Optimal Local Estimates of Visual Motion in a Natural Environment

Shiva R. Sinha,[1,‡] William Bialek,[2,3,*] and Rob R. de Ruyter van Steveninck[1,†]
[1]Department of Physics, Indiana University, Bloomington, Indiana 47405, USA
[2]Joseph Henry Laboratories of Physics, and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA
[3]Initiative for the Theoretical Sciences, The Graduate Center, City University of New York, 365 Fifth Avenue, New York, New York 10016, USA

(Received 22 October 2019; revised 4 July 2020; accepted 30 November 2020; published 4 January 2021)

Many organisms use visual signals to estimate motion, and these estimates typically are biased. Here, we ask whether these biases may reflect physical rather than biological limitations. Using a camera-gyroscope system, we sample the joint distribution of images and rotational motions in a natural environment, and from this distribution we construct the optimal estimator of velocity based on local image intensities. Over most of the natural dynamic range, this estimator exhibits the biases observed in neural and behavioral responses. Thus, imputed errors in sensory processing may represent an optimal response to the physical signals sampled from the environment.

DOI: 10.1103/PhysRevLett.126.018101

What limits the reliability of perception? On one hand, the visual system is capable of counting single photons [1]. On the other hand, perceptions are error prone, as illustrated by visual illusions. A well-studied example is visual motion estimation, where both neural [2–4] and behavioral [5,6] responses are systematically biased. These systematic errors could reflect limitations of the biological hardware. But it could also be that the brain performs a computation well matched to sensory signals, based on physical data that themselves are too limited to generate a veridical estimate [7]. In bright daylight, our visual impression of the world is that everything is crisp and clear, and it is hard to imagine that physical limits are relevant. Our goal here is to test this assumption by measuring the quality of inferences that can be drawn from visual data under reasonably natural conditions, on a scale relevant to brain function.

The best estimate of a feature of interest is determined by its joint distribution with the available data. Here, we consider angular velocity as the feature, while the data are the "movies" collected by the eye or a camera. Our goal is to sample their joint distribution directly and from these samples construct the function which optimally transforms visual inputs into velocity estimates. We simplify the problem by focusing on a small patch of the visual world and on situations where motion is dominated by rigid rotations of the observer. To facilitate comparisons with a biological example, we have built a camera that replicates the geometrical parameters of the blowfly visual system but with a larger collecting area, so we can measure intensities more reliably; rotations of the camera are measured with gyroscopes.

Early models for visual motion estimation, formulated by Hassenstein and Reichardt, were based on the behavioral responses of beetles to simple motion stimuli [8,9], and models for mammalian motion estimation are generalizations of this "Reichardt correlator" [10,11]. These models predict that velocity estimates are proportional to the square of image contrast at low contrast, consistent with measurements of behavioral [5] and neural [2,3] responses in flies, even as single neurons can encode motion with a precision close to the limits set by noise in the photoreceptor array [12]. In the primate cortex [4,13] and human perception [6], similar biases are observed. It is plausible that very low contrast images provide little data about the actual velocity, and so the best estimate will be biased toward the velocity which is most likely *a priori*, and this is zero. This argument can be made rigorous, showing that biases similar to those of the Reichardt correlator are features of the optimal motion estimator in the limit of low signal-to-noise ratios [14–16].

It is intriguing that correlatorlike computations, with their systematic errors, can emerge as optimal solutions to the motion estimation problem. But, to claim that this explains the estimation errors observed for real biological systems, we need independent measurements of the signal quality. More formally, at the core of all these theoretical discussions is the joint distribution of movies and motions, so we need a direct characterization of this distribution under conditions relevant to the organism. Interest in understanding human vision has led to considerable focus on primates, but there is renewed emphasis on insects in part because of genetic tools that make it possible to trace the circuits effecting particular computations [17]. It is an

