The relationship between optimal and biologically plausible decoding of stimulus velocity in the retina

Edmund C. Lalor,[1],* Yashar Ahmadian,[2] and Liam Paninski[2]

[1]*Trinity Centre for Bioengineering and Institute of Neuroscience, Trinity College Dublin, College Green, Dublin 2, Ireland*
[2]*Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, New York 10027, USA*
*Corresponding author: edlalor@tcd.ie

A major open problem in systems neuroscience is to understand the relationship between behavior and the detailed spiking properties of neural populations. We assess how faithfully velocity information can be decoded from a population of spiking model retinal neurons whose spatiotemporal receptive fields and ensemble spike train dynamics are closely matched to real data. We describe how to compute the optimal Bayesian estimate of image velocity given the population spike train response and show that, in the case of global translation of an image with known intensity profile, on average the spike train ensemble signals speed with a fractional standard deviation of about 2% across a specific set of stimulus conditions. We further show how to compute the Bayesian velocity estimate in the case where we only have some *a priori* information about the (naturalistic) spatial correlation structure of the image but do not know the image explicitly. As expected, the performance of the Bayesian decoder is shown to be less accurate with decreasing prior image information. There turns out to be a close mathematical connection between a biologically plausible "motion energy" method for decoding the velocity and the Bayesian decoder in the case that the image is not known. Simulations using the motion energy method and the Bayesian decoder with unknown image reveal that they result in fractional standard deviations of 10% and 6%, respectively, across the same set of stimulus conditions. Estimation performance is rather insensitive to the details of the precise receptive field location, correlated activity between cells, and spike timing. © 2009 Optical Society of America

OCIS codes: 330.4060, 330.4150.

1. INTRODUCTION

The question of how different attributes of a visual stimulus are represented by populations of cells in the retina has been addressed in a number of recent studies [1–8]. This field has received a major boost with the advent of methods for obtaining large-scale simultaneous recordings from multiple retinal ganglion neurons that almost completely tile a substantial region of the visual field [9,10]. The utility of this new method for understanding the encoding of behaviorally relevant signals was exemplified by [4], where the authors examined the question of how reliably visual motion was encoded in the spiking activity of a population of macaque parasol cells. These authors used a simple moving stimulus and attempted to estimate the velocity of that stimulus from the resulting spike train ensemble; this analysis pointed to some important constraints on the visual system's ability to decode image velocity given noisy spike train responses. We will explore these issues in more depth in this paper.

In parallel to these advances in retinal recording technology, significant recent advances have also been made in our ability to model the statistical properties of populations of spiking neurons. For example, a statistical model of a complete population of primate parasol retinal ganglion cells (RGCs) was recently described [7]. This model was fit using data acquired by the array recording techniques mentioned above and includes spike-history

effects and cross-coupling between cells of the same kind and of different kinds (i.e., ON and OFF cells). The authors demonstrated that the model accurately captures the stimulus dependence and spatiotemporal correlation structure of RGC population responses, and allows several insights to be made into the retinal neural code. One such insight concerns the role of correlated activity in preserving sensory information. Using pseudorandom binary stimuli and Bayesian inference, they reported that stimulus decoding based on the spiking output of the model preserved 20% more information when knowledge of the correlation structure was used than when the responses were considered independently [7].

At the psychophysical level, Bayesian inference has been established as an effective framework for understanding visual perception [11]; some recent notable applications to understanding visual velocity processing include [12–17]. In particular, [14] argued that a number of visual illusions actually arise naturally in a system that attempts to estimate local image velocity via Bayesian methods (though see also [18,19]).

Links between retinal coding and psychophysical behavior have also been recently examined using Bayesian methods; [20,21], for example, examine the contribution of turtle RGC responses to velocity and acceleration encoding. This study reported that the instantaneous firing rates of *individual* turtle RGCs contain information about

speed, direction, and acceleration of moving patterns. The firing-rate-based Bayesian stimulus reconstruction carried out in that study involved a couple of key approximations. These included the assumptions that RGCs generate spikes according to Poisson statistics and that they do so independently of each other. The work of [7] emphasizes that these assumptions are unrealistic, but the impact of detailed spike timing and correlation information on velocity decoding remains uncertain.

The primary goal of this paper is to investigate the fidelity with which the velocity of a visual stimulus may be estimated, given the detailed spiking responses of the primate RGC population model of [7], using Bayesian decoders, with and without full prior knowledge of the image. We begin by describing the mathematical construction of the Bayesian decoders, and then compare these estimates to those based on a biologically plausible "net motion signal" derived directly from the spike trains without any prior image information [4]. We derive a mathematical connection between these decoders and investigate the decoders' performance through a series of simulations.

2. METHODS

A. Model
The generalized linear model (GLM) [22,23] for the spiking responses of the sensory network used in this study was described in detail in [7]. It consists of an array of ON and OFF retinal ganglion cells (RGC) with specific baseline firing rates. Given the spatiotemporal image movie sequence, the model generates a mean firing rate for each cell, taking into account the temporal dynamics and the center-surround spatial stimulus filtering properties of the cells. Then, incorporating spike history effects and cross-coupling between cells of the same type and of the opposite type, it generates spikes for each cell as a stochastic point process.

In response to the visual stimulus $\mathbf{I}$, the ith cell in the observed population emits a spike train, which we represent by a response function

$$r_i(t) = \sum_{\alpha} \delta(t - t_{i,\alpha}), \tag{1}$$

where each spike is represented by a delta function, and $t_{i,\alpha}$ is the time of the αth spike of the ith neuron. We use the shorthand notation $\mathbf{r}_i$ and $\mathbf{r}$ for the response function of one neuron and the collective spike train responses of all neurons, respectively. The stimulus $\mathbf{I}$ represents the spatiotemporal luminance profile $I(\mathbf{n}, t)$ of a movie as a function of the pixel position $\mathbf{n}$ and time t.

In the GLM framework, the intensity functions (instantaneous firing rate) of the responses $\mathbf{r}_i$ are given by [7,24–26]

$$\lambda_i(t) \equiv f\left(b_i + J_i(t) + \sum_{j,\beta} h_{ij}(t - t_{j,\beta})\right), \tag{2}$$

where $f(\cdot)$ is a positive, strictly increasing rectifying function. As in [7], we adopt the choice $f(\cdot) = \exp(\cdot)$. The b_i represents the log of the baseline firing rate of the cell, the coupling terms h_{ij} model the within- and between-neuron spike history effects noted above, and the stimulus input

$J_i(t)$ is obtained from $\mathbf{I}$ by linearly filtering the spatiotemporal luminance,

$$J_i(t) = \int \int k_i(t - \tau, \mathbf{n}) I(\tau, \mathbf{n}) \mathrm{d}^2 \mathbf{n} \mathrm{d}\tau, \tag{3}$$

where $k_i(t, \mathbf{n})$ is the spatiotemporal receptive field of the cell i. The parameters for each cell were fit using 7 min of spiking data recorded during the presentation of a nonrepeating stimulus, with the baseline log firing rate being a constant and the various filter parameters being fit using a basis of raised cosine "bumps" [7]. Given Eq. (2), we can write down the point process log-likelihood in the standard way [27]

$$\log p(\mathbf{r}|\mathbf{I}) \equiv \sum_{i,\alpha} \log \lambda(t_{i,\alpha}) - \sum_i \int_0^T \lambda_i(t) \mathrm{d}t. \tag{4}$$

For movies arising from images rigidly moving with constant velocity $\mathbf{v}$ we have

$$I(t, \mathbf{n}) = x(\mathbf{n} - \mathbf{v}t), \tag{5}$$

where $x(\mathbf{n})$ is the luminance profile of a fixed image. Substituting Eq. (5) into Eq. (3) and shifting the integration variable $\mathbf{n}$ by $\mathbf{v}\tau$, we obtain

$$J_i(t) = \int \mathcal{K}_{i,\mathbf{v}}(t; \mathbf{n}) x(\mathbf{n}) \mathrm{d}^2 \mathbf{n}, \tag{6}$$

where we defined

$$\mathcal{K}_{i,\mathbf{v}}(t; \mathbf{n}) \equiv \int k_i(t - \tau, \mathbf{n} + \mathbf{v}\tau) \mathrm{d}\tau. \tag{7}$$

In the following we replace $p(\mathbf{r}|\mathbf{I})$ with its equivalent $p(\mathbf{r}|\mathbf{x}, \mathbf{v})$ [since, via Eq. (5), $\mathbf{I}$ is given in terms of $\mathbf{x}$ and $\mathbf{v}$] and use the short-hand matrix notation $\mathbf{J}_i = \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}$ for Eq. (6). An important point is that in the case of a convex and log-concave GLM nonlinearity, $f(\cdot)$ [conditions that are true for our choice, $f(\cdot) = \exp(\cdot)$], the GLM log-likelihood, Eq. (4), is a concave function of $\mathbf{x}(\mathbf{n})$.

B. Decoding
In order to estimate the speed of the moving bar given the simulated output spike trains $\mathbf{r}$ of our RGC population, we employed three distinct methods. The first method involved a Bayesian decoder with full image information, the second method utilized a Bayesian decoder with less than full image information, while the third method involved an "energy-based" algorithm introduced by [4] that used no explicit prior knowledge of the image. For reasons that will become clear, these decoders will be hereafter known as the optimal decoder, the marginal decoder, and the energy method, respectively. Given a simulated output spike train ensemble, we use each of these methods to estimate the speed of the stimulus that evoked the ensemble by maximizing some function across a range of possible or "putative" speeds.

1. Bayesian Velocity Estimation
To compute the optimal Bayesian velocity decoder we need to evaluate the posterior probability for the velocity

$p(\mathbf{v}|\mathbf{r})$ conditional on the observed spike trains $\mathbf{r}$. Given a prior distribution $p_v(\mathbf{v})$, from Bayes' rule we obtain

$$p(\mathbf{v}|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{v})\mathbf{p}_v(\mathbf{v})}{\sum_{\mathbf{v}'} p(\mathbf{r}|\mathbf{v}')\mathbf{p}_v(\mathbf{v}')}. \qquad (8)$$

If the image $\mathbf{x}$ (e.g., a narrow bar with a luminance distinct from the background) is known to the decoder, then we can replace $p(\mathbf{r}|\mathbf{v})$ with the likelihood function $p(\mathbf{r}|\mathbf{x},\mathbf{v})$, obtaining

$$p(\mathbf{v}|\mathbf{r},\mathbf{x}) = \frac{p(\mathbf{r}|\mathbf{x},\mathbf{v})\mathbf{p}_v(\mathbf{v})}{\sum_{\mathbf{v}'} p(\mathbf{r}|\mathbf{x},\mathbf{v}')\mathbf{p}_v(\mathbf{v}')}. \qquad (9)$$

$p(\mathbf{r}|\mathbf{x},\mathbf{v})$ is provided by the forward model Eq. (4), and therefore computation of the posterior probability is straightforward in this case.

Alternatively, if the image is not fully known, we represent the decoder's uncertain *a priori* knowledge regarding $\mathbf{x}$ with an image prior distribution $p_x(\mathbf{x})$. In this case, $p(\mathbf{r}|\mathbf{v})$ is obtained by marginalization over $\mathbf{x}$:

$$p(\mathbf{r}|\mathbf{v}) = \int p(\mathbf{r},\mathbf{x}|\mathbf{v})\mathrm{d}\mathbf{x} = \int p(\mathbf{r}|\mathbf{x},\mathbf{v})p_x(\mathbf{x})\mathrm{d}\mathbf{x}. \qquad (10)$$

Hence, we will refer to $p(\mathbf{r}|\mathbf{v})$ as the marginal likelihood. Given the marginal likelihood, Eq. (8) allows us to calculate Bayesian estimates for general velocity priors. The prior distribution $p_x(\mathbf{x})$ which describes the statistics of the image ensemble, can be chosen to have a naturalistic correlation structure. In our simulations in Section 3 we use a Gaussian image ensemble with power spectrum matched to observations in natural images [28,29].

In general, the calculation of the high-dimensional integral over $\mathbf{x}$ in Eq. (10) is a difficult task. However, when the integrand $p(\mathbf{r},\mathbf{x}|\mathbf{v})$ is sharply peaked around its maximum [which is the maximum *a posteriori* (MAP) estimate for $\mathbf{x}$—as the integrand is proportional to the posterior image distribution $p(\mathbf{x}|\mathbf{r},\mathbf{v})$ by Bayes' rule] the so-called "Laplace" approximation (also known as the "saddle-point" approximation) provides an accurate estimate for this integral [for applications of this approximation in the Bayesian setting, see e.g., [30]]. The Laplace approximation in the context of neural decoding is further discussed in, e.g., [31–35]. We briefly review this approximation here.

