

Nonlinear Circuits for Naturalistic Visual Motion Estimation - Supplementary Appendices

James E. Fitzgerald¹ and Damon A. Clark^{2,3}

¹Center for Brain Science, Harvard University, Cambridge, MA 02138

²Department of Molecular, Cellular, and Developmental Biology and

³Department of Physics, Yale University, New Haven, CT 06511

Appendix Figure Legends

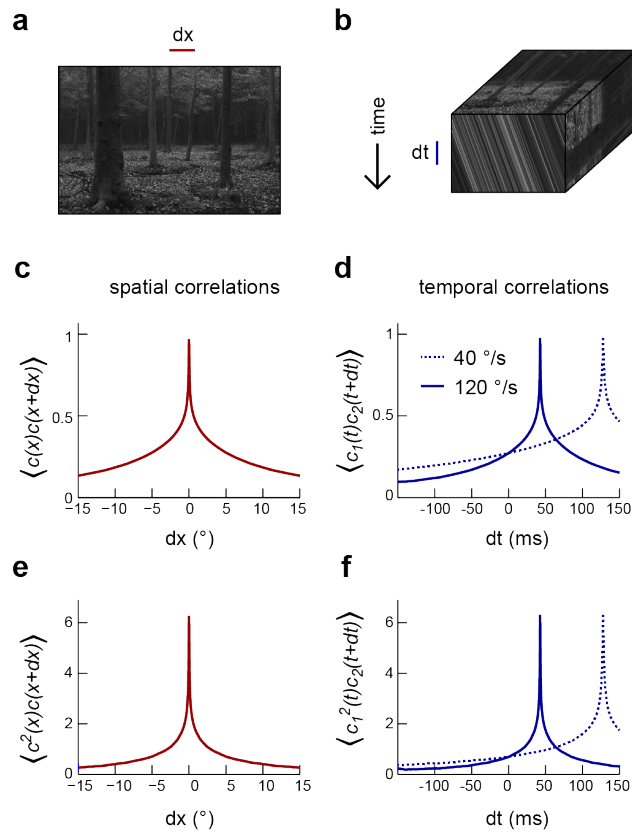
Appendix Figure 1: Motion transforms spatial correlations into temporal correlations. (a) An example natural image [1]. (b) When a natural image (*top face*) moves to the right, streaks in space-time (*front face*) indicate the direction and speed of the motion. Alternatively, motion influences the temporal correlation structure of visual signals (*side face*). (c) Second-order correlation function between pairs of spatially separated contrast signals (across the natural image ensemble [1]). (d) For constant velocity motion, the temporal correlation function between a pair of spatially separated points is shifted and stretched relative to the spatial correlation function. We separated the two points by *Drosophila*'s photoreceptor spacing (5.1°). (e) Example third-order spatial correlation function involving two points in space. (f) As with pairwise correlations, higher-order temporal correlations between spatially separated visual signals are shifted and stretched (relative to higher-order spatial correlation functions) in a manner that indicates the speed and direction of motion.

Appendix Figure 2: Correlations in binarized natural images. (a) We transformed each image in the van Hateren natural image database [1] with several binarizing nonlinearities. To implement the simplest binarizing nonlinearity, we set all pixels to +1 or -1 depending on whether that pixel exceeded or fell below the median intensity in the image. For the nonlinearity with two steps, the thresholds were at the 25th and 75th intensity percentiles. For the nonlinearity with three steps, the thresholds were at the 25th, 50th, and 75th intensity percentiles. When a pixel intensity exactly equaled a threshold, we considered its value below threshold. Binary nonlinearities with a larger number of steps produced grainier images that indicate a spatial decorrelation of the transformed image. (b) We computed second-order spatial correlation functions across the nonlinearly transformed natural image ensemble. This confirmed that each step in the binarizing nonlinearity further decorrelated the image ensemble. (c) In addition to decreasing the spatial extent of correlations, a larger number of transitions also the performance of the front-end nonlinearity model.

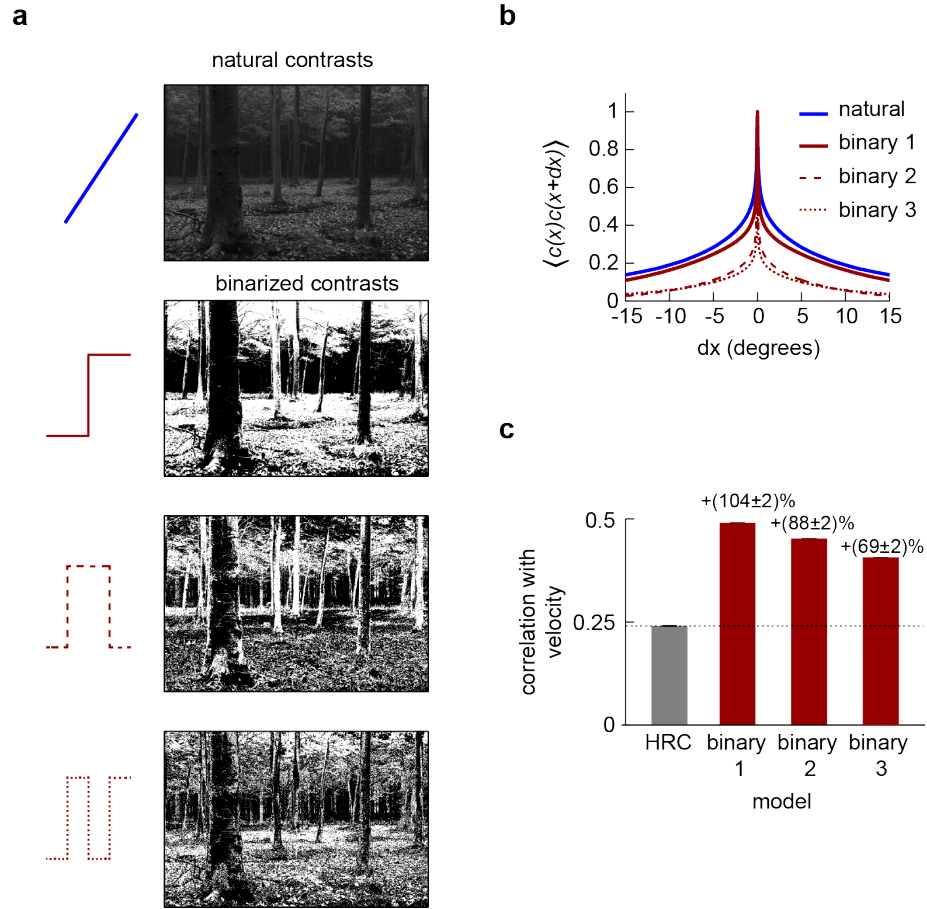
Appendix Figure 3: Accuracy of the weighted 4-quadrant model across model parameters. (a)-(b) We computed the correlation coefficient between the velocity and the response of the weighted 4-quadrant model for all possible sets of model parameters. Since rescaling the weight vector does not affect the correlation coefficient, we assumed that all model parameters satisfy $\sum_{a,b \in \{+, -\}} (w_{ab}^{(Q)})^2 = 1$. We color-coded each set of model parameters by its accuracy and projected the parameter space onto various subspaces. (a) We first examined the quadrant basis by projecting onto the $\{(--), (-+)\}$ (*left*) and $\{(+ -), (++)\}$ (*right*) subspaces. (b) We next examined the correlational basis by projecting onto the $\{\text{even} = 2, \text{odd}\}$ (*left*) and $\{\text{odd*}, \text{even} > 2\}$ (*right*) subspaces. These project into different linear combinations of the original quadrant weightings. One of the projections is the pure HRC ($\text{even} = 2$), while the other projections contain only odd correlations, of two different types (odd and odd*), or only even correlations of order greater than 2 ($\text{even} > 2$). This projection shows that accurate weighted 4-quadrant models always put positive weight into 2-point correlations and negative weight into odd-order correlations. Note that the glider responses predicted by the weighted 4-quadrant model mirror this pattern (**Fig. 3d**). See **Appendix VIII** for a more detailed interpretation of these plots.

Appendix Figure 4: The weighted 4-quadrant model in the basis of principal components. (a) We computed the covariance matrix of quadrant responses across the simulated ensemble of naturalistic motions. (b) The eigenvectors of the covariance matrix are called principal components (PCs). Signals from the $(++)$ and $(+-)$ quadrants primarily comprised the first two principal components, whereas the $(-+)$ and $(--)$ components comprised the third and fourth principal components. (c) The first two principal components accounted for the vast majority of the weighted 4-quadrant model’s response variance. (d) Each member of the ensemble of naturalistic motions comprised a velocity and a natural image, and both components contributed variance to the model response. Although the first two principal components accounted for most of the variance, little of that variance was associated with the velocity of motion. Instead, the third and fourth principal components best aided motion estimation, because they contributed the vast majority of the velocity-associated variance. See **Appendix IX** for a mathematical treatment of these points.

