INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **14** (2003) 629–645 PII: S0954-898X(03)62195-0

Detecting dynamical changes within a simulated neural ensemble using a measure of representational quality

Jadin C Jackson and A David Redish

Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455, USA

E-mail: redish@ahc.umn.edu

Received 12 August 2002, accepted for publication 28 April 2003
Published 28 July 2003
Online at stacks.iop.org/Network/14/629

Abstract

Technological advances allowing simultaneous recording of neuronal ensembles have led to many developments in our understanding of how the brain performs neural computations. One key technique for extracting information from neural populations has been population reconstruction. While reconstruction is a powerful tool, it only provides a value and gives no indication of the quality of the representation itself. In this paper, we present a mathematically and statistically justified measure for assessing the quality of a representation in a neuronal ensemble. Using a simulated neural network, we show that this measure can distinguish between system states and identify moments of dynamical change within the system. While the examples used in this paper all derive from a standard network model, the measure itself is very general. It requires only a representational space, measured tuning curves, and neural ensembles.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In recent years, technological advances have made simultaneous recording of neuronal ensembles possible, leading to many exciting developments in our understanding of how the brain performs neural computations. These breakthroughs have been critically dependent on the development of techniques for extracting features that describe population activity and relate it to cognitive and stimulus parameters (Georgopoulos *et al* 1983, Eskandar *et al* 1992, Wilson and McNaughton 1993, Salinas and Abbott 1994, Celebrini and Newsome 1994, Rieke *et al* 1997, Groh *et al* 1997, Brown *et al* 1998, Nicolelis *et al* 1998, Zhang *et al* 1998, Laubach *et al* 2000).

These techniques include *reconstruction* in which one attempts to calculate a behavioural or stimulus variable from neural firing patterns (Georgopoulos *et al* 1983, Wilson and McNaughton 1993, Salinas and Abbott 1994, Rieke *et al* 1997, Brown *et al* 1998, Nicolelis *et al* 1998, Zhang *et al* 1998, Laubach *et al* 2000, Averbeck *et al* 2002). The accuracy of this reconstruction has traditionally been assessed using *reconstruction error*, usually measured as the distance between the calculated value and the behavioural value actually experienced (Wilson and McNaughton 1993, Salinas and Abbott 1994, Brown *et al* 1998, Zhang *et al* 1998). However, in many reconstruction methods, reconstruction error is neither linear nor monotonic, so distance between the calculated and actual value is not a valid error measure. More importantly, reconstruction error does not provide information about the quality of the representation. In this paper, we derive a method for the detection of the quality of the representation.

Reconstruction techniques have also been employed to discriminate between pairs of represented stimuli (Celebrini and Newsome 1994, Groh *et al* 1997, Miller *et al* 1993, Gochin *et al* 1994, Deadwyler *et al* 1996). This method constitutes a psychometric approach to analysing neural data, which infers the represented stimulus from the firing patterns of one or more recorded cells. Success of the discrimination is measured as the percentage of times the stimulus is correctly inferred from the firing pattern. This is usually measured as a function of some quality of the stimulus (Celebrini and Newsome 1994). Although these methods ask how well an ensemble firing pattern discriminates between stimuli, these methods do not address the quality of the neuronal representations of those stimuli. And in particular, they do not address the dynamics of those representations.

Because neuronal representations are highly distributed, at any moment in time, cells could be firing in a manner that is generally consistent or generally inconsistent with the reconstructed value. Such inconsistencies may be expected during dynamic state transitions (Georgopoulos *et al* 1988, Redish *et al* 2000).

For example, in a mental-rotation task, in which monkeys were asked to reach to a target 90° anticlockwise from the cueing LED, the reconstructed direction (calculated as the weighted vector mean (Mardia 1972); also known as the *population vector* (Georgopoulos *et al* 1983)) rotates from the cued direction to the reaching direction during the reaction time (Georgopoulos *et al* 1988). The length of the population vector (a measure of representational quality) increases through the reaction time (Georgopoulos *et al* 1988). The length of the population vector is a limited measure of representational quality (Smyrnis *et al* 1992), but it only works on circular data sets and only tests for non-randomness of the distribution.

Redish *et al* (2000) presented a more general measure of representational quality and used it on hippocampal ensembles; they termed this measure *coherency*[1]. Redish *et al* used a three-step process: calculate the activity packet (a weighted sum of tuning curves based on the current neural activity), calculate the expected activity packet (a weighted sum of tuning curves based on the expected neural activity as predicted by the behavioural variable and the tuning curves), and then coherency was defined as the dot product between the two packets. Redish *et al* then measured which coordinate system was better represented within the ensemble as the ratio between the coherency in each of two coordinate systems (one aligned with the room, one aligned with the track). They observed that the ratio suggested that the hippocampal ensemble firing pattern represented position on the track at the beginning of the journey, but realigned to represent position in the room after a few seconds (Redish *et al* 2000). While this measure was capable of detecting the realignment of the neuronal representation, the measure itself was not mathematically or statistically justified.

[1] This term ‘coherency’ should not be confused with the term ‘coherence’ which measures the correlation across spectra (Rosenberg *et al* 1989, Kocsis *et al* 1999).

In this paper, we present a mathematically and statistically justified measure for neuronal ensembles. This method is also robust to non-uniformly sampled stimulus spaces. We show that this measure can distinguish between states of a simulated neural network and identify moments of dynamical change within the system. While the examples used in this paper all derive from a standard network model, the measure itself is very general. It requires only a representational space, measured tuning curves, and neural ensembles.

2. The network model

We used a well-studied attractor network (Wilson and Cowan 1973, Amari 1977, Kohonen 1984, Redish *et al* 1996, Zhang 1996, Redish 1999) in order to allow a better exploration of network states. This allowed a fuller study of the detection ability of the coherency method than would have been possible if we had used experimental data.

This type of network has been used to model a variety of neural structures, including the head-direction system of the rodent (Skaggs *et al* 1995, Redish *et al* 1996, Zhang 1996, Goodridge and Touretzky 2000, Sharp *et al* 2001), place cells within the hippocampus (McNaughton *et al* 1996, Shen and McNaughton 1996, Zhang 1996, Samsonovich and McNaughton 1997, Redish 1999, Káli and Dayan 2000, Guazzelli *et al* 2001), the formation of ocular-dominance columns (Obermayer *et al* 1992, Miller 1995), control of saccadic error in the superior colliculus (Sparks 1986, Munoz *et al* 1991, van Opstal and Kappen 1993, Arai *et al* 1994) and in the basal ganglia (Arbib and Dominey 1995), and memory storage within cortex (Wilson and Cowan 1973, Kohonen 1982, 1984).

Simulations were based on those presented in Redish (1999). The firing rate of each unit was continuous and normalized to be between 0 and 1. Continuous, random, Gaussian noise was added in order to better simulate physiological conditions. Seventy five excitatory neurons (voltage, V_k^{E}; firing rate, F_k^{E}; and synaptic drive, S_k^{E}) and one inhibitory neuron (voltage, V^{I}; firing rate, F^{I}; and synaptic drive, S^{I}) were used in these simulations. The excitatory neurons were arranged with uniform spacing of the preferred directions along a ring topology. This simplified boundary conditions and other calculations. Specific parameters are given in table 1.

