysiological data [11, 19, 44–47]. Aside from these assumptions, we do not impose a further specific structure on the solution to firing rates of the central ring. This moderate generality allows us to consider Equation (2) a general form of the solution and ensures our analysis to be valid for a broad range of ring attractor models with various particular ansatzs (including, but not limited to, the commonly used ones mentioned above). See Appendix 6.2.1 for a formal description of the properties of the ansatz function in Equation (2).

The persistent activity in the central ring also spreads to the rotation rings through synaptic connections, resulting in persistent activity in these rings, which can be described as follows:

$$\begin{aligned} r_{\mathrm{cw}}(t,\psi) &\triangleq r_{\mathrm{cw}}^{\star}(\psi,\theta(t),v(t)) = \sigma\left[W_{\text{c-cw}}(\psi) r_{\mathrm{c}}^{\star}(\psi-\theta(t)) + W_{\text{v-cw}}(\psi)(u_{\mathrm{cw}}^{0} - \alpha_{\mathrm{cw}} v(t)) - \bar{I}\right] \\ r_{\mathrm{ccw}}(t,\psi) &\triangleq r_{\mathrm{ccw}}^{\star}(\psi,\theta(t),v(t)) = \sigma\left[W_{\text{c-ccw}}(\psi) r_{\mathrm{c}}^{\star}(\psi-\theta(t)) + W_{\text{v-ccw}}(\psi)(u_{\mathrm{ccw}}^{0} + \alpha_{\mathrm{ccw}} v(t)) - \bar{I}\right] \end{aligned} \tag{3}$$

where r_{cw}, r_{ccw} denote the instantaneous firing rates of the CW, CCW rotation rings, $r_{\mathrm{cw}}^{\star}$,$r_{\mathrm{ccw}}^{\star}$ denote the assumed form of the solutions to these firing rates, u_{cw}^{0}, u_{ccw}^{0} denote the baseline firing rates of the CW, CCW velocity neurons (during the animal's immobility), α_{cw}, α_{ccw} denote the absolute value of the slopes of the velocity neurons' tuning curves (i.e., the absolute change in their firing rates per unit velocity of the animal), and $\bar{I}$ denotes global inhibition. With a properly set inhibition $\bar{I}$, the solutions $r_{\mathrm{cw}}^{\star}$,$r_{\mathrm{ccw}}^{\star}$ take the shape of a bump, similar to the ansatz $r_{\mathrm{c}}^{\star}$ for the central ring (see Appendix 6.2 for further details).

Collectively, Equations (2) and (3) constitute a solution to the ring attractor dynamics. That is, for a given velocity of the animal, the firing rates of all neurons in a given ring attractor network can be computed using the ansatz $r_{\mathrm{c}}^{\star}$ that describes the shape of the persistent activity bump within the central ring and θ that denotes the location of the bump. While the shape of the bump is determined by the profile of the recurrent weight function $W_{\text{c-c}}$ of the central ring, the location θ of the bump, i.e., the position representation (Fig. 2A), is controlled by external inputs I_{ext} to the central ring. Therefore, assuming that the ansatz $r_{\mathrm{c}}^{\star}$ persists at all times, similar to the previous work [41, 42], we can reduce the high dimensional dynamics of the ring attractor to a one dimensional differential equation that models dynamics of the position representation θ as a function of the external inputs. In Appendix 6.2, we derive this differential equation for the baseline case, that is, the animal is stationary in the absence of visual landmarks. According to this derivation, if the central ring receives balanced (i.e., symmetric) inputs from the rotation rings during immobility, a classical assumption in computational models of the continuous bump attractor networks [3, 11, 31, 48, 49], the position representation θ remains invariant.

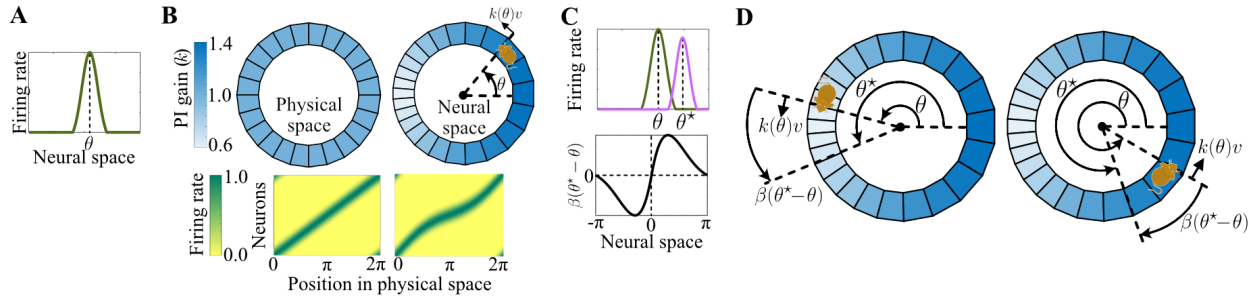

Figure 2: Models of the ring attractor's position representation. (A) The position representation θ, decoded from the peak location of an example ansatz solution $r_c^\star$ to the central ring's firing rates. (B) Model of path integration. Top left: Circular track with uniformly spaced points. Top right: internal representation of this track with a spatially inhomogenous path-integration (PI) gain that ranges from 0.6 at $\theta = \pi$ to 1.4 at $\theta = 0$. The position representation θ is visualized here by the rat. As the rat moves through physical space at velocity v, the representation moves through neural space at $k(\theta)v$. Bottom: Firing rate of uniformly distributed cells in the neural space as a function of the animal's position in physical space. Left shows a 'traditional' network model, including a single, global PI gain of 1. Right shows our unconstrained network model with the spatially inhomogenous PI gain in the top row. (C) Stabilizing visual feedback. Top: The central ring's activity bump $r_c^\star$ (green) and the bump-shaped synaptic input I_{vis} to the central ring from the visual ring (pink). The activities of both rings are aligned with respect to the same neural space, so in this example, the visual ring bump is "ahead of" the central ring bump. Bottom: The function β captures the stabilizing feedback from visual landmarks. Note that β operates on *the difference*, $\theta^\star - \theta$, so here, the x axis is a dummy variable. (D) Model of path integration with visual feedback. Left: the temporal change in the position representation is visualized as a cartoon rat, symbolizing the current representation as in (B), pulled by two arrows, one corresponding to updating by the path integration term $k(\theta)v$ and the other corresponding to updating by the visual feedback term $\beta(\theta^\star - \theta)$. Note that in this position, the PI gain is "low" and thus PI underestimates position relative to the landmarks. Right: Same as Left but PI overestimates position relative to the landmarks due to "high" PI gain in this position.

This invariance can be effectively modeled using the following ordinary differential equation:

$$\frac{d}{dt}\theta = 0. \tag{4}$$

3.1.2 Reduced-order path integration model reveals spatially distributed gain

Here we examine the case that the animal moves on a circular track in the absence of visual landmarks. Movement of the animal differentially modifies the activity of the CW and CCW rotation rings in a manner dependent on the movement direction and speed; for example, movement in the CW direction increases the firing rates of the CW rotation ring neurons that are also activated by the central ring and decreases corresponding neurons of the CCW rotation ring, both proportionally with the speed. Through synaptic connections, these movement-dependent changes in the activity of rotation rings are translated to a change in the synaptic inputs to the central ring such that the central ring no longer receives balanced excitation about its activity bump. This movement-based imbalance of excitation shifts the activity bump across the neural space, instantiating path integration.

To obtain a model that captures the temporal dynamics of the position representation during this process, we apply the dimensionality reduction protocol described in [41] (see Appendix 6.2.4 for details). Under the assumption that the magnitude of the movement-dependent changes is symmetric between the synaptic inputs from the CW and CCW rotation rings onto the central ring, this reduction protocol leads to an ordinary differential equation

$$\frac{d\theta}{dt} = k(\theta)v, \tag{5}$$

showing that the temporal change in the network's position representation θ is linearly related to the

animal's velocity v through a factor, termed the path-integration (PI) gain $k(\theta)$.

The analytical expression of the PI gain takes the form

$$k(\theta) = \frac{-b}{\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \frac{\partial^2 r_c^\star(\psi - \theta)}{\partial \psi^2} \sum_{i \in \{\text{cw},\text{ccw}\}} \alpha_i W_{i\text{-c}}(\psi) W_{\text{v-}i}(\psi) \, \text{sign}[r_i^\star(\psi, \theta, 0)] \, d\psi, \tag{6}$$

where i denotes the index of the summation, representing either the CW or CCW rotation ring, α denotes the absolute value of the slope of the velocity neurons' tuning curves, and b denotes the value of the offset in the connections between rotation rings and the central ring. This equation shows that the path integration gain $k(\theta)$ is a parameter determined by the network's functional properties (i.e., profiles of the activity bumps and the tuning slope of the velocity neurons) and structural properties (i.e., time constant of the neurons and synaptic weights of the velocity-to-rotation ring and rotation-to-central ring connections), thereby capturing the complex interaction between the network's persistent activity and its inputs from the velocity neurons. A comprehensive examination of how these network properties relate to the PI gain is provided later in Section 3.3 when we explore the mechanisms of gain recalibration.

As mentioned previously, in our model, we do not adopt the idealized assumption of the traditional models that constrains the synaptic weights of the velocity-to-rotation ring and rotation-to-central ring connections ($W_{\text{v-cw}}$,$W_{\text{v-ccw}}$ and $W_{\text{cw-c}}$,$W_{\text{ccw-c}}$, respectively) to be constants, each taking a uniform value in the entire neural space. Rather, we relax this constraint and treat the synaptic weights as functions that can vary along the neural space of the attractor network model, a heterogeneity likely to exist in biological networks. This functional treatment of the synaptic weights in our unconstrained model reveals an important characteristic of path integration in continuous bump attractor networks. That is, unlike traditional treatments that implicitly assume a single, spatially global PI gain, continuous bump attractor networks employ a distributed, possibly inhomogenous, gain factor that can vary as the position representation varies (Equation 6 and top row in Fig. 2B). This implies that, by employing an inhomogenous integrator gain, a bump attractor network can adjust its spatial resolution locally, which would result in 'overrepresentation' and/or 'underrepresentation' of certain locations (bottom row in Fig. 2B).

3.1.3 Landmark correction to path integration resembles a Kalman filter

Lastly, we examine the most general case: the animal moving on the circular track in the presence of landmarks. The simple model to be derived in this case will be crucial during our search for the algorithmic conditions of the gain recalibration in the next section. For this derivation, we must take into account the visual neurons that provide feedback from landmarks.

During path integration, visual feedback anchors the internal representation of position in a stable manner [22–25]. In traditional models of the ring attractor, this stabilizing feedback is achieved with an allocentrically anchored, bump-like synaptic current applied onto the central ring from the visual ring, encoding the animal's location relative to the landmarks [50, 51]. We adopt the same approach in our unconstrained ring attractor model and incorporate the following synaptic current from the visual ring

onto the central ring, namely

$$I_{vis}(t, \psi) = \rho_{\text{vis}}(\psi - \theta^\star), \tag{7}$$

where $\rho_{\text{vis}} : S^1 \to \mathbb{R}$ denotes a function describing the bump shape of this current, and $\theta^\star$ denotes the location of the peak current, corresponding to the animal's current position relative to landmarks. Note that, throughout the paper, we use the superscript * to distinguish a true value of a quantity (measured relative to the external world) from its internal value (e.g., $\theta^\star$ vs θ). For mathematical convenience, we assume that the function ρ_{vis} has the same general properties as the ansatz function $r_{\text{c}}^\star$, describing the central ring's persistent activity bump.

By applying the same dimensionality reduction technique [41], we can derive an equation that models how the velocity and visual inputs together change the central ring's position representation. As detailed in Appendix 6.2.5, under appropriate conditions, this equation takes the form

$$\frac{d\theta}{dt} = \beta\big(\theta^\star - \theta\big) + k(\theta)v, \tag{8}$$

where $\beta : S^1 \to \mathbb{R}^1$ is a function that, locally, takes the same sign as its argument (Fig. 2C). This sign property, a form of negative feedback, is crucial for stable control of the central ring's position representation; for example, in the absence of movement, the sign property of β ensures $\theta \to \theta^\star$. In the more general case, when the animal's velocity is nonzero, the position representation θ is changed by a combination of the (stabilizing) visual feedback $\beta : S^1 \to \mathbb{R}$ and the movement-related feedforward term $k(\theta)v$, as given in Equation (8). While the movement-related term, $k(\theta)v$, shifts the position representation proportionally with the animal's velocity, the function β acts in an additive manner to bring θ toward $\theta^\star$, the animal's true position relative to visual landmarks. This fusion between the feedforward path integration term and feedback correction from landmarks is illustrated in Fig. 2D. A properly balanced combination would stabilize θ around $\theta^\star$, with only minor deviations due to PI error. As a result, the ring attractor effectively "estimates" the position of the animal, a computation that bares striking similarity to a Kalman filter in engineering. Here, the state being estimated by the network is the animal's true position, $\theta^\star$, and the estimate is the location of the bump, θ.

3.2 Control Theory Reveals Algorithmic Conditions for Gain Recalibration

In the previous section, we presented a simple model that captures the temporal change in the ring attractor's position representation θ during the movements of the animal in the presence of landmarks. There are two key components of this model: 1) the PI gain k, which maps the animal's velocity v to a change in θ, and 2) the visual-feedback term β that corrects the error in θ relative to landmarks. Experiments showed that the PI gain is a plastic variable that can be recalibrated accurately by visual landmarks [34]. In this section, we use the simple model to seek analytical conditions that are necessary for this gain recalibration.

We begin by recalling the experimental conditions that brought about the recalibration of the PI gain [34]: An animal moved around a circular track while an array of visual landmarks was rotated around the track as a function of the animal's velocity and an experimentally controlled, visual gain

factor, $k^\star$. When $k^\star < 1$, the landmarks moved in the same direction as the animal, decreasing the animal's speed relative to the landmarks; when $k^\star > 1$, the landmarks moved in the opposite direction as the animal, increasing the animal's speed relative to the landmarks; when $k^\star = 1$ (veridical condition), the landmarks were stationary. Neural recordings showed that persistent exposure to these visual conditions recalibrated the animal's PI gain such that a tight correlation was observed between the average value of the PI gain measured over many laps after the landmarks were extinguished and the final value of the visual gain $k^\star$ before the landmarks were extinguished. Here, we analyze this recalibration using our simplified ring attractor model.

The visual conditions that gave rise to the gain recalibration can be simulated in a ring attractor with a visual synaptic drive revolving around the central ring at a rate proportional to the animal's velocity times the visual gain. Assuming such a visual drive, we extend Equation (8), the simple model for the attractor's position representation, with an additional variable $\theta^\star$ that denotes the visual drive's peak location. This extension leads to

$$\begin{aligned} \frac{d\theta^\star}{dt} &= k^\star v, \\ \frac{d\theta}{dt} &= \beta\big(\theta^\star - \theta\big) + k(\theta)v, \end{aligned} \tag{9}$$

where the first equation models the relation between the visual gain $k^\star$, the animal's velocity v, and the resulting temporal change in the visual drive's bump location $\theta^\star$, and the second equation models the change in the attractor's bump location θ based on the external velocity and visual inputs. Recall that, as we previously showed, the PI gain k of a ring attractor network is a spatially distributed parameter that may potentially take different values as the network's position representation θ varies with the animal's movement in the environment. However, in recalibration experiments [34], average neural activity over many laps was used to estimate the average value of the PI gain, providing no information about whether the PI gain took different values across the environment as predicted by the model. Therefore, from a theoretical perspective, the experimental result that the PI gain was recalibrated to the visual gain [34] only suggests that the *spatial average* of the PI gain converges to the visual gain $k^\star$. We can formulate such convergence via an equation

$$\lim_{t\to\infty} k_0(t) = k^\star, \tag{10}$$

where k_0 denotes the spatial average of the PI gain, namely

$$k_0 \triangleq \frac{1}{2\pi}\int_0^{2\pi} k(\theta)d\theta. \tag{11}$$

This asymptotic convergence of k_0 toward $k^\star$ necessitates k_0 to be updated over time during recalibration. The firing rate of neurons in the network is a biologically plausible candidate for controlling these updates. Therefore, to garner mathematical insight into this process, we searched for a general equation that could represent the updating of k_0 based on firing rates within the ring attractor, assuming an environment with spatially homogenous feedback from visual landmarks. Under very general assumptions detailed in Appendix 6.3.1, we found that updating of k_0 can be represented by a surprisingly simple equation, namely,

$$\frac{dk_0}{dt} = g_0(k_0, \theta^\star - \theta, v). \tag{12}$$

The update rule $g_0 : \mathbb{R} \times S^1 \times \mathbb{R} \to \mathbb{R}$ is an analytic function that instantiates the instantaneous change in k_0 based on three variables: the current gain k_0, the animal's velocity v and the difference between the visual drive's position representation $\theta^\star$ and the ring attractor's position representation θ, *without* directly depending on either θ or $\theta^\star$ (since the visual feedback is assumed to be uniform across the environment).

3.2.1 Necessity of matching sign in gain change and the product of error and velocity

The update rule g_0 could be any function that takes the form of Equation (12); however, some such functions may fail to result in gain recalibration, i.e., the PI gain's spatial average k_0 would not converge to the visual gain $k^\star$. What are the necessary properties of the gain update rule g_0 for convergence of k_0 to $k^\star$?

To seek these properties, we revisit Equation (8), the simple model of the central ring's and visual ring's position representations θ and $\theta^\star$, now taking into account that the PI gain's spatial average k_0 is time varying according to the gain update rule in Equation (12). Perfect convergence of k_0 to $k^\star$ through this update rule would imply that the error between these two gains,

$$\tilde{k} \triangleq k^\star - k_0, \tag{13}$$

approaches zero. When this gain error becomes zero with the completion of the recalibration process, it is intuitively expected that the difference $\tilde{\theta} \triangleq \theta^\star - \theta$, denoting the error in the attractor's position representation relative to the visual drive, also approaches zero. Hence, analyzing the temporal progression of these error terms—namely error in the PI gain and the error in the positional representation—provides an opportunity to garner insight into the algorithmic underpinnings of the gain recalibration process.

This analysis proceeds as follows. Let us assume that $dk^\star/dt = 0$, that is the visual gain is held constant. We can track how the spatial average of the PI gain recalibrates to this constant value by subtracting the second row of Equation (9) from its first row and subtracting Equation (12) from $dk^\star/dt$. These substractions yield the so-called *error dynamics*

$$\begin{aligned} \frac{d\tilde{\theta}}{dt} &= -\beta\big(\tilde{\theta}\big) + \tilde{k}v - k_{ac}(t,\theta)v \\ \frac{d\tilde{k}}{dt} &= -g_0(k_0, \tilde{\theta}, v), \end{aligned} \tag{14}$$

where $k_{ac}(\theta) := k(\theta) - k_0$ captures the potential variation of the spatially distributed PI gain from its spatial average.

Using tools from feedback control theory, we can analyze the error dynamics (Equation (14)) under this assumption to reach the ultimate goal of finding necessary conditions of the gain update rule for complete recalibration. First, we find that, if the ring attractor achieves and maintains zero gain error $\tilde{k}$=0, as required for complete recalibration, then it also maintains zero positional error $\tilde{\theta}$, rendering the origin $(\tilde{\theta}, \tilde{k}) = 0$ an equilibrium point of the error dynamics. For such convergence, however, the animal must be moving, else there exists no gain update rule that can achieve it. This is an unsurprising result,

because, were the animal stationary, the visual landmarks would correct all the positional error $\tilde{\theta}$, making the gain error $\tilde{k}$ imperceptible to the animal. That is, only through movement can errors in the path integration process become detectable. See Remark (10) for a simple mathematical proof. Assuming accordingly that the animal is always moving, we next analyze stability of the error dynamics. Such stability is necessary for convergence to zero gain error and zero positional error from different initial conditions and for different visual gains. This analysis reveals a sign requirement that must be satisfied by a gain update rule for complete recalibration: Except for a negligible number of gain values (in the sense that the values that are not one of these exceptions are infinitely many, compared to the exception values that are only finitely many.), the gain's spatial average k_0 must be updated in the same direction as the animal's velocity v times the ring attractor's positional error $\tilde{\theta}$. In the following proposition, we state this finding more formally:

Proposition 1 (Necessary conditions for complete gain recalibration)**.** *Let $\mathcal{K}$ be a bounded open interval in $\mathbb{R}$. If the PI gain's spatial average converges to the visual gain (i.e., $k_0^\infty = k^\star$) for all $k^\star$ in $\mathcal{K}$ and for all nonzero velocities of the animal (i.e., $\forall t, v(t) \neq 0$), then it requires*

$$g_0(k_0, 0, v) = 0, \quad \textit{for all } v \neq 0 \textit{ and } k_0, \tag{15}$$

as well as the sign requirement

$$\operatorname{sign}[g_0(k_0, \tilde{\theta}, v)] = \operatorname{sign}[\tilde{\theta}\, v], \ \textit{for all } v \neq 0 \textit{ and for almost all } k_0 \textit{ (except a finite set of isolated points)}, \tag{16}$$

in a non-empty open interval of the positional error $\tilde{\theta} = 0$ other than $\tilde{\theta} = 0$.

Proof. See Appendix 6.3.2. □

What if the biological system can be recalibrated only partially? That is, at the steady state, the PI gain's spatial average k_0 converges to a value biased towards, but not necessarily the same as, the visual gain $k^\star$. Under a simplifying assumption that the animal's velocity is constant, we can extend our findings in Prop. 1 for the complete recalibration to a more general case that covers both complete and partial recalibration. To this end, we analyzed the error dynamics (Equation 14) at the steady-state and found that convergence of the gain error $\tilde{k}$ to some value, say $\tilde{k}^\infty \triangleq \lim_{t\to\infty} \tilde{k}(t)$, after recalibration results in convergence of the ring attractor's positional error $\tilde{\theta}$ also, say to $\tilde{\theta}^\infty \triangleq \lim_{t\to\infty} \tilde{\theta}(t)$. If the gain recalibration is complete (i.e., $\tilde{k}^\infty = 0$), then this steady-state positional error $\tilde{\theta}^\infty$ is zero; if the recalibration is partial (i.e., $\tilde{k}^\infty \neq 0$), however, $\tilde{\theta}^\infty$ is nonzero, increasing with the product of the steady-state gain error $\tilde{k}^\infty$ and the animal's velocity v. We then analyzed stability of the error dynamics to understand the conditions that determine the convergence of the gain in the first place. We found a generalized sign requirement that must be satisfied by a gain update rule for recalibration: the gain's spatial average k_0 must be updated in the same direction as the animal's velocity v times the deviation of the current positional representation error $\tilde{\theta}(t)$ relative to a steady-state value $\tilde{\theta}^\infty$, namely,

$$\operatorname{sign}\big[g_0(k_0, \tilde{\theta}, v)\big] = \operatorname{sign}\big[(\tilde{\theta} - \tilde{\theta}^\infty)\, v\big]. \tag{17}$$

This generalized sign requirement captures both complete and partial recalibration; if, for example, the

steady-state positional error $\tilde{\theta}^{\infty}$ is zero, it reduces to Equation (16), the sign requirement for the special case of the complete gain recalibration. See Corollary 1 in Appendix 6.3.3 for a formal statement and derivation of these findings.

3.2.2 Sufficiency of positive gain change with respect to product of error and velocity

Thus far, we have presented the necessary algorithmic conditions for recalibration of the PI gain in Equations (15), (16) and (17). Although these conditions are required for gain recalibration, none of them are informative about whether a gain update rule would achieve gain recalibration.

Here, we present a sufficient condition for gain recalibration by again analyzing the stability of the error dynamics in Equation (14). The sufficient condition identifies a class of gain update rules that guarantee gain recalibration. Specifically, if a gain update rule has a positive slope with respect to the product of velocity v and positional error $\tilde{\theta}$ at its zero value (i.e., $g_0(\tilde{\theta}, v) = 0$), then at least partial gain recalibration is achieved for any visual gain $k^{\star}$ and a given velocity v of the animal. This sufficient condition is a small extension to the generalized necessary condition for recalibration, and gain recalibration is guaranteed if a gain update rule satisfies the generalized sign condition in Equation (17) together with the mild additional requirement that the gain update rule also enjoy a nonzero derivative with respect to the product of $\tilde{\theta}$ and v. For example, a linear update rule $g_0(k_0, \tilde{\theta}, v) = \mu\tilde{\theta}v$ with a positive slope μ satisfies the sufficient condition but a cubic rule $g_0(k_0, \tilde{\theta}, v) = \mu(\tilde{\theta}v)^3$ only satisfies the necessary condition. In the following proposition, we describe the sufficient condition more formally:

Proposition 2 (Sufficient conditions for gain recalibration)**.**
For a given velocity $v = v_0 \neq 0$ and visual gain $k^{\star}$, if the gain update rule satisfies the slope condition

$$\left.\frac{\partial g_0(k_0, \tilde{\theta}, v)}{\partial(\tilde{\theta}v)}\right|_{v=v_0} = \frac{\partial g_0(k_0, \tilde{\theta}, v_0)}{\partial\tilde{\theta}}\frac{1}{v_0} > 0 \tag{18}$$

everywhere in the solution set of $g_0(k_0, \tilde{\theta}, v_0) = 0$, then PI gain k_0 converges to a steady-state value.

Proof. See Appendix 6.3.4. □

We now would like to make this result more concrete with two specific examples. Although both satisfy the sufficient condition, they differ in the extent of resulting recalibration: Example 1 achieves complete recalibration, while example 2 achieves only partial recalibration. Details are below.

Example 1. *One of the simplest gain update rules that satisfy the slope condition in Equation (18) takes the form*

$$g_0(\tilde{\theta}, v) = \mu\,\tilde{\theta}\,v,$$

where μ denotes a positive learning rate. This update rule results in gain recalibration as per Proposition 2. Furthermore, the update rule takes the same sign as the product $\tilde{\theta}v$, thus also satisfying the necessary condition for complete recalibration (Fig. 3B).

Example 2. *This example is inspired from the modified ring attractor network that we propose in Section 4 as a model that achieves gain recalibration. Let μ denote a positive learning rate as before and η*

denote a constant. Consider the gain update rule

$$g_0(k_0, \tilde{\theta}, v) = \mu\big(\eta k_0 v^2 + \tilde{\theta}\, v\big).$$

Since its partial derivative is positive (i.e., $\frac{\partial g_0(k_0,\tilde{\theta},v)}{\partial(\tilde{\theta}v)} = \mu > 0$*), this update rule results in gain recalibration like Example 1. However, unlike Example 1, the recalibration can be complete or partial, depending on the value of* η*. If* $\eta = 0$*, the update rule takes the same sign as the product* $\tilde{\theta}v$*, thus satisfying the necessary condition for complete recalibration. Otherwise, it only satisfies the necessary condition for partial recalibration by taking the same sign as* $v(\tilde{\theta} - \tilde{\theta}^{\infty})$ *for* $\tilde{\theta}^{\infty} = \eta v$ *(Fig. 3C).*

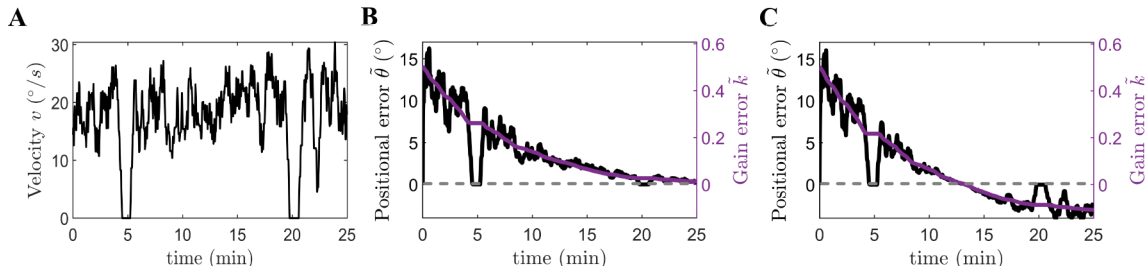

Figure 3: Numerical simulation of the two example gain update rules. For both simulations, we chose the initial condition $k_0(0) = 1$ and the parameters $\beta(\tilde{\theta}) = 0.66 \times \sin(\tilde{\theta})$, $k^{\star} = 1.5$, $\mu = 0.02$. The gain choices imply that the initial value of gain error is $\tilde{k}(0) = 0.5$. Additionally, we chose $\eta = 0.12$ for the second example. (A) Temporal progression of the smoothed animal's velocity from an experiment in [34]. (B) Simulated error trajectories under example gain update rule 1. As soon as the animal begins its movement at $t = 0$, the positional error $\tilde{\theta}$ (black line relative to the left y axis) quickly increases because of the nonzero gain error $\tilde{k}$ (purple line relative to the right y axis). As the animal recalibrates its gain, the gain error gradually converges to zero (i.e., $\tilde{k} \to 0$), accompanied by positional error gradually converging to zero also (i.e., $\tilde{\theta} \to 0$). In addition to these gradual convergent trends, the error trajectories include many fast, transitory changes. As can be seen from the black line, the instantaneous value of positional error $\tilde{\theta}$ is correlated with the animal's velocity v also. For example, when animal slows down, the positional error decreases, becoming zero when the velocity is zero. This is a reflection of the relatively increased landmark stabilization $\beta(\tilde{\theta})$ when path integration inputs kv are decreased. On the other hand, the temporal changes in the gain are correlated with the multiplication of the positional error $\tilde{\theta}$ and the animal's velocity v as determined by the gain update rule g_0. When the animal pauses temporarily around minutes 5 and 20 (i.e., $v = 0$), the positional error $\tilde{\theta}$ is completely corrected by landmarks (i.e., $\tilde{\theta} = 0$), causing the gain updates to pause also (i.e., $\frac{d\tilde{k}}{dt} = 0$). As the animal continues moving, the positional error and the velocity fine-tune the gain until the gain error converges to zero, demonstrating that the system can achieve complete gain recalibration. (C) Simulated error trajectories under example gain update rule 2. The convention is the same as panel B. The error trajectories exhibit similar trends to the panel B except that their final values do not converge to zero, demonstrating that the system can only achieve partial gain recalibration.

3.3 Mechanistic Constraints Reveal Instrumental Role of Positional Error Codes

In the previous section, we presented necessary and sufficient algorithmic conditions for recalibration of the PI gain under a broad class of neural mechanisms. The necessary conditions briefly state that spatial average of the PI gain (k_0) must change in the direction of the product of the animal's velocity (v) and the network's positional representation error ($\tilde{\theta}$). In this section, our objective is to identify mechanistic prerequisites for meeting this algorithmic requirement at the network level.

To achieve this goal, we investigate the analytical expression of the PI gain's spatial average k_0 which can be simply obtained by averaging the PI gain $k(\theta)$, in Equation (6), over θ as follows:

$$k_0 \triangleq \frac{-b}{2\pi\tau_c \big\| \frac{\partial r_c^{\star}(\psi)}{\partial \psi} \big\|^2} \int_0^{2\pi}\int_0^{2\pi} \frac{\partial^2 r_c^{\star}(\psi)}{\partial \psi^2} \sum_{i\in\{\text{cw},\text{ccw}\}} \alpha_i W_{i\text{-c}}(\psi) W_{\text{v-}i}(\psi)\, \text{sign}[r_i^{\star}(\psi,\theta,0)]\, d\psi\, d\theta \tag{19}$$

Specifically, we examine how the updating of k_0 through temporal changes in the terms of this expression satisfy the algorithmic requirement for gain recalibration in Equation (16). These terms include the following: (i) the offset b in the central-to-rotation ring connections, (ii) the synaptic time constant τ_c, (iii) the slope parameters α_{cw},α_{ccw} quantifying the absolute value of the tuning slopes of velocity neurons, (iv) the synaptic weight functions $W_{\text{v-cw}}$,$W_{\text{v-ccw}}$ of the velocity-to-rotation ring connections, (v) the synaptic weight functions $W_{\text{cw-c}}$,$W_{\text{ccw-c}}$ of the rotation-to-central ring connections, (vi) the function $r_{\text{c}}^{\star}$ describing the central ring's persistent activity bump, and (vii) the functions $r_{\text{cw}}^{\star}$, $r_{\text{ccw}}^{\star}$ describing solutions to the rotation ring's persistent activity bump. Note that we treat the CW and CCW counterparts of the same term as an inseparable pair, hence referring to them together as a single term. Out of the seven terms, we consider the last five terms, (iii)-(vii), as candidates that drive gain recalibration within our ring attractor model through temporal changes (e.g., synaptic plasticity), implicitly assuming that the first two terms, the synaptic time constant τ_c and the offset b, are "hardwired" (i.e., time-invariant).

