

# A Controlled Attractor Network Model of Path Integration in the Rat

John Conklin

Systems Design Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1,  
Canada, jmlconkl@uwaterloo.ca

Chris Eliasmith

Philosophy and Systems Design Engineering, University of Waterloo, Waterloo,  
Ontario, N2L 3G1, Canada, celiasmith@uwaterloo.ca, (519) 888-4567 x2638.  
Corresponding author.

## Keywords

Path Integration; attractor network; neural control; hippocampus; subiculum

## Abstract

Cells in several areas of the hippocampal formation show place specific firing patterns, and are thought to form a distributed representation of an animal's current location in an environment. Experimental results suggest that this representation is continually updated even in complete darkness, indicating the presence of a path integration mechanism in the rat. Adopting the Neural Engineering Framework (NEF) presented by Eliasmith and Anderson (2003) we derive a novel attractor network model of path integration, using heterogeneous spiking neurons. The network we derive incorporates representation and updating of position into a single layer of neurons, eliminating the need for a large external control population, and without making use of multiplicative synapses. An efficient and biologically plausible control mechanism results directly from applying the principles of the NEF. We simulate the network for a variety of inputs, analyze its performance, and give three testable predictions of our model.

## Introduction

Neurons which fire maximally at specific locations in an environment (place cells) have been found in several areas within the hippocampal region of freely moving rats (O'Keefe and Dostrovsky 1971; Taube 1995; Sharp 1997). These areas include the entorhinal cortex, dentate gyrus, subiculum, parasubiculum, and hippocampus proper. It has been suggested that place cells contain a representation of the rat's instantaneous location on a two dimensional map of the environment, also known as a "place code" (McNaughton et al. 1996; Redish and Touretzky 1997). Place specific firing in hippocampal place cells has been found to persist during active locomotion, after the removal of visual cues (Muller and Kubie 1987; O'Keefe and Speakman 1987), and even in complete darkness (Quirk et al. 1990; Markus et al. 1994). This suggests that

these cells may be involved in a form of path integration, reflected in the ability of a rat to return directly to its initial location using only idiothetic cues. Sensitivity to vestibular motion signals has also been observed in place cell firing (Sharp et al. 1995), which could be used by a path integration mechanism to update the animal's internal representation of position.

The ability to perform path integration has been shown in a wide variety of animals (see Redish 1999 or Etienne and Jeffery 2004 for a review). Rats in particular have demonstrated path integration both while walking (Tolman 1948; Whishaw and Maaswinkel 1997; Alyan et al. 1997; Alyan and McNaughton 1999) and swimming (Benhamou 1997). However, path integration is only one element of the rat navigation system. The system also requires a mechanism for consolidating visual and external cues with path integration information (Redish and Touretzky 1997), and is theorized to contain mechanisms for planning trajectories (Frank et al. 2000; Frank et al. 2001), and other various functions. We focus specifically on path integration, and provide a model that can be incorporated into more comprehensive models of the rat navigation system.

To generate this model, we apply the Neural Engineering Framework (NEF) developed by Eliasmith and Anderson (2003). Our goals are to provide a better control mechanism than previous path integration models, to increase the biological plausibility without sacrificing model functionality, and to make predictions about the behaviour of neurons involved in path integration.

To develop this model, we first describe the relevant anatomy and physiology of path integration, including which areas of the brain are involved, and the characteristics of cells in these areas. We then review two past attractor models of path integration, and state our goals for improvement on these models. Next we apply the NEF to formulate the mathematical statement of our model. Following this derivation, we present the results of our numerical simulations, covering a wide range of input velocity signals. The resulting performance of the model is judged against previous models, and against observed biological phenomena. Finally, the predictions and significance of the model are discussed.

## **Methods**

### **Anatomy and Physiology of Path Integration**

Single cell data collected from experiments on locomoting rats has lead to the conclusion that not all place cells are created equal. Place cells in different parts of the hippocampal formation have been found to have different characteristic tuning curves, to vary in their spatial sensitivity, and are thought to encode space in fundamentally different ways. Place fields in hippocampus are observed to be randomly reset upon entering a novel environment (Thompson and Best 1989), while subicular place cells have been observed to stretch or scale their place fields so that their spatial firing properties are similar across different environments (Sharp 1997). Spatial cells in the entorhinal cortex (which has reciprocal connections to subiculum) also show similar firing patterns in different environments (Quirk et al. 1992). These data support the idea that the subic-

ular and entorhinal cells have the ability to transfer an abstract spatial representation from one environment to another, consistent with path integration.

There have been several hypotheses as to which areas of the brain actually perform path integration in the rat. One hypothesis is that the hippocampus itself contains the path integration mechanism (McNaughton et al. 1996). However this theory is weakened by experimental data showing that rats with hippocampal lesions can still perform path integration effectively (Alyan et al. 1997; Alyan and McNaughton 1999). Further, recordings from medial entorhinal cortex (MEC) during random exploration (Fyhn et al. 2004) have shown that the activity of a small number of superficial MEC neurons is sufficient to accurately reconstruct the trajectory of the rat, and that bilateral hippocampal lesions have no significant effect on the spatial information rate of these neurons. These results are consistent with the hypothesis that a path integration circuit external to the hippocampus contains a coherent representation of position.

Sharp (1997) alternatively proposed that path integration could be performed in the subiculum. Because each environment is represented by the same set of place cells, the same mechanism could be used to update the place representation as the rat moves through any environment. This single update mechanism could be either learned at a young age, or genetically prewired (Redish 1999). This would not be possible in hippocampus where there is no correlation between place fields in different environments (Thompson and Best 1989). However, the subiculum does not project directly to the hippocampus (Kohler 1986; Kohler 1988; Witter et al. 1989; Witter et al. 1990). Thus if the subiculum alone is performing path integration, it is unclear how this path integration signal could be used as input to the place cells observed in hippocampus.

This has lead Redish (1999) to propose that path integration is distributed across three structures, the subiculum, the parasubiculum, and the superficial layers of the entorhinal cortex. These areas are connected in a loop, and together they meet the five requirements proposed by Redish for structures which could theoretically perform path integration:

- they are collectively able to represent the position of the animal;
- they receive input from the head direction system;
- they receive information about self-motion from the motor and vestibular systems;
- they update the representation as the animal moves around the environment; and
- they send output to the area associated with the place code.

Redish's proposal addresses the shortcomings of the subiculum hypothesis, as the superficial entorhinal cortex projects directly to the hippocampus, so the path integrator would be able to update the hippocampal place code. Also, the parasubiculum receives input from postsubiculum, which contains head direction information, and both parietal and cingulate cortex, which contain motor and directional information. Thus the path integrator could receive the required velocity inputs to update its internal representation of position.

Our model is intended to be consistent with this characterization of the path integrator. However, as with past models (e.g. Samsonovich and McNaughton 1997), the functionality of our model is likely consistent with other anatomical implementations.

## Past Models

Many types of path integration models have been proposed. For a review and summary of the major models, see Redish (1999). Here, we review only attractor network models, as they have a high degree of neural plausibility, and have been the most successful to date (Redish and Touretzky 1997).

Attractor models of path integration can be viewed as a generalization of ring attractor networks, which have been used to model head direction in rats (Zhang 1996; Redish et al. 1996; Goodridge and Touretzky 2000). The ring attractor model consists of a gaussian bump of neural activity on a cyclic ring of neurons, which represent the current direction of the animal's head. This form of activity is the only dynamically stable state of the network, and from random initial conditions the network rapidly converges to a bump at a random location on the ring. Generalizing to two dimensions yields a plane attractor network, consisting of a gaussian hill of activation (sometimes called an activity packet) on a sheet of neurons, which represents the animal's current location in the environment. Various strategies have been adopted to control updating of the location of the activity packet based on motion signals.

Zhang (1996) has presented, but not simulated, such a generalization of his one-dimensional attractor model. However, there are a number of problems with his approach. His model uses homogeneous, non-spiking neurons. In addition, his dynamic mechanism for updating the activity packet requires multiplicative synapses to produce an asymmetry in the neuron connection weights, and is thus not generally thought to be biologically plausible (c.f. Mel 1994).

Samsonovich and McNaughton (1997) present and simulate an improved model which divides the system into two stages: a P-stage, which is a two-dimensional sheet of neurons representing the current position of the animal, and an I-stage which controls updating of the position. The I-stage consists of a stack of two dimensional sheets, each having asymmetric connections to the P-stage, with displacement corresponding to a particular direction of motion. The model implements translation of the activity packet by using the head direction to select a particular layer of the I-stage as active, while other layers remain silent. The activity in the P-stage is projected directly to the active I-layer, which in turn excites a section of the P-stage which is slightly offset from the current position, in the direction corresponding to the selected I-layer. This causes the activity packet to move in the selected direction, due to the attractor dynamics of the P-stage.

While this model reproduced several observations about hippocampal place fields, the translation mechanism seems cumbersome, requiring many times more neurons for the translation mechanism (the I-stage) than for the representation itself (the P-stage). Presented simulations range from 6 to 200 I layers, with each layer containing from 30,000 to 300,000 neurons. This model also uses homogeneous units (i.e., identical biophysical parameters), which is inconsistent with the heterogeneity observed in real

neurons (i.e., widely varying biophysical parameters that result in, for example, a variety of peak and background firing rates).

## Goals of our Path Integration Model

Redish and Touretzky (1997) present a conceptual framework for the rodent navigation system. In this framework they explain the functions and interactions of several subsystems, including a subsystem for processing sensory queues, a place code subsystem, and a path integration subsystem. We take this to be a good analysis and have designed our model so that it can function as a subsystem in this more comprehensive model of rodent navigation, meeting only the specific functional requirements of path integration. Redish and Touretzky (1998) simulate a larger network model to explain the role played by the hippocampus in solving a hidden platform water maze. However, their emphasis is on the simulation of the entire navigation system and the interactions between subsystems, so they do not explicitly simulate the path integration mechanism. Instead they refer to the above mentioned published models (Zhang 1996; Samsonovich and McNaughton 1997) for examples of mechanisms that can be used to simulate path integration. As discussed above, we feel that both of these models have shortcomings which can be improved upon.

So, our chief goal is to provide a biologically plausible path integration system which could be useful in modeling of the rodent navigation system. We aim to improve upon past efforts by avoiding use of multiplicative synapses, using realistically diverse neurons, and using a translation mechanism that is less costly from a biological perspective (i.e. one that can use fewer neurons to control movement of the activity packet), while maintaining or improving upon the functionality of past models.

Physiological and behavioral data leads to several constraints on our model. Our network must be able to maintain a stable representation of position, while exhibiting experimentally observed properties of place cells. Specifically, the neurons in the model must have diverse tuning properties, distributed over space and with a variety of background firing rates and sensitivities to change in position. Further, drift error in our model must be low enough to provide an explanation for experiments in which rats have been observed to successfully perform path integration (e.g. Alyan and McNaughton 1999). Finally, the model must be able to integrate movements in any direction.

One common characteristic of the attractor networks discussed above is that the construction of each network is somewhat *ad hoc*. There is no systematic process that has been followed to generate the required network behavior. Rather, deriving control mechanisms has required significant insight by the authors into the structure and dynamics of their networks, learning from the behaviour and shortcomings of previously constructed networks.

