

Optimal decoding of stimulus velocity using a probabilistic model of ganglion cell populations in primate retina.

Edmund C. Lalor,^{1,*} Yashar Ahmadian,² and Liam Paninski²

¹*Department of Electronic and Electrical Engineering and Institute of Neuroscience,
Trinity College Dublin,*

College Green, Dublin 2, Ireland

²*Department of Statistics, Columbia University,
1255 Amsterdam Avenue, New York, N.Y. 10027, USA*

**Corresponding author: edlador@tcd.ie*

A major open problem in systems neuroscience is to understand the relationship between behavior and the detailed spiking properties of neural populations. In this work, 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, given complete information about the displayed image, the spike train ensemble signals speed with an average relative precision 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) 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 optimal Bayesian decoder in the case that the image is not known. Simulations using the motion energy method reveal that it results in an average relative precision of only 10% 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, 330.7310, 330.5310.

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 [10,13,14,23,24,29,30,32]. 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 [20,31]. The utility of this new method for understanding the encoding of behaviorally-relevant signals was exemplified by [14], who examined how reliably visual motion was encoded in the spiking activity of a population of macaque parasol cells. These authors used a simple velocity stimulus and attempted to estimate the stimulus velocity 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. [29] recently described a statistical model of a complete population of primate parasol retinal ganglion cells (RGCs). 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). [29] demonstrated that this model accurately captures the stimulus dependence and spatio-temporal 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 pseudo-random binary stimuli and Bayesian inference, [29] 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.

At the psychophysical level, Bayesian inference has been established as an effective framework for understanding visual perception [17]; some recent notable applications to understanding visual velocity processing include [3,33,35,42,43]. In particular, [42] 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 [15,39]).

Links between retinal coding and psychophysical behavior have also been recently examined using Bayesian methods; [37], 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 this 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 [29] 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 population RGC model of [29], using an optimal Bayesian decoder, with and without full prior knowledge of the image. We begin by describing the mathematical construction of this optimal decoder, and then compare the optimal estimates to those based on a “net motion signal” derived directly from the spike trains without any prior image information [14]. We derive a mathematical connection between these two decoders and investigate the decoders’ performance through a series of simulations.

2. Methods

2.A. Model

The generalized linear model (GLM) [8, 21] for the spiking responses of the sensory network used in this study was described in detail in [29]. 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 i -th cell in the observed population emits a spike train, which we represent by a response function

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

where each spike is represented by a delta function, and $t_{i,\alpha}$ is the time of the α -th spike of the i -th 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 [25, 26, 29, 40]

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

where $f(\cdot)$ is a positive, strictly increasing rectifying function (in this case, $f(\cdot) = \exp(\cdot)$). The b_i represents 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,

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

where $k_i(t, \mathbf{n})$ is the spatio-temporal receptive field of the cell i . Given Eq. (2), we can write down the point process log-likelihood in the standard way [34]

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

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

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

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

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

where we defined

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

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).

2.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 [14] which used no explicit prior knowledge of the image.

2.B.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

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

If the image \mathbf{x} (e.g. a narrow bar of nonzero contrast) 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

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

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

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

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 Sec. 3 we used a Gaussian image ensemble with power spectrum matched to observations in natural images [7, 12].

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., [16]). The Laplace approximation in the context of neural decoding is further discussed in, e.g., [2, 4, 9, 18, 28]. We briefly review this approximation here.

Following [7], we consider Gaussian image priors with zero mean and covariance, \mathcal{C}_x , chosen to match the power spectrum of natural images [12]. Let us define the function

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

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

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

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

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

where λ_i are given by Eqs. (2) and (6)–(7), and we made their dependence on \mathbf{x} and \mathbf{r} manifest. 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}_{MAP}). When the integrand is sharply concentrated around its maximum, we can Taylor expand \mathcal{L} , around \mathbf{x}_{MAP} , to the first non-vanishing 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

$$\mathcal{L}(\mathbf{x}, \mathbf{r}) \approx \mathcal{L}(\mathbf{x}_{\text{MAP}}, \mathbf{r}, \mathbf{v}) - \frac{1}{2}(\mathbf{x} - \mathbf{x}_{\text{MAP}})^{\text{T}} H(\mathbf{r}, \mathbf{v})(\mathbf{x} - \mathbf{x}_{\text{MAP}}), \quad (14)$$

where the negative Hessian matrix

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

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

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

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} [25,26,28], 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 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

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

for the marginal likelihood or its logarithm

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

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

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

Notice that this equation is nonlinear due to the appearance of \mathbf{x}_{MAP} inside the GLM nonlinearity on the right hand side. For the case of convex and log-concave GLM nonlinearity, $f(\cdot)$, (conditions that are true for our $f(\cdot) = \exp(\cdot)$) the objective function Eq. (11) becomes

concave and can be efficiently optimized using gradient-based optimization algorithms, such as the Newton-Raphson method. Once \mathbf{x}_{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 $\mathbf{x}_{\text{MAP}}(\mathbf{r}, \mathbf{v})$ 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}_* , that maximizes $p(\mathbf{v}|\mathbf{r})$ as the estimate; i.e., we use the MAP estimate for the velocity. As discussed in the Introduction, 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 only enters here 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 [42]). 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 choose the velocity, \mathbf{v} , which maximizes the likelihood $p(\mathbf{r}|\mathbf{x}, \mathbf{v})$.

2.B.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 [14]. This algorithm closely resembles the spatiotemporal energy models for motion processing introduced by [1]. 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 i -th 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 's 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 [14]). Let us define

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

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 [11]

$$\mathcal{E}(\mathbf{v}, \mathbf{r}) = \sum_{i,j} \int \tilde{r}_i \left(t + \frac{n_i}{v} \right) \tilde{r}_j \left(t + \frac{n_j}{v} \right) dt = \int \left[\sum_i \tilde{r}_i \left(t + \frac{n_i}{v} \right) \right]^2 dt. \quad (21)$$

In order to cancel the effect of spontaneous activity of the cells, in reference [14] a “net motion signal”, $N(\mathbf{v}, \mathbf{r})$, was obtained by subtracting energy of the left-shifted spike trains from that of the right-shifted responses: $N(\mathbf{v}, \mathbf{r}) \equiv \mathcal{E}(\mathbf{v}, \mathbf{r}) - \mathcal{E}(-\mathbf{v}, \mathbf{r})$. 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. Fig. 1 illustrates the basic idea of this method.

2.B.3. Connection between the Bayesian and energy-based methods

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

$$\mathbf{r}_i = b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x} + \boldsymbol{\epsilon}_i \quad (22)$$

where the noise term is Gaussian $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \Sigma)$. In the case that this noise terms for different cells are independent, we have $p_{\text{LG}}(\mathbf{r}|\mathbf{x}, \mathbf{v}) = \prod_i \mathcal{N}(b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}, \Sigma)$, though the generalization to correlated outputs is straightforward. We show in the appendix that in a certain regime the logarithm of the LG marginal likelihood is given by (see Eqs. (31)–(32) and Eq. (42))

$$\log p_{\text{LG}}(\mathbf{r}|\mathbf{v}) = \frac{1}{2} \sum_{i,j} \int R_i \left(t + \frac{n_i}{v} \right) R_j \left(t + \frac{n_j}{v} \right) dt + A(\mathbf{v}), \quad (23)$$

where $A(\mathbf{v})$ has no dependence on the observed spike trains, and only a weak dependence on \mathbf{v} ¹. The resemblance of the remaining term to equation (21) above is clear. Here, \mathbf{R}_i are smoothed versions of the spike trains \mathbf{r}_i (with the baseline firing rate subtracted out) and are given, as in Eq. (20), by

$$\mathbf{R}_i = w_{\text{LG}} * (\mathbf{r}_i - b_i), \quad (24)$$

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. (45) in the appendix), as we discuss in more depth below.

¹We find empirically that the term $A(\mathbf{v})$ in Eq. (23) grows with velocity, and therefore its inclusion shifts value of the maximum likelihood estimate towards 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.