The rationale behind excluding the first two terms arises, in part, from limitations of our modeling framework. First, the rate-based dynamical model, within which we describe the network, does not include any cellular and receptor details to capture possible temporal changes in the synaptic time constant τ_c. Instead, our model includes τ_c as a "lumped parameter" reduction of complex phenomena that governs the changes in membrane potential with ion flux through receptors; future work could use chemical kinetics modeling to investigate how changes in τ_c (e.g., activity dependent control of ion channel gating dynamics) could contribute to gain recalibration, but that is beyond the scope of the present study. Second, our model employs a simplified one-to-one connectivity between the rotation rings and the central ring such that one neuron in a rotation ring connects to only one neuron in the central ring according to a fixed b, rather than a one-to-all connectivity pattern required to capture plasticity in the offset b by continuously shifting the strength of connections across the neural space through gradual modulation of weights. Therefore, excluding τ_c and b from further consideration, we focused our analysis on the remaining five terms as the driver of gain recalibration.

By analyzing the mapping between the temporal change in any one of these candidate terms and the resulting temporal change in k_0, we find that rate-based encoding of the positional error plays an instrumental role in gain recalibration. That is, independent of which term drives the changes in the PI gain, a rate code of the positional error is required to satisfy the necessary algorithmic condition for the gain recalibration. However, the specific nature of the error code depends on which term drives the changes in the PI gain. In the next subsections, we first verbally describe the rationale and properties of the required error codes for each candidate term based on our theoretical analysis and then visualize each error code based on a numerical simulation of gain recalibration (Fig. 4A).

3.3.1 Plasticity in the velocity pathway requires a rate code of the instantaneous error

We first consider the scenario that the temporal change in the PI gain is driven by temporal changes in either set of synaptic weights along the pathway from velocity neurons to the central ring. These synaptic weights include the pair $W_{\text{v-cw}}, W_{\text{v-ccw}}$, describing the strength of velocity-to-rotation ring connections

(④ in Fig. 1A), and the pair $W_{\text{cw-c}}, W_{\text{ccw-c}}$, describing the strength of rotation-to-central ring connections (③ in Fig. 1A).

According to Equation (19), the CW and CCW components of these weight pairs additively influence the PI gain k_0. This additive influence suggests a possibility where CW and CCW components vary independently of one another during recalibration to tune k_0. However, this possibility is limited by the requirement that the central ring must receive balanced (i.e., symmetric) inputs from the CW and CCW rotation rings to keep the activity bump stationary during the animal's immobility as we show in Appendix 6.4. Thus, we assume hereafter that the CW and CCW components vary in a coordinated fashion to ensure that their individual contribution to the PI gain k_0 is symmetric. This symmetry assumption implies that if the overall value of k_0 changes as per the necessary algorithmic condition in (16), then the individual contribution to this change from both CW and CCW components must be in the direction of the product of the animal's velocity v and the positional error $\tilde{\theta}$. To identify the mechanistic underpinnings of such symmetric gain recalibration, we revisit Equation (19). By differentiating this equation with respect to time and considering Hebbian plasticity as the mechanism underlying the changes in the weight pairs $W_{\text{v-cw}}, W_{\text{v-ccw}}$ or $W_{\text{cw-c}}, W_{\text{ccw-c}}$, we find that the algorithmic condition translates into a mechanistic constraint. Details are below:

Hebbian plasticity of the velocity-to-rotation ring connections: A change in the synaptic weights of the velocity-to-rotation ring connections leads to a commensurate change in the speed at which the network's activity bump is shifted along the ring for a given speed of the animal. These commensurate changes suggest a positively correlated relationship between the synaptic weights of the velocity-to-rotation ring connections and the PI gain. Assuming the symmetry between CW and CCW components, we show in Appendix 6.4.1 that Equation (19), relating the weights $W_{\text{v-cw}}$, $W_{\text{v-ccw}}$ to the PI gain k_0 via positively-weighted integrals, can be reformulated as

$$k_0 \propto \frac{1}{2\pi}\int_0^{2\pi} W_{\text{v-cw}}(\psi)d\psi,$$
$$k_0 \propto \frac{1}{2\pi}\int_0^{2\pi} W_{\text{v-ccw}}(\psi)d\psi,$$

where the $\propto$ symbol denotes the existence of a positively-sloped, proportional relationship. Because of these positive correlations, satisfying the algorithmic condition for recalibration implies that the average strength of both CW and CCW velocity-to-rotation ring synapses is modified in the direction of the product of the animal's velocity (v) and the network's positional error ($\tilde{\theta}$), namely,

$$\text{sign}\left[\frac{1}{2\pi}\int_0^{2\pi} \dot{W}_{\text{v-cw}}(\psi)\right]d\psi = \text{sign}\left[\frac{1}{2\pi}\int_0^{2\pi} \dot{W}_{\text{v-ccw}}(\psi)\right] = \text{sign}[\tilde{\theta}\, v]. \tag{20}$$

Here, the dots appearing over the weights denote the temporal change in the weight. Recall that Hebbian plasticity of a synapse is driven by the joint activity of pre- and post-synaptic neurons. In the specific case of velocity-to-rotation ring synapses, the pre-synaptic side composed of the velocity neurons, designated to solely encode the animal's velocity v with tuning curves that have a negative slope for the CW velocity neuron and a positive slope for the CCW velocity neuron (previously shown in Fig. 1D). Therefore, in a manner matching these differential signs of v-encoding on the presynaptic side, the CW and CCW rotation

rings on the post-synaptic side *must monotonically decrease and increase their average firing rate with the instantaneous positional error* $\tilde{\theta}$ to satisfy the equality in Equation (20) (Fig. 4B). Mathematical proof and further details are provided in Appendix 6.4.1.

Hebbian plasticity of the rotation-to-central ring connections: Like the velocity-to-rotation ring connections discussed above, the rotation-to-central ring synaptic weight functions $W_{\text{cw-c}}$,$W_{\text{ccw-c}}$ enter linearly in the calculation of the PI gain in Equation (19). Therefore, similar to the previous case, the algorithmic condition for recalibration via $W_{\text{cw-c}}$,$W_{\text{ccw-c}}$ requires Hebbian plasticity to modify their average strength in the direction of the product of the animal's velocity (v) and the network's positional error ($\tilde{\theta}$), namely,

$$\text{sign}\left[\frac{1}{2\pi}\int_0^{2\pi}\dot{W}_{\text{cw-c}}(\psi)\right]d\psi = \text{sign}\left[\frac{1}{2\pi}\int_0^{2\pi}\dot{W}_{\text{ccw-c}}(\psi)\right] = \text{sign}[\tilde{\theta}\, v]. \tag{21}$$

In the previous case, the mechanistic prerequisite for meeting a similar sign requirement was error-encoding on the *postsynaptic* side since the pre-synaptic neurons were assumed to solely encode the velocity. In the present case, however, it is feasible to encode the error in either the *pre-* or *post-synaptic* side since neither side is subject to such a limitation. Therefore, when the animal is traveling in one direction (as was the case in the experiments that originally demonstrated the gain recalibration [34]), satisfying the equality in Equation (21) requires one of the following two quantities to *vary monotonically with the network's instantaneous positional error* (Fig. 4C): i) The mean firing rate of the rotation rings or ii) The mean firing rate of the central ring. However, unlike the previous case, our analysis of the present case does not provide conclusive information about the direction of these monotonic relations. Mathematical proof and further details are provided in Appendix 6.4.2.

Collectively, these findings show that Hebbian plasticity in the pathway carrying the external velocity information to the central ring requires a rate code of the network's instantaneous positional error to update the synaptic weights in the direction of the product of the animal's velocity and the error as per the algorithmic condition for gain recalibration in Equation (16). Consequently, when the animal is moving in one direction, the synaptic weights track the time-integral of the error (bottom row in Fig. 4A) as a proxy that encodes the PI gain.

3.3.2 Plasticity elsewhere requires a rate code of the time-integral of the error

We next consider the scenario that the synaptic weights along the pathway from velocity neurons to the central ring are hardwired (i.e., constant). This scenario implies that the gain recalibration is instead driven by temporal changes in one of the terms associated with the *firing rates* of neurons in the network. This includes the slope parameters α_{cw}, α_{ccw} quantifying the absolute value of the slopes of the CW, CCW velocity neurons' tuning curves, the ansatz functions describing the persistent activity bumps $r^{\star}_{\text{cw}}$,$r^{\star}_{\text{ccw}}$ of the rotation rings, or the ansatz of the persistent activity bump $r^{\star}_{\text{c}}$ of central ring. Independent of which of these terms undergoes temporal changes, the algorithmic condition for gain recalibration translates into a mechanistic constraint that, while still a rate code of error, differs in its fundamental characteristics as detailed below:

Changes in the slopes of velocity neurons' tuning curves: As shown previously in Fig. 1D, the CW and

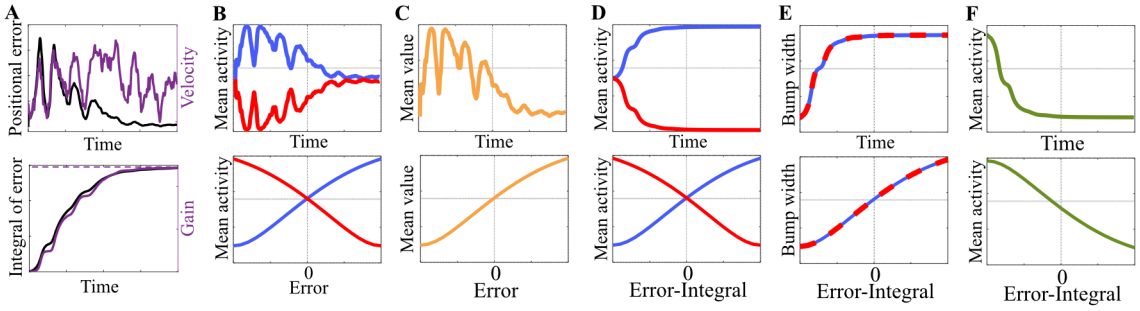

Figure 4: Visualization of mechanistic constraints for a numerical simulation based on a hypothetical gain update rule $g_0(k_0, \tilde{\theta}, v) = \mu\tilde{\theta}v$ Except the animal's velocity profile, we chose the parameters and initial conditions the same as Fig. 3B. For color coding, we use Fig. 1A as the reference, where red and blue denote the CW and CCW rotation rings, and green denotes the central ring. (A) Top graph shows the simulated velocity of the animal (purple line) on the right y axis and the temporal progression of the positional error (black line) on the left y axis. Notice the synchronous fluctuations in the positional error and the animal's velocity. As explained in Fig. 3B, these synchronous fluctuations occur because the positional error is correlated with the animal's velocity. Bottom graph shows the PI and visual gains with solid and dashed purple lines, respectively, on the right y axis and the time-integral of the positional error with the black line on the left y axis. Notice that as the PI gain gradually converges to the visual gain, the temporal progression of the time-integral of the positional error follows a very similar trajectory. This similarity indicates that the integration gain reflects the past accumulation of positional representation errors, thus opening up the possibility for the network to track the time-integral of the error as a proxy signal to encode the integration gain. (B) The mechanistic constraint for recalibration through plasticity of the velocity-to-rotation ring connections. Top graph shows the mean firing rates of the CCW and CW rotation rings over time with blue and red lines. Notice that they are similar to the trajectory of the positional error in panel A, except that the changes in the CW rotation ring's mean firing rate (red line) is the negative of those in the CCW rotation ring's mean firing rate. Bottom graph shows the direct relationship between mean firing rates and positional error in the attractor's representation. (C) The mechanistic constraint for recalibration through plasticity of the rotation-to-central ring connections. Top graph shows the mean firing rates of either the rotation rings or the central ring over time with the orange line. Notice that the changes in these firing rates follow a similar trend as the temporal progression of the positional error. Bottom graph shows this relationship directly (the positive correlation is chosen arbitrarily as our analysis does not provide a conclusive insight into the required direction). (D) The mechanistic constraint for recalibration through changes in the velocity neurons' slopes. Top graph shows the mean firing rate of the CCW and CW rotation rings, the same quantities as panel B. However, unlike panel B where the mean firing rates were similar to the instantaneous positional error, the mean firing rates in this panel are similar to the time-integral of the error. Bottom graph shows this relationship between the mean firing rates and the time-integral of the error directly. (E) The mechanistic constraint for recalibration through changes in the rotation rings' activity bumps. Top graph shows the bump width of both rotation rings over time. Similar to how mean firing rates of the rotation rings encode the time-integral of the positional error in panel D, the bump widths encode the time-integral of the error in this panel. Bottom graph shows this relationship directly. (F) The mechanistic constraint for recalibration through changes in the central ring's activity bump. Top graph shows the temporal progression of the mean firing rate of the central ring, which is tightly but negatively correlated with the temporal progression of the time-integral of the positional error. Bottom graph shows this relationship directly.

CCW velocity neurons are tuned to the animal's velocity with slopes in different directions. Throughout the paper, we assume these differential directions of the tuning slopes to be hard-wired (i.e., constant). Under this assumption, we limit ourselves in the present case to temporal changes in the absolute values of the slopes, denoted by α_{cw} and α_{ccw},. A change in α_{cw}, α_{ccw} parameters results in a commensurate change occurs in the movement speed of the persistent activity bump along the central ring for a given speed of the animal. This implies that α_{cw} and α_{ccw} are positively correlated with the PI gain. Indeed, a positively-correlated, linear relationship can be explicitly seen in Equation (19). Thus, under the symmetry assumption described in the previous section, the algorithmic condition for recalibration (Equation (16)) implies that α_{cw} and α_{ccw} must *change* in the direction of the product of the animal's velocity (v) and the network's positional error ($\tilde{\theta}$), namely,

$$\mathrm{sign}[\dot{\alpha}_{\mathrm{cw}}] = \mathrm{sign}[\dot{\alpha}_{\mathrm{ccw}}] = \mathrm{sign}[\tilde{\theta}\, v] \tag{22}$$

An implication of this requirement is that, when the animal is traveling in one direction, the *change in the slopes* is monotonically related to the positional error, reflecting its value on a moment-to-moment basis. Its sign, however, reflects not only the sign of the error but also the sign of the velocity as in Equation (22). When these changes are integrated over time, *the current value of the slopes* reflects the accumulation of positional errors through the past, depending monotonically on the time integral of the positional error. This dependence becomes a monotonically increasing relationship when the animal travels in a positive direction and a monotonically decreasing relationship when it travels in a negative direction. Through connections between the velocity neurons and the rotation rings, these monotonic relationships are also reflected in the mean firing rate of rotation rings. Since CW and CCW rotation rings selectively receive input from the respective velocity neurons, this reflection additionally carries over the directional differences in the slopes to the rotation rings. As a result, when the animal is traveling in one direction (say positive), the mean firing rate of the CCW rotation ring increases monotonically with the time-integral of the error, based on inputs from the CCW velocity neurons. Conversely, the mean firing rate of the CW rotation ring decreases monotonically with the time-integral of the error due to negative slope of the CW velocity neuron (Fig. 4D). If the animal travels in the other direction, the direction of these monotonic relationships is also reversed. Mathematical proof and further details are provided in Appendix 6.4.3.

Changes in the persistent activity bump of the rotation rings: Transmission of velocity information to the central ring occurs through a series of steps. First, the velocity neurons differentially modulate the rotation rings' activity bumps $r^{\star}_{\mathrm{cw}}$, $r^{\star}_{\mathrm{ccw}}$ during the animal's movement. Active neurons in the rotation rings then pass this differential velocity information down to the central ring, which in turn shifts the activity bump. In this transmission chain, an increase in the number of active neurons in the rotation rings (i.e., larger width of the bumps $r^{\star}_{\mathrm{cw}}$, $r^{\star}_{\mathrm{ccw}}$) would result in a commensurate increase in the number of central ring neurons that receive the differential velocity inputs. As a result, the bump of activity shifts faster along the central ring even when the animal's velocity is unchanged. This implies a positively correlated relationship between the gain and the widths of the rotation rings' activity bumps, reminiscent of the relationship between the gain and the slope parameters $\alpha_{\mathrm{cw}}, \alpha_{\mathrm{ccw}}$ (the previous case). Thus, similar

to the previous case, satisfying the algorithmic condition requires *the rotation rings' activity widths to change* in the direction of the product of the animal's velocity and the positional error. Consequently, when the animal is traveling in one direction (say positive), the widths of the rotation rings' activity bumps reflect the accumulation of positional errors through the past, *increasing monotonically with the time-integral of the error* (Fig. 4E). If the animal's travel direction is negative, this relationship turns into a monotonically decreasing function. Mathematical proof and further details are provided in Appendix 6.4.4.

Changes in the persistent activity bump of the central ring: Consider as an example that there are two networks with the same bump profile (e.g., Gaussian) but one has a higher peak firing rate. In this case, if all else is equal, the network with the higher firing requires higher velocity inputs to be shifted from point A to point B in the same time as the other network, thus indicating a lower PI gain. This example implies an inversely correlated relationship between the gain and the magnitude of the central ring's activity bump, which can also be verified from the calculation of PI gain in Equation (19) as the denominator includes a squared norm of the gradient of the central ring's activity bump. Aside from the inverse effect, this correlated relationship is similar to the previous case. Thus, satisfying the algorithmic condition for recalibration is subject to a mechanistic constraint that is similar to that case in spirit but slightly different due to the inverse effect: When the animal is traveling in one direction (say positive), the mean firing rate of the central ring's activity bump $r_c^\star$ must *decrease monotonically with the time-integral of the error* (Fig. 4F). If the animal's travel direction is negative, this relationship turns into a monotonically increasing function. Note that, when deriving this result, we assumed that the general shape of the central ring's recurrent connectivity pattern (i.e., local excitation and global inhibition) remains invariant. Mathematical proof and further details are provided in Appendix 6.4.5.

Collectively, these findings show that gain recalibration might be possible without plasticity in the velocity pathway carrying the external velocity information to the central ring, provided that there is a rate code of the time-integral of the positional error as opposed to the error itself. Our analysis does not provide any insights into the mechanisms necessary for such rate code and the temporal changes in the terms associated with it (e.g., slopes of velocity neurons). However, it may be the case that plasticity elsewhere than the velocity pathway is required. Independent of the underlying mechanism, however, with the absence of plasticity in the velocity pathway, the PI gain is no longer encoded in the synaptic weights, instead, it is encoded in the firing rates that track the time-integral of the error, a proxy of the PI gain (bottom row in Fig. 4A).

4 Implementing Gain Recalibration in a Ring Attractor

In this section, we propose a modified ring attractor model that can achieve gain recalibration through a mechanism devised based on the insights we have garnered so far. Briefly, the model relies on synaptic plasticity in the velocity-to-rotation ring connections and its mechanistic prerequisite(a rate code for the positional error instantiated in the rotation rings as shown in Fig. 4B) to achieve gain recalibration. We chose this mechanism, instead of other theoretical candidate mechanisms examined in the previous section, partly because we found it relatively easy to implement compared to others. We conjecture

that some of the other candidate mechanisms also have biologically feasible implementations. Thus, our model based on synaptic plasticity in the velocity-to-rotation ring connections should be viewed as an example that demonstrates the effectiveness of our theoretical analysis rather than the only network model that can achieve gain recalibration. In addition to gain recalibration, the model can reproduce two important aspects of biological path integration: (1) flexible association of visual landmarks to different positions in the neural space and (2) correction of accumulated PI errors by visual landmarks. In the next subsections, we describe the model in detail and explain the mechanisms by which it achieves these aspects.

4.1 A connectivity pattern yielding a rate code for error

Before delving into details of the model, we first propose a connectivity pattern that causes a neuron population to vary its firing rate as a monotonic function of the difference in the peak locations of the activity bumps within two other populations. As will be evident later, this connectivity pattern will play a crucial role in our model because the monotonic relationship allows us to achieve a rate code for the positional error, with the error being the difference in the bump locations of the visual drive and the attractor activity.

The connectivity pattern can be described by considering three distinct populations of neurons (X1-3 in Fig. 5A), each arranged on a circle similar to the populations in the ring attractor model. Suppose that populations X1 and X2 consist of excitatory and inhibitory neurons, respectively, and maintain their own activity bumps. The third population (X3), however, derives its activity based on the synaptic inputs from X1 and X2 to encode the difference in their bump locations. The excitatory inputs from X1 to X3 are routed through topographic connections that wire together the neurons at the same angular location in the neural space. The inhibitory inputs from X2 to X3, on the other hand, are routed through *offset* connections that wire together the neurons at different angular locations in the neural space. To understand how X3 population varies its firing rate monotonically with the difference in peak locations of X1 and X2's activity bumps, we can track the flow of neural activity across the proposed connectivity pattern (Fig. 5B).

Let x_1 and x_2 denote the angular locations of the activity bumps within the populations X1 and X2, respectively. When $x_2 - x_1 = 0$, the activity bumps of X1 and X2 are aligned in the circular neural space, but their synaptic inputs to X3 are misaligned because of the CCW offset in X2-to-X3 connections (the second column of Fig. 5B). When $x_2 - x_1 \neq 0$, however, the activity bumps of the X1 and X2 populations are initially misaligned (the first and third columns of Fig. 5B, corresponding to CW (i.e., $x_2 - x_1 < 0$) and CCW (i.e., $x_2 - x_1 > 0$) differences, respectively). Consider first the CW difference case (the first column of Fig. 5B): the bump of the X1 is assumed to be at the same location as the zero-difference condition (without loss of generality), thus the bump of the X2 is misaligned, being shifted in the CW direction. Because of the CCW offset in the X2-to-X3 connections, this misalignment decreases at the level of synaptic inputs, causing X3 to receive increased inhibition on its active neurons. This increased inhibition, in turn, decreases the firing rate of the X3 population. Consider next the CCW difference (i.e., $x_2 - x_1 > 0$ and the third column of Fig. 5C). In this case, the CCW offset in the X2-to-X3 connections

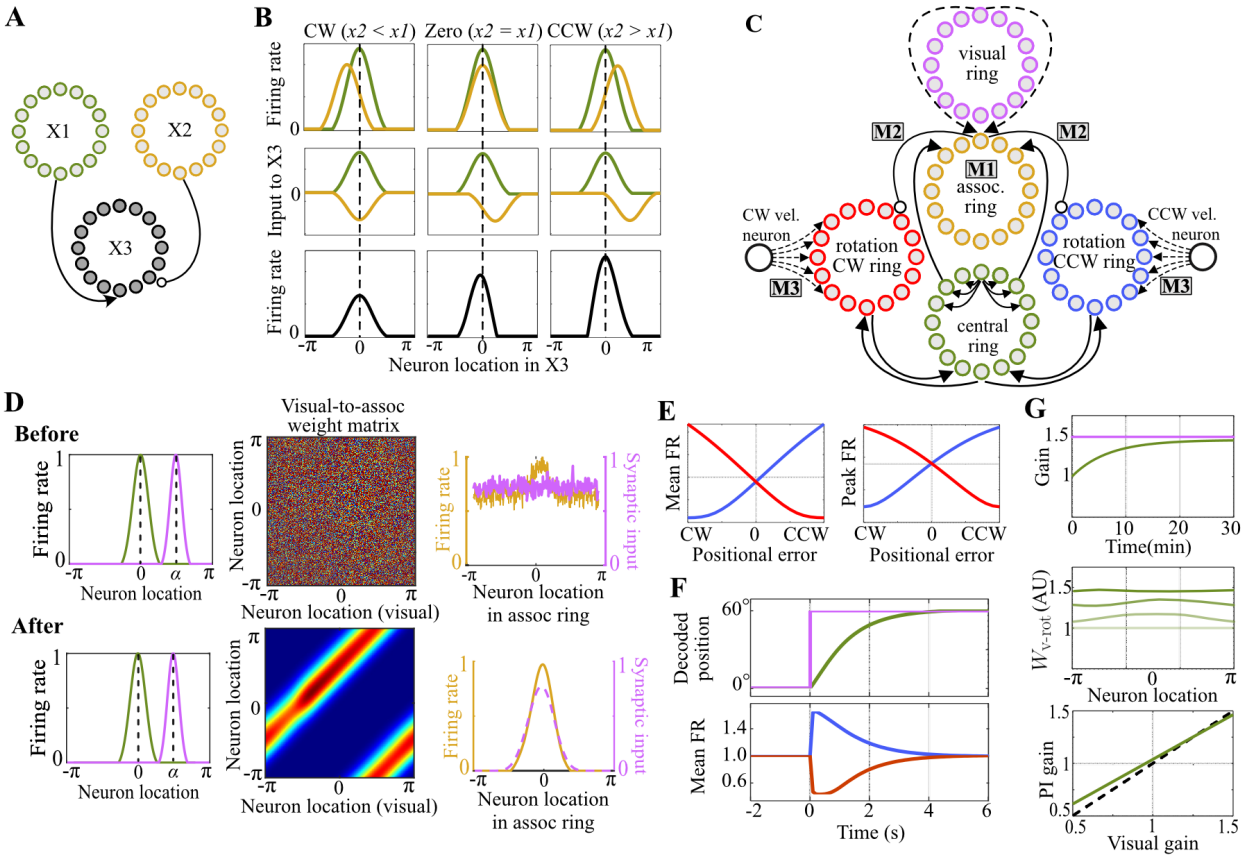

Figure 5: A modified ring attractor network model. (A) A proposed connectivity pattern that varies firing rate of population X3 as a monotonic function of the difference in the activity-bump locations of populations X1 and X2. Arrow and circle terminals denote excitatory and inhibitory connections, respectively. (B) Schematic diagrams depicting the computation within the proposed connectivity in panel (A). The first row shows the activity of X2 population (yellow) relative to the activity of the X1 population (green) for different conditions: CW difference (left column), no difference (middle column), and CCW difference (right column). The second row shows the synaptic inputs to X3 population from the excitatory X1 (green line) and inhibitory X2 (yellow line) populations. The fourth row shows the resulting firing rates of the X3 population. (C) Schematic representation of the model. Solid and dashed lines denote hardwired and plastic connections, respectively. The labels M1, M2, M3, correspond to the three modifications made to the classical ring attractor: M1 is the association ring, M2 refers to the hardwired, offset association-to-rotation connections, and M3 refers to the plasticity of the velocity-to-rotation-ring connections. (D) Numerical simulation demonstrating the association of the visual ring's activity with the central ring's activity through Hebbian plasticity. Top and bottom rows show the initial and final values of the simulated variables. The left column shows the firing rates of the central (green) and visual (pink) rings. The middle column visualizes the weight matrix describing the visual-to-association ring connections. The right column shows the firing rates of the association ring (yellow) and the synaptic inputs from the visual ring (pink). (E) Tuning curves depicting the relationships between the rotation rings' mean and peak firing rates vs. the error (left and right graph, respectively). The color coding is the same as panel C. (F) Numerical simulation demonstrating how visual landmarks correct positional error. The top panel shows the progression of the bump locations of the visual (pink) and central rings (green). Bottom panel shows the mean firing rate of CW (red) and CCW (blue) rotation rings over time. (G) Numerical simulations of the gain recalibration within the proposed model. The top shows the the recalibration of PI gain (green) toward the visual gain (pink) in a selected simulation. The middle shows the progression of the weights of the velocity-to-rotation ring connections (four samples normalized to initial condition of the weights with the opacity changes from the lightest ($t = 0$) and to the darkest green ($t = 30$) min corresponding to chronological order of the samples.). The bottom shows the final values of the PI gain for various visual gains (green line) and the hypothetical perfect recalibration (dashed black line).

redirects the inhbitory synaptic inputs from X2 further away from the active neurons of X3, thus causing X3 population to increase its overall firing rate. Through this mechanism, the proposed connectivity patterns vary the firing rate of X3 population as a monotonically increasing function of the difference in the bump locations of X1 and X2 populations. The direction of this monotonic relationship can be easily reversed if the offset in the X2-to-X3 connections is reversed from CCW to CW.

How can we make use of this connectivity pattern in the modified ring attractor model that we will present in the following subsections? Because the modified model will rely on plastic velocity-to-rotation ring connections for gain recalibration, it requires that the rotation rings include a rate code for the positional error, i.e., the difference in the attractor's representation relative to the visual drive. The proposed connectivity is well-suited for this purpose as it can vary the firing rates of the rotation rings (X3) as a monotonic function of the difference in the bump locations of the central ring (X1) and the visual drive (X2). Despite this suitability, however, the proposed connectivity is not directly applicable to the traditional ring attractor network as it requires an additional modification for a reason that will be clear in the next section.

4.2 Flexible Association of Landmarks to Positions through Hebbian Plasticity

In traditional models, visual inputs from external landmarks are incorporated into the attractor network via direct synaptic connections from the visual ring [20, 46, 52]. These connections undergo a Hebbian plasticity process that potentiates the weights of synapses between coactive neuron pairs encoding the same angular position of the animal, thereby resulting in a flexible associative mapping between the visual ring and the attractor network [20, 53].

However, this approach is not compatible with the connectivity pattern proposed in the previous section for the error-rate code. The incompatability arises because Hebbian plasticity wires together the neurons representing the same angular position, whereas the error code's proposed connectivity pattern necessitates the wiring of neurons representing different angular positions by a specific offset. This required wiring of non-coactive neurons cannot be achieved through Hebbian plasticity; instead, it requires hard-wired, topographic connections with an embedded offset. To resolve these issues, we propose a small modification to the network model. This extension deals with the plastic associative mapping problem outside the ring attractor while providing an opportunity to include the error code's connectivity pattern inside the ring attractor network.

The modification is as follows: We first remove the visual-to-central ring connections and introduce an intermediate ring of neurons, which we call an *association ring*. This intermediate ring associates the activity in the visual ring with that in the central ring by receiving input from both (see M1 in Fig. 5C). The connections from the central ring to the association ring are hardwired and weak, while those from the visual ring are plastic, hence capable of becoming strong. In a novel environment, where plastic connections are initially untuned and random, synaptic inputs from the visual ring to the association ring do not convey spatial selectivity to the activity in the visual ring. On the other hand, inputs from the central ring to the association ring, routed through hard-wired, topographic connections, preserve the spatial selectivity of the central ring's activity. This combination of malleable inputs from the visual ring

and weak but persistent (hardwired) inputs from the central ring biases the connections from co-active visual neurons to strengthen near the peak of the central ring's activity bump through Hebbian plasticity (top row in Fig. 5D). Eventually, synaptic inputs from the visual ring become sufficiently strong and aligned with the activity in the central ring, making the association ring's activity strongly visually driven such that it implicitly represents a flexible associative mapping between the representations in the visual ring and the attractor network (bottom row Fig. 5D). Note that the association ring in our model resembles so-called R neurons in the fly brain [54–56], which associate attractor-like compass neurons with sensory neurons that are tuned to visual features in an environment.