Eliasmith and Anderson (2003) have presented a theoretical framework for the construction of large-scale biologically plausible networks, the NEF. This framework gives a systematic procedure for implementing control of these networks, and has been applied to the control of spiking attractor networks including line attractors, ring attractors, cyclic attractors, and chaotic attractors (Eliasmith in press). By adopting this framework, we eliminate much of the guess work involved in getting our network to behave in the required manner. This gives our model a secondary purpose. Not only do

we provide a biologically plausible path integration system, but we also demonstrate the usefulness of the NEF.

## Theoretical Framework

The NEF describes how to create a population of neurons that can represent some value or function (in our case a two-dimensional gaussian bump), perform transformations on that representation (in our case a translation of the function), and incorporate the dynamics of the neural network over time (in our case stable control of translation on a two-dimensional map). The NEF can be summarized by the following three principles:

1. *Representation:* Neural representations are defined by a combination of non-linear encoding and optimal linear decoding
2. *Transformation:* Transformations of neural representations are functions of the variables that are represented by a population
3. *Dynamics:* Neural dynamics are characterized by considering neural representations as control theoretic state variables

We recall and elaborate on each principle where appropriate in the following derivation.

## Function representation

To meet the requirements of the model, we take the population to be able to represent two-dimensional gaussians, i.e. functions of the form:

$$x(\mu, \nu, t) = \exp\left(\left[-[\mu - \mu_0(t)]^2 - [\nu - \nu_0(t)]^2\right] / (2\sigma^2)\right) \quad (1)$$

where  $t$  is time,  $\mu$  and  $\nu$  are Cartesian coordinates on a plane normalized in both directions to the interval  $[-1, 1)$ ,  $(\mu_0(t), \nu_0(t))$  is the mean of the gaussian, and  $\sigma$  quantifies the width.

When defining the representation, it is important to determine what happens when the moving animal reaches the edge of the defined map. Three methods of dealing with this problem are discussed by Redish (1999). We have chosen to make our two-dimensional plane toroidal, (i.e. both axes are cyclical) as there is currently no strong biological evidence supporting one solution over another (although recent recordings from entorhinal cortex may suggest a toroidal representation, see Results).

To apply the NEF, it is useful to project the function representation onto an orthonormal basis and deal directly with the resulting vectors of coefficients, rather than the continuous functions of  $\mu$  and  $\nu$ . Because our spatial dimensions are cyclic, we have chosen the standard two-dimensional Fourier basis of bivariate sines and cosines. We can now identify the function  $x(\mu, \nu, t)$  by its vector of Fourier coefficients  $\mathbf{x}(t)$  where

$$x(\mu, \nu, t) = \sum_{m,n} x_{1,m,n}(t) \cos(\pi m \mu + \pi n \nu) + x_{2,m,n}(t) \sin(\pi m \mu + \pi n \nu). \quad (2)$$

The summations over  $m$  and  $n$  evenly tile the frequency space in a square grid about the origin. We use a finite number of basis functions, which will be the dimensionality  $D$  of the vector  $\mathbf{x}(t)$ . The number of basis functions needed is determined by the minimum variance of the set of gaussians we must represent. To represent gaussians with smaller variances, we require higher frequency components and thus a greater number of basis functions. To define our representations neurally, we refer to the first principle of the NEF:

**Representation:** Neural representations are defined by a combination of non-linear encoding and optimal linear decoding.

The following expression defines the encoding of  $\mathbf{x}(t)$  (the coefficient vector of the gaussian bump) by a network of spiking neurons:

$$\sum_n \delta(t - t_{in}) = G_i \left[ \alpha_i \langle \tilde{\phi}_i \cdot \mathbf{x}(t) \rangle + J_i^{bias} \right], \quad (3)$$

where  $G_i[\cdot]$  is a spiking nonlinearity which we take to be defined by the parameters of the  $i^{th}$  leaky integrate-and-fire (LIF) neuron. This function takes as its argument the soma input current on neuron  $i$  and produces a spike train for that neuron, where  $t_{in}$  is the time of the  $n^{th}$  spike from neuron  $i$ . The properties  $\alpha_i$  and  $J_i^{bias}$  represent the gain factor and constant background current for neuron  $i$ . The vector  $\tilde{\phi}_i$  is the preferred direction vector, or encoding vector, for neuron  $i$ , and determines the direction of maximal firing in the vector space. The  $\langle \cdot \rangle$  operator denotes the Euclidean inner product.

Because place cells in subiculum and entorhinal cortex exhibit roughly gaussian tuning curves, we choose the encoding vectors  $\tilde{\phi}_i$  to be Fourier coefficients of gaussians evenly distributed on the  $\mu, \nu$  plane. Noting that the inner product in (3) acts as a measure of similarity between a neuron's encoding vector  $\tilde{\phi}_i$  (corresponding to a particular gaussian on the plane) and the encoded vector  $\mathbf{x}(t)$  (corresponding to a gaussian at the rat's current location), the neurons will have roughly gaussian tuning curves, firing maximally when the encoded gaussian is coincident with the neuron's preferred direction vector. The maximum firing rate of each neuron is determined by randomly chosen LIF parameters, to match the observed heterogeneity of neuron firing rates. As the encoded gaussian is translated about the plane, a corresponding packet of roughly gaussian neural activity will translate as well. A similar idea has been used in the one-dimensional case to generate neurons with gaussian tuning curves in a model of working memory in macaque monkeys (Eliasmith and Anderson 2003; Eliasmith and Anderson 2001).

Having defined the encoding, we now define the decoded population estimate of  $\mathbf{x}(t)$ , recovered by optimal linear decoding:

$$\hat{\mathbf{x}} = \sum_{i,n} h(t - t_{in}) \phi_i^x. \quad (4)$$

Here, the output spike train of each neuron is convolved with a temporal filter  $h(t)$  defined by the post synaptic current (PSC), and weighted by the optimal linear decoders  $\phi_i^x$ . Using the PSC as a temporal filter gives similar information transfer characteristics

as optimal filters, with vastly improved biological plausibility (Eliasmith and Anderson 2003). The optimal linear decoders  $\phi_i^{\mathbf{x}}$  are the vectors which decode  $\mathbf{x}(t)$  with minimum mean square error (MSE) over the expected range of  $\mathbf{x}(t)$ , and are found by minimizing

$$E = \frac{1}{2} \left\langle [\mathbf{x} - \hat{\mathbf{x}}]^2 \right\rangle_{\mathbf{x}} \quad (5)$$

where  $\langle \cdot \rangle_{\mathbf{x}}$  denotes integration over the expected range of the vector  $\mathbf{x}$  (in our case, the coefficients of two-dimensional gaussians on the plane).

## Translation

We must now identify what happens to the coefficient vector  $\mathbf{x}(t)$  when the gaussian bump undergoes a translation  $x(\mu, \nu, t + \Delta t) = x(\mu + a, \nu + b, t)$ . If we substitute  $(\mu + a)$  for  $\mu$  and  $(\nu + b)$  for  $\nu$  in (2) and let  $\mathbf{x}(t + \Delta t)$  be the translated vector of Fourier coefficients, we can solve for  $\mathbf{x}(t + \Delta t)$  in terms of  $\mathbf{x}(t)$ , which shows us the effect of the translation in the frequency space. Using trigonometric identities, we obtain:

$$\begin{bmatrix} x_{1,m,n}(t + \Delta t) \\ x_{2,m,n}(t + \Delta t) \end{bmatrix} = \begin{bmatrix} \cos(\pi ma + \pi nb) & \sin(\pi ma + \pi nb) \\ -\sin(\pi ma + \pi nb) & \cos(\pi ma + \pi nb) \end{bmatrix} \begin{bmatrix} x_{1,m,n}(t) \\ x_{2,m,n}(t) \end{bmatrix} \quad (6)$$

This is recognizable as a two dimensional rotation of each Fourier coefficient pair. Interestingly then, any translation on the plane is equivalent to a rotation of each coefficient pair with coordinates  $(m, n)$  in the frequency space. Furthermore this rotation can be separated into two independent rotations, a rotation by  $\pi ma$  corresponding to translation by  $a$  in the  $\mu$  direction and a rotation by  $\pi nb$  corresponding to translation by  $b$  in the  $\nu$  direction. We can thus rewrite (6) as

$$\mathbf{x}(t + \Delta t) = \mathbf{R}_{\mathbf{ma}} \mathbf{R}_{\mathbf{nb}} \mathbf{x}(t). \quad (7)$$

Here  $\mathbf{R}_{\mathbf{ma}}$  and  $\mathbf{R}_{\mathbf{nb}}$  are  $D \times D$  matrices with  $2 \times 2$  rotation sub-matrices along the diagonal for each Fourier coefficient pair, such that left multiplication of  $\mathbf{x}(t)$  by  $\mathbf{R}_{\mathbf{ma}}$  or  $\mathbf{R}_{\mathbf{nb}}$  results in rotations by angles  $\pi ma$  or  $\pi nb$  of each Fourier coefficient pair.

## Dynamics of translation

To determine how to control and stabilize translation over time, consider the simplified case where the gaussian translates in one direction, say  $\mu$ . That is

$$\mathbf{x}(t + \Delta t) = \mathbf{R}_{\mathbf{ma}} \mathbf{x}(t). \quad (8)$$

This describes a system where the gaussian  $\mathbf{x}(t)$  translates a distance  $a$  in the  $\mu$  direction at each time step. We can think of  $a$  as an external linear velocity input for the  $\mu$  direction.

The dynamics of the system can be expressed as follows:



$$\dot{\mathbf{x}}(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{x}(t + \Delta t) - \mathbf{x}(t)}{\Delta t} \quad (9)$$

$$= \lim_{\Delta t \rightarrow 0} \frac{(\mathbf{R}_{\mathbf{m}a} - \mathbf{I})}{\Delta t} \mathbf{x}(t) \quad (10)$$

This equation requires that the rotation matrix  $\mathbf{R}_{\mathbf{m}a}$  be recalculated every time the velocity input  $a$  changes, which involves computation of many trigonometric functions at each time step, a task that is not at all straightforward in a neural network.

As an alternative method for controlling the speed of rotation, we set  $a = \delta$  in the above equation, where  $\delta$  is a very small constant. We now refer to the matrix as  $\mathbf{R}_{\mathbf{m}\delta}$  where multiplication by  $\mathbf{R}_{\mathbf{m}\delta}$  results in rotation of the Fourier coefficients in the state vector by  $\pi m\delta$ , which corresponds to a translation by  $\delta$  in the  $\mu$  direction. We can now use an external linear velocity input  $a(t)$  to essentially scale, or switch on and off, the translation:

$$\dot{\mathbf{x}}(t) = \frac{a(t)(\mathbf{R}_{\mathbf{m}\delta} - \mathbf{I})}{\Delta t} \mathbf{x}(t) \quad (11)$$

We have dropped the limit in the above equation and now consider  $\Delta t$  as the model time step, and  $a(t)$  as a time-varying linear velocity input normalized on  $[-1, 1]$ . The velocity of the bump varies directly with  $a(t)$ . For example, if  $a(t) = 0$  then  $\dot{\mathbf{x}}(t) = 0$  and the bump is static. If  $a(t) = 1$  then we recover equation (10) corresponding to maximum velocity translation in the  $\mu$  direction.