Following [29], we consider Gaussian image priors with zero mean and covariance $\mathcal{C}_x$ chosen to match the power spectrum of natural images [28]. Let us define the function

$$\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v}) \equiv \log p_x(\mathbf{x}) + \log p(\mathbf{r}|\mathbf{x},\mathbf{v}) + \tfrac{1}{2}\log(2\pi)^d|\mathcal{C}_x|, \qquad (11)$$

where d represents the number of pixels in our simulated image, and rewrite Eq. (10) as

$$p(\mathbf{r}|\mathbf{v}) = \frac{1}{\sqrt{(2\pi)^d|\mathcal{C}_x|}} \int e^{\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})}\mathrm{d}\mathbf{x}. \qquad (12)$$

Using Eq. (4) and $p_x(\mathbf{x}) = \mathcal{N}(0,\mathcal{C}_x)$, we obtain the expression

$$\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v}) = -\tfrac{1}{2}\mathbf{x}^{\mathrm{T}}\mathcal{C}_x^{-1}\mathbf{x} + \sum_i \left[\sum_\alpha \log \lambda_i(t_{i,\alpha};\mathbf{x},\mathbf{r}) \right. \\ \left. - \int \lambda_i(t;\mathbf{x},\mathbf{r})\mathrm{d}t \right], \qquad (13)$$

where λ_i are given by Eqs. (2), (6), and (7), and we made their dependence on $\mathbf{x}$ and $\mathbf{r}$ manifest. Since both terms in Eq. (13) are concave (see the closing remarks in Subsection 2.A), the log-posterior $\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})$ is concave in $\mathbf{x}$. To obtain the Laplace approximation, for fixed $\mathbf{r}$, we first find the value of $\mathbf{x}$ that maximizes $\mathcal{L}$ (i.e., the image MAP, $\mathbf{x}_{\mathrm{MAP}}$). When the integrand is sharply concentrated around its maximum, we can Taylor expand $\mathcal{L}$ around $\mathbf{x}_{\mathrm{MAP}}$ to the first nonvanishing order beyond the zeroth-order (i.e., its maximum value) and neglect the rest of the expansion. Since at the maximum the gradient of $\mathcal{L}$ and hence the first-order term vanish, we obtain

$$\mathcal{L}(\mathbf{x},\mathbf{r}) \approx \mathcal{L}(\mathbf{x}_{\mathrm{MAP}},\mathbf{r},\mathbf{v}) - \tfrac{1}{2}(\mathbf{x}-\mathbf{x}_{\mathrm{MAP}})^{\mathrm{T}}H(\mathbf{r},\mathbf{v})(\mathbf{x}-\mathbf{x}_{\mathrm{MAP}}), \qquad (14)$$

where the negative Hessian matrix

$$H(\mathbf{r},\mathbf{v}) \equiv -\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})|_{\mathbf{x}=\mathbf{x}_{\mathrm{MAP}}}, \qquad (15)$$

is positive semidefinite due to the maximum condition. Exponentiating this yields the Gaussian approximation (up to normalization)

$$e^{\mathcal{L}(\mathbf{x},\mathbf{r},\mathbf{v})} \propto p(\mathbf{x}|\mathbf{r},\mathbf{v}) \approx \mathcal{N}(\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}),\mathcal{C}_x(\mathbf{r},\mathbf{v})), \qquad (16)$$

where $\mathcal{N}(\mu,C)$ denotes a Gaussian density with mean μ and covariance C for the integrand of Eq. (12). [An important technical point here is that this Gaussian approximation is partially justified by the fact that the log-posterior (13) is a concave function of $\mathbf{x}$ [24,26,34] and therefore has a single global optimum, like the Gaussian (16).] Here, the posterior image covariance $\mathcal{C}_x(\mathbf{r},\mathbf{v})$ is given by the inverse of the negative Hessian matrix $H(\mathbf{r},\mathbf{v})$. (Note the dependence on both the observed responses $\mathbf{r}$ and the putative velocity $\mathbf{v}$.) The elementary Gaussian integration in Eq. (12) then yields

$$p(\mathbf{r}|\mathbf{v}) \approx \frac{e^{-\mathcal{L}(\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}),\mathbf{r},\mathbf{v})}}{\sqrt{|\mathcal{C}_x H(\mathbf{r},\mathbf{v})|}} \qquad (17)$$

for the marginal likelihood, or its logarithm

$$\log p(\mathbf{r}|\mathbf{v}) \approx -\mathcal{L}(\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}),\mathbf{r},\mathbf{v}) - \tfrac{1}{2}\log|\mathcal{C}_x H(\mathbf{r},\mathbf{v})|. \qquad (18)$$

The MAP itself is found from the condition $\nabla_x\mathcal{L}=0$, which in the case of exponential GLM nonlinearity $f(\cdot)=\exp(\cdot)$ yields the equation

$$\mathbf{x}_{\mathrm{MAP}}(\mathbf{n};\mathbf{r},\mathbf{v}) = \int \mathrm{d}^2\mathbf{n}'\mathcal{C}_x(\mathbf{n},\mathbf{n}')\sum_i \int \mathcal{K}_{i,\mathbf{v}}(t;\mathbf{n}')[r_i(t) \\ - \lambda_i(t;\mathbf{x}_{\mathrm{MAP}},\mathbf{r})]\mathrm{d}t. \qquad (19)$$

Note that this equation is nonlinear due to the appearance of $\mathbf{x}_{\mathrm{MAP}}$ inside the GLM nonlinearity on the right-hand side. As mentioned above, the objective function Eq. (11) is concave and can be efficiently optimized using

gradient-based optimization algorithms, such as the Newton–Raphson method. In particular, by exploiting the quasi-locality of the GLM likelihood we can implement the Newton–Raphson method such that, in cases where the image $\mathbf{x}(\mathbf{n})$ depends only on one component of $\mathbf{n}$, the MAP can be found in a computational time scaling only linearly with the spatial size of the image (see Appendix B for further elaboration on this point). Once $\mathbf{x}_{\mathrm{MAP}}$ is found, the Hessian at MAP and Eq. (17) can be calculated easily, and using Eq. (17), the approximate computation of $p(\mathbf{r}|\mathbf{v})$ is complete.

To recapitulate, in the case of an *a priori* uncertain image, given the observed spike trains $\mathbf{r}$, we numerically find $\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v})$ for a range of putative velocities $\mathbf{v}$ and using Eq. (17), we compute $p(\mathbf{r}|\mathbf{v})$, from which we may obtain $p(\mathbf{v}|\mathbf{r})$ via Eq. (8). We then take the value of velocity $\mathbf{v}_\star$ that maximizes $p(\mathbf{v}|\mathbf{r})$ as the estimate; i.e., we use the MAP estimate for the velocity.

As discussed in Section 1, our goal here was to critically examine the role of the detailed spiking structure of the GLM in constraining our estimates of the velocity. Since the spiking network model structure enters here only via the likelihood term $p(\mathbf{r}|\mathbf{v})$, we did not systematically examine the effect of strong *a priori* beliefs $p(\mathbf{v})$ on the resulting estimator (as discussed at further length, e.g., in [14]). Instead we used a simple uniform prior on velocity, which renders the MAP velocity estimate equivalent to the maximum (marginal) likelihood estimate, i.e., the value of $\mathbf{v}$ that maximizes $p(\mathbf{r}|\mathbf{v})$ given by the approximation Eq. (17) [or equivalently, its logarithm Eq. (18)]. Similarly, in the case of *a priori* known image $\mathbf{x}$ we chose the velocity $\mathbf{v}$ that maximizes the likelihood $p(\mathbf{r}|\mathbf{x},\mathbf{v})$.

2. Velocity Estimation Using the Energy Method

In order to assess the precision of our Bayesian estimates of velocity, we compared our estimates to those obtained using the correlation-based algorithm described in [4]. This algorithm closely resembles the spatiotemporal energy models for motion processing introduced by [36]. In order to understand the rationale behind this method, assume, hypothetically, that all the cells have exactly the same receptive fields up to the positioning of their centers and that they respond reliably and without noise to the stimulus. Then the RGCs' spike trains $\mathbf{r}_i$ in response to moving images would clearly be identical up to time translations. In other words, $r_i(t+n_i/v)$ would be equal for all i, where n_i is the center position of the ith cell's receptive field along the axis of motion, and v is the magnitude of $\mathbf{v}$. Thus even in the realistic, noisy situation, we expect the $\mathbf{r}_i$ for different i to have a large overlap if they are shifted in time as described, and in principle, we should be able to recover the true velocity by maximizing a smoothed version of this overlap. Inspired by this observation, an energy function is constructed as follows. First, the spike trains are convolved with a Gaussian filter $w(t) \propto \exp(-t^2/2\tau^2)$ (we chose τ to be 10 ms; see below and [4]). Let us define

$$\widetilde{\mathbf{r}}_i(t) = w \star \mathbf{r}_i = \sum_\alpha e^{-(t-t_{i,\alpha})^2/2\tau^2}. \tag{20}$$

Then, the "energy" function for the entire population of cells is determined by the sum of the overlaps of the

shifted and smoothed responses of all cells [37],

$$\mathcal{E}(\mathbf{v},\mathbf{r}) = \sum_{i,j} \int \widetilde{r}_i\left(t + \frac{n_i}{v}\right)\widetilde{r}_j\left(t + \frac{n_j}{v}\right)\mathrm{d}t$$

$$= \int \left[\sum_i \widetilde{r}_i\left(t + \frac{n_i}{v}\right)\right]^2 \mathrm{d}t. \tag{21}$$

In order to cancel the effect of spontaneous activity of the cells, in [4] a "net motion signal" $N(\mathbf{v},\mathbf{r})$ is obtained by subtracting energy of the left-shifted spike trains from that of the right-shifted responses:

$$N(\mathbf{v},\mathbf{r}) \equiv \mathcal{E}(\mathbf{v},\mathbf{r}) - \mathcal{E}(-\mathbf{v},\mathbf{r}). \tag{22}$$

Finally, $N(\mathbf{v},\mathbf{r})$ is calculated for $\mathbf{v}$ across a range of putative velocities, and the value that maximizes the net motion signal is taken as the velocity estimate. Figure 1 illustrates the basic idea of this method for ON cells, although it should be noted that OFF cells are also included in our analysis.

3. Connection between the Bayesian and Energy-Based Methods

An interesting connection can be drawn between Bayesian velocity decoding and the method of Subsection 2.B.2 based on the energy function Eq. (21). For simplicity,
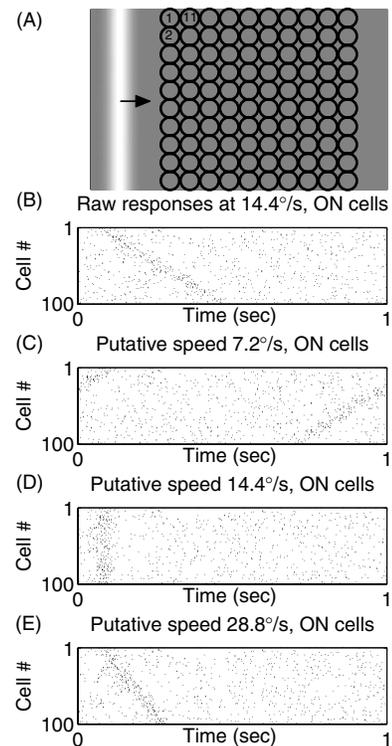

Fig. 1.   Ensemble motion signals. (A) Moving bar stimulus and cell layout. Cells in the left most column are numbered 1–10 from top to bottom, cells in the second column are numbered 11–20 from top to bottom, etc. (B) Raw responses from the ON cells for a moving bar with speed 14.4°/s. Each tick represents one spike and each row represents the response of a different cell. (C)–(E) Same spike trains circularly shifted by an amount equal to the time required for a stimulus with the indicated putative speed to move from an arbitrary reference location to the receptive field center. Responses from OFF cells were also included in this procedure.

imagine that spike trains are generated not by the GLM, but rather by a simpler linear-Gaussian (LG) model. In this case, it turns out that the marginal likelihood method is closely related to the energy function method described above. Specifically, we model the output spike trains as

$$\mathbf{r}_i = b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x} + \boldsymbol{\epsilon}_i, \tag{23}$$

where the noise term is Gaussian $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0,\Sigma)$. In the case that the noise terms for different cells are independent, we have $p_{LG}(\mathbf{r}|\mathbf{x},\mathbf{v}) = \Pi_i \mathcal{N}(b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}, \Sigma)$, though the generalization to correlated outputs is straightforward. We show in Appendix A that in a certain regime the logarithm of the LG marginal likelihood is given by [see Eqs. (A7) and (A8) and Eq. (A18)]

$$\log p_{LG}(\mathbf{r}|\mathbf{v}) = \frac{1}{2}\sum_{i,j} \int R_i\left(t + \frac{n_i}{v}\right) R_j\left(t + \frac{n_j}{v}\right) dt + A(\mathbf{v}), \tag{24}$$

where $A(\mathbf{v})$ has no dependence on the observed spike trains and only a weak dependence on $\mathbf{v}$. We find empirically that the term $A(\mathbf{v})$ in Eq. (24) grows with velocity, and therefore its inclusion shifts value of the maximum likelihood estimate toward higher velocities. Conversely, its absence in the energy function Eq. (21) causes the energy method estimate to have a negative bias. See Fig. 5 for an illustration of this effect. The resemblance of the remaining term to Eq. (21) above is clear. Here, $\mathbf{R}_i$ are smoothed versions of the spike trains $\mathbf{r}_i$ (with the baseline log firing rate subtracted out) and are given, as in Eq. (20), by

$$\mathbf{R}_i = w_{LG} * (\mathbf{r}_i - b_i), \tag{25}$$

where here the optimal smoothing filter w_{LG} is determined by the receptive fields k_i, the prior image correlation statistics, and the velocity [its explicit form is given in Eq. (A21) in Appendix A], as we discuss in more depth below.

