

Nonlinear circuits for naturalistic visual motion estimation

James E. Fitzgerald¹ and Damon A. Clark²

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

²Departments of Molecular, Cellular, and Developmental Biology and of Physics, Yale University, New Haven, CT 06511

Abstract

Many animals use visual signals to estimate motion. Canonical models suppose that animals estimate motion by cross-correlating pairs of spatiotemporally separated visual signals, but recent experiments indicate that humans and flies perceive motion from higher-order correlations that signify motion in natural environments. Here we show how biologically plausible processing motifs in neural circuits could be tuned to extract this information. We emphasize how known aspects of *Drosophila*'s visual circuitry could embody this tuning and predict fly behavior. We find that segregating motion signals into ON/OFF channels can enhance estimation accuracy by accounting for natural light/dark asymmetries. Furthermore, a diversity of inputs to motion detecting neurons can provide access to more complex higher-order correlations. Collectively, these results illustrate how non-canonical computations improve motion estimation with naturalistic inputs. This argues that the complexity of the fly's motion computations, implemented in its elaborate circuits, represents a valuable feature of its visual motion estimator.

Introduction

A major goal in neuroscience is to understand how the brain computes behaviorally relevant stimulus properties from streams of incoming sensory data¹. Visual motion guides

behaviors across the animal kingdom. To navigate, many vertebrates and invertebrates use visual data to estimate the velocity of full field motion, and they use that estimate to judge their motion with respect to their environment²⁻⁵. Spatially localized motion perception⁶⁻⁸ is also important, as it can indicate the presence of predators or prey in the environment⁹⁻¹², and spatial velocity gradients allow animals to judge relative distances¹³⁻¹⁶. In principle, different algorithms could be used to estimate different types of motion. However, data suggest that many animals compute local motion over an array of spatially localized elementary motion detectors, or EMDs, and then differentially pool those signals for use in different behaviors and neural operations^{6-8,12,17-19}.

The Hassenstein-Reichardt correlator (HRC) was introduced nearly sixty years ago to model the EMD underlying the beetle's optomotor response²⁰. It has since provided numerous insights into motion-guided behaviors across a variety of insect species. The HRC's successes are most striking in flies, where the HRC accurately predicts a wide variety of behavioral and neural responses^{8,21-23}, and even adaptation to stimulus statistics²⁴. The HRC's importance also extends to primates and vertebrates, where the EMDs are often described in terms of the motion energy model²⁵. In particular, although the HRC and motion energy models differ in terms of their intuition and neuronal bases, both models rely on the same mathematical fact about moving visual stimuli – motion causes pairs of spatially separated points to become correlated when a stimulus moves from one location towards the other^{25,26}.

Research on *Drosophila*'s motion detection system has progressed quickly in recent years. With an influx of novel genetic, anatomical, and physiological tools, *Drosophila* researchers are able to perform experiments that have revealed an intricate neural circuit whose details were not anticipated. For example, separate pathways process the motion of minimal light and dark moving edges²⁷⁻²⁹, and different neurons within these pathways coordinate the motion

response depending on the velocity of motion³⁰. Furthermore, connectomic analysis has revealed that more spatial and temporal channels converge onto the fly's motion computing neurons than had been predicted by the HRC's two-input architecture³¹. Going forward, it is critical that the field discovers which of these circuit details into are computationally relevant and which are not. Since many of these details go beyond the HRC's premise, we must consider alternate theories if we hope to understand how circuit details contribute to motion estimation.

A large body of theoretical and experimental work supports the hypothesis that visual systems are tailored for functionality in the animal's natural behavioral context³². For example, photoreceptors adapt effectively across the ecological range of light levels³³, the excess number of OFF versus ON retinal ganglion cells matches the excess information of dark versus light contrasts in natural images³⁴, and several learning algorithms predict the receptive fields of early cortical neurons when applied to natural images^{35,36}. These examples are special cases of the general hypothesis that the early visual system provides an efficient code for the natural visual environment, and recent research suggests that efficient coding accounts for certain aspects of higher-level coding and perception as well³⁷⁻³⁹.

Several recent studies have established connections between the biological algorithms used for visual motion estimation and the statistical demands of naturalistic motion estimation. Natural stimuli are intricately structured and light-dark asymmetric⁴⁰, and a variety of low and high order correlations characterize motion in such an environment. Although the HRC and motion energy models only respond to pairwise correlations in their inputs^{25,26}, the Bayes optimal visual motion estimator also incorporates a variety of higher-order correlations of both even and odd order^{41,42}. Accordingly, certain visual stimuli that contain only higher-order correlations induce motion percepts in both vertebrates and insects^{4,43-47}, and theoretical work

shows that the correlations that characterize these stimuli can also improve motion estimation in natural environments⁴³. This demonstrates that neither the HRC nor the motion energy model can account for the totality of experimentally observed motion percepts and suggests that departures from these canonical models might improve motion estimation accuracy. Relatively little is known about the neural basis of these higher-order motion percepts, although several studies have suggested intriguing commonalities across insect and primate species^{43,48}.

Here we investigate whether the computational demands imposed by accurate motion estimation in natural environments can illuminate the unexpected details of *Drosophila*'s motion estimation circuit or account for non-Reichardtian motion perception in flies. We study a sequence of five computational models, each of which considers a conceptually new aspect of the motion estimation problem. Since each model succeeds in improving estimation accuracy, these results provide a range of nonlinear circuit mechanisms that flies and other animals might incorporate into their motion estimators. We describe how observed elements of *Drosophila*'s motion estimation circuitry could support such computations (**Table 1**). Importantly, four of the five models also predict the signs and approximate magnitudes of known non-Reichardtian motion percepts in flies. Since the models were tuned exclusively for estimation accuracy, these results support the view that non-Reichardtian motion percepts probe ethologically relevant aspects of biological motion estimators. More generally, our results posit normative interpretations for some unexpected aspects of the fly's motion estimation circuit and behavior and suggest that non-Reichardtian aspects of fly circuitry and behavior might be closely linked through the statistics of natural scenes.

Results

Flies incorporate motion signals that the Hassenstein-Reichardt correlator (HRC) neglects.

The Hassenstein-Reichardt correlator (HRC) is the dominant model of motion computation in flies and other insects. In this paper we describe several generalizations of the HRC, but it is helpful to first review this canonical model. The HRC comprises three stages of processing. First, two different temporal filters (here, a low-pass filter ($f(t)$) and a high-pass filter ($g(t)$)) are applied to each of two spatially filtered visual input streams (**Figure 1a, Materials and Methods**). These four filtered signals are then paired and multiplied (**Figure 1a**). Finally, the HRC takes the difference between the two multiplied signals to obtain a mirror anti-symmetric motion estimator (**Figure 1a**). Because the HRC combines its two input channels via a multiplication operation, the average output of the HRC depends only on 2-point correlations in the visual stimulus. We thus refer to the HRC as a 2-point correlator, and we will return to this mathematical characterization of the HRC repeatedly throughout this work.

No motion estimator is perfect for every stimulus, and this paper explores the hypothesis that evolution has tuned *Drosophila*'s motion estimator for visual experiences that are likely to result from ordinary behavior in natural environments (**Appendix I**). We assessed the accuracy of the HRC and other motion estimators by approximating naturalistic motion as the rigid translation of natural images⁴³, with a velocity distribution that mimicked *Drosophila*'s natural behavior (**Figure 1b, Materials and Methods**)⁴⁹. We spatiotemporally filtered the input signals to simulate the responses of two neighboring photoreceptors (**Materials and Methods**). We quantified the performance of each model as the mean squared error between the input velocity and model output. However, we report each model's accuracy as the correlation coefficient between its output and the true velocity (**Figure 1c**), an intuitive metric that is equivalent to the mean squared error for correctly scaled model outputs (**Materials and Methods**). In isolation,

the local HRC was weakly correlated with the velocity of motion (**Figure 1d**). Although the HRC's performance can be improved by averaging over space and time^{43,50}, this study explores how alternate nonlinear processing can improve motion estimation accuracy without sacrificing spatial or temporal resolution⁴³.

Researchers can probe a fly's motion estimate by measuring its behavioral optomotor turning response^{3,8,20,51}. We previously measured optomotor responses from flies walking on a spherical treadmill by recording their turning responses to various visual stimuli (**Figure 1e**)⁴³. We emphasized binary stimuli called gliders⁴⁵ (**Figure 1f**), which enforce spatiotemporal correlations to interrogate the fly's motion estimation algorithm. For example, 2-point gliders contain only 2-point correlations (*first two stimuli, Figure 1f*). *Drosophila* turned in response to these stimuli⁴³ (*black bars, left, Figure 1g*), and the HRC correctly predicts that flies would respond to both positive and negative 2-point correlations (*gray bars, left, Figure 1g*). On the other hand, 3-point gliders contain 3-point correlations without 2-point correlations (*last four stimuli, Figure 1f*). These stimuli generated motion responses in flies (*black bars, right, Figure 1g*) that the HRC could not explain (*gray bars, right, Figure 1g*). Thus, behavioral responses to glider stimuli show that the HRC is an incomplete description of fly motion estimation and provide a useful benchmark for evaluating alternate models.