The neural analogues of voltage and firing rate are straightforward. The neural analogue of synaptic drive is derived by normalizing the synaptic α-function (here taken to be an exponential decay) by the synaptic weight (Pinto *et al* 1996). Thus, rather than calculating firing rate, multiplying that by the synaptic weight, and then producing a synaptic effect, we calculate firing rate, produce the synaptic effect, and then multiply that by the synaptic weight.

The voltage for each excitatory neuron was calculated from the synaptic drive of each excitatory neuron and from the synaptic drive for the inhibitory neuron. The weight matrix $W_{\mathrm{E}\leftarrow\mathrm{E}}$ was a Gaussian kernel of standard deviation 72°, providing a local excitation function. The weight $W_{\mathrm{E}\leftarrow\mathrm{I}}$ was a constant providing global inhibition:

$$V_k^{\mathrm{E}}(t) = \sum_j W_{\mathrm{E}\leftarrow\mathrm{E}}(\phi_k - \phi_j) S_j^{\mathrm{E}}(t) - W_{\mathrm{E}\leftarrow\mathrm{I}} S^{\mathrm{I}}(t) + \gamma^{\mathrm{E}} + \mathrm{IN}_k(t) \tag{1}$$

where ϕ_k and ϕ_j represent the preferred directions of neuron k and neuron j, respectively. Firing rate was taken as a simple sigmoidal function of voltage:

$$F_k^{\mathrm{E}}(t) = \frac{1 + \tanh V_k^{\mathrm{E}}(t)}{2}. \tag{2}$$

Noise was added to the excitatory neurons in the synaptic drive equation:

$$S_k^{\mathrm{E}}(t + \Delta t) = S_k^{\mathrm{E}}(t) + \frac{\Delta t}{\tau^{\mathrm{E}}}(F_k^{\mathrm{E}}(t) - S_k^{\mathrm{E}}(t)) + \xi_G(t) S_k^{\mathrm{E}}(t) \tag{3}$$

Table 1. Parameters used in the simulations ('var' indicates variables restricted to the range $[x, y]$).

V_k^{E}	Membrane voltage of excitatory neuron k	var
$F_k^{\mathrm{E}}(t)$	Normalized firing rate of excitatory neuron k	var $\in [0, 1]$
$S_k^{\mathrm{E}}(t)$	Synaptic drive from excitatory neuron k	var
$\mathrm{IN}_k(t)$	Extra-network excitatory input to excitatory neuron k	var $\in [0, 1]$
ξ_{G}	Gaussian noise added to S_k^{E}	$\mu = 0, \sigma = 0.1$
γ^{E}	Tonic inhibition to each excitatory neuron	-1.5
τ^{E}	Decay time constant for presynaptic effect of excitatory neuron	10 ms
V^{I}	Membrane voltage of inhibitory neuron	var
$F^{\mathrm{I}}(t)$	Normalized firing rate of inhibitory neuron	var $\in [0, 1]$
$S^{\mathrm{I}}(t)$	Synaptic drive from inhibitory neuron	var
γ^{I}	Tonic inhibition to inhibitory neuron	-7.5
τ^{I}	Decay time constant for presynaptic effect of inhibitory neuron	2 ms
$W_{\mathrm{E}\leftarrow\mathrm{E}}$	Synaptic weight kernel for E-to-E connections	See the text
$W_{\mathrm{E}\leftarrow\mathrm{I}}$	Synaptic weight for I-to-E connections	-8.0
$W_{\mathrm{I}\leftarrow\mathrm{E}}$	Synaptic weight for E-to-I connections	0.88
$W_{\mathrm{I}\leftarrow\mathrm{I}}$	Synaptic weight for I-to-I connections	-4
Δt	Time step	1 ms

where ξ_{G} was drawn randomly at each time step from a normal distribution with variance 0.1 and mean 0. Thus, the actual variance of the injected noise was signal dependent. Synaptic efficacy was limited to the range [0, 1].

Functions for the inhibitory interneuron were similar. However, for simplicity, noise was not included in the interneuron's synaptic drive:

$$V^{\mathrm{I}}(t) = \sum_j W_{\mathrm{I}\leftarrow\mathrm{E}} S_j^{\mathrm{E}}(t) - W_{\mathrm{I}\leftarrow\mathrm{I}} S^{\mathrm{I}}(t) + \gamma^{\mathrm{I}} \tag{4}$$

$$F^{\mathrm{I}}(t) = \frac{1 + \tanh V^{\mathrm{I}}(t)}{2} \tag{5}$$

$$S^{\mathrm{I}}(t + \Delta t) = S^{\mathrm{I}}(t) + \frac{\Delta t}{\tau^{\mathrm{I}}}(-S^{\mathrm{I}}(t) + F^{\mathrm{I}}(t)). \tag{6}$$

As noted above, this network has been well studied. In particular, the effects of starting conditions and extra-network input are well known (Wilson and Cowan 1973, Samsonovich and McNaughton 1997, Redish 1999). We will focus on three issues:

Issue 1: random noise versus stable activity mode

When started from random noise, the network settles to a stable mode of activity representing one direction. Cells with preferred directions near this direction will have much higher firing rates than those far away. The direction represented will be a consequence of the random input (Wilson and Cowan 1973, Kohonen 1977).

Issue 2: rotation versus jump

When a system is in a stable state (representing a single direction), and extra-network excitatory drive is provided near the represented direction (within 60° in our network), the represented direction will shift toward the input, passing through intermediate values (Redish *et al* 1996, Zhang 1996, Samsonovich and McNaughton 1997, Redish 1999). By continuously providing input that is offset by a constant from the current represented direction (offset by less than 60°), the network can be forced to shift its representation to any value. We will call this a *rotation*.

In contrast, when the system is in a stable state and sufficient extra-network excitatory drive is provided far away from the represented direction (beyond 60° in our simulations), the firing patterns of the system will change to encode a new value without encoding intermediate values in the interim (Zhang 1996, Samsonovich and McNaughton 1997, Redish 1999). We will call this a *jump*.

Issue 3: ambiguity

When the system is started from a random non-representational state (all firing rates near 0), and extra-network drive is provided at two inputs, the system will settle to represent a unique direction, depending on the random noise (Redish 1999).

3. Methods

3.1. Tuning curves

In order to generate the tuning curves for neurons in the attractor network, an attractor network simulation was run for 5000 time steps, and the centre of the population activity at each time step (given by the reconstructed direction) was used to align the activity to the same central direction. Since the attractor network is symmetric and all neurons have equivalent parameters, the result was equal to sampling a single neuron in multiple directions. Thus, the activity was averaged over all time steps to provide the expected mean firing rate for a cell for each direction sampled. For consistency, the tuning curve was then normalized so that it ranged from 0 to 1:

$$\mathsf{T}_k(\phi) = \frac{\bar{F}_k^{\mathrm{E}}(\phi)}{\max_\phi \bar{F}_k^{\mathrm{E}}(\phi)} \tag{7}$$

where $\bar{F}_k^{\mathrm{E}}(\phi)$ is the average firing rate in direction ϕ for excitatory neuron k (with preferred direction ϕ_0) over the stable state centred at direction ϕ_0. Thus $\mathsf{T}_k(\phi)$ is the tuning curve for neuron k with its maximum centred at preferred direction ϕ_0. (The tuning curve is a function over ϕ.) For all neurons, this generic tuning curve was rotated so that its maximum was aligned with the neuron's preferred direction.