4.3 Gain Recalibration by Landmarks through Hebbian Plasticity

In the previous section, we introduced the association ring, an additional population of neurons, that flexibly forms a mapping between the representations in the visual ring and the attractor network. With a properly mapped, strong visual drive available in the association ring, we are now ready to implement gain recalibration in our model via plasticity in the velocity-to-rotation ring connections.

As explained previously in Section 3.3, this recalibration mechanism requires the CCW and CW rotation rings to, respectively, increase and decrease their firing rates with the positional error in the representation of the central ring relative to the visual drive in the association ring. To obtain these error codes, we employ the connectivity pattern proposed in Section (4.1) as follows: We add CCW offset connections from the association ring to the CCW rotation ring and CW offset connections from the association ring to the CW rotation ring (M2 in Fig. 5C). These novel connections result in a monotonically increasing relationship between the CCW rotation ring's firing rate and the positional error and a monotonically decreasing relationship in the CW rotation ring (Fig. 5E).

Now that our network model includes the necessary error codes for gain recalibration through plasticity in the velocity-to-rotation ring connections, the natural next step toward gain recalibration is to incorporate plasticity in these connections (M3 in Fig. 5C) . However, it is not entirely clear whether this combination of the plasticity and the rate code of error would suffice to achieve gain recalibration as our findings only point to a necessity.

To clarify this, we first take an analytical approach based on Proposition 2, which states a sufficient condition that guarantees gain recalibration. Specifically, if the temporal change in the gain's spatial average k_0 has a positive slope with respect to the product of the animal's velocity v and the attractor's positional error $\tilde{\theta}$, recalibration is guaranteed. As discussed in Section 3.3.1, the gain's spatial average k_0 is positively correlated with the average strength of velocity-to-rotation ring connections. Therefore, the sufficient condition can be rephrased as follows: the gain recalibration is guaranteed if the change in the average strength of velocity-to-rotation ring connections has a positive slope with respect to the product of v and $\tilde{\theta}$. The presynaptic side of these plastic connections has velocity neurons encoding the animal's velocity v, while the postsynaptic side has the rotation ring neurons encoding the positional error $\tilde{\theta}$. Since both of these rate codes have the same slopes on the CW and CCW parts of the network (negative slope on the CW part, positive slope on the CCW part), their correlated activity, as the driver of Hebbian plasticity, modifies the strength of velocity-to-rotation ring connections with a positive slope

relative to the product of v and $\tilde{\theta}$, as required by the sufficient condition. Full mathematical proof of this finding can be found in Appendix 6.5. However, further analysis revealed that the resulting gain recalibration is imperfect, caused by the imperfect encoding of the positional error $\tilde{\theta}$ in the rotation rings whose firing rates are additionally modulated by the animal's velocity v. As a result, through plasticity in the velocity-to-rotation ring connections, our model is expected to tune its average PI gain k_0 to steady-state values that are close to, but not the same as, the visual gain, $k^\star$. It is further expected that the model goes through phases of spatially non-homogenous PI gain $k(\theta)$ as the plasticity at any moment modifies spatially distributed weights of the velocity-to-rotation ring connections non-homogeneously (because of non-homogenous firing rate of neurons in the rotation rings); however, as the animal runs laps, a spatially uniform error signal will eventually wash out spatial variation of the gain $k(\theta)$, leading to a nearly homogenous state.

We finally verify these analytical predictions by performing numerical simulations of the proposed ring attractor model for a simulated rat running laps on a circular track while visual landmarks were moved as per the visual gain $k^\star$. The model demonstrated imperfect yet stable gain recalibration for a range of visual gains (Fig. 5F).

4.4 Correction of Positional Errors by Landmarks through Rate Code of Error

In the previous sections, we addressed two of the three important aspects of biological path integration within our novel network model : i) flexible association of visual landmarks to different positions in the neural space ii) gain recalibration. Thus, it remains to show the model can also correct PI errors by visual landmarks.

The PI error manifests itself as a misalignment between the path-integration driven activity bump in the attractor ring and the visual drive. Traditional models correct for this misalignment by providing visual information *directly* onto the central ring (like in Fig. 1A), which in turn moves the central ring's activity bump toward the visual bump through attractor dynamics. In our model, however, we do not have such direct connections (as we previously removed them in Section 4.2). Instead, our model includes connections onto the rotation rings from the visual drive in the association ring with some offset as described in the previous section. The question then arises: are these offset connections sufficient for PI error correction or do we need to reinstate direct connections onto the central ring from the visual drive?

As explained in the previous section, the offset connections result in a monotonically increasing and decerasing rate codes of the positional error in the CCW and CW rotation rings, respectively. This differential modulation of the rotation rings' firing rates by the error is similar to their differential modulation by the animal's velocity; when the animal is moving, the firing rate of one rotation ring increases and that of the the other decreases, which in turn shifts the activity bump along the central ring. Therefore, by employing rate codes of the positional error, our model effectively transforms the error into a virtual velocity signal, shifting the attractor's activity bump in a manner so as to decrease this error—a manifestation of error correction provided by the visual landmarks. We verified this error correction mechanism in a numerical simulation of our ring attractor model. In the simulation, we

abruptly introduced a positional error between the activity bump of the ring attractor and that of the association ring, with a step change in the animal's position relative to visual landmarks. The differential changes in the firing rates of the rotation rings eliminated the error (Fig. 5E).

5 Discussion

Constant fine-tuning of a neural integration computation is crucial to maintain accurate representations of continuous variables since the relationship between the sensing of the relative change in a continuous variable and its actual value can fluctuate on a variety of timescales: throughout development, across different behavioral circumstances (e.g., swimming versus walking in the case of location coding), and due to dynamic biological processes such as alterations in the intrinsic electrical properties of neurons, hormonal and metabolic changes, variations in synaptic transmission (for instance due to circadian rhythm), etc. Physiological evidence for such fine-tuning was observed in hippocampal place cells, where persistent conflict between self-motion and external visual cues recalibrated the integrator gain [34]. In the present paper, we give the first theoretical examination of this phenomenon in continuous bump attractor networks, a prevailing network model for internal representations of continuous variables.

Our examination unveiled the algorithmic and mechanistic requirements for gain recalibration based on a ring attractor network, a representative CBAN model used especially for circular continuous variables. In CBAN models, when the integration gain is inaccurate, an internal representation of a continuous variable slightly drifts relative to its actual value, resulting in encoding errors. When absolute 'ground-truth' information, such as feedback from visual landmarks for location coding, is available, these errors are corrected through the internal dynamics of the network, without the need for an explicit representation of the error in the circuit. In contrast to this automatic error correction through network dynamics, we found that fine-tuning the integration gain based on errors requires an *explicit error signal*, i.e., firing rate of some neurons to encode the error in the CBAN's representation of a continuous variable relative to its actual value. Building upon this insight, we then proposed a ring attractor network model that shows how a network can recalibrate its integration gain through biologically known plasticity mechanisms. Although the ring attractor is a CBAN model for integration of a 1D circular continuous variable (e.g., location of an animal on a circular track), our analysis and modeling results can be readily extended to higher dimensions and other types of continuous variables. Overall, our findings suggest that a rate code for the error in the internal representation of a continuous variable is a core component of the bump attractor-type neural integrators, and that such a rate code plays an essential role in their gain recalibration.

5.1 Bump Attractor Network as an Adaptive Kalman Filter

To identify algorithmic requirements for recalibration of the integration gain, we simplified dynamics of the ring attractor through a dimensionality reduction protocol described in [41]. Similar approaches have been successfully applied in recent years to explore the neural dynamics capturing how the high-dimensional neural data evolves within low-dimensional topological structure [6, 14, 57]. In our specific case, the dimensionality reduction led to a simplified, one-dimensional model of the ring attractor, pro-

viding an accurate approximation to the temporal changes in the attractor's representation as a function of external inputs that encode the relative change information (e.g., animal's velocity for location coding) and the absolute ground-truth information (i.e., positional feedback from visual landmarks) [41, 42, 58].

Previous research showed that, when two external cues are presented as inputs, the ring attractor network fuses them optimally in the Bayesian sense [20, 59–61]. Furthermore, if one of the cues provides only relative information like the animal's velocity and needs to be integrated over time within the ring attractor, the network performs the Bayesian fusion recursively for each step of the integration [41]. This recursive computation is known as Kalman filtering and proposed as a model of cue integration in the entorhinal cortex of the mammalian brain [21]. Consistent with this prior work, our simplified model of the ring attractor revealed that the network operates as a Kalman filter updating the representation by a combination of integrated relative information (the internal model component of the Kalman filter) and the instantaneous feedback from absolute information (the measurement model component).

In engineered systems, the accuracy of a Kalman filter relies on precise knowledge of its internal model parameters; to address this issue, adaptive Kalman filters that automatically fine-tune their parameters have been proposed [62, 63]. Much like adaptive Kalman filters, we showed that a ring attractor network can fine-tune itself and maintain an accurate representation through gain recalibration. We also elucidated the necessary conditions for the gain recalibration, which showed that the integration gain must change in the same direction as the product of the animal's velocity and the error in the attractor's representation relative to the absolute 'ground-truth' information. Interestingly, this requirement resembles characteristics of a classical adaptation algorithm in Kalman filtering and feedback control systems known as the MIT rule [64]. Thus, a ring attractor with gain recalibration effectively operates as an adaptive Kalman filter, ensuring accurate updates to its representation through a finely tuned integration gain.

5.2 Necessity of a Rate-Based Explicit Error Signal in the Bump Attractor Networks

Satisfying the algorithmic requirements for recalibration of the integration gain is subject to certain mechanistic constraints. To find these constraints, we examined the algorithmic requirement at the level of full network dynamics. This examination led to a major finding of the present study: for gain recalibration, error in the internal representation of the continuous variable relative to its true value must be explicitly encoded via a rate code. Specifically, there must exist neurons whose firing rates vary monotonically with either the instantaneous error or the time-integral of the error. Without such explicit representations of error, the network does not have a teaching signal that can guide tuning of its gain for recalibration. This shows that, for CBAN networks, learning from encoding errors for recalibration of the integration gain is a very different neural process than correction of the encoding errors. In the case of error correction, input signals from absolute 'ground-truth' information sources, such as visual landmarks for a CBAN encoding location, are sufficient to trigger an error correction response from network dynamics. In contrast, learning from encoding errors for recalibration of the integration gain requires an additional neural signal that specifically encodes the error in the representation of the network

via a rate code.

This hypothesized error coding resembles reward and sensory prediction error signals within the mammalian brain. Dopamine neurons in the midbrain of mammals encode error in the internal predictions of reward via monotonic changes in their firing rates [65,66]; they exhibit elevated activity with more reward than predicted, remain at baseline activity for fully predicted rewards, and exhibit depressed activity with less reward than predicted. Climbing fiber inputs to Purkinje cells of the mammalian cerebellum encode errors in the predicted sensory consequences of motor commands relative to the actual sensory feedback via changes in the rate and duration of complex spikes [67, 68]. Both of these prediction error codes are thought to act as a teaching signal that fine-tunes the internal models, mapping the stimulus to reward prediction in the dopamine system and the motor commands to sensory prediction in the cerebellum, through plasticity, just like how error coding can act as a teaching signal that recalibrates the integration gain of a CBAN.

To verify this idea computationally, we proposed a modified ring attractor model that encodes the animal's position on a circular track and including a rate code of the instantaneous positional error within the rotation rings. Turning on Hebbian plasticity between the velocity neurons, providing the relative movement information, and the rotation rings was then enough to achieve gain recalibration.

There are two interesting connections between our model and the existing work in the literature. 1) A previous study used 180° offset connections in a dual ring attractor model to encode disparity between two cues via a rate code [69], similar to our model using offset connections, albeit much smaller than 180°, to encode error in the network's representation relative to 'ground-truth' signals. Despite this apparent similarity, the use of a smaller offset in our model led to a notably different rate code: In the model of [69], the neuronal firing rates vary symmetrically with the cue disparity, thus encoding only the magnitude of the disparity. In our model, however, neuronal firing rates vary asymmetrically (i.e., monotonically) with the error, thus encoding not only the magnitude of the error but also its sign. 2) Another previous study hypothesized a currently unknown plasticity rule as a mechanism for gain recalibration within the ring attractor network [32]; the hypothesized plasticity rule modifies synaptic weights of each neuron according to an implicit positional error signal, computed locally within each neuron through comparison of the synaptic inputs at basal and apical dendrites receiving, respectively, the absolute 'ground-truth' information and the network's current representation information. While it is unclear whether such a plasticity rule exists in the brain, our model demonstrates a biologically plausible alternative: Hebbian plasticity, combined with an explicit rate-based representation of the error, is sufficient to achieve gain recalibration. It remains as a future work to experimentally test if such an error signal exists in the brain circuits that are thought to employ CBANs for encoding continuous variables.

5.3 Implications of a Distributed, Inhomogenous Integration Gain

Prior CBAN models implicitly assumed the integration gain to be a single, global parameter of the network, independent of the value of the encoded continuous variable [11, 31]. Although the idea of different, hard-wired integration gains has previously been suggested in the context of location coding (as an explanation for the changes of spatial scale seen in the location code of place cells along the

dorsal-ventral axis of the hippocampus [27]), it is assumed that these integration gains are constant at all locations within the environment. In contrast, we showed that integration gain of a CBAN, specifically the ring-attractor network, is a distributed parameter instantiated in the network's array of synaptic weights, implying that the network can adopt unique integration gains for different values of the encoded continuous variable. This would be, for example, a CBAN changing its integration gain depending on the location of the animal in the context of spatial navigation or depending on the amount of accumulated evidence in the context of decision-making. In the latter case, the CBAN may look like it unevenly weights early or late evidence, which is a well-established phenomenon known as primacy and recency effects in the decision-making literature [70–72]. Compared to a network with a single, global integration gain, a network with a distributed, possibly inhomogeneous, gain can adjust its representation metric for the continuous variable locally, hence providing greater flexibility in representing different values of the continuous variable with uneven resolutions, depending on, for instance, their behavioral significance [73]. How might gain inhomogenoity arise? Theoretically, it can be a product of the recalibration process if the teaching signal (e.g., feedback from absolute 'ground-truth' information sources) is available unevenly across the values of the continuous variable. In the context of location coding, for example, such differences may occur between when the animal is nearby the boundaries of the environment, which is a relatively rich area in terms of external ground-truth information, vs. when it is near the center of the arena. We speculate that such spatially distributed recalibration of the integration gain may offer a mechanistic explanation to some of the experimental findings about local distortions and deformations in the activity patterns of entorhinal grid cells and hippocampal place cells, encoding the animal's location, during environmental manipulations [74–81]. According to this speculation, grid patterns might get distorted nearby environmental boundaries through changes in the local integration gain of the network; through the same mechanism, place cells might represent locations nearby landmarks and boundaries with a greater spatial resolution (also known as overrepresentation). Overall, inhomogenous integration gain of CBANs offer a potential explanation to an array of seemingly complex responses in spatial navigation as well as other brain functions.

6 Appendix

6.1 Model Setup: Ring attractor network

Dynamics of the ring attractor network can be formulated by modeling the firing rate of each neuron in response to its synaptic inputs.

6.1.1 Dynamics of the central ring

We begin with the central ring. Parameterizing neurons based on their locations $\psi \in S^1$ in the central ring, we can model their dynamics via

$$\tau_{\mathrm{c}} \frac{\partial r_{\mathrm{c}}(t,\psi)}{\partial t} = -r_{\mathrm{c}}(t,\psi) + \sigma\left[W_{\text{c-c}}(\psi) \circledast r_{\mathrm{c}}(t,\psi) + I_{\mathrm{ext}}(t,\psi)\right], \tag{23}$$

where $r_{\mathrm{c}}(t,\psi)$ denotes the firing rate of the central ring neuron ψ at time t, τ_{c} denotes the synaptic time constant, $\circledast$ denotes the circular convolution operation, σ is an activation function (chosen in the present work as a rectified linear unit, RELU), and $W_{\text{c-c}} : S^1 \to \mathbb{R}$ is a rotationally invariant synaptic weight function (chosen in the present work as a symmetric local excitation global inhibition, LEGI), and $I_{\mathrm{ext}}(t,\psi)$ denotes the external synaptic input to the central ring neuron ψ at time t from CW, CCW rotation rings and the visual ring.

We can write this combined external input to the central ring as

$$I_{\mathrm{ext}}(t,\psi) = I_{\text{cw-c}}(t,\psi) + I_{\text{ccw-c}}(t,\psi) + I_{\text{vis-c}}(t,\psi), \tag{24}$$

where $I_{\text{cw-c}}$, $I_{\text{ccw-c}}$, and $I_{\text{vis-c}}$ denote the inputs to the central ring from CW, CCW, and visual rings.[1] Using synaptic weights and firing rates, we can compute each of these inputs according to the equation

$$\begin{aligned} I_{\text{vis-c}}(t,\psi) &= \int_0^{2\pi} W_{\text{vis-c}}(\psi,\phi) r_{\mathrm{vis}}(t,\phi) d\phi, \\ I_{\text{cw-c}}(t,\psi) &= W_{\text{cw-c}}(\psi+b) r_{\mathrm{cw}}(t,\psi+b), \\ I_{\text{ccw-c}}(t,\psi) &= W_{\text{ccw-c}}(\psi-b) r_{\mathrm{ccw}}(t,\psi-b), \end{aligned} \tag{25}$$

where $W_{\text{vis-c}} : S^2 \to \mathbb{R}$ denotes the synaptic weight function describing the all-to-all connections between the visual ring neurons and the central ring neurons, r_{vis} denotes the firing rate of visual ring neurons, $W_{\text{cw-c}}, W_{\text{ccw-c}} : S^1 \to \mathbb{R}_{\geq 0}$ denotes the synaptic weight functions describing the one-to-one and differentially offset connections between the CW,CCW rotation ring neurons and the central ring neurons, and $b \in \mathbb{R}$ denotes the amount of this connection offset. Collectively, Equations (23)–(25) constitute the central ring model considered in this paper.

6.1.2 Dynamics of the rotation rings

We next model firing rates of the rotation rings' neurons. Similar to previous work [82], we assume that synaptic time constants of the rotation rings are much faster than that of the central ring. This allows us

[1] We follow the convention $F_{\text{x-y}}$ to denote a directional quantity F, such as a synaptic weight or current, from neurons of population x to neurons of population y.

to model the firing rates of rotation rings' neurons as algebraic quantities whose values are determined instantaneously by the synaptic inputs they receive, namely,

$$\begin{aligned} r_{\text{cw}}(t,\psi) &= \sigma\left[W_{\text{c-cw}}(\psi)r_{\text{c}}(t,\psi) + I_{\text{v-cw}}(t,\psi) - \bar{I}\right], \\ r_{\text{ccw}}(t,\psi) &= \sigma\left[W_{\text{c-ccw}}(\psi)r_{\text{c}}(t,\psi) + I_{\text{v-ccw}}(t,\psi) - \bar{I}\right], \end{aligned} \tag{26}$$

where $W_{\text{c-cw}}, W_{\text{c-ccw}} : S^1 \to \mathbb{R}_{\geq 0}$ denote the synaptic weights describing the one-to-one connections between the central ring and the each rotation ring, $\bar{I}$ denotes global inhibition, and $I_{\text{v-cw}}, I_{\text{v-ccw}}$ denote the synaptic inputs from the CW, CCW velocity neurons to the respective rotation rings, namely

$$\begin{aligned} I_{\text{v-cw}}(t,\psi) &= W_{\text{v-cw}}(\psi)u_{\text{cw}}(t), \\ I_{\text{v-ccw}}(t,\psi) &= W_{\text{v-ccw}}(\psi)u_{\text{ccw}}(t), \end{aligned} \tag{27}$$

where $W_{\text{v-cw}}, W_{\text{v-ccw}} : S^1 \to \mathbb{R}_{\geq 0}$ denote the synaptic weight functions describing the one-to-all connections between the velocity neurons and the respective rotation rings, and $u_{\text{cw}}, u_{\text{ccw}}$ denotes the firing rates of the velocity neurons. Following previous work [31], we assume that these firing rates simply follow a linear relationship with the animal's velocity v as follows:

$$\begin{aligned} u_{\text{cw}}(t) &= u^0 - \alpha_{\text{cw}}v(t), \\ u_{\text{ccw}}(t) &= u^0 - \alpha_{\text{ccw}}v(t), \end{aligned} \tag{28}$$

where $u^0 \geq 0$ denote the intercept of velocity neurons' tuning curves, and $\alpha_{\text{cw}} > 0$, $\alpha_{\text{ccw}} > 0$ denote the slopes of the velocity neurons' tuning curves. Substituting this model of velocity neurons into Equation (27), we obtain a more explicit expression for the synaptic inputs from the velocity neurons to the rotation rings as follows:

$$\begin{aligned} I_{\text{v-cw}}(t,\psi) &= W_{\text{v-cw}}(\psi)u_{\text{cw}}(t) = W_{\text{v-cw}}(\psi)\left(u^0 - \alpha_{\text{cw}}v(t)\right), \\ I_{\text{v-ccw}}(t,\psi) &= W_{\text{v-ccw}}(\psi)u_{\text{ccw}}(t) = W_{\text{v-ccw}}(\psi)\left(u^0 + \alpha_{\text{ccw}}v(t)\right). \end{aligned} \tag{29}$$

Collectively, Equations (26), and (29) constitute the model of the rotation rings.

6.2 Model reduction of network dynamics

6.2.1 An ansatz solution to the central ring

It is well known that when the recurrent weight function $W_{\text{c-c}}$ has the classical pattern of local excitation and global inhibition (LEGI), a persistent activity bump emerges in the central ring [1,38–40]. As done in the previous work [41,42], this bump of activity can be identified with a one-dimensional ansatz solution to Equation (23):

$$r_{\text{c}}(t,\psi) = r_{\text{c}}^{\star}(\psi - \theta), \tag{30}$$

where θ denotes the location of the activity bump according to the ansatz function $r_{\text{c}}^{\star}$ whose properties are stated in the Assumption below.

Assumption 1. *The function $r_c^{\star} : S^1 \to \mathbb{R}_{\geq 0}$ has even symmetry and compact support around its single*

peak $\psi = \theta$, namely,

$$r_c^\star(x) = r_c^\star(-x),$$

$$r_c^\star(x) = r_c^\star(x + 2\pi),$$

$$\text{supp}(r_c^\star) = [-\Omega_c, +\Omega_c],$$

where Ω_c denotes the limits of the support, corresponding to the half-width of the activity bump. Furthermore, the function $r_c^\star$ is twice differentiable, and there exists a non-empty open neighborhood $\mathcal{N}$ of the origin where the second derivative of $r_c^\star$ is negative, namely,

$$\frac{\partial^2 r_c^\star(x)}{\partial x^2} < 0, \ \forall x \in \mathcal{N}. \qquad \triangle$$

The LEGI pattern of recurrent weight function $W_{\text{c-c}}$ of the central ring maintains the shape of the activity bump $r_{\text{c}}^\star$, but does not control the location θ of the bump over the neural space. Rather, the external inputs I_{ext} to the central ring control θ. Therefore, for a given smoothly varying, external input $I_{\text{ext}} t, \psi$, the ansatz solution $r_{\text{c}}^\star$ to Equation (23) satisfies

$$-\tau_c \frac{d\theta}{dt} \frac{\partial r_{\text{c}}^\star(\psi - \theta)}{\partial \psi} = -r_{\text{c}}^\star(\psi - \theta) + \sigma\big[W_{\text{c-c}}(\psi) \circledast r_{\text{c}}^\star(\psi - \theta) + I_{\text{ext}}(t, \psi)\big], \tag{31}$$

since one can show (using the chain rule) that $\frac{\partial r_{\text{c}}^\star(\psi - \theta(t))}{\partial t} = -\frac{\partial r_{\text{c}}^\star(\psi - \theta(t))}{d\psi} \frac{\partial \theta}{dt}$.

6.2.2 An ansatz solution to the rotation rings

Substituting the ansatz solution $r_{\text{c}}^\star$ for the central ring dynamics into Equations (26) and (29), we can also solve for the firing rates of the rotation rings as follows:

$$\begin{aligned} r_{\text{cw}}(t, \psi) &= r_{\text{cw}}^\star(\psi, \theta(t), v) = \sigma\big[W_{\text{c-cw}}(\psi) r_{\text{c}}^\star(\psi - \theta(t)) + W_{\text{v-cw}}(\psi)(u^0 - \alpha_{\text{cw}} v) - \bar{I}\big], \\ r_{\text{ccw}}(t, \psi) &= r_{\text{ccw}}^\star(\psi, \theta(t), v) = \sigma\big[W_{\text{c-ccw}}(\psi) r_{\text{c}}^\star(\psi - \theta(t)) + W_{\text{v-ccw}}(\psi)(u^0 + \alpha_{\text{ccw}} v) - \bar{I}\big], \end{aligned} \tag{32}$$

where we introduced the functions $r_{\text{cw}}^\star(\psi, \theta, v)$, $r_{\text{ccw}}^\star(\psi, \theta, v)$ as a shorthand notation for the ansatz solutions of the activity of the rotation rings. Since firing rates of rotation rings' neurons depend on the ansatz solution $r_{\text{c}}^\star$ for the central ring, the ansatz solutions $r_{\text{cw}}^\star$, $r_{\text{ccw}}^\star$ for the rotation rings can have a compact support interval like $r_{\text{c}}^\star$, as formally stated below:

Assumption 2. *The global inhibition $\bar{I}$ to the CW and CCW rotation rings is sufficiently high such that*

1. *for a given velocity v and the bump location θ, the functions $r_{cw}^\star, r_{ccw}^\star : S^1 \to \mathbb{R}$ have compact support intervals over ψ:*

$$\begin{aligned} \text{supp}\,\{r_{cw}^\star(\cdot, \theta, v)\} &= [\theta - \Omega_{cw}^l(v, \theta), \theta + \Omega_{cw}^r(v, \theta)] \subset S^1, \\ \text{supp}\,\{r_{ccw}^\star(\cdot, \theta, v)\} &= [\theta - \Omega_{ccw}^l(v, \theta), \theta + \Omega_{ccw}^r(v, \theta)] \subset S^1. \end{aligned} \tag{33}$$

where the half widths $\Omega_{\{cw,ccw\}}^{\{l,r\}}(v, \theta)$ potentially depend on velocity v and bump location θ. We also

assume the total widths are not zero and do not cover the entire extent of S^1, namely,

$$0 < \Omega^l_{cw}(v,\theta) + \Omega^r_{cw}(v,\theta) < 2\pi,$$
$$0 < \Omega^l_{ccw}(v,\theta) + \Omega^r_{ccw}(v,\theta) < 2\pi.$$

2. *these compact support intervals are contained inside $\mathcal{N}$, the open set where the second derivative of the ansatz function $r^\star_c$ is negative as described in Assumption 1.* △

Notice that we assume the support intervals of the rotation rings' persistent activity bumps $r^\star_{\text{cw}}, r^\star_{\text{ccw}}$ during the animal's immobility are defined relative to the peak location θ of the central ring's activity bump $r^\star_{\text{c}}(\psi-\theta)$. However, unlike the central ring's activity bump, the widths of the rotation rings' activity bumps are not necessarily fixed. That is, the left and right half-widths of the activity bumps may vary with v and θ, depending on the animal's velocity and the spatial variance of the synaptic weight functions $W_{\text{c-cw}}$, $W_{\text{c-ccw}}$, $W_{\text{v-cw}}$, and $W_{\text{v-ccw}}$.

6.2.3 Dynamics of the position representation θ during animal's immobility in the absence of landmarks

During the animal's immobility (i.e., $v = 0$), Equation (32), the solution to rotation ring dynamics, reduces to

$$\begin{aligned} r^0_{\text{cw}}(t,\psi) &= r^\star_{\text{cw}}(\psi,\theta(t),0) = \sigma\big[W_{\text{c-cw}}(\psi)r^\star_{\text{c}}(\psi-\theta(t)) + W_{\text{v-cw}}(\psi)u^0 - \bar{I}\big], \\ r^0_{\text{ccw}}(t,\psi) &= r^\star_{\text{ccw}}(\psi,\theta(t),0) = \sigma\big[W_{\text{c-ccw}}(\psi)r^\star_{\text{c}}(\psi-\theta(t)) + W_{\text{v-ccw}}(\psi)u^0 - \bar{I}\big], \end{aligned} \tag{34}$$

where the superscript 0 denotes that a quantity is evaluated when the animal is stationary (i.e., $v = 0$) without visual landmarks. Furthermore, when there are no landmarks, the firing rate of visual ring neurons becomes $r^0_{\text{vis}}(t,\phi) = 0$ for all t, ϕ. Substituting these solutions into Equation (25), we obtain

$$\begin{aligned} I^0_{\text{vis-c}}(t,\psi) &= 0, \\ I^0_{\text{cw-c}}(t,\psi) &= W_{\text{cw-c}}(\psi+b)\sigma\big[W_{\text{c-cw}}(\psi+b)r^\star_{\text{c}}(\psi+b-\theta(t)) + W_{\text{v-cw}}(\psi+b)u^0 - \bar{I}\big], \\ I^0_{\text{ccw-c}}(t,\psi) &= W_{\text{ccw-c}}(\psi-b)\sigma\big[W_{\text{c-ccw}}(\psi-b)r^\star_{\text{c}}(\psi-b-\theta(t)) + W_{\text{v-ccw}}(\psi-b)u^0 - \bar{I}\big]. \end{aligned} \tag{35}$$

Combining these synaptic inputs according to Equation (24), we obtain the total external input to the central ring during the animal's immobility in the absence of landmarks as

$$\begin{aligned} I^0_{\text{ext}}(t,\psi) = {} & W_{\text{cw-c}}(\psi+b)\sigma\big[W_{\text{c-cw}}(\psi+b)r^\star_{\text{c}}(\psi+b-\theta) + W_{\text{v-cw}}(\psi+b)u^0 - \bar{I}\big] \\ & + W_{\text{ccw-c}}(\psi-b)\sigma\big[W_{\text{c-ccw}}(\psi-b)r^\star_{\text{c}}(\psi-b-\theta) + W_{\text{v-ccw}}(\psi-b)u^0 - \bar{I}\big]. \end{aligned} \tag{36}$$

The external synaptic input I_{ext} drives the temporal changes in θ according to the differential equation in (31). In a stable implementation of the path integration (PI), I^0_{ext} must be such that the equilibrium conditions of that differential equation is satisfied as shown below:

$$-\tau_c \frac{d\theta}{dt}\frac{\partial r^\star_{\text{c}}(\psi-\theta)}{\partial\psi} = -r^\star_{\text{c}}(\psi-\theta) + \sigma\big[W_{\text{c-c}}(\psi) \circledast r^\star_{\text{c}}(\psi-\theta) + I^0_{\text{ext}}(t,\psi)\big] = 0. \tag{37}$$

This implies that θ remains constant during immobility of the animal, namely,

$$\frac{d\theta(t)}{dt} = 0. \tag{38}$$

In the remark below, we describe a necessary property of I^0_{ext} to satisfy these equilibrium conditions.