In this study, we tune our models to optimize motion estimation accuracy, rather than to fit the behavioral data, for two main reasons. First, we want to explore the hypothesis that *Drosophila*'s glider responses follow from performance optimization within biologically plausible circuit architectures. Second, we seek models that will generalize well across visual stimuli, and the measured glider responses under-constrain possible motion estimation models. It's useful to illustrate our procedure with a simple example. The HRC does not account for 3-

point glider responses because it is insensitive to 3-point correlations. Nevertheless, 3-point correlations are present in natural stimuli^{43,52}, and their use might facilitate accurate motion estimation. We can explore this hypothesis by summing the HRC with a motion estimator designed to respond specifically to 3-point correlations. For instance, the mirror anti-symmetric “converging” 3-point correlator multiplies one high-pass filtered signal with two low-pass filtered signals (**Figure 1h**) and mimics the converging structure present in certain glider stimuli (*last two stimuli*, **Figure 1f**). We tune the model for motion estimation accuracy by choosing the weights of the HRC and the converging 3-point correlator to minimize the mean squared error (**Materials and Methods**). The resulting model is more accurate than the HRC (**Figure 1i**) and it predicts that flies should respond to glider stimuli in the observed directions (**Figure 1j**, **Materials and Methods**). Nevertheless, this simple model underestimates 3-point turning magnitudes (**Figure 1j**), indicating a discrepancy between the fly’s motion estimator and this performance-optimized model.

In this study, we apply this same basic model building procedure to a series of increasingly general model architectures. There are four benefits to this approach. First, each model incorporates a type of computation that was neglected by earlier models. Thus, we can compare model accuracies to quantify how important various computations are for naturalistic motion estimation. Second, each model has a distinct biological interpretation in terms of *Drosophila*’s motion estimation circuit (**Table 1**). This allows us to enumerate many directions for future experimental and computational research. Third, this set of models reveals several distinct principles of accurate naturalistic motion estimators, yet no single model illustrates every principle. Finally, by comparing the glider predictions of each model to behavioral data, we can gain insight into which principles underlie *Drosophila*’s known glider responses.

Nonlinear preprocessing of HRC inputs improves estimation but poorly predicts responses to gliders.

The HRC correlates pairs of photoreceptor signals (**Figure 1a**). We previously assumed that each photoreceptor's response was generated from incoming contrast signals through linear spatiotemporal filtering. However, real photoreceptors are linear over a limited range of inputs^{33,53} (**Table 1**). Our first model thus modifies the HRC by allowing the photoreceptor responses to become nonlinear (**Figure 2a**). More specifically, we consider models in which a static nonlinearity transforms the filtered contrast signals before a standard HRC is applied to the two input streams (**Figure 2a, Materials and Methods**). Since the nonlinearity occurs before the HRC, we refer to this model as the *front-end nonlinearity* model. By nonlinearly transforming the contrast signals, the front-end nonlinearity model is able to reshape natural sensory statistics. In particular, linear photoreceptor signals inherit complex non-Gaussian statistics from their natural inputs (**Figure 2b**), but front-end nonlinearities (**Figure 2c**) can produce transformed signals with alternate statistics (**Figure 2d, Materials and Methods**). Thus, optimal front-end nonlinearity models should reshape natural statistics into those that best suit the HRC. Previous studies have already demonstrated example front-end nonlinearity models that improve naturalistic motion processing by the HRC^{50,54}. Here we provide new theoretical insight into these improvements and their consequences for glider responses.

Although the statistics of natural images are complicated, the mean squared error between the HRC's output and the velocity of motion depends only on a few statistical quantities. Since the HRC is a 2-point correlator, the mean velocity signal decoded by an HRC is determined by the second-order statistics of the image ensemble⁵⁰. The variance of the motion

signal comes from the square of a quadratic signal, and thus the noise statistics of the HRC depend on the fourth-order statistics of the image ensemble (**Appendix II**). If the image ensemble is spatially uncorrelated, the situation simplifies further and the correlation between the estimated and true image velocity is determined entirely by the standardized fourth central moment of the input streams, a quantity known as kurtosis (**Appendix III**). A larger kurtosis results in a larger error in the motion estimate. Note that some authors use “kurtosis” to refer to the “excess kurtosis,” which shifts kurtosis values such that the Gaussian distribution has zero excess kurtosis. This shift is not relevant for our purposes. Because large positive contrasts are relatively probable, naturalistic inputs are highly kurtotic (kurtosis = 9.6 for the spatiotemporal filtering in our simulations) and are thus expected to hinder HRC performance (**Figure 2b**).

The Gaussian, uniform, and symmetric Bernoulli distributions have much lower kurtosis values (kurtosis = 3.0, 1.8, 1.0, respectively, **Figure 2d**). In fact, the symmetric Bernoulli distribution has the lowest kurtosis of any probability distribution⁵⁵. When we transformed the HRC’s inputs to have these statistics (**Materials and Methods**), we found that each nonlinearity substantially improved the accuracy of the HRC (**Figure 2e**). The *contrast equalizing* nonlinearity, which produces uniform outputs, performed best and also plays a prominent role in efficient coding theory⁵³. It is interesting that contrast equalization improved the accuracy of the HRC more than binarization (**Figure 2e**), even though it produced outputs with greater kurtosis. The reason for this is that natural images are spatially correlated, 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 (**Figure 2-figure supplement 1**), and spatial correlations can enhance the performance of the HRC (**Appendix IV-V**). Designing a

nonlinearity that optimally sculpts the correlation structure of natural images is not simple and goes beyond the scope of this study.

Each front-end nonlinearity model is sensitive to a variety of higher-order correlations (**Appendix VI**). We thus tested whether accurate front-end nonlinearity models would predict *Drosophila*'s glider response pattern. However, each front-end nonlinearity model performed poorly at this task (**Figure 2f-g**). None of the three models predicted that *Drosophila* would invert its response to positive and negative 3-point gliders (**Figure 2g**), even though they predicted that the 3-point glider responses would be nonzero. The simplest explanation for this observation is that the front-end nonlinearity models responded to fourth-order correlations that are common to the stimuli, rather than the third-order correlations that defined the glider stimuli and primarily drove the experimental response⁴³. Mechanistically, this result follows from the fact that the nonlinearities that reduced kurtosis (**Figure 2c**) were not strongly asymmetric around zero contrast (**Appendix VI**). The binarizing front-end nonlinearity model also failed to predict that *Drosophila* would respond less to negative 2-point glider stimuli than positive 2-point glider stimuli (**Figure 2f**). Since this effect was correctly predicted by the standard HRC (**Figure 1g**), this observation shows that accurate front-end nonlinearity models can distort the processing of 2-point correlations. Although the front-end nonlinearity model did not explain the phenomenon of fly glider perception, future work should investigate whether its merits make it functionally relevant for motion processing in other contexts or species.

Separating ON and OFF signals improves motion estimation and predicts responses to gliders.

Instead of a front-end nonlinearity, *Drosophila* could use an alternative non-Reichardtian motion estimation strategy that reflects natural sensory statistics, without necessarily requiring

nonlinear preprocessing. Previous computational analyses show that motion estimation strategies that distinguish light and dark information can enhance motion processing with natural inputs^{42,43,52}, and recent experiments indicate that flies use separate channels to process the motion of light and dark edges^{27,28,43,56-58} (**Table 1**). Our next model explores the hypothesis that *Drosophila* segregates ON and OFF signals in order to facilitate naturalistic motion estimation⁴³ (**Figure 3a, Materials and Methods**). There are four ways to pair the ON and OFF components of the two filtered signals that enter the HRC's multiplier. For example, one possibility is to pair the ON component of the low-pass filtered signals with the OFF component of the high-pass filtered signal. Since each pairing restricts the HRC's multiplier to a single quadrant of the Cartesian plane, we refer to these four signals as HRC-quadrants. If the quadrants are summed with equal weights, then this model is mathematically identical to the HRC^{20,43}. Unequal weighting coefficients enable the motion estimator to prioritize some quadrants over others, and here we select quadrant weightings that minimize the mean squared error between the model output and velocity (**Figure 3b, Materials and Methods**). More generally, we refer to any model that linearly combines the four HRC-quadrants as a *weighted 4-quadrant* model. The precise manner in which the four HRC-quadrants might map onto circuitry remains unclear; we do not suggest there exists separate circuitry for each quadrant. For instance, studies have identified only two motion-processing channels in the *Drosophila* brain, which might suggest that the fly only uses a subset of the quadrants^{29,59,60}. On the other hand, each channel appears imperfectly selective for light versus dark signals⁵⁶, which in principle enables these two channels to access all four quadrants (**Table 1**).

We began by examining how well individual quadrants predicted the velocity of motion. The four quadrants provided motion signals of strikingly different quality (*first four red bars,*

Figure 3c). The most accurate quadrant correlated negative low-pass filtered signals with negative high-pass filtered signals (($- -$) *bar*, **Figure 3c**). This isolated quadrant already outperformed the full HRC. The quadrant that correlated negative low-pass filtered signals with positive high-pass filtered signals also performed relatively well (($- +$) *bar*, **Figure 3c**), whereas the quadrants that involved positive low-pass filtered signals performed poorly (($+ +$) and ($+ -$) *bars*, **Figure 3c**). This shows that negative signals emanating from the low-pass filter better facilitate motion estimation, and the HRC's uniform weighting of all four quadrants is computationally detrimental.

We next considered all subsets of two, three, or four quadrants. The best subsets for each number of predictors were nested, and the quadrants were incorporated in the order (i) ($- -$); (ii) ($- +$); (iii) ($+ +$); (iv) ($+ -$). Although all four quadrants enhanced the accuracy of the weighted 4-quadrant model, the benefit of each added quadrant decreased with the number of quadrants (**Figure 3c**). It is possible to reparameterize the weighted 4-quadrant model in a form that isolates the contributions of various higher-order correlations to the model's accuracy (**Appendix VII**). Interestingly, this parameterization showed that nearly all the accuracy of the weighted 4-quadrant model can be obtained by supplementing the HRC with a set of odd-ordered correlations that account for the asymmetry between positive and negative low-pass filtered signals (**Figure 3-figure supplement 1, Appendix VIII**). Principal component analysis did not reveal this simple interpretation of the model's computation (**Appendix IX**).

