LETTER **Communicated by Jonathan Victor**

Dynamic Analyses of Information Encoding in Neural Ensembles

Riccardo Barbieri
barbieri@neurostat.mgh.harvard.edu
Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital/Harvard Medical School, Boston, MA 02114, U.S.A.

Loren M. Frank
loren@neurostat.mgh.harvard.edu
David P. Nguyen
dpnguyen@neurostat.mgh.harvard.edu
Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital; Division of Health, Sciences, and Technology, Harvard Medical School/MIT, Boston, MA 02114, U.S.A.

Michael C. Quirk
mquirk@ladyday.mwl.ai.mit.edu
Picower Center for Learning and Memory, Riken-MIT Neuroscience Research Center, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Victor Solo
vic@nmr.mgh.harvard.edu
School of Electrical Engineering and Telecommunications, University of New South Wales, 2052, Sydney, Australia; Martinos Center for Biomedical Imaging, Massachusetts General Hospital/Harvard Medical School, Boston, MA 02114, U.S.A.

Matthew A. Wilson
mwilson@mit.edu
Picower Center for Learning and Memory, Riken-MIT Neuroscience Research Center, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

Emery N. Brown
brown@neurostat.mgh.harvard.edu
Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital; Division of Health, Sciences, and Technology, Harvard Medical School/MIT, Boston, Massachusetts 02114-2698, U.S.A.

Neural spike train decoding algorithms and techniques to compute Shannon mutual information are important methods for analyzing how neural systems represent biological signals. Decoding algorithms are also one of several strategies being used to design controls for brain-machine interfaces. Developing optimal strategies to desig n decoding algorithms and compute mutual information are therefore important problems in computational neuroscience. We present a general recursive filter decoding algorithm based on a point process model of individual neuron spiking activity and a linear stochastic state-space model of the biological signal. We derive from the algorithm new instantaneous estimates of the entropy, entropy rate, and the mutual information between the signal and the ensemble spiking activity. We assess the accuracy of the algorithm by computing, along with the decoding error, the true coverage probability of the approximate 0.95 confidence regions for the individual signal estimates. We illustrate the new algorithm by reanalyzing the position and ensemble neural spiking activity of CA1 hippocampal neurons from two rats foraging in an open circular environment. We compare the performance of this algorithm with a linear filter constructed by the widely used reverse correlation method. The median decoding error for Animal 1 (2) during 10 minutes of open foraging was 5.9 (5.5) cm, the median entropy was 6.9 (7.0) bits, the median information was 9.4 (9.4) bits, and the true coverage probability for 0.95 confidence regions was 0.67 (0.75) using 34 (32) neurons. These findings improve significantly on our previous results and suggest an integrated approach to dynamically reading neural codes, measuring their properties, and quantifying the accuracy with which encoded information is extracted.

1 Introduction

Neural spike train decoding algorithms are commonly used methods for analyzing how neural systems represent biological signals (Georgopoulos, Kettner, & Schwartz, 1986; Bialek, Rieke, de Ruyter van Stevenick, & Warland, 1991; Wilson & McNaughton, 1993; Warland, Reinagel, & Meister, 1997; Brown, Frank, Tang, Quirk, & Wilson, 1998; Zhang, Ginzburg, McNaughton, & Sejnowski, 1998; Stanley, Li, & Dan, 1999; Pouget, Dayan, & Zemel, 2000). More recently, these algorithms are also one of several strategies being used to design controls for neural prosthetic devices (Chapin, Moxon, Markowitz, & Nicolelis, 1999; Wessberg et al., 2000; Shoham, 2001; Serruya, Hatsopoulos, Paninski, Fellows, & Donoghue, 2002; Taylor, Tillery, & Schwartz, 2002) and other brain-machine interfaces (Donoghue, 2002; Wickelgren, 2003). Developing optimal strategies for constructing and testing decoding algorithms is therefore an important question in computational neuroscience.

The rat hippocampus has specialized neurons known as place cells with spatial receptive fields that carry a representation of the animal's environment (O'Keefe & Dostrovsky, 1971). The spatial information carried by the

spiking activity of the hippocampal place cells is believed to be an important component of the rat's spatial navigation system (O'Keefe & Nadel, 1978). We previously derived the Bayes' filter algorithm to decode the position of a rat foraging in an open environment from the ensemble spiking activity of pyramidal neurons in the CA1 region of the hippocampus (Brown et al., 1998). We represented the spatial receptive field of each neuron as a two-dimensional gaussian surface and assumed that the position of the animal in the receptive field through time parameterized the rate function of an inhomogeneous Poisson process. We modeled the animal's path as a random walk. The Bayes' filter was a causal, recursive decoding algorithm developed by computing gaussian approximations to the well-known Bayes' rule and the Chapman-Kolmogorov system of equations for state-space modeling (Mendel, 1995; Kitagawa & Gersh, 1996). The algorithm computed position estimates and their confidence regions at 33 msec intervals. Because the algorithm depended critically on the form of the spatial gaussian model, it could not be used with other models of the place fields. While representing the place fields as gaussian surfaces and the path as a random walk was reasonable, these models gave only an approximate description of these data (Brown et al., 1998).

Methods to compute the Shannon mutual information (Skaggs, McNaughton, Gothard, Marcus, 1993; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997; Dan, Alonso, Usrey, & Reid, 1998; Strong, Koberle, de Ruyter van Steveninck, & Bialek, 1998; Reinagel & Reid, 2000; Reich, Melcher, & Victor, 2001; Johnson, Gruner, Baggerly, & Shehagiri, 2001; Nirenberg, Carcieri, Jacobs, & Latham, 2001; Victor, 2002), rather than decoding algorithms, are perhaps the most widely used techniques for analyzing how neural systems encode biological signals. None of these methods uses a parametric statistical model of the relation between spiking activity and the biological signal, and none computes mutual information estimates instantaneously (i.e., on the millisecond timescale of the decoded signal updates). Moreover, current analyses do not fully exploit the link between mutual information and decoding.

We derive a general recursive filter decoding algorithm by using a point process to represent the spiking activity of individual neurons and a linear stochastic state-space model (Smith & Brown, 2003) to represent the animal's path. From the algorithm, we derive new instantaneous estimates of the mutual information between position and the ensemble spiking activity, the entropy, and the entropy rate. We assess the accuracy of the algorithm by computing, along with the decoding error, the true coverage probability of the approximate 0.95 confidence regions for the individual position estimates. We illustrate the new algorithm by reanalyzing the position and ensemble neural spiking activity of CA1 hippocampal neurons from two rats foraging in an open circular environment (Brown et al., 1998). For the point process, we use an inhomogeneous Poisson model with spatial receptive fields modeled as both gaussian surfaces and Zernike polynomial

expansions (Born & Wolf, 1989), and for the linear state-space model of the path, we use a bivariate first-order autoregressive process. We compare the performance of these algorithms with a linear filter algorithm constructed by the widely-used reverse correlation method (Warland, Reinagel, & Meister, 1997; Stanley et al., 1999; Wessberg et al., 2000; Serruya et al., 2002).

2 Theory

In this section, we develop the theoretical framework for our approach. First, in the encoding analysis, we define the relation between position and spiking activity for individual place cell neurons with two inhomogeneous Poisson models. In the first model, the dependence of the rate function on space is defined as a gaussian surface, whereas in the second model, the spatial dependence is represented as an expansion in Zernike polynomials. We describe how the two models are fit to the experimental data by maximum likelihood. Second, in the decoding analysis, we describe the autoregressive model of the path and the recursive filter algorithm we will use to decode position from the ensemble place cell spiking activity. Third, we describe several properties of the recursive filter algorithm. Fourth, we present the reverse correlation algorithm, and finally, we present our algorithms for computing instantaneous entropy, entropy rate, and mutual information.

2.1 Encoding Analysis: Place Cell Model. To define the point process observation model, we consider two inhomogeneous Poisson models for representing the place cell spiking activity. In the first, we model the rate or conditional intensity function $\lambda_G^c(t \mid x(t), \zeta_G)$ as a two-dimensional Gaussian surface defined as

$$\lambda_G^c(t \mid x(t), \zeta_G^c) = \exp\left\{\alpha^c - \frac{1}{2}(x(t) - \mu^c)'(Q^c)^{-1}(x(t) - \mu^c)\right\}, \tag{2.1}$$

where $x(t) = (x_1(t), x_2(t))$ are the coordinates of the animal's position at time t, $\mu^c = (\mu_1^c, \mu_2^c))$ are the coordinates of the place field's center,

$$Q^c = \begin{bmatrix} (\sigma_1^c)^2 & 0 \\ 0 & (\sigma_2^c)^2 \end{bmatrix}, \tag{2.2}$$

is the scale matrix, α is the log maximum firing rate (Brown et al., 1998), and $\zeta_G^c = (\alpha^c, \mu^c, Q^c)'$ where c indexes a neuron, for $c = 1, \ldots, C$. As in Brown et al. (1998), we have taken the off-diagonal elements of Q^c to be zero because for each neuron, these parameter estimates were all close to zero.

In the second, we model the conditional intensity function $\lambda_z^c(t \mid x(t), \zeta_z^c)$ as an exponentiated linear combination of Zernike polynomials given as

$$\lambda_z^c(t \mid \zeta_z^c) = \exp\left\{\sum_{\ell=0}^{L}\sum_{m=-\ell}^{\ell} \zeta_{\ell,m}^c z_\ell^m(\rho(t), \phi(t))\right\}, \tag{2.3}$$

where $z_{\ell,m}$ is the mth component of the ℓth order Zernike polynomial, $\zeta^c_{\ell,m}$ is the associated coefficient, $\rho(t) = r^{-1/2}[(x_1(t) - \eta_1)^2 + (x_2(t) - \eta_2)^2]^{1/2}$, $\phi(t) = \tan^{-1}[(x_2(t) - \eta_2)(x_1(t) - \eta_1)^{-1}]$, (η_1, η_2) are the coordinates of the center of the circular environment, r is the radius of the circular environment, and $\eta_1 = \eta_2 = r = 35$cm. We have $\zeta^c_z = \{\{\zeta^c_{\ell,m}\}^{m=\ell}_{m=-\ell}\}^L_{\ell=0}$, the coefficients in equation 2.3. The Zernike polynomials form an orthogonal basis whose support is restricted to the circular environment (Born & Wolf, 1989) and are defined as

$$Z^m_\ell(\rho(t), \phi(t)) = \begin{cases} R^m_\ell(\rho(t)) \sin(m\phi(t)) & m > 0 \\ R^m_\ell(\rho(t)) \cos(m\phi(t)) & m < 0 \end{cases} \tag{2.4}$$

$$R^m_\ell(\rho(t)) = \begin{cases} \sum_{j=0}^{(\ell-|m|)/2} \frac{(-1)^j(\ell - j)!}{j!(\frac{\ell+m}{2} - j)!(\frac{\ell-m}{2} - j)!} \cdot \rho(t)^{\ell-2j} & (\ell - m)\ even \\ 0 & (\ell - m)\ odd, \end{cases} \tag{2.5}$$

where $0 \le \rho(t) \le 1$, $0 \le \phi(t) \le 2\pi$, $0 \le |m| \le \ell$. To balance the trade-off between model flexibility and computational complexity, we chose $L = 3$. Based on equation 2.4, there are 10 nonzero coefficients for $L = 3$ in equation 2.3.