3.2. Reconstruction

Many methods have been used for reconstruction (Georgopoulos *et al* 1983, Wilson and McNaughton 1993, Salinas and Abbott 1994, Rieke *et al* 1997, Zhang *et al* 1998). For a uniform ring topology such as the one used here, the simplest is the weighted vector mean (Mardia 1972, Batschelet 1981) also known as the *population vector* (Georgopoulos *et al* 1983, 1988). Let $\vec{v}_k$ be the preferred vector for excitatory neuron k, defined as the unit vector in the direction of the preferred direction ϕ_k of neuron k: $\vec{v}_k = \langle \cos(\phi_k), \sin(\phi_k) \rangle$. The reconstructed vector $\vec{R}$ is the weighted mean of the preferred vectors:

$$\vec{R} = \sum_k F_k^{\mathrm{E}} \vec{v}_k. \tag{8}$$

The reconstructed direction $\hat{\varphi}$ is then the orientation of the reconstructed vector $\vec{R}$. While we use this method for simplicity, it is important to note that coherency (defined below) works with any reconstruction method. It can also be used as a comparison between neural activity and expected neural activity given the real behavioural values.

3.3. Coherency

As stated earlier, coherency is simply a comparison of the actual and expected activity of the system, represented mathematically by the *actual* and *expected activity packet* ($A(x,t)$ and $\hat{A}(x,t)$, respectively). Note that, like a tuning curve, the activity packet is a function over the behavioural variable space x (not necessarily limited to only one or two dimensions); unlike the tuning curve, however, the activity packet is time dependent.

The activity packet is well defined for any given state of any system. Thus, comparing actual and expected activity packets can provide a measure of the state of the system in any given time step.

Actual and expected activity packets can be calculated from the tuning curves of each neuron and from the firing rates of each neuron at a time t. Let $\mathsf{T}_k(x)$ be the tuning curve of cell k over space x and $F_k(t)$ the firing rate of cell k at time t. Then the actual activity packet of a neuronal ensemble recorded at time t is given by a normalized, weighted sum of tuning curves:

$$A(x,t) = \frac{\sum_k \mathsf{T}_k(x) F_k(t)}{\sum_k \mathsf{T}_k(x)}. \tag{9}$$

Note that this is different from the definition of the activity packet used by Redish *et al* (2000) in that here the activity packet is normalized by a spatial function, the sum of all tuning curves in the ensemble. This allows a simple dependence upon the random variable $F_k(t)$. Thus, we can rewrite the activity packet as shown below:

$$A(x,t) = \sum_k B_k(x) F_k(t) \tag{10}$$

where

$$B(x) = \frac{\mathsf{T}_k(x)}{\sum_k \mathsf{T}_k(x)}. \tag{11}$$

Likewise, the expected activity packet of a neuronal ensemble recorded at time t is

$$\hat{A}(x,t) = \frac{\sum_k \mathsf{T}_k(x) \boldsymbol{E}[F_k(t)|\hat{x}(t)]}{\sum_k \mathsf{T}_k(x)} \tag{12}$$

where $\boldsymbol{E}[F_k(t)|\hat{x}(t)]$ is the expected firing rate observed in previous experiments given the actual or reconstructed behavioural value for the variable x at time t (designated here as $\hat{x}$). Note that $\boldsymbol{E}[F_k(t)|\hat{x}(t)]$ can be taken by evaluating the tuning curve at $\hat{x}$. Thus, we can rewrite the expected activity packet by substituting $\boldsymbol{E}[F_k(t)|\hat{x}(t)] = \mathsf{T}_k(\hat{x}(t))$:

$$\hat{A}(x,t) = \frac{\sum_k \mathsf{T}_k(x) \mathsf{T}_k(\hat{x}(t))}{\sum_k \mathsf{T}_k(x)} \tag{13}$$

or

$$\hat{A}(x,t) = \sum_k B_k(x) \mathsf{T}_k(\hat{x}(t)). \tag{14}$$

Also, note that in our simulations, we used the reconstructed direction $\hat{\varphi}$ as the expected representational direction (see equation (8)). However, in experimental data, the behavioural value actually experienced by the animal can be substituted in lieu of the reconstructed value (such as the actual position of the animal within a coordinate system, as used by Redish *et al* (2000)). Once the expected representational value is used to create an expected activity packet, the actual and expected activity packets are then compared to measure coherency.

The *coherency* of the ensemble is found by comparing the actual and expected activity packets in two steps. The first step is to determine the probability distribution of the comparison

function. The second is to calculate the probability of match between the expected and observed activity packets using the null hypothesis that the two packets are the same:

$$H_0 : A(x,t) = \hat{A}(x,t). \tag{15}$$

While earlier versions of this measurement used the dot product between the two activity packets (Redish *et al* 2000), we will demonstrate the root mean squared error (RMSE) method of comparison. The RMSE method effectively subtracts off the expected mean activity level and reveals the magnitude of the difference between the expected and actual activity packets. For this reason, RMSE appears to be much more sensitive than dot product. Large RMSE values imply a difference between the actual and expected activity packets:

$$\mathrm{RMSE}(t) = \frac{\sqrt{\int_x (A(x,t) - \hat{A}(x,t))^2 \,\mathrm{d}x}}{\int_x \hat{A}(x,t)\,\mathrm{d}x}. \tag{16}$$

While our analysis will use the RMS error measurement, it is important to note that there are many other equally valid and quite powerful comparison functions that can be used. The objective is to find a function that highlights the differences between the two activity packets.

The distribution of RMSE values is used to determine the probability of match between the expected and observed activity packets. As noted above, the reformulation of the activity packet allows a simple dependence on the random variable, the firing rate of each neuron. This reformulation facilitates the construction of the distribution of RMSE values by making the activity packet a linear function of the firing rates.

3.4. Calculating the coherency cumulative distribution function

The first step in our proposed coherency analysis method is to run a simple experiment to determine the tuning curves of the cells being recorded. For this purpose, we used the attractor simulations to substitute for our neural data. The data from this experiment can then be used to calculate RMS error values for each time interval (chosen to optimally estimate neuronal firing rates) for the ensemble.

Similarly, we used the 5000-step attractor simulation that generated our tuning curve data to calculate RMSE values for each time step. A histogram of these values is shown in figure 1(A). This distribution of RMSE values was smoothed, interpolated to a higher resolution, and normalized to represent an estimate of the probability distribution (pdf) of RMSE values.

The RMSE pdf was integrated to find the cumulative distribution (cdf) of RMSE values. The pdf and the cdf are shown in parts (B) and (C) of figure 1. The function cdf(RMSE) represents the probability of measuring an RMSE value smaller than a given value of error. Likewise, $1 - \mathrm{cdf}(\mathrm{RMSE})$ represents the probability of measuring an RMSE value larger than a given value, and therefore represents the likelihood that the expected activity packet and the actual activity packet are the same. So, evaluating $1 - \mathrm{cdf}(\mathrm{RMSE})$ for each value of RMSE measured in a new experiment from the same ensemble of cells will give a p-value representing the likelihood that the system has a coherent representation. This p-value corresponds to the probability of observing an RMSE value given the null hypothesis that the actual and expected activity packets are the same (as noted above, equation (15)).