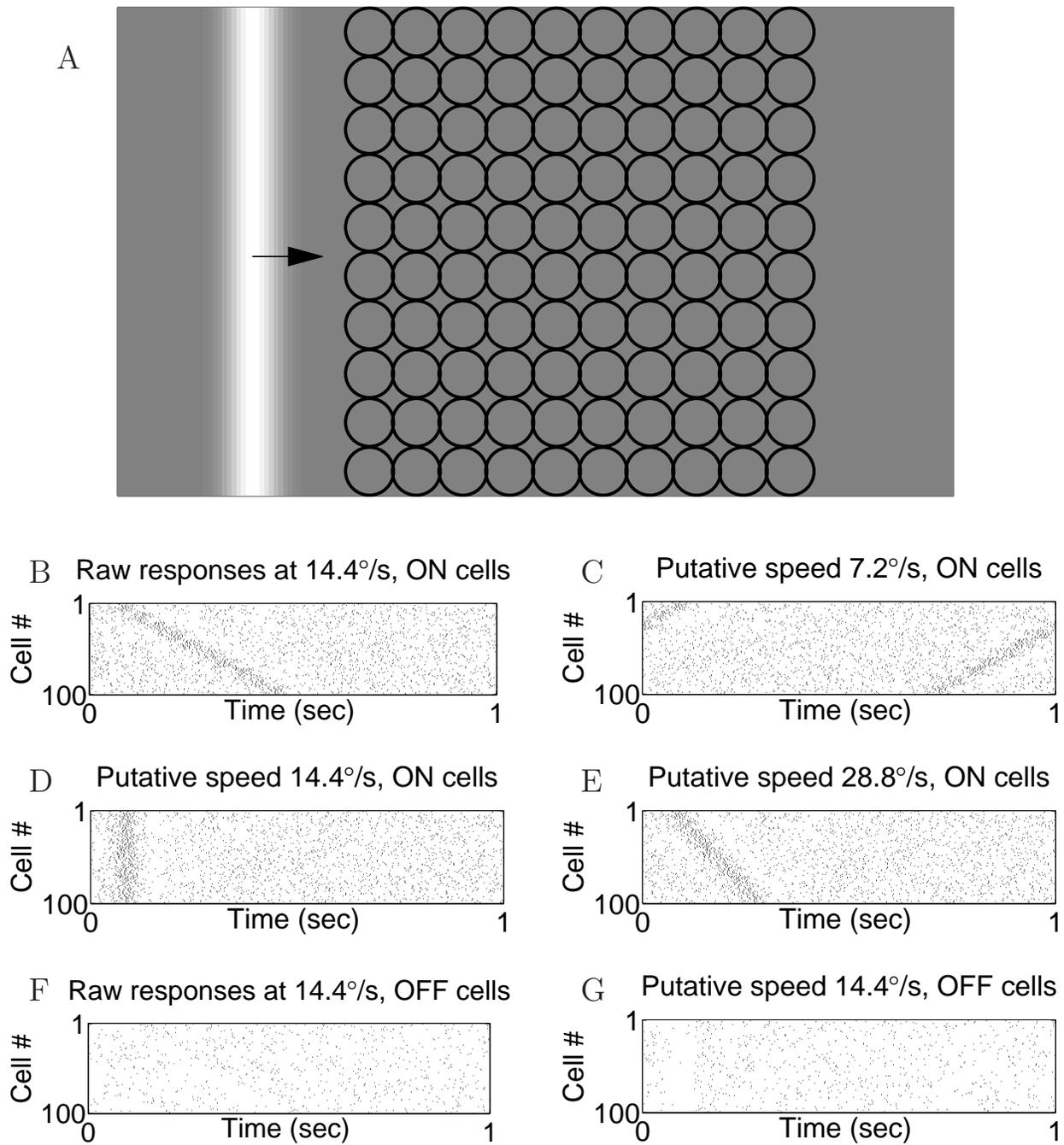


Fig. 1. Ensemble motion signals. (A) moving bar stimulus and cell layout. (B) and (F) show the raw responses from the ON and OFF cells, respectively, for a moving bar with speed $14.4^\circ/\text{s}$. Each tick represents one spike and each row represents the response of a different cell. (C-E) and (G) plot the 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.

Thus maximizing the marginal likelihood Eq. (23) is, to a good approximation, equivalent to maximizing the energy Eq. (21). The major difference between Eq. (21) and Eq. (23) 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). The dependence of this optimal filter as a function of v can be explained fairly intuitively, as we discuss at more length in the appendix, following Eq. (45). 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.

Fig. 2 illustrates these effects by plotting the optimal smoothing filters w_{LG} for a few different values of the velocity v . Interestingly, in the high-velocity limit, the analytically-derived optimal temporal filter width τ_w is on the order of 10 ms, which was the value chosen empirically for the optimal Gaussian filter used in [14]. 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 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 [14], 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 was borne out by our simulations, though it is worth noting that the improvement was significantly smaller than when the Bayesian decoder had access to the exact image.

2.C. Simulations

We simulated the presentation of a bar moving across the gray background of a CRT monitor refreshing at 120 Hz. The spatial profile of the bar in the direction of motion was a Gaussian function with a SD of 96 μm . The visual field was represented by a grid of 100 x 100 pixels

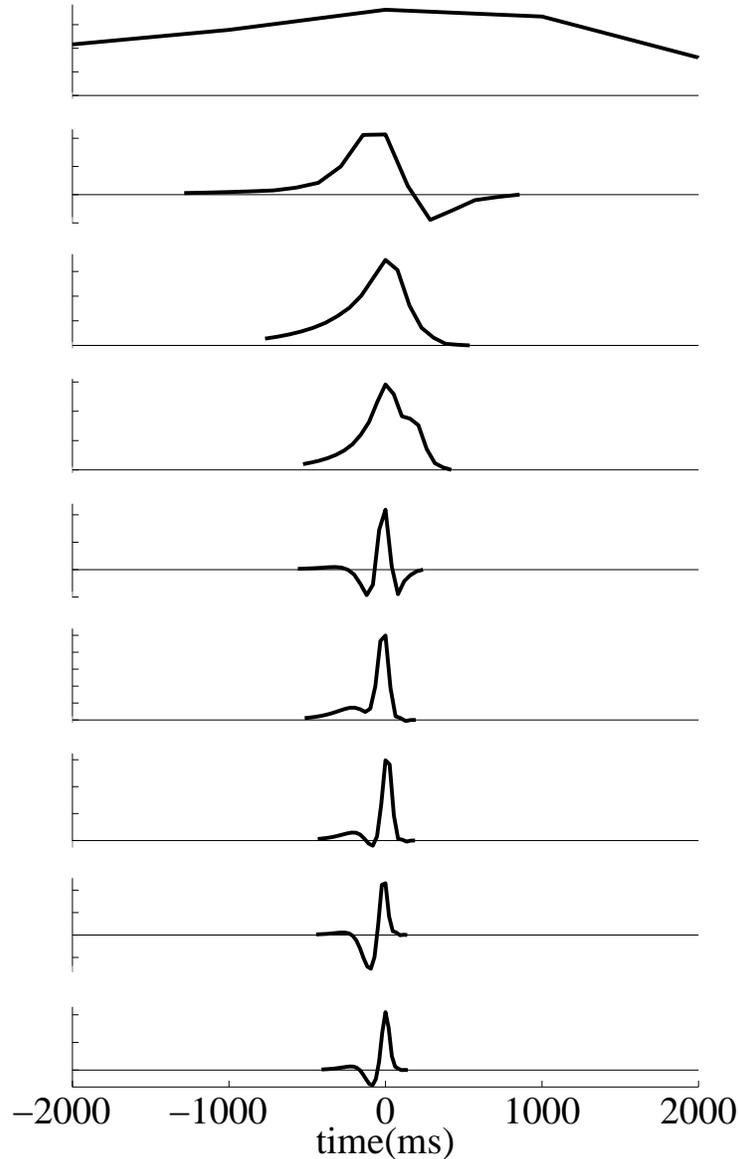


Fig. 2. Optimal linear spike train filter w_{LG} for velocities ranging from $0.2^\circ/s$ (top) to $9.8^\circ/s$ (bottom) in steps of $1.2^\circ/s$. The y axes are scaled in dimensionless units for clarity here. As discussed in section 2.B.3, there are three time scales that determine the time scale 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 wide (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 $\sim 10\text{--}20$ ms. For even higher velocities the shape of this filter remains essentially the same.

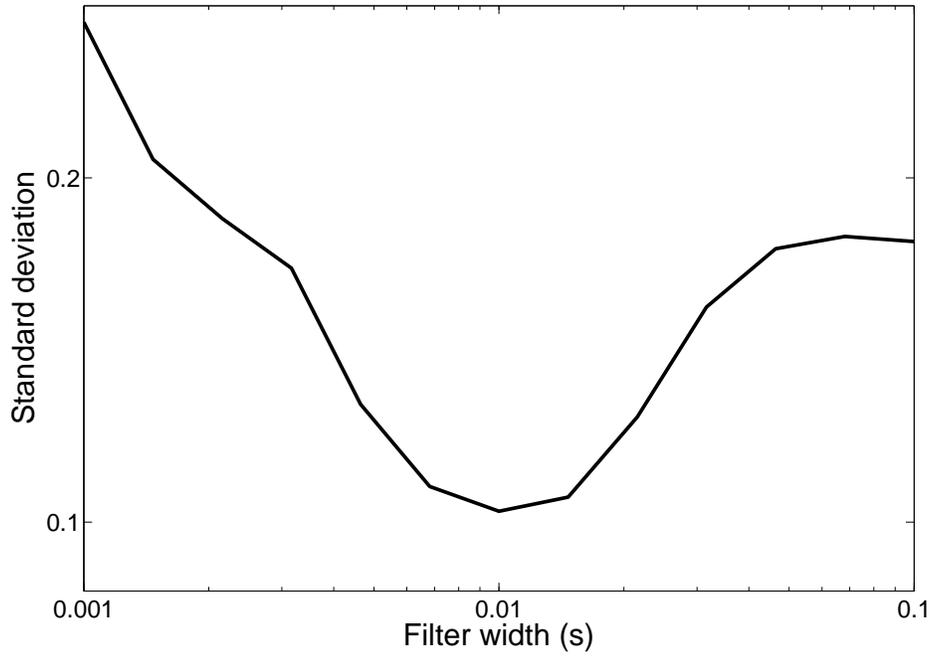


Fig. 3. Effect of filter width τ_w on the standard deviation of velocity estimates (obtained using the net motion signal described in section 2.B.2) across 100 presentations of a bar with luminance 0 moving at a speed of $28.8^\circ/\text{s}$. Note that a filter width of about $\tau_w \approx 10$ ms is optimal, in agreement with the findings of [14].