The performance-optimized weighted 4-quadrant model also offered an interesting interpretation of *Drosophila*'s glider response pattern. First note that the model preserved the HRC's response pattern to 2-point glider stimuli (compare *left* subpanels of **Figure 3d** and **Figure 1g**). More interestingly, the model predicted behavioral responses to 3-point glider

stimuli that matched the experimentally observed turning directions, and even the response magnitudes were similar between the model and the data (*right*, **Figure 3d**). Nevertheless, the model's predictions were imperfect. The primary qualitative discrepancy was that the model failed to predict that positive 3-point glider stimuli would generate smaller turning responses than negative 3-point glider stimuli. The simplest interpretation for this experimental result is that flies might incorporate both 3-point correlations and 4-point correlations into their motion estimation strategy. In particular, since the positive and negative 3-point glider stimuli have inverted 3-point correlations and matched 4-point correlations, third-order and fourth-order correlations would have the same sign for one parity and opposite signs for the other parity. This observation makes it easier to understand the glider predictions of the weighted 4-quadrant model. The optimized model does a good job accounting for the direction and approximate magnitude of the glider responses because it draws heavily on second-order and odd-order correlations, but it fails to predict the 3-point glider magnitude asymmetry because it finds little added utility in higher-order even correlations (**Figure 3-figure supplement 1, Appendix VIII**). This failure stems from architectural limitations in the weighted 4-quadrant model (**Figure 3-figure supplement 2**), so it is important to consider alternate model classes.

Drosophila circuitry contains additional elements that might facilitate motion estimation.

The previous section suggested that the segregation of light and dark signals by *Drosophila*'s motion estimation circuitry might enhance naturalistic motion estimation in a manner that also generates the observed glider responses. In this section, we introduce three hierarchical models to investigate other features of *Drosophila*'s circuit that might have functional consequences for the processing of natural stimuli and gliders (**Table 1**).

The first of these models recasts the HRC and the weighted 4-quadrant model in a more general architecture. This model is the class of mirror anti-symmetric models that apply a 2-dimensional nonlinearity to the low-pass filtered signal from one point in space and the high-pass filtered signal from a neighboring point in space (**Figure 4a**). Since the observed glider responses indicate that flies use higher-order correlations of both even and odd order, we model this 2-dimensional nonlinearity as a fourth-order polynomial (**Materials and Methods**). The HRC corresponds to the special case of this nonlinearity that multiplies the two inputs (*left*, **Figure 4b**). To emphasize how the model class in **Figure 4a** generalizes the HRC, we refer to it as the *non-multiplicative nonlinearity* model. In comparison, the weighted 4-quadrant model corresponds to a different non-multiplicative nonlinearity that separately scales a pure multiplication in each quadrant of the Cartesian plane. Compared to the HRC, the optimized forms of both the weighted 4-quadrant model and the non-multiplicative nonlinearity substantially attenuated positive low-pass filtered signals (*middle* and *right*, **Figure 4b**), though the non-multiplicative nonlinearity shows less attenuation. This model architecture provides enough flexibility to generate the glider response pattern (**Figure 4- figure supplement 1**).

The non-multiplicative nonlinearity model relaxes some restrictions of the 4-quadrant model. This is prudent because the exact nonlinear transformations implemented by neural circuits in the *Drosophila* brain remain poorly understood. For example, T4 and T5 are the first direction-selective neurons in the fly brain²⁹, but the mechanism by which they become direction-selective is not yet known. Furthermore, neurons upstream of T4 and T5 imperfectly segregate light and dark information⁵⁶ and show overlap between the two motion pathways⁶¹, suggesting that ON/OFF segregation may not be crisply realized. We will discuss this model's estimation accuracy and glider performance in the next section.

Drosophila's motion processing circuitry suggests two more generalizations of the non-multiplicative nonlinearity model. First, note that the non-multiplicative nonlinearity model inherits the HRC's assumption that each nonlinear unit only acts upon the low-pass filtered signal from one point in space and the high-pass filtered signal from the neighboring point (**Figure 4a**). In contrast, the converging 3-point correlator (**Figure 1h**) shows that the accuracy of motion estimation can sometimes be enhanced by nonlinearly combining both low-pass filtered signals (**Figure 1i**). Moreover, connectomic evidence conflicts with the non-multiplicative nonlinearity model's constraints, because each T4 cell receives synaptic connections from both the Mi1 cell and the Tm3 cells (T4's two major input channels) at overlapping points in space³¹. The *unrestricted nonlinearity* model removes this restriction of the non-multiplicative nonlinearity model by allowing a 4-dimensional nonlinearity to act on all four filtered signals (**Figure 4c**). Here, we again model this nonlinearity as a fourth-order polynomial (**Materials and Methods**). The unrestricted nonlinearity allows the motion estimator to nonlinearly combine multiple temporal channels from the same point in space. Recent experiments indicate that the Mi1 and Tm3 cells alone are insufficient to account for the motion processing of the T4 channel³⁰. Future work might generalize the unrestricted nonlinearity model to include three or more temporal channels at each point in space. Here we refrain from modifying the temporal filtering of the motion estimator.

The models presented so far operate only on a pair of neighboring photoreceptors, and the final generalization incorporates a third point in space. Averaging EMDs over space improves the accuracy of whole-field motion estimation⁵⁰, but *Drosophila*'s neural circuitry suggests that it might adopt a more sophisticated strategy to combine signals across space. In particular, single T4 cells receive synaptic inputs from Mi1 cells and Tm3 cells from more than

two retinotopic columns³¹. This arrangement could allow the circuit to incorporate higher-order correlations that are distributed across three or more spatial input channels. To explore whether this possibility has computational significance, we generalized the unrestricted nonlinearity model to provide unrestricted access to six temporal channels distributed across three points in space (**Figure 4d**). We refer to this model as the *extra input nonlinearity* model. We approximate its 6-dimensional nonlinearity as a fourth-order polynomial (**Materials and Methods**).

Elaborated circuit architectures improve motion estimation without sacrificing glider responses.

Having introduced the rationale behind the non-multiplicative, unrestricted, and extra input nonlinearity models, it is straightforward to examine their performance as motion estimators. First note that the polynomial non-multiplicative nonlinearity model was a better motion estimator (**Figure 4e**) than the weighted 4-quadrant model (**Figure 3c**). This implies that some useful signatures of naturalistic motion are not made accessible by simply segregating ON and OFF motion signals. Interestingly, this performance improvement is largely due to 3-point correlations, and models that exclude fourth-order polynomial terms still outperform the weighted 4-quadrant model (**Figure 4- figure supplement 2**). Third-order correlations are only useful for motion estimation because of light-dark asymmetries in natural stimulus statistics^{42,43}, so this result implies that ON/OFF segregation provides an imperfect way to account for the complexity of light-dark asymmetries found in the natural world. The non-multiplicative nonlinearity model also made novel use of low-order correlations to improve its motion estimate (**Appendix X**)

The three models are hierarchical because the non-multiplicative nonlinearity model is a special case of the unrestricted nonlinearity model, which is itself a special case of the extra

input nonlinearity model. Thus, we expect each model to perform at least as well as its predecessor, but it is possible that some circuit elaborations will not introduce useful computational cues. Nevertheless, we found that the unrestricted nonlinearity model performed better than the non-multiplicative nonlinearity model, and the extra input model performed better than the average of two neighboring unrestricted nonlinearity models (**Figure 4e**). Therefore, both models incorporated novel computational signatures with relevance for visual motion estimation. Although the relative improvements were fairly small, it's worth noting that the improvement from spatial averaging is also small, and it is possible that the fly brain builds an accurate motion estimator by combining a large number of weak predictors of motion.

Each of these three generalized models predicted 2-point glider responses (**Figure 4f**) that closely resembled the standard HRC (*left*, **Figure 1g**). Each model also correctly predicted the experimental turning directions to each of the 3-point glider stimuli (**Figure 4g**). The magnitudes of the 3-point glider turning responses did not unambiguously favor any of the three hierarchical models (**Figure 4g**) or the weighted 4-quadrant model (*right*, **Figure 3d**). Each model did better on some stimuli and worse on others. Nevertheless, the predicted glider responses did make several interesting points. First, the extra input nonlinearity model predicted a clear asymmetry between positive and negative 3-point gliders (**Figure 4g**). This shows that some of the even-ordered correlations found in 3-point glider stimuli have relevance for naturalistic motion estimation. Second, the observation that each model provides qualitatively similar glider response patterns illustrates that animals could use multiple nonlinear mechanisms to access ethologically relevant higher-order correlations. Future experiments should directly assess the functional relevance of the different models in the hierarchy. Finally, the qualitative

agreement between all of these predictions and the experimental data supports the general hypothesis that glider responses could reflect underlying nonlinear mechanisms that facilitate motion estimation in natural environments.

The extra input nonlinearity model retains the conceptual content of its predecessors.

In this paper, we sequentially introduced several models in order to isolate specific ideas about the relationships between *Drosophila*'s behavior, its motion estimation circuit, and the statistical demands of accurate motion estimation in natural environments. The front-end nonlinearity model explored an interesting candidate principle for visual motion estimation, but it conflicted sharply with fly behavior (**Figure 2g**) and excluded the conceptual insights offered by other models. For example, the front-end nonlinearities we considered eliminated the asymmetry between light and dark contrasts (**Figure 2d**), removing the need for separate ON and OFF processing. However, the remaining models embodied ideas that are complementary rather than exclusive, and these models should not be thought of as competitors. Instead, we will show here that the final, most general model incorporates the variety of conceptual points that were initially illustrated by specific models.