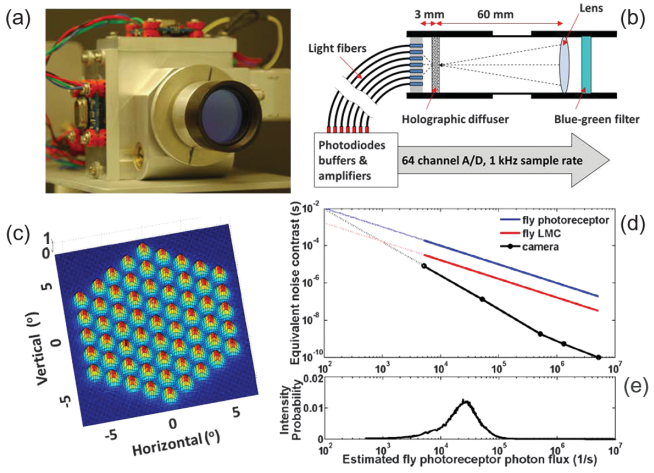

FIG. 1. Camera-gyroscope design. (a) Camera headstage with optical tube and gyroscopes (BEI Technology LCG50-00500-100 for yaw and pitch and Systron-Donner QRS130-01000-103 for roll) mounted on orthogonal surfaces, allowing measurement of yaw, pitch, and roll. (b) Schematic cross section of the camera, with optical components and hexagonal array of 61 fibers (Edmund Optics NT57-097) guiding light to 61 photodiodes (Hamamatsu S8729-10). Photodiode and gyroscope data were recorded simultaneously at 16-bit resolution (National Instruments PCI 6031). (c) Optical point spread functions and their position in the hexagonal array. (d) Equivalent contrast noise of the camera (black line), showing the much lower noise floor in the camera relative to fly photoreceptors (blue line) and their postsynaptic targets, large monopolar cells (LMCs) (red line), the putative inputs for motion estimation. (e) Probability density of light intensities in our experiment, converted to fly photoreceptor photon flux.

opportune time to ask not just what these circuits are computing, but why.

The forward-facing area of the fly eye samples the world through a hexagonal lattice of receptors with vertical row orientation and spacing $\phi_0 \sim 1.5°$, each with a Gaussian point spread function (width $\sigma \sim 0.5°$). We construct an imaging system matching these parameters (Fig. 1), with a 1000 samples per second readout of each photodetector [18].

Motion signals are due both to movement of the observer and to movement of objects in the environment. With objects far away, rotational self-motion is dominant, and our focus is on analyzing yaw in these conditions. We measure camera self-motion directly with a set of three gyroscopes aligned along the cardinal axes.

Flying the instrument in Fig. 1 along a trajectory taken by a real fly is challenging, and instead we take a half-hour walk in the woods, waving the instrument [19]. Azimuthal angular velocities are quite large, with a standard deviation of $\sim 100°/s$, so that the distribution of velocities (Fig. 2, left) covers the range experienced by flies in reasonably straight flight, though not in acrobatic flight [20]; with a correlation time of ~ 100 ms (Fig. 2, right), the fluctuations

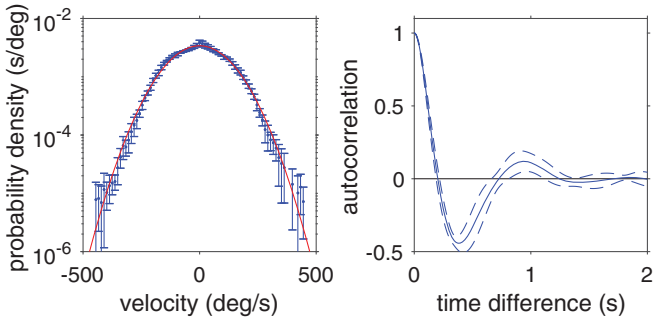

FIG. 2. Statistics of azimuthal (yaw) velocities. Left: Probability distribution of instantaneous velocities. Error bars are standard deviations across randomly chosen quarters of the half-hour walk, and the red line is a Gaussian with the same mean and variance as the data. Right: Normalized autocorrelation function. Dashed lines show $\pm$ one standard deviation across randomly chosen quarters of the data.

are a bit slower and perhaps more Gaussian than for flies. The distributions of light intensities and their (log) gradients are skewed, with roughly exponential tails (Fig. 3), as reported [21].

Our data represent samples out of the joint distribution of movies and motions, from which we must infer the structure of the optimal motion estimator. For the fly's brain, input data are the photoreceptor voltages $\{V_n(t)\}$. These are filtered, noisy versions of the light intensities in each pixel, $\{I_n(t)\}$, probabilistically related to the unknown angular velocity $v(t)$. All information about velocity is contained in the conditional distribution $P[v(t)|\{V_n(t)\}]$, and its structure constrains the computation needed to make optimal estimates [14].