Given this null hypothesis, we can set a threshold below which the expected and actual activity packets are *significantly* different. Since the RMSE values for a stable 'bump' of activity were used to define the frequency distribution (the *pdf*), a system activity consistent with the tuning curves should evenly explore the range from 0 to 1. Thus, a significance threshold of 0.05 could be defined such that the system should spend 5% of the time with

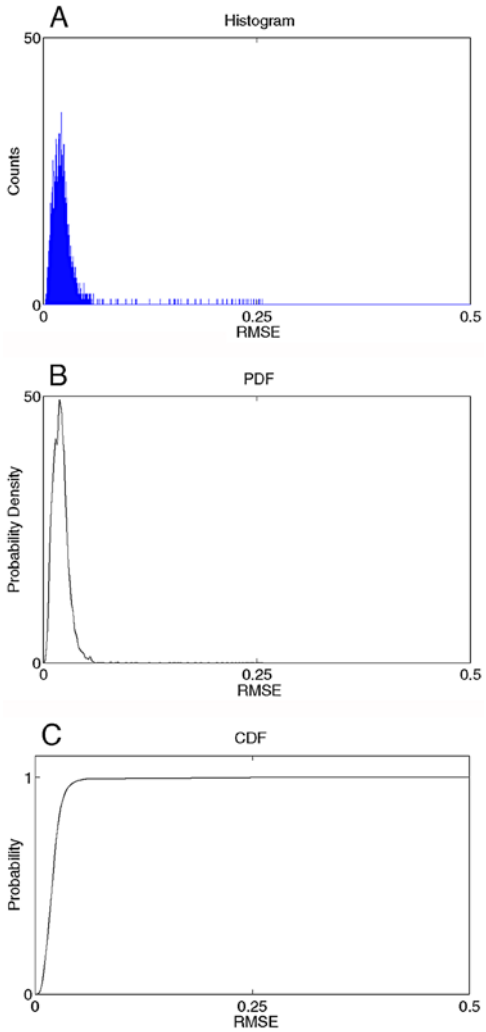

Figure 1. The RMSE probability density and the cumulative distribution. These curves were calculated from a 5000-step run as described in the text. The probability density function (B) was calculated by smoothing, interpolating, and normalizing the frequency distribution histogram (A) and was integrated to find the cumulative distribution function (C).

RMSE values that give a p-value less than 0.05. This was true in our simulations. In a simulation with a continuous directional input at 180°, 4.9% of the values out of 5000 time steps had RMSE values below the $\alpha = 0.05$ level.

For ensemble firing patterns that differ from the state predicted by the tuning curves, the p-value will be very close to zero, indicating a system state that is incompatible with the activity predicted by the tuning curves, even after accounting for the variability due to the stochasticity of neuronal firing rates. Due to the non-linearity of the RMSE measurement, a very small significance value is preferred to reduce false alarms. When the firing patterns are inconsistent with each other, they fall well below even reasonably small significance levels. Thus, we will use a significance level of $\alpha = 0.005$ for the rest of our simulations.

4. Results

4.1. Coherency differentiates representational quality (issue 1)

As noted above (issue 1), when an attractor network is started from random, noisy values, it settles to a stable state such that only cells with preferred directions near a specific orientation are active. Figure 2(A) shows an example of a network settling. Reconstruction always provided an orientation and did not differentiate the random and settled states (figure 2(B)).

In contrast, coherency differentiated the random and settled states (figures 2(C)–(E)). Because of the non-linearities of the measure, coherency detected the time of settling accurately, displaying a stark difference between the two states. While in the random state, the coherency measurement showed that the random state was significantly different from the expected 'bump' of activity ($p < 0.005$). After the network transitioned to a stable representational state, coherency showed a higher probability of match.

4.2. Coherency can detect dynamic changes in network activity (issue 2)

When an attractor network is in a stable state, providing synaptic input to neurons with preferred directions near the direction being represented by the network forces a rotation in the represented direction. Chaining this extra-network excitation to the represented direction forces the network to rotate continuously. In contrast, if sufficient extra-network excitation is provided to a group of neurons with preferred directions far from the encoded direction, the firing patterns of the system will change to encode the value consistent with the excitation without passing through intermediate values in the interim (Redish *et al* 1996, Zhang 1996, Samsonovich and McNaughton 1997, Redish 1999; issue 2, above).

Reconstruction showed a smooth transition through intermediate orientations in both conditions (figures 3(B), (G)). Reconstruction thus suggested that both of these transitions were simple rotations, yet the dynamics of these two transitions were fundamentally different (figures 3(A), (F)). Coherency, however, detected the difference (figures 3(C)–(E), (H)–(J)). In the jump condition, RMSE showed a strong transient increase at the time of transition (time steps 313–336, $p < 0.005$; see figures 3(H)–(J)), but no corresponding increase during the rotation (time steps 200–800, $p > 0.005$; see figures 3(C)–(E)).

4.3. Coherency can be used to detect the resolution of ambiguity (issue 3)

When the attractor network simulation begins with synaptic input to one group of neurons with nearby preferred directions, the network will settle to a bump of activity centred on that group. In other words, there will be one value represented by the network: the value consistent with the preferred direction of the most active neurons (Wilson and Cowan 1973, Kohonen 1977). However, when the attractor network simulation begins with synaptic input to two separate groups of neurons, the network will transiently display bimodal activity. In other words, there will temporarily be two bumps of activity, one at the centre of each input, and therefore two values represented. The system will quickly settle to a state where only one value is represented; the result is dependent on the separation of the inputs (Redish 1999).

When the two inputs are separated by a large enough distance (when the difference in preferred direction is greater than 60° in our network), the two values compete: the final represented value is equal to one or the other of the two input values. In contrast, when the network is primed with two inputs that are closer, the activity merges: the final represented value is the mean of the two input values (see figure 4).