We assume that during the experiment, the neural spiking activity of an ensemble of C neurons is recorded on an interval $(0, T]$. We take the subinterval $(0, T^e]$ to be the encoding interval and the subinterval $(T^e, T]$ to be the decoding interval. The spiking activity recorded during the encoding interval will be used to estimate the model parameters, whereas the spiking activity recorded during the decoding interval will be used to estimate, or decode, the animal's position from the ensemble neural spiking activity through time. We define the encoding interval to be the first 15 minutes of the experiment for Animal 1 and the first 13 minutes for Animal 2.

Let $0 < \mu^c_1 < \mu^c_2 <, \ldots, < \mu^c_{J^c} \le T$ be the set of J^c spike times from neuron c for $c = 1, \ldots, C$. For $t \in (0, T]$, let $N^c_{0:t}$ be the sample path of the spike times from neuron c in $(0, t]$. It is defined as the event $N^c_{0:t} = \{0 < u^c_1 < u^c_2 < \ldots, < u^c_j \le t \cap N^c(t) = j\}$, where $N^c(t)$ is the number of spikes in $(0, t]$ from neuron c and $j \le J^c$. To estimate the parameters, ζ^c_G for the gaussian model and ζ^c_Z for the Zernike models by maximum likelihood, we have to specify the likelihood function for these two models given the spiking activity in the encoding interval (Pawitan, 2001). It follows from the theory of point processes (Barbieri, Quirk, Frank, Wilson, & Brown, 2001; Brown, Barbieri, Eden, & Frank, 2003; Daley & Vere-Jones, 2003) that given the specification for neuron c of the conditional intensity function for the gaussian model in equation 2.1 and for the Zernike model in equation 2.3, the likelihood

function for the two models on the encoding interval can be written as

$$L(N^c_{0:T^e} \mid \zeta^c_j) = \exp\left\{\int_0^{T^e} \log(\lambda(u \mid \zeta^c_j) dN^c(u) - \int_0^{T^e} (\lambda(u \mid \zeta^c_j) du\right\}, \quad (2.6)$$

where $j = G$ and $j = Z$ index, respectively, the gaussian and the Zernike model parameters. The parameters of both models are estimated by maximum likelihood from the first 15 (13) minutes for Animal 1 (2) (Barbieri, Frank, Quirk, Wilson, & Brown, 2002). We used the Bayesian information criterion (BIC; Box, Jenkins, & Reinsel, 1994) to compare model goodness of fit.

2.2 Decoding Analysis: Path Model. The decoding interval contains the spiking activity and path data recorded in $(T_e, T]$. To simplify notation for the decoding analysis, we reexpress the decoding interval as the interval $(0, T^d]$, where $0 = T^e - T^e$ and $T^d = T - T^e$. That is, we denote the interval $(T_e, T]$ as the interval $(0, T^d]$. We define the updating lattice by choosing K large and by dividing $(0, T^d]$ into K intervals of equal width $\Delta = T^d/K$. To define the stochastic state-space model, we assume that the position of the rat during foraging obeys a bivariate pth-order linear gaussian autoregressive $AR(p)$ model defined as

$$x_k = \mu_x + Fx_{k-1} + R^{\frac{1}{2}}\varepsilon_k, \quad (2.7)$$

where x_k is the animal's position at time $k\Delta$, F is a $2p \times 2p$ matrix of system parameters, μ_x is a $2p \times 1$ vector of mean parameters, ε_k is a $2p \times 1$ gaussian random variable with mean zero and $2p \times 2p$ covariance matrix W_ε, and R is the learning rate scale factor (LRSF), a tuning parameter for the decoding algorithm defined below. Similarly, we estimate μ_x, F, and W_ε by maximum likelihood from 15 (13) minutes for Animal 1 (2) with $R = 1$. The state-space model used in the decoding algorithm is $p = 1$. We use the $AR(1)$ model because, like the random walk model used in our previous analysis, it is an elementary model that enforces stochastic continuity on the decoding. In addition, the $AR(1)$ model is stationary and therefore can be used in both the decoding and mutual information computations. The random walk model cannot be used in the mutual information estimation because it is a nonstationary process (Priestley, 1981).

2.3 Decoding Analysis: Recursive Filter Algorithm. In the encoding analysis, for a given animal, we estimate ζ^c for each of its place cells, for both the Zernike and gaussian models, and we estimate the parameters of its path F and W_ε. In the decoding analysis, we treat these estimated parameters as known and use them to recursively estimate the animal's position from the ensemble spiking activity. We derive our decoding algorithm from the well-known recursion relation between two coupled probability densities:

the posterior probability density and the Chapman-Kolmogorov (one-step prediction) probability density defined as (Mendel, 1995; Kitagawa & Gersh, 1996)

$$p(x_k \mid N_{0:k}) = \frac{p(x_k \mid N_{0:k-1})p(N_{k-1:k} \mid x_k, N_{0:k-1})}{p(N_{k-1:k} \mid N_{0:k-1})} \tag{2.8}$$

$$p(x_k \mid N_{0:k-1}) = \int p(x_k \mid x_{k-1})p(x_{k-1} \mid N_{0:k-1})dx_{k-1}, \tag{2.9}$$

where $N_{0:k}$ is the ensemble neural spiking activity from $(0, k\Delta]$ for $k = 1, \ldots, K$. These equations are a recursive system for computing the probability density of the position at $k\Delta$ given the ensemble spiking activity $(0, k\Delta]$. Under the assumption that the neurons are conditionally independent, the solutions to these equations using a gaussian approximation to equation 2.8 for the point process linear state-space model we study yield the following recursive filter algorithm:

(One-step prediction)

$$x_{k|k-1} = \mu_x + \hat{F}x_{k-1|k-1}, \tag{2.10}$$

(One-step prediction variance or learning rate)

$$W_{k|k-1} = \hat{F}W_{k-1|k-1}\hat{F}' + R\hat{W}_\varepsilon, \tag{2.11}$$

(Posterior mode)

$$x_{k|k} = x_{k|k-1} + W_{k|k-1}\sum_{c=1}^{C} \nabla \log \lambda^c(x_{k|k} \mid \hat{\xi}_j^c) \times [N_{k-1:k}^c - \lambda^c(x_{k|k} \mid \hat{\xi}_j^c)\Delta], \tag{2.12}$$

(Posterior variance)

$$W_{k|k}^{-1} = \left[W_{k|k-1}^{-1} - \sum_{c=1}^{C} [\nabla^2 \log \lambda^c(x_{k|k} \mid \hat{\xi}_j^c)[N_{k-1:k}^c - \lambda^c(x_{k|k} \mid \hat{\xi}_j^c)\Delta] - \nabla \log \lambda_c(x_{k|k} \mid \hat{\xi}_j^c)[\nabla \lambda_c(x_{k|k} \mid \hat{\xi}_j^c)\Delta]'] \right], \tag{2.13}$$

for $k = 1, \ldots, K$, where $\lambda^c(x_{k|k} \mid \hat{\xi}_j^c)$ is the conditional intensity (rate) function for either the spatial gaussian or Zernike model for neuron c, $\hat{\xi}_j^c$ is the associated model parameter for either the spatial gaussian ($j = G$) or Zernike ($j = Z$) model estimated from the encoding analysis, $\hat{F}$ and $\hat{W}_\varepsilon$ are, respectively, the estimates of the transition matrix and the white noise covariance matrix for the $AR(1)$ model from the encoding analysis, Δ (Δ^2) denotes the first (second) derivative of the indicated function with respect to x_k, and the

notation $j \mid k$ denotes the estimate at time $j\Delta$ given the spiking activity in $(0, k\Delta]$. The derivation of the algorithm is given in the appendix.

Equations 2.10 to 2.13 provide a recursive system of gaussian approximations to equations 2.8 and 2.9. Equations 2.10 and 2.11 give, respectively, the mean and covariance matrix for the gaussian approximation to $p(x_k \mid N_{0:k-1})$ in equation 2.9, whereas equations 2.12 and 2.13 give, respectively, the mean and covariance matrix for the gaussian approximation to $p(x_k \mid N_{0:k})$ in equation 2.8. Because we are computing gaussian approximations to $p(x_k \mid N_{0:k})$ and $p(x_k \mid N_{0:k-1})$, it suffices to recursively compute only their respective means and covariance matrices. We use a gaussian approximation because this is a standard first approach for approximating probability densities (Tanner, 1996; Pawitan, 2001). What is important about our analysis is that this gaussian approximation is made to the posterior probability density, yet the spiking activity enters into the computations in a very nongaussian way through the point process model (see equation A.7 in the appendix). This is very different from approximating the spiking activity as a gaussian process, an assumption that could not be justified over small time intervals for hippocampal neurons where the average spike rate is 0.75 to 2 Hz. We mention alternatives of the gaussian approximation in section 4.

The derivation of the algorithm follows, in the appendix, the arguments used in the maximum a posteriori derivation of the Kalman filter (Mendel, 1995; Brown et al., 1998; Fahrmeir & Tutz, 2001). The essential idea in the derivation is that the logarithm of equation 2.8 is approximated as a quadratic function (Tanner, 1996) at each step k of the algorithm. The mode $x_{k|k}$ in equation 2.12 is the point where the quadratic approximation is carried out. It is computed by differentiating the log of the posterior probability density (see equation A.7) and setting the derivative equal to zero. If the posterior probability density is approximately symmetric, then $x_{k|k}$ also approximates its mean and median. In this case, the recursive filter algorithm in equations 2.10 to 2.13 is an approximate optimal filter in both the mean square and mean absolute error sense. Equations 2.12 and 2.13 are nonlinear functions of $x_{k|k}$ that are solved iteratively using a Newton's procedure. The starting value for the Newton's procedure at step k is the previous estimate $x_{k-1|k-1}$.

The position update $x_{k|k}$ combines the one-step prediction $x_{k|k-1}$, based on the ensemble spiking activity from $(0, (k-1)\Delta]$, with the weighted sum of $[N^c_{k-1:k} - \lambda^c(x_{k|k} \mid \hat{\xi}^c_j)\Delta]$, the innovation or error signal from neuron c for $c = 1, \ldots, C$ multiplied by the learning rate (one-step prediction variance) (Brown et al., 1998; Brown, Nguyen, Frank, Wilson, & Solo, 2001). The innovations are the new information from the ensemble spiking activity. As is true for a Poisson process, $\lambda^c(x_{k|k} \mid \hat{\xi}^c_j)\Delta$ defines for a general point process model (see equation A.1) the probability of a spike in $((k-1)\Delta, k\Delta]$ (Daley & Vere-Jones, 2003). Each innovation compares the probability of a spike, $\lambda^c(x_{k|k} \mid \hat{\xi}^c_j)\Delta$, in $((k-1)\Delta, k\Delta]$ with $N^c_{k-1:k}$, which, for small Δ, is 1 if a

spike is observed from neuron c in $((k-1)\Delta, k\Delta]$ and 0 otherwise. Thus, for small Δ, the innovation gives a weight in the interval $(-1,1)$. A large positive weight arises if a spike occurs when a neuron has a low probability of a spike in $((k-1)\Delta, k\Delta]$, whereas a large negative weight arises if no spike occurs in $((k-1)\Delta, k\Delta]$ when the probability of a spike is high.