Remark 1. *In steady-state, Equation (37) requires the terms on its right-hand side to satisfy*

$$r_c^\star(\psi-\theta) = \sigma[W_{c\text{-}c}(\psi) \circledast r_c^\star(\psi-\theta) + I^0_{ext}(t,\psi)]. \tag{39}$$

This requirement can be translated into different requirements for the cases that ψ is inside or outside the support interval of $r_c^\star$ as follows:

$$\begin{aligned} r_c^\star(\psi-\theta) &= W_{c\text{-}c}(\psi) \circledast r_c^\star(\psi-\theta) + I^0_{ext}(t,\psi) \quad \forall \psi-\theta \in [-\Omega_c, \Omega_c], \\ 0 &\geq W_{c\text{-}c}(\psi) \circledast r_c^\star(\psi-\theta) + I^0_{ext}(t,\psi) \quad \forall \psi-\theta \notin [-\Omega_c, \Omega_c]. \end{aligned} \tag{40}$$

Since $r_c^\star$ and $W_{c\text{-}c}$ are even symmetric functions about the origin, the first line in Equation (40) implies that the total external synaptic input I^0_{ext} must also be even symmetric inside the interval $[-\Omega_c, \Omega_c]$ independent of the choice of b, namely,

$$I^0_{ext}(t,\phi+\theta) = I^0_{ext}(t,-\phi+\theta), \forall \phi \in [-\Omega_c, \Omega_c], \theta \in S^1. \tag{41}$$

△

Note that in the present work, we constrain the model parameters such that this symmetry property of I^0_{ext}, is satisfied to ensure stable PI.

6.2.4 Dynamics of the position representation θ without visual landmarks but with movement, i.e. pure path integration

In the present section, we follow the dimensionality reduction protocol described in [41] to derive the one-dimensional differential equation model of the position representation θ as a function of the animal's velocity v. Briefly, this protocol first computes the change in the external synaptic input of the central ring with the movement compared to the baseline case of animal's immobility, and then map this input change to a change in the position representation θ.

Computing the change in the synaptic inputs to the central ring with the movement

When the animal is immobile, CW and CCW velocity neurons fire at their baseline rates $u_{\text{cw}} = u^0$ and $u_{\text{ccw}} = u^0$. During movement of the animal, firing rates of these velocity neurons change in different directions:

$$\Delta u_{\text{cw}}(t) \triangleq u_{\text{cw}}(t) - u^0 = -\alpha_{\text{cw}} v(t),$$

$$\Delta u_{\text{ccw}}(t) \triangleq u_{\text{ccw}}(t) - u^0 = +\alpha_{\text{ccw}} v(t).$$

These directionally different changes in the firing rates of the velocity neurons carry over into their synaptic drive to the rotation rings via

$$\begin{aligned} \Delta I_{\text{v-cw}}(t,\psi) &\triangleq I_{\text{v-cw}}(t,\psi) - I^0_{\text{v-cw}}(t,\psi) = -\alpha_{\text{cw}} W_{\text{v-cw}}(\psi) v(t), \\ \Delta I_{\text{v-ccw}}(t,\psi) &= I_{\text{v-ccw}}(t,\psi) - I^0_{\text{v-ccw}}(t,\psi) = +\alpha_{\text{ccw}} W_{\text{v-ccw}}(\psi) v(t). \end{aligned} \tag{42}$$

Because the firing rates of the rotation ring neurons are determined instantaneously by their synaptic inputs, these input changes during the movement lead to instantaneous perturbation of their activity bumps during the immobility. To find an approximate analytical formulation of these perturbations, we follow the previous work [41] and linearize the algebraic model of rotation ring firing rates, given in Equation (26), about the immobility condition $v = 0$:

$$\begin{aligned} \Delta r_{\text{cw}}(t,\psi) &= \left.\frac{\partial r_{\text{cw}}(t,\psi)}{\partial\, I_{\text{v-cw}}}\right|_{v=0} \Delta I_{\text{v-cw}}(t,\psi) = -\left.\frac{\partial r_{\text{cw}}(t,\psi)}{\partial\, I_{\text{v-cw}}}\right|_{v=0} \alpha_{\text{cw}} W_{\text{v-cw}}(\psi) v(t), \\ \Delta r_{\text{ccw}}(t,\psi) &= \left.\frac{\partial r_{\text{ccw}}(t,\psi)}{\partial\, I_{\text{v-ccw}}}\right|_{v=0} \Delta I_{\text{v-ccw}}(t,\psi) = +\left.\frac{\partial r_{\text{ccw}}(t,\psi)}{\partial\, I_{\text{v-ccw}}}\right|_{v=0} \alpha W_{\text{v-ccw}}(\psi) v(t). \end{aligned} \tag{43}$$

Explicitly differentiating the algebraic model (26), we obtain

$$\begin{aligned} \left.\frac{\partial r_{\text{cw}}(t,\psi)}{\partial I_{\text{v-cw}}}\right|_{v=0} &= \left.\frac{\partial \sigma\left[W_{\text{c-cw}}(\psi) r_{\text{c}}(t,\psi) + I_{\text{v-cw}}(t,\psi) - \bar{I}\right]}{\partial\, I_{\text{v-cw}}}\right|_{v=0}, \\ \left.\frac{\partial r_{\text{ccw}}(t,\psi)}{\partial I_{\text{v-ccw}}}\right|_{v=0} &= \left.\frac{\partial \sigma\left[W_{\text{c-ccw}}(\psi) r_{\text{c}}(t,\psi) + I_{\text{v-ccw}}(t,\psi) - \bar{I}\right]}{\partial\, I_{\text{v-ccw}}}\right|_{v=0}. \end{aligned} \tag{44}$$

To find closed-form, analytical expressions of these terms, we follow the previous work [41] that approximates the derivative of the linear threshold-type activation function σ as follows:

$$\left.\frac{d\sigma[f(v)]}{dv}\right|_{v=0} \approx \text{sign}\left[\sigma[f(0)]\right] f'(0) = \begin{cases} f'(0) & \text{if } f(0) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Using this approximation and Equation (26), we obtain

$$\begin{aligned} \left.\frac{\partial r_{\text{cw}}(t,\psi)}{\partial I_{\text{v-cw}}}\right|_{v=0} &\approx \text{sign}[r^{\star}_{\text{cw}}(\psi,\theta,0)] \left(1 + W_{\text{c-cw}}(\psi) \left.\frac{\partial r_c(t,\psi)}{\partial I_{\text{v-cw}}}\right|_{v=0}\right), \\ \left.\frac{\partial r_{\text{ccw}}(t,\psi)}{\partial I_{\text{v-ccw}}}\right|_{v=0} &\approx \text{sign}[r^{\star}_{\text{ccw}}(\psi,\theta(t),0)] \left(1 + W_{\text{c-ccw}}(\psi) \left.\frac{\partial r_c(t,\psi)}{\partial I_{\text{v-ccw}}}\right|_{v=0}\right). \end{aligned} \tag{45}$$

Recall that firing rate of the central ring neurons change according the differential equation in (23), implying that they are independent of the received synaptic inputs at a given time, namely,

$$\frac{\partial r_c(t,\psi)}{\partial I_{\text{v-cw}}} = \frac{\partial r_c(t,\psi)}{\partial I_{\text{v-ccw}}} = 0.$$

Substituting this into Equation (45) leads to

$$\begin{aligned} \left.\frac{\partial r_{\text{cw}}(t,\psi)}{\partial I_{\text{v-cw}}}\right|_{v=0} &\approx \text{sign}\left[r^{\star}_{\text{cw}}(\psi,\theta(t),0)\right], \\ \left.\frac{\partial r_{\text{ccw}}(t,\psi)}{\partial I_{\text{v-ccw}}}\right|_{v=0} &\approx \text{sign}\left[r^{\star}_{\text{ccw}}(\psi,\theta(t),0)\right]. \end{aligned} \tag{46}$$

Substituting these expressions into Equation (43), we finally obtain closed-form, analytical expressions of the firing rate perturbations for rotation rings as

$$\begin{aligned}\Delta r_{\text{cw}}(t,\psi) &\approx -\alpha_{\text{cw}} v(t) W_{\text{v-cw}}(\psi)\,\text{sign}[r^{\star}_{\text{cw}}(\psi,\theta(t),0)],\\ \Delta r_{\text{ccw}}(t,\psi) &\approx +\alpha_{\text{ccw}} v(t) W_{\text{v-ccw}}(\psi)\,\text{sign}[r^{\star}_{\text{ccw}}(\psi,\theta(t),0).\end{aligned} \tag{47}$$

To map these firing rate perturbations into the disturbances in the synaptic inputs to the central ring, we return to Equation (25). First, we substitute $r_{\text{cw}}(t,\psi) = r^0_{\text{cw}}(t,\psi) + \Delta r_{\text{cw}}(t,\psi)$ into the first line of Equation (25) and $r_{\text{ccw}}(t,\psi) = r^0_{\text{ccw}}(t,\psi) + \Delta r_{\text{ccw}}(t,\psi)$ into the second line. After subtracting the results of these substitutions from the last two lines in Equation (35), we obtain

$$\begin{aligned}\Delta I_{\text{cw-c}}(t,\psi) &\triangleq I_{\text{cw-c}}(t,\psi) - I^0_{\text{cw-c}}(t,\psi) \approx W_{\text{cw}}(\psi+b)\Delta r_{\text{cw}}(t,\psi+b) \approx -\mathcal{H}_{\text{cw}}(\psi+b,\theta)v(t),\\ \Delta I_{\text{ccw-c}}(t,\psi) &\triangleq I_{\text{ccw-c}}(t,\psi) - I^0_{\text{ccw-c}}(t,\psi) \approx W_{\text{ccw}}(\psi-b)\Delta r_{\text{ccw}}(t,\psi-b) \approx +\mathcal{H}_{\text{ccw}}(\psi-b,\theta)v(t),\end{aligned} \tag{48}$$

where

$$\begin{aligned}\mathcal{H}_{\text{cw}}(\psi+b,\theta) &\triangleq \alpha_{\text{cw}} W_{\text{cw-c}}(\psi+b) W_{\text{v-cw}}(\psi+b)\,\text{sign}[r^{\star}_{\text{cw}}(\psi+b,\theta(t),0)],\\ \mathcal{H}_{\text{ccw}}(\psi-b,\theta) &\triangleq \alpha_{\text{ccw}} W_{\text{ccw-c}}(\psi-b) W_{\text{v-ccw}}(\psi-b)\,\text{sign}[r^{\star}_{\text{ccw}}(\psi-b,\theta(t),0)].\end{aligned} \tag{49}$$

We can compute the total disturbance by simply adding up these individual disturbances:

$$\Delta I_{\text{ext}}(t,\psi) \triangleq I_{\text{ext}}(t,\psi) - I^0_{\text{ext}}(t,\psi) = \Delta I_{\text{cw-c}}(t,\psi) + \Delta I_{\text{ccw-c}}(t,\psi) = \big(\mathcal{H}_{\text{ccw}}(\psi-b,\theta) - \mathcal{H}_{\text{cw}}(\psi+b,\theta)\big)v(t). \tag{50}$$

Relation between the synaptic input change and the change in θ

We next derive how this synaptic input disturbance modifies the location θ of the activity bump, described by the ansatz function $r^{\star}_{\text{c}}$. For this derivation, we substitute the above equation for the synaptic input disturbance into Equation (31), namely the dynamics of θ. This substitution yields

$$\tau_c \frac{\partial r^{\star}_{\text{c}}(\psi-\theta)}{\partial t} = -\tau_c \frac{d\theta}{dt}\frac{\partial r^{\star}_{\text{c}}(\psi-\theta)}{\partial \psi} = -r^{\star}_{\text{c}}(\psi-\theta) + \sigma\big[W_{\text{c-c}}(\psi) \circledast r^{\star}_{\text{c}}(\psi-\theta) + I^0_{\text{ext}} + \Delta I_{\text{ext}}(t,\psi)\big]. \tag{51}$$

Recall that, when $\Delta I_{\text{ext}}(t,\psi) = 0$ and I^0_{ext} satisfies the symmetry requirement in Equation (41), the above differential equation reduces to the equilibrium condition in Equation (38), namely,

$$\frac{\partial r^{\star}_{\text{c}}(\psi-\theta)}{\partial t} = \frac{d\theta}{dt} = 0.$$

Thus, the addition of dynamic synaptic input $\Delta I_{\text{ext}}(t,\psi)$ into Equation (51) perturbs the dynamics away from these equilibrium conditions.

As previously assumed, the attractor dynamics allow only rotation of the bump (i.e., change in θ) in response to external inputs. To find the resulting rotation in response to the disturbance ΔI_{ext}, we first project[2] ΔI_{ext} onto the tangent subspace of the ansatz function $r^{\star}_{\text{c}}(\psi-\theta)$, described by the gradient of

[2]Let $f, g : S^1 \to \mathbb{R}$. $\langle f, g\rangle \triangleq \int_{-\pi}^{+\pi} f(x)g(x)dx$ denote the inner product, and $\|f\| \triangleq \sqrt{\langle f, f\rangle}$ denote the L_2 norm. Projection operator is defined as $\text{proj}_f(g) = \frac{f}{\|f\|}\langle \frac{f}{\|f\|}, g\rangle = \frac{f}{\|f\|^2}\langle f, g\rangle$,.

$r_c^\star(\psi-\theta)$ with respect to θ, as follows:

$$\tilde{I}_{\mathcal{T}}(t,\psi) = \text{proj}_{\partial_\theta r_c^\star}(\Delta I_{\text{ext}}) = \frac{1}{\left\|\frac{\partial r_c^\star(\psi)}{\partial\theta}\right\|^2}\frac{\partial r_c^\star(\psi-\theta)}{\partial\theta}\left\langle\frac{\partial r_c^\star(\psi-\theta)}{\partial\theta}, \Delta I_{\text{ext}}(t,\psi)\right\rangle. \tag{52}$$

We then replace the term ΔI_{ext} in the central ring dynamics in Equation (51) with its tangential component $\tilde{I}_{\mathcal{T}}$, which leads to

$$\tau_c\frac{\partial r_c^\star(\psi-\theta)}{\partial t} = -\tau_c\frac{d\theta}{dt}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi} = -r_c^\star(\psi-\theta) + \sigma\big[W_{\text{c-c}}(\psi)\circledast r_c^\star(\psi-\theta) + I_{\text{ext}}^0 + \tilde{I}_{\mathcal{T}}(t,\psi)\big]. \tag{53}$$

Since the right hand side of this equation is zero when $\tilde{I}_{\mathcal{T}} = 0$ (see Equation (37)), we can reduce the right-hand side to

$$-\tau_c\frac{d\theta}{dt}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi} = \tilde{I}_{\mathcal{T}}(t,\psi). \tag{54}$$

See the remarks and the proposition at the end of this section for a formal proof of this reduction. Using Equation (52), we can rewrite this differential equation more explicitly as

$$-\tau_c\frac{d\theta}{dt}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi} = \frac{1}{\left\|\frac{\partial r_c^\star(\psi)}{\partial\psi}\right\|^2}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\left\langle\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}, \Delta I_{\text{ext}}(t,\psi)\right\rangle. \tag{55}$$

Eliminating the common terms in both sides simplifies this differential equation to

$$\frac{d\hat{\theta(t)}}{dt} = -\frac{1}{\tau_c\left\|\frac{\partial r_c^\star}{\partial\psi}\right\|^2}\left\langle\frac{\partial r_c^\star(\psi-\theta(t))}{\partial\psi}, \Delta I_{\text{ext}}(t,\psi)\right\rangle.$$

We then write the inner product more explicitly by using its integral-based definition, which yields

$$\frac{d\theta(t)}{dt} = -\frac{1}{\tau_c\left\|\frac{\partial r_c^\star}{\partial\psi}\right\|^2}\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta(t))}{\partial\psi}\Delta I_{\text{ext}}(t,\psi)d\psi. \tag{56}$$

Since ΔI_{ext} depends linearly on the animal's velocity v (see Equation (50)), this equation implies

$$\frac{d\theta}{dt} = k(\theta)v(t), \tag{57}$$

where $k : S^1 \to \mathbb{R}_{>0}$ denotes the PI gain and takes the form

$$k(\theta) \triangleq \frac{1}{\tau_c\left\|\frac{\partial r_c^\star}{\partial\psi}\right\|^2}\left(\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\mathcal{H}_{\text{cw}}(\psi+b,\theta)d\psi - \int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\mathcal{H}_{\text{ccw}}(\psi-b,\theta)d\psi\right). \tag{58}$$

Recasting the analytical expression of the PI gain into a simpler form

Next, we manipulate Equation (58) to garner insight into the analytical expression of the PI gain and to simplify it. To this end, we begin by adding and subtracting the term $\big(\mathcal{H}_{\text{ccw}}(\psi+b,\theta) - \mathcal{H}_{\text{cw}}(\psi+b,\theta)\big)/2$ to and from the first integrand and the term $\big(\mathcal{H}_{\text{cw}}(\psi-b,\theta) - \mathcal{H}_{\text{ccw}}(\psi-b,\theta)\big)/2$ to and from the second integrand, which yields

$$\begin{aligned} k(\theta) = \frac{1}{\tau_c\left\|\frac{\partial r_c^\star}{\partial\psi}\right\|^2}\Bigg[&\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\frac{\big(\mathcal{H}_{\text{cw}}(\psi+b,\theta)+\mathcal{H}_{\text{ccw}}(\psi+b,\theta)\big)-\big(\mathcal{H}_{\text{cw}}(\psi-b,\theta)+\mathcal{H}_{\text{ccw}}(\psi-b,\theta)\big)}{2}d\psi \\ &+\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\frac{\big(\mathcal{H}_{\text{cw}}(\psi+b,\theta)-\mathcal{H}_{\text{ccw}}(\psi+b,\theta)\big)+\big(\mathcal{H}_{\text{cw}}(\psi-b,\theta)-\mathcal{H}_{\text{ccw}}(\psi-b,\theta)\big)}{2}d\psi\Bigg]. \end{aligned} \tag{59}$$

Equation (59) reveals that two integral terms contribute to the PI gain:

- The first integral quantifies the contribution of the *differential* component of the positively and negatively offset (b) inputs to the central ring from the rotation rings.
- The second integral quantifies the contribution of the residual, *common-mode* component of the positively and negatively offset (b) inputs to the central ring from the rotation rings.

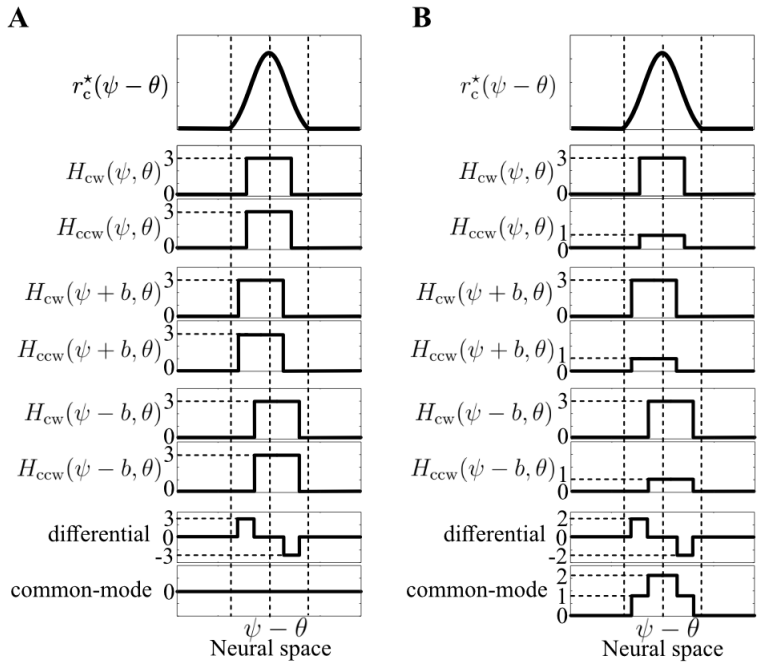

Figure 6: Examples showing that *common-mode* component results in an even function. (A) If the terms H_{cw},H_{ccw} are symmetric, the *common-mode* becomes zero, making the result of the second integral in Equation (59) also zero. (B) If the terms H_{cw},H_{ccw} are not symmetric, the *common-mode* results in a nonzero even function. Since the second-integral in Equation (59) is the inner product of the *common-mode* with an odd function, namely, the gradient of the ansatz $r_c^\star$, the result of the second integral again becomes zero.

As shown by the motivating examples in Fig. 6, the *differential* component is an odd function while the *common-mode* component is an even function. The integrals in Equation (59) compute the inner product of these components with an odd function, namely, the gradient of the ansatz $r_c^\star$. Thus, the first integral, including the *differential* component, yields a nonzero value but the second integral, including the *common-mode* component, yields zero. Based on these motivating examples, we assume that the second integral in Equation (59) has none or negligible effects on the gain, compared to the first integral.

Assumption 3. *We assume that parameters of the ring attractor network satisfy*

$$\frac{1}{\tau_c\left\|\frac{\partial r_c^\star}{\partial\psi}\right\|^2}\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\Big(\mathcal{H}_{cw}(\psi+b,\theta)-\mathcal{H}_{ccw}(\psi+b,\theta)+\mathcal{H}_{cw}(\psi-b,\theta)-\mathcal{H}_{ccw}(\psi-b,\theta)\Big)d\psi=0. \quad \triangle$$

Under this assumption, Equation (59) reduces to

$$k(\theta)=\frac{1}{\tau_c\left\|\frac{\partial r_c^\star}{\partial\psi}\right\|^2}\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\frac{\big(\mathcal{H}_{\rm cw}(\psi+b,\theta)+\mathcal{H}_{\rm ccw}(\psi+b,\theta)\big)-\big(\mathcal{H}_{\rm cw}(\psi-b,\theta)+\mathcal{H}_{\rm ccw}(\psi-b,\theta)\big)}{2}d\psi.$$

To further simplify this expression, we partition the integral into a difference of two integrals :

$$k(\theta)=\frac{1}{\tau_c\left\|\frac{\partial r_c^\star}{\partial\psi}\right\|^2}\left(\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\frac{\mathcal{H}_{\rm cw}(\psi+b,\theta)+\mathcal{H}_{\rm ccw}(\psi+b,\theta)}{2}d\psi-\int_0^{2\pi}\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}\frac{\mathcal{H}_{\rm cw}(\psi-b,\theta)+\mathcal{H}_{\rm ccw}(\psi-b,\theta)}{2}d\psi\right).$$

In this new form, we then perform a change of the integration variables to $z \triangleq \psi - \theta + b$ for the first integral and to $z \triangleq \psi - \theta - b$ for the second integral, which yields

$$k(\theta) = \frac{1}{\tau_c \| \frac{\partial r_c^\star}{\partial \psi} \|^2} \left(\int_0^{2\pi} \frac{\partial r_c^\star(z-b)}{\partial z} \frac{\mathcal{H}_{\text{cw}}(z+\theta,\theta) + \mathcal{H}_{\text{ccw}}(z+\theta,\theta)}{2} dz - \int_0^{2\pi} \frac{\partial r_c^\star(z-b)}{\partial z} \frac{\mathcal{H}_{\text{cw}}(z+\theta,\theta) + \mathcal{H}_{\text{ccw}}(z+\theta,\theta)}{2} dz \right).$$

Note that when performing this change of variables, the integration bounds do not need to be updated since the integrands are periodic, and the bounds cover their entire period. These two integrals can thus be combined into a single integral as follows:

$$k(\theta) = \frac{1}{2\tau_c \| \frac{\partial r_c^\star}{\partial \psi} \|^2} \int_0^{2\pi} \left(\frac{\partial r_c^\star(z-b)}{\partial z} - \frac{\partial r_c^\star(z-b)}{\partial z} \right) \big(\mathcal{H}_{\text{cw}}(z+\theta,\theta) + \mathcal{H}_{\text{ccw}}(z+\theta,\theta) \big) dz. \tag{60}$$

Assuming a sufficiently small offset b, we can employ the first-order Taylor series approximation to the difference term

$$\frac{\partial r_c^\star(z-b)}{\partial z} - \frac{\partial r_c^\star(z-b)}{\partial z} \approx -2\, b\, \frac{\partial^2 r_c^\star(z)}{\partial z^2}. \tag{61}$$

This approximation simplifies Equation (60), yielding a concise analytical expression of the PI gain

$$k(\theta) = \frac{-b}{\tau_c \| \frac{\partial r_c^\star}{\partial \psi} \|^2} \int_0^{2\pi} \frac{\partial^2 r_c^\star(z)}{\partial z^2} \big(\mathcal{H}_{\text{cw}}(z+\theta,\theta) + \mathcal{H}_{\text{ccw}}(z+\theta,\theta) \big) dz.$$

For notational consistency with the rest of the present paper, we make a change of variables once more and substitute $\psi \triangleq z + \theta$ into this equation, which yields

$$\boxed{k(\theta) = \frac{-b}{\tau_c \| \frac{\partial r_c^\star}{\partial \psi} \|^2} \int_0^{2\pi} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} \Big(\mathcal{H}_{\text{cw}}(\psi,\theta) + \mathcal{H}_{\text{ccw}}(\psi,\theta) \Big) d\psi,} \tag{62}$$

where the terms H_{cw},H_{ccw} are given as Equation (49). Throughout the paper, we mainly use this equation as an analytical expression for the PI gain. In the Remark below, we also would like to present an alternative expression that will be useful later when deriving mechanistic requirements for the gain recalibration.

Remark 2. *To derive the alternative expression, we examine the support interval of the terms constituting the integral in Equation (62) and re-define the boundaries of the integral accordingly.*

We begin this examination with the terms $\mathcal{H}_{cw}, \mathcal{H}_{ccw}$. By definition (see Equation (49)), these terms have the same compact support interval as the ansatz functions $r_{cw}^\star$, $r_{ccw}^\star$, describing the solution to activity of CW, CCW rotation ring during the animal's immobility ($v = 0$), namely,

$$supp_\psi(\mathcal{H}_{cw}^\star(\psi,\theta)) = [\theta - \Omega_{cw}^l(0,\theta), \theta + \Omega_{cw}^r(0,\theta)]$$
$$supp_\psi(\mathcal{H}_{ccw}^\star(\psi,\theta)) = [\theta - \Omega_{ccw}^l(0,\theta), \theta + \Omega_{ccw}^r(0,\theta)].$$

Since union of compact sets is compact, support of the summation $\mathcal{H}_{cw} + \mathcal{H}_{ccw}$ is also compact, i.e.,

$$supp_\psi(\mathcal{H}_{cw} + \mathcal{H}_{ccw}) = supp_\psi(\mathcal{H}_{cw}) \cup supp_\psi(\mathcal{H}_{cw}).$$

Assuming that the union of the support intervals of $\mathcal{H}_{cw}$ and $\mathcal{H}_{ccw}$ is a connected set, we can express the support interval of the summation $\mathcal{H}_{cw} + \mathcal{H}_{ccw}$ with a single interval as shown below:

$$supp_{\psi}(\mathcal{H}_{cw} + \mathcal{H}_{ccw}) = [\theta - \Omega^{l}_{rot}(\theta), \theta + \Omega^{r}_{rot}(\theta)],$$

where

$$\begin{aligned} \Omega^{l}_{rot}(\theta) &= \max\Big(\Omega^{l}_{cw}(0,\theta), \Omega^{l}_{ccw}(0,\theta)\Big) \\ \Omega^{r}_{rot}(\theta) &= \max\Big(\Omega^{r}_{cw}(0,\theta), \Omega^{r}_{ccw}(0,\theta)\Big). \end{aligned} \tag{63}$$

We next examine the second derivative of the ansatz function $r^{\star}_{c}$. Recall that, if a function has a compact support interval, then its derivatives enjoy the same interval. This implies

$$supp_{\psi}\left(\frac{\partial^2 r^{\star}_{c}(\psi-\theta)}{\partial\psi^2}\right) = supp_{\psi}(r^{\star}_{c}(\psi-\theta)) = [\theta - \Omega_c, \theta + \Omega_c].$$

It follows from the second part of Assumption 2 that the support of the summation term $\mathcal{H}_{cw} + \mathcal{H}_{ccw}$ lies inside the support of the derivative term $\frac{d^2 r^{\star}_{c}(\psi)}{d\psi^2}$. Therefore, the integrand term in Equation (62) attains the same compact support interval as the summation $\mathcal{H}_{cw} + \mathcal{H}_{ccw}$, namely,

$$supp_{\psi}\left(\frac{\partial^2 r^{\star}_{c}(\psi-\theta)}{\partial\psi^2}\Big(\mathcal{H}_{cw}(\psi,\theta) + \mathcal{H}_{ccw}(\psi,\theta)\Big)\right) = supp_{\psi}(\mathcal{H}_{cw} + \mathcal{H}_{ccw}) = [\theta - \Omega^{l}_{rot}(\theta), \theta + \Omega^{r}_{rot}(\theta)].$$

This lets us reduce the bounds of the integration in Equation (62) to

$$\boxed{k(\theta) = \frac{-b}{\tau_c\big\|\frac{\partial r^{\star}_{c}}{\partial\psi}\big\|^2}\int\limits_{\theta-\Omega^{l}_{rot}(\theta)}^{\theta+\Omega^{r}_{rot}(\theta)} \frac{\partial^2 r^{\star}_{c}(\psi-\theta)}{\partial\psi^2}\Big(\mathcal{H}_{cw}(\psi,\theta) + \mathcal{H}_{ccw}(\psi,\theta)\Big)d\psi} \tag{64}$$

Note that Equations (62) and (64) are analytically equivalent and used interchangeably throughout the paper as deemed appropriate. △

Finally, we describe in the Remark below that similar steps can be followed to rewrite the equality in Assumption 3 in a more compact form.

Remark 3. *By following the same steps as in between Equations (59)-(61), we obtain an alternate form for the equality in Assumption 3 as*

$$\frac{-b}{\tau_c\big\|\frac{\partial r^{\star}_{c}}{\partial\psi}\big\|^2}\int\limits_{0}^{2\pi} \frac{\partial^2 r^{\star}_{c}(\psi-\theta)}{\partial\psi^2}\big(\mathcal{H}_{cw}(\psi,\theta) - \mathcal{H}_{ccw}(\psi,\theta)\big)d\psi = 0.$$

Furthermore, by following the same steps as in Remark 2, we can choose the integration bounds more precisely as shown below:

$$\frac{-b}{\tau_c\big\|\frac{\partial r^{\star}_{c}}{\partial\psi}\big\|^2}\int\limits_{\theta-\Omega^{l}_{rot}(\theta)}^{\theta+\Omega^{r}_{rot}(\theta)} \frac{\partial^2 r^{\star}_{c}(\psi-\theta)}{\partial\psi^2}\big(\mathcal{H}_{cw}(\psi,\theta) - \mathcal{H}_{ccw}(\psi,\theta)\big)d\psi = 0.$$ △

Steps to reduce right-hand side of Equation (53) to right-hand side of Equation (54)

Remark 4. *As we mentioned previously, the ansatz function $r^{\star}_{c}$ and its derivative have the same compact*

support interval, namely,

$$supp_\psi\left(\frac{\partial r_c^\star(\psi-\theta(t))}{\partial\psi}\right) = supp_\psi(r_c^\star(\psi-\theta(t))) = [\theta(t)-\Omega_c, \theta(t)+\Omega_c].$$

Recall that, by definition in Equation (52), the tangential component $\tilde{I}_\mathcal{T}$ of the disturbance ΔI_{ext} lies in the same direction as the derivative $\frac{\partial r_c^\star(\psi-\theta)}{\partial\psi}$. This implies that

$$supp_\psi(\tilde{I}_\mathcal{T}(t,\psi)) = supp_\psi(r_c^\star(t,\psi-\theta(t))) = [\theta(t)-\Omega_c, \theta(t)+\Omega_c]. \qquad \triangle$$

Remark 5. *Let $f : S^1 \to [0,\infty)$ be a nonnegative function. Composition of the RELU activation function σ with f is equal to f, namely,*

$$\sigma[f(x)] = f(x), \quad \forall x. \qquad \triangle$$

Assumption 4. *Outside the support interval, the tangential component function $\tilde{I}_\mathcal{T}$ and the ansatz function $r_c^\star$ are equal to zero, namely,*

$$0 = \tilde{I}_\mathcal{T} t,\psi) = r_c^\star(\psi-\theta(t)) \quad \forall\psi \notin [\theta-\Omega_c, \theta+\Omega_c]. \tag{65}$$

Inside the support interval, they are not zero and assumed to satisfy

$$0 \le |\tilde{I}_\mathcal{T}(t,\psi)| \le r_c^\star(\psi-\theta(t)) \quad \forall\psi \in [\theta-\Omega_c, \theta+\Omega_c].$$

Together, these two conditions imply

$$\sigma[r_c^\star(\psi-\theta(t)) + \tilde{I}_\mathcal{T}(t,\psi)] = r_c^\star(\psi-\theta(t)) + \tilde{I}_\mathcal{T}(t,\psi). \qquad \triangle$$

Proposition 3. *The right-hand side of Equation (53) is equivalent to the right-hand side of Equation (54), namely,*

$$-r_c^\star(\psi-\theta(t)) + \sigma\big[W_{c\text{-}c}(\psi) \circledast r_c^\star(\psi-\theta(t)) + I_{ext}^0(t,\psi) + \tilde{I}_\mathcal{T}(t,\psi)\big] = \tilde{I}_\mathcal{T}(t,\psi). \tag{66}$$

Proof. We prove the statement in the proposition by showing that Equation (66) is true in a case by case manner for ψ outside and inside the support interval of the ansatz function $r_c^\star$.