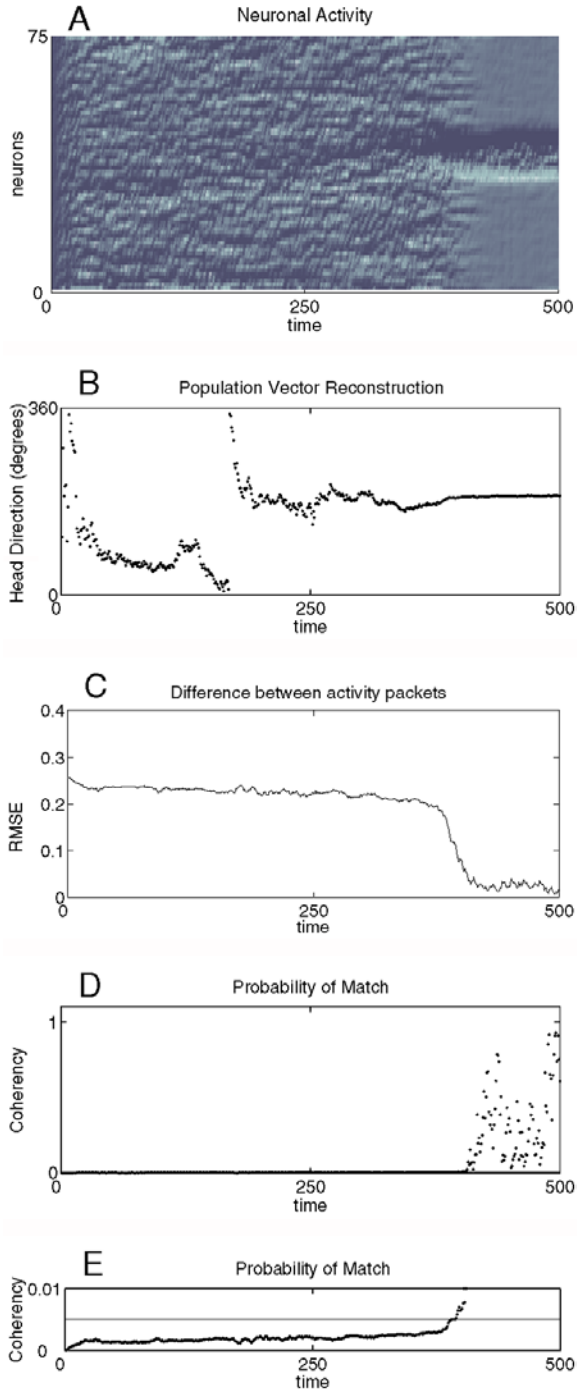

Figure 2. Random activity gives rise to a single coherent representation. (A) Neural activity. (B) Reconstructed values ($\hat{\varphi}(t)$). (C) RMSE. (D) $P(\text{match})$. (E) $P(\text{match})$ zoomed in: the line is drawn for the $\alpha = 0.005$ significance level. Note that while reconstruction did not detect the time of settling, the coherency method did.

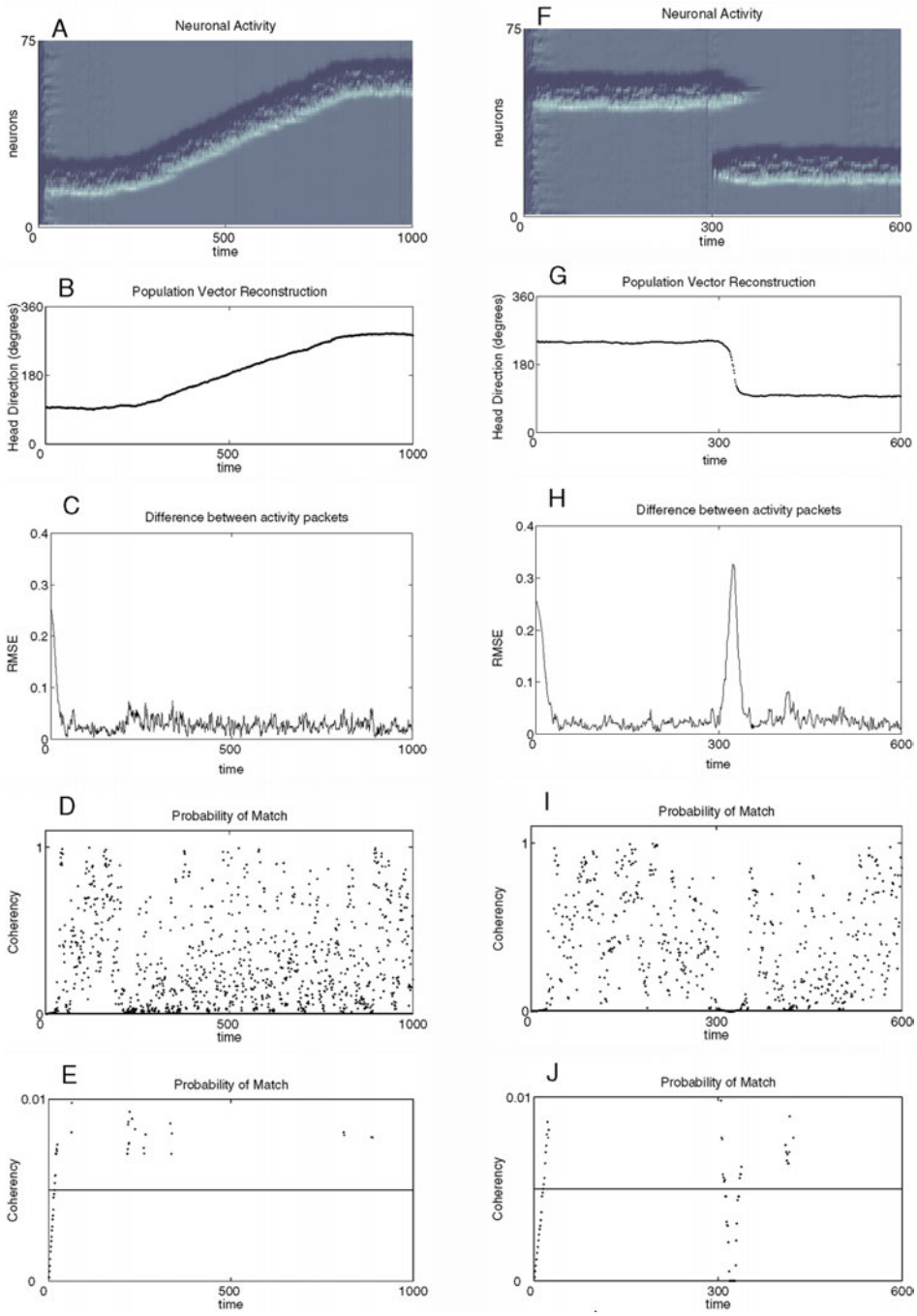

Figure 3. A rotation of the representation through continuously shifting input (left, (A)–(E)) and a jump caused by strong external input (right, (F)–(J)). (A) Neural activity. (B) Reconstructed values ($\hat{\varphi}(t)$). (C) RMSE. (D) $P(\text{match})$. (E) $P(\text{match})$ zoomed in: the line is drawn for the $\alpha = 0.005$ significance level. Note that the coherency showed that the representation remained valid through the transition. (F) Neural activity. (G) Reconstructed values ($\hat{\varphi}(t)$). (H) The RMSE and (I) $P(\text{match})$. (J) $P(\text{match})$ zoomed in: the line is drawn for the $\alpha = 0.005$ significance level. Note that the coherency showed that the representation became transiently invalid during the transition.