By equation 2.12, for a given value of R, how much the new information is weighted depends on the learning rate. If the one-step prediction variance is large, then the uncertainty about the position estimate at $k\Delta$ is large given the spiking activity in $(0, (k-1)\Delta]$, and more weight is given to the innovations in determining the new position. If the one-step prediction variance is small, then the uncertainty about the position estimate at $k\Delta$ is small given the spiking activity in $(0, (k-1)\Delta]$, and less weight is given to the innovations in determining the new position. While the learning rate is dynamic (i.e., it changes at each step), increasing R increases the weight given to the innovations. Therefore, for each animal, with its place field parameters fixed at the values estimated in the encoding analysis, we optimize the decoding algorithm by studying its performance as a function of R. The initial position, x_0, and the initial covariance matrix, W_0, are computed by local maximum likelihood (Brown et al., 1998) using the 1 sec of ensemble spiking activity prior to the start of the decoding period.

As we show in the appendix, the recursive filter algorithm in equations 2.10 to 2.13 provides a straightforward generalization of our previous Bayes' filter algorithm by using a linear gaussian state-space model and by using the conditional intensity function, the rate function of a point process, to describe the spiking activity of each neuron. Use of the conditional intensity function makes it possible to use point process models more general than the Poisson to represent the spiking activity of the ensemble neurons.

2.4 Reverse Correlation Decoding Analysis. We computed the reverse correlation filter by matrix regression of position on spiking activity from the first 15 (13) minutes of spike train and position data for Animal 1 (2) assuming a filter length of 1 sec and bin widths of 33 msec. The reverse correlation updates were computed at 33 msec during the 10 minutes of decoding.

2.5 Confidence Regions and Coverage Probabilities. At $k\Delta$, an approximate 0.95 confidence region for the true position x_k may be constructed as

$$(x_k - x_{k|k})' W_{k|k}^{-1} (x_k - x_{k|k}) \leq 6, \tag{2.14}$$

for $k = 1, \ldots, K$, where 6 is the 0.95 quantile of the χ^2 distribution with 2 degrees of freedom. If the decoding algorithm performs well, then approximately 0.95 of the confidence regions should cover the true position. We compute the dynamic estimate of the coverage probability by tracking the fraction $\frac{j_k}{k}$, where j_k is the number of times the true position is within

the confidence regions in $(0, k\Delta]$ for $k = 1, \ldots, K$. The estimate of the true coverage probability for the entire decoding period is $\frac{j_K}{K}$.

2.6 Instantaneous Entropy and Entropy Rate. From equations 2.8, 2.11, and 2.12, it follows that the posterior probability density of position x_k given the ensemble spiking activity in $(0, k\Delta]$ can be approximated as the gaussian probability density $p(x_k \mid N_{0:k})$ with mean $x_{k|k}$ (see equation 2.11) and covariance matrix $W_{k|k}$ (see equation 2.13). Another way to characterize the uncertainty in this probability density is to compute its entropy. We can express the entropy of this conditional probability density as (Cover & Thomas, 1991)

$$\begin{aligned} H(x_{k|k}) &= -\int \log_2 p(x_k \mid N_{0:k}) p(x_k \mid N_{0:k}) dx_k \\ &= 2^{-1} \log_2[(2\pi e)^2 |W_{k|k}|], \end{aligned} \tag{2.15}$$

where $x_{k|k}$ denotes the random variable x_k given $N_{0:k}$, $|W_{k|k}|$ is the determinants of $W_{k|k}$ and $\log_2$ is log base 2. To examine how the instantaneous entropy changes from one update to the next we compute the entropy rate which we define as

$$\begin{aligned} \Delta H(x_{k|k}) &= H(x_{k|k}) - H(x_{k-1|k-1}) \\ &= 2^{-1} \log_2 |W_{k|k}| / |W_{k-1|k-1}|. \end{aligned} \tag{2.16}$$

The final line in equation 2.16 follows from the gaussian approximation to $p(x_k \mid N_{0:k})$. Because $H(x_{k|k})$ defines in terms of number of bits, the uncertainty in the animal's position at time $k\Delta$ given the ensemble spiking activity in $(0, k\Delta]$, $\Delta H(k \mid k)$ is the number of bits the ensemble spiking activity provided in $((k-1)\Delta, k\Delta]$ and hence, $\Delta H(x_{k|k})$ is the entropy rate. The entropy rate in equation 2.16 has an intuitive interpretation in terms of the confidence region areas at each step of the decoding because, under the gaussian approximation to $p(x_k \mid N_{0:k})$, the area of the confidence region (see equation 2.14) at step k is proportional to the determinant of $W_{k|k}$. The entropy rate is negative if $|W_{k-1|k-1}| > |W_{k|k}|$ and is positive if $|W_{k-1|k-1}| < |W_{k|k}|$. If there is less uncertainty about the animal's position at step k than at step $k-1$, that is, $|W_{k-1|k-1}| > |W_{k|k}|$, then the entropy decreases at step k, whereas if there is more uncertainty about the animal's position at step k than at step $k-1$, that is, $|W_{k|k}| > |W_{k-1|k-1}|$, then the entropy at step k increases.

We can use the relation between the covariance matrices and the entropy rate to analyze whether the value of the entropy rate in $((k-1)\Delta, k\Delta]$ is due to either the $AR(1)$ model of the path or the spiking activity of the neural ensemble. By equation 2.11, $W_{k-1|k} = FW_{k-1|k-1}F' + RW_{\varepsilon}$. Therefore, if $|W_{k-1|k}| > |W_{k-1|k-1}|$, then the $AR(1)$ path model tends to increase the uncertainty and hence, the entropy about the animal's location at any step k of

the algorithm. That is, the evolution of the path as described by the $AR(1)$ model makes the entropy rate positive in $((k-1)\Delta, k\Delta]$. For the entropy rate to be negative or, equivalently, for $|W_{k|k}|$ to decrease, the ensemble spiking activity must cause the sum on the right-hand side of equation 2.13 to be negative. The activity of the neural ensemble can further increase the entropy rate beyond the increase due to the path model if the spiking activity makes the second term on the right-hand side of equation 2.13 positive. However, because the path model makes the entropy rate positive, a decrease in the entropy, or a negative entropy rate in $((k-1)\Delta, k\Delta]$ can be due only to the spiking activity of the neural ensemble in that interval. Because this conclusion is predicated on $|W_{k-1|k}| > |W_{k-1|k-1}|$, we will check this condition at each decoding step. As mentioned above, although we use a gaussian approximation to estimate the probability densities, the point process likelihood in equations A.2 and A.6, the ensemble spiking activity enters the computation of $W_{k|k}$ through equations 2.12 and 2.13 in a non-gaussian manner.

2.7 Mutual Information. Because the animal's position as a function of time is modeled as a bivariate gaussian $AR(1)$ process, its marginal probability density $p(x_k)$ is the gaussian probability density with mean $\mu = [1-F]^{-1}\mu_x$ and covariance matrix $W_x = [I - FF']^{-1}W_\varepsilon$. Using the definition of the Shannon mutual information and the gaussian approximation to $p(x_k \mid N_{0:k})$, the mutual information between position at $k\Delta$ and the ensemble spiking activity in $(0, k\Delta]$ is (Cover, & Thomas, 1991; Twum-Danso, 1997)

$$\begin{aligned} I(x_k; N_{0:k}) &= -\int p(x_k)\log_2[p(x_k)]dx_k \\ &\quad -\left[-\int\int p(x_k \mid N_{0:k})\log_2[p(x_k \mid N_{0:k})]dx_k p(N_{0:k})\,dN_{0:k}\right] \\ &= 2^{-1}\log_2(2\pi e)^2 \int |W_x| / |W_{k|k}|\, p(N_{0:k})dN_{0:k}, \end{aligned} \tag{2.17}$$

where $|W_x|$ is the determinant of W_x and $p(N_{0:k})$ is the marginal probability mass function of the neural ensemble in $(0, k\Delta]$. Equation 2.17 can be evaluated at each step k of the algorithm and defines on average, the number of bits the random variable $N_{0:k}$, i.e. the ensemble neural spiking activity in $(0, k\Delta]$, provides about the random variable x_k, the animal's position at $k\Delta$. This is in contrast to the conditional entropy at $k\Delta$ which describes the uncertainty about the animal's position at $k\Delta$ given the ensemble spiking activity in $k\Delta$.The mutual information describes a property of the system whereas the instantaneous entropy provides a measure of uncertainty at a given time in a given realization of the ensemble spiking activity (Twum-Danso, 1997). Given the inhomogeneous Poisson model for the spike train and the $AR(1)$ model for the path, we easily can compute the integral in the last line of equation 2.17 by Monte Carlo.

3 Application

As in our previous work (Brown et al., 1998), we divide the experimental data into two parts and the analysis into two stages: encoding analysis and decoding analysis. In the encoding analysis, we use the first part of the experimental data to estimate the relation between spiking activity and position for each neuron for both animals and to estimate for each animal the parameters of the $AR(1)$ path model. In the decoding analysis, we apply the path and place cell models with their parameters estimated in the encoding analysis to the second part of the experimental data to decode the position of each animal from the ensemble spiking activity of its respective hippocampal neurons. We perform the decoding analysis by comparing the gaussian and Zernike decoding models across a range of learning rates and and two update intervals. For the optimal model, we analyze the mutual information between the spiking activity and the path. We begin with a brief review of the experimental protocol.

3.1 Experimental Protocol. We applied our paradigm to place cell spike train and position data recorded from two Long-Evans rats freely foraging in a circular environment 70 cm in diameter with walls 30 cm high and a fixed visual cue. A multiunit electrode array was implanted into the CA1 region of the hippocampus of each animal. The simultaneous activity of 34 (32) place cells was recorded from the electrode array while the two animals foraged in the environment for 25 (23) minutes. Simultaneous with the recording of the place cell activity, the position of each animal was measured at 30 Hz by a camera tracking the location of two infrared diodes mounted on the animal's headstage (Brown et al., 1998).

3.2 Encoding Analysis. The maximum likelihood estimates of the spatial gaussian and Zernike inhomogeneous Poisson rate functions computed from the first 15 (13) minutes for Animal 1 (2) are shown in Figure 1. Neuron 34 for Animal 1 and neurons 29 to 32 for Animal 2 had split fields. Therefore, for the analysis with the gaussian model, the spiking activity of each of these neurons was fit with two gaussian surfaces to capture this feature. The spiking activity of each neuron was also fit with both a single Zernike model and two Zernike models for comparison. The spatial model fits in Figure 1 are displayed along with smoothed spatial rate maps. The empirical rate maps are displayed as an approximation to the raw data.