As a first step, we ignore the filtering and noise in the photoreceptors to work directly with the measured intensities $\{I_n(t)\}$. This is plausibly a good approximation for bright conditions; in a fuller analysis, we can add back the receptor noise, much of which is photon shot noise.

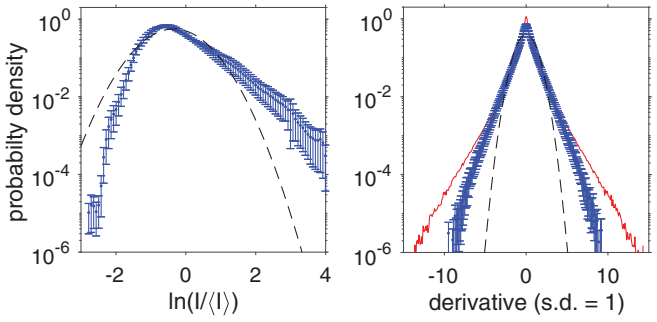

FIG. 3. Statistics of light intensities. Left: Distribution of (log) intensity, collected over all 61 pixels. Right: Distributions of spatial (blue line) and temporal (red line) derivatives of the (log) intensity. Error bars are standard deviations across randomly chosen quarters of the data; black dashed lines are Gaussians with the same mean and variance as the data.

With this approximation, all of the information available about velocity at a moment t_0 is contained in the distribution $P[v(t_0) = v|\{I_n(t)\}]$, where the notation reminds us that the optimal estimate depends on the pattern of light intensities over some window of time surrounding t_0. In the examples below, this distribution always has a single well-defined peak, so that the mean, median, and mode all are very similar. To extract a single optimal estimate, we choose the estimator that minimizes the mean square error, and this is the conditional mean

$$\hat{v}_{\rm opt}(t_0) = \int dv P[v(t_0) = v|\{I_n(t)\}]v. \tag{1}$$

The key idea is that, where a theory of optimal estimation asks for an integral over the relevant distribution, we approximate this as a sum over the measured samples, as in Monte Carlo simulations.

Equation (1) is complicated, because the best estimate of velocity depends on the dynamics of light intensities in all the pixels. In insects and in us, estimates of global rotational motion have long been thought to be built out of local motion estimates, and neurons that extract these local estimates now have been identified [17]. In the regular lattice structure of the insect visual system, there is direct evidence that motion estimation is dominated by comparisons between nearest neighbors [22], so that "local" really refers to a single pixel and its neighbors. From nearest neighbors, we can build lattice approximations to the first and second spatial derivatives in the two cardinal directions, but we find that once the first derivative is known the second derivative adds only a tiny amount of information about velocity [23], so we neglect that here. To remove dependence on the overall scene brightness, we take derivatives of the log of the intensity.

In the absence of receptor cell noise, there should be little need for temporal averaging, and we verify that the best estimates are based on averages over very short windows [24], allowing us to define a local approximation to the temporal derivatives. Information about motion in a particular direction and at a certain time is then dominated by the gradient in that direction and by local time derivatives, although motion in orthogonal directions can constitute an effective noise source (see below). Thus, the best local velocity estimate becomes

$$\hat{v}_{\rm opt} = \int dv P[v|\partial_\phi \ln I(t), \partial_t \ln I(t)]v, \tag{2}$$

where ∂_ϕ and ∂_t represent the relevant spatial and temporal derivatives, respectively, as described above. This is a map from the plane $[\partial_\phi \ln I(t), \partial_t \ln I(t)]$ to the velocity $\hat{v}_{\rm opt}$.

For a pattern moving rigidly across the array of detectors, $I(\phi, t) = f(\phi - vt)$, and we can recover the velocity by taking a ratio of derivatives:

$$\hat{v}_{\rm grad} = -\frac{\partial_t \ln I(t)}{\partial_\phi \ln I(t)}. \tag{3}$$

This "gradient model" of motion estimation gives veridical estimates in an idealized setting [26], but the derivatives and ratio make it susceptible to noise in the relation between intensities and velocities. The Reichardt correlator estimates velocity as the product of neighboring pixel intensities filtered with differing time constants [5,8,9]; with (anti)symmetrization, this approximates [27]

$$\hat{v}_{\rm cor} \propto \partial_t \ln I(t) \times \partial_\phi \ln I(t). \tag{4}$$

Notice that $\hat{v}_{\rm cor}$ is the behavior of an estimator, not the true velocity. One can see the gradient and correlation models as two limiting cases of a general optimal estimator [14], suggesting that biases of motion estimation *could* be features of optimal estimation [3,15,16,28,29]. To test the theory, we need independent evidence that the visual system operates in these limits.