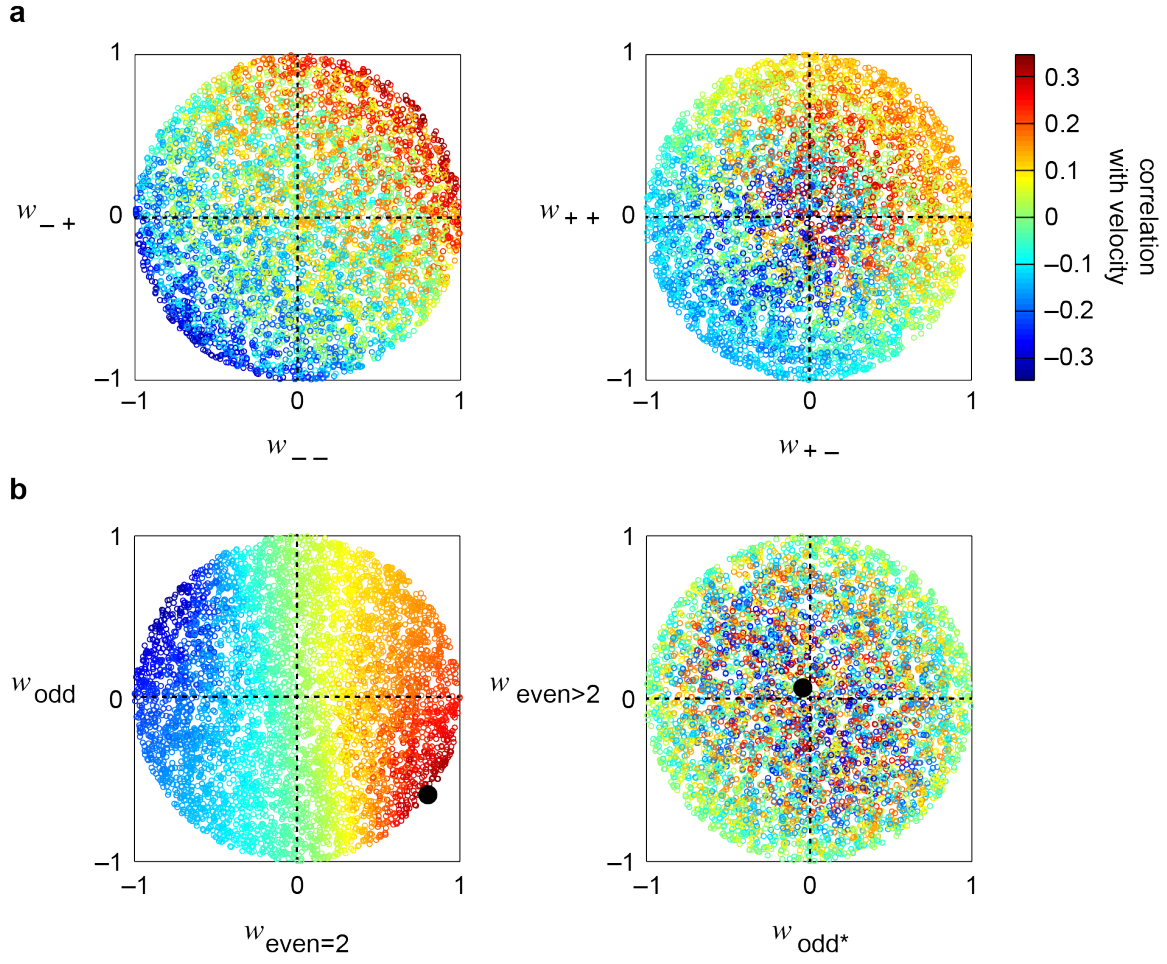
Appendix Figures



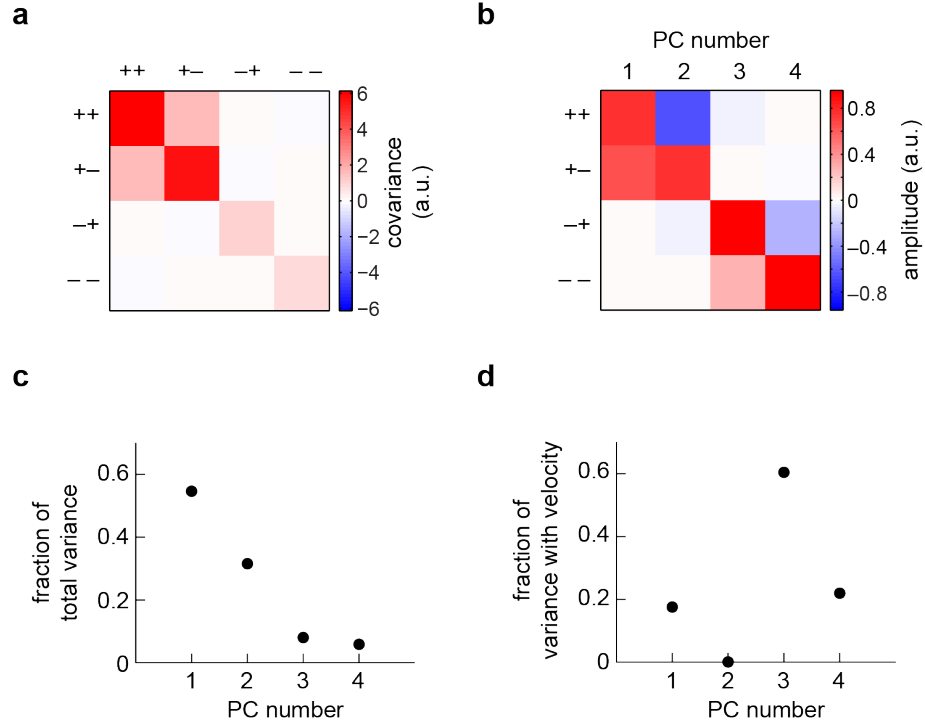
Appendix Figure 1: Motion transforms spatial correlations into temporal correlations.



Appendix Figure 2: Correlations in binarized natural images.



Appendix Figure 3: Accuracy of the weighted 4-quadrant model across model parameters.



Appendix Figure 4: The weighted 4-quadrant model in the basis of principal components.

Appendix I: Visual signatures of motion

The pattern of light that stimulates the retina encodes information about the relative motion between the retina and its visual environment. The manner in which this information is encoded depends on the geometry of the photoreceptor array, the statistics of self-motion, and the statistics of the visual environment. The principal goal of this paper is to illustrate several ways that the brain’s nonlinear processing of visual motion signals might be tuned to reflect specific features of the natural visual environment. We thus begin by enumerating some computational signatures of visual motion in natural environments, thereby exposing a diversity of stimulus features that visual system nonlinearities might aim to extract.

In the real world, animals encounter visual environments that are intricately structured and far from random (**App. Fig. 1a**) [1, 2, 3]. When an animal rotates with constant angular velocity through the environment, the spatiotemporal response profile of the photoreceptor array encodes the velocity of self-motion through the slope of oriented streaks in space-time (*front face*, **App. Fig. 1b**) [4]. Thus, a visual system with a dense array of noiseless photoreceptors could extract the angular velocity of an arbitrary image by computing the ratio of temporal and spatial derivatives [5]. The statistics of the image ensemble become relevant once multiple interpretations of the sensory world become logically consistent with the photoreceptor data. In particular, the optimal motion estimator depends on the statistics of the image ensemble when photoreceptors have noise [5, 6], and a nonzero spacing between photoreceptors introduces ambiguity via aliasing [5]. In these cases, the animal can use prior information regarding the sensory environment and its motion in order to weigh the plausibility of each sensory interpretation.

Full field motion transforms spatial features (*top face*, **App. Fig. 1b**) into temporal features (*side face*, **App. Fig. 1b**) in a manner that depends upon the velocity of motion. Consequently, one can also think about the visual signatures of motion in terms of spatiotemporal correlations between photoreceptors. The luminance contrast encoded by the i^{th} photoreceptor is $C_i(t) = (I_i(t) - I_0)/I_0$, where $I_i(t)$ is the luminance intensity seen by the i^{th} photoreceptor at time t and I_0 is the average luminance intensity over the visual field. Thus, the average contrast is zero, and the simplest correlation function corresponds to the product of two spatially separated contrast signals. Measured over an ensemble of natural images, this 2-point correlation function had a global maximum at zero spatial offset (**App. Fig. 1c**). Consequently, the velocity of motion is encoded by the peak of the temporal cross-correlation function between two neighboring photoreceptors, which occurs at the temporal offset that equals the photoreceptor spacing (5.1° for *Drosophila*) divided by the velocity of motion (**App. Fig. 1d**). Natural images also contain many higher-order correlations [2, 3]. For instance, the nonzero skewness of natural images implies that the third-order correlation that multiplies the contrast at one point with the squared contrast at a neighboring point also has a peak at zero spatial offset (**App. Fig. 1e**). Correspondingly, the peak of the temporal 3-point correlation function between neighboring photoreceptors encodes the velocity of motion (**App. Fig. 1f**). This argument generalizes to n^{th} -order correlation functions when the ensemble of natural images has a nonzero n^{th} moment. Note that this argument does not necessarily imply that a motion estimator would benefit from the incorporation of all nonzero correlation functions, because the velocity signals provided by one correlation function could be redundant with those provided by others.

Importantly, photoreceptor correlation functions also encode velocity information away from their peaks. For example, the velocity of motion influences the widths of the temporal cross-correlation functions between pairs of photoreceptors (**App Figs. 1d,f**). To see this, note that the values of the temporal correlation functions at zero temporal offsets are velocity independent, whereas the peak locations are closer to zero for larger speeds (**App Figs. 1d,f**). This implies a more rapid falloff for higher speeds. This fundamental effect occurs because nearby points are more correlated in natural environments and photoreceptors rapidly survey distant points when the speed of motion is high.

The description above illustrates how visual motion becomes encoded in photoreceptor correlations. A central goal of research in visual motion estimation is to understand how neural circuits invert (or decode) that encoding of velocity. Just as a broad class of functions can be represented as a power series, a broad class of motion estimators can be represented as a Volterra series [6, 7]. Each term in the Volterra series can be interpreted as a multipoint correlator that decodes velocity information from a specific correlation function [6]. For example, the HRC and the motion energy model are 2-point correlators that decode velocity from 2-point correlations, whereas the Bayes optimal motion estimator capitalizes on a wider variety of correlation functions [6]. Because multipoint correlators relate intuitively to measurable properties of the image ensemble, we will find that decomposing a motion estimator in terms of multipoint correlators is often illuminating. Moreover, we will use multipoint correlators as a common basis to compare the computations performed by mechanistically distinct models.

Appendix II: Accuracy of 2-point correlators

In this section we derive an expression for the accuracy of a general 2-point correlator in terms of the statistics of naturalistic motion.