In general, both the gaussian and Zernike model estimates of the place fields agree, with several noticeable exceptions. The Zernike place field estimates are concentrated in a smaller area and have a wider range of asymmetric shapes. The gaussian model attempts to capture the asymmetry by placing the center of the field outside the circle. The Zernike models can fit split fields with a single polynomial expansion, whereas two gaussian models are required to fit split fields. In a few cases when a place field is con-

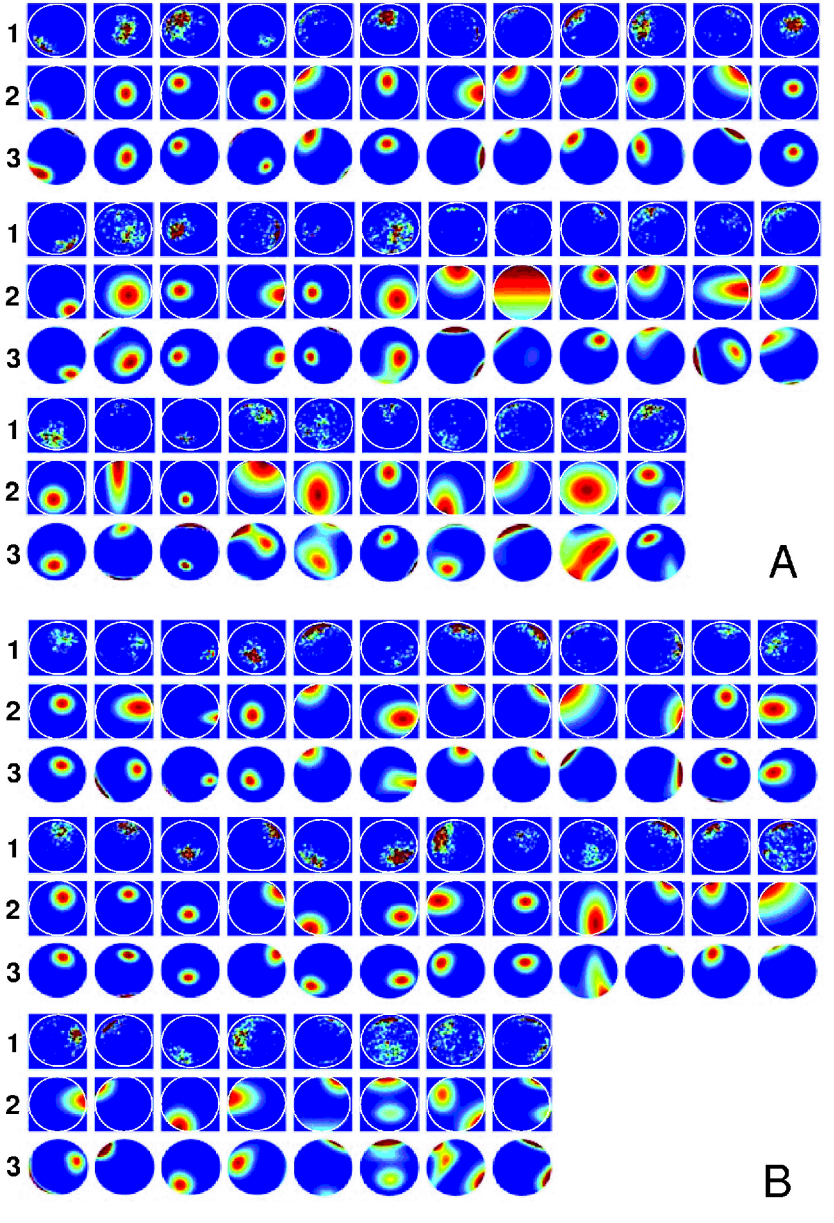


Figure 1: Spatial gaussian and Zernike place fields. Smoothed spatial spike histogram (row 1) and pseudocolor maps of place fields estimates from the spatial gaussian (row 2) and Zernike (row 3) models for the 34 place cells for Animal 1 (A) and the 32 place cells for Animal 2 (B).

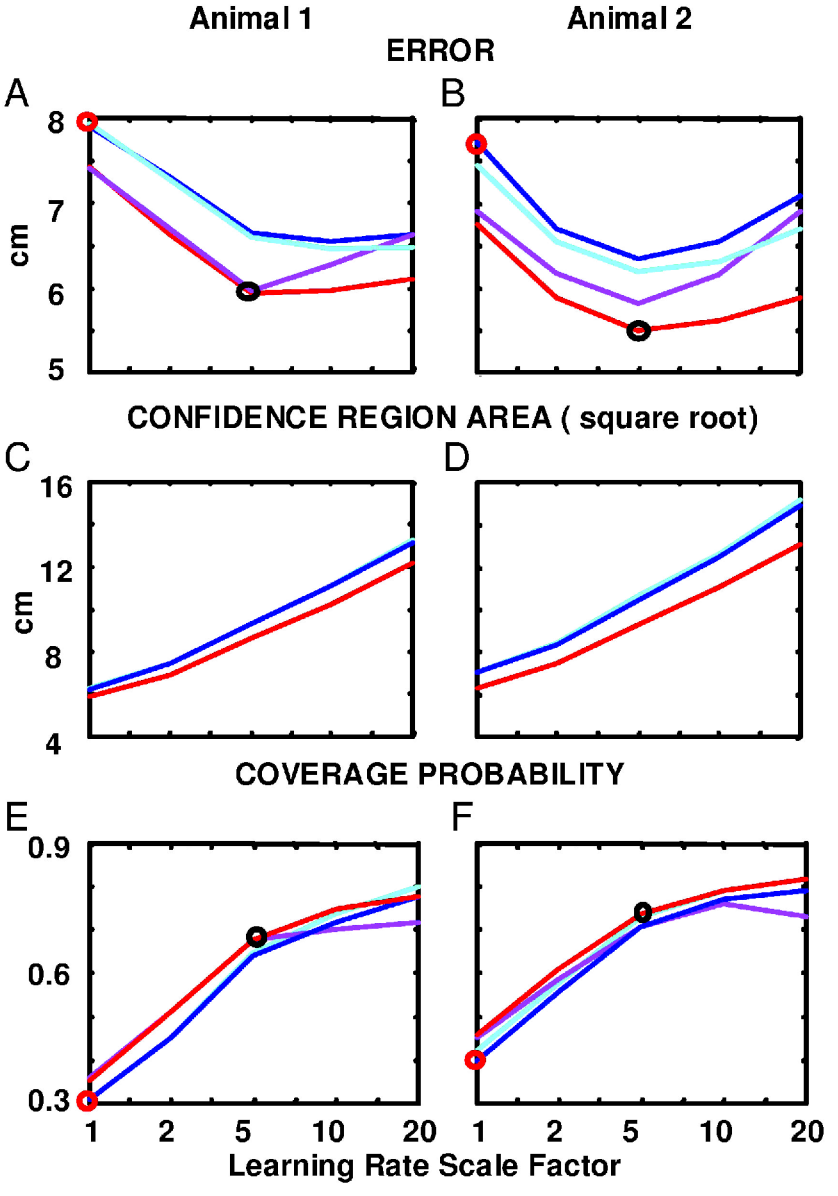
centrated on one border, the Zernike models produce an artifact of nonzero spiking activity on the opposite side of the environment. By the BIC, 29 (30) of the 34 (32) cells of Animal 1 (2) were best fit by the Zernike spatial model, and the remaining 5 (2) were best fit by the spatial gaussian model, suggesting that the former is an important improvement over the latter.

3.3 Decoding Analysis. We estimated the F and W_ε matrices for both Animal 1 and Animal 2 by maximum likelihood from the 13 (15) minutes of the respective path data in the encoding stage. The diagonal elements of the F matrices were 0.99 for both animals, and the off-diagonal elements were less than 0.01. This means that the decoding estimates from the $AR(1)$ model will be similar to those from the random walk model and that the determinants of the F matrices are close to 1. We decoded the position from the ensemble spiking activity of the 34 (32) neurons from the last 10 minutes of foraging for Animal 1 (2) using 20 different combinations of parameters and models for our algorithm (see equations 2.10–2.13). These were two rate function models, spatial gaussian or Zernike, two updating intervals, $\Delta = 3.3$ or 33.3 msec, and five values of LRSF with $R = 1, 2, 5, 10$ or 20 (see equation 2.10).

For both animals, the median decoding error was a minimum when the LRSF was 5 (see Figure 2), suggesting that below 5, the new spiking activity should be weighted more, whereas above 5, the new spiking activity should be weighted less. For the five LRSFs, the Zernike models with both 3.3 and 33 msec updating had smaller median decoding errors than the corresponding

Figure 2: *Facing page*. Summary of decoding algorithm performance. (A, B) Median decoding error computed as the difference between the true and estimated path at each decoding update. (C, D) Median square root of the 0.95 confidence region area. (E, F) Median coverage probability as a function of the learning rate scale factor for the spatial gaussian model with a 33 msec update (dark blue), the spatial gaussian model with a 3.3 msec update (light blue), the Zernike model with a 33 msec update (light red), and the Zernike model with a 3.3 msec update (dark red) for Animal 1 (left column) and Animal 2 (right column). The black circles in A and B indicate the minimum median error for the Zernike model with 3.3 msec updating and LRSF = 5. The red circles in A and B indicate the median error of the spatial gaussian model with a 33 msec updating and LRSF = 1. Although the spatial gaussian model also used the $AR(1)$ path model, the results are identical to the decoding results reported for this model in Brown et al. (1998) using the random walk path model. The black circles in E and F indicate the median coverage probability of the Zernike model with 3.3 msec updating and LRSF = 5. The red circles indicate the median coverage probability of the spatial gaussian model with 33 msec updating and LRSF = 1. Because the Zernike model with 3.3 msec updating and LRSF = 5 achieved for both animals nearly the highest median coverage probability with the smallest median error, we termed it the optimal Zernike model.

spatial gaussian models. For Animal 2, the 3.3 msec updating gave a smaller median error for both the Zernike and spatial gaussian models at all LRSFs, whereas for Animal 1, there was no difference in median error between the two sets of models for the 3.3 versus the 33 msec updating until the LRSF was greater than 5. The Zernike model with the 3.3 msec updating and an



LRSF of 5 gave the smallest median decoding error: 5.9 (5.5) cm for Animal 1 (2). This is an improvement of 2.0 (2.2) cm from the 7.9 (7.7) cm using the spatial gaussian model and either the random walk (Brown et al., 1998) or $AR(1)$ path models (the red circles in Figure 2). Of this improvement, 1.2 (1.3) cm for Animal 1 (2) was due to optimizing the learning rate, whereas 0.7 (0.7) cm was due to the switch from the spatial gaussian to the Zernike model.