covering the receptive fields of 2 layers of cells each arranged in a uniform 10 x 10 grid. One layer consisted of ON cells, while the other represented OFF cells. The pixel resolution used was 10 times that used in [29] resulting in a pixel size of 12 μm . The bar moved across the visual field in discrete steps of $\mathbf{v}^{\text{pixels}/\text{refresh}}$, although \mathbf{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 $^\circ/\text{s}$ using the approximation $200 \mu\text{m}/^\circ$ [27] with a pixel size of 12 x 12 μm . This meant that, with a refresh rate of 120 Hz, a speed of $1^{\text{pixel}/\text{refresh}}$ corresponded to a speed of $7.2^\circ/\text{s}$.

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 8 speeds (10.8, 14.4, 21.6, 28.8, 36.0, 43.2, 50.4 and $57.6^\circ/\text{s}$) by 6 luminance levels (0, 0.125, 0.25, 0.75, 0.875 and 1 on a gray-scale level where 0 is black, 1 is white and the background level was set at 0.5). For each of these trials, we obtained a set of spike trains \mathbf{r} . From these spike trains, it was possible to estimate the speed of the stimulus used. Thus, we could compare speed estimates across stimulus conditions, by examining the standard deviation (SD) of estimates across the 100 trials performed for each condition. As in [14], we focused on the fractional SD (SD divided by 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 conditions. As will be seen, the fractional bias and the fractional SD are roughly on 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 luminance levels between 0.25 and 0.75 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 above chance.

As outlined above, we used three different decoding methods to estimate the stimulus velocity from the simulated spike train ensembles. Specifically, we compare Bayesian velocity decoding, with and without complete prior information about the image, with velocity estimation using the energy method. In particular, 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, 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. See Fig. 6B for an illustration of this procedure. Short flashes were used instead of a continuous uninterrupted presentation, because in the latter case, the cells immediately filter out the fixed image contrast, and thus after

a brief interval (~ 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 \mathcal{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

3.A. Comparison of the different velocity decoders

In this section we compare the performance of the energy model with Bayesian velocity decoding, with and without complete prior image information, as described in Sec. 2. Fig. 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, \mathbf{r} , in response to the moving bar stimulus, as a function of putative stimulus speed \mathbf{v} . Here, the true stimulus speed was $36.0^\circ/\text{s}$. Fig. 4(B) shows the value of the net motion signal N as a function of putative speed for the same stimulus. The Bayesian decoder with an *a priori* known image successfully estimated the speed in the trial shown, however the energy method resulted in a velocity estimate of $37.44^\circ/\text{s}$.

The lower panels of Fig. 4 show the distribution of speed estimates across 100 presentations of a bar of luminance 1 moving at a speed of $36.0^\circ/\text{s}$ using both the Bayesian decoder with known image (C) and energy method (D). Also plotted are Gaussian fits to the distributions with a mean \pm SD of $36 \pm 0.3^\circ/\text{s}$ for the optimal decoder and $36 \pm 0.9^\circ/\text{s}$ 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 Bayesian decoder with full prior knowledge of the image, 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% and 11%, for the Bayesian decoder with fully known image and the energy method, respectively. Velocity estimation based on the energy method does not make use of the image profile at any stage, and therefore we expect its performance to be closer to that of the Bayesian decoding with unknown image. Indeed, the fractional SD and the root mean square fractional error of the Bayesian decoder with uncertain prior image information averaged across all simulated stimulus conditions, were 6.4% and 6.9%, respectively.

3.A.1. Accuracy as a function of stimulus speed

Because in our simulations the moving bar stimulus only makes one pass over the “visual field”, more time is spent traversing the field and more spike train information is obtained for slower moving stimuli. Fig. 5(A) illustrates the fractional SD of 100 speed estimates for both of the Bayesian methods and the energy method, at each of the 8 stimulus speeds,

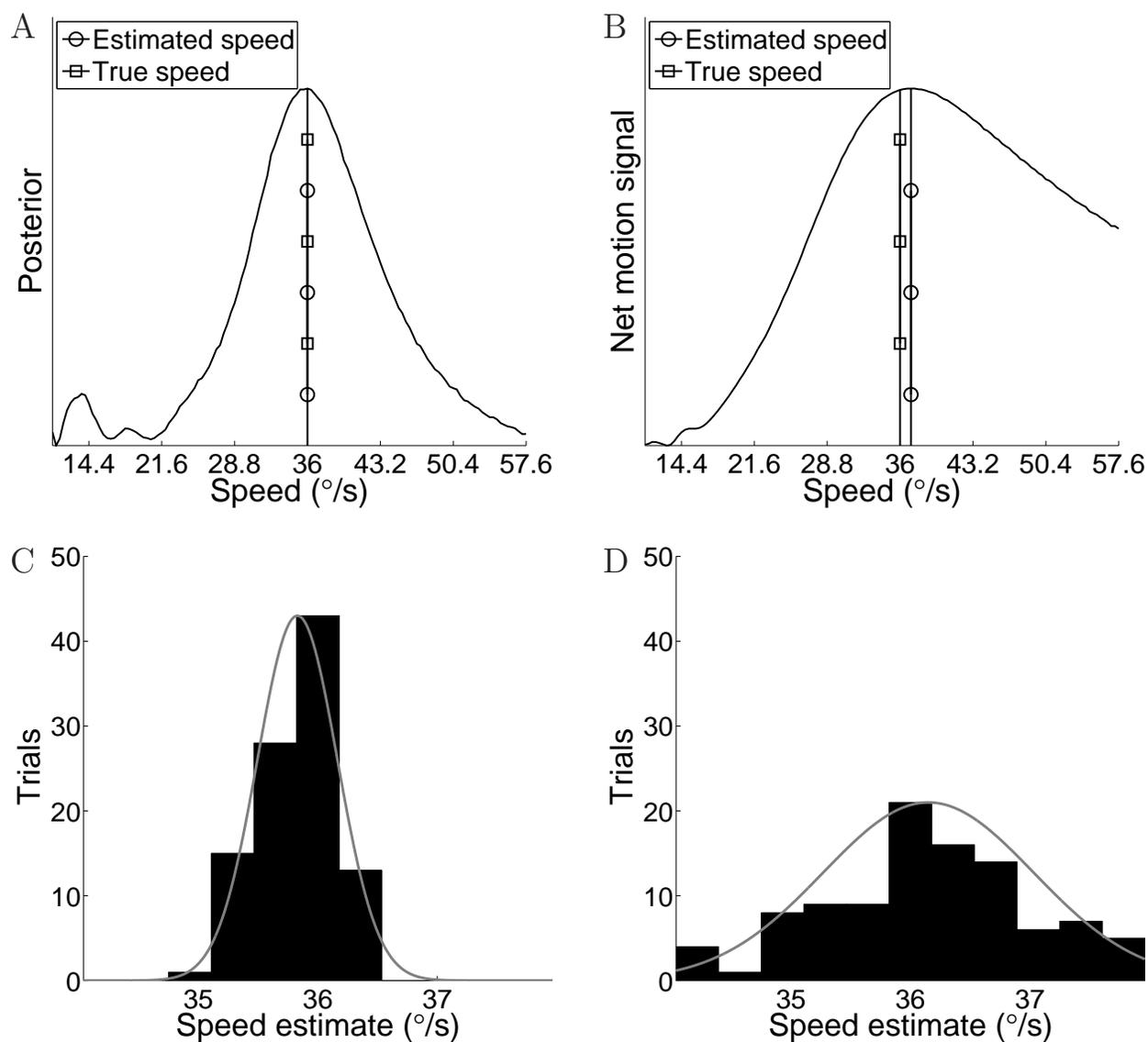


Fig. 4. The Bayesian method leads to more precise velocity estimates than does the energy-based “net motion signal” method. (A) Posterior, $p(\mathbf{r}|\mathbf{v})$ and (B) net motion signal, N , as a function of putative stimulus speed \mathbf{v} for spike trains generated using a stimulus with speed 36.0°/s. Distribution of speed estimates across 100 presentations of a bar moving at a speed of 36.0°/s using the posterior probability (C) and net motion signal (D). Also plotted are a Gaussian fits to the distributions with mean \pm SD of 35.8 ± 0.34 for the optimal decoder and 36.2 ± 0.89 for the net motion signal.

averaged across the 6 luminance 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. As expected, the advantage of the Bayesian decoder over the energy method is partly lost when its prior information about the image is uncertain.