We consider a general 2-point correlator that temporally correlates visual signals from the spatial points i and j . Mathematically, this estimator has the form

$$v_e^{(2)}(t) = \int dt_1 \int dt_2 k_{i,j}^{(2)}(t_1, t_2) V_i(t - t_1) V_j(t - t_2), \quad (1)$$

where the 2-point kernel, $k_{i,j}^{(2)}(t_1, t_2)$, defines the correlator by specifying how each 2-point correlation contributes to the motion estimate. We model the response of the i^{th} photoreceptor as

$$V_i(t) = \int d\tau T(\tau) \int d\theta M(\theta - \theta_i) c \left(\theta - \int_0^{t-\tau} dt' v(t') \right) \quad (2)$$

where T is a temporal integration kernel, M is the photoreceptor's spatial acceptance profile, θ_i is the location of the i^{th} photoreceptor, $c(\theta)$ is the spatial contrast pattern of the visual world, and $v(t)$ is the time-dependent velocity. This formula simplifies to the formula in the **Methods** when $v(t)$ is time-independent. If T is an invertible linear filter, then a more convenient representation of the photoreceptor signals is

$$U_i(t) = \mathcal{C} \left(\theta_i - \int_0^t dt' v(t') \right) \quad (3)$$

where $U_i = T^{-1} * V_i$, $\mathcal{C} = M * c$, and $*$ is the convolution operator [5]. We can rewrite the 2-point correlator in this representation as

$$v_e^{(2)}(t) = \int dt_1 \int dt_2 \kappa_{i,j}^{(2)}(t_1, t_2) U_i(t - t_1) U_j(t - t_2), \quad (4)$$

where

$$\kappa_{i,j}^{(2)}(t_1, t_2) \equiv \int dt_3 T(t_3) \int dt_4 T(t_4) k_{i,j}^{(2)}(t_1 - t_3, t_2 - t_4) \quad (5)$$

is the 2-point kernel that converts correlations in the U variables to a velocity estimate.

Recall that we quantify the performance of visual motion estimators based on the mean squared error between the true and estimated velocities

$$\epsilon \equiv \langle (v_e^{(2)}(t) - v(t))^2 \rangle = \sigma_v^2 - 2 \langle v(t) v_e^{(2)}(t) \rangle + \langle (v_e^{(2)}(t))^2 \rangle, \quad (6)$$

where $\sigma_v = 90^\circ/s$ is the standard deviation of the velocity distribution. For estimators that are scaled to minimize their mean squared error (**Methods**), this formula can be rewritten as

$$\epsilon = \sigma_v^2 (1 - r^2), \quad (7)$$

where

$$r = \frac{\langle v(t) v_e^{(2)}(t) \rangle}{\sqrt{\langle (v(t))^2 \rangle \langle (v_e^{(2)}(t))^2 \rangle}} = \frac{\langle v(t) v_e^{(2)}(t) \rangle}{\sigma_v \sqrt{\langle (v_e^{(2)}(t))^2 \rangle}}. \quad (8)$$

is the correlation coefficient between the estimated and true velocities. Thus, minimizing the mean squared error is mathematically equivalent to maximizing the correlation coefficient if all motion estimators are correctly scaled. We find the correlation coefficient to be a more intuitive error metric than the mean squared error, so many of our results will be presented in terms of correlation coefficients.

The numerator of the correlation coefficient is determined by the second-order statistics of the image ensemble,

$$\begin{aligned} \langle v(t) v_e^{(2)}(t) \rangle &= \int dt_1 \int dt_2 \kappa_{i,j}^{(2)}(t_1, t_2) \langle v(t) U_i(t - t_1) U_j(t - t_2) \rangle \\ &= \int dt_1 \int dt_2 \kappa_{i,j}^{(2)}(t_1, t_2) \left\langle v(t) \mathcal{C}^{(2)} \left(\Delta_{ij} + \int_{t_1}^{t_2} dt' v(t') \right) \right\rangle_v, \end{aligned} \quad (9)$$

where Δ_{ij} is the angular separation between the i^{th} and j^{th} photoreceptors, and

$$\mathcal{C}^{(2)}(\Delta) \equiv \langle \mathcal{C}(x) \mathcal{C}(x + \Delta) \rangle_{\mathcal{C}} \quad (10)$$

is the 2-point correlation function over the ensemble of spatially filtered natural scenes. Note that $\mathcal{C}^{(2)}$ is independent of x because reasonable image ensembles are translationally

invariant. Also note that the 2-point correlation function of filtered natural images is related to the correlation function of unfiltered images by

$$\mathcal{C}^{(2)}(\Delta) = \int dx' M(x') \int dx'' M(x'') \langle c(x - x') c(x + \Delta - x'') \rangle = ((M * M) * C^{(2)}) (\Delta), \quad (11)$$

where $C^{(2)}(\Delta)$ is the correlation function of unfiltered images, and we've assumed that M is a symmetric function. We model M as Gaussian with FWHM of 5.7° , so $M * M$ is also Gaussian with FWHM of $\sqrt{2} \times 5.7^\circ = 8.1^\circ$.

On the other hand, the denominator of the correlation coefficient is determined by fourth-order statistics of the image ensemble,

$$\begin{aligned} \langle (v_e^{(2)}(t))^2 \rangle &= \int dt_1 \int dt_2 \int dt_3 \int dt_4 \kappa_{i,j}^{(2)}(t_1, t_2) \kappa_{i,j}^{(2)}(t_3, t_4) \\ &\quad \times \langle U_i(t - t_1) U_j(t - t_2) U_i(t - t_3) U_j(t - t_4) \rangle \\ &= \int dt_1 \int dt_2 \int dt_3 \int dt_4 \kappa_{i,j}^{(2)}(t_1, t_2) \kappa_{i,j}^{(2)}(t_3, t_4) \\ &\quad \times \left\langle \mathcal{C}^{(4)} \left(\Delta_{ij} + \int_{t_1}^{t_2} dt' v(t'), \int_{t_1}^{t_3} dt' v(t'), \Delta_{ij} + \int_{t_1}^{t_4} dt' v(t') \right) \right\rangle_v, \end{aligned} \quad (12)$$

where

$$\mathcal{C}^{(4)}(\Delta_1, \Delta_2, \Delta_3) = \langle \mathcal{C}(x) \mathcal{C}(x + \Delta_1) \mathcal{C}(x + \Delta_2) \mathcal{C}(x + \Delta_3) \rangle_c \quad (13)$$

is the 4-point correlation function of the ensemble of filtered natural images. Notice that the second argument of $\mathcal{C}^{(4)}$ in Eq. (12) lacks the additive factor of Δ_{ij} because $U_i(t - t_1)$ and $U_i(t - t_3)$ correspond to the same point in space. As above, $\mathcal{C}^{(4)}$ is related to the unfiltered 4-point correlation function through a four-fold application of the photoreceptor spatial acceptance filter.

The preceding analysis shows that only the second-order and fourth-order statistics of the natural image ensemble contribute to the correlation coefficient between an arbitrary 2-point correlator and the true velocity. The same quantities also determine the mean squared error. Thus, the second-order and fourth-order statistics of the image ensembles are the critical determinants of a 2-point correlator's motion estimation accuracy. Note that both the HRC and the motion energy model fall into this important class of visual motion estimators, so our analysis is also important for understanding visual motion estimation by vertebrates.

Appendix III: Motion estimation without spatial correlations - the role of kurtosis on the accuracy of 2-point correlators.

In this section, we apply the results of **Appendix II** to the special case of normally distributed velocities and spatially uncorrelated image ensembles. This calculation reveals an

important role for kurtosis in motion estimation, and we discuss how nonlinearities in the early visual system could cope with highly kurtotic naturalistic inputs.

In this section, we assume that the velocity is time-independent (*i.e.* $v(t) = v$) and normally distributed

$$P_v(v) = \frac{1}{\sqrt{2\pi\sigma_v^2}} e^{-v^2/(2\sigma_v^2)}. \quad (14)$$

We also assume that the image ensemble is spatially uncorrelated. By this, we mean that the luminance contrast at each point in space is statistically independent of the luminance contrast at all other points in space. Thus, the second-order correlation function is

$$\mathcal{C}^{(2)}(\Delta) = \sigma_C^2 \delta(\Delta), \quad (15)$$

where σ_C is the standard deviation of the luminance contrast, and $\delta(\Delta)$ is the Dirac delta-function. The fourth-order correlation function is

$$\begin{aligned} \mathcal{C}^{(4)}(\Delta_1, \Delta_2, \Delta_3) = & \kappa_4 \sigma_C^4 \delta(\Delta_1) \delta(\Delta_2) \delta(\Delta_3) \\ & + \sigma_C^4 (\delta(\Delta_1) \delta(\Delta_2 - \Delta_3) + \delta(\Delta_2) \delta(\Delta_1 - \Delta_3) + \delta(\Delta_3) \delta(\Delta_1 - \Delta_2)) \end{aligned} \quad (16)$$

where κ_4 is the excess kurtosis of the contrast distribution. The excess kurtosis is zero for normally distributed contrasts. It can either be positive or negative for other contrast distribution. Note that we define the *kurtosis* of a probability distribution to be its fourth central moment normalized by the square of its second central moment. Thus, the kurtosis of a normal distribution is 3. We caution readers that some other sources use “kurtosis” to refer to the excess kurtosis.

With these assumptions, the signal term represented by Eq (9) is

$$\langle v(t) v_e^{(2)}(t) \rangle = \frac{\sigma_C^2 \Delta_{ij}}{\sqrt{2\pi\sigma_v^2}} \int dt_1 \int dt_2 \kappa_{i,j}^{(2)}(t_1, t_2) \frac{e^{-\Delta_{ij}^2/(2\sigma_v^2(t_2-t_1)^2)}}{(t_2 - t_1)|t_2 - t_1|}, \quad (17)$$

and the noise term represented by Eq. (12) is

$$\begin{aligned} \langle (v_e^{(2)}(t))^2 \rangle = & \frac{\sigma_C^4}{\Delta_{ij} \sqrt{2\pi\sigma_v^2}} \int dt_1 \int dt_2 \int dt_3 \int dt_4 \kappa_{i,j}^{(2)}(t_1, t_2) \kappa_{i,j}^{(2)}(t_3, t_4) \\ & \times \left(e^{-\Delta_{ij}^2/(2\sigma_v^2(t_1-t_4)^2)} \delta((t_1 - t_4) - (t_3 - t_2)) + e^{-\Delta_{ij}^2/(2\sigma_v^2(t_1-t_2)^2)} \delta((t_1 - t_2) - (t_3 - t_4)) \right. \\ & \left. + \kappa_4 \frac{|t_1 - t_2|}{\Delta_{ij}} e^{-\Delta_{ij}^2/(2\sigma_v^2(t_1-t_2)^2)} \delta(t_3 - t_1) \delta(t_4 - t_2) \right), \end{aligned} \quad (18)$$

where we’ve assumed that the 2-point correlator is mirror anti-symmetric,

$$\kappa_{i,j}^{(2)}(t_1, t_2) = -\kappa_{i,j}^{(2)}(t_2, t_1), \quad (19)$$

in order to ignore contributions from static signals. This mirror-symmetry assumption holds for the HRC and the motion energy model. Since the denominator of the correlation coefficient is set by $\sqrt{\langle (v_e^{(2)}(t))^2 \rangle}$, both the signal and the noise are proportional to σ_C^2 . Thus,

the only remaining dependence on the image ensemble is through the excess kurtosis. Note that

$$\frac{d \langle (v_e^{(2)}(t))^2 \rangle}{d\kappa_4} = \frac{\sigma_C^4}{\Delta_{ij}^2 \sqrt{2\pi\sigma_v^2}} \int dt_1 \int dt_2 (\kappa_{ij}^{(2)}(t_1, t_2))^2 |t_1 - t_2| e^{-\Delta_{ij}^2/(2\sigma_v^2(t_1-t_2)^2)} > 0. \quad (20)$$

Thus, the correlation coefficient is maximized by making κ_4 as small as possible.