1. Consider first that ψ is outside the support interval of the ansatz function $r_c^\star$. Since $\tilde{I}_\mathcal{T}(t,\psi)$ is zero in this case as per Equation (65), Equation (66) reduces to

$$-r_c^\star(\psi-\theta(t)) + \sigma\big[W_{c\text{-}c}(\psi) \circledast r_c^\star(\psi-\theta(t)) + I_{ext}^0\big] = 0, \quad \forall\psi \notin [\theta(t)-\Omega_c, \theta(t)+\Omega_c].$$

   By Equation (37), this is true.

2. Consider now the remaining case that ψ is inside the support interval of the ansatz function $r_c^\star$. In

this case, $\tilde{I}_{\mathcal{T}}(t,\psi)$ is not zero. Furthermore, as noted in Equation (40) the ansatz function satisfies

$$r_{\mathrm{c}}^{\star}(\psi-\theta(t)) = W_{\text{c-c}}(\psi) \circledast r_{\mathrm{c}}^{\star}(\psi-\theta(t)) + I_{\mathrm{ext}}^{0}(t,\psi), \quad \forall \psi \in [\theta(t)-\Omega_{\mathrm{c}}, \theta(t)+\Omega_{\mathrm{c}}].$$

Using this equation, we can simplify Equation (66) to

$$-r_{\mathrm{c}}^{\star}(\psi-\theta(t)) + \sigma[r_{\mathrm{c}}^{\star}(\psi-\theta(t)) + \tilde{I}_{\mathcal{T}}(t,\psi)] = \tilde{I}_{\mathcal{T}}(t,\psi), \quad \forall \psi \in [\theta-\Omega_{\mathrm{c}}, \theta+\Omega_{\mathrm{c}}]. \tag{67}$$

By Assumption 4, the activation function term on the left-hand side of this equation satisfies

$$\sigma[r_{\mathrm{c}}^{\star}(\psi-\theta(t)) + \tilde{I}_{\mathcal{T}}(t,\psi)] = r_{\mathrm{c}}^{\star}(\psi-\theta(t)) + \tilde{I}_{\mathcal{T}}(t,\psi).$$

Based on this equation, we can verify that Equation (67) is true. □

6.2.5 Dynamics of the position representation θ with visual landmarks

We continue following the dimensionality reduction protocol of [41] to extend Equation (57) of θ dynamics to the case where there is positional feedback from visual landmarks. To this end, we first extend the synaptic input disturbance ΔI_{ext} in Equation (50) by adding the visual component I_{vis}, namely,

$$\Delta I_{\mathrm{ext}}(t,\psi) = \Delta I_{\text{cw-c}}(t,\psi) + \Delta I_{\text{ccw-c}}(t,\psi) + I_{\text{vis-c}}(t,\psi). \tag{68}$$

We then plug this extended disturbance into the differential equation in (56), which yields

$$\frac{d\theta}{dt} = -\frac{1}{\tau_c \left\| \frac{\partial r_{\mathrm{c}}^{\star}}{\partial \psi} \right\|^2} \int_0^{2\pi} \frac{\partial r_{\mathrm{c}}^{\star}(\psi-\theta)}{\partial \psi} \big(\Delta I_{\mathrm{cw}}(t,\psi) + \Delta I_{\mathrm{ccw}}(t,\psi) + I_{\text{vis-c}}(t,\psi)\big) d\psi. \tag{69}$$

In the previous section, we analyzed this equation for $I_{\text{vis-c}}(t,\psi) = 0$ and found

$$\frac{d\theta}{dt} = -\frac{1}{\tau_c \left\| \frac{\partial r_{\mathrm{c}}^{\star}}{\partial \psi} \right\|^2} \int_0^{2\pi} \frac{\partial r_{\mathrm{c}}^{\star}(\psi-\theta)}{\partial \psi} \big(\Delta I_{\mathrm{cw}}(t,\psi) + \Delta I_{\mathrm{ccw}}(t,\psi)\big) d\psi = k(\theta) v(t).$$

This lets us simplify Equation (69) to

$$\frac{d\theta}{dt} = k(\theta) v - \frac{1}{\tau_c \left\| \frac{\partial r_{\mathrm{c}}^{\star}}{\partial \psi} \right\|^2} \int_0^{2\pi} \frac{\partial r_{\mathrm{c}}^{\star}(\psi-\theta)}{\partial \psi} I_{\text{vis-c}}(t,\psi) d\psi. \tag{70}$$

Next, we use Equation (7), expressing the assumed form of visual input $\Delta I_{\text{vis-c}}$, and obtain

$$\frac{d\theta}{dt} = k(\theta) v - \frac{1}{\tau_c \left\| \frac{d r_{\mathrm{c}}^{\star}(\psi)}{d\psi} \right\|^2} \int_0^{2\pi} \frac{\partial r_{\mathrm{c}}^{\star}(\psi-\theta)}{\partial \psi} \rho_{\mathrm{vis}}(\psi-\theta^{\star}) d\psi. \tag{71}$$

At the end of this section, we present three remarks. The first remark shows that the integral in Equation (71) can be reduced to a function, say $\beta : S^1 \to \mathbb{R}$, as follows:

$$\beta(\theta^{\star}-\theta) \triangleq -\frac{1}{\tau_c \left\| \frac{d r_{\mathrm{c}}^{\star}(\psi)}{d\psi} \right\|^2} \int_0^{2\pi} \frac{\partial r_{\mathrm{c}}^{\star}(\psi-\theta)}{\partial \psi} \rho_{\mathrm{vis}}(\psi-\theta^{\star}) d\psi. \tag{72}$$

Substituting this function into Equation (71) yields a simple form for the differential equation of θ

dynamics as follows:

$$\frac{d}{dt}\theta = kv + \beta(\theta^\star - \theta). \tag{73}$$

The second corollary shows that β is an odd function (i.e., $\beta(x) = -\beta(-x)$) because of even symmetry of the ansatz functions $r_c^\star$ and ρ_{vis}. Finally, note that, although we do not prove anything about the derivative of the function β, stable landmark control requires $\beta'(0) > 0$.

Remark 6. *Let f and g be bounded continuous functions from S^1 to $\mathbb{R}$. The dot product of $f(x - y_0)$ and $g(x - z_0)$ over x can be identified with a another function $h : S^1 \to \mathbb{R}$ that depends on the difference $y_0 - z_0 \in S^1$ as shown below:*

$$h(y_0 - z_0) = \int_0^{2\pi} f(x - y_0)g(x - z_0)dx. \tag{74}$$

To see this, let us expand f and g into their Fourier series representation

$$f(x - y_0) = \sum_{n=-\infty}^{+\infty} f_n e^{-in(x-y_0)}$$
$$g(x - z_0) = \sum_{m=-\infty}^{+\infty} g_m e^{-im(x-z_0)},$$

where f_n and g_m denote Fourier series coefficient of the functions f and g, respectively. Using these expansions, we can compute the integral as follows:

$$\int_0^{2\pi} f(x - y_0)g(x - z_0) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \int_0^{2\pi} f_n g_m \, e^{-i(n(x-y_0)+m(x-z_0))} dx.$$

When $m \neq -n$, the integral vanishes. Thus, we can reduce the double summation to a single summation with $m = -n$ and simplify the integral as follows:

$$\begin{aligned}\int_0^{2\pi} f(x - y_0)g(x - z_0) &= \sum_{n=-\infty}^{+\infty} \int_0^{2\pi} f_n \, g_{-n} e^{-i(n(x-y_0)-n(x-z_0))} dx \\ &= \sum_{n=-\infty}^{+\infty} \int_0^{2\pi} f_n \, g_{-n} e^{-in(z_0-y_0)} dx \\ &= \sum_{n=-\infty}^{+\infty} 2\pi f_n \, g_{-n} e^{-in(z_0-y_0)}.\end{aligned}$$

The resulting summation can be identified as the Fourier series expansion of another function, say $h : \mathbb{S}^1 \to \mathbb{R}$, such that

$$h(z_0 - y_0) = \sum_{n=-\infty}^{+\infty} h_n e^{-in(z_0-y_0)} \tag{75}$$

where $h_n := 2\pi f_n g_{-n}$ denotes the Fourier series coefficients of the function h. △

Remark 7. *If bounded continuous functions $f : S^1 \to \mathbb{R}$ and $g : S^1 \to \mathbb{R}$ have even and odd symmetry about the origin, then their dot product, as defined in Equation (74) of the Remark 6, is an odd function, namely,*

$$\forall x \in S^1 \quad f(x) = f(-x), g(x) = -g(-x) \implies h(x) = -h(-x).$$

To see this, let us assume, without loss of generality, that the function f has odd symmetry about the origin, and g has even symmetry about the origin. Fourier series coefficients of an odd function is also odd, implying $f_n = -f_{-n}$, while those of an even function is even, implying $g_n = g{-}n$. Previously in Equation (75), we found Fourier series coefficients of the function h to be $h_n = 2\pi f_n g_{-n}$. Taken together, these coefficients are odd $h_n = -h_{-n}$ and proves that the function h is also odd. △

Remark 8. *Let $f : S^2 \to \mathbb{R}$ be a bounded continuous function. Integral of $f(x - y_0, x - z_0)$ over x can be identified with another function h that depends on the difference $y_0 - z_0 \in S^1$ as follows:*

$$h(y_0 - z_0) = \int_0^{2\pi} f(x - y_0, x - z_0)dx. \tag{76}$$

To see this, let us expand f into its 2D Fourier series representation

$$f(x - y_0, x - z_0) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} f_{(n,m)} e^{-i(n(x-y_0)+m(x-z_0))},$$

where $f_{(n,m)}$ denotes the Fourier series coefficient for the nth and m^{th} harmonic over the first and second arguments of f, respectively. Using this expansion, we can compute the integral of f as follows:

$$\int_0^{2\pi} f(x - y_0)g(x - z_0) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} \int_0^{2\pi} f_{(n,m)}\, e^{-i(n(x-y_0)+m(x-z_0))}dx.$$

When $m \neq -n$, the integral vanishes. Thus, we can reduce the double summation to a single summation with $m = -n$ and simplify the integral as follows:

$$\begin{aligned}
\int_0^{2\pi} f(x - y_0, x - z_0) &= \sum_{n=-\infty}^{+\infty} \int_0^{2\pi} f_{(n,-n)} e^{-i(n(x-y_0)-n(x-z_0))}dx \\
&= \sum_{n=-\infty}^{+\infty} \int_0^{2\pi} f_{(n,-n)} e^{-in(z_0-y_0)}dx \\
&= \sum_{n=-\infty}^{+\infty} 2\pi f_{(n,-n)} e^{-in(z_0-y_0)}.
\end{aligned}$$

The resulting summation can be identified as the Fourier series expansion of another function, say $h : \mathbb{S}^1 \to \mathbb{R}$, such that

$$h(z_0 - y_0) = \sum_{n=-\infty}^{+\infty} h_n e^{-in(z_0-y_0)},$$

where $h_n := 2\pi f_{(n,-n)}$ denotes the Fourier series coefficients of the function h. △

6.3 Control Theory Reveals an Algorithmic Stability Criteria of Gain Recalibration

6.3.1 Derivation of a reduced-order model for the gain update rule

Here, we derive a reduced-order model for the gain update rule that determines the temporal changes in the PI gain. As will be evident below, the model can capture a broad class of neural mechanisms, including but not limited to activity dependent plasticity like Hebbian plasticity.

We begin our derivation by expressing Equation (62), an analytical expression of the PI gain, in a

more compact form, namely,

$$k(\theta) = \int_0^{2\pi} K(\psi, \theta) d\psi, \tag{77}$$

where K denotes a function that quantifies the contribution of each location ψ in the neural space to the PI gain for a given position representation θ as follows:

$$K(\psi, \theta) = \frac{-b}{\tau_c \| \frac{\partial r_c^\star}{\partial \psi} \|^2} \frac{\partial^2 r_c^\star(\psi - \theta)}{\partial \psi^2} \Big(\mathcal{H}_{\text{cw}}(\psi, \theta) + \mathcal{H}_{\text{ccw}}(\psi, \theta) \Big).$$

In the remainder of this section, we refer to K as the distributed gain function.

Gain recalibration requires the distributed gain function K to vary over time. To address this requirement, we extend the domain of K to depend explicitly on time, i.e., $K(t, \psi, \theta)$. We are interested in a function, termed the "gain update rule", that relates the temporal change in K across the entire neural space to a change in average PI gain. To find this relation, we first take the time-derivative of the spatial average of Equation (77), which yields

$$\frac{dk_0(t)}{dt} = \underbrace{\frac{1}{2\pi} \int_0^{2\pi} \int_0^{2\pi} \frac{dK(t, \psi, \theta)}{dt} d\psi \, d\theta}_{= \text{gain update rule}} \tag{78}$$

We then search for an explicit, closed-form solution to this integral, assuming that the temporal changes in K are driven by two factors:

1. Activity dependent factors: We consider two distinct forms of it:

   (a) As the first type of activity-dependent plasticity, we consider that the function K at each location ψ is modified locally by the firing rates of a finite number of neurons at a specific location relative to ψ. An example would be Hebbian plasticity of the synapses between two neurons at the same location ψ or at another location a certain distance away from ψ. What variables are needed to model the firing rates?

   Recalling the ansatz solutions to the firing rates of neurons in the central and rotation rings in Equations (30) and (32), we know that firing rate of any neuron in the rotation and central rings can be computed using three variables : $\psi - \theta$, ψ, and v. As for the visual ring and velocity neurons, we assume that $\psi - \theta^\star$ and v are sufficient to compute firing rate of any neuron, similar to the ansatz solutions in Equations (7) and (28). Therefore, four variables are sufficient to compute the firing rate of any neuron in the ring attractor network: i)$\psi - \theta$, ii)$\psi - \theta^\star$, iii)v, iv)ψ.

   Using these variables, we can model the local activity-dependent changes in the distributed gain function K via

   $$\frac{dK(t, \psi, \theta)}{dt} = l_{a_1} \times G_{a_1}(\psi - \theta, \psi - \theta^\star, \psi, v), \tag{79}$$

   where l_{a_1} denotes the learning rate, and $G_{a_1} : T^3 \times \mathbb{R} \to \mathbb{R}$ denotes the learning rule. Note that, although we determined the arguments of the function G_{a_1} based on the calculation of the firing rates, we do not enforce a specific form for it to make our model as general as possible

rather than restricting it to a certain learning rule like Hebbian plasticity between two specific neurons. This way, the learning rule G_{a_1} can imitate a broad class of neural mechanisms, including (but not limited to) firing rate dependent plasticity like Hebbian plasticity.

(b) As the second type of activity dependent plasticity, we consider an extended version of the first type. In this extended case, the distributed gain function K at neural location ψ is changed both locally by the firing rates of a finite number of neurons at a particular location and globally by all the other neurons' firing rates across the entire neural space. Two motivating examples can be described as follows: 1) Hebbian plasticity in the recurrent connections of the central ring 2) Plasticity mediated by neuromodulators whose release is controlled based on population activity.

Recall that the four variables, namely, i)$\psi - \theta$, ii)$\psi - \theta^\star$, iii)v, iv)ψ, are sufficient to compute the firing rate of any neuron in the network. Since we now consider plasticity that depends on both local activity and global activity, we can model the activity-dependent changes in the distributed gain function K at ψ via an integral of a function that depends on the same four variables, from which the local firing rates can be computed, as well as additional three variables, i)$\phi - \theta$, ii)$\phi - \theta^\star$, iii)ϕ, from which the firing rate of other neurons can be computed. This integral is over ϕ and takes the form

$$\frac{dK(t,\psi,\theta)}{dt} = l_{a_2} \times \int_0^{2\pi} G_{a_2}(\psi - \theta, \psi - \theta^\star, \psi, v, \phi - \theta, \phi - \theta^\star,\, \phi) d\phi, \tag{80}$$

where l_{a_2} denotes the learning rate, and $G_{a_2} : T^3 \times \mathbb{R} \times T^3 \to \mathbb{R}$ denotes the learning rule. It is again worth recalling that the learning rule G_{a_2} can capture a broad class of neural mechanisms since we do not enforce a particular functional form for it.

2. History dependent effects: To make our model even more general, we take possible history dependent effects into account. However, to limit the mathematical possibilities of how these effects can act, we consider a specific form inspired from the synaptic weight decay well known in the context of Hebbian plasticity [83]. Thus, we model history dependency as a linear decay of the distributed gain function K toward a preconfigured value as follows:

$$\frac{dK(t,\psi,\theta)}{dt} = l_h \times \big(K^{\text{pre}}(\psi) - K(t,\psi,\theta - \psi)\big). \tag{81}$$

Here, l_h denotes the forgetting rate of K decay toward its preconfigured value $K^{\text{pre}} : S^1 \to \mathbb{R}$ defined at each location ψ in the neural space.

Combining Equations (79), (80), and (81) via superposition, we obtain a general model for the gain

update rule in Equation (78), namely,

$$\begin{aligned}\frac{dk_0(t)}{dt} =& \frac{l_{a_1}}{2\pi}\int_0^{2\pi}\int_0^{2\pi} G_{a_1}(\psi-\theta,\psi-\theta^\star,\psi,v)\,d\psi\,d\theta \\ &+\frac{l_{a_2}}{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi} G_{a_2}(\psi-\theta,\psi-\theta^\star,\psi,v,\phi-\theta,\phi-\theta^\star,\phi)\,d\phi\,d\psi\,d\theta \\ &+\frac{l_h}{2\pi}\int_0^{2\pi}\int_0^{2\pi} K^{\text{pre}}(\psi)-K(t,\psi,\theta)\,d\psi\,d\theta.\end{aligned} \tag{82}$$

This equation shows that the gain update rule within the ring attractor admits a complicated analytical model. However, as we prove in the next proposition (Prop. 9), this complicated model can be equivalently reduced to a much simpler form. Proof of the proposition makes use of the Remark 8 presented at the end of the previous section.

Remark 9. *The gain update rule in Equation (82) can be equivalently written as*

$$\frac{dk_0}{dt} = g_0(k_0,\tilde{\theta},v), \tag{83}$$

where $g_0 : \mathbb{R}\times\mathbb{R}\times S^1 \to \mathbb{R}$ denotes a function that depends on the current gain k_0, the animal's velocity v, and the positional error $\tilde{\theta} \triangleq \theta^\star - \theta$ in the ring attractor's representation relative to the visual drive's representation. To see this, we can analyze each line in Equation (82) and show that the integral in each line admits a closed-form, simpler expression as a solution:

Line 1. By substituting $\theta^\star = \theta + \tilde{\theta}$ (note that $\tilde{\theta}$ denotes the error as in the main text) and changing the order of integration, we can rewrite the double integral in the first line as

$$\frac{l_{a_1}}{2\pi}\int_0^{2\pi}\int_0^{2\pi} G_{a_1}(\psi-\theta,\psi-\theta^\star,\psi,v)\,d\psi\,d\theta = \frac{l_{a_1}}{2\pi}\int_0^{2\pi}\int_0^{2\pi} G_{a_1}(\psi-\theta,\psi-\theta-\tilde{\theta},\psi,v)\,d\theta\,d\psi.$$

According to Remark 8, the inner integral reduces to a function that depends on three terms: $\tilde{\theta}$, as the difference between the first and second arguments of G_{a_1}, ψ as the third argument of G_{a_1}, and v as the last argument of G_{a_1}. The outer integral further reduces to a function that depends on two terms $\tilde{\theta}$ and v by marginalizing over ψ. These reductions can be expressed mathematically as follows:

$$\frac{l_{a_1}}{2\pi}\int_0^{2\pi}\int_0^{2\pi} G_{a_1}(\psi-\theta,\psi-\theta-\tilde{\theta},\psi,v)\,d\theta\,d\psi = \frac{l_{a_1}}{2\pi}\int_0^{2\pi} g_{a_1}(\tilde{\theta},\psi,v)d\psi = \frac{l_{a_1}}{2\pi}g^0_{a_1}(\tilde{\theta},v). \tag{84}$$

Here, $g_{a_1} : T^2\times\mathbb{R}\to\mathbb{R}$ denotes the function that describes the result of the inner integral, and $g^0_{a_1} : S^1\times\mathbb{R}\to\mathbb{R}$ denotes the function that describes the final result of the double integral.

Line 2. Performing the inner most integration over ϕ, we can reduce the triple integral in the second line to a double integral, namely,

$$\frac{l_{a_2}}{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi} G_{a_2}(\psi-\theta,\psi-\theta^\star,\psi,v,\phi-\theta,\phi-\theta^\star,\phi)\,d\phi\,d\psi\,d\theta = \frac{l_{a_2}}{2\pi}\int_0^{2\pi}\int_0^{2\pi} f_{a_2}(\psi-\theta,\psi-\theta^\star,\psi,v,\theta,\theta^\star)\,d\psi\,d\theta,$$

where $f_{a_2} : T^3 \times \mathbb{R} \times T^2 \to \mathbb{R}$ denotes a function that describes the result of the inner most integral. Since the last two arguments of f_{a_2} can be trivially computed from its first three arguments, we can omit the first two arguments of f_{a_2}, which simplifies the above equation to

$$\frac{l_{a_2}}{2\pi}\int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi} G_{a_2}(\psi-\theta, \psi-\theta^\star, \psi, v, \phi-\theta, \phi-\theta^\star, \phi)\, d\phi\, d\psi\, d\theta = \frac{l_{a_2}}{2\pi}\int_0^{2\pi}\int_0^{2\pi} f_{a_2}(\psi-\theta, \psi-\theta^\star, \psi, v)\, d\psi\, d\theta,$$

We then follow the same steps as Line 1. Briefly, we substitute $\theta^\star = \theta + \tilde{\theta}$, change the order of integration in the double integral, and use Remark 8 to show that there exists a function $g^0_{a_2} : S^1 \times \mathbb{R} \to \mathbb{R}$ such that

$$\frac{l_{a_2}}{2\pi}\int_0^{2\pi}\int_0^{2\pi} f_{a_2}(\psi-\theta, \psi-\theta^\star, \psi, v)\, d\psi\, d\theta = \frac{l_{a_2}}{2\pi} g^0_{a_2}(\tilde{\theta}, v). \tag{85}$$

Line 3. Let us denote the integral of $K^{pre} : S^1 \to \mathbb{R}$, the PI gain's preconfigured value, by k_0^{pre}, namely,

$$k_0^{pre} \triangleq \int_0^{2\pi} K^{pre}(\psi)\, d\psi.$$

Taking the spatial average of Equation (77) over θ, we can compute k_0 as

$$k_0 \triangleq \frac{1}{2\pi}\int_0^{2\pi}\int_0^{2\pi} K(t, \psi, \theta) d\psi\, d\theta$$

Using these equations, we can simplify the double integral in third line as follows:

$$\frac{l_h}{2\pi}\int_0^{2\pi}\int_0^{2\pi} K^{pre}(\psi) - K(t, \psi, \theta)\, d\psi\, d\theta = l_h(k_0^{pre} - k_0(t)). \tag{86}$$

Combining these simplifications, we obtain a much simpler, closed-form analytical expression for the gain update rule in Equation (82) :

$$\frac{dk_0(t)}{dt} = \frac{l_{a_1}}{2\pi} g^0_{a_1}(\tilde{\theta}, v) + \frac{l_{a_2}}{2\pi} g^0_{a_2}(\tilde{\theta}, v) + l_h(k_0^{pre} - k_0(t))$$

To make the right hand-side even more general rather than enforcing the above form, we can represent it with a function, say g_0, that depends on $\tilde{\theta}$, v, and k_0. Consequently, we find that the gain update rule in Equation (82) reduces to

$$\frac{dk_0(t)}{dt} = g_0(k_0, \tilde{\theta}, v). \qquad \triangle$$

6.3.2 Proof of Proposition 1 (Necessary conditions for complete gain recalibration)

Here we provide a formal proof for Prop. 1, which identifies necessary conditions for convergence to the origin of the error dynamics in Equation 14, reproduced below:

$$\begin{aligned} \frac{d\tilde{\theta}}{dt} &= -\beta(\tilde{\theta}) + \tilde{k}v - k_{ac}(t, \theta)v \\ \frac{d\tilde{k}}{dt} &= -g_0(k_0, \tilde{\theta}, v), \end{aligned} \tag{87}$$

where $\tilde{k} \triangleq k^\star - k_0$ as defined in Equation (13). Before doing so, however, we first present a remark and an assumption that will be useful later in the proof.

Remark 10. *A nonzero velocity $v \neq 0$ is necessary for stable recalibration, i.e., if the animal is stationary (i.e., $\forall t, v(t) = 0$), then k_0 does not stably converge to $k^\star$.*

This necessity can be easily verified with a counter example. For $v = 0$, and using the fact that β has odd symmetry about the origin as previously shown in Section 6.2.5, it is easy to show that $\tilde{\theta} = 0$ and $\tilde{k} = \text{const} \neq 0$, is a (constant) solution to the error dynamics in Equation (14), and thus $\tilde{k}$ does not converge to 0 (in fact, it does not even change). Intuitively, this is because without moving, the animal can never get any information about its gain. △

Remark 11. *Let $B_\delta(0) \subset \mathbb{R}^2$ be an open ball of radius δ centered at the origin. If $\frac{d\tilde{\theta}}{dt} = 0$ for all $\tilde{k}$ when $\tilde{\theta} = 0$ and either one of the following two conditions are satisfied, then the origin $\tilde{k} = \tilde{\theta} = 0$ is not asymptotically stable:*

$$\frac{d\tilde{k}}{dt} \geq 0 \quad \textit{for all } (\tilde{k}, \tilde{\theta}) \in B_\delta(0) \cap \{(\tilde{k}, \tilde{\theta}) : \tilde{k} \geq 0, \tilde{\theta} > 0\} \quad (*)$$

or

$$\frac{d\tilde{k}}{dt} \leq 0 \quad \textit{for all } (\tilde{k}, \tilde{\theta}) \in B_\delta(0) \cap \{(\tilde{k}, \tilde{\theta}) : \tilde{k} \geq 0, \tilde{\theta} < 0\} \quad (**),$$

This result can be verified by considering trajectories beginning in the ball $B_\delta(0) \subset \mathbb{R}^2$ in $(\tilde{\theta}, \tilde{k})$ coordinates. Without loss of generality, let us assume the first condition (labeled with $(*)$) given above is true. Asmyptotic stability of the origin requires any point in the ball $B_\delta(0) \subset \mathbb{R}^2$ to converge to the origin. For trajectories beginning inside the first quadrant, $\tilde{k}$ is positive and constrained to increase further because of $\frac{d\tilde{k}}{dt} \geq 0$. To avoid $\tilde{k}$ increasing indefinitely, trajectories must cross the horizontal or vertical axis and go into a different quadrant. However, because $\frac{d\tilde{\theta}}{dt} = 0$ on the vertical axis and $\frac{d\tilde{k}}{dt} \geq 0$ on the horizontal axis, trajectories are confined to the first quadrant, where $\tilde{k}$ increases monotonically. Thus, the origin cannot be asymptotically stable. △

Assumption 5. *The AC magnitude, k_{ac}, of the PI gain function is upper bounded by the distance between the PI gain function's current average (denoted by $k_0(t)$) and its average at the steady state (denoted by $k_0^\infty \triangleq \lim_{t\to\infty} k_0(t)$), namely,*

$$\max_\theta |k_{ac}(t, \theta)| \leq |k_0(t) - k_0^\infty| \quad \forall t.$$

This implies, $k_{ac}(t, \theta) = 0$, when $\frac{d\tilde{k}}{dt} = 0$. △

Remark 12. *The stabilizing feedback $\beta : S^1 \to \mathbb{R}$ is an odd function that crosses zero at the origin (i.e., $\beta(0) = 0$) with a positive slope (i.e.,$\beta'(0) > 0$). Since this function is odd periodic, it attains a symmetric maximum and minimum, namely,*

$$\beta_{\max} = \max_\psi \beta(\psi) > 0$$

$$-\beta_{\max} = \min_\psi \beta(\psi) < 0.$$

Remark 13. *Let $\tilde{k}_0$ and $v_0 \neq 0$ be constants that satisfy $\tilde{k}_0 v_0 \in [-\beta_{\max}, \beta_{\max}]$. If $d\tilde{k}/dt = 0$ at $\tilde{k} = \tilde{k}_0 v_0$, then $k_{ac} = 0$ by Assumption (5), and there exists a solution to $d\tilde{\theta}/dt = -\beta(\tilde{\theta}) + \tilde{k}_0 v_0 = 0$ at $\tilde{\theta} = \beta^{-1}(\tilde{k}_0 v_0)$.*

This implies $(\tilde{k}_0, \beta^{-1}(\tilde{k}_0 v_0))$ *is an equilibrium point.* $\triangle$

Restatement of Proposition 1. *Let* $\mathcal{K}$ *be a bounded open interval in* $\mathbb{R}$*. If the PI gain's spatial average converges to the visual gain (i.e.,* $k_0^\infty = k^\star$*) for all* $k^\star$ *in* $\mathcal{K}$ *and for all nonzero velocities of the animal (i.e.,* $\forall t, v(t) \neq 0$*), then it requires*

$$g_0(k_0, 0, v) = 0, \quad \textit{for all } v \neq 0 \textit{ and } k_0, \tag{88}$$

as well as the sign requirement

$$\text{sign}[g_0(k_0, \tilde{\theta}, v)] = \text{sign}[\tilde{\theta}\, v], \textit{ for all } v \neq 0 \textit{ and for almost all } k_0 \textit{ (except a finite set of isolated points)}, \tag{89}$$

in a non-empty open interval of the positional error $\tilde{\theta} = 0$ *other than* $\tilde{\theta} = 0$*.*

Proof of Equation (88). We first make an observation: If k_0 converges to $k^\star$, then $\tilde{k}$ and k_{ac} converge to zero, resulting in exponentially stable $\tilde{\theta}$ dynamics, i.e., $\frac{d\tilde{\theta}}{dt} = -\beta(\tilde{\theta})$, because of $\beta'(0) > 0$. Thus, convergence of k_0 to $k^\star \in D$ implies that the origin $(\tilde{\theta}, \tilde{k}) = 0$ is asymptotically stable for any $k^\star \in D$.