3.A.2. Accuracy as a function of stimulus luminance

Lowering the luminance 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 [29] lacks explicit luminance- or contrast-gain control effects; thus, these results should be interpreted in terms of local modifications around a fixed luminance pedestal which 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 6 luminance levels across the 8 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.A.3. Effect of luminance 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 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 where humans tend to choose the slowest motion that explains the incoming information [41], i.e., we have a bias toward slower speeds. As can be seen in Fig. 5(C), the energy method is biased towards lower velocity estimates at higher stimulus speeds. The Bayesian decoder with full image information shows a very slight tendency in this direction also. On the other hand, the Bayesian decoder without full prior knowledge of image has a positive bias towards higher velocities. The second phenomenon of interest was that where stimuli with low contrast are typically perceived as moving slower than those with high contrast [36, 38]. Fig. 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, $\langle \mathbf{v}_* \rangle$ normalized by \mathbf{v} versus the stimulus luminance for both the Bayesian decoder and the energy method against the stimulus luminance, averaged across all speeds tested in our simulations. There appears to be a slight trend towards 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

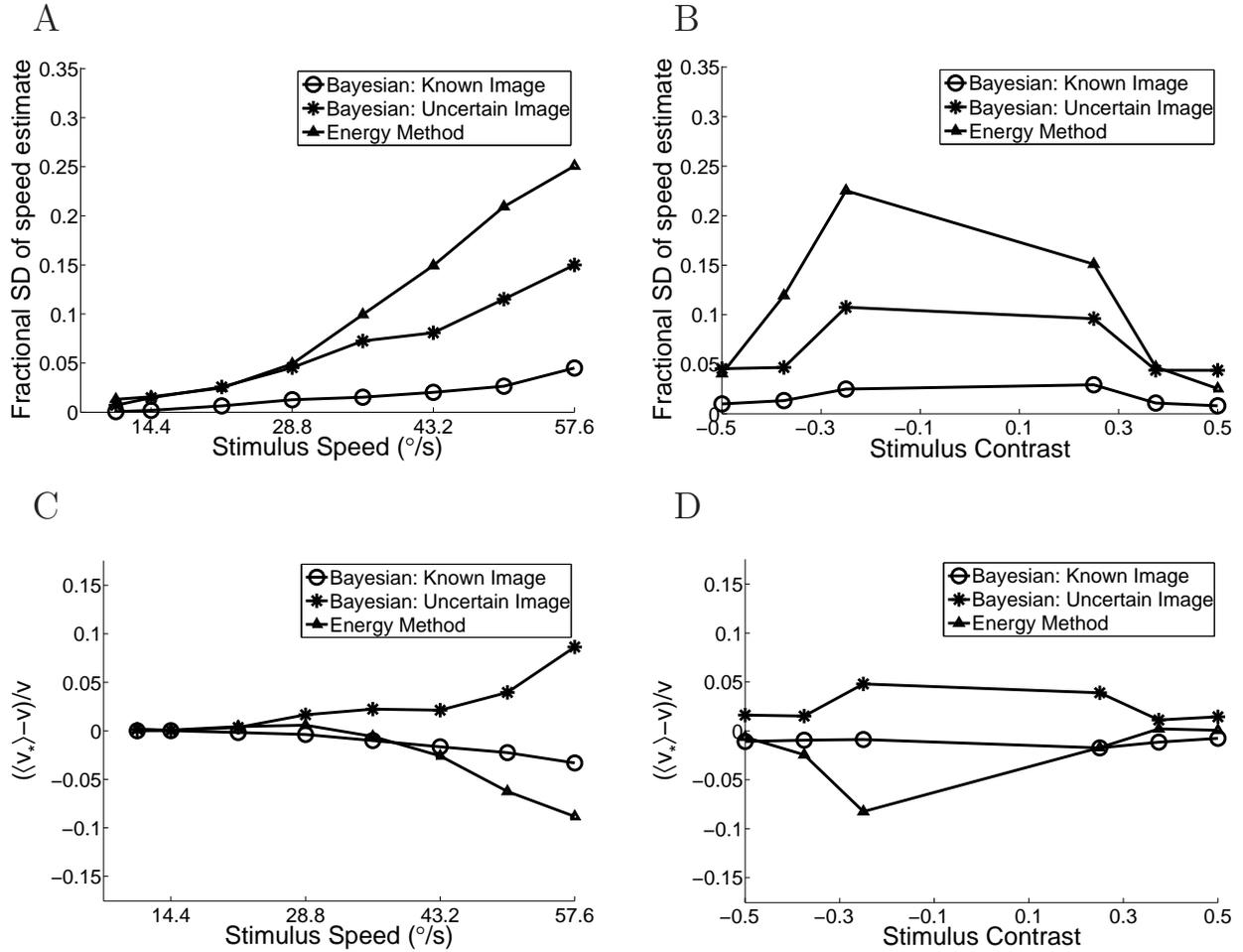


Fig. 5. Fractional standard deviation of speed estimates versus: (A) stimulus speed and (B) stimulus luminance, for the Bayesian decoder with full image information, the Bayesian decoder with incomplete image information and the energy method. (C) and (D) plot the difference between the mean estimated speed, $\langle v_* \rangle$, and the true stimulus speed, v , normalized by v versus the true stimulus speed and stimulus luminance, respectively. Note that the Bayesian decoder provides more precise estimates than the energy method at all levels, with performance improving with prior image information.

difficult to say anything definitive about a relationship between stimulus contrast and speed estimate bias.

3.A.4. Effect of prior image information

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 fully known image (shown by dashed lines). 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.

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 Sec. 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 the GLM-based Bayesian decoder was largely lost when the latter decoder’s prior knowledge of the image became too uncertain.

3.B. *Effects of manipulating model parameters*

3.B.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 representing 100 trials, for each of the 48 conditions mentioned above (i.e., 8 different speeds and 6 different luminance 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 to the output from the full model, which would have had a confounding effect on the results.

The results are shown in Fig. 7(A) and 7(B) for the Bayesian decoder and the energy method, respectively, and are plotted versus the fractional standard deviation of the speed

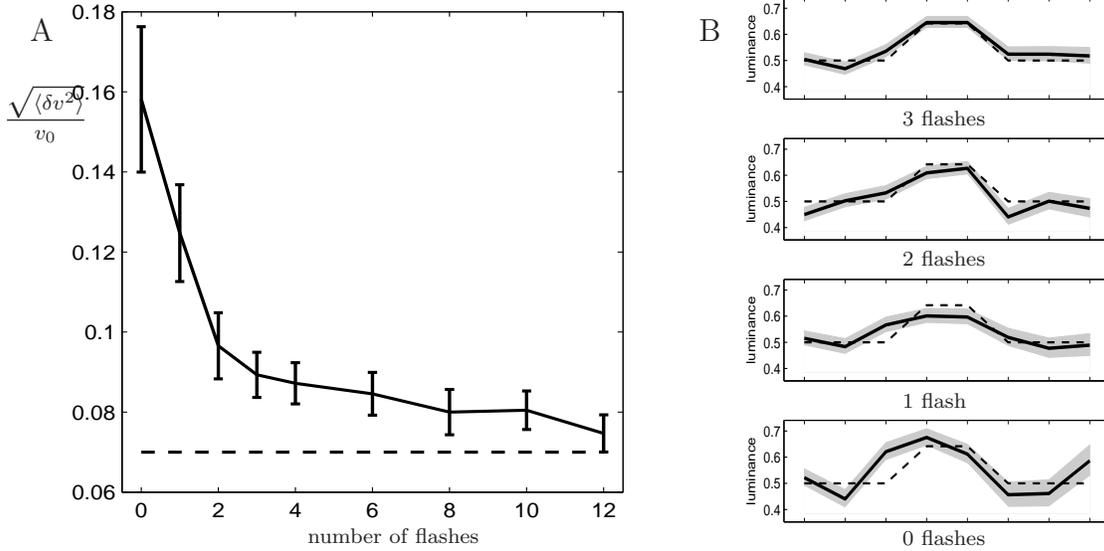


Fig. 6. Effect of decreasing image uncertainty on accuracy of Bayesian velocity estimation. See section 2.C for a detailed description of this simulation. A) The solid line with error bars shows the drop in the fractional rms error of the velocity estimate for an *a priori* unknown image, as the number of preview flashes increases. The dashed line is the fractional error for the case of *a priori* known image. The true velocity was $28.8^\circ/\text{sec}$ and the image contrast, 0.3. B) The plots show the maximum *a posteriori* estimate of the image luminance profile (solid line) 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 line shows the actual image profile.

estimate for the same 48 conditions using the spike train ensembles obtained directly from the model. The diagonal lines in Fig. 7(A) and 7(B) 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 (c.f. Fig. 2 in [29]), this trial-shuffling procedure did not significantly hurt the performance of either velocity estimator; in fact, if anything, there is a slight bias in Fig. 7(A) and 7(B), 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.