In conclusion, if the image ensemble is spatially uncorrelated (at second and fourth-order), then the image ensemble only affects the correlation coefficient between the velocity and a 2-point correlator through its kurtosis. The best accuracy is achieved when the kurtosis is minimized. In reality, the assumption that the image ensemble is spatially uncorrelated is clearly wrong. Natural images are strongly correlated, and even if they weren't, they'd become correlated once they are filtered by the photoreceptors' spatial acceptance filter. Nevertheless, **Fig. 2e** empirically shows that introducing several front-end nonlinearities that decrease the kurtosis also improve the accuracy of naturalistic motion estimation. Thus, kurtosis provides a useful guide for the design of neuronal nonlinearities. On the other hand, **Figs. 2d,e** demonstrate that it's too simplistic to assume that the kurtosis is the only relevant factor for the accuracy of a 2-point correlator. As we'll discuss in the next section, spatial correlations in the image ensemble also affect the accuracy of 2-point correlators.

Appendix IV: The HRC benefits from spatially correlated input signals

When we applied a contrast-equalizing or binarizing nonlinearity to naturalistic inputs before evaluating the HRC, we found that both nonlinearities substantially improved the accuracy of the HRC (**Fig. 2e**). Interestingly, contrast equalization improved the accuracy of the HRC more than binarization (**Fig. 2e**), even though it produced outputs with greater kurtosis. The reason for this is that natural images are correlated (**App. Fig. 1**), and the accuracy of the HRC over a general image ensemble depends on the ensemble's spatial correlation structure (**Appendix II**). Binarization attenuated spatial correlations more strongly than contrast equalization over the natural image ensemble (**Fig. 2-supp. 1**), which leads us to hypothesize that correlations present in the natural image ensemble might benefit the HRC's performance. In **Appendix V** we will provide theoretical support for this idea. Here we begin with a less mathematical argument that also supports our hypothesis.

A comparison between the estimation performance of binarizing and equalizing front-end nonlinearities was complicated by the fact that the models produced outputs that differed in both their point statistics and their correlation structures. To gain more direct insight into how spatial correlations affect motion estimation performance, it would be helpful to compare front-end nonlinearity models that differ *only* through their output correlation structures. We implemented this comparison using a family of binarizing front-end nonlinearities that undergo multiple steps between +1 and -1 (**App. Fig. 2a**). Although these nonlinearities are not physiologically realistic, they are conceptually useful because they each produced a stimulus ensemble that minimized the kurtosis yet achieved distinct correlation structures (**App. Fig. 2b**). These nonlinearities thus allow us to assess directly whether spatial

decorrelation of inputs degrades the motion estimation performance of the HRC. We found that each binarizing front-end nonlinearity model outperformed the original HRC (**App. Fig. 2c**). However, we found that the magnitude of the improvement decreased with the number of steps (**App. Fig. 2c**). Since spatial cross-correlations also decreased as a function of the number of steps (**App. Fig. 2b**), these results support our hypothesis that the correlations present in natural visual inputs aid the functionality of the standard HRC.

The HRC correlates two signals that are offset in space and differentially delayed in time. One intuition that researchers often apply to this computation is that the correlation operation effectively detects times when two signals that are offset in space and time are equal. However, a motion estimator that strictly obeyed this intuition would be agnostic to the spatial correlation structures present in the input signals, and our results show that the HRC is not (see also **Appendix V**). Instead, the HRC also generates motion signals when its two input channels are imperfectly aligned, and these signals depend strongly on the correlation structure of the inputs (**App. Fig. 1d**). Our results thus show that the HRC’s ability to detect imperfect coincidences contributes significantly to its performance as a motion estimator, as was suggested intuitively in **Appendix 1**.

Appendix V: Motion estimation with Gaussian image statistics - the role of spatial correlations on the accuracy of 2-point correlators

In this section, we apply the results of **Appendix II** to the special case of normally distributed velocities and normally distributed image ensembles. This model formalizes how spatial correlations in the natural world affect the accuracy of motion estimation by 2-point correlators and shows how spatial decorrelation can adversely affect estimation accuracy. For example, we’ll show that the simplest HRC is unable to extract motion signals from high frequency components of the image ensemble, yet those components still lead to variability in the motion estimator. Thus, this HRC works best when the image ensemble is correlated in a manner that avoids high frequency components in the signal, and spatial low-pass filtering at the photoreceptor level can help to eliminate the high-frequency image components that hurt the HRC’s accuracy.

Here we use the same velocity distribution that we used in **Appendix III** (*i.e.* Eq. (14)). However, we now allow the two point correlation function to have arbitrary structure

$$\mathcal{C}^{(2)}(\Delta) = \sum_{k=0}^{\infty} S_k \cos(k\Delta), \quad (21)$$

where S_k are the Fourier coefficients for $\mathcal{C}^{(2)}(\Delta)$, and we have noted that the image ensemble is 2π -periodic. Note that S_k is called the power spectrum of the image ensemble, and uncorrelated ensembles correspond to the special case where $S_k = \text{constant}$. With these assumptions

$$\left\langle v(t) \mathcal{C}^{(2)} \left(\Delta_{ij} + \int_{t_1}^{t_2} dt' v(t') \right) \right\rangle_v = \sum_{k=0}^{\infty} S_k \langle v \cos(k(\Delta_{i,j} + v(t_2 - t_1))) \rangle_v. \quad (22)$$

By evaluating the integral, we find that this velocity expectation is

$$\langle v \cos(k(\Delta_{ij} + v(t_2 - t_1))) \rangle_v = k(t_1 - t_2) \sigma_v^2 \sin(k\Delta_{ij}) e^{-\frac{1}{2}k^2(t_2 - t_1)^2 \sigma_v^2}. \quad (23)$$

Thus, if we define

$$\gamma_k = k \sigma_v^2 \sin(k\Delta_{ij}) \int dt_1 \int dt_2 \kappa_{ij}^{(2)}(t_1, t_2) (t_1 - t_2) e^{-\frac{1}{2}k^2(t_2 - t_1)^2 \sigma_v^2}, \quad (24)$$

then

$$\langle v(t) v_e^{(2)}(t) \rangle = \sum_{k=0}^{\infty} \gamma_k S_k. \quad (25)$$

Each frequency component of the image ensemble linearly contributes to the correlation between the 2-point correlator's response and the velocity. The weight of each frequency component is determined by the structure of the 2-point correlator and the width of the velocity distribution.

We compute the fourth-order moment of the image ensemble using Wick's theorem for Gaussian moments, which says

$$\begin{aligned} \langle \mathcal{C}(x_1) \mathcal{C}(x_2) \mathcal{C}(x_3) \mathcal{C}(x_4) \rangle &= \langle \mathcal{C}(x_1) \mathcal{C}(x_2) \rangle \langle \mathcal{C}(x_3) \mathcal{C}(x_4) \rangle + \langle \mathcal{C}(x_1) \mathcal{C}(x_3) \rangle \langle \mathcal{C}(x_2) \mathcal{C}(x_4) \rangle \\ &\quad + \langle \mathcal{C}(x_1) \mathcal{C}(x_4) \rangle \langle \mathcal{C}(x_2) \mathcal{C}(x_3) \rangle. \end{aligned} \quad (26)$$

This immediately implies that

$$\begin{aligned} \mathcal{C}^{(4)}(\Delta_1, \Delta_2, \Delta_3) &= \mathcal{C}^{(2)}(\Delta_1) \mathcal{C}^{(2)}(\Delta_3 - \Delta_2) + \mathcal{C}^{(2)}(\Delta_2) \mathcal{C}^{(2)}(\Delta_3 - \Delta_1) \\ &\quad + \mathcal{C}^{(2)}(\Delta_3) \mathcal{C}^{(2)}(\Delta_2 - \Delta_1). \end{aligned} \quad (27)$$

Once again, it's convenient to rewrite this expression in the Fourier domain

$$\begin{aligned} \mathcal{C}^{(4)}(\Delta_1, \Delta_2, \Delta_3) &= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} S_{k_1} S_{k_2} \left(\cos(k_1 \Delta_1) \cos(k_2 (\Delta_3 - \Delta_2)) \right. \\ &\quad \left. + \cos(k_1 \Delta_2) \cos(k_2 (\Delta_3 - \Delta_1)) + \cos(k_1 \Delta_3) \cos(k_2 (\Delta_2 - \Delta_1)) \right) \end{aligned} \quad (28)$$

With these assumptions

$$\begin{aligned} &\left\langle \mathcal{C}^{(4)}\left(\Delta_{ij} + \int_{t_1}^{t_2} dt' v(t'), \int_{t_1}^{t_3} dt' v(t'), \Delta_{ij} + \int_{t_1}^{t_4} dt' v(t')\right) \right\rangle_v \\ &= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} S_{k_1} S_{k_2} \left\langle \cos(k_1 (\Delta_{ij} + v(t_2 - t_1))) \cos(k_2 (\Delta_{ij} + v(t_4 - t_3))) \right. \\ &\quad \left. + \cos(k_1 v(t_3 - t_1)) \cos(k_2 v(t_4 - t_2)) \right. \\ &\quad \left. + \cos(k_1 (\Delta_{ij} + v(t_4 - t_1))) \cos(k_2 (\Delta_{ij} + v(t_2 - t_3))) \right\rangle_v. \end{aligned} \quad (29)$$