If we repeat this derivation for equation (7) in the case where the bump is translating in both the directions  $\mu$  and  $\nu$ , we arrive at a similar dynamic equation to (11):

$$\dot{\mathbf{x}}(t) = \frac{(\mathbf{R}_{\mathbf{m}\delta}\mathbf{R}_{\mathbf{n}\delta} - \mathbf{I})}{\Delta t} \mathbf{x}(t) \quad (12)$$

Here we face a problem because the product of the two rotation matrices makes it impossible to independently control movement in each direction by scaling. We use the following approximation to independently scale the two rotation matrices:

$$\dot{\mathbf{x}}(t) = \frac{a(t)(\mathbf{R}_{\mathbf{m}\delta} - \mathbf{I})}{\Delta t} \mathbf{x}(t) + \frac{b(t)(\mathbf{R}_{\mathbf{n}\delta} - \mathbf{I})}{\Delta t} \mathbf{x}(t) \quad (13)$$

$$= \frac{1}{\Delta t} [(\mathbf{R}_{\mathbf{m}\delta} - \mathbf{I})a(t) + (\mathbf{R}_{\mathbf{n}\delta} - \mathbf{I})b(t)] \mathbf{x}(t) \quad (14)$$

where  $b(t)$  is a linear velocity input in the  $\nu$  direction, and  $\mathbf{R}_{\mathbf{n}\delta}$  is a rotation matrix analogous to  $\mathbf{R}_{\mathbf{m}\delta}$ . For cases where the bump is translating in only one direction ( $a(t) = 0$  or  $b(t) = 0$ ), this equation reduces to equation (11) and the bump translates as required. For inputs where both  $a(t)$  and  $b(t)$  are non-zero, we can work backward from (14) to see that the approximation is a good estimate of the correct translation.

As an example, consider the case where  $a(t) = b(t) = 1$ , where we want to translate the bump positively in both directions. We can rewrite equation (14) as

$$\dot{\mathbf{x}}(t) = \lim_{\Delta t \rightarrow 0} \frac{(\mathbf{R}_{\mathbf{m}\delta} + \mathbf{R}_{\mathbf{n}\delta} - \mathbf{I}) \mathbf{x}(t) - \mathbf{x}(t)}{\Delta t}. \quad (15)$$

From the definition of the derivative, we can then infer

$$\mathbf{x}(t + \Delta t) = \mathbf{R}_{\mathbf{m}\delta} \mathbf{x}(t) + \mathbf{R}_{\mathbf{n}\delta} \mathbf{x}(t) - \mathbf{x}(t). \quad (16)$$

In the function space, this is a summation of three gaussians, one displaced by  $\delta$  in the  $\mu$  direction, one displaced by  $\delta$  in the  $\nu$  direction, and one scaled by  $-1$  in the previous location of the encoded gaussian. For small  $\delta$  relative to the width of the gaussians, this summation is a good approximation of a single gaussian displaced by  $\delta$  in both the  $\mu$  and  $\nu$  directions. Specifically our model uses gaussians of width  $1/3$  and a value of  $\delta = 1/5000$ , so the distortion of our encoded gaussian is negligible. Furthermore, there is no cumulative error, as any distortion is rapidly corrected by the attractor dynamics.

Thus, the final dynamic equation for our model is:

$$\dot{\mathbf{x}}(t) = \frac{1}{\Delta t} [(\mathbf{R}_{\mathbf{m}\delta} - \mathbf{I}) a(t) + (\mathbf{R}_{\mathbf{n}\delta} - \mathbf{I}) b(t)] \mathbf{x}(t) + \mathbf{u}(t) \quad (17)$$

We normalize the velocity inputs  $a(t)$  and  $b(t)$  such that  $a^2(t) + b^2(t) \leq 1$ , so that the maximum velocity of the bump is the same in all directions. The maximum velocity of the model is determined by the value of the small increment  $\delta$  and model time step  $\Delta t$ . The vector  $\mathbf{u}(t)$  is the frequency domain representation of any external input to the population, such as feedback from hippocampus (for error correction based on visual cues).

### Implementing translation

Observe that equation (17) requires multiplication of the state vector  $\mathbf{x}(t)$  by the velocity inputs  $a(t)$  and  $b(t)$ . To avoid the use of multiplicative synapses we refer to the second principle of the NEF:

**Transformation:** Transformations of neural representations are functions of the variables that are represented by a population. Any transformation can be determined using an alternately weighted linear decoding.

Accordingly, in order to compute the products  $a(t)\mathbf{x}(t)$  and  $b(t)\mathbf{x}(t)$ , we must include representations of the variables  $a(t)$  and  $b(t)$  in the population. This is accomplished by increasing the dimensionality of the state vector  $\mathbf{x}(t)$  to include the signals  $a(t)$  and  $b(t)$ .

In order for the population to represent these two scalar control signals as well as the current position, we must increase the dimensionality of our encoding vectors  $\tilde{\phi}_i$  accordingly. In general, for scalar representation, the encoder for each neuron is chosen to be either 1 or  $-1$  (the only directions in a one-dimensional space). Thus the final encoding vector  $\tilde{\phi}_i$  for a given neuron in this model's population consists of the  $D$  Fourier coefficients of a gaussian at that neuron's preferred location on the plane, and two elements which are chosen randomly to be either 1 or  $-1$ . As a result, our units are sensitive to both the position and velocity of the animal.

The required products of the Fourier elements of  $\mathbf{x}(t)$  and  $a(t)$  or  $b(t)$  can be computed by linear decoding, as suggested by the transformation principle. Specifically, we require two sets of alternate decoding weights, one to decode the function  $a(t)\mathbf{x}(t)$  (which is simply  $x_{D+1}\mathbf{x}_{1\dots D}$ ), and one to decode the function  $b(t)\mathbf{x}(t)$  (which is  $x_{D+2}\mathbf{x}_{1\dots D}$ ). The decoding is analogous to (4). For example, the decoded estimate of  $a(t)\mathbf{x}(t)$  can be expressed mathematically as:

$$f(\mathbf{x}(t)) = a(t)\mathbf{x}(t) \quad (18)$$

$$\hat{f}(\mathbf{x}(t)) = \sum_{i,n} h(t - t_{in})\phi_i^{a\mathbf{x}}, \quad (19)$$

where the function decoders  $\phi_i^{a\mathbf{x}}$  are found by minimizing the MSE of the decoded function estimate over the expected range of  $\mathbf{x}(t)$  and  $a(t)$ . The MSE is defined by

$$E = \frac{1}{2} \left\langle \left[ f(\mathbf{x}(t)) - \hat{f}(\mathbf{x}(t)) \right]^2 \right\rangle_{\mathbf{x},a}, \quad (20)$$

where  $\langle \cdot \rangle_{\mathbf{x},a}$  denotes integration over the expected ranges of  $\mathbf{x}(t)$  (all gaussians on the plane) and  $a(t)$  (the interval  $[-1, 1]$ ). The product  $b(t)\mathbf{x}(t)$  is decoded in an identical manner.

An advantage of this path integration model becomes evident from this discussion. Our translation mechanism is embedded directly in the population that represents the rat's current location. Thus, we require only a single layer of neurons to tile the spatial plane. This layer of neurons can both represent the current location of the animal with a packet of neural activity, and at the same time control the movement of this activity according to external velocity inputs. This is in stark contrast to the model presented by Samsonovich and McNaughton (1997), in which one layer of neurons (the P layer) represents the position of the rat, and many additional layers are required (the I layers) in order to control the translation of the activity packet.

## Neural Dynamics

Equation (17) defines the dynamics of our bump translation system. To derive a neural implementation of this dynamic system we refer to principle three of the NEF:

**Dynamics:** Neural dynamics are characterized by considering neural representations as control theoretic state variables. The dynamics of neurobiological systems can thus be analyzed using control theory.

Eliasmith and Anderson (2003) have used modern control theory to show that a dynamic system defined by

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (21)$$

can be implemented by a neural population through the encoding

$$\sum_n \delta(t - t_{in}) = G_i \left[ \alpha_i \left\langle \tilde{\phi}_i \cdot (h(t) * [\mathbf{A}' \mathbf{x}(t) + \mathbf{B}' \mathbf{u}(t)]) \right\rangle + J_i^{bias} \right] \quad (22)$$

where

$$\mathbf{A}' = \tau \mathbf{A} + \mathbf{I} \quad (23)$$

$$\mathbf{B}' = \tau \mathbf{B} \quad (24)$$

and  $\tau$  is the synaptic time constant of the units in the population.

Considering our dynamic equation (17) in the form of equation (21), we have the  $\mathbf{A}$  and  $\mathbf{B}$  matrices required to compute  $\mathbf{A}'$  and  $\mathbf{B}'$ , giving:

$$\mathbf{A}' = \frac{\tau}{\Delta t} [(\mathbf{R}_{m\delta} - \mathbf{I})a(t) + (\mathbf{R}_{n\delta} - \mathbf{I})b(t)] + \mathbf{I} \quad (25)$$

$$\mathbf{B}' = \tau \mathbf{I} \quad (26)$$

Substituting these matrices into (22) gives:

$$\sum_n \delta(t - t_{in}) = G_i \left[ \alpha_i \left\langle \tilde{\phi}_i \cdot (h(t) * \left[ \frac{\tau}{\Delta t} [(\mathbf{R}_{m\delta} - \mathbf{I})a(t) + (\mathbf{R}_{n\delta} - \mathbf{I})b(t) + 1] \mathbf{x}(t) + \tau \mathbf{u}(t) \right]) \right\rangle + J_i^{bias} \right] \quad (27)$$

Substituting the right hand side of equation (4) for  $\mathbf{x}(t)$ , the right hand side of equation (19) for  $a(t)\mathbf{x}(t)$ , and an analogous expression for  $b(t)\mathbf{x}(t)$  into equation (27) gives a complete description of our model:

$$\sum_n \delta(t - t_{in}) = G_i \left[ h(t) * \left( \sum_j w_{ij} a_j + \sum_k w_{ik} a_k + \sum_l w_{il} a_l + \alpha_i \left\langle \tilde{\phi}_i \cdot \tau \mathbf{u}(t) \right\rangle \right) + J_i^{bias} \right] \quad (28)$$

where

$$w_{ij} = \alpha_i \left\langle \tilde{\phi}_i \cdot \left( \frac{\tau}{\Delta t} (\mathbf{R}_{m\delta} - \mathbf{I}) \phi_j^{ax} \right) \right\rangle \quad (29)$$

$$w_{ik} = \alpha_i \left\langle \tilde{\phi}_i \cdot \left( \frac{\tau}{\Delta t} (\mathbf{R}_{n\delta} - \mathbf{I}) \phi_k^{bx} \right) \right\rangle \quad (30)$$

$$w_{il} = \alpha_i \left\langle \tilde{\phi}_i \cdot \phi_l^x \right\rangle \quad (31)$$

$$a_i = \sum_n h(t - t_{in}). \quad (32)$$

These equations complete the model derivation. Here  $\mathbf{u}(t)$  is the vector representation of the input to the population. The first  $D$  dimensions of  $\mathbf{u}(t)$  comprise the Fourier

decomposition of input to the function space, and can be used to control the initial location of neural activity, or to update or correct the position of the encoded gaussian (e.g. through hippocampal feedback of visual information). The last two dimensions of  $\mathbf{u}(t)$  are the linear velocity input components corresponding to the  $\mu$  and  $\nu$  directions.