3.B.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% [29]. 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 [14]. 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, 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 last 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 filters 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 Fig. 7(C) and 7(D) for the Bayesian 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

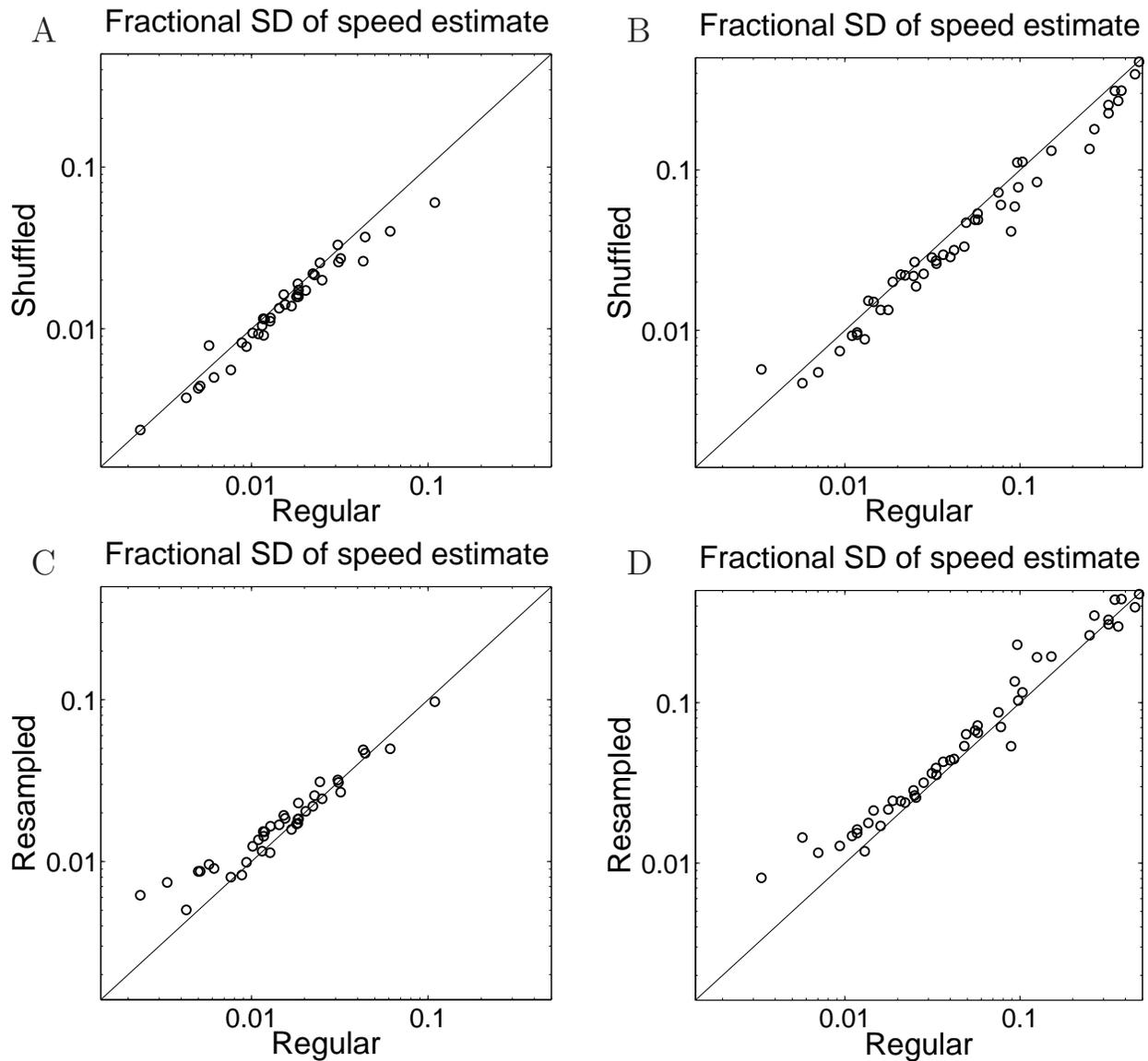


Fig. 7. Effect of correlated activity and spike timing structure on speed estimates. Fractional SD of speed estimates using shuffled responses plotted as a function of that obtained using regular simulated data for the Bayesian decoder (A) and energy method (B). Fractional SD of speed estimate using resampled spike trains plotted as a function of that obtained using regular simulated data for the Bayesian decoder (C) and energy method (D). Diagonal lines indicate equality. Note that the performance of the decoders is relatively unaffected by these rather drastic manipulations of spike timing.

above the identity line.

3.B.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 firing rate (b_i in Eq. (2)) of 2 and all OFF cells a baseline firing rate of 3, corresponding to the mean values obtained when fitting the model [29]. In order to examine a somewhat more biologically realistic case we jittered the center location of the cells as in Fig. 8(B) and randomly selected the baseline firing rates of the ON and OFF cells from uniform distributions on interval 1 to 3 and 1.5 to 4.5, respectively.

Fig. 8(C) and (D) illustrates the speed estimates over 100 trials for a stimulus with speed of $28.8^\circ/s$ and luminance of 0 using the regular cell arrangement and uniform baseline firing rates (left) versus the jittered cell arrangement and random baseline firing rates (right). No significant difference in performance is apparent.

While randomly jittering the baseline firing rates around the mean caused no change in estimation accuracy, this does not allow us to comment on the possible effects of changes in the *mean* baseline firing rate. To assess this, we also carried out 100 simulations, using a stimulus with speed of $28.8^\circ/s$ and a luminance of 0, where the cells were arranged in the original simple grid and the ON and OFF cells were given baseline firing rates of 4 and 6, respectively. This was compared to the distribution of speed estimates for 100 trials, using the same stimulus and cell arrangement, but where the baseline firing rates were 2 and 3 for the ON and OFF cells, respectively. Fig. 8(E) illustrates the significantly improved estimation performance obtained by inflating the baseline firing rates compared to the fitted values used throughout the rest of this study.

4. Discussion

The model of [29] 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 somewhat 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 only have 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 [1,6].