The data we collect are samples from the joint distribution $P[v, \partial_\phi \ln I(t), \partial_t \ln I(t)]$ [24]. We discretize the measured gradients into bins and compute the optimal local motion estimator from Eq. (2); results are shown in Fig. 4. At large values of the spatial gradient, contours of constant velocity are approximately linear (e.g., the yellow contour follows the black line), as expected in a gradient model [Eq. (3)]. But at smaller values of the spatial gradient, contours of constant velocity bend into curves that approximate hyperbolas, which is correlatorlike [Eq. (4)]. Importantly, the bulk of the data that we collect is in the regime where curvature of the constant velocity contours is prominent, echoing observations in Ref. [30].

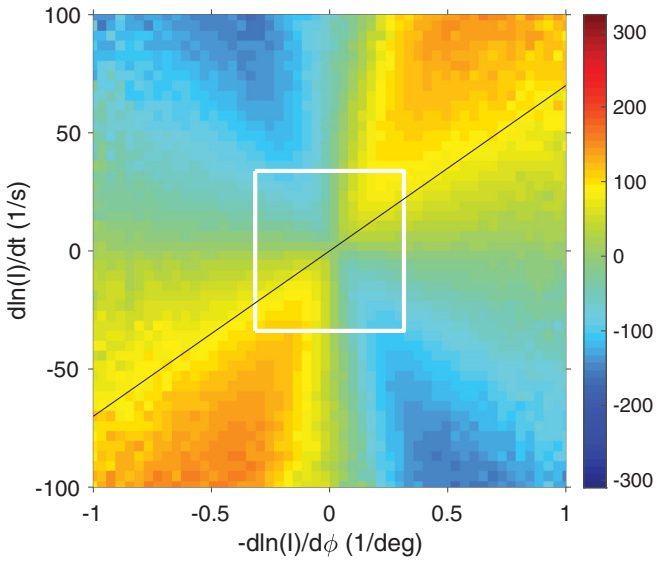

FIG. 4. Optimal estimator of velocity as a function of local spatial and temporal derivatives, from Eq. (2). The white box encloses 90% of the data. The black line is $\hat{v}_{\rm grad} = 70$ deg/s.

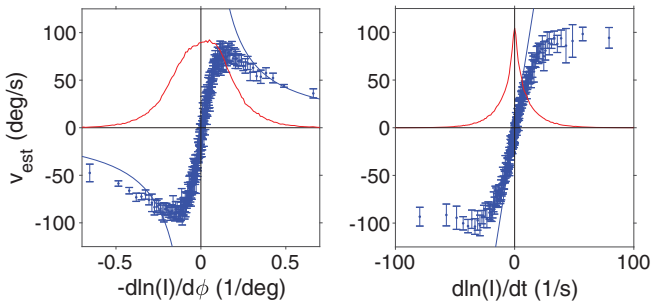

FIG. 5. Optimal motion estimates at constant temporal (left: $\partial_t \ln I = 21\ \mathrm{s}^{-1}$) or spatial (right: $-\partial_\phi \ln I = 0.13\ \mathrm{deg}^{-1}$) derivatives. Blue points are optimal estimates, slices through Fig. 4; error bars are standard deviations across random quarters of the data. Blue lines are the predictions of the gradient model [Eq. (3)], and red lines show the distribution of derivatives along the slice (scaled for clarity).

We draw attention to this in Fig. 4 by outlining a region that represents 90% of the data.

If we hold the time derivative of the local (log) light intensity fixed and vary the spatial derivative, then correlatorlike models predict that the velocity estimate will vary linearly [Eq. (4)], while gradient models predict that the estimate will vary inversely [Eq. (3)]. Both models predict that, if we hold the spatial derivative fixed, the estimate should vary linearly with the temporal derivative. In Fig. 5, we see linear dependencies at small values of the derivatives, along both slices. We see signs of the inverse dependence on the spatial derivative expected in the gradient model but only at large derivatives, in the tail of the distribution. In the bulk, the velocity estimate is linear in both spatial and temporal derivatives and, hence, quadratic in the overall image contrast, which is the essential signature of the correlator model.