Because increasing the LRSF increases the variance of the position estimate (see equations 2.11 and 2.13), the areas of the confidence regions necessarily grow with increasing LRSF. Beyond an LRSF of 2, this increase is faster for all algorithms for both animals (see Figure 2). For all the algorithms, the confidence region areas are smaller for the Zernike models than those of their corresponding spatial gaussian models. At an LRSF of 5, the median confidence region area for the Zernike model was 8.62 (9.32) cm^2 compared with 9.32 (10.52) cm^2 for Animal 1 (2). The true coverage probability for the 0.95 confidence regions (see equation 2.13) also increases with increasing LRSF. However, at an LRSF of 5, it reaches an approximate plateau for both animals for all algorithms. Up to an LRSF of 5, the coverage probability for both animals is greater for any Zernike model compared with its corresponding spatial gaussian model. The coverage probabilities plateau after 5 because even though the confidence region areas increase with higher learning rates (see Figures 2C and 2D), the higher learning rates beyond this point give less accurate decoding (see Figures 2A and 2B). At an LRSF of 5, for the Zernike model with 3.3 msec, the coverage probability of the 0.95 confidence regions was 0.67 (0.75) for Animal 1 (2) (black circles in Figures 2E and 2F). This is a significant improvement over the 0.31 (0.40) for the spatial gaussian model with either the random walk or $AR(1)$ path models (red circles in Figures 2E and 2F). We term the decoding algorithm with Zernike model with LRSF = 5 and updated at 3.3 msec the optimal Zernike model.

Illustrations of the recursive filter decoding algorithm with the spatial gaussian and optimal Zernike models are shown, along with a reverse correlation analysis, for a 60 sec continuous segment of the path from Animal 1 in Figure 3. The estimated trajectories from the optimal Zernike model resembled more closely the true trajectories than did those of the spatial gaussian model and had the smallest median error in each of the four 15 sec segments. In contrast, the reverse correlation algorithm performed poorly over the entire segment. The differences in performance of the algorithms in decoding the spike train data for both animals can be best appreciated in the videos of this analysis on our Web site (http://neurostat.mgh.harvard.edu).

To quantify further the relative performance of the decoding algorithms, we computed the decoding error defined for each algorithm as the difference between the true and estimated path at each 33 msec time step for the 10 min of decoding. The box plot summaries of these error distributions for the three decoding algorithms are shown in Figure 4. For Animal 1 (2),

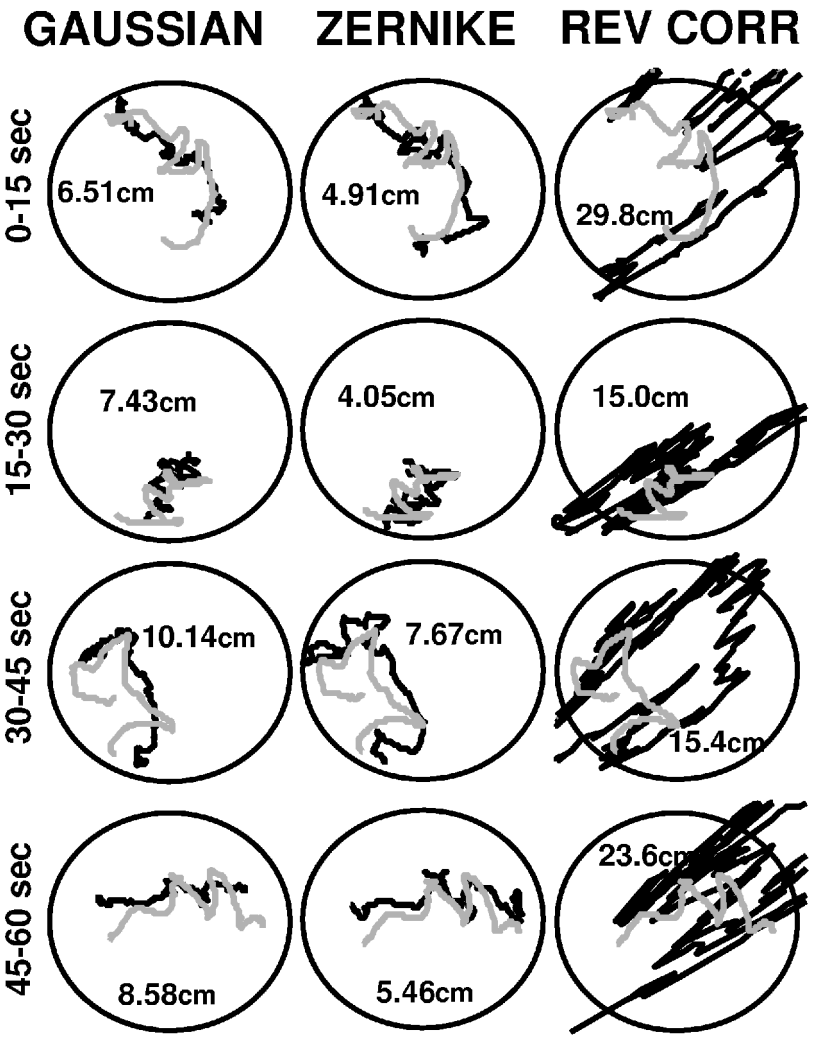


Figure 3: Continuous 60 sec segment of the true path (gray) displayed in four 15 sec intervals with the spatial gaussian decoding estimate with LRSF = 1 and 33 msec update (black, left column), the optimal Zernike model (black, center column), and the reverse correlation (black, right column). Numbers are the median error for the segment.

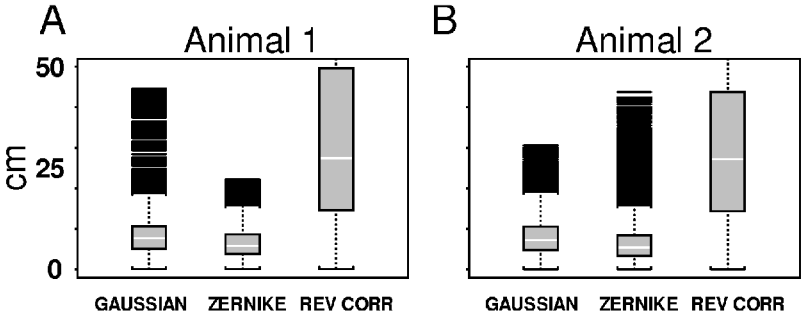


Figure 4: Box plot summary of decoding error (True position—estimated position) distributions. Box plot summaries of the decoding error distributions for the spatial gaussian model with LRSF = 1 and 33 msec (left), the optimal Zernike model (center), and the reverse correlation (right) for Animals 1(A) and 2(B). The lower border of the box is the 25th percentile of the distribution, and the upper border is the 75th percentile. The white bar within the box is the median of distribution. The distance between the 25th and 75th percentiles is the interquartile range (IQR). The lower (upper) whisker is at 1.5 the IQR below (above) the 25th (75th) percentile. All the black bars below (above) the lower (upper) whiskers are far outliers. For reference, less than 0.35% of the observations from a gaussian distribution would lie beyond the 75th percentile plus 1.5 × IQR.

the median decoding error for the spatial gaussian model was 7.7 (7.9) cm, with a 25th percentile of 5.2 (4.9) cm, a 75th percentile of 10.7 (10.9) cm, a minimum of 0.12 (0.06) cm, and a maximum of 44.6 (30.7) cm. For Animal 1 (2), the median decoding error for the Zernike model was 5.9 (5.5) cm, with a 25th percentile of 3.9 (3.3) cm, a 75th percentile of 8.6 (8.4) cm, a minimum of 0.8 (0.4) cm, and a maximum of 22.3 (43.8) cm. The decoding error distribution for Animal 1 was smaller for the optimal Zernike model compared with the spatial gaussian model (see Figure 4). For Animal 2, the decoding error distribution for the Zernike model was smaller up to the upper whisker of the box plot; however, beyond this point, this error distribution had a larger tail compared to the one for the corresponding spatial gaussian model. For both animals, the reverse correlation decoding errors were much larger. For Animal 1 (2), the median decoding error for the reverse correlation method was 27.5 (27.3) cm, with a 25th percentile of 14.7 (14.3) cm, a 75th percentile of 49.7 (43.8) cm, a minimum of 0.2 (0.7) cm, and a maximum of 143.3 (146.8) cm.

The optimal Zernike model was also noticeably more accurate than the spatial gaussian model near the borders of the environment (see Figure 5). This is because the Zernike model gave a more accurate description of the place fields (see Figure 1) and because, by equation 2.5, the support of the Zernike model is restricted to the circle. The more realistic trajectory es-

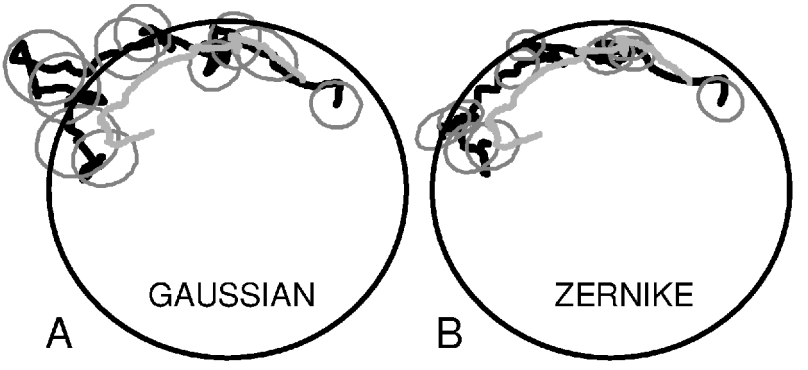


Figure 5: Decoding analysis: 95% confidence regions. Fifteen sec of true path (gray) optimal Zernike model estimate (black, A), the corresponding spatial gaussian model estimate (black, B), and 0.95 confidence regions (ellipses) computed at position estimates spaced 1.5 sec apart. The median decoding error for the Zernike (gaussian) model is 5.99 (7.28) cm, and the median confidence region area is 9.32 (10.62) cm^2 on this 5 sec segment.

timates, the smaller decoding error, and the greater coverage probability with smaller confidence regions (see Figures 2B and 2C, 4, and 5) show that the recursive filter decoding algorithm with the Zernike model was more accurate than using either this algorithm with the spatial gaussian model or the reverse correlation procedure.