The connection weights implied by this derivation have the form of two-dimensional center surround connections, reminiscent of the weights in one-dimensional ring attractor models (Redish et al. 1996; Zhang 1996; Goodridge and Touretzky 2000). Each neuron has strong local connections to neurons with nearby place fields, and weak negative connections to neurons with more distant place fields. Figure 1 shows the connection weights for a typical neuron in this model.

## Biological plausibility of the vector representation

It is important to note that the projection into the Fourier space is merely a convenient mathematical abstraction of the state of the network, which allows us to apply control theory to the dynamics of the system. This vector representation does not map directly onto neural activity, which still has the form of a gaussian activity packet. This is reflected in the connection weights (figure 1), which have the structure we would expect given past work with one-dimensional attractor networks that did not use any kind of orthonormal projection to define control. Further, the choice of the Fourier basis is not required to derive these weights. Choosing another orthonormal basis spanning the set of gaussians on the toroid (e.g. products of trigonometric functions) results in the same connection weights, given the same transformation in the function space. Choosing a different orthonormal basis is analogous to choosing a different vector space coordinate system, which is irrelevant to the function space transformation.

Similarly,  $\mathbf{u}(t)$  is an abstract representation of the neuron level soma inputs from external populations. That is, applying gaussian stimulation directly to the grid of neurons in our model (the type of input we would expect from hippocampal place cells) is equivalent to applying an input vector  $\mathbf{u}(t)$  which is the frequency domain representation of that same gaussian.

We thus emphasize that the use of a vector representation to define the control of the system does not detract from the biological plausibility of the network. This is most evident from the fact that the derived model exhibits biologically plausible tuning curves, connections, and behaviors.

## Results

For all model runs, the same population of 3969 spiking LIF neurons was used. In order to reproduce the broad tuning curves observed in subicular neurons by Sharp (1997), we estimate the variance of the encoded gaussian to be  $1/3$ . To accurately encode such a gaussian, a Fourier decomposition of 25 components is required. Thus our neurons encode a 27 dimensional space ( $D = 27$ ), where the first 25 dimensions encode the Fourier coefficients of a gaussian at the rat's current location on the plane, and the last two dimensions encode the rat's instantaneous velocity.

The encoding vectors  $\tilde{\phi}_i$  are chosen as Fourier coefficients of gaussians with variance  $1/3$  evenly distributed on the  $\mu, \nu$  plane in a  $63 \times 63$  grid. The last two dimensions of each  $\tilde{\phi}_i$  are chosen to be either 1 or  $-1$  in order to encode the velocity signal. We use a decaying exponential model of the PSC as our temporal filter:

$$h(t) = e^{-\frac{t}{\tau}}, \quad (33)$$

where  $\tau$  is the synaptic time constant of our neurons. For all runs we have used  $\tau = 5ms$ , and a model time step of  $0.1ms$ .

### Demonstration of a controlled attractor

When the network is initialized by being briefly stimulated with gaussian white noise, a stable packet of gaussian activity forms at a random location on the map (figure 2). The activity packet is clearly wider than that of the model by Samsonovich and McNaughton (1997). This is a result of our tuning of the units to match the broad place fields observed in subiculum (e.g. figure 2 in Sharp 1997). The activity packet also appears noisy, because the parameters for gain and bias current are chosen randomly (within a suitable range) for each neuron, mimicking the observed heterogeneity of neurons in this area.

We have tested the stability of the stationary bump after initial formation by simulating the network (with no external input) for 180 seconds. The activity packet retains its height (firing rates) and width (variance), with a small amount of drift in the mean of the bump (drift was equal to 8% of length of the plane over 180 seconds).

In addition, the packet retains its height and width when given a constant external velocity input. This results in a constant translation of the mean of the activity packet (see figures 3 and 4). These observations indicate the existence of a controlled stable attractor.

### Performance of the path integration model

In this section we characterize the dynamics of the control system and the accuracy of the path integration performed by the model. If the model is to be a useful path integrator, it must be able to change speed and direction as the vestibular input to the system changes.

In figure 4 the model is given a constant velocity input of 0.5 (one half maximum speed) in the negative  $\mu$  direction for two seconds, which is abruptly changed to 1 (maximum speed) in the negative  $\nu$  direction for a duration of 1 second. The figure shows the model output against the mathematically ideal path integration for the changing input. Over the run, the average drift is less than 5% of the total width of the plane (RMSE=2.73%). So, while there is some error throughout the run, the estimated position is a good approximation to the actual position of the animal. The error is small enough that it could easily be corrected by occasional, weak input from visual cues (see the subsequent section for details). Both the directional change and the velocity change are effectively captured by the model.

To compare our model with that of Samsonovich and McNaughton (1997), the model is given velocity input corresponding to circular motion of an animal. In figure 5 the simulated rat starts at the top of the circle, and moves clockwise around the circle at a constant speed. The decoded trajectory is noisy, due to the spiking representation of the encoded bump. Despite the noise, however, the model gives a good estimate of the simulated position over one circuit of the circle. Being able to integrate a circular path demonstrates that the model is able to integrate velocity inputs in any direction. However, since the end point of the trajectory does not meet the starting point, error due to drift may accumulate over further circuits of the circle, and would need to be corrected by visual input. Note that the drift in our 3969 neuron model after completing one circuit is dramatically less than that of the MPI model with 300,000 neurons per layer (see figure 10D in Samsonovich and McNaughton 1997). If we consider the location of the  $\nu$  coordinate ( $X$  in the MPI model) when the  $\mu$  coordinate ( $Y$  in the MPI model) has returned to its initial location as a measure of the drift after one cycle, then the MPI model is off by approximately 100% of the full diameter of the circle, while our model is off by only 11% of the circle’s diameter. Further, because the representational error of the model scales inversely with the number of neurons, the amount of drift in our model can be adjusted to match observed drift data (not currently available) by increasing or decreasing the number of units in the model. This could lead to predictions of how many neurons are actually involved in path integration.

To provide a summary of the error over the function space, we also plot the error surface obtained by integrating the error gradient on the  $\mu, \nu$  plane. The gradient is obtained via the encoding and decoding of a set of gaussians evenly tiling the plane, and taking the difference of the encoded and decoded means. The local minima of the error surface are the attractor points in the function space. We can see from figure 6 that there is a relatively even distribution of attractor points across the plane. This implies that the above simulations are not idiosyncratic, and that the stability of the bump for arbitrary velocity inputs will be similar for every location on the plane.

## Robustness to noise in the derived connection weights

We next demonstrate that the model does not require a precisely tuned connection weight matrix. This is significant as integrator circuits are often characterized as requiring particularly precise tuning of connection strengths (Seung et al. 2000; Koulakov et al. 2002; Goldman et al. 2003). To evaluate the robustness of our model, we randomly vary the connection weights. The noisy connection weight matrix is given by

$$\tilde{w}_{ij} = w_{ij} (1 + \eta_{ij}) \quad (34)$$

where  $w_{ij}$  is the derived connection weight between neurons  $i$  and  $j$ , and  $\eta_{ij}$  are random variables chosen from a gaussian distribution of mean 0 and variance  $\sigma^2$ .

We examine the effect of this noise on the representation by simulating the integration of a circular path (as above) for increasing values of the noise variance  $\sigma^2$  (figure 7). The root mean square error (RMSE) monotonically increases with noise variance. However, the performance of the  $\sigma^2 = 0.1$  and  $\sigma^2 = 0$  simulations are roughly comparable. Both trajectories are good approximations to the ideal, and the increase in

RMSE due to the addition of noise is less than 10%. As noise increased further, the drift in the simulations becomes more evident, and the RMSE increases more rapidly. However, even when  $\sigma^2 = 0.5$  (i.e. 50% noise) the decoded trajectory is a reasonable approximation to the circle, especially when compared to the performance of the MPI model (Samsonovich and McNaughton 1997).

### **The role of visual input in calibrating the path integrator**

Experiments by Gothard et al. (1996) have demonstrated two ways in which visual cue information and path integration interact competitively to update the rat's internal representation of its location. For small mismatches between visual cues and path integration, the internal representation was observed to shift smoothly through all intervening states (at an accelerated speed) until the internal representation 'caught up' with the location indicated by the visual input. For large mismatches, the internal representation jumped abruptly from its present state to the state indicated by visual input, and the intervening states were skipped.

We assume that visual input reaches our model in the form of gaussian stimulation centred on the location suggested by a visual cue (this input could be projected to subiculum from hippocampus), and show that our model supports both of these correction mechanisms.

First, when the visual stimulation is relatively weak, we observe a smooth acceleration of the translating activity packet toward the location of the stimulation (figure 8). This can be compared to the case where the packet is translating at the same velocity, but without external stimulation (figure 3). The external stimulation clearly accelerates the activity toward the location of stimulation.

Second, when the stimulation is relatively strong, we see the activity packet jump abruptly from its previous position on the plane to the location of the stimulation (figure 9). These effects demonstrate that sensory information can be used appropriately to correct error in our path integration model.

### **Reproduction of the effects of theta oscillations**

As reported by O'Keefe and Recce (1993) (see also Skaggs et al. 1996), hippocampal place cell firing is modulated by the hippocampal theta rhythm, a background oscillation typically in the range of 7-12 Hz. The observed effect of this modulation (stated in terms of neural activity on a two dimensional map) is that the activity packet moves ahead of the rat's actual location on the map and increases in width over the course of each theta cycle, before returning to the actual location and initial width at the beginning of the next cycle (Samsonovich and McNaughton 1997). Theta modulated place cells are also found in the subiculum and superficial entorhinal cortex (Sharp 1997; Frank et al. 2001).

We qualitatively reproduce these effects by introducing a global excitatory oscillation to the soma input of all neurons, at a frequency of 10 Hz. This theta current oscillates from zero to 2 nanoamperes. The introduction of this oscillation into the activity of all units causes the width and speed of the activity packet to oscillate with the theta rhythm (figure 10).



## Directionality of place cells

Radial maze experiments (McNaughton et al. 1983; Muller et al. 1994; Markus et al. 1995) demonstrate that most hippocampal place cells show a clear directional selectivity when the rat travels a linear path in a radial maze. These cells fire at a high rate when the rat passes through their place field in one direction, and fire more slowly (or are silent) when the rat returns in the opposite direction. The remaining cells show only positional sensitivity. In the same experiments it was found that most hippocampal place cells appear non-directional when the rat randomly forages in an open cylindrical environment.

To compare the directionality of our model place cells to those observed in hippocampus, we perform a simulation where the rat moves in a straight line at maximum speed for one second, and then returns along the same path in the opposite direction. We observe that most of our units have higher firing rates in one direction, and have lower firing rates (or are silent) in the other. We also find units which appear to fire independently of the direction of travel. Figure 11 shows firing for one neuron of each type in the simulation. Thus neurons in our model have directional firing properties similar to those of hippocampal place cells during linear movements.