Thus maximizing the marginal likelihood Eq. (24) is, to a good approximation, equivalent to maximizing the energy Eq. (21). The major difference between Eq. (21) and Eq. (24) is in the filter we apply to the spike trains: $\tilde{\mathbf{r}}_i$ has been replaced by $\mathbf{R}_i$. The key point is that $\mathbf{R}_i$ depends on the stimulus filters k_i, the velocity $\mathbf{v}$, and the image prior in an *optimal* manner, unlike the smoothing in Eq. (20). (Note that, while changes in optimal filters at differing light levels have been discussed in terms of motion estimation in fly vision [38], no account of varying light levels was taken here.) The dependence of this optimal filter as a function of v can be explained fairly intuitively, as we discuss at more length in Appendix A following Eq. (A21). We find that τ_w, the time scale of the smoothing filter w_{LG}, is dictated by three major time scales, some of which depend on the velocity v: τ_k, the width of the time window in which each RGC integrates its input; l_k/v, where l_k is the spatial width of the receptive field; and l_{corr}/v where l_{corr} is the correlation length of natural images. At low velocities, l_k/v and l_{corr}/v are large, and the smoothing time scale τ_w is also large, since in this case we gain more information about the underlying firing rates by averaging over a longer time window. At high velocities, on the other

hand, τ_k dominates l_k/v and l_{corr}/v, and $\tau_w \sim \tau_k$. This setting of τ_w makes sense because although the image movie I can vary quite quickly here, the filtered input $J_i(t)$ induces a firing rate correlation time of order τ_k, and examining the responses at a temporal resolution finer than τ_k only decreases the effective signal-to-noise.

Figure 2 illustrates these effects by plotting the optimal smoothing filters w_{LG} for several different values of the velocity v. Interestingly, in the high-velocity limit, the analytically derived optimal temporal filter width τ_w is of the order of 10 ms, which was the value chosen empirically for the optimal Gaussian filter used in [4]. We recomputed the optimal empirical filter for our simulated data here by plotting the standard deviation of the velocity estimates obtained using the net motion signal [defined in Eq. (22)] against the filter width (Fig. 3). For this velocity (28.8°/s) the optimal filter is of the order of 10 ms; thus, we used a filter of width 10 ms when comparing the energy method to the Bayesian decoder.

To summarize, maximizing the likelihood marginalized over the unknown image is very closely related to maximizing the energy function introduced by [4], if we replace the GLM with the simpler linear Gaussian model. Since the actual spike train generation is much better modeled by the GLM than by the Gaussian model, we expect Bayesian velocity estimation (even with uncertain prior knowledge of the image) based on the correct GLM to be more accurate. This expectation is borne out by our simulations, though it is worth noting that the improvement is significantly smaller than when the Bayesian decoder has access to the exact image.

C. Simulations
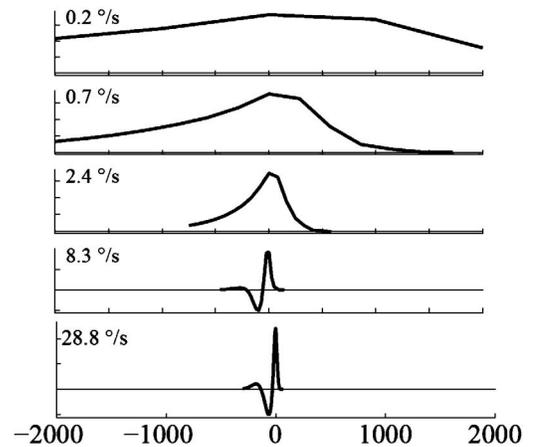We simulated the presentation of a bar moving across the gray background of a CRT monitor refreshing at 120 Hz.

Fig. 2.   Optimal linear spike train filter w_{LG} for a range of velocities from 0.2°/s to 28.8°/s (bottom) in exponential steps. The y axes are scaled in dimensionless units for clarity here. As discussed in Subsection 2.B.3, there are three time scales that determine the time scale of our filter w_{LG}. At low velocities, shown in the upper panels, the width of $w(t)$ is determined by the two scales x_k/v and x_{corr}/v and is thus quite large (since the denominator v is small). At the higher velocities shown in the lower panels, the optimal filter width is dominated by the time scale of the receptive field τ_k, and is of the order of τ_k, which is $\sim 10-20$ ms. For even higher velocities the shape of this filter remains essentially the same as in the bottom panel.
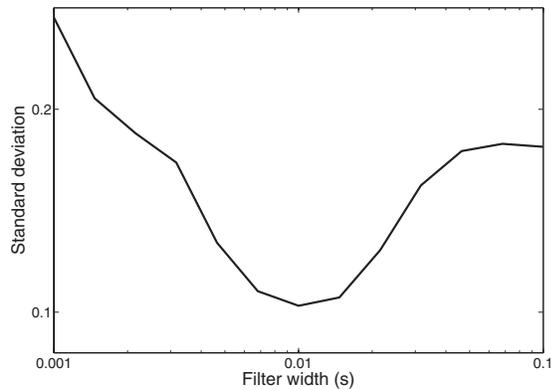
Fig. 3. Effect of filter width τ_w on the standard deviation of velocity estimates [obtained using the signal defined in Eq. (22)] across 100 presentations of a black bar moving at a speed of 28.8°/s across a gray background. Note that a filter width of $\tau_w \approx 10$ ms is optimal, in agreement with the findings of [4] and with the width of the optimal filter shown in Fig. 2.

The spatial profile of the bar in the direction of motion was a Gaussian function with a standard deviation (SD) of 96 μm. The visual field was represented by a grid of 100×100 pixels covering the receptive fields of two layers of cells each arranged in a uniform 10×10 grid. One layer consisted of ON cells, while the other represented OFF cells. The pixel resolution used was 10 times that used in [7] resulting in a pixel size of 12 μm. The bar moved across the visual field in discrete steps of **v** pixels/refresh, although **v** was not restricted to integer values. On each trial, the bar traversed the entire visual field once at a constant velocity. (Therefore, low-velocity trials lasted longer than high-velocity trials; this will affect some of our analyses below.) Stimulus dimensions and speeds were converted to degrees/second using the approximation 200 μm/° [39] with a pixel size of 12×12 μm. This meant that, with a refresh rate of 120 Hz, a speed of 1 pixel/refresh corresponded to a speed of 7.2°/s.

Then, to investigate the fidelity with which speed was encoded by our model, we ran simulations using a variety of stimulus parameter settings. Specifically, we conducted 100 trials at each of 48 stimulus conditions. These 48 conditions were made up of eight speeds (10.8, 14.4, 21.6, 28.8, 36.0, 43.2, 50.4 and 57.6°/s) by six luminance levels (0, 0.125, 0.25, 0.75, 0.875 and 1 on a grayscale level where 0 is black and 1 is white, and the background level was set at 0.5). We also refer to the six different luminance levels in terms of the contrast of the bar with respect to the background. More precisely we define the contrast as $(I_\text{bar} - I_\text{background})/I_\text{background}$, where I_bar and $I_\text{background}$ denote the bar and the background luminance, respectively. These six luminance levels thus become contrast levels −1, −0.75, −0.5, 0.5, 0.75, and 1.

For each of these trials, we obtained a set of spike trains **r**. From these spike trains, it was possible to estimate the speed of the stimulus used as being one of a number of putative speeds. The putative speeds tested in our simulations ranged from 7.2 to 108°/s in steps of 0.36°/s. Thus, we could compare speed estimates across stimulus conditions by examining the SD of estimates across the 100 trials performed for each condition. As in [4], we focused on the fractional SD (SD of velocity esti-

mate divided by the true stimulus speed) of estimates to assess the fidelity of retinal speed signals, as any systematic bias in speed estimate can in principle be compensated for by downstream processing. However, we will also present the dependence of the estimate bias on stimulus condition. As will be seen, the fractional bias and the fractional SD are roughly of the same order and thus both contribute to the total root-mean-square fractional error of the velocity estimate. The latter is given by the square root of the sum of the squared fractional bias and squared fractional SD. It should be noted that other contrast levels between −0.5 and 0.5 were also tested but are not presented, as for some combinations of decoder and speed, the velocity estimation performance at these low contrasts was not significantly above chance.

As outlined above, we used three different decoding methods to estimate the stimulus velocity from the simulated spike train ensembles; we compared Bayesian velocity decoding with (optimal decoder) and without (marginal decoder) complete prior information about the image with velocity estimation using the energy method. In particular, in Subsection 3.A.4 we discuss the effect of prior image uncertainty on the performance of the Bayesian decoder in more detail. In order to parametrically vary the prior information available to the decoder, in the simulations used in that section, the image was flashed a number of times to the cells while it was held fixed, and the image prior $p(I)$ was updated according to the observed spike train data elicited by the flashes (no preview flashes were used in the simulations discussed in other sections). See Fig. 6(b) below for an illustration of this procedure. Short flashes were used instead of a continuous uninterrupted presentation, because in the latter case, the cells rapidly filter out the fixed image contrast, and thus after a brief interval ($\sim 20–30$ ms), the spike trains cease to carry extra information about the image. The more times the image is flashed, the smaller the decoder's uncertainty C_x when the image starts moving. This allows the decoder to better estimate the velocity when it finally sees the same image in motion.

3. RESULTS

A. Comparison of the Different Velocity Decoders
In this section we compare the performance of the energy model with the optimal and marginal decoders, as described in Section 2. Figure 4(a) plots the velocity posterior $p(\mathbf{v}|\mathbf{r},\mathbf{x})$ for the case of an *a priori* known image (the moving bar described above) given a specific observed population spike train **r** in response to the moving bar stimulus as a function of putative stimulus speed **v**. Here, the true stimulus speed was 36.0°/s. Figure 4(b) shows the log of the marginal likelihood in the case where the image is not completely known, and Fig. 4(c) shows the value of the net motion signal N again as a function of putative speed and for the same stimulus. All three decoders successfully estimated the speed in the trial shown; however, it is clear from the figure that the net motion signal is much less sharply peaked around the stimulus speed than for the Bayesian decoders. The consequences of these findings are reflected in the lower panels of Fig. 4, which show that the distribution of speed estimates
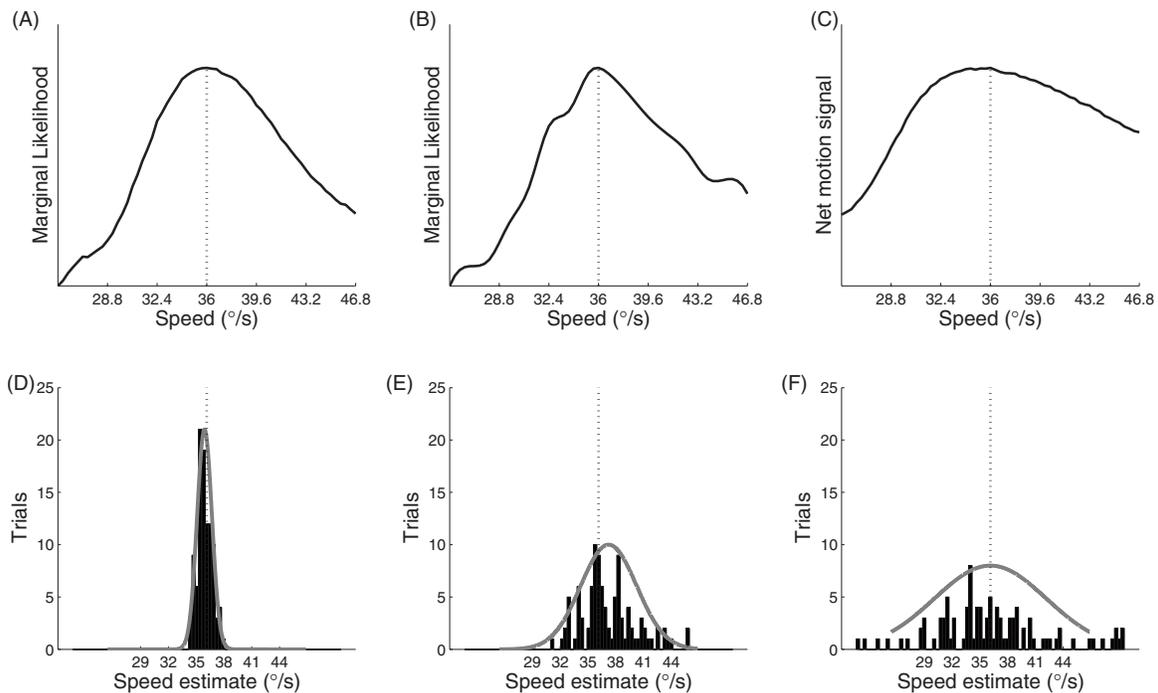
Fig. 4.   Optimal decoder leads to the most precise velocity estimates, while the marginal decoder outperforms the energy-based "net motion signal" method, in terms of precision. (A) Posterior for the optimal decoder. (B) Log marginal likelihood for the marginal decoder. (C) Net motion signal N as a function of putative stimulus speed $\mathbf{v}$ for spike trains generated using a stimulus with speed 36.0°/s and contrast −0.5 for a trial where all methods successfully estimate the stimulus speed. It can be seen that the nonmarginal likelihood (A) is more sharply peaked around the stimulus speed than the marginal likelihood (B) and the net motion signal (C). Distribution of speed estimates across 100 presentations of a bar moving at a speed of 36.0°/s using (D) the optimal posterior probability, (E) the marginal likelihood, and (F) the net motion signal. Also plotted are Gaussian fits to the distributions with mean±SD of 35.76±0.81 for the optimal decoder, 37.1±3.05 for the marginal decoder, and 36±6.09 for the net motion signal.

across 100 presentations of a bar of contrast −0.5 moving at a speed of 36.0°/s is most precise using the optimal decoder [Fig. 4(d)], rather than either the marginal decoder (E) or the energy method (F). Also plotted are Gaussian fits to the distributions with a mean ±SD of 35.76±0.81°/s for the optimal decoder, 37.1±3.05°/s for the marginal decoder and 36±6.09°/s for the net motion signal. The fractional SD averaged across all conditions simulated in this study was 1.6% of the stimulus speed for the optimal decoder, 6.4% for the marginal decoder, and 10% of the stimulus speed for the energy method. Since the estimators are not unbiased, their root-mean-square error is larger than their SD, as the error receives a contribution from the bias as well. The root-mean-square fractional errors averaged across all stimulus conditions were 2%, 6.9%, and 11%, for the optimal decoder, marginal decoder and the energy method, respectively. Because the velocity estimation based on the energy method does not make use of the image profile at any stage, these results are as expected with the performance of the marginal decoder being intermediate between that of the optimal decoder and the energy method.