Based on this observation, we can prove the proposition by showing that *Equations (88) and (89) are necessary for asymptotic stability of the origin* $(\tilde{\theta}, \tilde{k}) = 0$ *for any* $k^\star \in D$*.*

We first prove the necessity of Equation (88) for the asymptotic stability of the origin. This equation describes the behavior of g_0 at $\tilde{\theta} = 0$. Since the stability of the origin $\tilde{\theta} = \tilde{k} = 0$ implies that the origin is an equilibrium, namely that the error dynamics in Equation (87) must satisfy

$$\begin{aligned} \frac{d\tilde{\theta}}{dt} &= -\beta(0) = 0 \\ \frac{d\tilde{k}}{dt} &= -g_0(k_0, 0, v) = 0. \end{aligned}$$

The first equation is automatically satisfied since β function has odd symmetry about the origin. The second equation, however, requires

$$g_0(k_0, 0, v)\big|_{\tilde{\theta}=0} = 0 \quad \forall v \neq 0 \text{ and } \forall k_0.$$

Therefore, gain update rule g_0 satisfies the requirement (88) in the statement of the proposition. $\square$

Proof of Equation (89). We prove the necessity of Equation (89) for asymptotic stability of the origin through contradiction, namely, by showing the contrapositive statement:
The origin $(\tilde{\theta}, \tilde{k}) = 0$ *is not asymptotically stable if there exists a constant* $v = v_0 \neq 0$ *and a set* $\mathcal{K}$ *of* k_0*'s that is infinite or includes a limit point* $\bar{k}_0$ *such that*

$$\text{sign}[g_0(k_0, \tilde{\theta}, v_0)] \neq \text{sign}[\tilde{\theta}\, v_0] \quad \textit{for all } k_0 \in \mathcal{K} \subset D$$

in any non-empty open interval of $\tilde{\theta} = 0$*.*

Recalling that, by Bolzano-Weierstrass theorem, every bounded infinite set in $\mathbb{R}$ has at least one limit

point, we find it sufficient to prove this contrapositive statement for a set $\mathcal{K}$ of k_0's that has a limit point $\bar{k}_0$. Note that the limit point $\bar{k}_0$ is not necessarily included in $\mathcal{K}$, but, according to the definition of a limit point, in every neighborhood of $\bar{k}_0$, there exists a point in $\mathcal{K}$ other than $\bar{k}_0$.

For each k_0^* in $\mathcal{K}$, the following is true based on the italicized contrapositive statement above:

$$\forall \alpha > 0, \exists\, \tilde{\theta}^* \in [-\alpha, 0) \cup (0, \alpha] \quad \text{such that} \quad \text{sign}[g_0(k_0^*, \tilde{\theta}^*, v_0)] \neq \text{sign}[\tilde{\theta}^*\, v_0]. \tag{90}$$

Without loss of generality, let us assume $v_0 > 0$. Under this assumption, the inequality in (90) implies either one of the following, mutually exclusive cases:

- For some k_0^* in $\mathcal{K}$, there exists a point in every open interval of $\tilde{\theta} = 0$ other than $\tilde{\theta} = 0$, where g_0 is zero, namely,

$$\forall \alpha > 0, \exists\, \tilde{\theta}^* \in [-\alpha, 0) \cup (0, \alpha] \quad \text{such that} \quad g_0(k_0^*, \tilde{\theta}^*, v_0) = 0. \tag{91}$$

  This implies that the zero set of the restriction of the function g_0 to $k_0 = k_0^*$ and $v = v_0$, i.e., $g_0(k_0^*, \cdot, v_0) : S^1 \to \mathbb{R}$, has a limit point at $\tilde{\theta} = 0$. By analyticity of g_0, we infer that this restriction is a zero function, namely,

$$g_0(k_0^*, \tilde{\theta}, v_0) = 0 \quad \forall \tilde{\theta} \in \mathbb{R}. \tag{92}$$

- If Equation (91) is not true for some k_0^* in $\mathcal{K}$, then there exists an open interval of $\tilde{\theta} = 0$ where g_0 is not zero anywhere but $\tilde{\theta} = 0$. Taken together with Equation (90), this implies

$$\begin{aligned} \text{either} \quad & \exists\, A > 0, \forall \tilde{\theta}^* \in [-A, 0) \quad \text{such that} \quad g_0(k_0^*, \tilde{\theta}^*, v_0) > 0 \\ \text{or} \quad & \exists\, A > 0, \forall \tilde{\theta}^* \in (0, A] \quad \text{such that} \quad g_0(k_0^*, \tilde{\theta}^*, v_0) < 0. \end{aligned} \tag{93}$$

Let us define

$$\mathcal{K}_+ = \{k_0^* \in \mathcal{K},\ \ \forall \tilde{\theta^*} \in \mathbb{R}\ g_0(k_0^*, \tilde{\theta}^*, v_0) = 0,\ \ \forall \tilde{\theta}^* \in (0, A]\ g_0(k_0^*, \tilde{\theta}^*, v_0) < 0\}$$

as the set of points in $\mathcal{K}$ that satisfy either Equation (92) or the second line in Equation (93). Adopting a similar notation, we can define the set of points in $\mathcal{K}$ that satisfy either Equation (92) or the first line in Equation (93) as

$$\mathcal{K}_- = \{k_0^* \in \mathcal{K},\ \ \forall \tilde{\theta^*} \in \mathbb{R}\ g_0(k_0^*, \tilde{\theta}^*, v_0) = 0,\ \ \forall \tilde{\theta}^* \in [-A, 0)\ g_0(k_0^*, \tilde{\theta}^*, v_0) > 0\}.$$

Note that the union of these sets cover the entire $\mathcal{K}$, namely, $\mathcal{K}_+ \cup \mathcal{K}_- = \mathcal{K}$. Moreover, a limit point $\bar{k}_0$ of the set $\mathcal{K}$ is also a limit point of $\mathcal{K}_+$ or $\mathcal{K}_-$. To see this, suppose $\bar{k}_0$ is not a limit point $\mathcal{K}_-$. Thus, there exists $\epsilon^\star > 0$ such that the interval $I_{\epsilon^\star} = (\bar{k}_0 - \epsilon^\star, \bar{k}_0) \cup (\bar{k}_0, \bar{k}_0 + \epsilon^\star)$ does not include any point in $\mathcal{K}_-$, i.e., $I_\epsilon \cap \mathcal{K}_- = \emptyset$. Since $\bar{k}_0$ is a limit point of $\mathcal{K}$ and $\mathcal{K} = \mathcal{K}_+ \cup \mathcal{K}_-$, for every ϵ that is $0 < \epsilon < \epsilon^\star$, the interval I_ϵ includes an element of $\mathcal{K}_+$, i.e., $(\bar{k}_0 - \epsilon^\star, \bar{k}_0) \cup (\bar{k}_0, \bar{k}_0 + \epsilon^\star) \cap \mathcal{K}_+ \neq \emptyset$, implying that $\bar{k}_0$ is a limit point of $\mathcal{K}_+$.

Without loss of generality, let us assume $\bar{k}_0$ is a limit point of the set $\mathcal{K}_+$. Recall that a limit point of a set is not necessarily included in the set. We show that, regardless of whether the limit point $\bar{k}_0$ is

included in $\mathcal{K}_+$ or not, the origin $(\tilde{\theta}, \tilde{k}) = 0$ is not asymptotically stable for some visual gain $k^\star \in D$. To this end, consider an open ball $B_{p_0}(r)$ centered at $p_0 = (0, \bar{k}_0)$ with radius of radius $r < \beta_{\max}$. For a given r, pick $\delta_r > 0$ that solves $\delta_r^2 + \left(\beta^{-1}(\delta_r v_0)\right)^2 = r^2$ (the solution exists since $r < \beta_{\max}$). This δ_r ensures that the interval $I_r := [\bar{k}_0 - \delta_r, \bar{k}_0) \cup (\bar{k}_0, \bar{k}_0 + \delta_r]$ is contained in the ball $B_{p_0}(r)$. For every $r > 0$, this non-empty interval exists. Since $\bar{k}_0$ is a limit point of $\mathcal{K}_+$, there exists a point $k_0^\star$ in $I_r \cap \mathcal{K}_+$ that satisfies either Equation (92) or the second row of Equation (93) for every $r > 0$. Below, we investigate each case and show that the origin $(\tilde{\theta}, \tilde{k}) = 0$ is not asymptotically stable for some $k^\star \in D$.

Case 1. Consider first that $k_0^\star$ satisfies Equation (92). Let us choose the visual gain $k^\star = \bar{k}_0$. Equation (92) implies that the change in the gain error $\tilde{k}$ is zero, namely,

$$\frac{d\tilde{k}}{dt} = g_0(k_0^\star, \tilde{\theta}, v_0) = 0 \quad \forall \tilde{\theta} \in \mathbb{R}.$$

When this equation is satisfied, $k_{\text{ac}} = 0$ by Assumption 5. It further follows from Remark 13 that the point $(\tilde{\theta}, \tilde{k}) = (\beta^{-1}\big((\bar{k}_0 - k_0^\star)v_0\big), (\bar{k}_0 - k_0^\star)) \neq 0$ is an equilibrium. Thus, the origin cannot be asymptotically stable as there exists another equilibrium point other than the origin in any open ball of radius $r > 0$.

Case 2. Consider next that, for some $r^\star > 0$, there is no $k_0^\star \in I_r$ satisfying Equation (92), i.e., Case 1 is not true. Thus, for all $r \in (0, r^\star]$, there exists $k_0^\star \in I_r$ satisfying the second row of Equation (93) such that

$$\exists A^* > 0, \quad \frac{dk_0}{dt} = g_0(k_0^\star, \tilde{\theta}, v_0) < 0, \quad \forall \tilde{\theta} \in (0, A^\star].$$

On the other hand, depending on whether the limit point $\bar{k}_0$ is inside the set $\mathcal{K}_+$ or not, the change in the gain k_0 can be positive or negative, namely, there exists $\bar{A} > 0$ such that

$$\begin{aligned} \text{either } \bar{k}_0 \notin \mathcal{K}_+ &\implies \frac{dk_0}{dt} = g_0(\bar{k}_0, \tilde{\theta}, v_0) > 0, \quad \forall \tilde{\theta} \in (0, \bar{A}] \\ \text{or} \quad \bar{k}_0 \in \mathcal{K}_+ &\implies \frac{dk_0}{dt} = g_0(\bar{k}_0, \tilde{\theta}, v_0) leq 0, \quad \forall \tilde{\theta} \in (0, \bar{A}] \end{aligned}$$

Subcase 1. In the subcase of $\bar{k}_0 \notin \mathcal{K}_+$, continuity of g_0 implies that there exists an intermediate point k_0^m in the open interval connecting $\bar{k}_0$ and $k_0^\star$ such that

$$\frac{dk_0}{dt} = g_0(k_0^m, \tilde{\theta}, v_0) = 0, \quad \forall \tilde{\theta} \in (0, A].$$

where $A = \min(A^\star, \bar{A})$. Analyticity of g_0 further requires the function to be zero for all values of $\tilde{\theta}$ at $k_0 = k_0^m$, namely,

$$\frac{dk_0}{dt} = g_0(k_0^m, \tilde{\theta}, v_0) = 0, \quad \forall \tilde{\theta} \in \mathbb{R}.$$

This raises a contradiction, as we assumed that Case 1 is not true. Thus, this subcase is not possible.

Subcase 2. Consider the remaining subcase $\bar{k}_0 \in \mathcal{K}_+$. For $v = v_0$, this case implies the function g_0 is nonpositive everywhere in a half-closed region, namely, $(\tilde{\theta}, k_0) \in Q_k \triangleq \big(0, \min(A^\star, \bar{A})\big] \times I_k$ where I_k is the open interval connecting $\bar{k}_0$ and $k_0^\star$. Choose a visual gain in this open interval,

i.e., $k^\star \in I_k$. Since $\tilde{k} = k^\star - k_0$, this implies

$$\forall(\tilde{\theta}, k_0) \in Q_k, \quad \frac{d\tilde{k}}{dt} = -g_0(k_0, \tilde{\theta}, v_0) \geq 0.$$

One can find an open ball $B_{\delta^*}(0)$ that is entirely contained in Q_k in the right-half plane. Since $d\tilde{\theta}/dt \geq 0$ in the right-half of this ball and $d\tilde{\theta}/dt = 0$ on the vertical axis $\tilde{\theta} = 0$ by Equation (88), we can employ Remark 11 to conclude the origin is not asymptotically stable.

Therefore, gain update rule g_0 satisfies the requirement (89) in the statement of the proposition. □

6.3.3 Extension of Proposition 1 to Partial Recalibration

Corollary 1. *Let the animal's velocity be constant $v = v_0 \neq 0$ and $\mathcal{K}$ be a bounded open interval in $\mathbb{R}$. If the PI gain's spatial average converges to $k_0^\infty \in \mathcal{K}$ for a given $k^\star$, then the following statements are true.*

1. *There is an an equilibrium point $(\tilde{\theta}, \tilde{k}) = (\tilde{\theta}^\infty, \tilde{k}^\infty)$ that satisfies*

$$\begin{aligned} \tilde{\theta}^\infty &\triangleq \beta^{-1}((k^\star - k_0^\infty)v_0) \\ \tilde{k}^\infty &\triangleq k^\star - k_0^\infty. \end{aligned} \tag{94}$$

2. *Assume $\beta'(\tilde{\theta}^\infty) > 0$. The gain update rule g_0 satisfies*

$$g_0(k_0, \tilde{\theta}^\infty, v_0) = 0, \quad \textit{for all } k_0, \tag{95}$$

as well as the sign requirement

$$\operatorname{sign}[g_0(k_0, \tilde{\theta}, v_0)] = \operatorname{sign}\left[(\tilde{\theta} - \tilde{\theta}^\infty)\, v_0\right], \quad \textit{for almost all } k_0 \textit{ (except a finite set of isolated points)}, \tag{96}$$

in a non-empty open interval of the positional error $\tilde{\theta} = \tilde{\theta}^\infty$ other than $\tilde{\theta} = \tilde{\theta}^\infty$.

Proof of Statement 1. Convergence of k_0 to k_0^∞ implies that $\tilde{k}$ converge to

$$\tilde{k}^\infty \triangleq k^\star - k_0^\infty.$$

After convergence, $\tilde{k}$ dynamics satisfy

$$\frac{d\tilde{k}}{dt} = -g_0(k_0, \tilde{\theta}, v_0)\big|_{k_0 = k_0^\infty} = 0.$$

On this nullcline of $\tilde{k}$, the PI gain's AC component k_{ac} is zero by Assumption 5. Substituting $k_{ac} = 0$ into $\tilde{\theta}$ dynamics in (87), we obtain $\tilde{\theta}$ dynamics on the nullcline of $\tilde{k}$ as

$$\frac{d\tilde{\theta}}{dt} = -\beta(\tilde{\theta}) + \tilde{k}^\infty v_0.$$

Assuming $|\tilde{k}^\infty v_0| \leq \max_{\tilde{\theta}} \beta(\tilde{\theta})$, we can solve $\frac{d\tilde{\theta}}{dt} = 0$ for $\tilde{\theta}$, namely,

$$\tilde{\theta}^\infty = \beta^{-1}(\tilde{k}^\infty v_0).$$

Therefore, $(\tilde{\theta}, \tilde{k}) = (\tilde{\theta}^\infty, \tilde{k}^\infty)$ is an equilibrium point of the dynamics in Equation (87). □

Proof of Statement 2. Before proving this statement, we first perform the change of coordinates $z_\theta \triangleq \tilde{\theta} - \tilde{\theta}^\infty$, $z_k \triangleq \tilde{k} - \tilde{k}^\infty$. Under this coordinate change, the point $(\tilde{\theta}^\infty, \tilde{k}^\infty)$ in $(\tilde{\theta}, \tilde{k})$ coordinates maps to the origin $(0, 0)$ in (z_θ, z_k) coordinates. By Statement 1, the origin $(z_\theta, z_k) = 0$ is an equilibrium point. Also, note that $\frac{dz_\theta}{dt} = \frac{d\tilde{\theta}}{dt}$, $\frac{dz_k}{dt} = \frac{d\tilde{k}}{dt}$ by the chain rule.

In these new coordinates, wee can reformulate the dynamics in Equation (87) as

$$\begin{aligned} \frac{dz_\theta}{dt} &= -\mathcal{B}(z_\theta) + (z_k - k_{ac}(t, \theta))v_0 \\ \frac{dz_k}{dt} &= -h_0(k_0, z_\theta, v_0). \end{aligned} \tag{97}$$

where

$$\begin{aligned} \mathcal{B}(z_\theta) &\triangleq \beta(z_\theta + \tilde{\theta}^\infty) - \tilde{k}^\infty v_0 = \beta(\tilde{\theta}) - \tilde{k}^\infty v_0 \\ h_0(k_0, z_\theta, v_0) &\triangleq g_0(k_0, z_\theta + \tilde{\theta}^\infty, v_0) = g_0(k_0, \tilde{\theta}, v_0). \end{aligned} \tag{98}$$

Recall that the PI gain's AC component k_{ac} in this equation satisfy the inequality in Assumption 5. By adding subtracting $k^\star$ to the right-hand side of that inequality, we can reformulate it in the new coordinates as

$$\max_\theta |k_{ac}(t, \theta)| \leq |z_k(t)| \quad \forall t. \tag{99}$$

Furthermore, definition of $\mathcal{B}$ and the fact $\beta'(\tilde{\theta}^\infty) > 0$ imply that the function $\mathcal{B}$ satisfy $\mathcal{B}(0) = 0$ and $\mathcal{B}'(0) > 0$, the same properties possessed by the function β (i.e., $\beta(0) = 0$ and $\beta'(0) > 0$).

Based on these reformulations, we can rephrase the second statement in the corollary as follows: *Let $\mathcal{K}_z$ be a bounded open interval in $\mathbb{R}$ and $\mathcal{B}'(0) > 0$. If z_k converges to the origin for any $k_0^\infty \in \mathcal{K}_z$, then the function h_0 satisfies*

$$h_0(k_0, z_\theta, v_0) = 0, \quad \textit{for all } k_0, \tag{100}$$

as well as the sign requirement

$$\text{sign}[h_0(k_0, z_\theta, v_0)] = \text{sign}\big[z_\theta\, v_0\big], \quad \textit{for almost all } k_0 \textit{ (except a finite set of isolated points)}, \tag{101}$$

in a non-empty open interval of $z_\theta = 0$ other than $z_\theta = 0$.

Note that (z_θ, z_k) dynamics in Equation (97) are isomorphic to $(\tilde{\theta}, \tilde{k})$ dynamics in Equation (87). Moreover, the rephrased statement is a special case of the statement in Prop. 1 for $v = v_0 \neq 0$ and $k^\star = k_0^\infty$. Therefore, using Prop. 1, we prove the above rephrased statement of the corollary. □

6.3.4 Proof of Proposition 2 (Sufficient Conditions for Gain Recalibration)

At the end of this section we provide a formal proof for Prop. 2, which identifies sufficient conditions for convergence of the error dynamics to constant, steady-state values. Throughout the proof, we make use of the following remarks.

Remark 14. *As stated previously in Equation (99), the term k_{ac} is bounded by the absolute value of z_k. We can model this relation more directly by including z_k into the arguments of the function k_{ac}, namely $k_{ac}(t, \theta, z_k)$.* △

Remark 15. *Suppose visual gain $k^\star$ and the animal's velocity v_0 are given. In this case, $\theta^\star = k^\star v_0 t$ is determined completely by the time t. Since one can compute $\theta = k^\star v_0 t - \tilde{\theta}$, the gain's spatial variation can be modeled, without loss of any generality, as a function that depends on $\tilde{\theta}$ instead of θ, i.e., $k_{ac}(t, \tilde{\theta}, z_k)$.*

Moreover, since the gain's spatial variation is zero-mean and periodic over $\theta^\star$ for a given $\tilde{\theta}$, the function $k_{ac}(t, \tilde{\theta}, z_k)$ is zero-mean and periodic over time t with period $T = 2\pi/(k^\star v_0)$, namely,

$$\forall t, \quad \begin{aligned} &\int_0^T k_{ac}(t, \tilde{\theta}, z_k) dt = 0 \\ &k_{ac}(t, \tilde{\theta}, z_k) = k_{ac}(t + T, \tilde{\theta}, z_k). \end{aligned} \qquad \triangle$$

Finally, since $z_\theta \triangleq \tilde{\theta} - \tilde{\theta}^\infty$, the second argument of this function can be considered as a variable that depends on z_θ for a fixed $\tilde{\theta}^\infty$ instead of $\tilde{\theta}$, namely, $k_{ac}(t, z_\theta, z_k)$. $\triangle$

Restatement of Proposition 2. *For a given velocity $v = v_0 \neq 0$ and visual gain $k^\star$, if the gain update rule satisfies the slope condition*

$$\left.\frac{\partial g_0(k_0, \tilde{\theta}, v)}{\partial(\tilde{\theta} v)}\right|_{v=v_0} = \frac{\partial g_0(k_0, \tilde{\theta}, v_0)}{\partial \tilde{\theta}} \frac{1}{v_0} > 0 \tag{102}$$

everywhere in the solution set of $g_0(k_0, \tilde{\theta}, v_0) = 0$, then PI gain k_0 converges a steady-state value.

Proof. Let us assume, without loss of generality, that the steady-state value of k_0 is k_0^∞, corresponding to $\tilde{k}^\infty \triangleq k^\star - k_0^\infty$ as the steady-state value of $\tilde{k}$. A necessary condition for convergence to these steady-state values were previously given in Corollary 1. According to this corollary, the function g_0 must satisfy

$$\operatorname{sign}\big[g_0(k_0, \tilde{\theta}, v_0)\big] = \operatorname{sign}\big[(\tilde{\theta} - \tilde{\theta}^\infty)\, v_0\big] \quad \forall k_0,$$

where $\tilde{\theta}^\infty \triangleq \beta^{-1}(\tilde{k}^\infty v_0)$ denotes the steady-state value of $\tilde{\theta}$. This requirement implies that $\tilde{\theta} = \tilde{\theta}^\infty$ constitutes the solution set of $g_0(k_0, \tilde{\theta}, v_0) = 0$ for all k_0, namely,

$$g_0(k_0, \tilde{\theta}, v_0)\big|_{\tilde{\theta}=\tilde{\theta}^\infty} = 0 \quad \forall k_0. \tag{103}$$

With this in mind, we now proceed with an investigation of Equation (102), describing the slope condition, to prove its sufficiency for asymptotic stability of the point $(\tilde{\theta}^\infty, \tilde{k}^\infty)$. For this investigation, we take a multi-step approach.

1. We first perform a change of coordinates $z_\theta \triangleq \tilde{\theta} - \tilde{\theta}^\infty$, $z_k \triangleq \tilde{k} - \tilde{k}^\infty$, which leads to the same non-autonomous system dynamics as in Equation (97). This change of variables maps the point $(\tilde{\theta}^\infty, \tilde{k}^\infty)$, representing steady-state values in the $(\tilde{\theta}, \tilde{k})$ coordinates, to the origin in the (z_θ, z_k) coordinates.

   Under this change of coordinates, k_0, the first argument of the function h_0, can be expressed as $k_0 = k^\star - \tilde{k}^\infty - z_k$. Because $\tilde{k}^\infty$ is fixed as a coordinate of the equilibrium point and because visual gain $k^\star$ is given, we can re-organize the arguments of the function h_0 without loss of generality as

$h_0(z_k, z_\theta, v_0)$. In this new form, the slope condition in Equation (102) can be written as

$$\left.\frac{\partial h_0(z_k, z_\theta, v_0)}{\partial z_\theta}\right|_{z_\theta=0} \frac{1}{v_0} > 0 \quad \forall z_k, \tag{104}$$

where we used $\frac{\partial z_\theta}{\partial} z_\theta = 0$ identifies the solution set of $h_0(z_k, z_\theta, v_0) = 0$ in the new form (see Equation (103) for the respective solution in the old form). This solution set implies

$$\left.\frac{\partial h_0(z_k, z_\theta, v_0)}{\partial z_k}\right|_{z_\theta=0} = 0, \quad \forall z_k, v_0. \tag{105}$$

2. In the next step, we work with the term k_{ac}, which denotes the AC component of the PI gain. Recall that, by Equation (99), when $z_k = 0$, we have $k_{ac} = 0$ for all t, z_θ, which implies

$$\left.\frac{\partial k_{ac}(t, z_\theta, z_k)}{\partial \theta^\star}\right|_{z_k=0} = \left.\frac{\partial k_{ac}(t, z_\theta, z_k)}{\partial z_\theta}\right|_{z_k=0} = 0, \quad \forall t, z_\theta. \tag{106}$$

3. Finally, on the basis of our findings in the previous steps, we will show that the system dynamics in Equation (97) is stable as claimed in the proposition. To this end, we first modify Equation (97) by replacing the function g_0 with the function h_0 and by updating the functional form of k_{ac} as described in Remark 15, which yields

$$\begin{aligned} \frac{dz_\theta}{dt} &= -\mathcal{B}(z_\theta) + \big(z_k - k_{ac}(t, z_\theta, z_k)\big) v_0 \\ \frac{dz_k}{dt} &= -h_0(z_k, z_\theta, v_0). \end{aligned} \tag{107}$$

For the stability assessment, we employ the Averaging Theorem from dynamical systems theory that exploits the fact that k_{ac} is a periodic function over time with zero-mean and period $T = 2\pi/(k^\star v_0)$. According to this theorem, if the autonomous system without the periodic perturbation k_{ac} is exponentially stable at the origin, then Equation (107), the nonautonomous system with periodic perturbation k_{ac}, attains an exponentially stable periodic orbit around the origin. Moreover, as stated in [84], that premise makes the origin an exponentially stable equilibrium point of the perturbed, nonautonomous system in Equation (107) due to $k_{ac}(t, z_\theta, z_k)|_{(z_\theta, z_k)=0} = 0$. Therefore, to prove the proposition, it is sufficient to show that the origin is an exponentially stable equilibrium point of the autonomous system without the periodic perturbation k_{ac} when Equation (104) is true. (See Theorem 10.3 in [84] and Theorem 4.1.1 in [85] for more formal, detailed statements of the Averaging Theorem.)

For this specific task, we first ignore the k_{ac} in Equation (107), which leads to the autonomous system without any periodic perturbation, namely

$$\begin{aligned} \frac{dz_\theta}{dt} &= -\mathcal{B}(z_\theta) + z_k v_0 \\ \frac{dz_k}{dt} &= -h_0(z_k, z_\theta, v_0). \end{aligned} \tag{108}$$

To assess stability of this autonomous system, we linearize it about the origin $(z_\theta, z_k) = 0$, leading

to the following linear time-invariant system:

$$\frac{dz_\theta}{dt} = -\mathcal{B}'(0)z_\theta + v_0 z_k$$
$$\frac{dz_k}{dt} = -\left(\frac{\partial h_0(z_k, z_\theta, v_0)}{\partial z_\theta}\bigg|_{(z_\theta, z_k)=0}\right) z_\theta - \left(\frac{\partial h_0(z_k, z_\theta, v_0)}{\partial z_k}\bigg|_{(z_\theta, z_k)=0}\right) z_k.$$

Substituting Equation (105) into this system, we obtain

$$\frac{d}{dt}\begin{bmatrix} z_\theta \\ z_k \end{bmatrix} = \begin{bmatrix} -\mathcal{B}'(0) & v_0 \\ -\frac{\partial h_0(z_k, z_\theta, v_0)}{\partial z_\theta}\bigg|_{(z_\theta, z_k)=0} & 0 \end{bmatrix} \begin{bmatrix} z_\theta \\ z_k \end{bmatrix}. \tag{109}$$

Both eigenvalues of this system are negative, namely,

$$\lambda_{1,2} = -\frac{\mathcal{B}'(0)}{2} \pm \sqrt{\frac{\mathcal{B}'(0)^2}{4} - v_0 \frac{\partial h_0(z_k, z_\theta, v_0)}{\partial z_\theta}\bigg|_{(z_\theta, z_k)=0}}.$$

The inequality in Equation (104) ensures that these eigenvalues are negative, i.e., $\lambda_{1,2} < 0$, rendering the origin an exponentially stable equilibrium point of the linear system in Equation (109). Since this linear system was obtained by linearizing Equation (108) about the origin, it follows from Hartman-Grobman theorem [84] that the origin is an exponentially stable equilibrium point of the autonomous system in Equation (108). Moreover, by using the aferomentioned Averaging theorem, we can conclude that the origin is also an exponentially stable equilibrium point of the nonautonomous system in Equation (107). This proves that the PI gain k_0 converges to a steady-state value, namely k_0^∞, as claimed in the proposition. □

6.4 Mechanistic Constraints Reveal Instrumental Role of Positional Error Codes

In this section, we analyze analytical expressions of the average PI gain k_0 to identify the mechanistic constraints for the gain recalibration. We derive two alternative expressions by averaging Equations (62) and (64) over θ, which yields

$$k_0 = \frac{-b}{2\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi}\int_0^{2\pi} \frac{d^2 r_c^\star(\psi - \theta)}{dz^2} \left(\mathcal{H}_{\text{cw}}(\psi, \theta) + \mathcal{H}_{\text{ccw}}(\psi, \theta)\right) d\psi d\theta \tag{110}$$

$$k_0 = \frac{-b}{2\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi}\int_{\theta-\Omega_{\text{rot}}^l(\theta)}^{\theta+\Omega_{\text{rot}}^r(\theta)} \frac{d^2 r_c^\star(\psi - \theta)}{dz^2} \left(\mathcal{H}_{\text{cw}}(\psi, \theta) + \mathcal{H}_{\text{ccw}}(\psi, \theta)\right) d\psi d\theta, \tag{111}$$

respectively. Also recall that H_{cw}, H_{ccw} the terms must satisfy the symmetry constraint in Remark 3. For the expressions above, this implies

$$k_0 = \frac{-b}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi}\int_0^{2\pi} \frac{d^2 r_c^\star(\psi - \theta)}{dz^2} \mathcal{H}_{\text{cw}}(\psi, \theta) d\psi d\theta$$
$$k_0 = \frac{-b}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi}\int_0^{2\pi} \frac{d^2 r_c^\star(\psi - \theta)}{dz^2} \mathcal{H}_{\text{ccw}}(\psi, \theta) d\psi d\theta$$

and

$$k_0 = \frac{-b}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_{\theta-\Omega_{\text{rot}}^l(\theta)}^{\theta+\Omega_{\text{rot}}^r(\theta)} \frac{d^2 r_c^\star(\psi-\theta)}{dz^2} \, \mathcal{H}_{\text{cw}}(\psi,\theta) d\psi d\theta$$

$$k_0 = \frac{-b}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_{\theta-\Omega_{\text{rot}}^l(\theta)}^{\theta+\Omega_{\text{rot}}^r(\theta)} \frac{d^2 r_c^\star(\psi-\theta)}{dz^2} \, \mathcal{H}_{\text{ccw}}(\psi,\theta) d\psi d\theta,$$

respectively. Using the closed-form analytical expressions of the terms H_{cw}, H_{ccw} in Equation (49), we can rewrite these expressions more explicitly as

$$\begin{aligned} k_0 &= \frac{-b\,\alpha_{\text{cw}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_0^{2\pi} \frac{d^2 r_c^\star(\psi-\theta)}{dz^2} \, W_{\text{cw-c}}(\psi) W_{\text{v-cw}}(\psi) \text{sign}[r_{\text{cw}}^\star(\psi,\theta,0)] d\psi d\theta \\ k_0 &= \frac{-b\,\alpha_{\text{ccw}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_0^{2\pi} \frac{d^2 r_c^\star(\psi-\theta)}{dz^2} \, W_{\text{ccw-c}}(\psi) W_{\text{v-ccw}}(\psi) \text{sign}[r_{\text{ccw}}^\star(\psi,\theta,0)] d\psi d\theta \end{aligned} \tag{112}$$

and

$$\begin{aligned} k_0 &= \frac{-b\,\alpha_{\text{cw}}}{2\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_{\theta-\Omega_{\text{rot}}^l(\theta)}^{\theta+\Omega_{\text{rot}}^r(\theta)} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} W_{\text{cw-c}}(\psi) W_{\text{v-cw}}(\psi) d\psi d\theta \\ k_0 &= \frac{-b\,\alpha_{\text{ccw}}}{2\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_{\theta-\Omega_{\text{rot}}^l(\theta)}^{\theta+\Omega_{\text{rot}}^r(\theta)} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} W_{\text{ccw-c}}(\psi) W_{\text{v-ccw}}(\psi) d\psi d\theta \end{aligned} \tag{113}$$

6.4.1 Temporal changes in the synaptic weights of connections between the velocity neurons and the rotation rings

In this subsection, we examine how gain recalibration can be driven by temporal changes in the functions $W_{\text{v-cw}}$ and $W_{\text{v-ccw}}$, describing the synaptic weight of connections between the velocity neurons and the rotation rings. To consider the temporal effects of these functions on k_0, we extend Equation (112) to be time-dependent as given below:

$$\begin{aligned} k_0(t) &= \frac{-b\,\alpha_{\text{cw}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_0^{2\pi} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} W_{\text{cw-c}}(\psi) W_{\text{v-cw}}(t,\psi) \text{sign}[r_{\text{cw}}^\star(\psi,\theta,0)] d\psi d\theta \\ k_0(t) &= \frac{-b\,\alpha_{\text{ccw}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_0^{2\pi} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} W_{\text{ccw-c}}(\psi) W_{\text{v-ccw}}(t,\psi) \text{sign}[r_{\text{ccw}}^\star(\psi,\theta,0)] d\psi \, d\theta \end{aligned} \tag{114}$$

On the other hand, we assume that the remaining parameters of our model are hardwired (i.e., time-invariant), initialized with values inspired by the classical ring attractor models such that

$$\begin{aligned} \alpha_{\text{cw}} &= \alpha_{\text{ccw}} = \alpha_{\text{rot}} \\ W_{\text{cw-c}}(\psi) &= W_{\text{ccw-c}}(\psi) = W_{\text{rot-c}} \\ W_{\text{c-cw}}(\psi) &= W_{\text{c-ccw}}(\psi) = W_{\text{c-rot}}. \end{aligned} \tag{115}$$

Since the integrands in Equation (114) are finite and attain their extreme values over a compact space, their double integrals in these equations lead to finite constants. This lets us change the order of

integration. Doing so and re-arranging the terms, we obtain

$$
\begin{aligned}
k_0(t) &= \frac{-b\alpha_{\text{rot}} W_{\text{rot-c}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} W_{\text{v-cw}}(t,\psi) R_{\text{cw}}(\psi) d\psi \\
k_0(t) &= \frac{-b\alpha_{\text{rot}} W_{\text{rot-c}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} W_{\text{v-ccw}}(t,\psi) R_{\text{ccw}}(\psi) d\psi,
\end{aligned}
\tag{116}
$$

where the functional terms R_{cw} and R_{ccw} denote

$$
\begin{aligned}
R_{\text{cw}}(\psi) &= \int_0^{2\pi} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} \text{sign}[r_{\text{cw}}^\star(\psi,\theta,0)]\, d\theta \\
R_{\text{ccw}}(\psi) &= \int_0^{2\pi} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} \text{sign}[r_{\text{ccw}}^\star(\psi,\theta,0)]\, d\theta.
\end{aligned}
$$

Notice that, by virtue of Assumption 2, the integrands in these terms are negative inside the compact support intervals of $r_{\text{cw}}^\star$, $r_{\text{ccw}}^\star$ and zero elsewhere. Thus, the functional terms R_{cw}, R_{ccw} are negative.