We evaluate the expectations over velocity by noting that each has the form

$$\begin{aligned} \langle \cos(k_1(\Delta + v\delta_1)) \cos(k_2(\Delta + v\delta_2)) \rangle_v = & \frac{1}{2} e^{-\frac{1}{2}(k_1\delta_1 + k_2\delta_2)^2 \sigma_v^2} \left(\cos(\Delta(k_1 + k_2)) \right. \\ & \left. + e^{2k_1k_2\delta_1\delta_2\sigma_v^2} \cos(\Delta(k_1 - k_2)) \right) \end{aligned} \quad (30)$$

for some spatial offset Δ and temporal offsets $\{\delta_1, \delta_2\}$. Thus, if we define

$$\begin{aligned} \Gamma_{k_1k_2} = & \int dt_1 \int dt_2 \kappa_{ij}^{(2)}(t_1, t_2) \int dt_3 \int dt_4 \kappa_{ij}^{(2)}(t_3, t_4) \\ & \left(\frac{1}{2} e^{-\frac{1}{2}(k_1(t_2-t_1) + k_2(t_4-t_3))^2 \sigma_v^2} \left(\cos(\Delta_{ij}(k_1 + k_2)) + e^{2k_1k_2(t_2-t_1)(t_4-t_3)\sigma_v^2} \cos(\Delta_{ij}(k_1 - k_2)) \right) \right. \\ & + \frac{1}{2} e^{-\frac{1}{2}(k_1(t_4-t_1) + k_2(t_2-t_3))^2 \sigma_v^2} \left(\cos(\Delta_{ij}(k_1 + k_2)) + e^{2k_1k_2(t_4-t_1)(t_2-t_3)\sigma_v^2} \cos(\Delta_{ij}(k_1 - k_2)) \right) \\ & \left. + \frac{1}{2} e^{-\frac{1}{2}(k_1(t_3-t_1) + k_2(t_4-t_2))^2 \sigma_v^2} \left(1 + e^{2k_1k_2(t_3-t_1)(t_4-t_2)\sigma_v^2} \right) \right), \end{aligned} \quad (31)$$

then

$$\langle (v_e^{(2)}(t))^2 \rangle = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \Gamma_{k_1k_2} S_{k_1} S_{k_2}. \quad (32)$$

Power spectrum components contribute to the 2-point correlator's variance quadratically.

Putting these pieces together, the expected squared error achieved by a 2-point correlator is a quadratic function of the power spectrum

$$\epsilon = \sigma_v^2 - 2 \sum_{k=0}^{\infty} \gamma_k S_k + \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \Gamma_{k_1k_2} S_{k_1} S_{k_2}. \quad (33)$$

We're interested to know whether spatial correlations can enhance the accuracy of the 2-point correlator. This will be the case unless a uniform power spectrum minimizes ϵ . Note that every physically meaningful power spectrum is non-negative

$$S_k \geq 0. \quad (34)$$

Thus, the minimum of ϵ either occurs at an extremum point or on the boundary of admissible solutions. If the minimum occurs on the boundary, then a subset of the S_k are exactly equal to zero. In particular, the power spectrum would not be constant, which implies that the image ensemble would be spatially correlated. At an extremum point, we must find

$$0 = \frac{\partial \epsilon}{\partial S_k} = -2\gamma_k + 2 \sum_{k'=0}^{\infty} \Gamma_{kk'} S_{k'} \quad (35)$$

for every k . A uniform power spectrum can only satisfy this condition if

$$\gamma_k = \beta \sum_{k'=0}^{\infty} \Gamma_{kk'}, \quad (36)$$

where $\beta > 0$ is the (constant) value of each power spectrum component. This is generally not the case, so correlations exist that would help typical 2-point correlators.

For example, the simplest HRC, which replaces the low-pass and high-pass filters with pure time delays, is

$$\tilde{R} = A (U_1(t - \tau)U_2(t) - U_1(t)U_2(t - \tau)), \quad (37)$$

where A is a constant with units of $^\circ/s$. For this model,

$$\kappa_{1,2}^{(2)}(t_1, t_2) = A (\delta(t_1 - \tau)\delta(t_2) - \delta(t_1)\delta(t_2 - \tau)). \quad (38)$$

Substituting this expression into the above formulas, we find

$$\gamma_k = 2Ak\tau\sigma_v^2 \sin(k\Delta_0) e^{-\frac{1}{2}k^2\tau^2\sigma_v^2} \quad (39)$$

and

$$\begin{aligned} \Gamma_{k_1 k_2} = & A^2 \left(3 \sin(k_1 \Delta_0) \sin(k_2 \Delta_0) (e^{-\frac{1}{2}(k_1 - k_2)^2 \tau^2 \sigma_v^2} - e^{-\frac{1}{2}(k_1 + k_2)^2 \tau^2 \sigma_v^2}) \right. \\ & \left. + (1 - \cos(k_1 \Delta_0) \cos(k_2 \Delta_0)) (2 - e^{-\frac{1}{2}(k_1 - k_2)^2 \tau^2 \sigma_v^2} - e^{-\frac{1}{2}(k_1 + k_2)^2 \tau^2 \sigma_v^2}) \right), \end{aligned} \quad (40)$$

where Δ_0 is the spacing between adjacent photoreceptors. Note that

$$\lim_{k \rightarrow \infty} \gamma_k = 0. \quad (41)$$

On the other hand,

$$\lim_{k_2 \gg k_1} \Gamma_{k_1 k_2} = 2A^2 (1 - \cos(k_1 \Delta_0) \cos(k_2 \Delta_0)). \quad (42)$$

This does not approach zero, even for large values of k_1 . Therefore, $\sum_{k'=0}^{\infty} \Gamma_{kk'}$ diverges and $\gamma_k \neq \beta \sum_{k'=0}^{\infty} \Gamma_{kk'}$. In this model, high frequency components lack signal but contribute noise. It's helpful if these frequency components are absent from the image ensemble. Future work should more fully investigate the role of spatial correlations in naturalistic motion estimation.

Appendix VI: Front-end nonlinearities give the HRC access to higher-order correlations

The response of the front-end nonlinearity model to a 3-point glider stimulus is determined by the higher-order correlations that it detects in the stimulus. Furthermore, we argued in **Appendix I** and **Fig. 1i** that higher-order correlations can contribute to the accuracy of visual motion estimators. We now describe how front-end nonlinearities provide pair-correlation mechanisms with access to certain types of higher-order correlations.

We suppose that the front-end nonlinearity, denoted h , has a power series expansion:

$$h(x) = \sum_{n=0}^{\infty} h_n x^n. \quad (43)$$

Then the cross-correlation function between two non-linearly transformed input streams, denoted y_1 and y_2 , is

$$\langle y_1(t)y_2(t+\tau) \rangle = \langle h(V_1(t))h(V_2(t+\tau)) \rangle = \sum_{n,m=0}^{\infty} h_n h_m V_1^n(t) V_2^m(t+\tau), \quad (44)$$

where V_1 and V_2 are linear photoreceptor signals. This substitution explicitly demonstrates that the front-end nonlinear transformation enables pair correlation mechanisms to incorporate higher-order correlations of the form $V_1^n(t)V_2^m(t+\tau)$. The choice of nonlinearity specifies the expansion coefficients, h_n , which in turn determines the pattern of higher-order correlations that the pair correlator incorporates into its velocity estimate. For example, sensitivity to odd-ordered correlations demands that h_n be large for some even values of n . These expansion coefficients would manifest themselves in the structure of the front-end nonlinearity as asymmetries between positive and negative contrasts, but strong asymmetries were not needed to eliminate kurtosis in natural image ensembles (**Fig. 2c**). Inversely, one could use this equation to determine whether a set expansion coefficients exist that would implement a desired series of multipoint correlators. The preceding argument implies that strongly asymmetric front-end nonlinearities would be needed to account for the 3-point glider responses.

Appendix VII: Expansion of the weighted 4-quadrant model

In this **Appendix**, we rewrite the weighted 4-quadrant in a basis that isolates its dependence on 2-point correlations, on higher-even-ordered correlations, and on two types of odd-ordered correlations. In **Appendix VIII**, we'll discuss the motion estimation performance of the weighted 4-quadrant model in this basis in order to gain insight into why performance-optimized weighted 4-quadrant models also predict 3-point glider responses that resemble *Drosophila* behavior.

The weighted 4-quadrant model supposes that the input signals are segregated into four separate streams:

$$\begin{aligned} Q_{++} &= [f * V_1]_+[g * V_2]_+ - [g * V_1]_+[f * V_2]_+ \\ Q_{+-} &= [f * V_1]_+[g * V_2]_- - [g * V_1]_-[f * V_2]_+ \\ Q_{-+} &= [f * V_1]_-[g * V_2]_+ - [g * V_1]_+[f * V_2]_- \\ Q_{--} &= [f * V_1]_-[g * V_2]_- - [g * V_1]_-[f * V_2]_- \end{aligned} \quad (45)$$

where Q_{ab} denotes the (ab) quadrant for $a, b \in \{+, -\}$, $[x]_+$ is x for $x > 0$ and is zero otherwise, and $[x]_-$ is x for $x < 0$ and is zero otherwise. The HRC is equal to

$$R = Q_{++} + Q_{+-} + Q_{-+} + Q_{--}. \quad (46)$$

More generally, we suppose that *Drosophila* could estimate motion as any linear combination of these signals, and we define the weighted 4-quadrant model as

$$Q = w_{++}^{(Q)} Q_{++} + w_{+-}^{(Q)} Q_{+-} + w_{-+}^{(Q)} Q_{-+} + w_{--}^{(Q)} Q_{--}, \quad (47)$$

where $w_{++}^{(Q)}$, $w_{+-}^{(Q)}$, $w_{-+}^{(Q)}$, and $w_{--}^{(Q)}$ are linear weighting coefficients that specify the computation performed by the model. Since this section, and the next two, focus entirely on the weighted 4-quadrant model, we simplify notation by dropping the superscript (Q) .