The structure of the non-multiplicative nonlinearity models can be directly plotted (**Figure 4b**), but is not easy to visualize the 6-dimensional nonlinearity that defines the extra input nonlinearity model. We therefore need an alternate technique to illustrate its computations. We proceed by leveraging three ideas. First, a wide variety of visual motion estimators can be expanded as an abstract series of multipoint correlators (*e.g.* see Poggio & Reichardt, 1980⁶², Fitzgerald et al., 2011⁴², and **Appendices VI, VII, XI**), and it is straightforward to pictorially represent a multipoint correlator (*e.g.* see **Figures. 1a, 1h**, and more to come). In the extra input

nonlinearity model, this expansion is immediate because we have already parameterized its 6-dimensional nonlinearity as a polynomial. Importantly, this expansion should be considered at the algorithmic level⁶³, and we do not suggest that the wiring of brain circuits will reflect a large number of higher-order correlators. To the contrary, a large number of higher-order multipoint correlators may be implemented implicitly by high-dimensional nonlinearities suggested by *Drosophila*'s visual circuitry. Second, we note that certain multipoint correlators can be recombined into a 2-dimensional non-multiplicative nonlinearity that facilitates easy comparisons with the HRC and weighted 4-quadrant models (e.g. see **Figure 4b**). Taken together, these two points mean that we can represent the performance-optimized extra input nonlinearity model in terms of non-multiplicative nonlinearity models and multipoint correlators, each of which are easy to represent graphically.

This graphical representation could be unwieldy because of the sheer number of higher-order correlators in the model. Thus the third and final point is that we need a way to identify a relatively small number of terms that substantially improve the accuracy of motion estimation and illustrate the conceptual content of the model. To achieve this, we use lasso regression⁶⁴ to identify sets of models with fewer multipoint correlators that still enable accurate motion estimation (**Materials and Methods**). This analysis revealed that fewer than half of all multipoint correlators were needed to account for the full accuracy of the extra input nonlinearity model (*rightmost bars*, **Figure 5a**). In fact, the accuracy of naturalistic motion estimation increased rapidly as the few correlators were sequentially added (*left bars*, **Figure 5a**), and a model that used 16 out of the 209 possible predictors was already able to produce 74% of the gain offered by the full extra input nonlinearity model (*red bar*, **Figure 5a**).

The leading 16 predictors compactly illustrated how the extra input nonlinearity model recapitulates the conceptual advances offered by the other models (**Figure 5b**). Four of the predictors combine to implement a mirror-symmetric non-multiplicative nonlinearity model that acts on the first and second points in space (*first term*, **Figure 5b**). The dominant contribution to the nonlinearity is the HRC's multiplier, but an additional third-order term breaks the symmetry between positive and negative low-pass filtered signals (**Figure 5- figure supplement 1**). Thus, the extra input nonlinearity model approximately correlates neighboring points in space, as the HRC would suggest, but it differentially weights positive and negative low-pass filtered signals, like the weighted 4-quadrant model. It also replicates the main insight from the non-multiplicative nonlinearity model: the best treatment of asymmetric light and dark information need not be as simple as pure ON/OFF segregation. The model used another eight predictors to construct two more non-multiplicative nonlinearity models, one that surveyed the second and third points in space and another that surveyed the first and third points (*first and second terms*, **Figure 5b**, **Figure 5- figure supplement 1**). These components make the previously highlighted conceptual points and add the observation that spatial averaging improves estimates.

The final four predictors implemented two mirror anti-symmetric multipoint correlators (*third and fourth terms*, **Figure 5b**). In particular, two predictors went towards implementing a converging 3-point correlator that spanned the first and third spatial points (*third term*, **Figure 5b**). This estimator made the model's asymmetric treatment of light and dark signals more nuanced than permitted by the non-multiplicative nonlinearity model, and it also incorporated motion signals that combine multiple temporal signals from the same point in space. This latter point was the main conceptual motivation for the unrestricted nonlinearity model. Finally, the last two predictors implemented a 4-point correlator that combined temporal signals from three

distinct points in space (*fourth term*, **Figure 5b**). This component reinforces the conceptual motivation for the extra input nonlinearity model and gives a concrete example of a computationally relevant higher-order correlator that is distributed across three points in space. It's interesting that the leading fourth-order correlator spanned three spatial points, because the extra input nonlinearity model was the first performance-optimized model that generated a substantially asymmetric response to positive and negative 3-point gliders (**Figure 4g**).

This paper set out with the goal of exploring whether the statistical demands of naturalistic motion estimation could provide a useful lens for interpreting features of *Drosophila*'s behavior and neural circuitry that push beyond the canonical HRC. Although we have considered several interesting classes of visual motion estimators, the space of possible motion estimators is much larger (**Figure 5c**). For instance, these models have not explored the impact of temporal filter choice on naturalistic motion estimation. Nor have they assessed the possibility of more than two temporal filters, which is suggested by anatomical³¹ and physiological³⁰ experiments. More generally, the neural circuits contributing to *Drosophila*'s motion estimator are still incompletely known, and the extent to which the fly brain's biological complexity reflects computational sophistication remains an open question. Theoretical considerations will be critical for resolving that question and pinpointing the most relevant principles underlying visual motion estimation.

Discussion

Ongoing research is providing an increasingly detailed picture of the anatomy and physiology of the visual circuitry that implements motion processing in *Drosophila*. Through the combination of genetic silencing experiments, connectomic analysis, and functional recordings,

researchers have identified many individual neurons in the fly brain that contribute to visual motion processing⁶⁵. Although the HRC provided the initial theoretical impetus for these experiments, specific experimental outcomes have often been unanticipated. For instance, the fly brain contains multiple pathways that segregate different types of motion information^{27,28,61}; its direction-selective neurons receive inputs from more than two neighboring points in visual space³¹; and the biological substrates for reverse-phi signals, which were fundamental to the formulation of the HRC, remain poorly understood^{27,60,66}. Theoretical work to illuminate the computational significance of these various discrepancies is critical for understanding *Drosophila*'s motion estimator.

The results presented in this paper provide a new theoretical perspective on these experimental results. A recurring theme of our models was that motion-processing circuits should treat light and dark signals differently. We first showed that visual systems could use ON and OFF processing channels that separately correlate light and dark signals to improve the accuracy of motion estimation (**Figure 3**). This model was inspired by the experimental observation that *Drosophila*'s motion processing channels distinguish between light increments and decrements^{27,28}, but this study is the first to explicitly demonstrate how such processing channels can improve the accuracy of motion estimation. Furthermore, our model shows that both the phi channels (*i.e.* the $(++)$ and $(--)$ and quadrants) and the reverse-phi channels (*i.e.* the $(+-)$ and $(-+)$ quadrants) can contribute productively to motion estimation in natural environments. Since many animals experience similar sensory statistics and ON and OFF visual processing channels are pervasive across visual systems^{67,68}, these mechanisms might be very general. Ultimately, the performance gains from weighted quadrants were a consequence of statistical asymmetries between light and dark contrasts in natural images, and our models

showed that neural circuits could perform even better if they made distinctions between light and dark signals that were subtler than simple ON/OFF segregation (**Figure 4**). Recent experimental evidence indicates that the fly's motion processing channels are imperfectly selective for ON versus OFF information^{56,57,61}, and it is important that future experiments characterize such subtleties in the computations performed by these circuits.

Our most general model contained three spatial inputs and showed that spatial averaging of local motion detectors was suboptimal (**Figure 4e**). Anatomy suggests that single T4 cells receive inputs from several different retinotopic columns, and also from multiple neuron types in a single retinotopic column³¹. Our modeling suggests that these two forms of circuit heterogeneity could enhance motion estimation by facilitating computations that go beyond averaging to compute higher-order correlations that are distributed across multiple points in space (**Figure 5b**). Overall, our results demonstrate how the subtleties of neural nonlinearities can improve motion detection with naturalistic inputs. It therefore seems likely that some of the complexities of *Drosophila*'s circuitry are critical to its performance under natural conditions.

It is remarkable that our approximation of natural motion by the rigid translation of natural images revealed substantial utility for higher-order correlations in motion processing. Truly naturalistic motion would include spatial velocity gradients, occlusion, expansion, and contraction, yet the simplified naturalism we used to optimize our models already sufficed to account for many aspects of the fly's glider responses. This may be because the rotational optomotor response measured in the fly experiments is thought to be sensitive primarily to full-field rotations, which our naturalism emulates well. However, since other higher-order correlations may be associated with non-rigid translation⁵², one might expect a different set of glider sensitivities to be optimal in the context of other motion-guided behaviors, such as

looming responses^{12,69,70}. Since a common elementary motion detector might underlie many or all motion-guided behaviors, incorporating more complex optic flow patterns may even diminish discrepancies between our models and *Drosophila*'s behavior.

The approach of this study is also relevant to vertebrate vision, where researchers typically model motion estimation using the motion energy model²⁵. Like the HRC, the motion energy model only responds to 2-point correlations in the visual stimulus. Consequently, many of the theoretical considerations in this paper apply directly to the motion energy model. Furthermore, each of our computational models can be straightforwardly generalized to the architecture of the motion energy model. For example, one could incorporate non-multiplicative nonlinearities by replacing the squaring operation of the motion energy model with a more flexible nonlinearity. Nevertheless, the numerical benefits offered by each modification to the motion energy model might differ from those found for the HRC because the motion energy model and HRC use distinct spatial and temporal filtering. Such differences could in principle manifest themselves as a different pattern of predicted glider responses^{43,45}, but comparative electrophysiology experiments in macaques and dragonflies currently suggest that similarities between primate and insect motion processing are abundant⁴⁸.