This similarity may not extend to random foraging, where hippocampal place cells appear largely non-directional. Because our encoding of space is independent of the geometry of the environment, we predict that neurons involved in path integration will have similar velocity sensitivity in both linear path and random foraging experiments. This does not mean, however, that the observed directionality of these cells will be identical in these two environments. If the rat has different velocity profiles within place fields in different environments, the observed directional sensitivity will differ as well. For example, the rat's average velocity when passing through a place field in an arm of a radial maze is likely to be high because the rat usually travels directly from the center of the maze to the food site at the end of the arm (without stopping on the way) and back. The average velocity of the rat while visiting a given place field in the random foraging experiments is likely to be lower, as the rat stops at random locations in the environment to eat food pellets. Because our model encodes velocity, rather than independent representations of speed and direction (in the latter case direction *cannot* affect firing when speed is zero), the velocity profile of the rat must be taken into account when analyzing directionality in the data.

## Reproduction of subicular firing rate maps

We next demonstrate that the tuning curves of neurons in our model are similar to those observed experimentally of subicular place cells. Sharp (1997) has illustrated firing rate maps for subicular and hippocampal place cells in both cylindrical and square environments. These maps are generated by recording the location and time of each spike from a given cell, while the rat randomly forages for pellets during a 20 minute recording session. The number of spikes was then averaged over the length of the session to give the firing rate at each location in the environment. We have simulated this random foraging by using band-limited white noise as our head direction input,

and having our rat randomly explore its environment until all sections of the map have been visited. Firing rate maps were then calculated in the same manner as Sharp.

Figure 12 shows a comparison of the observed tuning curves with those generated by our model. Notice that the model tuning curves are not perfect gaussians, but are distorted due to the random path of the rat (which affects cell firing due to directional sensitivity), and the noise introduced by neural spiking. These noisy model tuning curves are qualitatively similar to those observed experimentally, as we require.

### **Multiple place fields in medial entorhinal cortex**

Recent data from Fyhn et al. (2004) shows strong place selectivity in neurons in the dorsolateral and intermediate MEC of rats exploring a rectangular environment. Cells in dorsolateral MEC showed sharp coherent place fields with multiple peaks (median number of 4). As seen in figure 2B of that paper, the multiple place fields for each cell form a relatively even tiling of the environment.

This tiling is consistent with the toroidal representation of the environment implemented in our model. If the encoded spatial plane were normalized to one quarter of the area of the rectangular chamber, we would expect to see precisely the type of sharp multi-peaked tuning curves observed in this experiment. Simulated place fields would appear narrow as a result of normalizing our representation to cover a smaller area, and multiple firing fields would result naturally as the rat travels from one edge of the toroid to the other.

Fyhn et al. (2004) also report that most cells in the intermediate MEC have broader, less coherent place fields, and are less likely to have multiple peaks. These cells seem to be tuned similarly to the subicular place cells discussed above. This raises the interesting possibility that there could be multiple attractor maps at different spatial resolutions (high resolution in dorsolateral MEC, low resolution in intermediate MEC, subiculum, and other areas), simultaneously encoding position of the rat. While we do not explore this idea here, the decoding of position from multiple attractor maps using the NEF would be a straightforward generalization of our model.

## **Discussion**

The analysis of our model has led to three testable predictions. Firstly, we predict that neurons involved in path integration must be sensitive to the instantaneous velocity of the animal. Secondly, we predict that the relative spatial relationship between place fields of cells involved in path integration will be the same or similar in different environments. Thirdly, we predict that the head direction signal cannot be used as a direct input signal to the path integrator, and is used instead to modulate an allocentric velocity input.

### **Velocity sensitive neurons**

The control mechanism of our model requires that the population performing path integration not only encode a representation of the rat's position on the plane, but also

a two-dimensional in-plane velocity vector. Thus our model predicts the existence of place cells in the subiculum, parasubiculum, superficial entorhinal cortex loop that will show sensitivity to the instantaneous velocity of the animal. Further we predict that this sensitivity will persist regardless of the rat's environment and trajectory, in contrast to the observed behavior of hippocampal place cells. As discussed above, it is important to consider that this is a prediction of velocity sensitivity (not just direction sensitivity), i.e. we predict that the directional sensitivity of neurons will not be evident when the rat is moving slowly relative to its maximum speed.

### **Spatial relationship between place fields in differing environments**

The model we have derived requires that the relative spatial location of the place fields of cells performing path integration must be constant, independent of the environment. That is, place fields of path integration cells will always have the same position relative to each other, although the plane of place fields may undergo translation, rotation, or scaling.

As discussed above, it is observed experimentally that hippocampal place fields are randomly reset upon entering a novel environment, while the place fields of cells in subiculum and entorhinal cortex are similar across different visually similar environments. As a result, this prediction is consistent with the hypothesis that subiculum and entorhinal cortex are involved in path integration. Note however that the data is from experiments where the differing environments were visually similar (Sharp 1997; Quirk et al. 1992). Experiments showing similar relative spatial location of place fields in visually different environments would further support this hypothesis.

### **Role of the head direction system and availability of velocity input**

Previous schemes for modelling path integration in the rat have used as their directional input a combination of a head direction signal, and a self motion signal which is proportional to the speed of the animal in the direction it is currently facing (Samsonovich and McNaughton 1997; Redish 1999; Redish and Touretzky 1997).

Our derivation has led to a different requirement for velocity input. Our model requires the input of an allocentric velocity vector ( $\frac{d\mu}{dt}, \frac{d\nu}{dt}$ ). There is evidence to suggest that the monkey vestibular system has the ability to determine an unambiguous translational acceleration relative to the current direction of the head based on inputs from the otoliths and semi-circular canals (Angelaki et al. 1999). Experiments by Hess and Dieringer (1991) suggest that rats have the same ability. Given the presence of a translational acceleration signal in the vestibular system, a similar velocity signal relative to current head direction could easily be computed by integration, which is thought to be a common computation across brain areas (Seung et al. 2000; Seung 1996; Askay et al. 2001; Askay et al. 2000; Douglas et al. 1995). Neurons sensitive to direction of motion, speed, and turning angle are known to exist in posterior parietal cortex, which is reciprocally connected to posterior cingulate cortex (Chen et al. 1994). Further, Page and Duffy (2003) have found that neurons in the dorsal segment of the medial superior temporal sulcus (MSTd) of Rhesus monkeys are sensitive to rotational and translational self-motion even in darkness, and have theorized these cells to be involved in a path

integration mechanism. We hypothesize that posterior parietal cortex contains a representation of the animal's velocity relative to current head direction (likely in MSTd), and that this signal is projected to posterior cingulate cortex, where the global head direction signal from postsubiculum is used to resolve the velocity vector into allocentric map components. Posterior cingulate cortex is also connected to motor cortex (Cho and Sharp 2001) which supports the idea that it may be involved in path integration or navigational motor planning. We thus propose that our required velocity input reaches parasubiculum via posterior cingulate cortex.

This proposal does not discount the important role of the head direction system. Specifically we consider the purpose of the head direction system to be the alignment of head direction relative velocities to global map coordinates  $\mu$  and  $\nu$ . It is thus essential for modulating the velocity input to our path integration system, rather than providing direct input.

It should be possible to test which form of velocity input the path integration system receives. According to the schemes proposed in Samsonovich and McNoughton (1997) and Redish and Touretzky (1997) the path integration system can only function correctly if the animal moves in the direction its head is currently facing. However, our model predicts that path integration should function correctly regardless of correlation between the rat's head direction and its direction of motion. Recordings from a rat which moves in darkness in a direction other than the one it is currently facing (e.g. the rat could move forward while its head was angled to one side) would provide evidence to support one theory or the other. If the place code in hippocampus was destroyed or distorted, this would support head direction as a direct input to the path integration system. If the place code was maintained, this would suggest that the path integration system may receive a global map relative velocity input such as the one proposed here.

## Conclusions

Our numerical simulations confirm that an attractor map implementation of path integration is possible without the use of a large external control population or multiplicative synapses. We have demonstrated that the control mechanism for translating the activity packet of neural activity can in fact be incorporated into the same population which stores the activity packet. This results in a single population which can act as a path integration subsystem in a larger model of rat navigation, such as (Redish and Touretzky 1997).

We have also demonstrated how the Neural Engineering Framework can provide a systematic and insightful solution to a significant problem in attractor dynamics and the rat navigation system. As discussed, this particular application has generated three testable predictions regarding path integration. Because the framework is general, and provides a useful method for dealing with complex control problems involving attractor networks, it should prove useful for analyzing other neural systems as well.

## Acknowledgements

We are grateful to Neil Burthiergess, David Redish, Ray Singh, Phil Smith, and Bryan Tripp for helpful discussions. We also thank our two anonymous reviewers for thithierer constructive feedback. This work is supported by the National Science and Engineering Research Council of Canada (261453-03), the Canadian Foundation for Innovation (3358401), and the Ontario Innovation Trust (3358501).

## References

- Alyan, S. H. and G. L. McNaughton (1999). Hippampectomized rats are capable of homing by path integration. *Behavioural Neuroscience* 113, 19–31.
- Alyan, S. H., B. M. Paul, E. Ellsworth, R. D. White, and G. L. McNaughton (1997). Is the hippocampus required for path integration? *Society for Neuroscience Abstracts* 23, 504.
- Angelaki, D. E., M. Q. McHenry, J. D. Dickman, S. D. Newlands, and B. J. M. Hess (1999). Computation of inertial motion: neural strategies to resolve ambiguous otolith information. *Journal of Neuroscience* 19, 316–327.
- Askay, E., R. Baker, H. S. Seung, R. Baker, and D. Tank (2000). Anatomy and discharge properties of pre-motor neurons in the goldfish medulla that have eye-position signals during fixations. *Journal of Neurophysiology* 84, 1035–1049.
- Askay, E., G. Gamkrelidze, H. S. Seung, R. Baker, and D. Tank (2001). In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nature Neuroscience* 4, 184–193.
- Benhamou, S. (1997). Path integration by swimming rats. *Animal Behaviour* 54, 321–327.
- Chen, L. L., L. H. Lin, C. A. Barnes, and B. L. McNaughton (1994). Head-direction cells in the rat posterior cortex: II. Contributions of visual and ideothetic information to the directional firing. *Experimental Brain Research* 101, 24–34.
- Cho, J. and P. Sharp (2001). Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. *Behavioural Neuroscience* 115, 3–25.
- Douglas, R. J., C. Koch, M. Mahowald, K. A. C. Martin, and H. H. Suarez (1995). Recurrent excitation in neocortical circuits. *Science* 269, 981–985.
- Eliasmith, C. (in press). A unified approach to building and controlling spiking attractor networks. *Neural computation*.
- Eliasmith, C. and C. H. Anderson (2001). Beyond bumps: Spiking networks that store sets of functions. *Neurocomputing* 38, 581–586.
- Eliasmith, C. and C. H. Anderson (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.
- Etienne, A. S. and K. J. Jeffery (2004). Path integration in mammals. *Hippocampus* 14, 180–192.