The weighted 4-quadrant model can be rewritten in an alternate form that facilitates an understanding of how various correlation types contribute to its motion estimates. We begin by noting that

$$[x]_+ = \frac{x}{2}(1 + \text{sgn}(x)), \quad (48)$$

$$[x]_- = \frac{x}{2}(1 - \text{sgn}(x)), \quad (49)$$

where $\text{sgn}(x)$ is +1 for positive arguments and -1 for negative arguments. We thus see that

$$\begin{aligned} Q_{ab} &= [f * V_1]_a [g * V_2]_b - [g * V_1]_b [f * V_2]_a \\ &= \frac{(f * V_1)(g * V_2)}{4} (1 + a \text{sgn}(f * V_1) + b \text{sgn}(g * V_2) + ab \text{sgn}(f * V_1) \text{sgn}(g * V_2)) \\ &\quad - \frac{(g * V_1)(f * V_2)}{4} (1 + b \text{sgn}(g * V_1) + a \text{sgn}(f * V_2) + ab \text{sgn}(g * V_1) \text{sgn}(f * V_2)). \end{aligned} \quad (50)$$

Therefore, the complete weighted 4-quadrant model is

$$\begin{aligned} Q &= \frac{w_{++} + w_{+-} + w_{-+} + w_{--}}{4} ((f * V_1)(g * V_2) - (g * V_1)(f * V_2)) \\ &\quad + \frac{w_{++} + w_{+-} - w_{-+} - w_{--}}{4} ((f * V_1) \text{sgn}(f * V_1)(g * V_2) - (g * V_1)(f * V_2) \text{sgn}(f * V_2)) \\ &\quad + \frac{w_{++} - w_{+-} + w_{-+} - w_{--}}{4} ((f * V_1)(g * V_2) \text{sgn}(g * V_2) - (g * V_1) \text{sgn}(g * V_1)(f * V_2)) \\ &\quad + \frac{w_{++} - w_{+-} - w_{-+} + w_{--}}{4} ((f * V_1) \text{sgn}(f * V_1)(g * V_2) \text{sgn}(g * V_2) \\ &\quad \quad - (g * V_1) \text{sgn}(g * V_1)(f * V_2) \text{sgn}(f * V_2)). \end{aligned} \quad (51)$$

This expression for the weighted 4-quadrant model groups the four weighting coefficients into four alternate terms. The first term is proportional to a standard HRC, which computes second-order correlations. We denote its associated coefficient as

$$w_{\text{even}=2} = \frac{1}{4} (w_{++} + w_{+-} + w_{-+} + w_{--}). \quad (52)$$

The second and third terms invert sign and retain magnitude under contrast inversion. Therefore, they only compute odd-ordered correlations:

$$w_{\text{odd}} = \frac{1}{4} (w_{++} + w_{+-} - w_{-+} - w_{--}), \quad (53)$$

$$w_{\text{odd}*} = \frac{1}{4} (w_{++} - w_{+-} + w_{-+} - w_{--}). \quad (54)$$

The fourth term is unaffected by contrast inversion. Thus, it only computes even-ordered correlations. We'll soon see that the lowest-order contribution from this term is fourth-order, so we denote its coefficient as

$$w_{\text{even}>2} = \frac{1}{4} (w_{++} - w_{+-} - w_{-+} + w_{--}). \quad (55)$$

These four coefficients define the correlational basis considered in **Fig. 3-supp. 1**. For example, note that **Fig. 3-supp. 1a** shows the transformation defined by Eqs. (52)-(55).

Because $\text{sgn}(x)$ is a non-analytic function, it is still somewhat opaque how the weighted 4-quadrant model relates to specific higher-order correlations in the visual stimulus. We thus rewrite $\text{sgn}(x)$ as the limit of an analytic function:

$$\text{sgn}(x) = \lim_{\beta \rightarrow \infty} \text{erf}(\beta x) \quad (56)$$

where

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dy e^{-y^2} \quad (57)$$

is the Gauss error function. The Gauss error function is entire, which means that it has a power series expansion for any value x . Also note that since real biological nonlinearities are not infinitely sharp, a more realistic weighted 4-quadrant model would fix β at a finite value. We thus consider the follow approximation,

$$\text{sgn}(x) \approx \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n (\beta x)^{2n+1}}{n!(2n+1)} = \frac{2}{\sqrt{\pi}} \left(\beta x - \frac{(\beta x)^3}{3} + O((\beta x)^5) \right), \quad (58)$$

where $\beta \in (0, \infty)$. Although high-order terms might not be negligible in this expansion, the contributions of low-order correlations to visual motion estimation are set by low-order terms. In particular, the contributions of second, third, and fourth-order correlations to the weighted four quadrant model are determined by the leading terms in the expansion,

$$\begin{aligned} F = & w_{\text{even}=2} ((f * V_1)(g * V_2) - (g * V_1)(f * V_2)) \\ & + w_{\text{odd}} \frac{2\beta}{\sqrt{\pi}} ((f * V_1)^2(g * V_2) - (g * V_1)(f * V_2)^2) \\ & + w_{\text{odd}*} \frac{2\beta}{\sqrt{\pi}} ((f * V_1)(g * V_2)^2 - (g * V_1)^2(f * V_2)) \\ & + w_{\text{even}>2} \frac{4\beta^2}{\pi} ((f * V_1)^2(g * V_2)^2 - (g * V_1)^2(f * V_2)^2) + O(\beta^3 V^5). \end{aligned} \quad (59)$$

Thus, the third-order term associated with w_{odd} squares the low-pass filtered signal and might help to account for light-dark asymmetries in the low-pass filtered signal. The third-order term associated with $w_{\text{odd}*}$ squares the high-pass filtered signal. Finally, note that this formula confirms that the lowest-order term associated with $w_{\text{even}>2}$ is fourth-order.

Appendix VIII: The weighted 4-quadrant model improves motion estimation with odd-ordered correlations

In the main text we quantitatively characterized the weighted 4-quadrant model by discussing its accuracy given various subsets of the four quadrants (**Fig. 3c**). Here we consider

the performance of the weighted 4-quadrant model in the correlational basis defined in **Appendix VII** and **Fig. 3-suppl. 1a**. These results lead to a simple interpretation of the computation performed by performance optimized weighted 4-quadrant models.

Models that oriented all of their weight along the even=2 axis outperformed models that focused their weight along any other correlational axis (**Fig. 3-suppl. 1b**). This reinforces the foremost importance of second-order correlations for motion estimation. In isolation, odd-ordered correlations were weaker predictors of motion than higher even-ordered correlations (**Fig. 3-suppl. 1b**). Nevertheless, the odd class best complemented the HRC, and the full accuracy of the weighted 4-quadrant model was obtained by linearly combining the even=2 and odd correlation classes (*best 2 bar*, **Fig. 3-suppl. 1b**). This result suggests that the weighted 4-quadrant model has two relevant dimensions. In particular, accurate models combine an HRC with odd-ordered correlations that account for statistical light-dark asymmetries in the HRC’s low-pass filtered branch. Although the even>2 correlation class predicts motion in isolation, our results indicate that its motion signals are largely redundant with those better captured by the even=2 and odd correlation classes.

Since the weighted 4-quadrant model only has four parameters, it’s possible to exhaustively study its parameter dependence. We have in mind models that are correctly scaled, in which case the mean squared error is determined by the correlation coefficient (**Appendix II**). Since the value of the correlation coefficient is unchanged when all four weighting coefficients are scaled by the same positive factor, it suffices to consider weighting coefficients drawn from the 3-sphere, such that $w_{++}^2 + w_{+-}^2 + w_{-+}^2 + w_{--}^2 = 1$. Because the 3-sphere has a finite volume, we were able to densely sample the correlation coefficient for all parameter values (**App. Fig. 3**). This function has one global maximum, corresponding to the optimal weight vector discussed in the main text. Its global minimum occurs on the polar opposite side of the 3-sphere, where the weighted 4-quadrant model is strongly anti-correlated with the velocity. More generally, correlation coefficients corresponding to model parameters on opposite poles of the 3-sphere always have the same magnitude and opposite sign. Both models explain the same amount of variance about the velocity, and they become equivalent after they’re correctly scaled. Thus, we henceforth focus our discussion on the hemisphere where the correlation coefficient was positive.

Weighted 4-quadrant models were most accurate when w_{-+} and w_{--} were large (**App. Fig. 3a, left**) and w_{++} and w_{+-} were small (**App. Fig. 3a, right**). In the correlational basis, the HRC is the model with maximum weight in $w_{\text{even}=2}$ and with zero weight in w_{odd} , $w_{\text{odd}*}$, and $w_{\text{even}>2}$. Thus, this basis makes it easy to compare the accuracy of the HRC to other weighted 4-quadrant models (**App. Fig. 3b**). Furthermore, this basis clearly sorts the weighted 4-quadrant models according to their accuracy and confirms that the accuracy of a weighted 4-quadrant model is largely determined by $w_{\text{even}=2}$ and w_{odd} (**App. Fig. 3b, left**). Higher even-ordered correlations and odd-ordered correlations that account for light-dark asymmetries in the high-pass filtered visual signals did not contribute prominently to the accuracy of the weighted 4-quadrant model (**App. Fig. 3b, right**). Interestingly, **App. Fig. 3a** shows that there is a diversity of ways to combine the four quadrants in order to improve the accuracy of the HRC, which translates into a diversity of correlational responses (**App. Fig. 3b**). Similarly, the HRC is only one of many models that achieve a comparable level of accuracy. Every other motion estimator that achieves the HRC’s performance level incorporates higher-order correlations into its estimate.

Appendix IX: The weighted 4-quadrant model in the basis of principal components

Principal components analysis is a popular method to reduce the dimensionality of neural population recordings. In this section, we conceptualize the four quadrants as a small neural population and study how each principal component accounts for variance in the system and contributes to motion estimation. We show that most of the weighted 4-quadrant model’s variance is due to two of the four principal components. Interestingly, most of this variance is not velocity-related, and we show that the two low-variance principal components are the ones that dominate motion estimation.