1. Accuracy as a Function of Stimulus Speed
Because in our simulations the moving bar stimulus makes only one pass over the visual field, more time is spent traversing the field and more spike train information is obtained for slower moving stimuli. Figure 5(a) illustrates the fractional SD of 100 speed estimates for both of the Bayesian methods and the energy method, at each

of the eight stimulus speeds, averaged across the six contrast levels. As expected, performance declines with increasing speed for all three methods. The Bayesian decoders provide more precise estimates than the energy method at all speeds. Again, the advantage of the Bayesian decoder over the energy method is partly lost when its prior information about the image is uncertain.

2. Accuracy as a Function of Stimulus Contrast
Lowering the contrast of the moving bar causes a reduction in the number of stimulus-related spikes generated by the GLM model, according to Eqs. (2) and (3). As with increasing stimulus speed, this obviously results in a reduction in stimulus-related information with which to estimate the stimulus speed. (Note that the model of [7] lacks explicit luminance-or contrast-gain control effects; thus, these results should be interpreted in terms of local modifications around a fixed luminance pedestal that are sufficiently small to avoid engaging classical luminance gain-control mechanisms.) To examine this relationship, we averaged the SD of the 100 speed estimates at each of the six contrast levels across the eight stimulus speeds. The results are shown in Fig. 5(b) and illustrate the expected increase in performance with increasing stimulus contrast. Again, the Bayesian decoders clearly outperform the energy method at all levels.

3. Effect of Contrast and Speed on Mean Speed Estimate
While we were primarily concerned with the precision of speed estimates in the current study, a number of well-
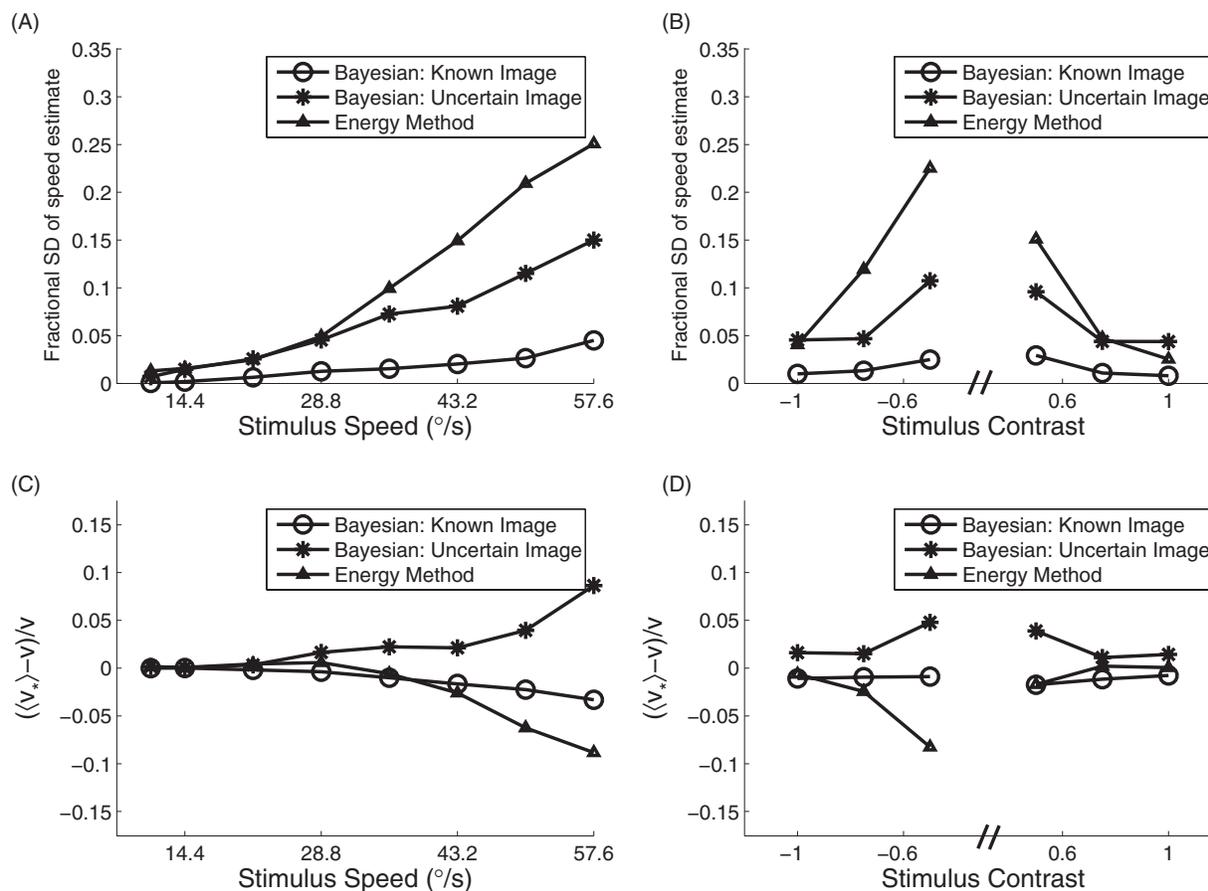
Fig. 5.   Fractional standard deviation of speed estimates versus (A) stimulus speed and (B) stimulus contrast for the Bayesian decoder with full image information (optimal decoder), the Bayesian decoder with incomplete image information (marginal decoder), and the energy method. (C), (D) plot the difference between the mean estimated speed $\langle \mathbf{v}_\star \rangle$ and the true stimulus speed $\mathbf{v}$ normalized by $\mathbf{v}$ against the true stimulus speed and stimulus contrast, respectively. Note that the Bayesian decoder provides more precise estimates than the energy method at all levels, with performance improving with prior image information. Furthermore, it should be noted that for contrasts lower than ±0.5, particularly at high speeds, the dearth of information in the spike train ensemble resulted in the estimate from an inordinate number of trials being either the highest or lowest input putative speed. Accordingly, performance for *s* in this range were not calculated and not plotted.

researched visual phenomena concerning the relationship between the mean visual speed perceived, i.e., the bias, and the properties of the visual stimulus prompted us to investigate this in our simulations. The first phenomenon of interest was that in which humans tend to choose the slowest motion that explains the incoming information [40], i.e., we have a bias toward slower speeds. As can be seen in Fig. 5(c), the energy method is biased toward lower velocity estimates at higher stimulus speeds. The optimal decoder shows a very slight tendency in this direction also. On the other hand, the marginal decoder has a positive bias toward higher velocities. The second phenomenon of interest was that in which stimuli with low contrast are typically perceived as moving slower than those with high contrast [41,42]. Figure 5(d) plots the fractional bias of the speed estimate, i.e., the difference between the true stimulus speed $\mathbf{v}$ and the mean estimated speed $\langle \mathbf{v}_\star \rangle$ normalized by $\mathbf{v}$, versus the stimulus contrast for both the Bayesian decoder and the energy method against the stimulus contrast, averaged across all speeds tested in our simulations. There appears to be a slight trend toward greater bias at low contrast, although it should be noted that this is due to a strong bias at low

negative contrast, while at low positive contrast, the bias is close to zero. The fact that the fractional SD of the speed estimate at this low negative contrast value is so large makes it difficult to say anything definitive about a relationship between stimulus contrast and speed estimate bias. It is important to remember that a uniform prior on velocity was used in this study when considering the lack of any clear effect.

4. Effect of Prior Image Information
Here we discuss the effect of preview flashes of the fixed image on the velocity decoding performance of the marginal decoder. As mentioned above, the more times the image is flashed or "shown" to the cells, the less will be the decoder's uncertainty about it and the better the velocity estimate made by the decoder when it finally sees the same image in motion. This effect is shown in Fig. 6 where panel (A) shows the decrease in the relative error of the velocity estimate as the number of flashes increases. For a large number of flashes the error asymptotically reaches the level for the fully known image
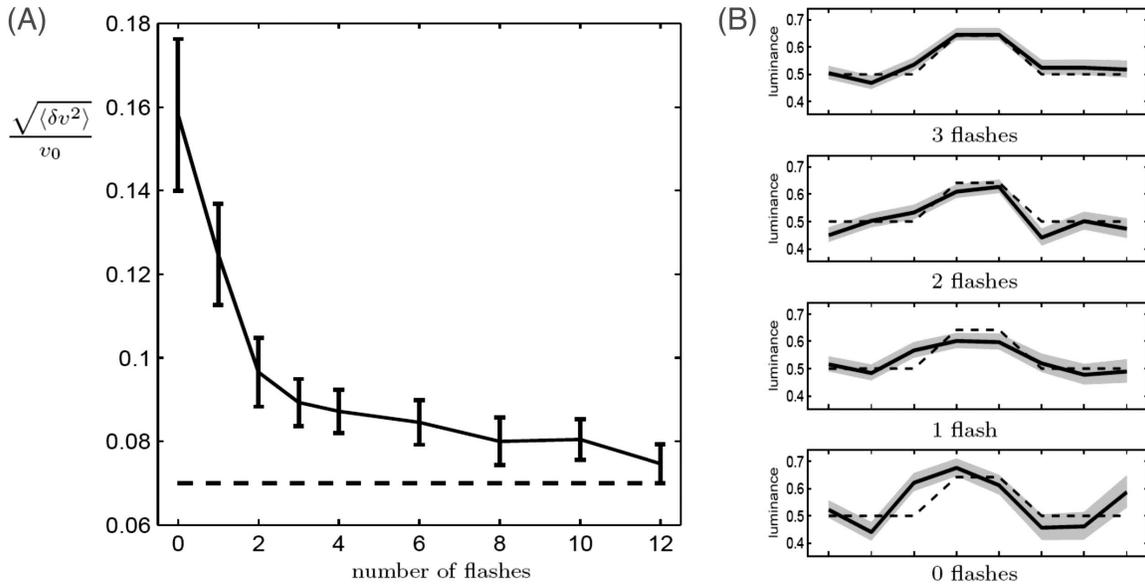
Fig. 6.   Effect of decreasing image uncertainty on accuracy of Bayesian velocity estimation. See Subsection 2.C for a detailed description of this simulation. (A) The solid curve with error bars shows the drop in the fractional rms error of the velocity estimate for an *a priori* unknown image as the number of preview flashes increases. The dashed curve is the fractional error for the case of an *a priori* known image. The true velocity was 28.8°/sec and the bar contrast 0.6. (B) The plots show the maximum *a posteriori* estimate of the image luminance profile (solid curve) in four trials with different numbers of preview flashes (indicated below each plot). The gray areas indicate the marginal uncertainty of the estimated luminance, and the dashed curve shows the actual image profile.

(shown by dashed curves). Panel (B) shows the convergence of the estimated luminance profile $\mathbf{x}_{MAP}$ to that of the actual bar image as the number of preview flashes increases. Figure 6 shows this effect for a particular stimulus velocity and contrast, but we note that the effect is qualitatively the same for all other values of these stimulus parameters; as the number of flashes increases beyond a few, the accuracy of the marginal decoder's estimate approaches that of the optimal decoder's.

As seen here and above, the efficiency of the GLM-based Bayesian decoder can be significantly deteriorated when the prior information about the image is too incomplete. As we showed in Subsection 2.B.3, Bayesian decoding with uncertain prior image information is, except for the replacement of the GLM with the LG model, closely related to the energy model. Indeed, in our simulations, the disparity between the performances of the energy model and those of the GLM-based Bayesian decoder was largely lost when the latter decoder's prior knowledge of the image became too uncertain.

B. Effects of Manipulating Model Parameters

1. Importance of Correlation Between Cells
In order to investigate the importance of correlated activity between cells, we wished to remove the interaction between neighboring spike trains without reducing the overall spiking rate. We used a straightforward trial-shuffling approach: we generated 200 individual spike trains, one for each cell, using 200 distinct presentations of the stimulus to the full model. We then constructed a single trial surrogate population spike train by serially assigning each independent spike train recorded on simulated trial i as the observed spike train in cell i. We repeated this 100 times to obtain spike ensembles repre-

senting 100 trials for each of the 48 conditions mentioned above (i.e., eight different speeds and six different contrast levels). This allowed us to determine the fractional standard deviation of the speed estimate for each of the 48 different stimulus conditions. It should be noted that this (somewhat involved) procedure was carried out in preference to simply removing the coupling between cells, as that would have resulted in a different average number of population wide spikes compared with the output from the full model, which would have had a confounding effect on the results.