Our models make predictions that are testable with new experiments. Researchers hypothesize that the T4 and T5 neurons in the fly lobula nonlinearly combine visual inputs across space and time to become the first direction-selective neurons in *Drosophila*'s visual system²⁹. In accordance with the HRC model, conventional wisdom says that these neurons will multiply their input channels. In contrast, we predict that T4 and T5 will combine their visual input streams with non-multiplicative nonlinearities that facilitate accurate motion estimation in natural sensory environments. It's crucial to note that subtle differences between biology's

nonlinearity and a pure multiplication can correspond to substantial functional effects. In particular, the optimized nonlinearity that we found here (**Figure 4b**) is superficially similar a simple multiplication, yet its subtle distinctions manifest themselves by improving the local estimation accuracy of the HRC by an impressive margin (**Figure 2d**).

In this paper, we studied several simple models to most clearly illustrate the computational consequences of fundamental nonlinear circuit operations. Each of these operations individually provided a way for *Drosophila* to improve their motion estimation accuracy in natural environments, but they are not necessarily exclusive. For example, if a front-end nonlinearity does not fully remove the asymmetry between light and dark contrasts, then subsequent ON and OFF processing might further improve estimation accuracy. Similarly, non-multiplicative nonlinearities might enable an even better combination of ON and OFF signals for motion estimation. The general approach that we adopted here is to restrict the space of candidate models to those that have immediate biological relevance and to identify interesting models by optimizing the model's estimation accuracy over naturalistic stimuli. Future models should incorporate more biological details to better emulate the specifics of *Drosophila*'s visual circuitry, which is rapidly being dissected through unprecedented anatomical, functional, and behavioral experiments⁶⁵.

Materials and Methods

Simulated ensemble of naturalistic motions

We simulated the linear responses of neighboring photoreceptors to naturalistic motion using methods similar to previous work⁴³. We began with a database of natural images⁷¹. We converted each natural image to a contrast scale, $C(\vec{x}) = (I(\vec{x}) - I_0)/I_0$, where $C(\vec{x})$ is the

contrast at the spatial point \vec{x} , $I(\vec{x})$ is its intensity, and I_0 is the average intensity across the image. Since we only consider horizontal motion, we emulated the spatial blurring of *Drosophila*'s photoreceptors in the vertical dimension by filtering across rows with a Gaussian kernel (FWHM = 5.7°). We then took the central row of each filtered image to represent a one-dimensional variant of the natural image, denoted $c(x)$. We applied reflective boundary conditions to generate images that covered 360° and down-sampled each resulting image to 1° pixels by averaging. Photoreceptor blurring from signals in the horizontal dimension depends on the velocity of motion. In particular, we model the response of the i^{th} photoreceptor as

$$V_i(t) = \int dt' T(t') \int dx M(x - x_i) c(x - v(t - t')),$$

where T is a causal exponential kernel (timescale = 10 ms), M is a Gaussian kernel (FWHM = 5.7°), x_i is the location of the i^{th} photoreceptor, and v is the velocity of motion.

Each naturalistic motion comprised a randomly selected one-dimensional natural image, an offset to set the initial location of the photoreceptors, and a velocity drawn from a zero-mean normal distribution with a standard deviation of $90^\circ/\text{s}$. In this manner, we simulated the responses of three horizontally adjacent photoreceptors (spaced by 5.1°) to 5×10^5 naturalistic motions (each with duration = 800 ms, time step = 5 ms). We then explicitly enforced left-right symmetry in the naturalistic ensemble by pairing each naturalistic motion with a new simulated motion, in which the natural image is reflected, the velocity is inverted, and the offset is chosen such that $\{V_1(t), V_2(t), V_3(t)\}$ in the new naturalistic motion is exactly $\{V_3(t), V_2(t), V_1(t)\}$ from its partner. The final symmetric ensemble thus consists of 10^6 naturalistic motions.

The Hassenstein-Reichardt Correlator (HRC)

The HRC applies two temporal filters to its photoreceptor inputs. We denote the kernels of the low-pass and high-pass filters as f and g , respectively, such that the output of a local HRC is

$$R(t) = (f * V_1)(t)(g * V_2)(t) - (g * V_1)(t)(f * V_2)(t)$$

(Figure 1a). We consider the HRC's velocity estimate for a given naturalistic motion as its value at the final time point of the simulation. We model the filter kernels as

$$f(t) = te^{-t/\tau}, t \geq 0$$

and

$$g(t) = \frac{df(t)}{dt},$$

where $\tau = 20$ ms and $g(t)$ is comparable to lamina monopolar cell responses^{27,56}. We built the alternate motion estimators considered in this work from the same four filtered signals, $\{(f * V_1), (g * V_1), (f * V_2), (g * V_2)\}$, always considering the estimator's output at the final time point as its velocity estimate. Thus, none of our models modified the spatial or temporal processing of the HRC, reflecting our emphasis on how nonlinear processing might be tuned for naturalistic motion estimation. The global output of an array of HRCs would be obtained by pooling signals across space. Here we focus on spatiotemporally local strategies for motion estimation and at most pool motion signals across two neighboring motion detectors.

Relationship between the mean squared error and the correlation coefficient

We quantify motion estimators by the mean squared error between their output and the true velocity. To minimize the mean squared error of the HRC, we scale its output by $r^{(R)}\sigma_v/\sigma_R$, where $r^{(R)}$ is the correlation coefficient between the HRC's output and the velocity of motion, σ_v

615 is the standard deviation of the velocity distribution, and σ_R is the standard deviation of the
 616 HRC’s output. Once the HRC is scaled in this manner, its mean squared error is

$$\epsilon = \sigma_v^2 \left(1 - \left(r^{(R)}\right)^2\right).$$

617 More generally, this equation rewrites the mean squared error of any optimally scaled motion
 618 estimator in terms of its correlation coefficient with the velocity. All motion estimators
 619 considered in this paper are optimally scaled, and we find the correlation coefficient to be more
 620 intuitive than the mean squared error. We thus always report the performance of each motion
 621 estimator in terms of the correlation coefficient between the true and estimated velocity.

622

623 *Model fitting procedure*

624 We fit the linear weighting parameters in the models of **Figures. 1i, 3a, 4a, 4c and 4d** to
 625 maximize the estimation accuracy over a simulated ensemble of naturalistic motions. The
 626 formulas provided in subsequent sections of the **Materials and Methods** will cast each motion
 627 estimation scheme as a linear combination of a variety of motion predictors,

$$v_e = \sum_i w_i x_i,$$

628 where the x_i are nonlinear combinations of $\{(f * V_1), (g * V_1), (f * V_2), (g * V_2), (f * V_3), (g * V_3)\}$
 629 that depend on the model architecture, and the w_i are associated weighting coefficients. We
 630 chose the weights to minimize the mean squared error between the true and predicted velocity,
 631 which is the standard scenario considered by ordinary least-squares regression. The same
 632 weights maximize the correlation coefficient between the true and predicted velocity, and we
 633 typically present model accuracies as correlation coefficients.

634 We used 2-fold cross-validation to protect against over-fitting. In particular, we randomly
 635 divided the ensemble of naturalistic motions into a training set of 500,000 symmetrically paired

examples and a testing set of the remaining 500,000 examples. We determined the weighting coefficients by minimizing the empirical error over the training set, and we reported accuracies over the test set. To estimate error bars for each model's accuracy, we computed twenty random divisions of the naturalistic motion ensemble and calculated the standard deviation of the estimation accuracy.

Model responses to glider stimuli

We generated 25 random instantiations of each glider stimulus considered by our previous experimental work (**Figure 1f**, duration = 3 s, update rate = 40 Hz, pixel size = 5°)^{43,45}. We evaluated the response of each model to these stimuli by averaging the outputs of 60 identical local motion estimators (each separated by 5.1°) over the last two seconds of visual stimulation. Glider predictions were equal and opposite for the left and right variants of the stimuli, so we pooled leftward and rightward stimuli in all figures (**Figure 1f** shows the rightward variants). We scaled each model's output such that the average response to the positive 2-point glider was 1. All figures associated with glider responses show the mean and standard error of each model's response across the 25 glider instantiations.

Front-end nonlinearity model

The model in **Figure 2a** replaces the linear photoreceptor signals, V_1 and V_2 , with nonlinear photoreceptor signals

$$y_i = h(V_i)$$

where h is some nonlinear function. Thus, the motion estimate from the front-end nonlinearity model is

$$F = (f * y_1)(g * y_2) - (g * y_1)(f * y_2).$$

To implement the contrast equalizing nonlinearity, we replaced values of $V_i(t)$ by their rank-order (scaled and shifted to range between -1 and +1). Note that all $V_i(t)$ were sorted together (*i.e.* including all spatial points, temporal points, and simulated naturalistic motions). When multiple $V_i(t)$ had the same value, they were given the same rank. To implement binarizing nonlinearities, we again sorted the $V_i(t)$ and found the values corresponding to the threshold locations. For example, to calculate the binarizing nonlinearity with two steps (**Appendix IV**): (i) we found the $V_i(t)$ values corresponding to the 25th and 75th percentiles; (ii) signals below the 25th percentile or above the 75th percentiles were assigned the value of -1; and (iii) signals between 25th and 75th percentiles were assigned the value of +1. To implement the Gaussianizing nonlinearity, we again rank-ordered the $V_i(t)$ (scaled to range between 0 and 1) and applied the inverse Gaussian cumulative distribution function to these ranks. The HRC is the special case of this model where the front-end nonlinearity is linear.