We began by directly applying principal component analysis to the weighted 4-quadrant model. We computed the 4×4 covariance matrix of the four quadrants over the ensemble of simulated motions (**App. Fig. 4a**). The eigenvectors of the covariance matrix are called the principal components (**App. Fig. 4b**), and the associated eigenvalues specify the amount of variance accounted for by each principal component (**App. Fig. 4c**). We found that the first two principal components accounted for 86.3% of the variance, whereas the third and fourth principal components each contributed about 7% of the variance (**App. Fig. 4c**). The high-variance eigenvectors roughly corresponded to a sum and a difference of the $(++)$ and $(+-)$ quadrants, whereas the low-variance principal components roughly corresponded to a sum and a difference of the $(-+)$ and $(--)$ quadrants (**App. Fig. 4b**). The $(--)$ and $(-+)$ quadrants best facilitated motion estimation (**Fig. 3c**). Thus, the low-variance principal components were most important for motion estimation.

This result is counter to one’s usual intuition, but it is a straightforward consequence of the mathematics of linear regression and principal component analysis. We want to linearly combine the principal component signals to best predict the velocity:

$$\beta = \operatorname{argmin} \left\langle (v - \beta^T x)^2 \right\rangle, \quad (60)$$

where β is a four-dimensional column vector of weights, v denotes the velocity, the superscript T denotes the matrix transpose, and x is the 4-vector of principal component signals. The solution to this problem is well-known from the theory of linear regression:

$$\beta = M^{-1}U, \quad (61)$$

where $M_{ij} = \langle x_i x_j \rangle$ is the covariance matrix of the predictors, and $U_i = \langle v x_i \rangle$ is the covariance of each predictor with the velocity. In practice, we estimate these expectations from the empirical data, and principal components are uncorrelated over the naturalistic motion ensemble by construction

$$M_{ij} = \lambda_i \delta_{ij}, \quad (62)$$

where λ_i is the variance associated with i^{th} principal component, and δ_{ij} is the Kronecker δ -function. Thus,

$$\beta_i = \frac{\langle v x_i \rangle}{\langle x_i^2 \rangle} = \frac{\sigma_v \sqrt{\lambda_i} r_i}{\lambda_i} = \frac{\sigma_v r_i}{\sqrt{\lambda_i}} \quad (63)$$

where σ_v is the standard deviation of the velocity signal, and r_i is the correlation coefficient between the velocity and the i^{th} principal component.

It is also easy to calculate the correlation coefficient between the true velocity and the estimated velocity. First note that

$$\langle v\beta^T x \rangle = \beta^T \langle vx \rangle = \sum_i \frac{\sigma_v r_i}{\sqrt{\lambda_i}} \sigma_v \sqrt{\lambda_i} r_i = \sigma_v^2 \sum_i r_i^2 \quad (64)$$

$$\langle (\beta^T x)^2 \rangle = \sum_{i,j} \beta_i \beta_j \langle x_i x_j \rangle = \sum_i \frac{\sigma_v^2 r_i^2}{\lambda_i} \lambda_i = \sigma_v^2 \sum_i r_i^2. \quad (65)$$

Thus the squared of the correlation coefficient between the true and estimated velocities is

$$r^2 = \frac{(\langle v\beta^T x \rangle)^2}{\langle v^2 \rangle \langle (\beta^T x)^2 \rangle} = \sum_i r_i^2. \quad (66)$$

Because the principal components are uncorrelated, each contributes independently to the motion estimator's accuracy. The amount that each principal component contributes to the estimation accuracy is determined by its correlation with the velocity, and all dependence on the total amount of variance associated with the principal component has dropped out entirely. These conclusions are also true when we look at the squared error directly

$$\epsilon = \langle (v - \beta^T x)^2 \rangle = \sigma_v^2 + \langle (\beta^T x)^2 \rangle - 2\langle v\beta^T x \rangle = \sigma_v^2 \left(1 - \sum_i r_i^2 \right). \quad (67)$$

As would be expected from this formula, the third and fourth principal components account for much more of the velocity-associated variance than the first and second principal components (**App. Fig. 4d**). Nevertheless, the first principal component does account for a significant portion of the velocity-associated variance (**App. Fig. 4d**), so the basis of principal components does not fully reveal the structure that was apparent in the correlational basis (**Appendix VIII**).

Appendix X: Novel use of low-order signatures for motion estimation

The non-multiplicative nonlinearity model (**Fig. 4a**) relaxed the assumption that *Drosophila*'s motion estimator multiplies its inputs and substantially improved the accuracy of visual motion estimation (**Fig. 4e**). Surprisingly, the non-multiplicative nonlinearity model slightly outperformed the HRC when we parameterized it as a second-order polynomial (**Fig. 4-suppl. 2**). This indicates that there are useful low-order correlations that the HRC neglects. In this section, we will explain how visual motion estimators can sometimes productively incorporate computational signatures that do not nonlinearly combine signals across space.

This section considers computational signatures that clash harshly with our usual intuition for visual motion estimation, and we need to unpack *how* the motion estimator in **Fig.**

4-suppl. 2 works before we can understand *why* it works. The observed improvement results from a linear combination of the HRC

$$R = (f * V_1)(g * V_2) - (g * V_1)(f * V_2) \quad (68)$$

with a linear transformation of the photoreceptor signals

$$L = g * V_1 - g * V_2. \quad (69)$$

We thus must consider the motion estimator

$$v_e^{(\text{auto})} = \beta_R R + \beta_L L, \quad (70)$$

where β_R and β_L are the weighting coefficients that minimize the mean-squared error. Note that L linearly combines signals from multiple points in space. Like the HRC, it is mirror anti-symmetric:

$$\{V_1(t), V_2(t)\} \mapsto \{V_2(t), V_1(t)\} \implies L \mapsto -L. \quad (71)$$

It is useful to take a detour to abstractly consider how motion estimation performance depends on the joint statistics of R , L , and the velocity of motion, v . All three of these quantities are zero mean. We denote their variances as

$$\sigma_R^2 = \langle R^2 \rangle, \quad \sigma_L^2 = \langle L^2 \rangle, \quad \sigma_v^2 = \langle v^2 \rangle \quad (72)$$

and their cross-correlation coefficients as

$$r^{(R)} = \frac{\langle vR \rangle}{\sigma_v \sigma_R}, \quad r^{(L)} = \frac{\langle vL \rangle}{\sigma_v \sigma_L}, \quad c^{(RL)} = \frac{\langle RA \rangle}{\sigma_R \sigma_L}. \quad (73)$$

The optimal weighting coefficients are determined by these quantities (see Eq. (61)):

$$\beta_R = \frac{\sigma_v (r^{(R)} - c^{(RL)} r^{(L)})}{\sigma_R (1 - (c^{(RL)})^2)}, \quad \beta_L = \frac{\sigma_v (r^{(L)} - c^{(RL)} r^{(R)})}{\sigma_L (1 - (c^{(RL)})^2)}; \quad (74)$$

as is the correlation coefficient between the true velocity and $v_e^{(\text{auto})}$:

$$r^{(\text{auto})} = \sqrt{\frac{(r^{(R)})^2 + (r^{(L)})^2 - 2c^{(RL)} r^{(R)} r^{(L)}}{1 - (c^{(RL)})^2}}. \quad (75)$$

Across the simulated ensemble of naturalistic motion we empirically found that $r^{(R)} \approx 0.24$, $r^{(L)} \approx -0.0017$, and $c^{(RL)} \approx -0.28$. Thus, we note that $|r^{(L)}| \ll |r^{(R)}|$ and approximate the correlation coefficient as

$$\frac{r^{(\text{auto})}}{r^{(R)}} \approx \sqrt{\frac{1}{1 - (c^{(RL)})^2}}. \quad (76)$$

Thus, we expect the inclusion of the linear term L to improve the accuracy of motion estimation by about 4.3% (compare to **Fig. 4-suppl. 2**). Interested readers can find a complete derivation of these equations in Section V of the Supplemental Materials for [8].

With this machinery in hand, we can start to understand the utility of the linear term. First, note that this term was only weakly correlated with the velocity across the simulated ensemble of motions. Furthermore, the correlation would have been *exactly* zero if $\langle v(g * V_1) \rangle$ had been equal to $\langle v(g * V_2) \rangle$, as would have been the case for an ensemble that was perfectly translationally invariant. So the small correlation we observed between L and v is nothing more than residual noise resulting from a finitely sized data sample that did not explicitly enforce translation invariance. Nevertheless, it's critical to realize that Eq. (76) treated $r^{(L)}$ as if it *were* zero, yet it still managed to account for the results of **Fig. 4-suppl. 2**. Thus, this residual sampling noise has nothing to do with the improvements offered by the hybrid estimator. As intuitively expected, the linear term is completely uncorrelated with the velocity of motion.

Eq. (76) suggests that a linear term, which is itself uncorrelated with the velocity of motion, can nevertheless help velocity estimation. However, this improvement demands that it be combined with another motion estimator that: (i) is correlated with the velocity (*i.e.* $r^{(R)} \neq 0$); and (ii) is correlated with the linear term (*i.e.* $c^{(RL)} \neq 0$). Our numerical results indicate that the HRC is an example of such a motion estimator. The HRC obviously satisfies the first condition. To examine the second condition, we note that correlation between the HRC and the linear term is nonzero if and only if

$$\begin{aligned} \langle RL \rangle = & \langle (f * V_1)(g * V_1)(g * V_2) \rangle + \langle (g * V_1)(f * V_2)(g * V_2) \rangle \\ & - \langle (f * V_1)(g * V_2)^2 \rangle - \langle (g * V_1)^2(f * V_2) \rangle \end{aligned} \quad (77)$$

is nonzero. As long as the image ensemble is light-dark asymmetric, there are no symmetry principles that force this number to vanish for a general choice of f and g . Our numerical results show that the associated correlation coefficient is far from zero for natural inputs and our choices of filters. Fundamentally, this correlation can be nonzero because the HRC's response depends on the pattern that is moving, as does the linear response. Because image-induced variability is partially shared between the HRC and the linear term, the latter can help to eliminate image-induced noise from the HRC, thereby improving the motion estimate.