- Frank, L. M., E. N. Brown, and M. Wilson (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* 27, 169–178.
- Frank, L. M., E. N. Brown, and M. A. Wilson (2001). A comparison of the firing properties of putative excitatory and inhibitory neurons from ca1 and the entorhinal cortex. *Journal of Neurophysiology* 86, 2029–2040.
- Fyhn, M., S. Molden, M. Witter, E. Moser, and M. Moser (2004). Spatial representation in the entorhinal cortex. *Science* 305, 1258–1264.
- Goldman, M. S., J. H. Levine, G. Major, D. W. Tank, and H. S. Seung (2003). Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cerebral cortex* 13, 1185–1195.
- Goodridge, J. P. and D. S. Touretzky (2000). Modeling attractor deformation in the rodent head-direction system. *Journal of Neurophysiology* 83, 3402–3410.
- Gothard, K. M., W. E. Skaggs, and B. L. McNaughton (1996). Dynamics of mismatch correlations between stimuli to spatial correlations between attractors in neural networks. *Journal of Neuroscience* 16, 8027–8040.
- Hess, B. J. and N. Dieringer (1991). Spatial organization of linear vestibuloocular reflexes of the rat: responses during horizontal and vertical linear acceleration. *Journal of Neurophysiology* 66, 1805–1818.
- Kohler, C. (1986). Intrinsic connections of the retrohippocampal region in the rat brain. II. The medial entorhinal area. *Journal of Comparative Neurology* 246, 149–169.
- Kohler, C. (1988). Intrinsic connections of the retrohippocampal region in the rat brain. III. The lateral entorhinal area. *Journal of Comparative Neurology* 271, 208–228.
- Koulakov, A. A., S. Raghavachari, A. Kepecs, and J. E. Lisman (2002). Model for a robust neural integrator. *Nature Neuroscience* 5, 775–782.
- Markus, E. J., C. A. Barnes, B. L. McNaughton, V. L. Gladden, and W. E. Skaggs (1994). Spatial information content and reliability of hippocampal CA1 neurons: effects of visual input. *Hippocampus* 4, 410–421.
- Markus, E. J., Y. Qin, B. Leonard, W. E. Skaggs, B. L. McNaughton, and C. A. Barnes (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *Journal of Neuroscience* 15, 7079–7094.
- McNaughton, B. L., C. A. Barnes, J. L. Gerrard, K. Gothard, M. W. Jung, J. J. Knierim, H. Kudrimoti, Y. Qin, W. E. Skaggs, M. Suster, and K. L. Weaver (1996). Deciphering the hippocampal polygot: The hippocampus as a path integration system. *Journal of Experimental Biology* 199, 173–185.
- McNaughton, B. L., C. A. Barnes, and J. O’Keefe (1983). The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Experimental Brain Research* 52, 41–49.
- Mel, B. (1994). Information processing in dendritic trees. *Neural Computation* 6, 1031–1085.

- Muller, R. U., E. Bostock, J. S. Taube, and J. L. Kubie (1994). On the directional firing properties of hippocampal place cells. *Journal of Neuroscience* 14, 7235–7251.
- Muller, R. U. and J. L. Kubie (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience* 7, 1935–1950.
- O’Keefe, J. and J. Dostrovsky (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. *Brain Research* 34, 171–175.
- O’Keefe, J. and M. L. Recce (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317–330.
- O’Keefe, J. and A. Speakman (1987). Single unit activity in the rat hippocampus during a spatial memory task. *Experimental Brain Research* 68, 1–27.
- Page, W. K. and C. J. Duffy (2003). Heading representation in MST: sensory interactions and population encoding. *Journal of Neurophysiology* 89, 1994–2013.
- Quirk, G. J., R. U. Muller, and J. L. Kubie (1990). The firing of hippocampal place cells in the dark depends on the rat’s previous experience. *Journal of Neuroscience* 10, 2008–2017.
- Quirk, G. J., R. U. Muller, J. L. Kubie, and J. B. R. Jr. (1992). The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *Journal of Neuroscience* 12, 1945–1963.
- Redish, A. D. (1999). *Beyond the cognitive map*. Cambridge, MA: MIT Press.
- Redish, A. D., A. N. Elga, and D. S. Touretzky (1996). A coupled attractor model of the rodent head direction system. *Network: Computation in Neural Systems* 7, 671–685.
- Redish, A. D. and D. S. Touretzky (1997). Cognitive maps beyond the hippocampus. *Hippocampus* 7, 15–35.
- Redish, A. D. and D. S. Touretzky (1998). The role of the hippocampus in solving the morris water maze. *Neural Computation* 10, 73–111.
- Samsonovich, A. and B. L. McNaughton (1997). Path integration and cognitive mapping in a continuous attractor model. *Journal of Neuroscience* 17, 5900–5920.
- Seung, H. S. (1996). How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences, USA* 93, 13339–13344.
- Seung, H. S., D. Lee, B. Reis, and D. Tank (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26, 259–271.
- Sharp, P. E. (1997). Subicular cells generate similar spatial firing patterns in two geometrically and visually distinctive environments: Comparison with hippocampal place cells. *Behavioral Brain Research* 85, 71–92.

- Sharp, P. E., H. T. Blair, D. Etkin, and D. B. Tzvetos (1995). Influences of vestibular and visual motion information on the spatial firing patterns of hippocampal place cells. *Journal of Neuroscience* 15, 173–189.
- Skaggs, W. E., B. L. McNaughton, M. A. Wilson, and C. A. Barnes (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6, 149–172.
- Taube, J. S. (1995). Place cells recorded in the parasubiculum in freely moving rats. *Hippocampus* 5, 569–583.
- Thompson, L. T. and P. J. Best (1989). Place cells and silent cells in the hippocampus of freely-behaving rats. *Journal of Neuroscience* 9, 2382–2390.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review* 55, 189–208.
- Whishaw, I. Q. and H. Maaswinkel (1997). Absence of dead reckoning in hippocampal rats. *Society for Neuroscience Abstracts* 23, 1839.
- Witter, M. P., H. J. Groenewegen, F. H. L. da Silva, and A. H. M. Lohman (1989). Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Progress in Neurobiology* 33, 161–253.
- Witter, M. P., R. H. Ostendorf, and H. J. Groenewegen (1990). Heterogeneity in the dorsal subiculum of the rat. Distinct neuronal zones project to different cortical and subcortical targets. *European Journal of Neuroscience* 2, 718–725.
- Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience* 16, 2112–2126.



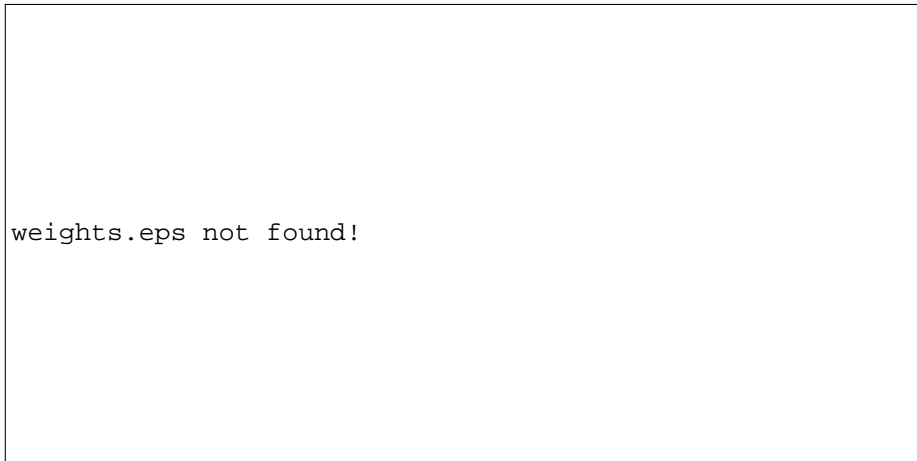


Figure 1: Connection weight matrix for a typical neuron in the derived model. Neurons are arranged in a square matrix by the locations of their place fields on the plane.

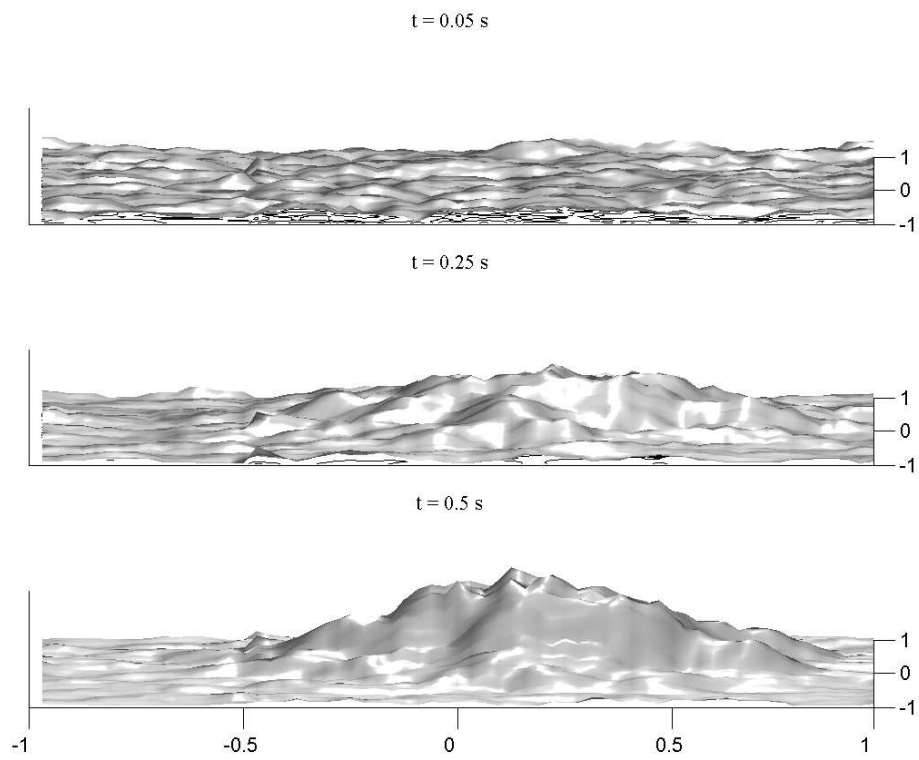


Figure 2: Self-focusing of network activity. For these and all subsequent activity plots, neurons are arranged on the plane by location of their place fields. Firing rates of spiking neurons have been smoothed on the plane using a 5 point moving average.