3.4 Instantaneous Entropy, Entropy Rate, and Mutual Information. Along with the decoding estimate at each time step, we computed from the decoding analysis with the optimal Zernike model the instantaneous entropy, entropy rate, and the Shannon mutual information between the animal's position and the ensemble spiking activity. Figure 6 shows the instantaneous entropy, entropy rate, confidence region area, decoding error, and coverage probability for the 60 sec segment of data from Animal 1 considered in Figure 3. By equations 2.14 and 2.15, the number of bits the entropy of position given the ensemble spiking activity conveys about position (see Figure 6A) is directly proportional to the area of the 0.95 confidence region at that instant (see Figure 6C). The entropy provided a different assessment of the algorithm's performance compared with the decoding error. For example, at 12 sec, the decoding error was 20 cm (see Figure 6D), the confidence region area was 17^2 cm^2 (see Figure 6C), and the entropy was 9.3 bits (see Figure 6A). However, at 44 sec, the decoding error was nearly the same at 18^2 cm^2, yet the confidence region area was 7^2 cm^2 and the entropy was 8.0 bits. Because the entropy rate approximates the derivative of the entropy, it was positive when the instantaneous entropy increased and neg-

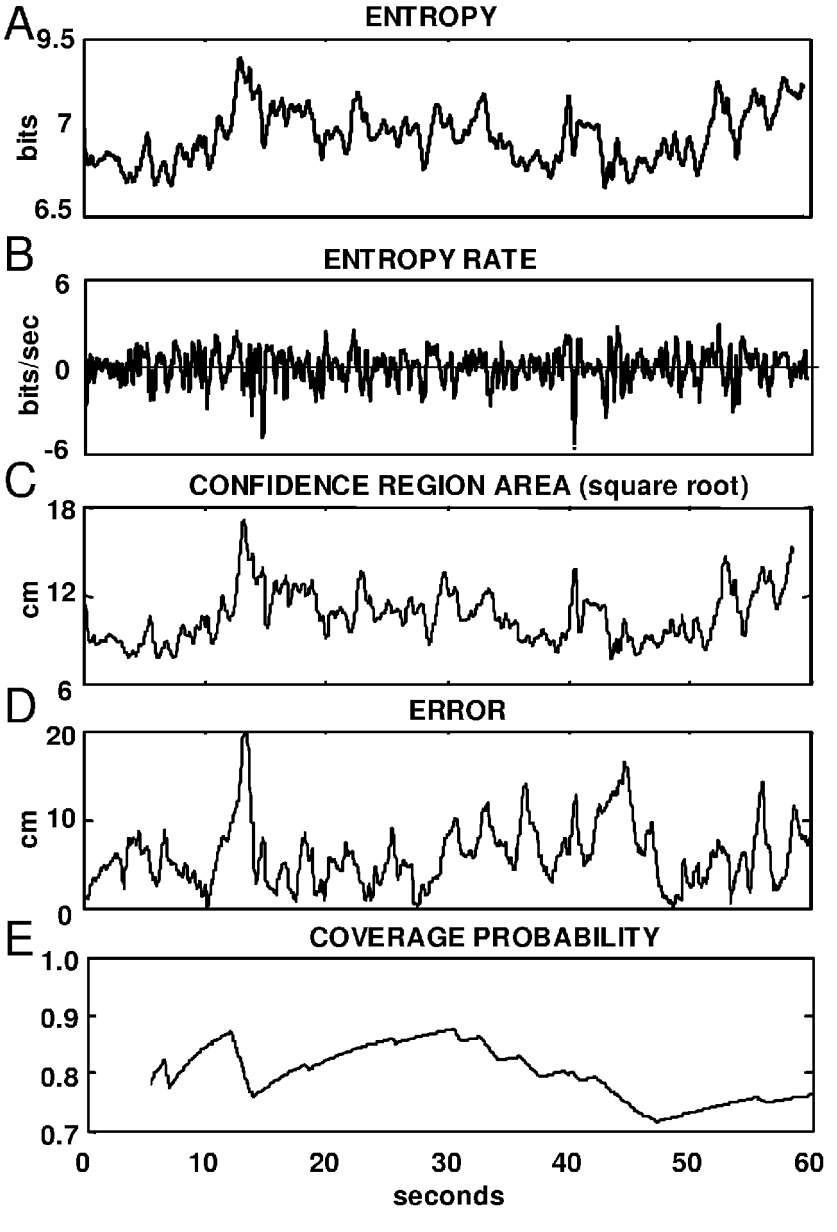


Figure 6: Decoding and mutual information analysis. (A) Instantaneous entropy. (B) Entropy rate. (C) Square root of the confidence region area. (D) Decoding error. (E) Coverage probability for the continuous 60 sec segment in Figure 3 computed by decoding with the optimal Zernike model.

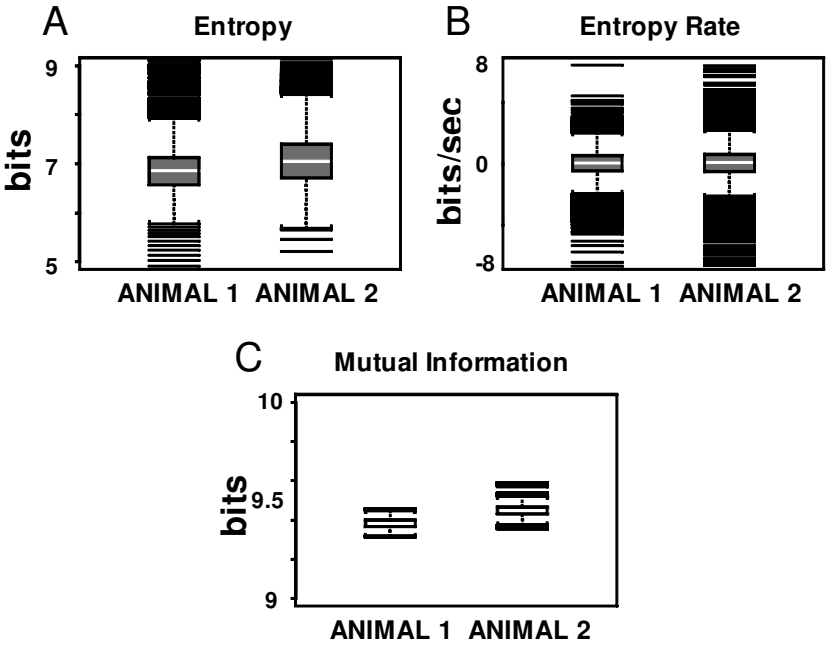


Figure 7: Box plots of the (A) Instantaneous entropy, (B) entropy rate, and (C) Shannon mutual information distributions for the optimal Zernike model for Animal 1 (left) and Animal 2 (right) for the 10 min of decoding. See Figure 4 for an interpretation of the box plots.

ative when it decreased (see Figure 6B). For both animals, we verified that at each time step k of the algorithm, $|W_{k-1|k}| > |W_{k-1|k-1}|$. Therefore, as stated in section 2.6, the negative entropy rates were due solely to the ensemble spiking activity. The mean and median of the entropy rate were zero during this 60 sec segment, suggesting that there was no overall increase or decrease in the uncertainty about position given the ensemble spiking activity.

The instantaneous coverage probabilities for this 60 sec segment range from 0.71 to 0.85 (see Figure 6E). The value of 0.73 at the end of the segment is closer to the total decoding stage coverage probability for this animal of 0.67. Together, the decoding error, the entropy, the entropy rate, confidence regions, and the coverage probability give a comprehensive assessment of algorithm performance not appreciated from applying any single measure alone.

To appreciate the range of entropy and entropy rates observed during the 10 minutes of decoding with the optimal Zernike model, we summarized as box plots in Figure 7 the distributions of these values for each animal. During 10 min of decoding with the optimal Zernike model, the median

entropy was 6.9 (7.0) bits for Animal 1 (2) with a minimum of 3.5 (5.2) bits and a maximum of 9.3 (12.0) bits (see Figure 7). For the entire 10 min of decoding stage, as in the analysis of the 60 sec segment of the ensemble spiking activity in Figure 6, the median information rate was 0.08 (−0.09) and the mean information rate was 0.002 (0.004) for Animal 1 (2). The corresponding 25th and 75th percentiles of the entropy rate distributions were −0.57 (−0.63) and 0.67 (0.73) for Animal 1 (2), respectively. These findings show that there is no overall growth or decline in the entropy of position given the ensemble spiking activity during the decoding time interval.

We computed the mutual information in equation 2.17 by using Monte Carlo methods to simulate 50 realizations of the path (equation 2.7) and the ensemble spiking activity given the path model (equation 2.3) and the maximum likelihood estimate of the Zernike model parameters computed in the encoding analysis. The median and mean mutual information between position and ensemble spiking activity is 9.38 (9.45) with 25th and 75th percentile of 9.37 (9.43) and 9.41 for Animal 1 (2), respectively.

4 Discussion

Our Bayes' filter previously estimated the rat's position with a median error of 8 cm during 10 min of decoding using ~ 30 place cells with a coverage probability of ~ 0.35 (Brown et al., 1998). The recursive point process decoding algorithm with the optimal Zernike model reduced the median error to less than 6 cm and increased the coverage probability to ~ 0.70, suggesting that the ensemble spiking activity carries more information than we previously estimated. At any instant, approximately 30% of the 10^5 (3×10^4) neurons in the hippocampus are active in a given environment (Wilson & McNaughton, 1993). Therefore, our results based on ~ 30 neurons suggest that this brain region maintains a precise, dynamic representation of the animal's spatial location.

4.1 Decoding, Coverage Probability, Entropy, and Mutual Information. The central concept in our decoding paradigm is the recursive estimation of the posterior probability density (see equation 2.8). The decoding estimates, entropy, mutual information, confidence regions, and coverage probability are all functions of the posterior probability density that can be computed directly once this density has been estimated. The relation among these functions is the difference between the message in the code, certain properties of the code, and the accuracy with which the algorithm reads the code. The decoded position estimate together with its 0.95 confidence region give a read-out of the message in the code at a given instant. In contrast, the entropy of the animal's position given the ensemble spiking activity, and the area of the 0.95 confidence region at that instant, are second-order properties of the code. The smaller the number of bits in the entropy, the smaller is the confidence region. In this way, the entropy rate,

like the changes in the areas of the confidence regions, identifies when the spiking activity contributes "meaningfully" to the decoding estimation. Because the mutual information is computed by integrating the entropy, with respect to the marginal probability mass function of the ensemble spiking activity (equation 2.17), it measures the average number of bits at each time point the ensemble provides about the path. Hence, the entropy provides a measure of uncertainty for the current experiment whereas the mutual information defines the number of bits on average, i.e. over many realizations, that the neural ensemble conveys about the animal's position.

The decoding error and the coverage probability measure the accuracy with which the algorithm reads the neural code. The coverage probability measures simultaneously the accuracy of the algorithm's first-order (position estimate) and second-order (position uncertainty) properties. While the coverage probability for no model reached the expected 0.95, the coverage probability of each Zernike model was greater, with confidence region areas that were smaller than those of the corresponding spatial gaussian model. Moreover, the coverage probability of the optimal Zernike model was greater than that of our original spatial gaussian model with either the random walk or $AR(1)$ path models. These findings suggest that the new algorithm with the new spatial model is more accurate.

Our computation of the instantaneous entropy and the entropy rate follows directly from applying the definition of entropy to the Bayes' rule Chapman-Kolmogorov updating formula in equations 2.8 and 2.9. We used these calculations to make explicit the relation between recursively computing a decoding estimate and computation of mutual information. In the current experiments, the entropy was constant at approximately 7 bits and the median and mean entropy rates were zero. This suggests that the uncertainty in position determined by the ensemble of hippocampal neurons did not change during the experiment. Because place fields do evolve as an animal moves through both familiar (Mehta, Lee, & Wilson, 2002) and novel environments (Frank, Stanley, & Brown, 2003), it would not have been surprising to see an increase in entropy and more positive entropy rates as the decoding proceeded. This is because, as the place fields evolve, their estimates computed during the encoding stage no longer represent the spatial location that the neuron encoded during the decoding stage. To carry out decoding in this case would require combining our decoding algorithms with adaptive estimation algorithms (Brown et al., 2001) to track place field evolution during decoding.