Although our results indicate that a linear term can improve local motion estimation, its benefits do not sum over space. In particular, imagine an ensemble of elementary motion detectors that combine a local HRC and a local linear estimator:

$$v_{e,i}^{(\text{auto})} = \beta_R ((f * V_i)(g * V_{i+1}) - (g * V_i)(f * V_{i+1})) + \beta_L (g * V_i - g * V_{i+1}), \quad (78)$$

where i indexes the first point in space surveyed by the i^{th} local estimator. A whole field motion percept could be found by averaging these local motion signals over space

$$v_e^{(\text{auto})} = \frac{1}{N} \sum_{i=1}^N v_{e,i}^{(\text{auto})}, \quad (79)$$

where N denotes the total number of local motion detectors. However, the second term in the linear estimator at point i cancels the first term in the linear term at point $i + 1$. Thus, spatial averaging eliminates most of the dependence on the linear term

$$v_e^{(\text{auto})} = \frac{\beta_R}{N} \sum_{i=1}^N ((f * V_i)(g * V_{i+1}) - (g * V_i)(f * V_{i+1})) + \frac{\beta_L}{N} (g * V_1 - g * V_{N+1}). \quad (80)$$

All that remains of the linear term is a boundary term that depends on photoreceptor activity at the edges of the visual field. Furthermore, the magnitude of this contribution decreases with N . Thus, linear estimators have little utility for full-field motion estimation. Nevertheless, it's conceivable that such terms could play a role in *Drosophila*'s motion estimation circuit, because the same elementary motion detector is thought to underlie a wide variety of motion-guided behaviors, and the inclusion of this locally beneficial term is not detrimental to whole field motion estimation.

Finally we note that the principles discussed in the context of linear motion estimators also apply in other counterintuitive contexts. For example, consider an autocorrelator,

$$A = (f * V_1)(g * V_1) - (f * V_2)(g * V_2), \quad (81)$$

which correlates visual signals from the same point in space. Like the HRC, it is mirror anti-symmetric:

$$\{V_1(t), V_2(t)\} \mapsto \{V_2(t), V_1(t)\} \implies A \mapsto -A, \quad (82)$$

but it is uncorrelated with the velocity. Nevertheless, the autocorrelator's correlation with the HRC is determined by

$$\begin{aligned} \langle RA \rangle = & \langle (f * V_1)^2 (g * V_1)(g * V_2) \rangle + \langle (g * V_1)(f * V_2)^2 (g * V_2) \rangle \\ & - \langle (f * V_1)(f * V_2)(g * V_2)^2 \rangle - \langle (f * V_1)(g * V_1)^2 (f * V_2) \rangle \end{aligned} \quad (83)$$

and need not be zero. Empirically, we find the relevant correlation coefficient to be -0.40 across the ensemble of naturalistic motions, so Eq. (76) implies that this autocorrelator would enhance the HRC by 8.9%. However, such improvements do not sum over space. Thus, autocorrelators might be relevant for local motion estimates, but not for motion estimates that average over space.

Appendix XI: Regarding the computational problem of visual motion estimation

Throughout this paper, we have illustrated connections between the computations performed by our models and spatiotemporal correlations. These links are important for both practical and theoretical reasons. First, the many experimental successes of the HRC already suggest that the fly's computation of motion is organized around spatiotemporal correlations in the stimulus [11]. Thus, by relating our models to spatiotemporal correlations, we were able to discern how each model generalizes this canonical model. For example, **Fig. 3-suppl. 1b** shows that the optimal weighted 4-quadrant model supplements the standard HRC with a specific subclass of odd-ordered correlations, an observation that both reiterates the importance of the HRC and highlights the most critical signals that it lacks. Second, spatiotemporal correlations provide a fundamental connection between the motion estimation strategies used by invertebrates and vertebrates [4, 12]. In particular, although the HRC and motion energy models differ in their architectural details, both models are ultimately driven by 2-point correlations in the stimulus. Therefore, general arguments framed in terms of

spatiotemporal correlations are easy to investigate in the specific context of either the HRC or motion estimation model. Third, an understanding of the spatiotemporal correlations computed by each model facilitates the design of psychophysical experiments that test the models. For example, glider stimuli [10] provide a flexible experimental tool to probe how specific correlations contribute to motion percepts. Future work will lead to a variety of more realistic models that can also be characterized by the stimulus correlations that they detect. These models can be distinguished by carefully designed glider experiments.

From a theoretical point of view, correlation functions are important because they provide a mathematical basis in which to decompose neural computations [6, 7, 13]. David Marr famously proposed that neural computation must be understood at several levels [14]. He described his second level as that at which the algorithms that implement a computation are characterized. Our emphasis on correlation functions is directed towards unraveling motion estimation at this algorithmic level. As illustrated concretely by **Fig. 3-suppl. 1**, it’s possible for an algorithm to have a simple characterization in terms of correlation functions, even when the fundamental computational units (*e.g.* the quadrants) do not actually compute correlations. Furthermore, correlation functions intuitively relate the visual signatures of motion to measurable features of natural visual environments (**App. Fig. 1**). Nevertheless, it’s possible that correlation functions will ultimately provide an inefficient basis for representing the algorithms of visual motion estimation. For example, although the weighted 4-quadrant model is well understood in terms of the correlations that it detects, it would be nontrivial to discern its underlying simplicity based solely on its responses to glider stimuli, because the constraints relating various higher order correlators would be *a priori* unknown. Overall, we consider correlation functions to provide a useful lens for characterizing and understanding the algorithms of visual motion estimation, but research should also consider visual motion estimation in alternate bases that might reflect the brain’s biological substrates more directly [15].

Our characterization of visual motion estimation in terms of correlation functions provides an interesting perspective on the computational problem faced by *Drosophila*’s visual motion estimator in natural environments. Natural images contain many low and high-order correlations [3], and this implies that the fly brain could in principle use a wide array of correlations for visual motion estimation (**App. Fig. 1**). However, each correlation is only weakly associated with the velocity of motion in naturalistic settings [8, 9]. The reason for this is that the specific structure of the scene that is moving acts as a nuisance parameter that hinders the unambiguous assignment of a velocity to pattern of light input. For example, it’s well known that the temporal frequency of a moving sinusoidal grating determines the HRC’s output [16], thereby conflating the velocity with the grating’s spatial frequency. More generally, the variability of a multipoint correlator across an ensemble of moving scenes is determined by higher-order statistics of the image ensemble (*e.g.* see **Appendix II**). The fact that the same natural image drives every multipoint correlator also implies that the correlators co-vary with each other across the naturalistic motion ensemble. This shared variability can sometimes enable higher-order multipoint correlators to compensate effectively for image-induced noise that contaminates the HRC [8].

Questions of how the brain compute behaviorally relevant stimulus features from sensory inputs are central to neuroscience, but they are extraordinarily difficult to answer, even in principle. In the context of *Drosophila*’s visual motion estimator, the ensemble of photore-

ceptor signals contains many nonlinear cues that are weakly correlated with the stimulus velocity and with each other under naturalistic conditions. There are many ways to pool these signals into an improved motion estimate. The space of possible stimuli is astronomically large, so it is impossible for experiments to sample it completely. Nevertheless, synthetic laboratory stimuli can be designed to rule out specific algorithms that the brain might use to estimate motion. Thus, to deconstruct a neural computation, one must find ways to dramatically restrict the space of candidate models and to identify interesting models that can be experimentally ruled out. It's important to note that we did not construct our models to reproduce the behavioral data, even though this is a straightforward exercise (**Fig. 4-suppl. 1**). Instead we aimed for a predictive framework that can relate behavioral responses to the statistics of natural sensory inputs, the statistics of natural behavior, and the constraints imposed by the neural circuits that implement the computation. Such constructions are complicated and depend on features of neural circuits that are incompletely known. Nevertheless, we hope that this added complexity will eventually pay off in computational models that have a rational structure from the viewpoint of the stimulus, the animal, and the brain.

References

- [1] van Hateren JH, van der Schaaf (1997) Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc R Soc Lond B* 265:359-366.
- [2] Ruderman DL, Bialek W (1994) Statistics of natural images: scaling in the woods. *Phys Rev Lett* 73:814-817.
- [3] Geisler WS (2008) Visual perception and the statistical properties of natural scenes. *Annu Rev Psychol* 59:167-192.
- [4] Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284-299.
- [5] Potters M, Bialek W (1994) Statistical mechanics and visual signal processing. *J Phys I France* 4:1755-1775.
- [6] Fitzgerald JE, Katsov AY, Clandinin TR, Schnitzer MJ (2011) Symmetries in stimulus statistics shape the form of visual motion estimators. *Proc Natl Acad Sci USA* 108:12909-12914.
- [7] Poggio T, Reichardt W (1980) On the representation of multi-input systems: computational properties of polynomial algorithms. *Biol Cybern* 37:167-186.
- [8] Clark DA, Fitzgerald JE, Ales JM, Gohl DM, Silies MA, Norcia AM, Clandinin TR (2014) Flies and humans share a motion estimation strategy that exploits natural scene statistics. *Nat Neurosci* 17(2):296-303.
- [9] Dror DO, O'Carroll DC, Laughlin SB (2001) Accuracy of velocity estimation by Reichardt correlates. *J Opt Soc Am A* 18:241-252.

- [10] Hu Q, Victor JD (2010) A set of high-order spatiotemporal stimuli that elicit motion and reverse-phi percepts. *J Vis* 10(3):9, 1-16.
- [11] Silies M, Gohl DM, Clandinin TR (2014) Motion-detecting circuits in flies: coming into view. *Annu Rev Neurosci* 37:307-327.
- [12] van Santen JPH, Sperling G (1985) Elaborated Reichardt detectors. *J Opt Soc Am A* 2:300-320.
- [13] Poggio T, Reichardt W (1973) Considerations of models of movement detection. *Kybernetik* 13:223-227.
- [14] Marr D, Poggio T (1976) From understanding computation to understanding circuitry. *Massachusetts Institute of Technology Artificial Intelligence Laboratory* A.I. Memo 357.
- [15] Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neuro* 9:1421-1431.
- [16] Egelhaaf M, Borst A, Reichardt W (1989) Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. *J Opt Soc Am A* 6:1070-1087.