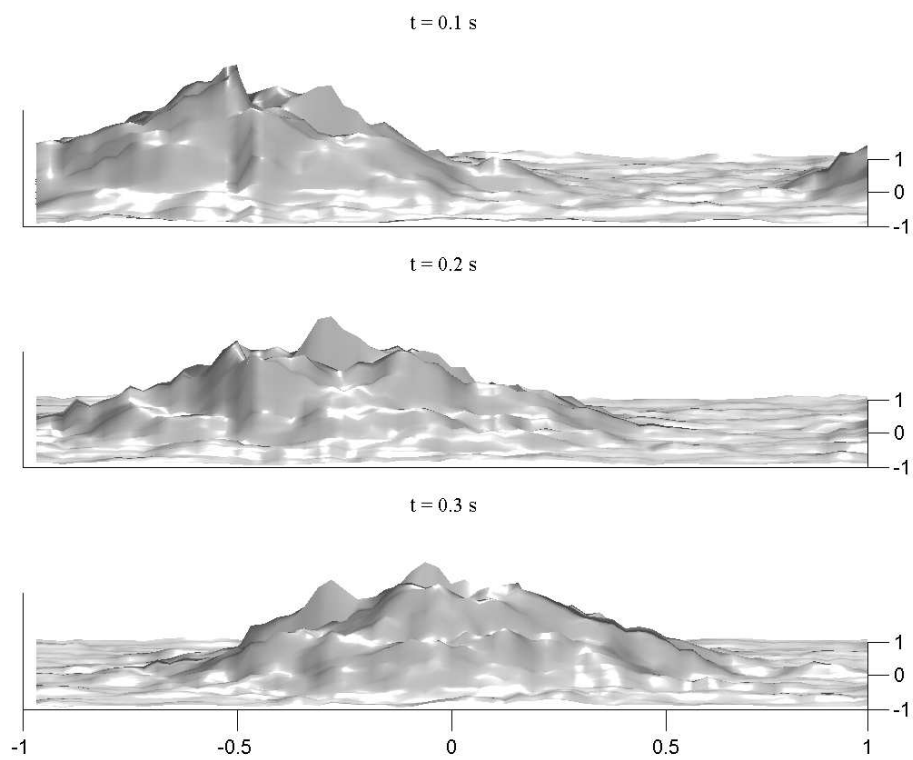


Figure 3: Propagation of the activity packet at a constant velocity. Velocity input is constant and to the right, at one half maximum speed.

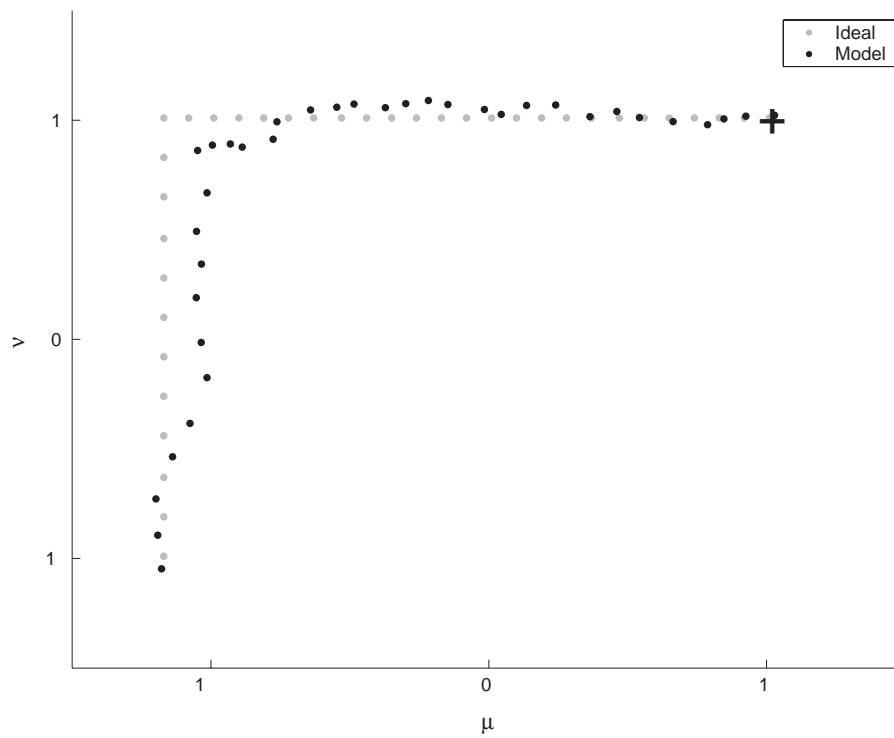


Figure 4: Path integration of velocity input with changing speed and direction. Ideal output (grey dots) is the actual position of the simulated rat. Model output (black dots) is the decoded mean of the gaussian represented by the network. The rat starts in the upper right (marked by the cross). Dots are separated in time by 100 ms.

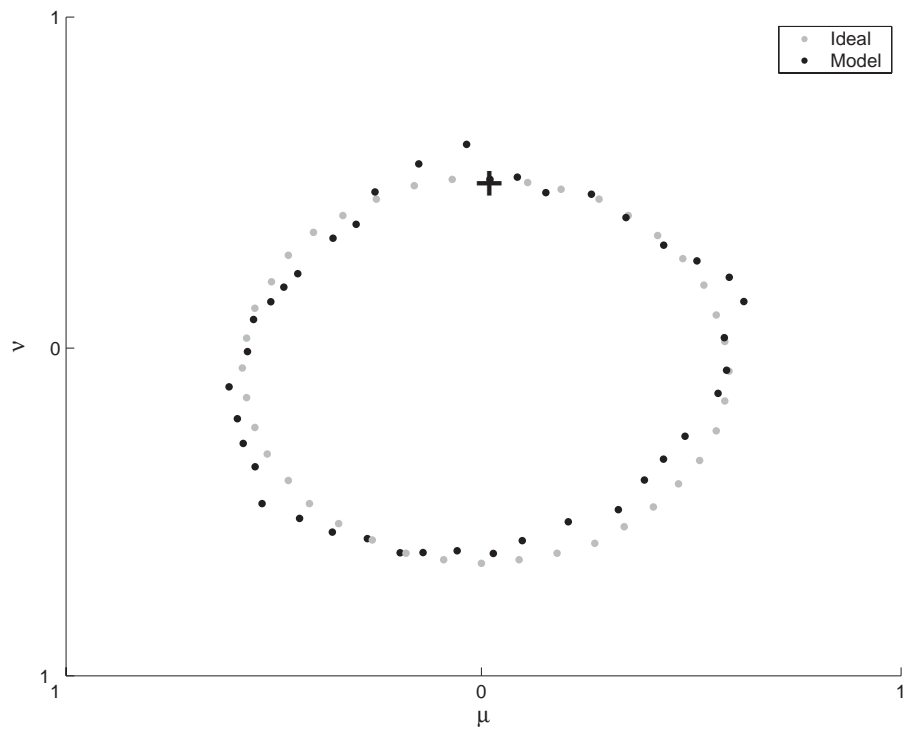


Figure 5: Path integration of circular motion. Ideal output (grey dots) is a perfectly circular trajectory. Model output (black dots) is the decoded mean of the gaussian represented by the network. The rat starts at the top of the circle (marked by the cross) and moves in the clockwise direction. Dots are separated in time by 50 ms.

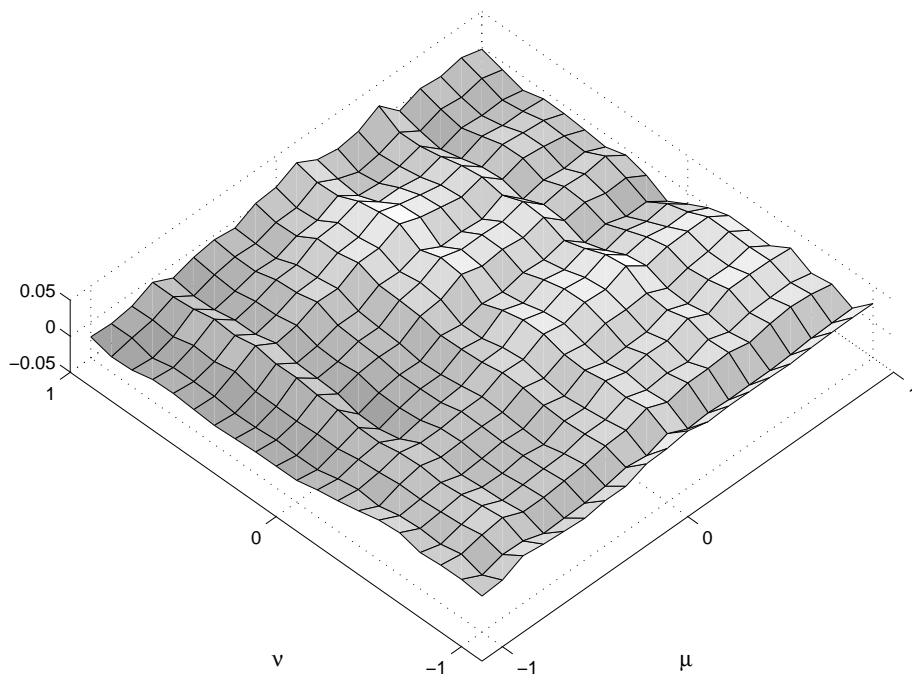


Figure 6: Error surface of the function representation. Local minima represent attractor points in the function space. Drift of the encoded gaussian is proportional to the slope of the surface.

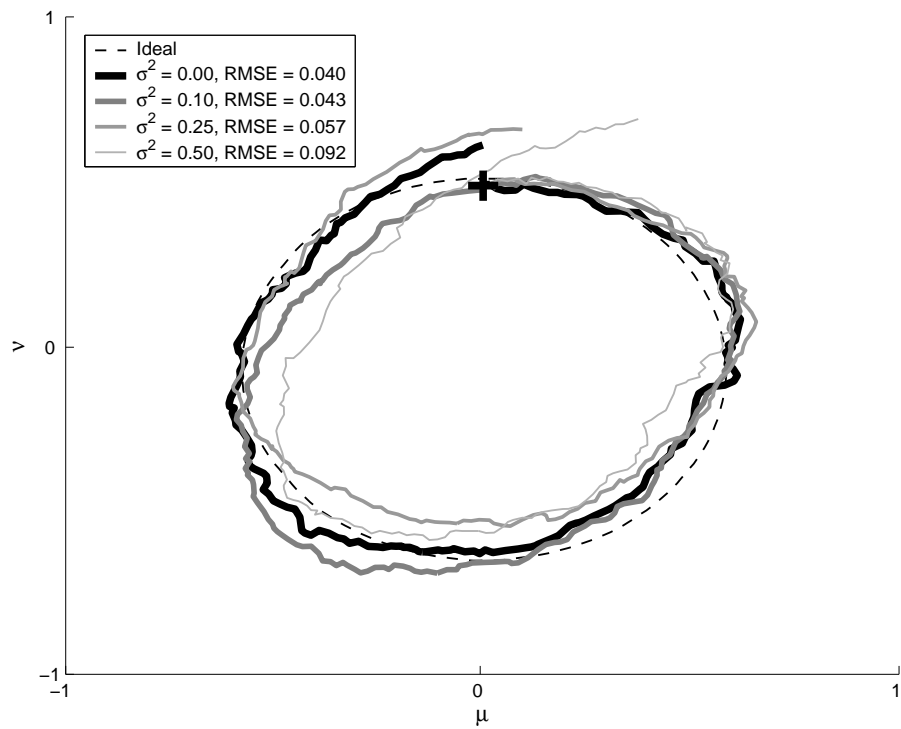


Figure 7: Integration of a circular path after addition of noise to the connection weights. Ideal output (dashed line) is a perfectly circular path. For each simulation the variance of the connection weight noise is  $\sigma^2$ . The RMSE of the decoded trajectory for each simulation is calculated over the complete length of the run (2 seconds).