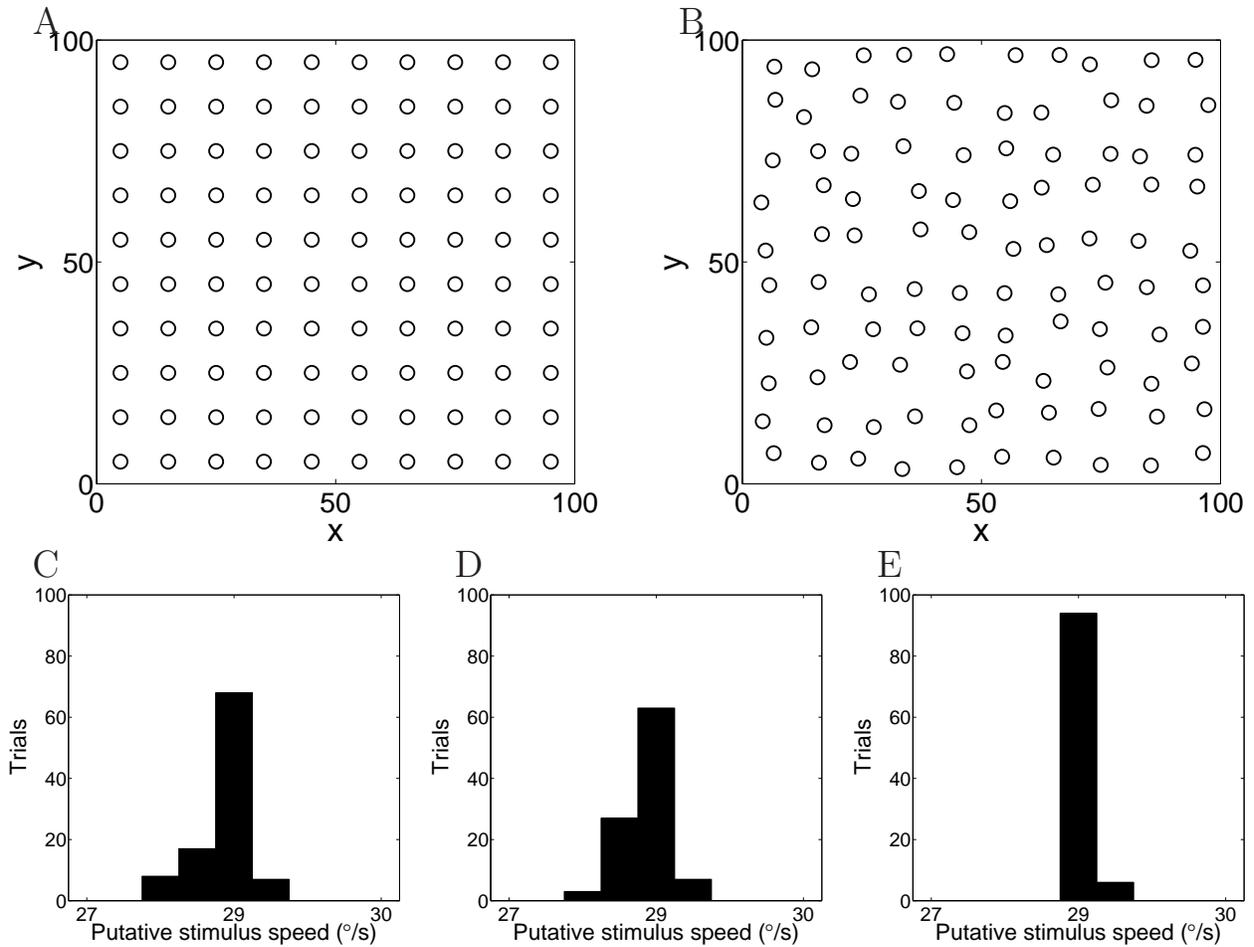


Fig. 8. Simple rectangular grid cell arrangement (A), jittered cell arrangement (B). Histograms illustrating the velocity estimates over 100 trials for a stimulus with velocity $28.8^\circ/s$ and luminance of 0 using the regular cell arrangement and uniform baseline firing rates (C) and the jittered cell arrangement and random baseline firing rates (D). Similar performance was obtained with both the rectangular-grid and randomized spatial layouts. (E) illustrates the improved estimation performance obtained by doubling the baseline firing rates from 2 and 3 to 4 and 6 for the ON and OFF cells respectively.

A connection between Bayesian velocity estimation and the energy method of [1] has been noted before [33]. 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 [1], a link was drawn between the two methods. To the best of our knowledge, however, a mathematical solution to the Bayesian GLM decoding problem we solve here has not been previously described. Furthermore, we believe our work on the marginal likelihood decoding of static images in the LG case to be novel.

Because of the connection between the two methods, we have compared the precision of speed estimates obtained using the optimal Bayesian decoder, with full image knowledge, to that obtained using the energy method. In all simulations performed in the present study, the optimal Bayesian decoder outperforms the energy method. Using our particular set of 48 stimulus conditions, we found that the optimal decoder achieved an average relative precision of 2%, with the energy method only realizing 10% relative precision. This result is not surprising given the extra image information available to the former. It is interesting, however, to compare the estimation performance using our model to that obtained using similar stimuli with real cells [14]. 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 somewhat 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 [14], i.e., speeds of (10.8, 14.4, 28.8, and 57.6 °/s) and luminance levels (0 and 1), we obtain a value of 2.8% which is of the same order as their result.

Also reported in [14] was the finding that the optimal filter for velocity estimation from cell population responses was of the order of 10 ms. This implies that the elementary motion signal was conveyed in a timespan comparable to the interspike interval of RGCs. In the present study, our analytically-derived optimal filter is shown to be of similar width in the case where stimulus velocities are above about 5 °/s (Fig. 2). Replicating the finding of [14], an optimal width of 10 ms was also demonstrated using simulations on our model (Fig. 3).

We examined the precision of our speed estimates as a function of both stimulus speed and stimulus luminance. As expected, decoding performance improves with increasing luminance and with decreasing speed (Fig. 5). Fig. 5(A) illustrates that our model approximately followed a Weber-Fechner law with visual speed discrimination being roughly proportional to speed [22]. As discussed in Sec.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. If the bar moves twice as fast, we might reasonably

expect to have approximately half as much “signal” and, thus, the fact that the relationship between speed and estimated speed precision appears to be roughly linear is not surprising. Supporting this notion, [14] presented a simple model of speed estimate precision that proposed a quadratic relationship between estimated speed variability and speed, i.e., a linear relationship between fractional SD and 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. [35] used an optimal Bayesian observer model 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 due to the fact that the broader likelihood (based on psychophysical measurements), when multiplied by a prior favouring low speeds [5], resulted in a larger shift towards zero than multiplication by a narrower likelihood. Thus, low contrast stimuli, giving noisier measurements, result in an underestimation of stimulus speed, agreeing with psychophysical reports [36]. 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 towards 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)). It is worth emphasizing that this is not the same phenomenon as described in the Bayesian

model of [42], where the bias in the Bayesian estimate was due to a strong prior term which preferentially weighted slow speeds; as discussed in the Methods section, we are employing a MAP estimator with a uniform prior, which is equivalent to using a maximum likelihood estimator and ignoring the prior term completely. Instead, the results shown here 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 unknown image Bayesian decoder and the energy-based estimate is clarified by the connection between these two methods as described in Sec. 2.B.3 and the appendix. Specifically, see the discussion after Eqs. (23)–(24) of Sec. 2.B.3, and Eqs. (31)–(32) of the appendix).

[29] found that, when comparing the full RGC model with an uncoupled version (retaining spike history effects), Bayesian stimulus decoding recovered 20% more information, using pseudo-random stimuli. They 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 to speed estimation precision. We followed the procedure employed in [14] and, as in that study, it appeared that the shuffled, uncorrelated spike trains surprisingly resulted in a weak improvement in estimation precision. We also replicated their test of how precise spike timing might effect speed estimation precision [14]. Again, as in their study, we found similar results. Specifically, decoding speed using the resampled spike trains resulted in a slight decrease in performance. However, despite the fact that we have completely abolished the intra- and inter-neuronal non-stimulus-driven correlation structure here, these decreases in performance were quite small, indicating that velocity decoding does not depend strongly on the fine spike train structure here. 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 resample 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 virtually no difference to the accuracy of the estimated velocity (not shown).

For the majority of the simulations performed in the present study, the model cells were arranged in a simplistic grid pattern (Fig. 8(A)), all ON cells were assigned one baseline firing rate and all OFF cells were assigned another. In order to make our model more biologically realistic we manipulated both the physical arrangement and the baseline spiking rate of the cells (Fig. 8(B)). We tested the speed estimation performance of the optimal Bayesian decoder using cell location's which were randomly jittered around their original location and whose baseline firing rates were randomized around the original values. No change in performance was apparent (Fig. 8(C,D)). This is not surprising given that the decoder was furnished with the locations of the cells in the new arrangement and that the total number of spikes generated by the model was not altered. [14] found improved speed estimation performance using a cell arrangement where cells were more dispersed along the axis of motion, however there was no difference between the amount of dispersal along the axis of motion in our two cell arrangements. While randomizing the baseline firing rates around the data-fitted values did not result in any change in estimation performance, a population-wide increase in firing rate caused a significant improvement. Fig. 8(E,F) illustrates the improvement obtained by doubling the baseline firing rates. Again, this is the expected result considering that the increased spiking rate leads to a higher signal to noise ratio and results in a greater amount of information about the stimulus in the spiking activity.

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, in which [13] employed the energy method (Sec. 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, suggesting that readout of the retinal population code can be extremely efficient when exposure to the moving stimulus is very brief. 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 a strategy based on marginalization over the uncertain image seems to be more consistent with the available data.

As in the present study, Bayesian inference was recently used to estimate properties, including velocity, of a visual motion stimulus from ensemble spike train responses [19, 37]. This study reported that individual ganglion cells in the turtle retina encode velocity and even acceleration. The authors employed Bayesian inference to determine the MAP estimate of the stimulus speed using the stimulus speed prior and the response likelihood, based on

average firing rates in specified time bins in response to different speeds. They assumed that cell responses were independent of each other and determined the likelihood as the product of single-neuron likelihoods. Our study differs in a number of ways. First of all, our Bayesian decoder does not operate on binned firing rates but on individual spike times. This allows for greater investigation of the importance of the specific spike timing structure in determining stimulus velocity. Secondly, our study explicitly takes account of both spike history effects and correlations between cells in estimating speed. Finally, the lone, relatively low spatial frequency stimulus used in our study was chosen to investigate the fidelity of global velocity encoding across the entire population of RGCs. [37] used a stimulus with a much higher spatial frequency content. Using such a stimulus, an increase in translation speed equates to an increase in the number of on/off and off/on stimulus transitions seen by each cell, per unit time. Presumably, this would cause a corresponding increase in firing rate in a certain percentage of cells. Given that the Bayesian decoder used is based on average binned firing rates, the possibility exists that the reported encoding of velocity by individual cells is somewhat influenced by the change in the number of discrete stimulus events occurring per unit time that accompanies a change in velocity. Further work using our model may serve to address this issue.

5. Conclusion

Optimal Bayesian decoding with full image information has been shown to outperform a “motion energy” method that uses no prior image information, which in turn was shown to outperform human psychophysical performance [13]. A mathematical description of the connection between these two decoders indicates that, in addition to the extra information about the image used by the Bayesian estimator, information about the network’s spatio-temporal 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.

Acknowledgements

Thanks to J. Pillow for providing us with the parameters for the network model introduced in [29], and to E.J. Chichilnisky and E.P. Simoncelli for many useful comments. YA and LP are partially supported by NEI Grant R01 EY018003 and by a McKnight Scholar award to LP. YA is additionally supported by a Patterson Trust Fellowship in Brain Circuitry. EL is supported by an IRCSET Government of Ireland Postdoctoral Research Fellowship.