The results are shown in Figs. 7(a)–7(c) for the optimal decoder, marginal decoder and the energy method, respectively, and are plotted versus the fractional standard deviation of the speed estimate for the same 48 conditions using the spike train ensembles obtained directly from the model. The diagonal lines indicate equality between the fractional SD of the speed estimates obtained using the shuffled responses and that obtained directly from the model. Somewhat surprisingly, given the significant correlations in this data (cf. Fig. 2 in [7]), this trial-shuffling procedure did not significantly hurt the performance of any of the three velocity estimators. For the marginal decoder, there is a noticeable reduction in performance for those stimulus conditions with more precise speed estimates, while for conditions with higher fractional SD, most points lie just below the line. However, for the other two decoders, if anything, there is a slight bias in Fig. 7(a) and 7(c), with data points tending to lie a bit below the identity line in both plots, indicating that the shuffling procedure happened to lead to velocity estimates with slightly reduced variability. These results are consistent with the conclusions of [2], that treating retinal ganglion cells as independent encoders leads only to a minor loss of information.
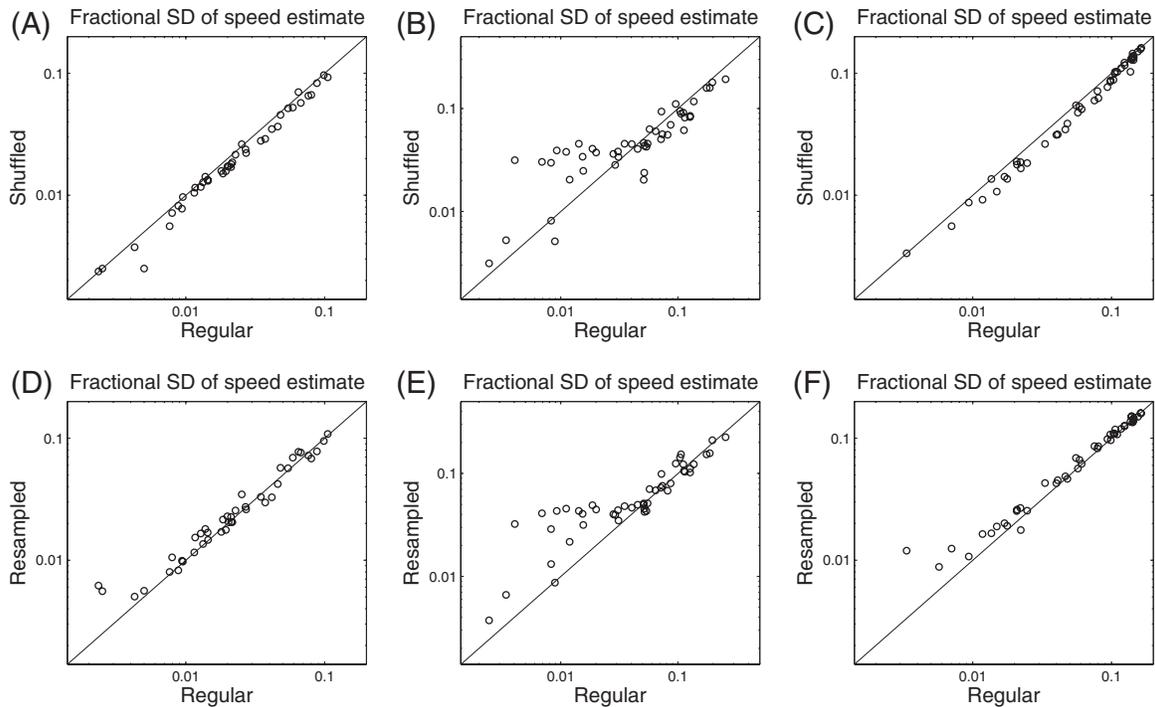
Fig. 7.   Effect of correlated activity and spike timing structure on speed estimates. Fractional SD of speed estimates using shuffled responses plotted as a function of that obtained using regular simulated data for (A) the Bayesian decoder with full image information, (B) the Bayesian decoder with incomplete image information, and (C) the energy method. Fractional SD of speed estimate using resampled spike trains plotted as a function of that obtained using regular simulated data for (D) the Bayesian decoder with full image information, (E) the Bayesian decoder with incomplete image information, and (F) the energy method. Diagonal lines indicate equality. Each circle represents a different one of the 48 speed-by-contrast stimulus conditions. Note that the performance of the decoders is relatively unaffected by these rather drastic manipulations of spike timing.

2. Timing Structure of Spike Trains

The question of whether cell spiking activity can be accurately modeled as a simple Poisson process with a time-varying rate or whether the intrinsic temporal structure of retinal spike trains plays an important role in communication has a long history in systems neuroscience. Simulations with the retinal ganglion cell model used in this study have demonstrated that preserving the spike history and cross-coupling effects can increase stimulus decoding performance by up to 20% [7]. We wished to examine the effect of removing the specific timing information of the individual spike trains. This was carried out using the method of [4]. Specifically, we generated a spike train for each cell for 100 trials of the moving bar stimulus. We then randomly selected spike times for each cell, with replacement, from that cell's spike distribution [its peri-stimulus time histogram (PSTH)], such that the number of spikes in each resampled spike train was equal to the average number of spikes in the corresponding original spike trains. This results in a spike train for each cell where spikes occur according to the marginal mean firing rate only, with no consideration given to spike history effects such as action potential refractoriness. Note that this process is even more disruptive of spike timing information than the shuffling procedure described in the previous subsection, since now we are destroying spike train structure both between and within cells. Again, this convoluted process was carried out in preference to simply removing the spike history filters h_{ij} from the model before generating the spike trains, as removal of those fil-

ters would have resulted in a greater number of total spikes and would thus have resulted in a misleadingly good speed estimation performance. This process of generating a spike train ensemble through resampling was carried out for each of the 48 stimulus conditions mentioned above.

The results are shown in Figs. 7(d)–7(f) for the optimal decoder, marginal decoder and the energy method, respectively, and are plotted versus the fractional standard deviation of the speed estimate for the same 48 conditions using spike train ensembles obtained directly from the model. Once again, the effects of this spike timing disruption on the performance of the velocity estimators was fairly minimal, with the resampled spike trains appearing to give a marginally worse performance as indicated by the preponderance of data points slightly above the identity line. Again for the marginal decoder the decreased performace is more pronounced for stimulus conditions with better performance.

3. Parameters of Cell Population

In the simulations above, two simple assumptions were made about the parameters of the cell population. First, the cells were arranged in an oversimplistic grid as in Fig. 8(a). And second, all ON cells were given a baseline log firing rate [b_i in Eq. (2)] of 2 spikes/s and all OFF cells a baseline log firing rate of 3 spikes/s, corresponding to the mean values obtained when fitting the model [7]. In order to examine a somewhat more biologically realistic case we jittered the center location of the cells as in Fig. 8(b) and
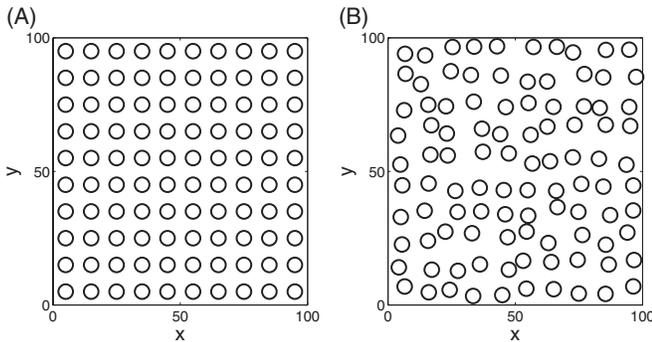
Fig. 8. (A) Simple rectangular grid cell arrangement. (B) Jittered cell arrangement.

randomly selected the baseline log firing rates of the ON and OFF cells from uniform distributions on intervals 1 to 3 spikes/s and 1.5 to 4.5 spikes/s, respectively.

Figure 9 illustrates the speed estimates over 100 trials for a stimulus with speed of 28.8°/s and contrast of −1 using the regular cell arrangement and uniform baseline log firing rates (left column) versus the jittered cell arrangement and random baseline log firing rates (middle column) for all three methods. The performances of the optimal decoder, marginal decoder and energy method are shown in the top, middle, and bottom row, respectively. No significant difference in performance between the regular and jittered arrangements is apparent for any method.

While randomly jittering the baseline log firing rates around the mean caused no obvious change in estimation accuracy, this does not allow us to comment on the pos-
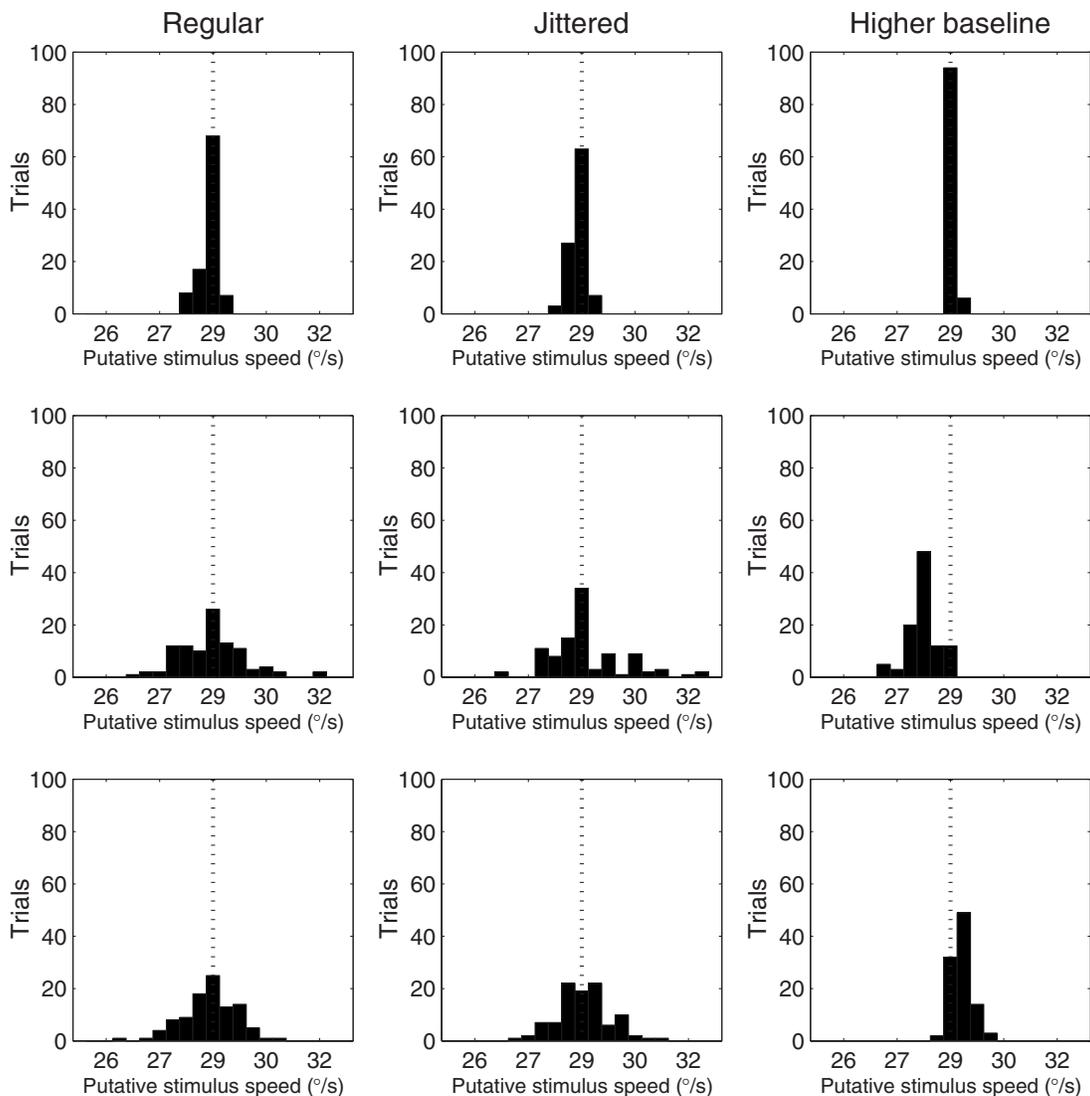

Fig. 9. Histograms illustrating the velocity estimates over 100 trials for a stimulus with velocity 28.8°/s and contrast of −1 using the regular cell arrangement and uniform baseline log firing rates (left column) and the jittered cell arrangement and random baseline log firing rates (middle column). The top row represents the performance of the optimal Bayesian decoder, the middle row represents that of the marginal decoder, and the bottom row presents that of the energy method. Similar performance was obtained with both the rectangular-grid and randomized spatial layouts for all three methods. The right column illustrates the improved estimation performance obtained for all three methods by doubling the baseline log firing rates from 2 and 3 spikes/s to 4 and 6 spikes/s for the ON and OFF cells respectively.

sible effects of changes in the *mean* baseline log firing rate. To assess this, we also carried out 100 simulations using a stimulus with speed of 28.8°/s and a contrast of −1, where the cells were arranged in the original simple grid and the ON and OFF cells were given baseline log firing rates of 4 and 6 spikes/s, respectively. The right column of Fig. 9 illustrates the significantly improved estimation performance obtained by inflating the baseline log firing rates compared to the fitted values used throughout the rest of this study for all three methods.