Equation 2.16 and our analysis in section 2.6 showed that by the construction of our algorithm, a negative entropy rate at any step of the algorithm or, equivalently, a decrease in the entropy came solely from the spiking activity and not from the $AR(1)$ path model. This was because the $AR(1)$ path model contributed at each step to a positive entropy rate because, at each step, $|W_{k|k-1}| > |W_{k-1|k-1}|$. Furthermore, because under the Poisson assumption used in the current analysis, the spiking activity of neurons in

nonoverlapping intervals is independent, the decrease in entropy due to the spiking activity did not arise from correlated activity. The point process conditional intensity framework is well suited to carry out the important next step of repeating the current entropy and mutual information analysis with models using history-dependent spiking activity. Our parametric modeling paradigm provides dynamic estimates of entropy, the entropy rate, and mutual information using the Bayes' rule and Chapman-Kolmogorov equations (see equations 2.8 and 2.9) to model explicitly the temporal evolution of the relation between position and ensemble spiking activity. The paradigm relates decoding and mutual information analyses and offers a plausible alternative to static mutual information analyses computed over long time intervals. Moreover, our approach to estimating entropy and mutual information obviates concerns about choosing word lengths and the accuracy of information estimation, particularly when word lengths are large relative to the length of the spike train (Victor, 2002).

4.2 The State-Space Paradigm Applied to Neural Spike Train Decoding. The current decoding algorithm is part of the paradigm we have been developing to conduct state estimation from point process measurements (Smith & Brown, 2003). We reviewed there the relation of our paradigm to other approaches to state estimation from point process observations. We previously compared our Bayes' filter algorithm to several other approaches (Brown et al., 1998). In this study, we compared our new decoding algorithm to the reverse correlation because it was not included in our previous comparison and because it is one of the most widely used methods for decoding ensemble neural spike train activity (Warland et al., 1997; Stanley et al., 1999; Serruya et al., 2002). The reverse correlation algorithm, while simple to implement, performed significantly worse than any version of our algorithm because it made no use of the spatial and temporal structure in this problem. This finding suggests that application of our algorithm to other decoding and neural prosthetic control problems where the reverse correlation methods have performed successfully (Bialek et al., 1991; Warland et al., 1997; Stanley et al., 1999; Wessberg et al., 2000; Serruya et al., 2002) may yield important improvements. The decoding error, confidence regions, and coverage probabilities used together are more meaningful measures of algorithm performance than the R^2 coefficient typically reported in reverse correlation analyses (Wessberg et al., 2000; Serruya et al., 2002; Taylor et al., 2002).

Equations 2.10 to 2.13 generalize our Bayes' filter decoding algorithm to one using arbitrary point process and linear stochastic state-space models. To define our new algorithm or, equivalently, the posterior probability density (see equation 2.8), it suffices to specify these two models. This general formulation was necessary for our current analysis because our original algorithm depended critically on the form of the spatial gaussian model (Brown et al., 1998) and, hence, could not be applied to the Zernike model. The current formulation of the algorithm can use any point process model

that can be defined in terms of a conditional intensity function (see equation A.1). The inhomogeneous Poisson observation model based on the Zernike expansion gave more accurate decoding results than the one based on the gaussian model for two reasons. First, the Zernike model was more flexible and represented more realistically the complex place field shapes in the circular environment, as indicated by the plots in Figure 1 and the BIC goodness-of-fit assessments. In particular, the support of the Zernike models was restricted to the circular environment, whereas the support of the gaussian model was all of $\mathbb{R}^2$. Second, the algorithm with the Zernike Poisson model decoded more accurately because this model more accurately estimated the probability of a neuron's spiking or not in a given location. The decoding update (see equation 2.11) "listens" to both spiking and silent neurons in each time window. Because few spikes occur within the 3.3 and 33 msec time windows, the new information (innovations) at each update comes mostly from neurons that do not spike. Use of the information from the neurons that do not spike in order to decode was a feature of our Bayes' filter algorithm (Brown et al., 1998). Wiener and Richmond (2003) recently noted this feature as well. This observation is consistent with the idea that once a brain region develops a representation, which neurons do not spike, or are inhibited, can convey as much information as those that do spike.

We chose an $AR(1)$ model as our linear stochastic state-space model. There were no differences between the decoding results for the random walk model compared with the $AR(1)$ model. However, because the $AR(1)$ model is stationary, it could be used, unlike the random walk model, to compute meaningful estimates of mutual information. Scaling the model's white noise variance (see equation 2.7) scaled the learning rate. This scaling enhanced significantly the algorithm's accuracy by balancing the weight given to the prediction based on the previous estimate (see equation 2.10) and the modification produced by the innovations (see equation 2.12).

Further improvements in our algorithm can be readily incorporated into the current framework. First, these include combining the Zernike spatial model with models of known hippocampal temporal dynamics such as bursting (Quirk & Wilson, 1999; Barbieri et al., 2001), theta phase modulation (Skaggs, McNaughton, Wilson, & Barnes, 1996; Brown et al., 1998; Jensen & Lisman, 2000), and phase precession (O'Keefe & Reece, 1993). Second, we can consider both higher-order linear state-space models (Shoham, 2001; Gao, Black, Bienenstock, Shoham, & Donoghue, 2002; Smith & Brown, 2003) or, perhaps, nonlinear state-space models to describe the path dynamics more accurately (Kitagawa & Gersh, 1996). Finally, we developed gaussian approximations to evaluate the Bayes' rule and Chapman-Kolmogorov equations in equations 2.8 and 2.9 as a plausible initial approach. We can improve our algorithm and evaluate the accuracy of this approximation by applying nongaussian approximations (Pawitan, 2001), exact numerical integration techniques (Genz & Kass, 1994; Kitagawa & Gersh, 1996), and Monte Carlo methods (Doucet, DeFreitas, Gordon, & Smith, 2001; Shoham, 2001)

to evaluate these equations. We suggest our point process linear state-space model framework as a broadly applicable approach for studying decoding problems as well as for designing algorithms to control neural prosthetic devices and brain-machine interfaces.

Appendix: Derivation of the Decoding Algorithm

During the decoding interval $(0, T^d]$, we record the spiking activity of C neurons. Let $0 < u_1^c < u_2^c <, \ldots, < u_{J^c}^c \leq T^d$ be the set of J^c spike times from neuron c for $c = 1, \ldots, C$. For $t \in (0, T^d]$, let $N_{0:t}^c$ be the sample path of the spike times from neuron c in $(0, t]$. It is defined as the event $N_{0:t}^c = \{0 < u_1^c < u_2^c <, \ldots, < u_j^c \leq t \cap N^c(t) = j\}$, where $N^c(t)$ is the number of spikes in $(0, t]$ from neuron c and $j \leq J^c$. Let $N_{0:t} = \{N_{0:t}^1, \ldots, N_{0:t}^c\}$ be the ensemble spiking activity in $(0, t]$. To define a general point process probability model for a neuron, it suffices to specify its conditional intensity function for $t \in (0, T^d]$, defined as

$$\lambda^c(t \mid N_{0:t}) = \lim_{\Delta \to 0} \frac{\Pr(N^c(t + \Delta) - N^c(t) = 1 \mid N_{0:t})}{\Delta}. \tag{A.1}$$

The conditional intensity function is a history-dependent rate function that generalizes the definition of the Poisson rate (Barbieri et al., 2001; Brown et al., 2001; Brown et al., 2002; Brown et al., 2003; Daley & Vere-Jones, 2003). If the point process is an inhomogeneous Poisson process, then the conditional intensity is $\lambda^c(t \mid N_{0:t}) = \lambda^c(t)$. It follows that $\lambda^c(t \mid N_{0:t})\Delta$ is the probability of a spike in $[t, t + \Delta)$ when there is history dependence in the spike train.

From equation A.1, we denote the point process observation model on the updating lattice by the conditional intensity function $\lambda^c(x_k \mid N_{0:k})$ for neuron c, $c = 1, \ldots, C$ evaluated at position x_k, the animal's position at $k\Delta$, and given $N_{0:k}$, the spiking history in $(0, k\Delta]$. For small Δ, and under the assumption that the spiking activity of the neurons in the ensemble is conditionally independent, the joint probability density or the likelihood function of the spike train observations in the interval $((k-1)\Delta, k\Delta]$ is (Brown et al., 1998; Smith & Brown, 2003)

$$\begin{aligned} p(N_{k-1:k} \mid x_k, N_{0:k}) &= \prod_{c=1}^{C} p(N_{k-1:k}^c \mid x_k, N_{0:k}) \\ &= \exp\left[\sum_{c=1}^{C} N_{k-1:k}^c \log \lambda^c(x_k \mid N_{0:k}) \right. \\ &\qquad \left. - \sum_{c=1}^{C} \lambda^c(x_k \mid N_{0:k})\Delta\right]. \end{aligned} \tag{A.2}$$

The probability model in equation A.2 depends on the parameter $\hat{\zeta}^c$, estimated in the encoding analysis for each neuron. We have omitted this parameter to simplify the notation.