Appendix: 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 Linear Gaussian (LG) model of the RGC's is closely related to the energy function of the reference [14], and thus for this model the Bayesian velocity decoding is nearly equivalent to the energy model approach. 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. (22). Thus we have $p_{\text{LG}}(\mathbf{r}|\mathbf{x}, \mathbf{v}) = \prod_i \mathcal{N}(b_i + \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}, \Sigma)$. 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 Sec. 2.B.1. For the LG model, the log-posterior function is given by

$$\begin{aligned} \mathcal{L}_{\text{LG}}(\mathbf{x}, \mathbf{r}, \mathbf{v}) &\equiv \log [p_x(\mathbf{x})p_{\text{LG}}(\mathbf{r}|\mathbf{x}, \mathbf{v})] = \\ &-\frac{1}{2}\mathbf{x}^T \mathcal{C}_x^{-1} \mathbf{x} - \frac{1}{2} \sum_i (\mathbf{r}_i - b_i - \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x})^T \Sigma^{-1} (\mathbf{r}_i - b_i - \mathcal{K}_{i,\mathbf{v}} \cdot \mathbf{x}) + \text{const.}, \end{aligned} \quad (25)$$

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

$$p_{\text{LG}}(\mathbf{r}|\mathbf{v}) = \int e^{\mathcal{L}_{\text{LG}}(\mathbf{x}, \mathbf{r}, \mathbf{v})} d\mathbf{x}, \quad (26)$$

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

$$\mathbf{x}_{\text{MAP}}(\mathbf{r}, \mathbf{v}) = H(\mathbf{v})^{-1} \sum_i \mathcal{K}_{i,\mathbf{v}}^T \cdot \Sigma^{-1} \cdot (\mathbf{r}_i - b_i). \quad (27)$$

Here, the negative Hessian is given by

$$H(\mathbf{v}) = -\nabla_{\mathbf{x}} \nabla_{\mathbf{x}} \mathcal{L}_{\text{LG}} = \mathcal{C}_x^{-1} + \sum_i \mathcal{K}_{i,\mathbf{v}}^T \cdot \Sigma^{-1} \cdot \mathcal{K}_{i,\mathbf{v}}, \quad (28)$$

which is now independent of the observed spike trains \mathbf{r} . Using Eqs. (27)–(28), we can rearrange the terms in Eq. (25) to complete the square for \mathbf{x} , and obtain

$$\begin{aligned} \mathcal{L}_{\text{LG}}(\mathbf{x}, \mathbf{r}, \mathbf{v}) &= -\frac{1}{2}(\mathbf{x} - \mathbf{x}_{\text{MAP}})^T H(\mathbf{v})(\mathbf{x} - \mathbf{x}_{\text{MAP}}) \\ &\quad - \frac{1}{2} \sum_i \delta \mathbf{r}_i^T \Sigma^{-1} \delta \mathbf{r}_i + \frac{1}{2} \sum_{ij} \mathbf{X}_i^T \mathcal{C}_x(\mathbf{v}) \mathbf{X}_j + \text{const.}, \end{aligned} \quad (29)$$

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

$$\mathbf{X}_i \equiv \mathcal{K}_{i,\mathbf{v}}^T \Sigma^{-1} \delta \mathbf{r}_i. \quad (30)$$

The marginalization in Eq. (26) is thus a standard Gaussian integration, which yields

$$\log p_{\text{LG}}(\mathbf{r}|\mathbf{v}) = \frac{1}{2} \sum_{ij} \mathbf{X}_i^T \mathcal{C}_x(\mathbf{v}) \mathbf{X}_j - \frac{1}{2} \log |\mathcal{C}_x H(\mathbf{v})| + \text{const.} \quad (31)$$

(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. (31) 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 $\mathbf{x}_{\text{MAP}}(\mathbf{r}, \mathbf{v})$. Unlike Eq. (18), however, although the second term on the right hand side of Eq. (31) 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 's) is the first, which we denote by $\mathcal{E}_{\text{LG}}(\mathbf{v}, \mathbf{r})$. We will see that this term corresponds closely to the energy function introduced in [14]. More explicitly, we have

$$\mathcal{E}_{\text{LG}}(\mathbf{v}, \mathbf{r}) \equiv \frac{1}{2} \sum_{ij} \mathbf{X}_i^{\text{T}} \mathcal{C}_x(\mathbf{v}) \mathbf{X}_j = \frac{1}{2} \sum_{ij} \int \int X_i(\mathbf{n}_1) \mathcal{C}_x(\mathbf{n}_1, \mathbf{n}_2; \mathbf{v}) X_j(\mathbf{n}_2) d^2 \mathbf{n}_1 d^2 \mathbf{n}_2. \quad (32)$$

In the following we will rewrite Eq. (32) 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 [29]). From Eq. (30) and the definition of $\mathcal{K}_{i,\mathbf{v}}$, Eq. (7), we then obtain the explicit form

$$X_i(\mathbf{n}) = \frac{1}{\sigma^2} \int dt \int d\tau k_i(t - \tau, \tau \mathbf{v} + \mathbf{n}) \delta r_i(t). \quad (33)$$

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 1-d case, which we are studying in this paper, the image profile $\mathbf{x}(\mathbf{n})$, and hence the prior image covariance, only depend 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 perform the integrals over n_{\perp} in Eq. (32), and rewrite it as

$$\mathcal{E}_{\text{LG}}(\mathbf{v}, \mathbf{r}) = \frac{1}{2} \sum_{ij} \int \int \tilde{X}_i(n_1) \mathcal{C}_x(n_1, n_2) \tilde{X}_j(n_2) dn_1 dn_2, \quad (34)$$

$$\tilde{X}_i(n) \equiv \int X_i(\mathbf{n}) dn_{\perp} = \frac{1}{\sigma^2} \int dt \int d\tau \tilde{k}_i(t - \tau, \tau v + n) \delta r_i(t), \quad (35)$$

where $v \equiv |\mathbf{v}|$, and we defined $\tilde{k}_i(t, n) \equiv \int k_i(t, \mathbf{n}) dn_{\perp}$. 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 ($\sim 1^\circ$). Hence, if we define

$$q_i(t, n) \equiv \tilde{k}_i(t, n + n_i) = \int k_i(t, \mathbf{n} + \mathbf{n}_i) dn_{\perp} \quad (36)$$

(where $n_i \equiv \mathbf{n}_i \cdot \hat{\mathbf{v}}$), $q_i(t, n)$ vanishes when $|n| \gg 1^\circ$; for all cells, q_i 's 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 Sec. 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 $\tilde{R}_i(t) \equiv \tilde{X}_i(n_i - vt)$ (equivalently, $\tilde{X}_i(n) = \tilde{R}_i((-n + n_i)/v)$), and rewrite Eq. (34) by changing the integration variables from $n_{1(2)}$ to $vt_{1(2)}$:

$$\mathcal{E}_{\text{LG}}(\mathbf{v}, \mathbf{r}) = \frac{1}{2v^2} \sum_{ij} \int \int \tilde{R}_i \left(-t_1 + \frac{n_i}{v} \right) \mathcal{C}_x(vt_1, vt_2) \tilde{R}_j \left(-t_2 + \frac{n_j}{v} \right) dt_1 dt_2. \quad (37)$$

Using Eq. (35) and the definition (36), we write $R_i(t_1)$ explicitly as

$$\begin{aligned} \tilde{R}_i(t_1) &\equiv \tilde{X}_i(n_i - vt_1) = \frac{1}{\sigma^2} \int dt \int d\tau \tilde{k}_i(t - \tau, v\tau - vt_1 + n_i) \delta r_i(t) \\ &= \frac{1}{\sigma^2} \int dt \int d\tau q_i(t - \tau, v(\tau - t_1)) \delta r_i(t). \end{aligned} \quad (38)$$

Exploiting the translation invariance of the prior image ensemble which 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

$$\mathcal{C}_x(n_1 - n_2) = \int \mathcal{B}_x(n_1 - n) \mathcal{B}_x(n_2 - n) dn. \quad (39)$$

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 [7]². Substituting definition (39) (after renaming the integration variable n to vt) in Eq. (37), we rewrite the latter as

$$\begin{aligned} \mathcal{E}_{\text{LG}}(\mathbf{v}, \mathbf{r}) &= \quad (40) \\ \frac{1}{2v} \sum_{ij} \int \int \int \tilde{R}_i \left(-t_1 + \frac{n_i}{v} \right) \mathcal{B}_x(v(t_1 - t)) \mathcal{B}_x(v(t_2 - t)) \tilde{R}_j \left(-t_2 + \frac{n_j}{v} \right) dt_1 dt_2 dt &= \\ \frac{1}{2v} \sum_{ij} \int \int \int \tilde{R}_i(t_1) \mathcal{B}_x \left(v \left(t + \frac{n_i}{v} - t_1 \right) \right) \mathcal{B}_x \left(v \left(t + \frac{n_j}{v} - t_2 \right) \right) \tilde{R}_j(t_2) dt_1 dt_2 dt. \end{aligned}$$