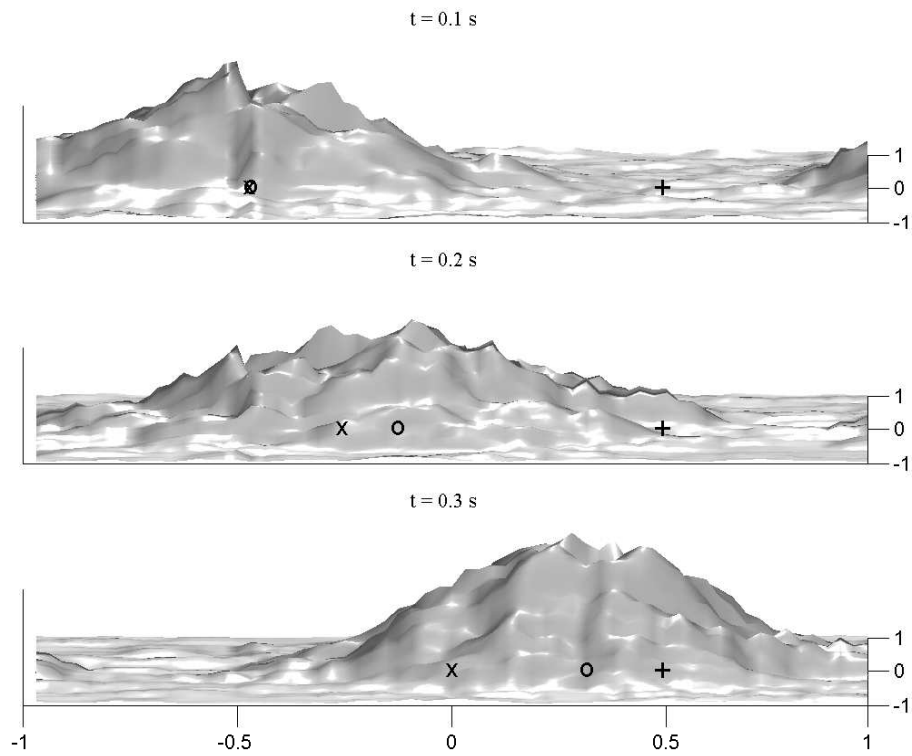


Figure 8: Weak sensory input causes smooth acceleration of the activity packet. Velocity input is to the right at one half maximum speed. '+' marks the horizontal position of the sensory stimulation. 'o' marks the horizontal mean of the activity packet. 'x' marks the horizontal mean of the activity packet from figure 3 without sensory stimulation at the same time index.



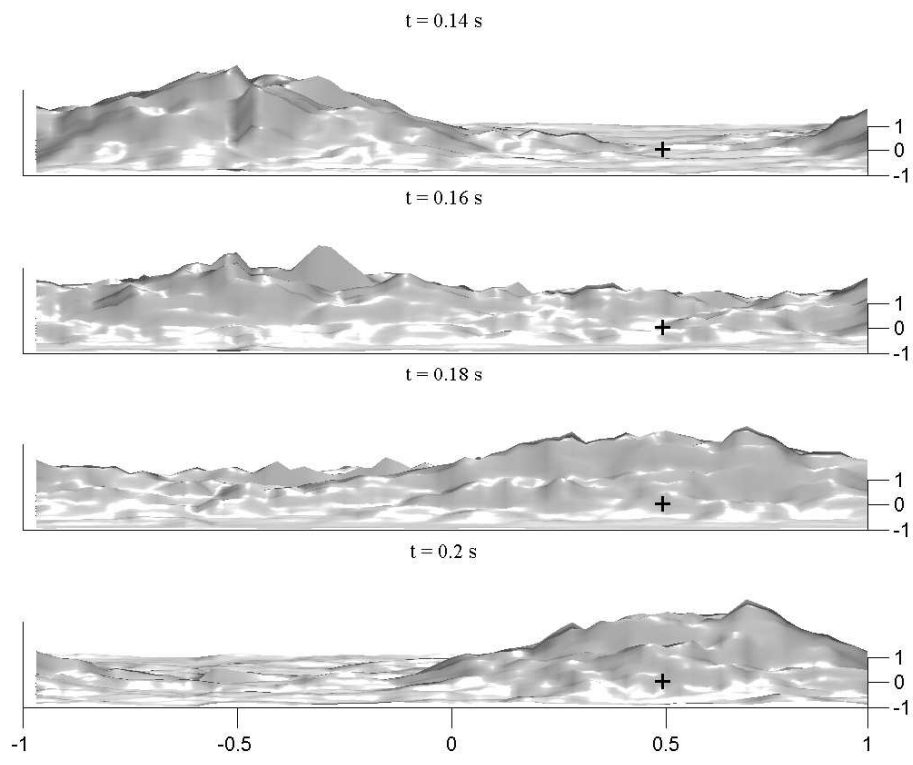


Figure 9: Strong sensory input causes an abrupt jump of the activity packet. The cross marks the location of sensory stimulation.

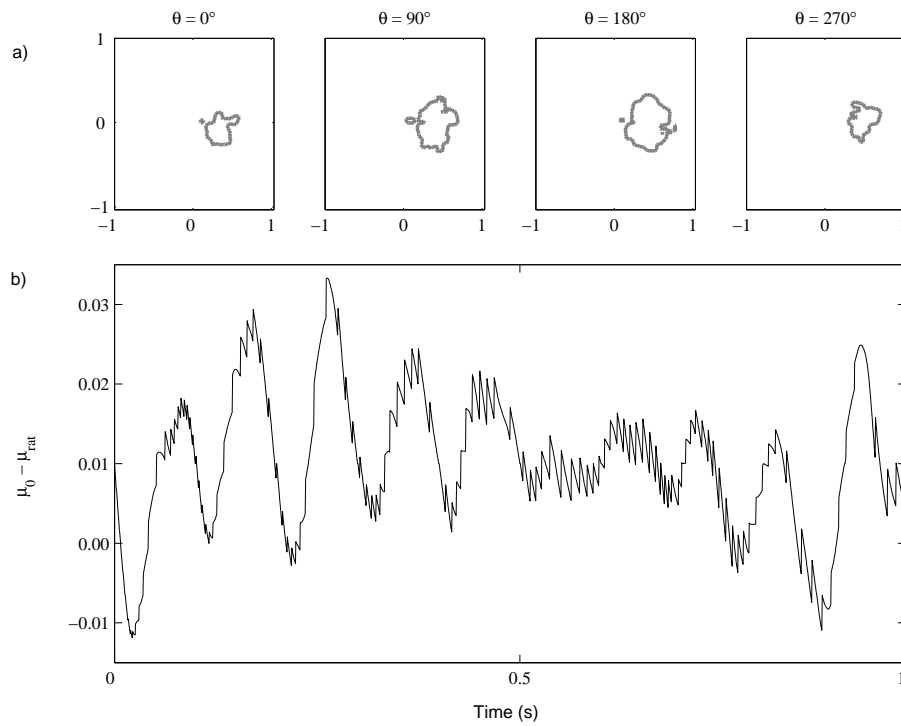


Figure 10: a) Width of the activity packet oscillates with the theta rhythm. A constant frequency contour is shown for four consecutive phases of a theta cycle. The direction of motion is to the right. b) The center of the activity packet oscillates ahead of the rat's actual position in the direction of motion over the course of each cycle. Here  $\mu_0$  is the position of the center of the activity packet, and  $\mu_{rat}$  is the simulated rat's actual position on the plane. The rat moves at a constant velocity in the positive  $\mu$  direction.

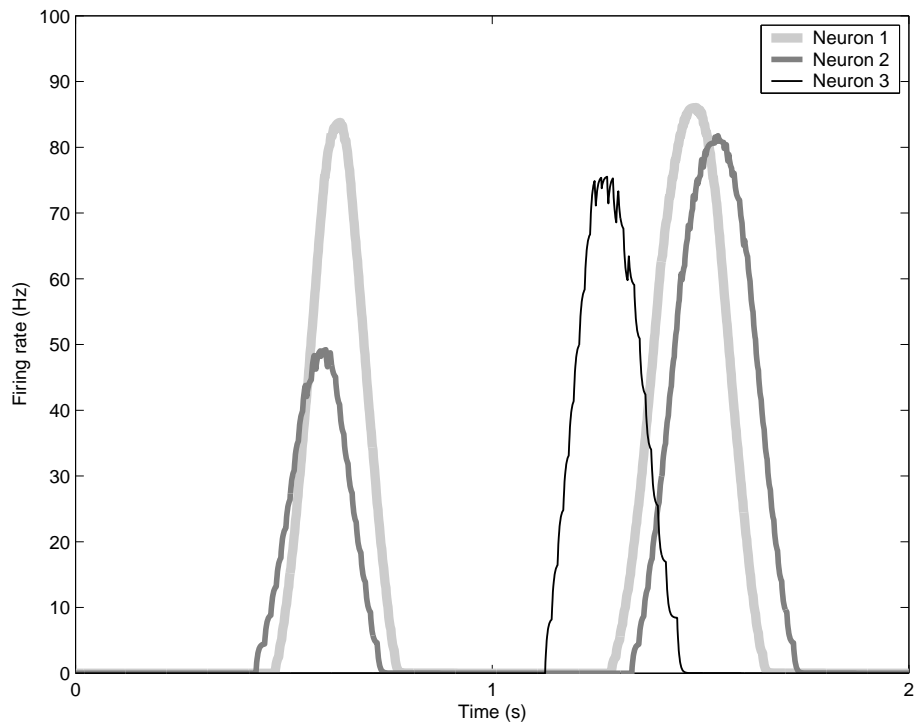


Figure 11: Firing of three neurons in a linear path simulation. For the first second the rat travels in a straight line at maximum speed. For the next second the rat travels back along the same line in the opposite direction at maximum speed. Neuron 1 fires independently of direction. Neuron 2 fires at a higher rate in one direction than the other. Neuron 3 is active in one direction and silent in the other. Firing rates are found by binning spikes and smoothing with a running average.

Experimental Tuning Curves



Simulated Tuning Curves

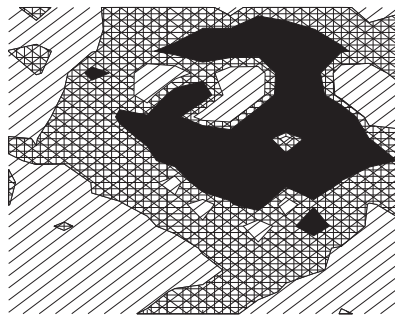
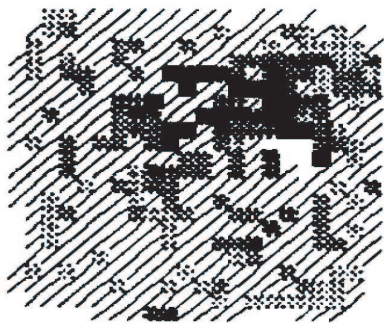
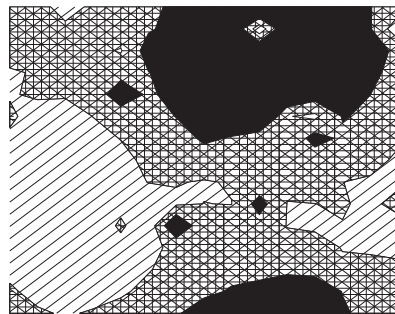


Figure 12: Tuning curves of observed subicular neurons (left; Sharp, 1997, Copyright Elsevier permission pending) and simulated neurons (right). Darker shading indicates higher firing rate.