Optimal estimation is always a trade-off, and systematic errors are optimal only insofar as they protect the estimate against random errors. By construction, photon noise is not a major source of randomness in our case (see Fig. 1), but intensity variations induced by motion along other directions, such as pitch, may also contribute noise. Motion-sensitive neurons in the fly encode velocities with significantly less precision in the presence of motions orthogonal to their preferred direction [31], and our data (not shown) indicate that this source of effective noise is significant for the optimal estimator as well. If noise drives the optimal estimator into the correlator regime, then the estimates themselves should be noisy, and indeed, comparing the time series predicted by our optimal estimator to the velocity measured by the gyroscope, signal-to-noise ratios are low, rising above unity only below 2 Hz, to a maximum of ~ 3.

In summary, the relationship between the local dynamics of images and movement velocities is sufficiently noisy that optimal estimates are driven into a regime where systematic errors are significant. In this regime, the optimal estimator is approximately a correlator or motion energy estimator. We have emphasized the connection to fly vision, but similar considerations apply to primate vision when the visual cortex computes motion on the scale of $\sim 1.5°$–$3°$ [32]. The idea that apparent errors of motion computation might be optimal responses to physically limited signals is an old one, both in flies [14] and in humans [15,16]. Figure 4 provides direct evidence that motion estimation in a naturalistic context really is in the regime where correlation is optimal.

Emphasizing the connections between statistical physics and inference [33], our approach replaces the integrals which appear in the theory of optimal estimation with sums over samples from the natural environment, as in Monte Carlo simulations. Along this path, there is much more to be done. The crossover between correlatorlike and gradientlike estimation should depend on the signal-to-noise ratio, which we can vary by adding back photon shot noise or focusing on periods with different typical values of image contrast. Asymmetries in the underlying distributions should lead to asymmetries in the optimal estimator [34], which are barely visible in Fig. 4 and should be connected to the separate processing of on and off signals [35]. It also will be interesting to understand the rules for optimal combination of these local estimators into wide-field motion signals. Finally, an important challenge is to explore the relation between optimal motion estimators and the structure of neural computation, quantitatively.

This work supported in part by the National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030) and Grants No. IIS-0423039 and No. PHY-1607612, and by the W. M. Keck Foundation.

*wbialek@princeton.edu
†deruyter@indiana.edu
‡Present address: Cranial & Spinal Technologies, Medtronic, Louisville Colorado, 80027.

[1] S. Hecht, S. Shlaer, and M. H. Pirenne, J. Gen. Physiol. **25**, 819 (1942); F. Rieke and D. A. Baylor, Rev. Mod. Phys. **70**, 1027 (1998).
[2] M. Egelhaaf, A. Borst, and W. Reichardt, J. Opt. Soc. Am. A **6**, 1070 (1989).
[3] R. R. de Ruyter van Steveninck, W. Bialek, M. Potters, and R. H. Carlson, *Proceedings of IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 1994* (1994), pp. 302–307, https://ieeexplore.ieee.org/document/399855.
[4] H. W. Heuer and K. H. Britten, J. Neurophysiol. **88**, 3398 (2002).
[5] W. Reichardt and T. Poggio, Q. Rev. Biophys. **9**, 311 (1976).
[6] J. P. H. van Santen and G. Sperling, J. Opt. Soc. Am. A **1**, 451 (1984).

[7] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton NJ, 2012).

[8] S. Hassenstein and W. Reichardt, Z. Naturforsch. **11B**, 513 (1956).

[9] W. Reichardt, in *Sensory Communication*, edited by W. A. Rosenblith (MIT Press, Cambridge MA, 1961), pp. 303–317.

[10] E. H. Adelson and J. R. Bergen, J. Opt. Soc. Am. A **2**, 284 (1985).

[11] J. P. H. van Santen and G. Sperling, J. Opt. Soc. Am. A **2**, 300 (1985).

[12] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, 1997).

[13] K. H. Britten, W. T. Newsome, M. N. Shadlen, S. Celebrini, and J. A. Movshon, Visual Neurosci. **13**, 87 (1996).

[14] M. Potters and W. Bialek, J. Phys. I (France) **4**, 1755 (1994).

[15] Y. Weiss, E. P. Simoncelli, and E. H. Adelson, Nat. Neurosci. **5**, 598 (2002).

[16] A. A. Stocker and E. P. Simoncelli, Nat. Neurosci. **9**, 578 (2006).