4. DISCUSSION

The model of [7] employed stochastic checkerboard stimuli in order to accurately capture both the stimulus dependence and detailed spatio temporal correlation structure of responses from a population of retinal ganglion cells. In this study, we have examined responses from this model to a more behaviorally relevant coherent velocity stimulus. Specifically, we have used these responses to assess how faithfully speed is encoded in a population of neurons using an optimal Bayesian decoder, with complete knowledge of the stimulus image. We have also shown how to compute the Bayesian velocity estimate in the case where we have only a limited amount of information about the stimulus image, and how the Bayesian estimate in this case is closely related to a biologically plausible motion-energy-based method [36,43].

A connection between Bayesian velocity estimation and the energy method of [36] has been noted before [12,15]. In that work, a Bayesian model of local motion information was described. It was shown that this model could be represented using a number of mathematical "building blocks" that qualitatively resembled direction-selective complex cells. Given that models of those cells have been based on the energy method of [36], a link was drawn between the two methods. Furthermore, previous work has sought to optimally estimate instantaneous motion from spike train ensembles in the fly [38,44]. However, to the best of our knowledge, in the case of non local estimation of rigid motion, the mathematical connection revealed here between the energy method and the Bayesian method based on the marginal image likelihood in the LG case has not been previously described.

In terms of biological plausibility, it is unlikely that the brain performs optimal Bayesian inference with full knowledge of the image in order to estimate velocity. This is supported by a recent study, which [8] employed the energy method (Subsection 2.B.2) to examine the efficiency of the code from a population of primate RGCs. They did this by comparing the estimate of the velocity of a stimulus using the spiking activity in the cell population with psychophysical estimates made by human observers. While the energy model consistently outperformed the human observers, it was shown that at very brief presentation times, i.e., <100 ms the difference in estimation performance between the energy method and the human behavior was much smaller than at longer presentation times. This suggests that readout of the retinal population code can be extremely efficient when exposure to the moving stimulus is very brief, but less efficient over long trials when storing information over a long time is re-

quired by the optimal Bayesian decoder. In this study, having used longer presentation times (125–675 ms), and given that the optimal Bayesian decoder significantly outperforms even the energy method, it seems clear that human observers do not decode using a known image in this task. Instead, given the relationship presented here between the marginal decoder and the energy method, it appears that a strategy equivalent to marginalization over the uncertain image seems to be more consistent with the available data.

A couple of factors in the relationship between the marginal decoder and the energy method are worthy of further discussion. First, the optimal filter width for velocity estimation from cell population responses when using the energy method on real data was reported to be of the order of 10 ms [4]. This implies that the elementary motion signal was conveyed with a time resolution comparable to the interspike interval of RGCs. A similar filter width was empirically shown using our simulated data (Fig. 3), which was not unexpected given the large amount of variance captured by the model in peristimulus time histograms in response to novel stimuli [7]. Of more interest, however, the optimal filter derived analytically for our (LG) marginal decoder is also shown to be of similar width, at least in the case where stimulus velocities are above about 5°/s (Fig. 2). This lends further weight to the biological plausibility of the marginal decoder. The notion that optimal filters based on stimulus filters, natural image prior, and velocity could have a biological instantiation seems reasonable. Second, in this study, we have assessed the performance of the energy method relative to the marginal decoder where the spike trains were generated not by a simple LG model, but by a GLM model. Because of the resulting improvement in spike train modeling we saw a significant improvement in velocity estimation for the marginal decoder relative to the energy method. Obviously the optimal decoder outperformed both other methods given the extra image information with which it was furnished.

In terms of performance specifics, the optimal Bayesian decoder achieved an average relative precision of 2% across all 48 stimulus conditions, with the marginal decoder achieving 6.4% and the energy method realizing only 10% relative precision. It is interesting to compare the estimation performance using our model to that obtained using similar stimuli with real cells in [4]. The authors of that study reported that the ensemble activity of around 100 RGCs signaled speed with a precision of the order of 1%. The precision of 10% obtained using the same decoder on our model output spike trains is higher than that result. One likely reason for this is that our stimulus range included much lower contrast stimuli. If we restrict our precision estimate to those conditions that most closely resemble those used by [4], i.e., speeds of 10.8, 14.4, 28.8, and 57.6°/s and contrast levels of −1 and 1, we obtain a value of 2.8% using the energy method which is of the same order as their result.

We examined the precision of our speed estimates as a function of both stimulus speed and stimulus contrast. As expected, decoding performance improves with increasing contrast and with decreasing speed (Fig. 5). Figure 5(a) illustrates that our model approximately followed a

Weber–Fechner law with visual speed discrimination being roughly proportional to speed [45]. As discussed in Subsection 3.A.1, the faster the moving bar traverses the retina, the less time spent stimulating the cells, and the smaller the total number of spikes we have with which to decode the stimulus speed. Similarly, the precision of the speed estimate improves with increasing absolute contrast, which increases the effective signal-to-noise of the retinal output [see Fig. 5(b)]. The nonlinear function $f(\cdot)$ used in Eq. (2) for this study was chosen to be $\exp(\cdot)$. Given that in determining the firing intensities $\lambda_i(t)$, this function operates on the stimulus input (as well as the baseline firing rates and spike history and cross-coupling effects), any increase in stimulus contrast would be expected to have a strong impact on the stimulus-related firing rates; similar conclusions may be drawn from an analysis of the Fisher information in this model [26].

As mentioned earlier, Bayesian modeling has been employed in a number of studies investigating how visual speed perception is affected by properties of the visual stimulus. In [16] an optimal Bayesian observer model was used to examine human psychophysical data in terms of stimulus noise characteristics and prior expectations. They reported that the perception that low-contrast stimuli move more slowly than high-contrast stimuli was well modeled by an ideal Bayesian observer. This was because the broader likelihood (based on psychophysical measurements), when multiplied by a prior favoring low speeds [46], resulted in a larger shift toward zero than multiplication by a narrower likelihood. In the present study, a uniform prior was used for the speed of the moving bar. Thus, we would not expect a widening of the likelihood distribution by lowering the stimulus contrast to shift the location of the posterior probability distribution. As such, we would not expect any relationship between stimulus contrast and the mean (or median) of the speed estimate distribution. This appeared to be the case, with no straightforward relationship seen to exist between speed estimate bias and contrast [Fig. 5(d)]. There did appear to be a very slight trend toward greater bias to low speeds at low contrasts for the energy method, but given the much higher variance in the speed estimate at this contrast [Fig. 5(b)], we are disinclined to draw any deeper conclusions from these results.

In terms of a relationship between speed estimate bias and stimulus speed, however, our results indicate a clear trend. Specifically, there appears to be a systematic bias in speed estimation tending to underestimate speed at high stimulus velocities for both the energy method and the Bayesian decoder with known image, while tending to overestimate speed at the same high stimulus velocities for the Bayesian decoder with uncertain image [Fig. 5(c)]. This can be explained by the well-known fact that likelihood-based estimators can display bias in low-information settings (as the high-speed setting is here, since effectively less time is available to observe spiking data during the stimulus presentation). In the low-speed, high-information setting, the bias of the likelihood-based estimator is negligible, as expected. The discrepancy between the biases of the marginal decoder and the energy-based estimate is clarified by the connection between these two methods as described in Subsection 2.B.3 and

Appendix A. Specifically, see the discussion after Eqs. (24) and (25) of Subsection 2.B.3 and Eqs. (A7) and (A8) of Appendix A.

In [7] it was found that, when comparing the full RGC model with an uncoupled version (retaining spike history effects), Bayesian stimulus decoding recovered 20% more information (about the spatiotemporal light intensity profile) using pseudorandom stimuli. The authors also noted that additionally ignoring spike history effects further reduced the recovered information by 6%. Thus, we wished to examine the importance of correlations between cells and of the intrinsic timing structure of the spike trains for speed estimation precision. We followed the procedure employed in [4] and, as in that study, it appeared that, for most stimulus conditions, the shuffled, uncorrelated spike trains surprisingly resulted in a weak improvement in estimation precision. The one exception to this was in the case of the marginal decoder, where for high-information stimulus conditions (i.e., low speeds and high contrast), the shuffled spike trains resulted in a significant reduction in performance. We also replicated their test of how precise spike timing might affect speed estimation precision [4]. Again, as in their study, we found only a very slight reduction in performance for the optimal Bayesian decoder and the energy method. Similar to the reshuffled case, the deterioration in the performance of the marginal decoder was more pronounced. Given that we have completely abolished the intraneuronal and interneuronal nonstimulus-driven correlation structure here, these small decreases in performance indicate that velocity decoding does not depend strongly on the fine spike train structure—at least in the very simple case of a moving bar. It should be noted that for the results plotted in Fig. 7, all spike train ensembles were decoded using the full model. That is, coupling filters and spike history effects were assumed and accounted for when calculating λ_i in the decoding step. Given that coupling effects were removed by our shuffling procedure and that both coupling effects and spike history effects were removed by our resampling procedure, it is possible that decoding the spike trains with an appropriately reduced model might provide more accurate speed estimation for these manipulated spike train ensembles. To that end, we used a model without coupling filters to decode the speed of the shuffled spike train ensembles and a model with all h_{ij} set to zero to decode the speed of the resampled spike train ensembles. It is interesting to note that incorporating this knowledge about the presence or absence of cell coupling and spike history effects into the decoding made no qualitative difference to the accuracy of the estimated velocity (not shown).

5. CONCLUSION

Optimal Bayesian decoding with full image information has been shown to outperform a "motion energy" method that uses no prior image information. A method for performing Bayesian decoding without full image information has been described and has demonstrated performance intermediate between that of the optimal decoder and the energy method. All of these methods appear to outperform human psychophysical performance [8], par-

ticularly in experiments in which the motion stimulus was visible for an extended period of time. A mathematical description of the connection between the Bayesian decoder with less than full image information and the energy method indicates that, in addition to the extra information about the image used by the Bayesian estimator, information about the network's spatiotemporal stimulus filtering properties also plays an important role in optimal velocity estimation. The results of a number of simulations indicate a good correspondence between the speed encoding performance of the model and that of a population of real RGCs. This work thus provides a rigorous framework with which to explore the factors limiting the estimation of velocity in vision. Future work will seek to utilize these methods to investigate motion decoding using more complex stimuli moving in nontranslational ways, perhaps incorporating real-world issues such as occlusions and accelerations. Also, we aim to employ the methods to investigate real data.

APPENDIX A: MARGINAL LIKELIHOOD IN THE LINEAR GAUSSIAN MODEL

In this appendix we show that the logarithm of the marginal likelihood $p(\mathbf{r}|\mathbf{v})$ for a simple LG model of the RGCs is closely related to the energy function of [4], and thus for this model the Bayesian velocity decoding is nearly equivalent to the energy method. In the linear Gaussian model, the response of cell i, $\mathbf{r}_i$, is given linearly in terms of the image intensity profile $\mathbf{x}$ up to additive Gaussian noise with covariance Σ, as in Eq. (23). Thus we have

$$p_{\mathrm{LG}}(\mathbf{r}|\mathbf{x},\mathbf{v}) = \prod_i \mathcal{N}(b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}, \Sigma).$$

Using this and $p_x(\mathbf{x}) = \mathcal{N}(0, \mathcal{C}_x)$ as the Gaussian image prior, we repeat the steps in Eqs. (11)–(19) of Subsection 2.B.1. For the LG model, the log-posterior function is given by

$$\mathcal{L}_{\mathrm{LG}}(\mathbf{x},\mathbf{r},\mathbf{v}) \equiv \log[p_x(\mathbf{x})] + \log[p_{\mathrm{LG}}(\mathbf{r}|\mathbf{x},\mathbf{v})] = -\tfrac{1}{2}\mathbf{x}^{\mathrm{T}}\mathcal{C}_x^{-1}\mathbf{x}$$
$$-\tfrac{1}{2}\sum_i (\mathbf{r}_i - b_i - \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x})^{\mathrm{T}}\Sigma^{-1}(\mathbf{r}_i - b_i - \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x})$$
$$+ \text{const.} \tag{A1}$$

instead of Eq. (11), and the marginal distribution $p_{\mathrm{LG}}(\mathbf{r}|\mathbf{v})$ by

$$p_{\mathrm{LG}}(\mathbf{r}|\mathbf{v}) = \int e^{\mathcal{L}_{\mathrm{LG}}(\mathbf{x},\mathbf{r},\mathbf{v})}\mathrm{d}\mathbf{x}, \tag{A2}$$

similar to Eq. (10). As before, setting $\nabla_x \mathcal{L}_{\mathrm{LG}} = 0$ yields the equation for $\mathbf{x}_{\mathrm{MAP}}$, which unlike Eq. (19) is linear, and can be easily solved to yield

$$\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v}) = H(\mathbf{v})^{-1}\sum_i \mathcal{K}_{i,\mathbf{v}}^{\mathrm{T}} \cdot \Sigma^{-1} \cdot (\mathbf{r}_i - b_i). \tag{A3}$$