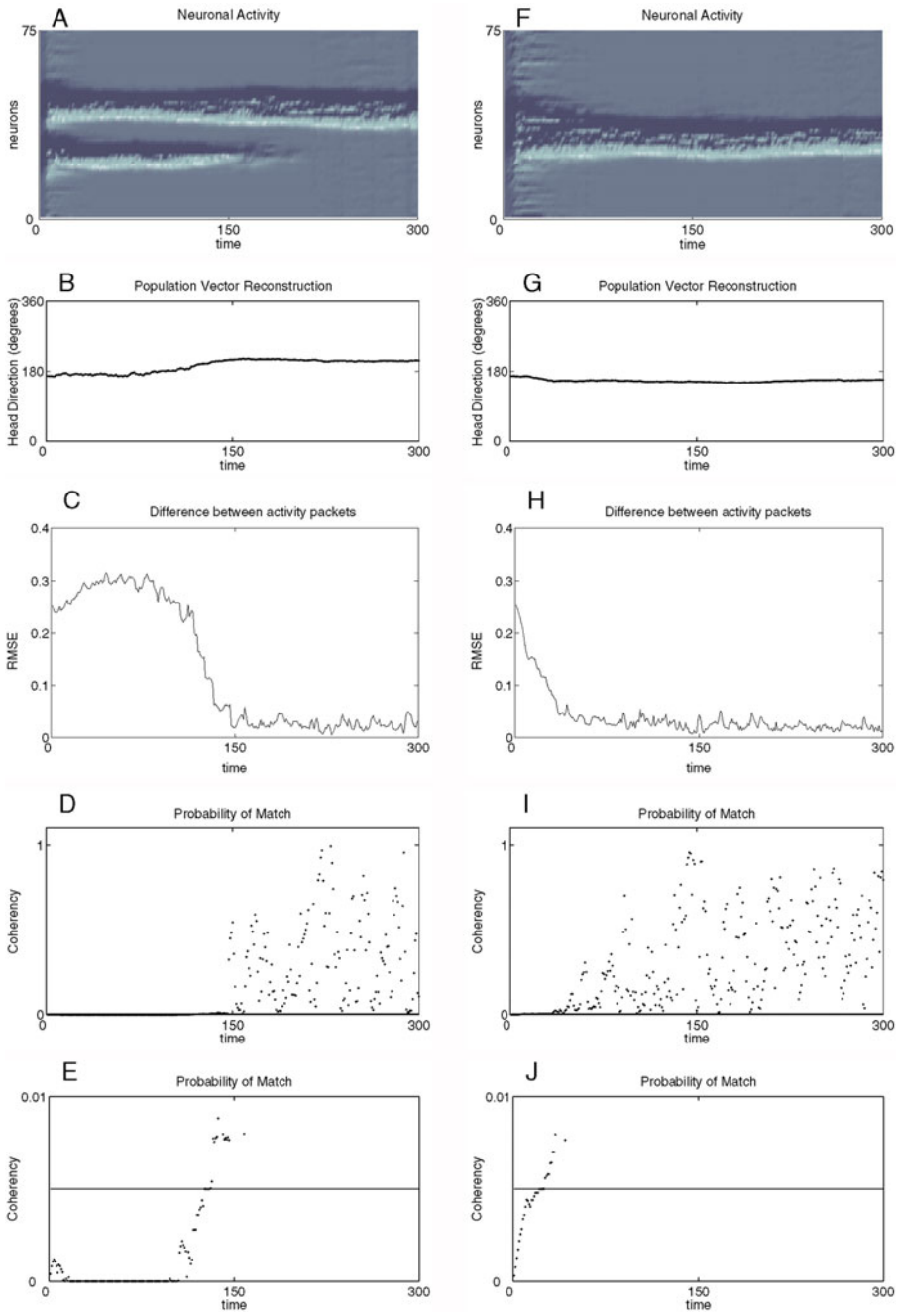

Figure 4. Ambiguous inputs with small separation compete (left, (A)–(E)) while ambiguous inputs with large separation merge (right, (F)–(J)). (A) Neural activity of merging inputs. (B) Reconstructed values ($\hat{\varphi}(t)$). (C) RMSE. (D) $P(\text{match})$. (E) $P(\text{match})$ zoomed in: the line is drawn for the $\alpha = 0.005$ significance level. Note that the reconstruction cannot detect that the neural activity is actually ambiguous before time step 126. Coherency differentiates the ambiguous and non-ambiguous representations. (F) Neural activity. (G) Reconstructed values ($\hat{\varphi}(t)$). (H) RMSE. (I) $P(\text{match})$. (J) $P(\text{match})$ zoomed in: the line is drawn for the $\alpha = 0.005$ significance level. Note that the reconstruction cannot detect that the neural activity is actually ambiguous before time step 27. Coherency differentiates the ambiguous and non-ambiguous representations.

In both of these cases, the network can be seen as resolving ambiguity by forcing a multimodal input to settle to a unimodal firing pattern. The representation of the eventual steady state of the system can be measured using a standard reconstruction algorithm. However, as in the earlier examples, reconstruction provided no information as to the time at which the system resolved the ambiguity. Although reconstruction did not vary in the near-input case (figure 4(G)), reconstruction provided incorrect results at time steps 0–126 in the distantly separated case (figure 4(B)).

Coherency, on the other hand, did provide that timing information. When the representation was still ambiguous, the system was incoherent ($p < 0.005$; figures 4(C)–(E), (H)–(J)). In the well-separated case (in which the inputs compete; figure 4(A)), coherency transitioned from a highly significant, low probability of match to within the non-significant range at time step 126, indicating that the network had successfully resolved the ambiguity (figures 4(C)–(E)). The low probability of match ($p < 0.005$) at time steps 0–126 indicated that reconstruction provided invalid results, while the high probability of match afterwards indicated that the reconstructed directions after time step 126 were valid. Thus, coherency successfully detected the time of transition to a stable state. Similarly, for the close-input case (in which the inputs merge; figure 4(F)), coherency transitioned to the non-significant range at time step 27 (figures 4(H)–(J)), indicating that reconstructed directions after time step 27 were valid. Again, coherency successfully detected the time of transition to a stable state.

5. Discussion

Our results have shown that coherency is a robust measure of the state of a neural ensemble. It distinguished between consistent and inconsistent representational states: there was a stark difference between a random noise state and a coherent representational state. Coherency detected transient changes in the state of the neural ensemble: it distinguished between smooth rotations, where intermediate values were encoded, and jumps, where intermediate values were not encoded. These two transitions were ambiguous using traditional reconstruction techniques. Coherency also detected ambiguity.

We have derived a statistically justified measurement of coherency. In doing so, we can directly measure the significance of coherency values. Thus, the random noise state was significantly different from the stable state of the attractor network. Similarly, during the jump transition, coherency indicated that the network was transiently significantly different from before and after the transition.

In order to avoid the strict assumptions made by standard parametric statistics, our method requires only the assumption that the data used to construct the tuning curves and RMSE cumulative distribution function be taken during a stationary period where it was unlikely that any dynamic instabilities occurred. This is not a new or special assumption; it is fundamental to the concept of the tuning curve. After this measurement is made, coherency can be used to detect non-stationary events.

Redish *et al* (2000) used the ratio between two coherency measurements in order to determine the time at which the hippocampal ensemble realigned between two coordinate systems. While they were able to derive expected coherency ratios under those two conditions, the dot-product-based coherency did not provide direct statistics. By redefining coherency as the probability of match, the statistical difference of the two states can be directly assessed. This should allow the statistical measurement of coherency within two coordinate systems, extending the usefulness of the coherency measure.