We now employ the first mean value theorem for definite integrals. According to this theorem, since the functions R_{cw}, $R_{\text{v-ccw}}$ are negative everywhere (i.e., not changing sign), there exists constants $\psi_{\text{cw}}^0(\theta)$, $\psi_{\text{ccw}}^0(\theta)$ for each θ such that Equation (116) reduces to

$$
\begin{aligned}
k_0(t) &= \frac{-b\alpha_{\text{rot}} W_{\text{rot-c}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} R_{\text{cw}}(\psi_{\text{cw}}^0) \int_0^{2\pi} W_{\text{v-cw}}(t,\psi) d\psi, \\
k_0(t) &= \frac{-b\alpha_{\text{rot}} W_{\text{rot-c}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} R_{\text{ccw}}(\psi_{\text{ccw}}^0) \int_0^{2\pi} W_{\text{v-ccw}}(t,\psi) d\psi.
\end{aligned}
\tag{117}
$$

Differentiating this equation with respect to time, we obtain the relation between the temporal change in the synaptic weights and the temporal change in the average gain, namely,

$$
\begin{aligned}
\frac{dk_0(t)}{dt} &= \frac{-b\alpha_{\text{rot}} W_{\text{rot-c}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} R_{\text{cw}}(\psi_{\text{cw}}^0) \int_0^{2\pi} \frac{dW_{\text{v-cw}}(t,\psi)}{dt} d\psi, \\
\frac{dk_0(t)}{dt} &= \frac{-b\alpha_{\text{rot}} W_{\text{rot-c}}}{\pi\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} R_{\text{cw}}(\psi_{\text{ccw}}^0) \int_0^{2\pi} \frac{dW_{\text{v-ccw}}(t,\psi)}{dt} d\psi,
\end{aligned}
$$

Since the leading coefficient of the integral in each line is positive, Equation (89) that describes the necessary condition for gain recalibration, translates to

$$
\text{sign}\left[\frac{dk_0(t)}{dt}\right] = \text{sign}\left[\int_0^{2\pi} \frac{dW_{\text{v-cw}}(t,\psi)}{dt} d\psi\right] = \text{sign}\left[\int_0^{2\pi} \frac{dW_{\text{v-ccw}}(t,\psi)}{dt} d\psi\right] = \text{sign}[v\,\tilde{\theta}] \tag{118}
$$

Assuming Hebbian plasticity, we can transform this requirement on the synaptic weights into a

requirement on the firing rates of the rotation rings and velocity neurons, namely,

$$\begin{aligned}\operatorname{sign}\left[\eta\left(u_{\mathrm{cw}}(t)-u_{\mathrm{cw}}^{\mathrm{th}}\right)\int_0^{2\pi}\left(r_{\mathrm{cw}}(t,\psi)-r_{\mathrm{cw}}^{\mathrm{th}}\right)d\psi\right]&=\operatorname{sign}[v\,\tilde{\theta}]\\ \operatorname{sign}\left[\eta\left(u_{\mathrm{ccw}}(t)-u_{\mathrm{ccw}}^{\mathrm{th}}\right)\int_0^{2\pi}\left(r_{\mathrm{ccw}}(t,\psi)-r_{\mathrm{ccw}}^{\mathrm{th}}\right)d\psi\right]&=\operatorname{sign}[v\,\tilde{\theta}],\end{aligned}\tag{119}$$

where $\eta > 0$ denotes the learning rate, and the terms with the superscript th denotes the threshold activity-levels. These thresholds must be chosen properly to halt the plasticity when either v or $\tilde{\theta}$ is zero. Thus, we choose the activity thresholds of velocity neurons as follows:

$$\begin{aligned}u_{\mathrm{cw}}^{\mathrm{th}}&=u_{\mathrm{cw}}(0)\big|_{v(t)=0}=u_0\\ u_{\mathrm{ccw}}^{\mathrm{th}}&=u_{\mathrm{ccw}}(0)\big|_{v(t)=0}=u_0.\end{aligned}$$

According to the model of velocity neurons in Equation (28), this implies

$$\begin{aligned}\operatorname{sign}\left[u_{\mathrm{cw}}(t)-u_{\mathrm{cw}}^{\mathrm{th}}\right]&=-\operatorname{sign}[v]\\ \operatorname{sign}\left[u_{\mathrm{ccw}}(t)-u_{\mathrm{ccw}}^{\mathrm{th}}\right]&=\operatorname{sign}[v].\end{aligned}$$

Using this implication and the fact $\eta > 0$, we can refine the sign requirement in Equation (119) to

$$\begin{aligned}\operatorname{sign}\left[r_{\mathrm{cw}}^{\mathrm{avg}}(t)-r_{\mathrm{cw}}^{\mathrm{th}}\right]&=-\operatorname{sign}[\tilde{\theta}]\\ \operatorname{sign}\left[r_{\mathrm{cwc}}^{\mathrm{avg}}(t)-r_{\mathrm{cw}}^{\mathrm{th}}\right]&=\operatorname{sign}[\tilde{\theta}],\end{aligned}\tag{120}$$

where $r_{\mathrm{cw}}^{\mathrm{avg}}(t) \triangleq \frac{1}{2\pi}\int_0^{2\pi} r_{\mathrm{cw}}(t,\psi)d\psi$ and $r_{\mathrm{ccw}}^{\mathrm{avg}}(t) \triangleq \frac{1}{2\pi}\int_0^{2\pi} r_{\mathrm{ccw}}(t,\psi)d\psi$ denote the average firing rate of CW and CCW rotation rings, respectively. The sign requirement in this equation implies that these average firing rates encode the positional error $\tilde{\theta}$ via

$$\begin{aligned}r_{\mathrm{cw}}^{\mathrm{avg}}(t)&=-f_{\mathrm{cw}}(\tilde{\theta})\\ r_{\mathrm{ccw}}^{\mathrm{avg}}(t)&=f_{\mathrm{ccw}}(\tilde{\theta}),\end{aligned}$$

where $f_{\mathrm{cw}} : U_{\mathrm{cw}} \to \mathbb{R}$, $f_{\mathrm{ccw}} : U_{\mathrm{ccw}} : \mathbb{R}$ denote strictly increasing functions in the open neighborhoods U_{cw}, U_{ccw} of the origin.

6.4.2 Temporal changes in the synaptic weights of connections between the rotation rings and the central ring

In this subsection, we examine how gain recalibration can be driven by temporal changes in the functions $W_{\text{cw-c}}$ and $W_{\text{ccw-c}}$, describing the synaptic weight of connections between the rotation rings and the central ring. To consider the temporal effects of these functions on k_0, we extend Equation (112) to be

time-dependent as given below:

$$
\begin{aligned}
k_0(t) &= \frac{-b}{\pi\tau_c\|\frac{\partial r_c^\star}{\partial\psi}\|^2}\int_0^{2\pi}\int_0^{2\pi}\frac{\partial^2 r_c^\star(\psi-\theta)}{\partial\psi^2}\alpha_{\text{cw}}W_{\text{cw-c}}(t,\psi)W_{\text{v-cw}}(\psi)\text{sign}[r_{\text{cw}}^\star(\psi,\theta,0)]d\psi d\theta \\
k_0(t) &= \frac{-b}{\pi\tau_c\|\frac{\partial r_c^\star}{\partial\psi}\|^2}\int_0^{2\pi}\int_0^{2\pi}\frac{\partial^2 r_c^\star(\psi-\theta)}{\partial\psi^2}\alpha_{\text{ccw}}W_{\text{ccw-c}}(t,\psi)W_{\text{v-ccw}}(\psi)\text{sign}[r_{\text{ccw}}^\star(\psi,\theta,0)]d\psi d\theta
\end{aligned}
\tag{121}
$$

On the other hand, we assume that the remaining parameters of our model are hardwired (i.e., time-invariant), initialized with values inspired by the classical ring attractor models such that

$$
\begin{aligned}
\alpha_{\text{cw}} &= \alpha_{\text{ccw}} = \alpha_{\text{rot}} \\
W_{\text{v-cw}}(\psi) &= W_{\text{v-ccw}}(\psi) = W_{\text{v-rot}} \\
W_{\text{c-cw}}(\psi) &= W_{\text{c-ccw}}(\psi) = W_{\text{c-rot}}.
\end{aligned}
\tag{122}
$$

We now apply the steps between Equation (116) and (118), but this time keeping the now plastic weights $W_{\text{cw-c}}$, $W_{\text{ccw-c}}$ inside the integral and instead moving the now hard-wired weights $W_{\text{v-cw}}(\psi) = W_{\text{v-ccw}}(\psi) = W_{\text{v-rot}}$ outside. This yields the stability requirement of gain recalibration for the temporal changes in the plastic weights $W_{\text{cw-c}}$, $W_{\text{ccw-c}}$, namely,

$$
\text{sign}\left[\frac{dk_0(t)}{dt}\right] = \text{sign}\left[\int_0^{2\pi}\frac{dW_{\text{cw-c}}(t,\psi)}{dt}d\psi\right] = \text{sign}\left[\int_0^{2\pi}\frac{dW_{\text{ccw-c}}(t,\psi)}{dt}d\psi\right] = \text{sign}[v\,\tilde{\theta}]
$$

We next assume Hebbian plasticity as the mechanism of these temporal changes in $W_{\text{cw-c}}$, $W_{\text{ccw-c}}$. This assumption lets us transform this sign requirement on the synaptic weights into a requirement on the firing rates of the central ring and rotation ring neurons, namely,

$$
\begin{aligned}
\text{sign}\left[\eta\int_0^{2\pi}\big(r_{\text{c}}(t,\psi)-r_{\text{c}}^{\text{th}}\big)\big(r_{\text{cw}}(t,\psi)-r_{\text{cw}}^{\text{th}}\big)d\psi\right] &= \text{sign}[v\,\tilde{\theta}] \\
\text{sign}\left[\eta\int_0^{2\pi}\big(r_{\text{c}}(t,\psi)-r_{\text{c}}^{\text{th}}\big)\big(r_{\text{ccw}}(t,\psi)-r_{\text{ccw}}^{\text{th}}\big)d\psi\right] &= \text{sign}[v\,\tilde{\theta}],
\end{aligned}
$$

where $\eta > 0$ denotes the learning rate, and the terms with the superscript th denote the threshold activity-levels. Using the fact $\eta > 0$ and the superscript avg as a shorthand notation for the average firing rates, we can simplify this equation to

$$
\begin{aligned}
\text{sign}\left[\int_0^{2\pi} r_{\text{c}}(t,\psi)r_{\text{cw}}(t,\psi)d\psi - 2\pi r_{\text{cw}}^{\text{th}}r_{\text{cw}}^{\text{avg}}(t) - 2\pi r_{\text{c}}^{\text{th}}r_{\text{c}}^{\text{avg}}(t) + 2\pi r_{\text{c}}^{\text{th}}r_{\text{cw}}^{\text{th}}\right] &= \text{sign}[v\,\tilde{\theta}] \\
\text{sign}\left[\int_0^{2\pi} r_{\text{c}}(t,\psi)r_{\text{ccw}}(t,\psi)d\psi - 2\pi r_{\text{cw}}^{\text{th}}r_{\text{ccw}}^{\text{avg}}(t) - 2\pi r_{\text{c}}^{\text{th}}r_{\text{c}}^{\text{avg}}(t) + 2\pi r_{\text{c}}^{\text{th}}r_{\text{ccw}}^{\text{th}}\right] &= \text{sign}[v\,\tilde{\theta}].
\end{aligned}
\tag{123}
$$

Recall that firing rates r_{c},r_{cw}, and r_{ccw} are positive. This lets us employ the first-mean value theorem for the definite integral in Equation (123). According to this theorem, there exists constants x_{cw}^0 and

x^0_{ccw} such that

$$\int_0^{2\pi} r_{\text{c}}(t,\psi) r_{\text{cw}}(t,\psi) d\psi = r_{\text{c}}(t, x^0_{\text{cw}}) \int_0^{2\pi} r_{\text{cw}}(t,\psi) d\psi = 2\pi\, r_{\text{c}}(t, x^0_{\text{cw}}) r^{\text{avg}}_{\text{cw}}$$
$$\int_0^{2\pi} r_{\text{c}}(t,\psi) r_{\text{ccw}}(t,\psi) d\psi = r_{\text{c}}(t, x^0_{\text{ccw}}) \int_0^{2\pi} r_{\text{ccw}}(t,\psi) d\psi = 2\pi\, r_{\text{c}}(t, x^0_{\text{ccw}}) r^{\text{avg}}_{\text{ccw}}.$$

Under the assumption that the shape of the central ring's activity bump, denoted by $\mu_{\text{c}}(\psi) \triangleq \frac{r^\star_{\text{c}}(\psi)}{\|r^\star_{\text{c}}\|}$, remains constant thanks to attractor dynamics, there exists a constant gain factor β^0_{cw} for a given x^0_{cw} such that $r_{\text{c}}(t, x^0_{\text{cw}}) = \beta^0_{\text{cw}} r^{\text{avg}}_{\text{c}}(t)$. Using this, we can rewrite the above equation as

$$\int_0^{2\pi} r_{\text{c}}(t,\psi) r_{\text{cw}}(t,\psi) d\psi = 2\pi \beta^0_{\text{cw}}\, r^{\text{avg}}_{\text{c}}(t) r^{\text{avg}}_{\text{cw}}(t)$$
$$\int_0^{2\pi} r_{\text{c}}(t,\psi) r_{\text{ccw}}(t,\psi) d\psi = 2\pi \beta^0_{\text{ccw}}\, r^{\text{avg}}_{\text{c}}(t) r^{\text{avg}}_{\text{ccw}}(t).$$

Substituting this equation into (123), we obtain the following simplified sign requirement:

$$\begin{aligned} \text{sign}\left[2\pi\beta^0_{\text{cw}}\, r^{\text{avg}}_{\text{c}}(t) r^{\text{avg}}_{\text{cw}}(t) - 2\pi r^{\text{th}}_{\text{cw}} r^{\text{avg}}_{\text{cw}}(t) - 2\pi r^{\text{th}}_{\text{c}} r^{\text{avg}}_{\text{c}}(t) + 2\pi r^{\text{th}}_{\text{c}} r^{\text{th}}_{\text{cw}}\right] &= \text{sign}[v\,\tilde{\theta}] \\ \text{sign}\left[2\pi\beta^0_{\text{ccw}}\, r^{\text{avg}}_{\text{c}}(t) r^{\text{avg}}_{\text{ccw}}(t) - 2\pi r^{\text{th}}_{\text{cw}} r^{\text{avg}}_{\text{ccw}}(t) - 2\pi r^{\text{th}}_{\text{c}} r^{\text{avg}}_{\text{c}}(t) + 2\pi r^{\text{th}}_{\text{c}} r^{\text{th}}_{\text{ccw}}\right] &= \text{sign}[v\,\tilde{\theta}]. \end{aligned} \tag{124}$$

Satisfying each line this equation at a given time t and velocity v requires that the central ring's average firing rate $r^{\text{avg}}_{\text{c}}$ or the rotation rings' average firing rates $r^{\text{avg}}_{\text{cw}}$, $r^{\text{avg}}_{\text{ccw}}$ must encode the positional error $\tilde{\theta}$ via strictly monotonic changes. More formally, there exists strictly monotonic functions $f_{\text{c}} : U_{\text{c}} \to \mathbb{R}$, $f_{\text{cw}} : U_{\text{cw}} \to \mathbb{R}$, $f_{\text{ccw}} : U_{\text{ccw}} \to \mathbb{R}$ in non-empty open neighborhoods U_{c},U_{cw},U_{ccw} of the origin such that at least one of the

$$\begin{aligned} r^{\text{avg}}_{cw}(t) &= f_{\text{cw}}(\tilde{\theta}) \\ r^{\text{avg}}_{ccw} &= f_{\text{ccw}}(\tilde{\theta}) \\ r^{\text{avg}}_{\text{c}} &= f_{\text{c}}(\tilde{\theta}). \end{aligned}$$

6.4.3 Temporal changes in the tuning slope of velocity neurons

In this subsection, we examine how gain recalibration can be driven by temporal changes in the terms α_{cw} and α_{ccw}, denoting the slope of the CW and CCW velocity neuron's tuning curves, respectively. To consider the temporal effects of these terms on k_0, we extend Equation (112) to be time-dependent and re-arrange terms, which yields

$$\begin{aligned} k_0(t) &= C_{\text{cw}}\, \alpha_{\text{cw}}(t) \\ k_0(t) &= C_{\text{ccw}}\, \alpha_{\text{ccw}}(t), \end{aligned} \tag{125}$$

where the positive constants C_{cw}, C_{ccw} take the form

$$C_{\text{cw}} \triangleq \frac{-b}{\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_{\theta-\Omega_{\text{rot}}^l(\theta)}^{\theta+\Omega_{\text{rot}}^r(\theta)} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} W_{\text{cw-c}}(\psi) W_{\text{v-cw}}(\psi) d\psi d\theta$$

$$C_{\text{ccw}} \triangleq \frac{-b}{\tau_c \left\| \frac{\partial r_c^\star}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_{\theta-\Omega_{\text{rot}}^l(\theta)}^{\theta+\Omega_{\text{rot}}^r(\theta)} \frac{\partial^2 r_c^\star(\psi-\theta)}{\partial \psi^2} W_{\text{ccw-c}}(\psi) W_{\text{v-ccw}}(\psi) d\psi d\theta.$$

On the other hand, we assume that the remaining parameters of our model are hardwired (i.e., time-invariant), initialized with values inspired by the classical ring attractor models such that

$$\begin{aligned} W_{\text{v-cw}}(\psi) &= W_{\text{v-ccw}}(\psi) = W_{\text{v-rot}} \\ W_{\text{cw-c}}(\psi) &= W_{\text{ccw-c}}(\psi) = W_{\text{rot-c}} \\ W_{\text{c-cw}}(\psi) &= W_{\text{c-ccw}}(\psi) = W_{\text{c-rot}} \end{aligned} \tag{126}$$

Note that, under these assumptions, the left- and right- widths of rotation rings' activity bumps during immobility become equal and independent of θ, namely,

$$\forall \theta, \quad \Omega_{\text{cw}}^l(0,\theta) = \Omega_{\text{cw}}^r(0,\theta) = \Omega_{\text{ccw}}^l(0,\theta) = \Omega_{\text{ccw}}^r(0,\theta) \triangleq \Omega_{\text{rot}}.$$

Differentiating Equation (125) with respect to time and using Equation (89), describing the necessary condition for gain recalibration, we obtain the requirements that must be satisfied by the temporal changes in α_{cw}, α_{ccw} for gain recalibration as follows:

$$\text{sign}\left[\frac{dk_0(t)}{dt}\right] = \text{sign}\left[C_{\text{cw}} \frac{d\alpha_{\text{cw}}(t)}{dt}\right] = \text{sign}\left[C_{\text{ccw}} \frac{d\alpha_{\text{ccw}}(t)}{dt}\right] = \text{sign}[v\,\tilde{\theta}]. \tag{127}$$

Since C_{cw} and C_{ccw} are positive, this simplifies to

$$\text{sign}\left[\frac{d\alpha_{\text{cw}}(t)}{dt}\right] = \text{sign}\left[\frac{d\alpha_{\text{ccw}}(t)}{dt}\right] = \text{sign}[v\,\tilde{\theta}]. \tag{128}$$

In previous sections, we examined the neural mechanisms of plasticity that can satisfy similar equalities through temporal changes in the synaptic weights. However, for the velocity neurons, our model does not include details to examine the neural mechanisms governing the temporal changes in the slopes of their tuning curves. Alternatively, we can garner insight into the consequences of these temporal changes for the rest of the neurons in our model, assuming that the above equalities are met. In the remainder of this section, we examine how average firing rate of rotation rings are affected by the temporal changes in the slopes of the velocity neurons.

Implications of Equation (127) on the average firing rate of rotation rings It can be trivially seen from Equation (26) that the average firing rate of rotation rings varies proportionally with

the slopes α_{cw},α_{ccw}. Specifically,

$$\text{sign}\Big[\frac{dr_{\text{cw}}^{\text{avg}}}{dt}\Big] = -\text{sign}\Big[\frac{d\alpha_{\text{cw}}}{dt}v\Big]$$
$$\text{sign}\Big[\frac{dr_{\text{ccw}}^{\text{avg}}}{dt}\Big] = +\text{sign}\Big[\frac{d\alpha_{\text{ccw}}}{dt}v\Big].$$

Plugging Equation (128) into the right-hand side, we can further refine this requirement :

$$\text{sign}\Big[\frac{dr_{\text{cw}}^{\text{avg}}}{dt}\Big] = -\text{sign}\Big[v^2\tilde{\theta}\Big]$$
$$\text{sign}\Big[\frac{dr_{\text{ccw}}^{\text{avg}}}{dt}\Big] = +\text{sign}\Big[v^2\tilde{\theta}\Big].$$

According to these equations, for a given v, the temporal change in the rotation rings' average firing rates must be in the same direction as $\tilde{\theta}$, namely,

$$\text{sign}\Big[\frac{dr_{\text{cw}}^{\text{avg}}}{dt}\Big] = -\text{sign}[\tilde{\theta}]$$
$$\text{sign}\Big[\frac{dr_{\text{ccw}}^{\text{avg}}}{dt}\Big] = +\text{sign}[\tilde{\theta}].$$

This implies that temporal changes in the rotation rings' average firing rates can be identified with strictly monotone functions of $\tilde{\theta}$ in some neighborhood of the origin. More specifically, we can find two strictly increasing functions $f_{\text{cw}} : U_{\text{cw}} \to \mathbb{R}$, $f_{\text{ccw}} : U_{\text{ccw}} \to \mathbb{R}$ in the open neighborhoods U_{cw}, U_{ccw} of the origin such that

$$\frac{dr_{\text{cw}}^{\text{avg}}}{dt} = -f_{\text{cw}}(\tilde{\theta})$$
$$\frac{dr_{\text{ccw}}^{\text{avg}}}{dt} = +f_{\text{ccw}}(\tilde{\theta}).$$

What can we tell about the average firing rates $r_{\text{ccw}}^{\text{avg}}$,$r_{\text{cw}}^{\text{avg}}$ when their temporal changes are strictly monotone functions of $\tilde{\theta}$? To answer this question, we restrict ourselves to simple $\tilde{\theta}(t)$ trajectories that do not change sign by crossing zero. Under this condition, one can show that average firing rates of the CW and CCW rotation rings encode the time-integral $I_{\tilde{\theta}} \triangleq \int_0^t \tilde{\theta}(\lambda)d\lambda$ of the positional error via

$$r_{\text{cw}}^{\text{avg}}(t) - r_{\text{cw}}^{\text{avg}}(0) = -F_{\text{cw}}(I_{\tilde{\theta}}(t))$$
$$r_{\text{ccw}}^{\text{avg}}(t) - r_{\text{ccw}}^{\text{avg}}(0) = +F_{\text{ccw}}(I_{\tilde{\theta}}(t)),$$

where $F_{\text{cw}} : \mathcal{U}_{\text{cw}} \to \mathbb{R}$, $F_{\text{ccw}} : \mathcal{U}_{\text{ccw}} \to \mathbb{R}$ denote strictly increasing functions of $I_{\tilde{\theta}}$ in some non-empty neighborhoods $\mathcal{U}_{\text{cw}}$, $\mathcal{U}_{\text{ccw}}$ of the origin.

6.4.4 Temporal changes in the width of the rotation ring's activity bump

In this subsection, we examine how gain recalibration can be driven by temporal changes in the terms Ω_{rot}^{l} and Ω_{rot}^{r}, respectively, denoting the left- and right- width of the rotation rings' activity bumps during immobility. As a first step toward this examination, we extend Equation (111), the analytical expression of the average gain k_0, to be time-dependent as given below:

$$k_0(t) = \frac{-b}{\tau_c\big\|\frac{\partial r_{\text{c}}^{\star}}{\partial\psi}\big\|^2}\int_0^{2\pi}\int_{\theta-\Omega_{\text{rot}}^{l}(t,\theta)}^{\theta+\Omega_{\text{rot}}^{r}(t,\theta)}\frac{\partial^2 r_{\text{c}}^{\star}(\psi-\theta)}{\partial\psi^2}\Big(\alpha_{\text{cw}}W_{\text{cw-c}}(\psi)W_{\text{v-cw}}(\psi) + \alpha_{\text{ccw}}W_{\text{ccw-c}}(\psi)W_{\text{v-ccw}}(\psi)\Big)d\psi d\theta \tag{129}$$

On the other hand, we assume that the remaining parameters of our model are hardwired (i.e., time-invariant), initialized with values inspired by the classical ring attractor models such that

$$\begin{aligned} W_{\text{v-cw}}(\psi) &= W_{\text{v-ccw}}(\psi) = W_{\text{v-rot}} \\ W_{\text{cw-c}}(\psi) &= W_{\text{ccw-c}}(\psi) = W_{\text{rot-c}} \\ W_{\text{c-cw}}(\psi) &= W_{\text{c-ccw}}(\psi) = W_{\text{c-rot}} \\ \alpha_{\text{cw}} &= \alpha_{\text{ccw}} = \alpha_{\text{rot}}. \end{aligned} \tag{130}$$

Under these assumptions, Ω^l_{rot},Ω^r_{rot} become independent of θ, implying that they can change only over time, i.e. $\Omega^l_{\text{rot}}(t)$, $\Omega^r_{\text{rot}}(t)$. Taking into account this implication and the assumptions in Equation (130), we can simplify Equation (129) to

$$k_0(t) = \frac{-2\, b\, \alpha_{\text{rot}}\, W_{\text{rot-c}}\, W_{\text{v-rot}}}{\tau_c \left\| \frac{\partial r^\star_{\text{c}}}{\partial \psi} \right\|^2} \int_0^{2\pi} \int_{\theta-\Omega^l_{\text{rot}}(t)}^{\theta+\Omega^r_{\text{rot}}(t)} \frac{\partial^2 r^\star_{\text{c}}(\psi-\theta)}{\partial \psi^2}\, d\psi d\theta \tag{131}$$

If the widths Ω^l_{rot},Ω^r_{rot} are time-invariant, k_0 in Equation (131) is a constant, implying that the temporal change in θ does not lead to any change in k_0. Therefore, taking time-derivative of Equation (131) and ignoring the changes due to θ, we obtain the relationship between the temporal change in k_0 and that in the widths Ω^l_{rot},Ω^r_{rot} as follows:

$$\frac{dk_0}{dt} = \frac{-4\pi\, b\, \alpha_{\text{rot}}\, W_{\text{rot-c}}\, W_{\text{v-rot}}}{\tau_c \left\| \frac{\partial r^\star_{\text{c}}}{\partial \psi} \right\|^2} \left(\frac{d\Omega^r_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=\Omega^r_{\text{rot}}} + \frac{d\Omega^l_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=-\Omega^l_{\text{rot}}} \right)$$

Note that, when deriving this equation, we applied the Leibniz's integral rule since the widths Ω^l_{rot},Ω^r_{rot} appear in the bounds of an integral in Equation (131).