Here, the negative Hessian is given by

$$H(\mathbf{v}) = -\nabla_x \nabla_x \mathcal{L}_{\mathrm{LG}} = \mathcal{C}_x^{-1} + \sum_i \mathcal{K}_{i,\mathbf{v}}^{\mathrm{T}} \cdot \Sigma^{-1} \cdot \mathcal{K}_{i,\mathbf{v}}, \tag{A4}$$

which is now independent of the observed spike trains $\mathbf{r}$. Using Eqs. (A3) and (A4), we can rearrange the terms in Eq. (A1) to complete the square for $\mathbf{x}$, and obtain

$$\mathcal{L}_{\mathrm{LG}}(\mathbf{x},\mathbf{r},\mathbf{v}) = -\tfrac{1}{2}(\mathbf{x} - \mathbf{x}_{\mathrm{MAP}})^{\mathrm{T}}H(\mathbf{v})(\mathbf{x} - \mathbf{x}_{\mathrm{MAP}})$$
$$-\tfrac{1}{2}\sum_i \delta\mathbf{r}_i^{\mathrm{T}}\Sigma^{-1}\delta\mathbf{r}_i + \frac{1}{2}\sum_{ij} \mathbf{X}_i^{\mathrm{T}}\mathcal{C}_x(\mathbf{v})\mathbf{X}_j + \text{const.}, \tag{A5}$$

where $\mathcal{C}_x(\mathbf{v}) = H^{-1}(\mathbf{v})$ is the posterior covariance over the fixed image, and we defined the mean-adjusted response $\delta\mathbf{r}_i \equiv \mathbf{r}_i - b_i$ and the prefiltered response

$$\mathbf{X}_i \equiv \mathcal{K}_{i,\mathbf{v}}^{\mathrm{T}}\Sigma^{-1}\delta\mathbf{r}_i. \tag{A6}$$

The marginalization in Eq. (A2) is thus a standard Gaussian integration, which yields

$$\log p_{\mathrm{LG}}(\mathbf{r}|\mathbf{v}) = \tfrac{1}{2}\sum_{ij} \mathbf{X}_i^{\mathrm{T}}\mathcal{C}_x(\mathbf{v})\mathbf{X}_j - \tfrac{1}{2}\log|\mathcal{C}_x H(\mathbf{v})| + \text{const.} \tag{A7}$$

(the constant term is independent of $\mathbf{v}$, and therefore irrelevant for estimating it). The decomposition into the two terms on the right-hand side of Eq. (A7) is similar to that in Eq. (18). In both equations the second term arose from a Gaussian integration over $\mathbf{x}$ [an approximation in the case of Eq. (18)], and the first was (up to a constant in $\mathbf{v}$) the value of the logarithm of the joint distribution of $\mathbf{x}$ and $\mathbf{r}$, given $\mathbf{v}$, at $\mathbf{x}_{\mathrm{MAP}}(\mathbf{r},\mathbf{v})$. Unlike Eq. (18), however, although the second term on the right-hand side of Eq. (A7) depends on $\mathbf{v}$, it is nevertheless independent of the observed response $\mathbf{r}$. The only term that modulates the velocity posterior depending on $\mathbf{r}$ (through the implicit dependence of $\mathbf{X}_i$) is the first, which we denote by $\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r})$. We will see that this term corresponds closely to the energy function introduced in [4]. More explicitly, we have

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) \equiv \frac{1}{2}\sum_{ij} \mathbf{X}_i^{\mathrm{T}}\mathcal{C}_x(\mathbf{v})\mathbf{X}_j$$
$$= \frac{1}{2}\sum_{ij} \int\int X_i(\mathbf{n}_1)\mathcal{C}_x(\mathbf{n}_1,\mathbf{n}_2;\mathbf{v})X_j(\mathbf{n}_2)\mathrm{d}^2\mathbf{n}_1\mathrm{d}^2\mathbf{n}_2. \tag{A8}$$

In the following we will rewrite Eq. (A8) in a form which is explicitly akin to Eq. (21). For simplicity, we assume that the noise covariance is white, i.e., $\Sigma = \sigma^2\mathbf{1}$. Physiologically, this implies that we are ignoring stimulus-conditional correlations and history dependences in the network (as, e.g., in the uncoupled model discussed in [7]). From Eq. (A6) and the definition of $\mathcal{K}_{i,\mathbf{v}}$, Eq. (7), we then obtain the explicit form

$$X_i(\mathbf{n}) = \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau k_i(t - \tau, \tau\mathbf{v} + \mathbf{n})\delta r_i(t). \tag{A9}$$

If we further assume that the spike train observation has not revealed much information about the identity of the fixed image (as happens, e.g., for low contrasts or short

presentation times), then the posterior distribution over $\mathbf{x}$ will not be very different from the prior $p_x(\mathbf{x})$. Therefore, we can use the approximation $\mathcal{C}_x(\mathbf{v}) \approx \mathcal{C}_x$. In the one-dimensional case, which we are studying in this paper, the image profile $\mathbf{x}(\mathbf{n})$, and hence the prior image covariance, depend only on the component of $\mathbf{n}$ parallel to the direction of motion $\hat{\mathbf{v}} = \mathbf{v}/|\mathbf{v}|$ and are constant in the perpendicular direction. Denoting the former component by $n(=\mathbf{n} \cdot \hat{\mathbf{v}})$ and the latter by $n_\perp(=\mathbf{n} - n\hat{\mathbf{v}})$, we can then perfom the integrals over $n_\perp$ in Eq. (A8) and rewrite it as

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) = \frac{1}{2}\sum_{ij} \int \int \widetilde{X}_i(n_1)\mathcal{C}_x(n_1,n_2)\widetilde{X}_j(n_2)\mathrm{d}n_1\mathrm{d}n_2,$$
(A10)

$$\widetilde{X}_i(n) \equiv \int X_i(\mathbf{n})\mathrm{d}n_\perp = \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau \widetilde{k}_i(t-\tau, \tau v + n)\delta r_i(t),$$
(A11)

where $v \equiv |\mathbf{v}|$, and we defined $\widetilde{k}_i(t,n) \equiv \int k_i(t,\mathbf{n})\mathrm{d}n_\perp$. For each cell i, we specify a fixed point $\mathbf{n}_i$ positioned at its receptive field center, so that $k_i(t,\mathbf{n}_i + \Delta\mathbf{n})$ vanishes when $|\Delta\mathbf{n}|$ gets considerably larger than the size of the receptive field surround ($\sim 1°$). Hence, if we define

$$q_i(t,n) \equiv \widetilde{k}_i(t,n+n_i) = \int k_i(t,\mathbf{n}+\mathbf{n}_i)\mathrm{d}n_\perp,$$
(A12)

(where $n_i \equiv \mathbf{n}_i \cdot \hat{\mathbf{v}}$), $q_i(t,n)$ vanishes when $|n| \gg 1°$; for all cells, q_i are localized (up to the above scale) around the origin, as opposed to around the position of their respective receptive field centers along $\mathbf{v}$. In order to make the comparison with the energy model of Subsection 2.B.2 clearer, we also switch to the time domain (recalling that space n and time t are linked here via the velocity v); we define $\widetilde{R}_i(t) \equiv \widetilde{X}_i(n_i - vt)$ [equivalently, $\widetilde{X}_i(n) = \widetilde{R}_i((-n+n_i)/v)$] and rewrite Eq. (A10) by changing the integration variables from $n_{1(2)}$ to $vt_{1(2)}$:

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) = \frac{1}{2v^2}\sum_{ij} \int \int \widetilde{R}_i\left(-t_1 + \frac{n_i}{v}\right)\mathcal{C}_x(vt_1,vt_2)$$
$$\times \widetilde{R}_j\left(-t_2 + \frac{n_j}{v}\right)\mathrm{d}t_1\mathrm{d}t_2.$$
(A13)

Using Eq. (A11) and the definiton (A12), we write $R_i(t_1)$ explicitly as

$$\widetilde{R}_i(t_1) \equiv \widetilde{X}_i(n_i - vt_1)$$
$$= \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau \widetilde{k}_i(t-\tau, v\tau - vt_1 + n_i)\delta r_i(t)$$
$$= \frac{1}{\sigma^2}\int \mathrm{d}t \int \mathrm{d}\tau q_i(t-\tau, v(\tau-t_1))\delta r_i(t).$$
(A14)

Exploiting the translation invariance of the prior image ensemble that dictates $\mathcal{C}_x(n_1,n_2) = \mathcal{C}_x(n_1-n_2)$, we define $\mathcal{B}_x$ to be the operator square root of $\mathcal{C}_x$, in the sense that

$$\mathcal{C}_x(n_1-n_2) = \int \mathcal{B}_x(n_1-n)\mathcal{B}_x(n_2-n)\mathrm{d}n.$$
(A15)

In general, given an explicit form of $\mathcal{C}_x(n_1-n_2)$, $\mathcal{B}_x$ can be computed in the Fourier domain by taking the square root of the power spectrum [29]. In particular, for $\mathcal{C}_x(n_1-n_2) = c^2 e^{-|n_1-n_2|/l_{\mathrm{corr}}}$, we have $\mathcal{B}_x(n) = c\sqrt{2/l_{\mathrm{corr}}}\theta(n)e^{-n/l_{\mathrm{corr}}}$, where c is the image contrast, l_{corr} is the correlation length of typical images in the naturalistic prior ensemble, and $\theta(t)$ is the Heaviside step function. In the simulations of Subsection 3.A.4 we used this particular form of $\mathcal{C}_x$, as it yields (for spatial frequencies f larger than the inverse of the correlation length l_{corr} but smaller than the inverse image pixel size) a power spectrum $\propto 1/f^2$, as observed in natural images. Substituting definition (A15) (after renaming the integration variable n to vt) in Eq. (A13), we rewrite the latter as

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) = \frac{1}{2v}\sum_{ij} \int \int \int \widetilde{R}_i\left(-t_1 + \frac{n_i}{v}\right)\mathcal{B}_x(v(t_1-t))$$
$$\times \mathcal{B}_x(v(t_2-t))\widetilde{R}_j\left(-t_2 + \frac{n_j}{v}\right)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t$$
$$= \frac{1}{2v}\sum_{ij} \int \int \int \widetilde{R}_i(t_1)\mathcal{B}_x\left(v\left(t + \frac{n_i}{v} - t_1\right)\right)$$
$$\times \mathcal{B}_x\left(v\left(t + \frac{n_j}{v} - t_2\right)\right)\widetilde{R}_j(t_2)\mathrm{d}t_1\mathrm{d}t_2\mathrm{d}t.$$
(A16)

We derived the last line by renaming the integration variables as $t_1 \to n_i/v - t_1$, $t_2 \to n_j/v - t_2$, and $t \to -t$. Finally, defining

$$R_i(t) \equiv \frac{1}{\sqrt{v}}\int \mathcal{B}_x(v(t-t_1))\widetilde{R}_i(t_1)\mathrm{d}t_1,$$
(A17)

we obtain

$$\mathcal{E}_{\mathrm{LG}}(\mathbf{v},\mathbf{r}) = \frac{1}{2}\sum_{ij} \int R_i\left(t + \frac{n_i}{v}\right)R_j\left(t + \frac{n_j}{v}\right)\mathrm{d}t.$$
(A18)

Equation (A18) is akin to the energy function used in [4], and together with Eq. (A7) yields Eq. (24) of Subsection 2.B.3. To find the explicit form of the smoothing filter in Eq. (25), we compare that equation, in the form

$$R_i(t) = \int w_{\mathrm{LG}}(t-t')\delta r_i(t')\mathrm{d}t',$$
(A19)

with definition (A17):

$$R_i(t) = \frac{1}{\sigma^2\sqrt{v}}\int \mathrm{d}t_1 \int \mathrm{d}t' \int \mathrm{d}\tau \mathcal{B}_x(v(t-t_1))$$
$$\times q_i(t'-\tau, v(\tau-t_1))\delta r_i(t'),$$
$$= \frac{1}{\sigma^2\sqrt{v}}\int \mathrm{d}t_1 \int \mathrm{d}t' \int \mathrm{d}\tau \mathcal{B}_x(v(t-t'-t_1))$$
$$\times q_i(-\tau, v(\tau-t_1))\delta r_i(t'),$$
(A20)

[where we used Eq. (A14) to write the first line, and

shifted τ and t_1 by t' to derive the second], and obtain

$$w_{\text{LG}}(t) = \frac{1}{\sigma^2 \sqrt{v}} \int dt_1 \int d\tau \mathcal{B}_x(v(t-t_1)) q_i(-\tau, v(\tau-t_1)).$$

(A21)

Thus, $R_i(t)$ is a version of the response function of the cell i offset by its baseline log firing rate b_i and smoothed out on the time scale dictated by the largest of the spatiotemporal scales of the receptive fields (via q_i) or the correlation length of typical images (via $\mathcal{B}_x$)—with spatial scales converted to time scales by dividing by v. To see this more precisely, let us define $\Delta\tau_1 \equiv \tau$, $\Delta\tau_2 \equiv t_1 - \tau$, and $\Delta\tau_3 \equiv t - t_1$, such that $t = \Delta\tau_1 + \Delta\tau_2 + \Delta\tau_3$. On the other hand, because of the finite support of the factors of its integrand, the double integral Eq. (A21) receives nonzero contributions only when $|\Delta\tau_1| \lesssim \tau_k$, $|\Delta\tau_2| \lesssim l_k/v$, and $|\Delta\tau_3| \lesssim l_{\text{corr}}/v$ [where τ_k and l_k are the typical temporal and spatial size of the receptive field filters $k_i(t,n)$, respectively, and l_{corr} is the correlation length of typical images in the naturalistic prior ensemble]. Thus if $|t| = |\tau_1 + \Delta\tau_2 + \Delta\tau_3|$ is much larger than the sum of the three scales τ_k, l_k/v, and l_{corr}/v, the filter $w(t)$ is bound to vanish. This leads to the discussion of Subsection 2.B.3 following Eq. (25).