Weighted 4-quadrant model

The weighted 4-quadrant model in **Figure 3a** separately correlates bright and dark signals. Mathematically, it is

$$Q = \sum_{a \in \{+, -\}} \sum_{b \in \{+, -\}} w_{ab}^{(Q)} Q_{ab},$$

where $w_{ab}^{(Q)}$ are adjustable weights that parameterize the model,

$$Q_{ab} = [(f * V_1)]_a [(g * V_2)]_b - [(g * V_1)]_b [(f * V_2)]_a,$$

$[x]_+$ is equal to x when x is positive and zero otherwise, and $[x]_-$ is equal to x when x is negative and zero otherwise. The HRC is the special case of this model where $w_{++}^{(Q)} = w_{+-}^{(Q)} = w_{-+}^{(Q)} = w_{--}^{(Q)}$.

Non-multiplicative nonlinearity model

The non-multiplicative nonlinearity model in **Figure 4a** replaces the HRC's multiplication step with a more flexible two-dimension nonlinearity. In particular, it is

$$N = \eta((f * V_1), (g * V_2)) - \eta((f * V_2), (g * V_1)),$$

where we approximate the nonlinearity, η , as a fourth-order polynomial

$$\eta(x, y) = \sum_{i=0}^4 \sum_{j=0}^{4-i} w_{ij}^{(N)} x^i y^j,$$

and $w_{ij}^{(N)}$ are adjustable weights that parameterize the model. We include terms up to fourth order in this model to ensure that it is flexible enough to describe the published glider response data. In particular: (i) the second-order terms accommodate responses to 2-point glider stimuli; (ii) the third-order terms accommodate parity-inverting responses to 3-point glider stimuli; and (iii) the fourth-order terms enable the model to respond with unequal magnitude to positive and negative parity 3-point glider stimuli (**Figure 3- figure supplement 1**). Thus, this model has 14 parameters. The HRC is the special case of this model where only $w_{11}^{(N)}$ is nonzero.

Unrestricted nonlinearity model

Here we model the 4-dimensional nonlinearity in **Figure 4c** as a fourth-order polynomial of the four filtered signals in the HRC. In general, this motion estimator is

$$S = \sum_{i=0}^4 \sum_{j=0}^{4-i} \sum_{k=0}^{4-i-j} \sum_{l=0}^{4-i-j-k} w_{ijkl}^{(S)} (f * V_1)^i (g * V_1)^j (f * V_2)^k (g * V_2)^l,$$

where $w_{ijkl}^{(S)}$ are adjustable weights that parameterize the model, and we set $w_{0000}^{(S)} = 0$ because this term has no utility for naturalistic motion estimation. Thus, this model has 69 parameters. The HRC is the special case of this model where $w_{1001}^{(S)} = -w_{0110}^{(S)} \neq 0$, and all other parameters are zero.

Extra input nonlinearity model

Here we model the 6-dimensional nonlinearity in **Figure 4d** as a fourth-order polynomial of the six filtered signals in two neighboring HRCs. In general, this motion estimator is

$$E = \sum_{i=0}^4 \sum_{j=0}^{4-i} \sum_{k=0}^{4-i-j} \sum_{l=0}^{4-i-j-k} \sum_{m=0}^{4-i-j-k-l} \sum_{n=0}^{4-i-j-k-l-m} w_{ijklmn}^{(E)} (f * V_1)^i (g * V_1)^j \\ \times (f * V_2)^k (g * V_2)^l (f * V_3)^m (g * V_3)^n,$$

where $w_{ijklmn}^{(E)}$ are adjustable weights that parameterize the model, and we set $w_{000000}^{(E)} = 0$ because this term has no utility for naturalistic motion estimation. Thus, this model has 209 parameters. The average of two neighboring HRCs is the special case of this model where $w_{100100}^{(E)} = -w_{011000}^{(E)} = w_{001001}^{(E)} = -w_{000110}^{(E)} \neq 0$, and all other parameters are zero.

Lasso regression for predictor selection

Lasso regression augments the squared error with an L_1 penalty on nonzero weighting coefficients that favors sparse solutions⁶⁴. We used lasso regression to identify subsets of predictors that might enable accurate motion estimation (**Figure 5a**). Once we identified a predictor subset using lasso regression, we refit the nonzero model weights using ordinary least squares regression (*i.e.* without the weight penalty).

Acknowledgments

The authors thank Ruben Portugues, Haim Sompolinsky, and Florian Engert for helpful conversations. JEF acknowledges fellowship support from the Swartz Foundation. DAC acknowledges support from a Searle Scholar Award and the Smith Family Foundation.

References

- 1 Sejnowski, T. J. *et al.* Computational neuroscience. *Science* **241**, 1299-1306 (1988).
- 2 Sperry, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology* **43**, 482 (1950).
- 3 Reichardt, W. & Poggio, T. Visual control of orientation behaviour in the fly: Part I. A quantitative analysis. *Q. Rev. Biophys.* **9**, 311-375 (1976).
- 4 Orger, M. B. *et al.* Perception of Fourier and non-Fourier motion by larval zebrafish. *Nat. Neurosci.* **3**, 1128-1133 (2000).
- 5 Kalmus, H. Animals as mathematicians. (1964).
- 6 Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106 (1962).
- 7 Barlow, H. B. & Hill, R. M. Selective sensitivity to direction of movement in ganglion cells of the rabbit retina. *Science* (1963).
- 8 Buchner, E. Elementary movement detectors in an insect visual system. *Biol. Cybern.* **24**, 85-101 (1976).
- 9 Reichardt, W. *et al.* Figure-ground discrimination by relative movement in the visual system of the fly. *Biol. Cybern.* **46**, 1-30 (1983).
- 10 Nordström, K. *et al.* Insect detection of small targets moving in visual clutter. *PLoS Biol.* **4**, e54 (2006).
- 11 Zhang, Y. *et al.* The most numerous ganglion cell type of the mouse retina is a selective feature detector. *Proc. Natl. Acad. Sci.* **109**, E2391-E2398 (2012).
- 12 Gabbiani, F. *et al.* Computation of object approach by a wide-field, motion-sensitive neuron. *The Journal of neuroscience* **19**, 1122-1141 (1999).
- 13 Pick, S. & Strauss, R. Goal-Driven Behavioral Adaptations in Gap-Climbing *Drosophila*. *Curr. Biol.* **15**, 1473-1478 (2005).
- 14 Rogers, B. & Graham, M. Motion parallax as an independent cue for depth perception. *Perception* **8**, 125-134 (1979).
- 15 Srinivasan, M. *et al.* Range perception through apparent image speed in freely flying honeybees. *Visual neuroscience* **6**, 519-535 (1991).
- 16 Kral, K. Behavioural-analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behav. Processes* **64**, 1-12 (2003).
- 17 Rust, N. C. *et al.* How MT cells analyze the motion of visual patterns. *Nat. Neurosci.* **9**, 1421-1431 (2006).
- 18 Franz, M. O. & Krapp, H. G. Wide-field, motion-sensitive neurons and matched filters for optic flow fields. *Biol. Cybern.* **83**, 185-197 (2000).

- 19 Britten, K. H. *et al.* The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience* **12**, 4745-4765 (1992).
- 20 Hassenstein, B. & Reichardt, W. Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers *Chlorophanus*. *Zeitschrift für Naturforschung* **11**, 513–524 (1956).
- 21 Egelhaaf, M. & Borst, A. Transient and steady-state response properties of movement detectors. *JOSA A* **6**, 116-127 (1989).
- 22 Haag, J. *et al.* Fly motion vision is based on Reichardt detectors regardless of the signal-to-noise ratio. *Proc. Natl. Acad. Sci. USA* **101**, 16333 (2004).
- 23 Goetz, K. G. Flight control in *Drosophila* by visual perception of motion. *Biol. Cybern.* **4**, 199-208 (1968).
- 24 Borst, A. *et al.* Adaptation without parameter change: Dynamic gain control in motion detection. *Proc. Natl. Acad. Sci.* **102**, 6172-6176 (2005).
- 25 Adelson, E. & Bergen, J. Spatiotemporal energy models for the perception of motion. *JOSA A* **2**, 284-299 (1985).
- 26 Van Santen, J. P. H. & Sperling, G. Elaborated reichardt detectors. *J. Opt. Soc. Am. A* **2**, 300-320 (1985).
- 27 Clark, D. A. *et al.* Defining the Computational Structure of the Motion Detector in *Drosophila*. *Neuron* **70**, 1165-1177 (2011).
- 28 Joesch, M. *et al.* ON and OFF pathways in *Drosophila* motion vision. *Nature* **468**, 300-304 (2010).
- 29 Maisak, M. S. *et al.* A directional tuning map of *Drosophila* elementary motion detectors. *Nature* **500**, 212-216 (2013).
- 30 Ammer, G. *et al.* Functional Specialization of Neural Input Elements to the *Drosophila* ON Motion Detector. *Curr. Biol.* (2015).
- 31 Takemura, S.-y. *et al.* A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* **500**, 175-181 (2013).
- 32 Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193-1216 (2001).
- 33 Juusola, M. & Hardie, R. C. Light Adaptation in *Drosophila* Photoreceptors I. Response Dynamics and Signaling Efficiency at 25° C. *The Journal of general physiology* **117**, 3-25 (2001).
- 34 Ratliff, C. P. *et al.* Retina is structured to process an excess of darkness in natural scenes. *Proc. Natl. Acad. Sci.* **107**, 17368-17373 (2010).
- 35 Olshausen, B. A. & Field, D. J. Natural image statistics and efficient coding*. *Network: Comput. Neural Syst.* **7**, 333-339 (1996).
- 36 Bell, A. & Sejnowski, T. The “independent components” of natural scenes are edge filters. *Vision Res.* **37**, 3327 (1997).
- 37 Tkačik, G. *et al.* Local statistics in natural scenes predict the saliency of synthetic textures. *Proc. Natl. Acad. Sci.* **107**, 18149-18154 (2010).
- 38 Hermundstad, A. M. *et al.* Variance predicts salience in central sensory processing. *eLife* **3**, e03722 (2014).
- 39 Yu, Y. *et al.* Visual processing of informative multipoint correlations arises primarily in V2. *eLife* **4**, e06604 (2015).
- 40 Geisler, W. S. Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* **59**, 167-192 (2008).