For gain recalibration, this temporal change in k_0 must satisfy Equation (89), which implies

$$\text{sign}\left[\frac{dk_0}{dt}\right] = \text{sign}\left[\frac{-4\pi\, b\, \alpha_{\text{rot}}\, W_{\text{rot-c}}\, W_{\text{v-rot}}}{\tau_c \left\| \frac{\partial r^\star_{\text{c}}}{\partial \psi} \right\|^2} \left(\frac{d\Omega^r_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=\Omega^r_{\text{rot}}} + \frac{d\Omega^l_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=-\Omega^l_{\text{rot}}} \right)\right] = \text{sign}\big[v\tilde{\theta}\big].$$

Since the terms (α, b, τ_c, $\left\| \frac{\partial r^\star_{\text{c}}(\psi)}{\partial \psi} \right\|$, $W_{\text{rot-c}}$, $W_{\text{v-rot}}$) are positive, we can simplify this requirement to

$$\text{sign}\left[\frac{dk_0}{dt}\right] = \text{sign}\left[\frac{d\Omega^r_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=\Omega^r_{\text{rot}}} + \frac{d\Omega^l_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=-\Omega^l_{\text{rot}}} \right] = -\text{sign}\big[v\tilde{\theta}\big]. \tag{132}$$

In addition to this equation, the widths Ω^l_{rot},Ω^r_{rot} must satisfy the symmetry constraint in Remark 3. Using Equations (49), (130) and applying the steps between Equation (131) and (132), we can express this symmetry constraint in a form that explicitly depends on Ω^l_{rot},Ω^r_{rot} as follows:

$$\frac{d\Omega^r_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=\Omega^r_{\text{rot}}} = \frac{d\Omega^l_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=-\Omega^l_{\text{rot}}}$$

Combining this symmetry constraint with Equation (132), we find that the necessary condition for gain recalibration requires the temporal changes in Ω^l_{rot} and Ω^r_{rot} to change according to

$$\begin{aligned} \text{sign}\left[\frac{d\Omega^r_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=\Omega^r_{\text{rot}}} \right] &= \text{sign}[v\,\tilde{\theta}] \\ \text{sign}\left[\frac{d\Omega^l_{\text{rot}}}{dt} \frac{d^2 r^\star_{\text{c}}(\psi)}{d\psi^2}\bigg|_{\psi=-\Omega^l_{\text{rot}}} \right] &= \text{sign}[v\,\tilde{\theta}]. \end{aligned}$$

Recall that, by Assumptions 1 and 2, the second derivative $\frac{d^2 r_c^\star(\psi)}{d\psi^2}$ takes a negative value at the points $\psi = \Omega_{\text{rot}}^r$, $\psi = -\Omega_{\text{rot}}^l$. Thus, the above equation can be simplified to

$$\begin{aligned}\text{sign}\left[\frac{d\,\Omega_{\text{rot}}^r}{dt}\right] &= \text{sign}[\tilde{\theta} v]\\ \text{sign}\left[\frac{d\,\Omega_{\text{rot}}^l}{dt}\right] &= \text{sign}[\tilde{\theta} v].\end{aligned}$$

According to these equations, for a given v, the temporal change in the rotation rings' widths must be in the same direction as $\tilde{\theta}$, namely,

$$\begin{aligned}\text{sign}\left[\frac{d\,\Omega_{\text{rot}}^r}{dt}\right] &= \text{sign}[\tilde{\theta}]\\ \text{sign}\left[\frac{d\,\Omega_{\text{rot}}^l}{dt}\right] &= \text{sign}[\tilde{\theta}].\end{aligned} \tag{133}$$

This implies that there exists strictly increasing functions $f_{\text{rot}}^l : U_{\text{rot}}^l \to \mathbb{R}$, $f_{\text{rot}}^r : U_{\text{rot}}^r \to \mathbb{R}$ in the open neighborhoods U_{rot}^l,U_{rot}^r of the origin such that

$$\begin{aligned}\text{sign}\left[\frac{d\,\Omega_{\text{rot}}^r}{dt}\right] &= f_{\text{rot}}^r(\tilde{\theta})\\ \text{sign}\left[\frac{d\,\Omega_{\text{rot}}^l}{dt}\right] &= f_{\text{rot}}^l(\tilde{\theta}).\end{aligned}$$

What can we tell about the left- and right- widths Ω_{rot}^r,Ω_{rot}^l when their temporal changes are strictly increasing functions of $\tilde{\theta}$? To answer this question, we restrict ourselves to simple $\tilde{\theta}(t)$ trajectories that do not change sign by crossing zero. Under this condition, one can show that left- and right- widths of the rotation rings encode the time-integral $I_{\tilde{\theta}} \triangleq \int_0^t \tilde{\theta}(\lambda) d\lambda$ of the positional error via

$$\begin{aligned}\Omega_{\text{rot}}^r(t) - \Omega_{\text{rot}}^r(0) &= F_{\text{rot}}^r(I_{\tilde{\theta}}(t))\\ \Omega_{\text{rot}}^l(t) - \Omega_{\text{rot}}^l(0) &= +F_{\text{rot}}^l(I_{\tilde{\theta}}(t)),\end{aligned}$$

where $F_{\text{rot}}^r : \mathcal{U}_{\text{rot}}^r \to \mathbb{R}$, $F_{\text{rot}}^l : \mathcal{U}_{\text{rot}}^l \to \mathbb{R}$ denote strictly increasing functions of $I_{\tilde{\theta}}$ in some non-empty neighborhoods $\mathcal{U}_{\text{rot}}^r$, $\mathcal{U}_{\text{rot}}^l$ of the origin.

6.4.5 Temporal changes in the norm of the central ring's persistent activity bump

In this subsection, we examine how gain recalibration can be driven by temporal changes in the term $r_c^\star$, describing the central ring's persistent activity bump. Specifically, we assume that the activity bump's shape $\mu_c(\psi) \triangleq \frac{r_c^\star(\psi)}{\|r_c^\star\|}$ remains invariant while the temporal changes in the norm $N_c \triangleq \|r_c^\star\|$ drive the gain recalibration. On the other hand, we assume that the remaining parameters of our model are hardwired (i.e., time-invariant), initialized with values inspired by the classical ring attractor models such that

$$\begin{aligned}W_{\text{v-cw}}(\psi) &= W_{\text{v-ccw}}(\psi) = W_{\text{v-rot}}\\ W_{\text{cw-c}}(\psi) &= W_{\text{ccw-c}}(\psi) = W_{\text{rot-c}}\\ W_{\text{c-cw}}(\psi) &= W_{\text{c-ccw}}(\psi) = W_{\text{c-rot}}\\ \alpha_{\text{cw}} &= \alpha_{\text{ccw}} = \alpha_{\text{rot}}.\end{aligned} \tag{134}$$

An implication of these assumptions is that Ω^l_{rot},Ω^r_{rot} are independent of θ. Re-organizing Equation (111), the analytical expression of k_0, in accordance with these conditions, we obtain

$$k_0(t) = \frac{1}{N_c(t)}\left(\frac{-2\,b\,\alpha_{\text{rot}}\,W_{\text{rot-c}}\,W_{\text{v-rot}}}{\tau_c\|\mu_c'\|^2}\int_0^{2\pi}\int_{-\Omega^l_{\text{rot}}}^{\Omega^r_{\text{rot}}}\mu_c''(\psi)\,d\psi d\theta\right) = \frac{1}{N_c(t)}C, \tag{135}$$

where C denotes a positive constant, and $N_c(t)$ denotes the time-varying norm of the persistent activity bump $r_c^\star$ as the driver of the gain recalibration. Also note that, while reorganizing to obtain this equation, we used the following facts

$$\mu_c'(\psi) \triangleq \frac{1}{N_c(t)}\frac{\partial r_c^\star(\psi)}{\partial\psi}$$
$$\mu_c''(\psi) \triangleq \frac{1}{N_c(t)}\frac{\partial^2 r_c^\star(\psi)}{\partial\psi^2}.$$

In order to find the relation between the temporal change in N_c and the resulting change in k_0, we simply differentiate Equation (135) with respect to time, which yields,

$$\frac{dk_0}{dt} = -\frac{C}{N_c^2}\frac{dN_c}{dt}.$$

For gain recalibration, this temporal change in k_0 must satisfy Equation (89). This implies

$$\text{sign}\left[\frac{dN_c}{dt}\right] = -\text{sign}[\tilde{\theta}v].$$

According to this equation, for a given v, the temporal change in the norm of the central ring's activity bump must be in the opposite direction of $\tilde{\theta}$, namely,

$$\text{sign}\left[\frac{dN_c}{dt}\right] = -\text{sign}[\tilde{\theta}].$$

This implies that there exists a strictly increasing function $f_c : U_c \to \mathbb{R}$ in the open neighborhood U_c of the origin such that

$$\text{sign}\left[\frac{dN_c}{dt}\right] = -f_c(\tilde{\theta})$$

What can we tell about the norm N_c when its temporal changes is a strictly decreasing function of $\tilde{\theta}$? To answer this question, we restrict ourselves to simple $\tilde{\theta}(t)$ trajectories that do not change sign by crossing zero. Under this condition, one can show that the norm N_c of the central ring's activity bump encode the time-integral $I_{\tilde{\theta}} \triangleq \int_0^t \tilde{\theta}(\lambda)d\lambda$ of the positional error via

$$N_c(t) - N_c(0) = -F_c(I_{\tilde{\theta}}(t)) \tag{136}$$

where $F_c : \mathcal{U}_c \to \mathbb{R}$ denotes a strictly increasing function of $I_{\tilde{\theta}}$ in some non-empty neighborhood $\mathcal{U}_c$of the origin.

Note that the norm N_c is related to the average firing rate r_c^{avg} via

$$r_c^{\text{avg}}(t) = \frac{1}{2\pi}\int_0^{2\pi} r_c^\star(\psi)d\psi = \frac{1}{2\pi}\int_0^{2\pi}\mu_c(\psi)d\psi N_c(t).$$

Since the integral $\int_0^{2\pi}\mu_c(\psi)$ is constant (i.e., time-invariant), Equation (136) implies that the average

firing rate $r_{\mathrm{c}}^{\mathrm{avg}}$ of the central ring encodes the time-integral of $\tilde{\theta}$ via similar monotonic changes.

6.5 Analytical proof of gain recalibration in the proposed ring attractor model

The proposed model employs plasticity in the connections between the velocity neurons and the rotation ring neurons. Assuming Hebbian plasticity, we can model this plasticity through

$$\frac{dW_{\text{v-cw}}(t,\psi)}{dt} = \eta\big(u_{\mathrm{cw}}(t) - u_{\mathrm{cw}}^{\mathrm{th}}\big)\big(r_{\mathrm{cw}}(t,\psi) - r_{\mathrm{cw}}^{\mathrm{th}}\big)$$
$$\frac{dW_{\text{v-ccw}}(t,\psi)}{dt} = \eta\big(u_{\mathrm{ccw}}(t) - u_{\mathrm{ccw}}^{\mathrm{th}}\big)\big(r_{\mathrm{ccw}}(t,\psi) - r_{\mathrm{ccw}}^{\mathrm{th}}\big),$$

where $\eta > 0$ denotes the learning rate, and the terms with the superscript th denotes the threshold activity-levels. As we previously derived in Appendix 6.4.1, both synaptic weights $W_{\text{v-cw}}$ and $W_{\text{v-ccw}}$ are related to the average gain k_0 via Equation (117). Differentiating each line in this equation, averaging over these time-derivatives of the lines, and then using the model of Hebbian plasticity (given above), we obtain an equation for how firing rates of velocity and rotation rings neurons drive the gain recalibration in our proposed attractor model, namely,

$$\frac{dk_0}{dt} = 2\pi\eta\Big(C_{\mathrm{cw}}\big(u_{\mathrm{cw}}(t) - u_{\mathrm{cw}}^{\mathrm{th}}\big)\big(r_{\mathrm{cw}}^{\mathrm{avg}}(t) - r_{\mathrm{cw}}^{\mathrm{th}}\big) + C_{\mathrm{ccw}}\big(u_{\mathrm{ccw}}(t) - u_{\mathrm{ccw}}^{\mathrm{th}}\big)\big(r_{\mathrm{ccw}}^{\mathrm{avg}}(t) - r_{\mathrm{ccw}}^{\mathrm{th}}\big)\Big), \tag{137}$$

where the terms C_{cw}, C_{ccw} denote the positive constants

$$C_{\mathrm{cw}} = \frac{-b\alpha_{\mathrm{rot}}W_{\text{rot-c}}}{\pi\tau_c\big\|\frac{\partial r_{\mathrm{c}}^{\star}}{\partial\psi}\big\|^2}R_{\mathrm{cw}}(\psi_{\mathrm{cw}}^0)$$
$$C_{\mathrm{ccw}} = \frac{-b\alpha_{\mathrm{rot}}W_{\text{rot-c}}}{\pi\tau_c\big\|\frac{\partial r_{\mathrm{c}}^{\star}}{\partial\psi}\big\|^2}R_{\mathrm{ccw}}(\psi_{\mathrm{ccw}}^0).$$

Recall that, firing rate of velocity neurons encode the velocity v via

$$u_{\mathrm{cw}}(t) = u^0 - \alpha_{\mathrm{rot}}v$$
$$u_{\mathrm{ccw}}(t) = u^0 + \alpha_{\mathrm{rot}}v,$$

where α_{rot} denotes the absolute value of the tuning curves' slopes. Moreover, the rotation rings encode v and error $\tilde{\theta}$ conjunctively. This conjunctive coding takes the form of a superposition, which can be approximated as a linear model, namely,

$$r_{\mathrm{cw}}^{\mathrm{avg}}(t) = r_{\mathrm{rot}}^0 - \beta_v v - \beta_{\tilde{\theta}}\tilde{\theta}$$
$$r_{\mathrm{ccw}}^{\mathrm{avg}}(t) = r_{\mathrm{rot}}^0 + \beta_v v + \beta_{\tilde{\theta}}\tilde{\theta},$$

where β_v and $\beta_{\tilde{\theta}}$ denote absolute value of slopes due to v and $\tilde{\theta}$, respectively. Using these firing-rate equations and the threshold choices

$$u_{\mathrm{cw}}^{\mathrm{th}} = u_{\mathrm{ccw}}^{\mathrm{th}} = u^0$$
$$r_{\mathrm{cw}}^{\mathrm{th}} = r_{\mathrm{rot}}^0/\beta_{\mathrm{cw}}$$
$$r_{\mathrm{ccw}}^{\mathrm{th}} = r_{\mathrm{rot}}^0/\beta_{\mathrm{ccw}},$$

we can reduce Equation (137) to a gain update rule g_0 that is a function of $\tilde{\theta}$ and v, namely,

$$\frac{dk_0}{dt} = g_0(\tilde{\theta}, v) = 2\pi\eta \left(C_{\mathrm{cw}} + C_{\mathrm{ccw}}\right)\Big(\alpha_{\mathrm{rot}}\beta_{\mathrm{v}} v^2 + \alpha_{\mathrm{rot}}\beta_{\tilde{\theta}} v\tilde{\theta}\Big)$$

For a given $v \neq 0$, this gain update rule satisfies

$$\frac{dg_0(\tilde{\theta}, v)}{d\tilde{\theta}} \frac{1}{v} = 2\pi\eta \left(C_{\mathrm{cw}} + C_{\mathrm{ccw}}\right)\alpha_{\mathrm{rot}}\beta_{\tilde{\theta}} > 0.$$

As per Prop. 2, this inequality guarantees that the proposed model achieves stable gain recalibration.

References

[1] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, "Theory of orientation tuning in visual cortex.," *Proceedings of the National Academy of Sciences*, vol. 92, no. 9, pp. 3844–3848, 1995.

[2] K. Wimmer, D. Q. Nykamp, C. Constantinidis, and A. Compte, "Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory," *Nature neuroscience*, vol. 17, no. 3, pp. 431–439, 2014.

[3] A. Compte, N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang, "Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model," *Cerebral cortex*, vol. 10, no. 9, pp. 910–923, 2000.

[4] J. M. Esnaola-Acebes, A. Roxin, and K. Wimmer, "Flexible integration of continuous sensory evidence in perceptual estimation tasks," *Proceedings of the National Academy of Sciences*, vol. 119, no. 45, p. e2214441119, 2022.

[5] M. Furman and X.-J. Wang, "Similarity effect and optimal control of multiple-choice decision making," *Neuron*, vol. 60, no. 6, pp. 1153–1168, 2008.

[6] E. H. Nieh, M. Schottdorf, N. W. Freeman, R. J. Low, S. Lewallen, S. A. Koay, L. Pinto, J. L. Gauthier, C. D. Brody, and D. W. Tank, "Geometry of abstract learned knowledge in the hippocampus," *Nature*, vol. 595, no. 7865, pp. 80–84, 2021.

[7] W. Wojtak, S. Coombes, D. Avitabile, E. Bicho, and W. Erlhagen, "A dynamic neural field model of continuous input integration," *Biological Cybernetics*, vol. 115, no. 5, pp. 451–471, 2021.

[8] W. Skaggs, J. Knierim, H. Kudrimoti, and B. McNaughton, "A model of the neural basis of the rat's sense of direction," *Advances in neural information processing systems*, vol. 7, 1994.

[9] H. Blair, "Simulation of a thalamocortical circuit for computing directional heading in the rat," *Advances in neural information processing systems*, vol. 8, 1995.

[10] A. D. Redish, A. N. Elga, and D. S. Touretzky, "A coupled attractor model of the rodent head direction system," *Network: computation in neural systems*, vol. 7, no. 4, p. 671, 1996.

[11] K. Zhang, "Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory," *Journal of Neuroscience*, vol. 16, no. 6, pp. 2112–2126, 1996.

[12] J. J. Knierim and K. Zhang, "Attractor dynamics of spatially correlated neural activity in the limbic system," *Annual review of neuroscience*, vol. 35, pp. 267–285, 2012.

[13] K. Yoon, M. A. Buice, C. Barry, R. Hayman, N. Burgess, and I. R. Fiete, "Specific evidence of low-dimensional continuous attractor dynamics in grid cells," *Nature neuroscience*, vol. 16, no. 8, pp. 1077–1084, 2013.

[14] R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M.-B. Moser, and E. I. Moser, "Toroidal topology of population activity in grid cells," *Nature*, vol. 602, no. 7895, pp. 123–128, 2022.

[15] Z. Ajabi, A. T. Keinath, X.-X. Wei, and M. P. Brandon, "Population dynamics of head-direction neurons during drift and reorientation," *Nature*, vol. 615, no. 7954, pp. 892–899, 2023.

[16] S. S. Kim, H. Rouault, S. Druckmann, and V. Jayaraman, "Ring attractor dynamics in the drosophila central brain," *Science*, vol. 356, no. 6340, pp. 849–853, 2017.

[17] D. Turner-Evans, S. Wegener, H. Rouault, R. Franconville, T. Wolff, J. D. Seelig, S. Druckmann, and V. Jayaraman, "Angular velocity integration in a fly heading circuit," *Elife*, vol. 6, p. e23496, 2017.

[18] D. B. Turner-Evans, K. T. Jensen, S. Ali, T. Paterson, A. Sheridan, R. P. Ray, T. Wolff, J. S. Lauritzen, G. M. Rubin, D. D. Bock, *et al.*, "The neuroanatomical ultrastructure and function of a biological ring attractor," *Neuron*, vol. 108, no. 1, pp. 145–163, 2020.

[19] A. Bicanski and N. Burgess, "Environmental anchoring of head direction in a computational model of retrosplenial cortex," *Journal of Neuroscience*, vol. 36, no. 46, pp. 11601–11618, 2016.

[20] K. J. Jeffery, H. J. Page, and S. M. Stringer, "Optimal cue combination and landmark-stability learning in the head direction system," *The Journal of physiology*, vol. 594, no. 22, pp. 6527–6534, 2016.

[21] M. G. Campbell, S. A. Ocko, C. S. Mallory, I. I. Low, S. Ganguli, and L. M. Giocomo, "Principles governing the integration of landmark and self-motion cues in entorhinal cortical codes for navigation," *Nature neuroscience*, vol. 21, no. 8, pp. 1096–1106, 2018.

[22] R. U. Muller and J. L. Kubie, "The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells," *Journal of Neuroscience*, vol. 7, no. 7, pp. 1951–1968, 1987.

[23] J. S. Taube, R. U. Muller, and J. B. Ranck, "Head-direction cells recorded from the postsubiculum in freely moving rats. ii. effects of environmental manipulations," *Journal of Neuroscience*, vol. 10, no. 2, pp. 436–447, 1990.

[24] J. P. Goodridge, P. A. Dudchenko, K. A. Worboys, E. J. Golob, and J. S. Taube, "Cue control and head direction cells.," *Behavioral neuroscience*, vol. 112, no. 4, p. 749, 1998.

[25] J. J. Knierim, H. S. Kudrimoti, and B. L. McNaughton, "Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells," *Journal of neurophysiology*, vol. 80, no. 1, pp. 425–446, 1998.

[26] F. Savelli, J. Luck, and J. J. Knierim, "Framing of grid cells within and beyond navigation boundaries," *Elife*, vol. 6, p. e21354, 2017.

[27] A. Terrazas, M. Krause, P. Lipa, K. M. Gothard, C. A. Barnes, and B. L. McNaughton, "Self-motion and the hippocampal spatial metric," *Journal of Neuroscience*, vol. 25, no. 35, pp. 8085–8096, 2005.

[28] L. Tcheang, H. H. Bülthoff, and N. Burgess, "Visual influence on path integration in darkness indicates a multimodal representation of large-scale space," *Proceedings of the National Academy of Sciences*, vol. 108, no. 3, pp. 1152–1157, 2011.

[29] M. C. Fuhs and D. S. Touretzky, "A spin glass model of path integration in rat medial entorhinal cortex," *Journal of Neuroscience*, vol. 26, no. 16, pp. 4266–4276, 2006.

[30] J. Conklin and C. Eliasmith, "A controlled attractor network model of path integration in the rat," *Journal of computational neuroscience*, vol. 18, pp. 183–203, 2005.

[31] Y. Burak and I. R. Fiete, "Accurate path integration in continuous attractor network models of grid cells," *PLoS computational biology*, vol. 5, no. 2, p. e1000291, 2009.

[32] P. Vafidis, D. Owald, T. D'Albis, and R. Kempter, "Learning accurate path integration in ring attractor models of the head direction system," *Elife*, vol. 11, p. e69841, 2022.

[33] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser, "Path integration and the neural basis of the'cognitive map'," *Nature Reviews Neuroscience*, vol. 7, no. 8, pp. 663–678, 2006.

[34] R. P. Jayakumar, M. S. Madhav, F. Savelli, H. T. Blair, N. J. Cowan, and J. J. Knierim, "Recalibration of path integration in hippocampal place cells," *Nature*, vol. 566, no. 7745, pp. 533–537, 2019.

[35] A. Shimbo, E.-I. Izawa, and S. Fujisawa, "Scalable representation of time in the hippocampus," *Science Advances*, vol. 7, no. 6, p. eabd7013, 2021.

[36] J. P. Goodridge and D. S. Touretzky, "Modeling attractor deformation in the rodent head-direction system," *Journal of neurophysiology*, vol. 83, no. 6, pp. 3402–3410, 2000.

[37] A. Samsonovich and B. L. McNaughton, "Path integration and cognitive mapping in a continuous attractor neural network model," *Journal of Neuroscience*, vol. 17, no. 15, pp. 5900–5920, 1997.

[38] S.-i. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological cybernetics*, vol. 27, no. 2, pp. 77–87, 1977.

[39] H. R. Wilson and J. D. Cowan, "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue," *Kybernetik*, vol. 13, no. 2, pp. 55–80, 1973.

[40] G. B. Ermentrout and J. D. Cowan, "A mathematical theory of visual hallucination patterns," *Biological cybernetics*, vol. 34, no. 3, pp. 137–150, 1979.

[41] S. A. Ocko, K. Hardcastle, L. M. Giocomo, and S. Ganguli, "Emergent elasticity in the neural code for space," *Proceedings of the National Academy of Sciences*, vol. 115, no. 50, pp. E11798–E11806, 2018.

[42] C. A. Fung, K. M. Wong, and S. Wu, "A moving bump in a continuous manifold: a comprehensive study of the tracking dynamics of continuous attractor neural networks," *Neural Computation*, vol. 22, no. 3, pp. 752–792, 2010.

[43] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[44] E. Salinas and L. F. Abbott, "A model of multiplicative neural responses in parietal cortex," *Proceedings of the national academy of sciences*, vol. 93, no. 21, pp. 11956–11961, 1996.

[45] T. Solstad, E. I. Moser, and G. T. Einevoll, "From grid cells to place cells: a mathematical model," *Hippocampus*, vol. 16, no. 12, pp. 1026–1031, 2006.

[46] Y. Yan, N. Burgess, and A. Bicanski, "A model of head direction and landmark coding in complex environments," *PLoS computational biology*, vol. 17, no. 9, p. e1009434, 2021.

[47] C. Lyu, L. Abbott, and G. Maimon, "Building an allocentric travelling direction signal via vector computation," *Nature*, vol. 601, no. 7891, pp. 92–97, 2022.

[48] S. Stringer, T. Trappenberg, E. Rolls, and I. Araujo, "Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells," *Network: Computation in Neural Systems*, vol. 13, no. 2, pp. 217–242, 2002.

[49] C. D. Brody, R. Romo, and A. Kepecs, "Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations," *Current opinion in neurobiology*, vol. 13, no. 2, pp. 204–211, 2003.

[50] B. McNaughton, L. Chen, and E. Markus, ""dead reckoning," landmark learning, and the sense of direction: a neurophysiological and computational hypothesis," *Journal of Cognitive Neuroscience*, vol. 3, no. 2, pp. 190–202, 1991.

[51] B. L. McNaughton, J. J. Knierim, and M. A. Wilson, *Vector encoding and the vestibular foundations of spatial cognition: neurophysiological and computational mechanisms*, p. 585–595. The MIT Press, 2020.

[52] B. L. McNaughton, C. A. Barnes, J. L. Gerrard, K. Gothard, M. W. Jung, J. J. Knierim, H. Kudrimoti, Y. Qin, W. Skaggs, M. Suster, *et al.*, "Deciphering the hippocampal polyglot: the hippocampus as a path integration system," *Journal of Experimental Biology*, vol. 199, no. 1, pp. 173–185, 1996.

[53] M. J. Milford, G. F. Wyeth, and D. Prasser, "Ratslam: a hippocampal model for simultaneous localization and mapping," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, vol. 1, pp. 403–408, 2004.

[54] J. D. Seelig and V. Jayaraman, "Feature detection and orientation tuning in the drosophila central complex," *Nature*, vol. 503, no. 7475, pp. 262–266, 2013.

[55] Y. E. Fisher, J. Lu, I. D'Alessandro, and R. I. Wilson, "Sensorimotor experience remaps visual input to a heading-direction network," *Nature*, vol. 576, no. 7785, pp. 121–125, 2019.

[56] S. S. Kim, A. M. Hermundstad, S. Romani, L. Abbott, and V. Jayaraman, "Generation of stable heading representations in diverse visual scenes," *Nature*, vol. 576, no. 7785, pp. 126–131, 2019.

[57] I. I. Low, A. H. Williams, M. G. Campbell, S. W. Linderman, and L. M. Giocomo, "Dynamic and reversible remapping of network representations in an unchanging environment," *Neuron*, vol. 109, no. 18, pp. 2967–2980, 2021.

[58] S. Wu, K. Hamaguchi, and S.-i. Amari, "Dynamics and computation of continuous attractors," *Neural computation*, vol. 20, no. 4, pp. 994–1025, 2008.

[59] A. Kutschireiter, M. A. Basnak, R. I. Wilson, and J. Drugowitsch, "Bayesian inference in ring attractor networks," *Proceedings of the National Academy of Sciences*, vol. 120, no. 9, p. e2210622120, 2023.

[60] A. Pouget, K. Zhang, S. Deneve, and P. E. Latham, "Statistically efficient estimation using population coding," *Neural computation*, vol. 10, no. 2, pp. 373–401, 1998.

[61] W.-H. Zhang, A. Chen, M. J. Rasch, and S. Wu, "Decentralized multisensory information integration in neural systems," *Journal of Neuroscience*, vol. 36, no. 2, pp. 532–547, 2016.

[62] R. Mehra, "On the identification of variances and adaptive kalman filtering," *IEEE Transactions on automatic control*, vol. 15, no. 2, pp. 175–184, 1970.

[63] A. Mohamed and K. Schwarz, "Adaptive kalman filtering for ins/gps," *Journal of geodesy*, vol. 73, pp. 193–203, 1999.

[64] K. J. Åström and B. Wittenmark, *Adaptive control*. Courier Corporation, 2013.

[65] W. Schultz, "Predictive reward signal of dopamine neurons," *Journal of neurophysiology*, 1998.

[66] J. R. Hollerman and W. Schultz, "Dopamine neurons report an error in the temporal prediction of reward during learning," *Nature neuroscience*, vol. 1, no. 4, pp. 304–309, 1998.

[67] J. F. Medina and S. G. Lisberger, "Links from complex spikes to local plasticity and motor learning in the cerebellum of awake-behaving monkeys," *Nature neuroscience*, vol. 11, no. 10, pp. 1185–1192, 2008.

[68] S. Kitazawa, T. Kimura, and P.-B. Yin, "Cerebellar complex spikes encode both destinations and errors in arm movements," *Nature*, vol. 392, no. 6675, pp. 494–497, 1998.

[69] W.-H. Zhang, H. Wang, A. Chen, Y. Gu, T. S. Lee, K. M. Wong, and S. Wu, "Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation," *Elife*, vol. 8, p. e43753, 2019.

[70] S. Cheadle, V. Wyart, K. Tsetsos, N. Myers, V. De Gardelle, S. H. Castañón, and C. Summerfield, "Adaptive gain control during human perceptual choice," *Neuron*, vol. 81, no. 6, pp. 1429–1441, 2014.

[71] W. Keung, T. A. Hagen, and R. C. Wilson, "A divisive model of evidence accumulation explains uneven weighting of evidence over time," *Nature Communications*, vol. 11, no. 1, p. 2160, 2020.

[72] B. W. Brunton, M. M. Botvinick, and C. D. Brody, "Rats and humans can optimally accumulate evidence for decision-making," *Science*, vol. 340, no. 6128, pp. 95–98, 2013.

[73] G. Ginosar, J. Aljadeff, L. Las, D. Derdikman, and N. Ulanovsky, "Are grid cells used for navigation? on local metrics, subjective spaces, and black holes," *Neuron*, 2023.

[74] J. Krupic, M. Bauza, S. Burton, C. Barry, and J. O'Keefe, "Grid cell symmetry is shaped by environmental geometry," *Nature*, vol. 518, no. 7538, pp. 232–235, 2015.

[75] H. Stensola, T. Stensola, T. Solstad, K. Frøland, M.-B. Moser, and E. I. Moser, "The entorhinal grid map is discretized," *Nature*, vol. 492, no. 7427, pp. 72–78, 2012.

[76] C. Barry, R. Hayman, N. Burgess, and K. J. Jeffery, "Experience-dependent rescaling of entorhinal grids," *Nature neuroscience*, vol. 10, no. 6, pp. 682–684, 2007.

[77] C. Barry, L. L. Ginzberg, J. O'Keefe, and N. Burgess, "Grid cell firing patterns signal environmental novelty by expansion," *Proceedings of the National Academy of Sciences*, vol. 109, no. 43, pp. 17687–17692, 2012.

[78] A. T. Keinath, R. A. Epstein, and V. Balasubramanian, "Environmental deformations dynamically shift the grid cell spatial metric," *Elife*, vol. 7, p. e38169, 2018.

[79] P. A. Hetherington and M. L. Shapiro, "Hippocampal place fields are altered by the removal of single visual cues in a distance-dependent manner.," *Behavioral neuroscience*, vol. 111, no. 1, p. 20, 1997.

[80] H. Eichenbaum, S. I. Wiener, M. Shapiro, and N. Cohen, "The organization of spatial coding in the hippocampus: a study of neural ensemble activity," *Journal of Neuroscience*, vol. 9, no. 8, pp. 2764–2775, 1989.

[81] M. Sato, K. Mizuta, T. Islam, M. Kawano, Y. Sekine, T. Takekawa, D. Gomez-Dominguez, A. Schmidt, F. Wolf, K. Kim, *et al.*, "Distinct mechanisms of over-representation of landmarks and rewards in the hippocampus," *Cell reports*, vol. 32, no. 1, 2020.

[82] M. Noorman, B. K. Hulse, V. Jayaraman, S. Romani, and A. M. Hermundstad, "Accurate angular integration with only a handful of neurons," *bioRxiv*, pp. 2022–05, 2022.

[83] P. Dayan and L. F. Abbott, *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.

[84] H. Khalil, *Nonlinear Systems*. Pearson Education, Prentice Hall, 2002.

[85] J. Guckenheimer and P. Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, vol. 42. Springer Science & Business Media, 2013.