APPENDIX B: $\mathcal{O}(D)$ DECODING

Here, we discuss how to implement the Newton–Raphson optimization algorithm such that it finds the maximum *a posteriori* estimate for the image $\mathbf{x}_{\text{MAP}}$ [satisfying Eq. (19)] in cases where the image depends only on one spatial dimension and in a computational time that scales only linearly with the spatial dimensionality of the image vector d. The Newton–Raphson algorithm for minimizing the function $\mathcal{L}(\mathbf{x})$ works as follows [we have in mind the objective function defined in Eq. (13), but for simplicity we drop $\mathbf{r}$ and $\mathbf{v}$ from its arguments in this appendix]. At each iteration of this algorithm, starting from the vector $\mathbf{x}$, we change this vector by an amount $\Delta\mathbf{x}$ which is found by solving the set of linear equations $H(\mathbf{x})\Delta\mathbf{x} = \nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x})$. Here, the right-hand side is the gradient of $\mathcal{L}(\mathbf{x})$, and $H(\mathbf{x})$ is its negative Hessian matrix [as in Eq. (15)], both evaluated at $\mathbf{x}$. In general, the solution of a set of d linear equations can be calculated in $\mathcal{O}(d^3)$ elementary operations [47]. This would make the decoding of images with even moderate angular extension forbidding. Fortunately, as we will now explain, the quasi-locality of the GLM model allows us to overcome this limitation. The negative Hessian of $\mathcal{L}(\mathbf{x})$, Eq. (13), is given by

$$H(\mathbf{x}) = \mathcal{C}_x^{-1} + \sum_{i,t} J_{i,t}(\mathbf{x}),$$

(B1)

where the matrices $J_{i,t}(\mathbf{x})$ have the elements

$$J_{i,t}(n_1, n_2; \mathbf{x}) = \mathcal{K}_{i,\mathbf{v}}(t;n_1)\mathcal{K}_{i,\mathbf{v}}(t;n_2)\lambda_i(t;\mathbf{x})dt,$$

(B2)

and $\mathcal{K}_{i,\mathbf{v}}(t;n)$ was defined in Eq. (7) in terms of the receptive field filter of the cell i, $k_i(\tau,n)$ [as we are considering the one-dimensional image case, $k_i(\tau,n)$ is understood to be the full receptive field integrated along the transverse spatial dimension]. Here, we turned the integral over t in Eq. (13) into a discrete sum in Eq. (B1), as is done in the

numerical implementation, and for simplicity we wrote Eq. (B2) for the case of exponential GLM nonlinearity. Generalization of this equation and the rest of the argument to general nonlinearities is straightforward. Because of the finite spatial size of the receptive field and the finite duration of the temporal filter, $k_i(\tau,n)$ is nonzero only when $\tau \in [0,T_k]$ and $n \in [n_{\min}^i, n_{\max}^i]$, where T_k and $\Delta n \equiv n_{\max}^i - n_{\min}^i$ are upper bounds on the cells' temporal integration windows and the size of the cells' receptive field surrounds, respectively. It follows then from Eq. (7) that $\mathcal{K}_{i,\mathbf{v}}(t;n)$ vanishes unless $n_{\min}^i - vt \le n \le n_{\max}^i - vt + vT_k$ (we assumed $v > 0$, but generalization to $v < 0$ is straightforward). Thus $J_{i,t}(n_1, n_2; \mathbf{x})$ vanishes if $|n_1 - n_2| > \Delta n + vT_k$, regardless of i and t. In other words, for all (i,t), $J_{i,t}(\mathbf{x})$ are banded matrices with a band width of $\Delta n + vT_k$, and so is their sum. If we further use a prior covariance $\mathcal{C}_x$ with a banded inverse, then the full Hessian Eq. (B1) will be banded [e.g., the naturalistic prior covariance introduced after Eq. (A15) can be defined as the inverse of a tridiagonal matrix in the numerical implementation].

Unlike in the general case, the solution of a set of d linear equations with a banded equation matrix of band width B can be found in a computational time $\propto B^2 d$—i.e., in our case, in a computational time scaling only linearly (as opposed to cubically) with the image size d. On the other hand, we have observed empirically that the number of necessary Newton–Raphson iterations is more or less constant and does not scale with d. Hence the overall optimization procedure for finding $\mathbf{x}_{\text{MAP}}$ can be performed in $\mathcal{O}(d)$ computational time. This allows us to decode the velocity of large moving images. Similar methods with $\mathcal{O}(d)$ computational cost have been used in inference and estimation problems involving state–space models [48,49], e.g., in applications to neural data analysis [50].

ACKNOWLEDGMENTS

REFERENCES

1. M. Meister, L. Lagnado, and D. Baylor, "Concerted signaling by retinal ganglion cells," Science **270**, 1207–1210 (1995).
2. S. Nirenberg, S. Carcieri, A. Jacobs, and P. Latham, "Retinal ganglion cells act largely as independent encoders," Nature **411**, 698–701 (2002).
3. E. Chichilnisky and R. Kalmar, "Functional asymmetries in ON and OFF ganglion cells of primate retina," J. Neurosci. **22**, 2737–2747 (2002).

4. E. Frechette, A. Sher, M. Grivich, D. Petrusca, A. Litke, and E. Chichilnisky, "Fidelity of the ensemble code for visual motion in the primate retina," J. Neurophysiol. **94**, 119–135 (2005).

5. E. Schneidman, M. Berry, R. Segev, and W. Bialek, "Weak pairwise correlations imply strongly correlated network states in a neural population," Nature **440**, 1007–1012 (2006).

6. J. Shlens, G. Field, J. Gauthier, M. Grivich, D. Petrusca, A. Sher, A. Litke, and E. Chichilnisky, "The structure of multi-neuron firing patterns in primate retina," J. Neurosci. **26**, 8254–8266 (2006).

7. J. Pillow, J. Shlens, L. Paninski, A. Sher, A. Litke, E. Chichilnisky, and E. Simoncelli, "Spatio-temporal correlations and visual signalling in a complete neuronal population," Nature **454**, 995–999 (2008).

8. E. S. Frechette, M. I. Grivich, R. S. Kalmar, A. M. Litke, D. Petrusca, A. Sher, and E. J. Chichilnisky, "Retinal motion signals and limits on speed discrimination," J. Vision **4**, 570 (2004).

9. A. Litke, N. Bezayiff, E. Chichilnisky, W. Cunningham, W. Dabrowski, A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, R. Kalmar, K. Mathieson, D. Petrusca, M. Rahman, and A. Sher, "What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity," IEEE Trans. Nucl. Sci. **51**, 1434–1440 (2004).

10. R. Segev, J. Goodhouse, J. Puchalla, and M. Berry, "Recording spikes from a large fraction of the ganglion cells in a retinal patch," Nat. Neurosci. **7**, 1154–1161 (2004).

11. D. Knill and W. Richards, eds., *Perception as Bayesian Inference* (Cambridge Univ. Press, 1996).

12. E. P. Simoncelli, "Distributed analysis and representation of visual motion," Ph.D. thesis (Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1993). Also available as MIT Media Laboratory Vision and Modeling Technical Report #209.

13. D. Ascher and N. Grzywacz, "A Bayesian model for the measurement of visual velocity," Vision Res. **40**, 3427–3434 (2000).

14. Y. Weiss, E. Simoncelli, and E. Adelson, "Motion illusions as optimal percepts," Nat. Neurosci. **5**, 598–604 (2002).

15. E. P. Simoncelli, "Local analysis of visual motion," in *The Visual Neurosciences*, L. M. Chalupa and J. S. Werner, eds. (MIT Press, 2003), Chap. 109, pp. 1616–1623.

16. A. Stocker and E. Simoncelli, "Noise characteristics and prior expectations in human visual speed perception," Nat. Neurosci. **9**, 578–585 (2006).

17. A. E. Welchman, J. M. Lam, and H. H. Bulthoff, "Bayesian motion estimation accounts for a surprising bias in 3D vision," Proc. Natl. Acad. Sci. U.S.A. **105**, 12087–12092 (2008).

18. F. Hurlimann, D. Kiper, and M. Carandini, "Testing the Bayesian model of perceived speed," Vision Res. **42**, 2253–2257 (2002).

19. P. Thompson, K. Brooks, and S. Hammett, "Speed can go up as well as down at low contrast: Implications for models of motion perception," Vision Res. **46**, 782–786 (2005).

20. A. Thiel, M. Greschner, C. Eurich, J. Ammermüller, and J. Kretzberg, "Contribution of individual retinal ganglion cell responses to velocity and acceleration encoding," J. Neurophysiol. **98**, 2285–2296 (2007).

21. J. Kretzberg, I. Winzenborg, and A. Thiel, "Bayesian analysis of the encoding of constant and changing stimulus velocities by retinal ganglion cells," presented at Frontiers in Neuroinformatics 2008, Stockholm, September 7–9, 2008.

22. D. Brillinger, "Maximum likelihood analysis of spike trains of interacting nerve cells," Biol. Cybern. **59**, 189–200 (1988).

23. P. McCullagh and J. Nelder, *Generalized Linear Models* (Chapman & Hall, 1989).

24. L. Paninski, "Maximum likelihood estimation of cascade point-process neural encoding models," Network Comput. Neural Syst. **15**, 243–262 (2004).

25. W. Truccolo, U. Eden, M. Fellows, J. Donoghue, and E. Brown, "A point process frame-work for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects," J. Neurophysiol. **93**, 1074–1089 (2005).

26. L. Paninski, J. Pillow, and J. Lewi, "Statistical models for neural encoding, decoding, and optimal stimulus design," in *Computational Neuroscience: Progress in Brain Research*, P. Cisek, T. Drew, and J. Kalaska, eds. (Elsevier, 2007).

27. D. Snyder and M. Miller, *Random Point Processes in Time and Space* (Springer-Verlag, 1991).

28. D. Field, "Relations between the statistics of natural images and the response profiles of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).

29. D. H. Brainard, D. R. Williams, and H. Hofer, "Trichromatic reconstruction from the interleaved cone mosaic: Bayesian model and the color appearance of small spots," J. Vision **8**, 1–23 (2008).

30. R. Kass and A. Raftery, "Bayes factors," J. Am. Stat. Assoc. **90**, 773–795 (1995).

31. E. Brown, L. Frank, D. Tang, M. Quirk, and M. Wilson, "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 (1998).

32. W. Bialek and A. Zee, "Coding and computation with neural spike trains," J. Stat. Phys. **59**, 103–115 (1990).

33. S. Koyama and S. Shinomoto, "Empirical Bayes interpretations of random point events," J. Phys. A **38**, 531–537 (2005).

34. J. Pillow, Y. Ahmadian, and L. Paninski, "Model-based decoding, information estimation, and change-point detection in multi-neuron spike trains," submitted to Neural Comput.

35. Y. Ahmadian, J. Pillow, and L. Paninski, "Efficient Markov chain Monte Carlo methods for decoding neural spike trains," submitted to Neural Comput.

36. E. Adelson and J. Bergen, "Spatiotemporal energy models for the perception of motion," J. Opt. Soc. Am. A **2**, 284–99 (1985).

37. E. Chichilnisky and R. Kalmar, "Temporal resolution of ensemble visual motion signals in primate retina," J. Neurosci. **23**, 6681–6689 (2003).

38. W. Bialek (Princeton University, bbrinker@princeton.edu) and R. de Ruyter van Steveninck (Indiana University, deruyter@indiana.edu) (personal communication, 2003).

39. V. Perry and A. Cowey, "The ganglion cell and cone distributions in the monkey's retina: implications for central magnification factors," Vision Res. **25**, 1795–1810 (1985).

40. S. Ullman, *The Interpretation of Visual Motion* (MIT Press, 1979).

41. P. Thompson, "Perceived rate of movement depends on contrast," Vision Res. **22**, 377–380 (1982).

42. L. Stone and P. Thompson, "Human speed perception is contrast dependent," Vision Res. **32**, 1535–1549 (1992).

43. D. C. Bradley and M. S. Goyal, "Velocity computation in the primate visual system," Nat. Rev. Neurosci. **9**, 686–695 (2008).

44. M. Potters and W. Bialek, "Statistical mechanics and visual signal processing," J. Phys. I France **4**, 1755–1775 (1994).

45. S. McKee, G. Silvermann, and K. Nakayama, "Precise velocity discrimination despite random variations in temporal frequency and contrast," Vision Res. **26**, 609–619 (1986).

46. M. Blakemore and R. Snowden, "The effect of contrast upon perceived speed: a general phenomenon?" Perception **28**, 33–48 (1999).

47. W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C* (Cambridge Univ. Press, 1992).

48. N. Shephard and M. Pitt, "Likelihood analysis of non-Gaussian measurement time series," Biometrika **84**, 653–667 (1997).

49. R. Davis and G. Rodriguez-Yam, "Estimation for state-space models: an approximate likelihood approach," Stat. Sin. **15**, 381–406 (2005).

50. L. Paninski, Y. Ahmadian, D. Ferreira, S. Koyama, K. Rahnama, M. Vidne, J. Vogelstein, and W. Wu, "A new look at state-space models for neural data," J. Comput. Neurosci. (to be published). Epub ahead of print, doi 10.1007/s10827-009-0179-x.