We derived the last line by renaming the integration variables as $t_{1(2)} \rightarrow n_{i(j)}/v - t_{1(2)}$, and $t \rightarrow -t$. Finally, defining

$$R_i(t) \equiv \frac{1}{\sqrt{v}} \int \mathcal{B}_x(v(t - t_1)) \tilde{R}_i(t_1) dt_1, \quad (41)$$

²In particular, for $\mathcal{C}_x(n_1 - n_2) = c^2 e^{-\frac{|n_1 - n_2|}{l_{\text{corr}}}}$, we have $\mathcal{B}_x(n) = c \sqrt{\frac{2}{l_{\text{corr}}}} \theta(n) e^{-n/l_{\text{corr}}}$, 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 Sec. 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.

we obtain

$$\mathcal{E}_{\text{LG}}(\mathbf{v}, \mathbf{r}) = \frac{1}{2} \sum_{ij} R_i \left(t + \frac{n_i}{v} \right) R_j \left(t + \frac{n_j}{v} \right) dt. \quad (42)$$

Equation (42) is akin to the energy function used in Frechette *et al*, and together with Eq. (31) yields Eq. (23) of Sec. 2.B.3. To find the explicit form of the smoothing filter in Eq. (24), we compare that equation, in the form

$$R_i(t) = \int w_{\text{LG}}(t - t') \delta r_i(t') dt', \quad (43)$$

with definition (41)

$$\begin{aligned} R_i(t) &= \frac{1}{\sigma^2 \sqrt{v}} \int dt_1 \int dt' \int d\tau \mathcal{B}_x(v(t - t_1)) q_i(t' - \tau, v(\tau - t_1)) \delta r_i(t'), \\ &= \frac{1}{\sigma^2 \sqrt{v}} \int dt_1 \int dt' \int d\tau \mathcal{B}_x(v(t - t' - t_1)) q_i(-\tau, v(\tau - t_1)) \delta r_i(t'), \end{aligned} \quad (44)$$

(where we used Eq. (38) to write the first line, and shifted τ and t_1 by t' to derive the second), and obtain

$$w_{\text{LG}}(t) = \frac{1}{\sigma^2 \sqrt{v}} \int dt_1 \int d\tau \mathcal{B}_x(v(t - t_1)) q_i(-\tau, v(\tau - t_1)). \quad (45)$$

Thus, $R_i(t)$ is a version of the response function of the cell i , offset by its baseline firing rate b_i , and smoothed out on the time scale dictated by the largest of the spatio-temporal 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, due to the finite support of the factors of its integrand, the double integral Eq. (45) 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 Sec. 2.B.3, following Eq. (24).

References

1. E.H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A*, 2(2):284–99, 1985.
2. Y. Ahmadian, J. Pillow, and L. Paninski. Efficient markov chain monte carlo methods for decoding neural spike trains. *Under review, Neural Computation*, 2008.

3. D. Ascher and N.M. Grzywacz. A bayesian model for the measurement of visual velocity. *Vision Res.*, 40:3427–3434, 2000.
4. W. Bialek and A. Zee. Coding and computation with neural spike trains. *Journal of Statistical Physics*, 59:103–115, 1990.
5. M.R. Blakemore and R.J. Snowden. The effect of contrast upon perceived speed: a general phenomenon? *Perception*, 28:33–48, 1999.
6. David C. Bradley and Manu S. Goyal. Velocity computation in the primate visual system. *Nature Reviews Neuroscience*, 9:686–695, 2008.
7. 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. *Journal of Vision*, To appear, 2008.
8. D. Brillinger. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cyberkinetics*, 59:189–200, 1988.
9. 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. *Journal of Neuroscience*, 18:7411–7425, 1998.
10. E.J. Chichilnisky and R.S. Kalmar. Functional asymmetries in ON and OFF ganglion cells of primate retina. *J Neurosci*, 22(7):2737–2747, 2002.
11. E.J. Chichilnisky and R.S. Kalmar. Temporal resolution of ensemble visual motion signals in primate retina. *J Neurosci*, 23:6681–6689, 2003.
12. D. Field. Relations between the statistics of natural images and the response profiles of cortical cells. *Journal of the Optical Society of America A*, 4:2379–2394, 1987.
13. Eric S. Frechette, Matthew I. Grivich, Rachel S. Kalmar, Alan M. Litke, Dumitru Petrusca, Alexander Sher, and E. J. Chichilnisky. Retinal motion signals and limits on speed discrimination. *J. Vis.*, 4(8):570, 2004.
14. E.S. Frechette, A. Sher, M.I. Grivich, D. Petrusca, A.M. Litke, and E.J. Chichilnisky. Fidelity of the ensemble code for visual motion in the primate retina. *J Neurophysiol*, 94(1):119–135, 2005.
15. F. Hurlimann, D. Kiper, and M. Carandini. Testing the bayesian model of perceived speed. *Vision Research*, 42:2253–2257, 2002.
16. R. Kass and A. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.
17. D. Knill and W. Richards, editors. *Perception as Bayesian Inference*. Cambridge University Press, 1996.
18. S. Koyama and S. Shinomoto. Empirical Bayes interpretations of random point events. *J. Phys. A*, 38:531–537, 2005.
19. J. Kretzberg, I. Winzenborg, and A. Thiel. Bayesian analysis of the encoding of constant

- and changing stimulus velocities by retinal ganglion cells. *Frontiers in Neuroinformatics.*, Conference Abstract: Neuroinformatics, 2008.
20. A.M. Litke, N. Bezaiff, E.J. Chichilnisky, W. Cunningham, W. Dabrowski, A.A. Grillo, M. Grivich, P. Grybos, P. Hottowy, S. Kachiguine, R.S. 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(4):1434–1440, 2004.
 21. P. McCullagh and J. Nelder. *Generalized linear models*. Chapman and Hall, London, 1989.
 22. S. McKee, G. Silvermann, and K. Nakayama. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*, 26:609–619, 1986.
 23. M. Meister, L. Lagnado, and D.A. Baylor. Concerted signaling by retinal ganglion cells. *Science*, 270:1207–1210, 1995.
 24. S. Nirenberg, S. Carcieri, A. Jacobs, and P. Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 411:698–701, 2002.
 25. L. Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15:243–262, 2004.
 26. L. Paninski, J. Pillow, and J. Lewi. Statistical models for neural encoding, decoding, and optimal stimulus design. In P. Cisek, T. Drew, and J. Kalaska, editors, *Computational Neuroscience: Progress in Brain Research*. Elsevier, 2008.
 27. V.H. Perry and A. Cowey. The ganglion cell and cone distributions in the monkey’s retina: implications for central magnification factors. *Vision Research*, 25:1795–1810, 1985.
 28. J. Pillow and L. Paninski. Model-based decoding, information estimation, and change-point detection in multi-neuron spike trains. *Under review, Neural Computation*, 2008.
 29. J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, and E.P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454:995–999, 2008.
 30. 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.
 31. R. Segev, J. Goodhouse, J. Puchalla, and M. Berry. Recording spikes from a large fraction of the ganglion cells in a retinal patch. *Nature Neuroscience*, 7:1154–1161, 2004.
 32. J. Shlens, G.D. Field, J.L. Gauthier, M.I. Grivich, D. Petrusca, A. Sher, A.M. Litke, and E.J. Chichilnisky. The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.*, 26:8254–8266, 2006.
 33. E P Simoncelli. Local analysis of visual motion. In L M Chalupa and J S Werner, editors,

- The Visual Neurosciences*, chapter 109, pages 1616–1623. MIT Press, January 2003.
34. D. Snyder and M. Miller. *Random Point Processes in Time and Space*. Springer-Verlag, 1991.
 35. A.A. Stocker and E.P. Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–585, 2006.
 36. L. Stone and P. Thompson. Human speed perception is contrast dependent. *Vision Research*, 32:1535–1549, 1992.
 37. A. Thiel, M. Greschner, C.W. Eurich, J. Ammermüller, and J. Kretzberg. Contribution of individual retinal ganglion cell responses to velocity and acceleration encoding. *J Neurophysiol*, 98(2):2285–2296, 2007.
 38. P. Thompson. Perceived rate of movement depends on contrast. *Vision Research*, 22:377–380, 1982.
 39. 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 Research*, 46:782–786, 2005.
 40. W. Truccolo, U. Eden, M. Fellows, J. Donoghue, and E. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble and extrinsic covariate effects. *Journal of Neurophysiology*, 93:1074–1089, 2005.
 41. S. Ullman. *The Interpretation of Visual Motion*. MIT Press, 1979.
 42. Y. Weiss, E. Simoncelli, and E. Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5:598–604, 2002.
 43. Andrew E. Welchman, Judith M. Lam, and Heinrich H. Bulthoff. Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences*, 105(33):12087–12092, 2008.