The examples of network dynamics shown in this paper may be observable *in vivo*. For example, this network has been used to model the rodent head-direction system (Redish *et al*

1996, Zhang 1996). Under normal circumstances, the head-direction system should provide a reliable representation of the rodent's current head direction: only those neurons with preferred directions similar to the current head direction should be firing (Taube *et al* 1990a, 1996, Sharp *et al* 2001). When the rodent is turning, the system should show a smooth rotation to update the system's representation to match the rodent's state (Redish *et al* 1996, Zhang 1996) as in figure 3(A).

The head-direction system is generally studied in a 1 m diameter high-wall open field with a cue-card subtending 100° (Ranck 1984, Taube *et al* 1990a, 1990b). When an animal is returned to the field with the card rotated, the head-direction cells generally rotate with the cue-card (Taube *et al* 1990b). However, when the animal is disoriented before each experience, the head-direction cells no longer follow the cue-card (Knierim *et al* 1995). This is presumably because, during disorientation, the network destabilizes and resettles to a random orientation (as in figure 2(A)), and the orientation of the cue-card is no longer at a constant orientation relative to the internal coordinate system defined by the head-direction system (Knierim *et al* 1995). The head-direction system has also been studied under reorientation paradigms in which the cues are rotated in the presence of the animal (Blair and Sharp 1996, Knierim *et al* 1998, Zugaro *et al* 2003). Whether these reorientation protocols produce smooth rotations (as in figure 3(A)) or jump transitions (as in figure 3(F)) is an open question that will require ensemble recording from head-direction cells to answer. In cases in which the wall and floor do not rotate in synchrony (e.g. the wall rotates 90° but the floor remains unrotated), this will produce a cue-conflict situation (Taube and Burton 1995, Blair and Sharp 1996, Knierim *et al* 1998), which will require the resolution of ambiguity (as in figure 4).

There have been statistically based methods that use the length of the reconstructed population vector (Moore 1980, Smyrnis *et al* 1992, Ashe *et al* 1993) but these methods require nearly complete revision when attempting to apply them to other systems with different symmetries. For example, applying such a framework based on unimodal tuning curves to systems with bimodal, multimodal, or spatial tuning curves (O'Keefe and Conway 1978, O'Keefe and Speakman 1987, Sharp 1996, Blair *et al* 1997) would require a complete rederivation of the statistics. In contrast, the coherency method proposed here will work for any shape of tuning curve without modification.

While this paper has focused on using coherency to analyse an attractor network with a uniform ring topology, it should be noted that the coherency method is applicable to any neural system. Thus, multidimensional systems with bimodal, multimodal, or even heterogeneous tuning curves are accessible to the coherency method. For example, some hippocampal place cells exhibit multiple place fields. The coherency method presented here requires no modifications to accommodate an ensemble with a mixture of multimodal and unimodal responses; the only requirement is the ability to calculate a tuning curve for each individual cell. Knowing the tuning curve allows the construction of an expected activity packet, to which the actual activity packet can be compared.

Any simultaneously recorded neural ensemble that can be used to reconstruct a behavioural variable can also be used to measure coherency. Reconstruction with simultaneously recorded neural ensembles has been used in place cells (Wilson and McNaughton 1993, Brown *et al* 1998, Zhang *et al* 1998) and in motor cortex (Salinas and Abbott 1994, Averbeck *et al* 2002). Accurate coherency measures can be taken from as few cells as are needed to provide adequate coverage of the parameter space by their tuning curves (i.e. enough to have reliable reconstruction). In our simulations, we found accurate detection of the three dynamic transitions discussed above with as few as 10 cells. With modern recording technologies, simultaneous recordings of 20–100 cells are common. Coherency can be used with any of these recordings to detect both differences in representational quality as well as transient representational events.

Acknowledgments

We thank Dr C Bingham and Dr J Baxter for helpful discussions on statistical issues and probability theory, as well as N C Schmitzer-Torbert for generally helpful dialogues. This work was supported by NIH MH68029-01. JCJ was supported by NSF-IGERT training grant No 9870633.

References

Amari S I 1977 Dynamics of pattern formation in lateral-inhibition type neural fields *Biol. Cybern.* **27** 77–87

Arai K, Keller E L and Edelman J A 1994 Two-dimensional neural network model of the primate saccadic system *Neural Networks* **7** 1115–35

Arbib M A and Dominey P F 1995 Modeling the roles of basal ganglia in timing and sequencing saccadic eye movements *Models of Information Processing in the Basal Ganglia* ed J C Houk, J L Davis and D G Beiser (Cambridge, MA: MIT Press) pp 149–62

Ashe J, Taira M, Smyrnis N, Pellizer G, Georgakopoulos T, Lurito J T and Georgopoulos A P 1993 Motor cortical activity preceding a memorized movement trajectory with an orthogonal bend *Exp. Brain Res.* **95** 118–30

Averbeck B B, Chafee M V, Crowe D A and Georgopoulos A P 2002 Parallel processing of serial movements in prefrontal cortex *Proc. Natl Acad. Sci. USA* **99** 13172–7

Batschelet E 1981 *Circular Statistics in Biology* (New York: Academic)

Blair H T, Lipscomb B W and Sharp P E 1997 Anticipatory time intervals of head-direction cells in the anterior thalamus of the rat, implications for path integration in the head-direction circuit *J. Neurophysiol.* **78** 145–59

Blair H T and Sharp P E 1996 Visual and vestibular influences on head-direction cells in the anterior thalamus of the rat *Behav. Neurosci.* **110** 643–60

Brown E N, Frank L M, Tang D, Quirk M C and Wilson M A 1998 A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells *J. Neurosci.* **18** 7411–25

Celebrini S and Newsome W T 1994 Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey *J. Neurosci.* **14** 4109–24

Deadwyler S A, Bunn T and Hampson R E 1996 Hippocampal ensemble activity during spatial delayed-nonmatch-to-sample performance in rats *J. Neurosci.* **16** 354–72

Eskandar E N, Richmond B J and Optican L M 1992 Role of inferior temporal neurons in visual memory I. Temporal encoding of information about visual images, recalled images, and behavioral context *J. Neurophysiol.* **68** 1277–95

Georgopoulos A P, Caminiti R, Kalaska J F and Massey J T 1983 Spatial coding of movement: a hypothesis concerning the coding of movement direction by motor cortical populations *Exp. Brain Res. (Suppl.)* 327–36

Georgopoulos A P, Kettner R E and Schwartz A B 1988 Primate motor cortex and free arm movements to visual targets in three-dimensional space: II. Coding of the direction of movement by a neuronal population *J. Neurosci.* **8** 2928–37

Gochin P M, Colombo M, Dorfman G A, Gerstein G L and Gross C G 1994 Neural ensemble coding in inferior temporal cortex *J. Neurophysiol.* **71** 2325–37

Goodridge J P and Touretzky D S 2000 Modeling attractor deformation in the rodent head-direction system *J. Neurophysiol.* **83** 3402–4310

Groh J M, Born R T and Newsome W T 1997 How is a sensory map read out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements *J. Neurosci.* **17** 4312–30

Guazzelli A, Bota M and Arbib M A 2001 Competitive Hebbian learning and the hippocampal place cell system: modeling the interaction of visual and path integration cues *Hippocampus* **11** 216–39