798 41 Potters, M. & Bialek, W. Statistical mechanics and visual signal processing. *Journal de*
799 *Physique I* **4**, 1755-1775 (1994).

800 42 Fitzgerald, J. E. *et al.* Symmetries in stimulus statistics shape the form of visual motion
801 estimators. *Proc. Natl. Acad. Sci.* **108**, 12909-12914 (2011).

802 43 Clark, D. A. *et al.* Flies and humans share a motion estimation strategy that exploits natural
803 scene statistics. *Nat. Neurosci.* **17**, 296-303 (2014).

804 44 Chubb, C. & Sperling, G. Drift-balanced random stimuli: a general basis for studying non-
805 Fourier motion perception. *JOSA A* **5**, 1986-2007 (1988).

806 45 Hu, Q. & Victor, J. D. A set of high-order spatiotemporal stimuli that elicit motion and
807 reverse-phi percepts. *Journal of vision* **10** (2010).

808 46 Quenzer, T. & Zanker, J. Visual detection of paradoxical motion in flies. *Journal of*
809 *Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*
810 **169**, 331-340 (1991).

811 47 Zanker, J. M. Theta motion: a paradoxical stimulus to explore higher order motion
812 extraction. *Vision Res.* **33**, 553-569 (1993).

813 48 Nitzany, E. I. *et al.* in *Cosyne* (Salt Lake City, 2014).

814 49 Katsov, A. & Clandinin, T. Motion processing streams in *Drosophila* are behaviorally
815 specialized. *Neuron* **59**, 322-335 (2008).

816 50 Dror, R. O. *et al.* Accuracy of velocity estimation by Reichardt correlators. *JOSA A* **18**, 241-
817 252 (2001).

818 51 Götz, K. & Wenking, H. Visual control of locomotion in the walking fruitfly *Drosophila*. *J.*
819 *Comp. Physiol. A* **85**, 235-266 (1973).

820 52 Nitzany, E. I. & Victor, J. D. The statistics of local motion signals in naturalistic movies.
821 *Journal of vision* **14**, 10 (2014).

822 53 Laughlin, S. B. A simple coding procedure enhances a neuron's information capacity. *Z.*
823 *Naturforsch* **36**, 51 (1981).

824 54 Brinkworth, R. S. & O'Carroll, D. C. Robust models for optic flow coding in natural scenes
825 inspired by insect biology. *PLoS Comp. Biol.* **5**, e1000555 (2009).

826 55 DeCarlo, L. T. On the meaning and use of kurtosis. *Psychological methods* **2**, 292 (1997).

827 56 Behnia, R. *et al.* Processing properties of ON and OFF pathways for *Drosophila* motion
828 detection. *Nature* **512**, 427-430 (2014).

829 57 Strother, J. A. *et al.* Direct Observation of ON and OFF Pathways in the *Drosophila* Visual
830 System. *Curr. Biol.* **24**, 976-983 (2014).

831 58 Meier, M. *et al.* Neural Circuit Components of the *Drosophila* OFF Motion Vision Pathway.
832 *Curr. Biol.* **24**, 385-392 (2014).

833 59 Eichner, H. *et al.* Internal structure of the fly elementary motion detector. *Neuron* **70**, 1155-
834 1164 (2011).

835 60 Joesch, M. *et al.* Functional Specialization of Parallel Motion Detection Circuits in the Fly. *J.*
836 *Neurosci.* **33**, 902-905 (2013).

837 61 Silies, M. *et al.* Modular use of peripheral input channels tunes motion-detecting circuitry.
838 *Neuron* **79**, 111-127 (2013).

839 62 Poggio, T. & Reichardt, W. On the representation of multi-input systems: computational
840 properties of polynomial algorithms. *Biol. Cybern.* **37**, 167-186 (1980).

841 63 Marr, D. & Poggio, T. From understanding computation to understanding neural circuitry.
842 (1976).

- 64 Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288 (1996).
- 65 Silies, M. *et al.* Motion-detecting circuits in flies: Coming into view. *Annu. Rev. Neurosci.* **37**, 307-327 (2014).
- 66 Tuthill, J. C. *et al.* Neural correlates of illusory motion perception in *Drosophila*. *Proc. Natl. Acad. Sci.* **108**, 9685-9690 (2011).
- 67 Westheimer, G. The ON-OFF dichotomy in visual processing: From receptors to perception. *Progress in retinal and eye research* **26**, 636-648 (2007).
- 68 Schiller, P. H. The ON and OFF channels of the visual system. *Trends Neurosci.* **15**, 86-92 (1992).
- 69 Card, G. & Dickinson, M. H. Visually Mediated Motor Planning in the Escape Response of *Drosophila*. *Curr. Biol.* **18**, 1300-1307 (2008).
- 70 Tammero, L. F. & Dickinson, M. H. Collision-avoidance and landing responses are mediated by separate pathways in the fruit fly, *Drosophila melanogaster*. *J. Exp. Biol.* **205**, 2785-2798 (2002).
- 71 van Hateren, J. H. & van der Schaaf, A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 359-366 (1998).
- 72 Laughlin, S. B. The role of sensory adaptation in the retina. *J. Exp. Biol.* **146**, 39-62 (1989).
- 73 Van Hateren, J. & Snippe, H. Phototransduction in primate cones and blowfly photoreceptors: different mechanisms, different algorithms, similar response. *J. Comp. Physiol. A* **192**, 187-197 (2006).
- 74 Juusola, M. & Hardie, R. C. Light Adaptation in *Drosophila* Photoreceptors II. Rising Temperature Increases the Bandwidth of Reliable Signaling. *The Journal of general physiology* **117**, 27-42 (2001).
- 75 Koch, C. *Biophysics of computation: information processing in single neurons*. (Oxford university press, 2004).

Table 1: The different models used in this paper, experimental results that support each model, and references for those results.

Model	Supporting Evidence
Front-end Nonlinearity	<ul style="list-style-type: none"> Photoreceptors show nonlinear responses to contrast changes^{33,72-74} Some neurons in the early visual system have nonlinear responses that make their output signals nearly uniform⁵³
Weighted 4 Quadrant Model	<ul style="list-style-type: none"> Visual processing is divided early into ON and OFF channels^{27,28,56-58} The two output channels (T4/T5) are sensitive to light and dark edges²⁹, but their inputs are incompletely rectified⁵⁶ Stimuli targeting the four quadrants are differentially represented in neural substrates^{27,60}
Non-multiplicative Nonlinearity	<ul style="list-style-type: none"> Pure multiplication is not a trivial neural operation⁷⁵ Inputs to T4/T5 are nonlinearly transformed⁵⁶, which also contributes to the biologically implemented non-multiplicative nonlinearity
Unrestricted Nonlinearity	<ul style="list-style-type: none"> The direction-selective neurons T4 receives inputs from more than two types of neurons^{30,31} T4 receives inputs from both its major input channels at overlapping points in space³¹
Extra Input Nonlinearity	<ul style="list-style-type: none"> The direction-selective neuron T4 receives inputs from more than two discrete retinotopic locations³¹

Figure legends

Figure 1. The Hassenstein-Reichardt correlator (HRC) model is an incomplete description of *Drosophila*'s motion estimator **(a)** Diagram of the HRC model. **(b)** We assessed motion estimation performance across an ensemble of naturalistic motions, each of which consisted of a natural image⁷¹ and a velocity chosen from a normal distribution. **(c)** We quantified model accuracy by comparing the model response to the true velocity using the mean squared error. **(d)** We summarized the error with the correlation coefficient between the model output and the true velocity. **(e)** In previous work⁴³, we used a panoramic display to measure the rotational responses of *Drosophila* to visual stimuli. **(f)** We presented flies with binary stimuli called gliders⁴⁵, which imposed specific 2-point and 3-point correlations⁴³. **(g)** Flies turned in response to 3-point glider stimuli, but these responses cannot be predicted by the standard HRC. **(h)** Diagram of the converging 3-point correlator, which is designed to detect higher-order motion signals like those found in 3-point glider stimuli. **(i)** Adding the converging 3-point correlator to the HRC improved motion estimation performance with naturalistic inputs. We optimized weighting coefficients to minimize the mean squared error over the ensemble of naturalistic motions and used cross-validation to protect against over-fitting. **(j)** This model predicted that *Drosophila* would weakly turn in response to 3-point glider stimuli.

Figure 2. Front-end nonlinearities improved naturalistic motion estimation but did not reproduce the psychophysical results. **(a)** Diagram of the front-end nonlinearity model. The nonlinearity occurs after the spatiotemporal filtering of photoreceptors but before the temporal filtering of the HRC. **(b)** The distribution of contrast signals after photoreceptor filtering had a kurtosis of 9.6. The kurtosis of unfiltered pixels in the image database was 7.8. **(c)** Three different nonlinearities

that transformed this input distribution into a Gaussian distribution, a uniform distribution, and a binary distribution. **(d)** After these transformations, the kurtosis of the contrast signal was reduced to 3, 1.8, and 1, respectively. **(e)** Each front-end nonlinearity model improved the HRC's estimation accuracy, and uniform output signals worked best. **(f,g)** The front-end nonlinearity models reproduced the sign of the negative 2-point glider psychophysical responses but did not reproduce the pattern of psychophysical responses to 3-point gliders.