To evaluate the system in equations 2.8 and 2.9, we apply the maximum a posteriori derivation of the Kalman filter and approximate the $p(x_k \mid N_{0:k})$ as gaussian probability densities by recursively computing their means (modes) and negative inverse of the Hessian matrices (covariance matrices) (Mendel, 1995; Brown et al., 1998). To initiate our construction of the recursive algorithm, we assume that the mean $x_{k-1|k-1}$ and covariance matrix $W_{k-1|k-1}$ have been estimated at time $(k-1)\Delta$. That is, we take $p(x_{k-1} \mid N_{0:k-1})$, the posterior probability density at time $(k-1)\Delta$, to be the gaussian probability density with mean $x_{k-1|k-1}$ and covariance matrix $W_{k-1|k-1}$. The next step is to compute $p(x_k \mid N_{0:k-1})$, the one-step prediction probability density at time $k\Delta$. This is, the probability density of the predicted position at $k\Delta$ given the spiking activity in $(0, (k-1)\Delta]$. To do this, we must evaluate equation 2.9. By equation 2.7, the state model, $p(x_k \mid x_{k-1})$, is the gaussian probability density with mean $\mu_x + Fx_{k-1}$ and covariance matrix RW_ε. Because both $p(x_{k-1} \mid N_{0:k-1})$ and $p(x_k \mid x_{k-1})$ are gaussian probability densities and x_{k-1} enters linearly into the mean of both, $p(x_{k-1} \mid N_{0:k-1})$ is also a gaussian probability density and suffices to compute the mean $x_{k|k-1}$ and covariance matrix $W_{k|k-1}$ in order to define it. It follows from standard properties of integrals of gaussian functions that the mean and covariance matrix are respectively defined as

$$x_{k|k-1} = E(x_k \mid N_{0:k}) = \mu_x + Fx_{k-1|k-1} \tag{A.3}$$

$$W_{k|k-1} = \mathrm{Var}(x_k \mid N_{0:k}) = FW_{k-1|k-1}F' + RW_\varepsilon. \tag{A.4}$$

Equations A.3 and A.4 are, respectively, the one-step prediction estimate and the one-step prediction variance, or learning rate, in equations 2.10 and 2.11 of the algorithm. The probability density defined by equations A.3 and A.4 is

$$p(x_k \mid N_{0:k-1}) = (2\pi^{\frac{1}{2}}|W_{k|k-1}|)^{-p} \exp\left\{-\frac{1}{2}(x_k - x_{k|k-1})' W_{k|k-1}^{-1}(x_k - x_{k|k-1})\right\}. \tag{A.5}$$

Substituting equations A.2 and A.5 into equation 2.8 and neglecting the denominator and other terms that do not depend on position, we obtain an explicit expression for the probability density of position at time $k\Delta$ given the spiking activity in $(0, k\Delta]$:

$$p(x_k \mid N_{0:k}) \propto \exp\left\{-\frac{1}{2}(x_k - x_{k|k-1})' W_{k|k-1}^{-1}(x_k - x_{k|k-1})\right\} \times \exp\left[\sum_{c=1}^{C} N_{k-1:k}^c \log \lambda^c(x_k \mid N_{0:k}) - \sum_{c=1}^{C} \lambda^c(x_k \mid N_{0:k})\Delta\right]. \tag{A.6}$$

To evaluate equation 2.8, or equivalently equation A.6, we derive a gaussian approximation to it by deriving a quadratic approximation to $\log p(x_k \mid N_{0:k})$. This corresponds to taking the first two terms in its Taylor series expansion of $\log p(x_k \mid N_{0:k})$ (Tanner, 1996; Brown et al., 1998). From equation A.6, $\log p(x_k \mid N_{0:k})$ is

$$\log p(x_k \mid N_{0:k}) \propto -\frac{1}{2}(x_k - x_{k|k-1})' W_{k|k-1}(x_k - x_{k|k-1}) + \sum_{c=1}^{C} N_{k-1:k}^{c} \log \lambda^c(x_k \mid N_{0:k}) - \sum_{c=1}^{C} \lambda^c(x_k \mid N_{0:k})\Delta. \quad \text{(A.7)}$$

The gradient and Hessian matrix required for the quadratic approximation are, respectively,

$$\nabla \log p(x_k \mid N_{0:k}) = -W_{k|k-1}^{-1}(x_k - x_{k|k-1}) + \sum_{c=1}^{C} \nabla \log \lambda^c(x_k \mid N_{0:k}) \times [N_{k-1:k}^{c} - \lambda^c(x_k \mid N_{0:k})\Delta], \quad \text{(A.8)}$$

$$\nabla^2 \log p(x_k \mid N_{0:k}) = -W_{k|k-1}^{-1} + \sum_{c=1}^{C} \nabla^2 \log \lambda^c(x_k \mid N_{0:k}) \times [N_{k-1:k}^{c} - \lambda^c(x_k \mid N_{0:k})\Delta] - \nabla \log \lambda^c(x_k \mid N_{0:k})[\nabla \lambda^c(x_k \mid N_{0:k})\Delta]'. \quad \text{(A.9)}$$

Solving for x_k in equation A.8 yields the posterior mode in equation 2.12, and because under the gaussian approximation the covariance matrix is the negative inverse of the Hessian (Tanner, 1996; Brown et al., 1998), equation A.9 yields equation 2.13. Equations A.8 and A.9 are generalizations of equations A.12 and A.13 in Brown et al. (1998) for an arbitrary point process model.

Acknowledgments

Support was provided in part by NIH grants MH59733, MH61637, MH65018, and DA015644; NSF grant 0081548; and grants from DARPA and ONR. We are grateful to Jonathan Victor for helpful discussions and to an anonymous referee whose suggestions helped significantly improve the work presented in this article.

References

Barbieri, R., Frank, L. M., Quirk, M. C., Wilson, M. A., & Brown, E. N. (2002). Construction and analysis of non-gaussian place field models of neural spiking activity. *Neurocomputing, 44–46*, 309–314.

Barbieri, R., Quirk, M. C., Frank, L. M., Wilson, M. A., & Brown, E. N. (2001). Construction and analysis of non-Poisson stimulus-response models of neural spiking activity. *J. Neurosci. Methods, 105,* 25–37.

Bialek, W., Rieke, F. de Ruyter van Stevenick, R. R., & Warland, D. (1991). Reading a neural code. *Science, 252,* 1854–1857.

Born, M., & Wolf, E. (1989). *Principles of optics* (6th ed.). New York: Pergamon.

Box, G.E.P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis, forecasting and control* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Brown, E. N., Barbieri, R., Eden, U. T., & Frank, L. M. (2003). Likelihood methods for neural data analysis (pp. 253–286). In J. Feng (Ed.), *Computational neuroscience: A comprehensive approach.* London: Chapman and Hall.

Brown E. N., Barbieri R., Ventura V., Kass, R. E., & Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comp., 14*(2), 325–346.

Brown, E. N., Frank, L. M., Tang, D., Quirk, M., & 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,* 7411–7425.

Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., & Solo V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *Proceedings of the National Academy of Sciences, 98,* 12261–12266.

Chapin, J. K., Moxon, K. A., Markowitz, R. S., & Nicolelis, M.A.L. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neurosci., 7,* 664–670.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory.* New York: Wiley.

Daley, D., & Vere-Jones, D. (2003). *An introduction to the theory of point process* (2nd ed.). New York: Springer-Verlag.

Dan, Y., Alonso, J. M., Usrey, W. M., & Reid, R. C. (1998). Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neuroscience, 1*(6), 501–507.

Donoghue, J. P. (2002). Connecting cortex to machines: Recent advances in brain interfaces. *Nature Neurosci. Supplement., 5,* 1085–1088.

Doucet, A., DeFreitas, N., Gordon, N., & Smith, A. (2001). *Sequential Monte Carlo methods in practice.* New York: Springer-Verlag.

Fahrmeir, L., & Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models* (2nd ed.). New York: Springer-Verlag.

Frank, L. M., Stanley, G. B., & Brown E. N. (2003, November). *New places and new representations: Place field plasticity in the hippocampus.* Paper presented at the Society for Neuroscience conference, New Orleans, LA.

Gao, Y., Black, M. J., Bienenstock, E., Shoham, S., & Donoghue, J. (2002). Probabilistic inference of hand motion from neural activity in motor cortex. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems, 14.* Cambridge, MA: MIT Press.

Genz, A., & Kass, R. E. (1994). Subregion-adaptive integration of functions having a dominant peak. *J. Comp. Graph. Stat., 6,* 92–11.

Georgopoulos, A. P., Kettner, R. E., & Schwartz, A. B. (1986). Neuronal population coding of movement direction. *Science, 233*, 1416–1419.

Jensen, O., & Lisman, J. E. (2000). Position reconstruction from an ensemble of hippocampal place cells: Contribution of theta phase coding. *J. Neurophysiol., 83*, 2602–2609.

Johnson, D. H., Gruner, C. M., Baggerly, K., & Shehagiri, C. (2001). Information-theoretic analysis of neural coding. *J. Comp. Neurosci., 10*, 47–69.

Kitagawa, G., & Gersh, W. (1996). *Smoothness priors analysis of time series.* New York: Springer-Verlag.

Mehta, M. R., Lee, A. L., & Wilson, M. A. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature, 417*(6), 741–746.

Mendel, J. M. (1995). *Lessons in estimation theory for signal processing, communications, and control.* Englewood Cliffs, NJ: Prentice Hall.

Nirenberg, S., Carcieri, S. M., Jacobs, A. L., & Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature, 411*, 698–701.

O'Keefe J., & Dostrovsky J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Res., 34*, 171–175.

O'Keefe, J., & Nadel, N. (1978). *The hippocampus as a cognitive map.* New York: Oxford University Press.

O'Keefe, J., & Reece, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus, 3*, 317–330.

Pawitan, Y. (2001). *All likelihood: Statistical modelling and inference using likelihood*. New York: Oxford University Press.

Pouget, A., Dayan, P., & Zemel R. (2000). Information processing with population codes. *Nature Reviews Neurosci., 1*, 125–132.

Priestley, M. B. (1981). *Spectral analysis and time series*. Orlando, FL: Academic Press.

Quirk, M. C., & Wilson, M. A. (1999). Interaction between spike waveform classification and temporal sequence detection. *J. Neurosci. Methods, 94*, 41–52.

Reich, D. S., Melcher, F., & Victor, J. D. (2001). Independent and redundant information in nearby cortical neurons. *Science, 294*, 2566–2568.

Reinagel, P., & Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *J. Neurosci., 20*, 5392–5400.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., & Donoghue, J. P. (2002). Instant neural control of a movement signal. *Nature, 416*, 141–142.

Shoham, S. (2001). *Advances towards an implantable motor cortical interface.* Unpublished doctoral dissertation, University of Utah.

Skaggs, W. E., McNaughton, B. L. Gothard, K., & Marcus, E. (1993). An information-theoretic approach to deciphering the hippocampal code. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing, 5* (pp. 1030–1037). San Mateo, CA: Morgan Kaufmann.

Skaggs, W. E., McNaughton, B. L., Wilson, M. A., & Barnes, C. A. (1996). Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus, 6*, 149–172.

Smith, A. C., & Brown, E. N. (2003). Estimating a state-space model from point process observations. *Neural Comp., 15*, 965–991.

Stanley, G. B., Li, F. F., & Dan Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J. Neurosci., 19*, 8036–8042.

Strong, S. P., Koberle, R., Ruyter van Steveninck, R. R. de, & Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Let., 80*, 197–200.

Tanner, M. A. (1996). *Tools for statistical inference*. New York: Springer-Verlag.

Taylor, D. M. Tillery, S.I.H., & Schwartz, A. B. (2002). Direct cortical control of 3D neuroprosthetic devices. *Science, 296*, 1829–1832.

Twum-Danso, N. T. (1997). Estimation, information and neural signals. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Victor, J. D. (2002). Binless strategies for estimation of information from neural data. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys., 66*(5–1), 051903.

Warland, D. K., Reinagel, P., & Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol., 78*, 2336–2350.

Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., & Nicolelis, M. A. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature, 408*, 361–365.

Wickelgren, I. (2003). Tapping the mind. *Science, 299*, 496–499.

Wiener, M. C., & Richmond, B. J. (2003). Decoding spike trains instant by instant using order statistics and the mixture-of-Poissons model. *J. Neurosci., 23*, 2394–2406.

Wilson, M. A. & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science, 261*, 1055–1058.

Zhang, K., Ginzburg, I., McNaughton, B. L., & Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol., 79*, 1017–1044.

Received April 2, 2003; accepted July 30, 2003.