Káli S and Dayan P 2000 The involvement of recurrent connections in area CA3 in establishing the properties of place fields: a model *J. Neurosci.* **20** 7463–77

Knierim J J, Kudrimoti H S and McNaughton B L 1995 Place cells, head direction cells, and the learning of landmark stability *J. Neurosci.* **15** 1648–59

Knierim J J, Kudrimoti H S and McNaughton B L 1998 Interactions between idiothetic cues and external landmarks in the control of place cells and head direction cells *J. Neurophysiol.* **80** 425–46

Kocsis B, Bragin A and Buzsaki G 1999 Interdependence of multiple theta generators in the hippocampus: a partial coherence analysis *J. Neurosci.* **19** 6200–12

Kohonen T 1977 *Associative Memory: a System-Theoretical Approach* (New York: Springer)

Kohonen T 1982 Self-organized formation of topologically correct feature maps *Biol. Cybern.* **43** 59–69

Kohonen T 1984 *Self-Organization and Associative Memory* (New York: Springer)
Laubach M, Wessberg J and Nicolelis M A L 2000 Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task *Nature* **405** 567–71
Mardia K V 1972 *Statistics of Directional Data* (New York: Academic)
McNaughton B L, Barnes C A, Gerrard J L, Gothard K, Jung M W, Knierim J J, Kudrimoti H, Qin Y, Skaggs W E, Suster M and Weaver K L 1996 Deciphering the hippocampal polyglot: the hippocampus as a path integration system *J. Exp. Biol.* **199** 173–86
Miller E K, Li L and Desimone R 1993 Activity of neurons in anterior inferior temporal cortex during a short-term memory task *J. Neurosci.* **13** 1460–78
Miller K D 1995 Receptive fields and maps in the visual cortex: models of ocular dominance and orientation columns *Models of Neural Networks* vol 3, ed E Domany, J L van Hemmen and K Schulten (New York: Springer)
Moore B R 1980 A modification of the Rayleigh test for vector data *Biometrika* **67** 175–80
Munoz D P, Pélisson D and Guitton D 1991 Movement of neural activity on the superior colliculus motor map during gaze shifts *Science* **251** 1358–60
Nicolelis M A L, Ghazanfar A A, Stambaugh C R, Oliveira L M O, Laubach M, Chapin J K, Nelson R J and Kaas J H 1998 Simultaneous encoding of tactile information by three primate cortical areas *Nat. Neurosci.* **1** 621–30
Obermayer K, Blasdel G G and Schulten K 1992 Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps *Phys. Rev.* A **45** 7568–88
O'Keefe J and Conway D H 1978 Hippocampal place units in the freely moving rat: why they fire where they fire *Exp. Brain Res.* **31** 573–90
O'Keefe J and Speakman A 1987 Single unit activity in the rat hippocampus during a spatial memory task *Exp. Brain Res.* **68** 1–27
Pinto D J, Brumberg J C, Simons D J and Ermentrout G B 1996 A quantitative population model of whisker barrels: re-examining the Wilson–Cowan equations *J. Comput. Neurosci.* **3** 247–64
Ranck J B Jr 1984 Head-direction cells in the deep cell layers of dorsal presubiculum in freely moving rats *Soc. Neurosci.* **10** 599 (abstracts)
Redish A D 1999 *Beyond the Cognitive Map: from Place Cells to Episodic Memory* (Cambridge, MA: MIT Press)
Redish A D, Elga A N and Touretzky D S 1996 A coupled attractor model of the rodent head direction system *Network: Comput. Neural Syst.* **7** 671–85
Redish A D, Rosenzweig E S, Bohanick J D, McNaughton B L and Barnes C A 2000 Dynamics of hippocampal ensemble realignment: time versus space *J. Neurosci.* **20** 9289–309
Rieke F, Warland D, de Ruyter van Steveninck R and Bialek W 1997 *Spikes* (Cambridge, MA: MIT Press)
Rosenberg J R, Amjad A M, Breeze P, Brillinger D R and Halliday D M 1989 The Fourier approach to the identification of functional coupling between neuronal spike trains *Prog. Biophys. Mol. Biol.* **53** 1–31
Salinas E and Abbott L 1994 Vector reconstruction from firing rates *J. Comput. Neurosci.* **1** 89–107
Samsonovich A V and McNaughton B L 1997 Path integration and cognitive mapping in a continuous attractor neural network model *J. Neurosci.* **17** 5900–20
Sharp P E 1996 Multiple spatial/behavioral correlates for cells in the rat postsubiculum: multiple regression analysis and comparison to other hippocampal areas *Cereb. Cortex* **6** 238–59
Sharp P E, Blair H T and Cho J 2001 The anatomical and computational basis of the rat head-direction cell signal *Trends Neurosci.* **24** 289–94
Shen B and McNaughton B L 1996 Modeling the spontaneous reactivation of experience-specific hippocampal cell assembles during sleep *Hippocampus* **6** 685–93
Skaggs W E, Knierim J J, Kudrimoti H S and McNaughton B L 1995 A model of the neural basis of the rat's sense of direction *Advances in Neural Information Processing Systems 7,* ed G Tesauro, D S Touretzky and T K Leen (Cambridge, MA: MIT Press) pp 173–80
Smyrnis M, Taira M, Ashe J and Georgopoulos A P 1992 Motor cortical activity in a memorized delay task *Exp. Brain Res.* **92** 139–51
Sparks D L 1986 Translation of sensory signals into commands for control of saccadic eye movements: role of primate superior colliculus *Physiol. Rev.* **66** 118–71
Taube J S and Burton H L 1995 Head direction cell activity monitored in a novel environment and during a cue conflict situation *J. Neurophysiol.* **74** 1953–71
Taube J S, Goodridge J P, Golob E J, Dudchenko P A and Stackman R W 1996 Processing the head direction cell signal: a review and commentary *Brain Res. Bull.* **40** 477–86
Taube J S, Muller R U and Ranck J B Jr 1990a Head direction cells recorded from the postsubiculum in freely moving rats: I. Description and quantitative analysis *J. Neurosci.* **10** 420–35
Taube J S, Muller R U and Ranck J B Jr 1990b Head direction cells recorded from the postsubiculum in freely moving rats: II. Effects of environmental manipulations *J. Neurosci.* **10** 436–47

van Opstal A J and Kappen H 1993 A two-dimensional ensemble coding model for spatial–temporal transformation of saccades in monkey superior colliculus *Network: Comput. Neural Syst.* **4** 19–38

Wilson H R and Cowan J D 1973 A mathematical theory of the functional dynamics of cortical and thalamic tissue *Kybernetik* **13** 55–80

Wilson M A and McNaughton B L 1993 Dynamics of the hippocampal ensemble code for space *Science* **261** 1055–8

Zhang K 1996 Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory *J. Neurosci.* **16** 2112–26

Zhang K, Ginzburg I, McNaughton B L and Sejnowski T J 1998 Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells *J. Neurophysiol.* **79** 1017–44

Zugaro M B, Arleo A, Berthoz A and Wiener S I 2003 Rapid spatial reorientation and head direction cells *J. Neurosci.* **23** 3478–82