[17] J. Rister, D. Pauls, B. Schnell, C.-Y. Ting, C.-H. Lee, I. Sinakevitch, J. Morante, N. J. Strausfeld, K. Ito, and M. Heisenberg, Neuron **56**, 155 (2007); S. Takemura *et al.*, Nature (London) **500**, 175 (2013); M. S. Maisak *et al.*, Nature (London) **500**, 212 (2013); Y. W. Fisher, J. C. Leong, K. Sporar, M. D. Ketkar, D. M. Gohl, T. R. Clandinin, and M. Silies, Curr. Biol. **25**, 3178 (2015).

[18] To spatially calibrate the optical channels [Fig. 1(c)], a moving dot was scanned in a raster pattern on a Tektronix 608 oscilloscope. Temporal response, gain, and noise [Fig. 1(d)] characteristics were calibrated by generating a white noise sequence on a high-intensity LED.

[19] S. R. S. took the walk in Dunn's woods, a wooded area on the campus of Indiana University Bloomington. We chose a sunny, cloud-free summer day, around noon, for the brightest conditions possible.

[20] C. Schilstra and J. H. van Hateren, J. Exp. Biol. **202**, 1481 (1999), https://jeb.biologists.org/content/202/11/1481.short.

[21] D. L. Ruderman and W. Bialek, Phys. Rev. Lett. **73**, 814 (1994).

[22] E. Buchner, Biol. Cybern. **24**, 85 (1976); B. Pick and E. Buchner, J. Comp. Physiol. **134**, 45 (1979); S. Roy and R. de Ruyter van Steveninck, J. Vis. **16**, 8 (2016).

[23] Concretely, we select points in the plane in Fig. 4 and compute the extra information that the second derivative conveys about the velocity when the first spatial and temporal derivatives are fixed. The result is ∼0.02 bits. We will give a fuller analysis of these small effects in a longer paper.

[24] We smooth the gradients in time to improve the quality of estimates, minimizing $\langle|\hat{v}-v|^2\rangle$. The optimal $\tau\sim 10$ ms, close to the fastest transient responses of motion-sensitive neurons in the blowfly [3,25]. The optimal τ is longer if we add back the effects of photon shot noise.

[25] R. R. de Ruyter van Steveninck, W. H. Zaagman, and H. A. K. Mastebroek, Biol. Cybern. **54**, 223 (1986).

[26] J. O. Limb and J. A. Murphy, Comput. Graph. Image Proc. **4**, 311 (1975).

[27] The classical correlator model multiplies the image intensity at one point, or the corresponding receptor signal, by a filtered version of the signal at a neighboring receptor and then (anti)symmetrizes [8,9]. For a linear array

$$\hat{v}_{\rm cor} \propto \sum_n I_n(t)\int d\tau f(\tau) I_{n+1}(t-\tau) - \sum_n I_{n+1}(t)\int d\tau f(\tau) I_n(t-\tau). \quad (5)$$

In the limit that the integration time of the filter is small, we have

$$\hat{v}_{\rm cor} \propto \sum_n \frac{\partial I_n(t)}{\partial t}\{[I_{n+1}(t)-I_n(t)]+[I_n(t)-I_{n-1}(t)]\} = \sum_n [\partial_t I(\phi,t)\times\partial_\phi I(\phi,t)]|_{\phi=\phi_n}, \quad (6)$$

where ∂_ϕ is the lattice derivative.

[28] W. Bialek and R. de Ruyter van Steveninck, arXiv:q-bio/0505003.

[29] J. Burge and W. S. Geisler, Nat. Commun. **6**, 7900 (2015).

[30] R. O. Dror, D. C. O'Carroll, and S. B. Laughlin, J. Opt. Soc. Am. A **18**, 241 (2001).

[31] S. Roy, S. R. Sinha, and R. de Ruyter van Steveninck, J. Neurosci. **35**, 6481 (2015).

[32] K. H. Britten and H. W. Heuer, J. Neurosci. **19**, 5074 (1999).

[33] *From Statistical Physics to Statistical Inference and Back*, edited by P. Grassberger and J.-P. Nadal (Kluwer, Dordrecht, 1994).

[34] J. E. Fitzgerald, A. Y. Katsov, T. R. Clandinin, and M. J. Schnitzer, Proc. Natl. Acad. Sci. U.S.A. **108**, 12909 (2011).

[35] R. Behnia, D. A. Clark, A. Carter, T. R. Clandinin, and C. Desplan, Nature (London) **512**, 427 (2014).