Figure 3. The weighted 4-quadrant model improved estimation performance and reproduced the directionality of psychophysical results. **(a)** Diagram of the weighted 4-quadrant model. Similar to ON/OFF processing in the visual system, the weighted 4-quadrant model splits the four differentially filtered signals into positive and negative components. As in the HRC, these component signals are paired, multiplied, and subtracted to produce four mirror anti-symmetric signals. We refer to these signals as HRC-quadrants. The model output is a weighted sum of the quadrant signals. We identify quadrants by whether they respond to the positive or negative components of each filtered signal and denote the four quadrants as $(+ +)$, $(+ -)$, $(- +)$, and $(- -)$. In this notation, the first index refers to the sign of the low-pass filtered signal (emanating from $f(t)$), and the second refers to the high-pass filtered signal (emanating from $g(t)$). **(b)** We measured the response of each quadrant to naturalistic motions and chose the quadrant weightings to minimize the mean squared error between the model output and the true velocity. **(c)** Comparison of the estimation performance of individual quadrants, multiple quadrants, and the HRC. The best two quadrants were $(- -)$ and $(- +)$; the best three also included $(+ -)$. **(d)** The performance-optimized weighted 4-quadrant model reproduced the signs and approximate magnitudes of the psychophysical results.

Figure 4. Several biologically motivated generalizations of the motion estimator further improved estimation performance without sacrificing glider responses. See **Table 1** for a description of the biological rationales behind these models. **(a)** The “non-multiplicative nonlinearity” model substitutes a 2-dimensional nonlinearity for the pure multiplication of the HRC. Here, we approximated the nonlinearity with a 4th order polynomial. **(b)** Two-dimensional nonlinearities underlying the HRC, the weighted 4-quadrant model, and the non-multiplicative nonlinearity model. The latter models reflect optimized cases, in which the weighting coefficients maximized estimation performance with natural inputs. Iso-output lines are shown in each plot, and the horizontal and vertical limits are chosen to include 95% of the naturalistic input signals. **(c)** Another generalization, the “unrestricted nonlinearity” model allows all 4 input signals to be combined nonlinearly. We approximate this 4-dimensional nonlinearity with a 4th order polynomial. **(d)** A final generalization, the “extra input nonlinearity” model, relaxes the restriction that the motion estimator only uses 2 spatial inputs. We approximate this 6-dimensional nonlinearity with a 4th order polynomial. **(e)** Comparison of the estimation performance of these models to the HRC. We compare the extra input nonlinearity model to the average of two neighboring motion estimators. **(f, g)** The three models correctly predicted the directions of psychophysical responses. The pattern of 3-point responses differed somewhat across the models, and the extra input nonlinearity model was the first to predict a large asymmetry between positive and negative 3-point glider responses.

Figure 5. Computational interpretation of the extra input nonlinearity model. **(a)** We used lasso regression to select subsets of predictors that might enable accurate estimation (see **Materials and Methods**). With only 16 predictors, the model improved naturalistic performance over the

HRC by 68%, and including fewer than half of the predictors improved it by the full 92%. The maximum number of predictors corresponds to the number of polynomial coefficients that were fit in the full model. **(b)** We visualized the 6-dimensional nonlinearity as the sum of several simpler computational modules. When only 16 predictors were used (red bar in (a)), the model used four distinct types of computations. In particular, the model included nearest-neighbor and next-nearest-neighbor non-multiplicative nonlinearities (*top row*). It also included a converging 3-point correlator from the two furthest photoreceptors and a 4-point correlator that combined three spatial inputs (*bottom row*). **(c)** Venn diagram illustrating the hierarchical nesting of models used in this paper. All models in this paper contain sets of parameters that reproduce the HRC (gray dot). The weighted 4-quadrant model is a subset of non-multiplicative nonlinearity models, which are themselves a subset of unrestricted nonlinearity models. The extra input nonlinearity encompasses all the models. When we approximated the nonlinearities with 4th order polynomials, we restricted them to a smaller portion of the model space. The 4-quadrant nonlinearities only overlapped with the 4th-order polynomial approximation at the HRC, because the weighted 4-quadrant model is infinite order when expanded as a polynomial (see **Appendix VII**).

Supplement Figure legends

Figure 2-figure supplement 1. Front-end nonlinearities modify the correlations present in natural scenes. **(a)** Example images with no front-end nonlinearity (*top*), with an equalizing front-end nonlinearity (*middle*), and with a binarizing front-end nonlinearity (*bottom*). **(b)** The covariance between contrasts at 2 horizontally separated points is plotted as a function of distance between the points. The binary nonlinearity attenuated spatial correlations.

Figure 3-figure supplement 1. Separate ON and OFF processing improved motion estimation by supplementing the HRC with odd-ordered correlations. **(a)** By summing and subtracting the four quadrants (*top labels, e.g. “++”*) in four different patterns, we isolated the contributions of various correlation types (*side labels, e.g. “odd”*) to the weighted 4-quadrant model (**Appendix VII**). For example, the uniform sum of the four quadrants is the HRC, and we denote this quadrant combination as “even=2.” (*top row of matrix*). The other three rows of the matrix define quadrant combinations that are sensitive to two different classes of third and higher odd-ordered correlations (“*odd*” and “*odd**” rows) and to fourth and higher even-ordered correlations (“*even>2*” row). The factor of 1/4 merely sets the magnitude of the quadrant contributions to match the formulas in **Appendix VII** and is without conceptual importance. **(b)** The “even=2” correlation class worked best in isolation. Nevertheless, the “even=2” and “odd” classes were highly synergistic (their weighted sum is notated “best 2”), and these classes together made the “odd*” and the “even > 2” classes irrelevant.

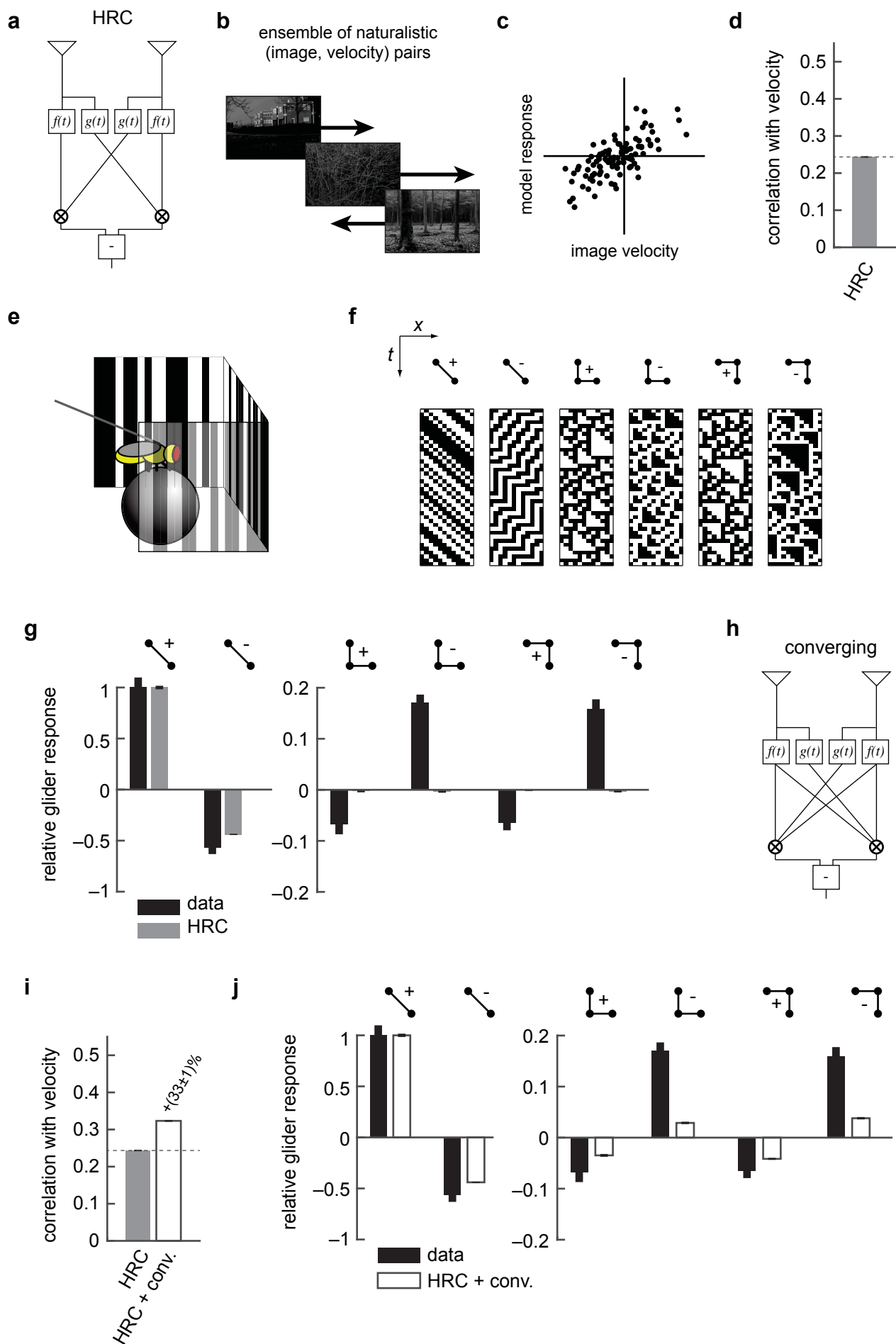
Figure 3-figure supplement 2. The weighted 4-quadrant model cannot reproduce the positive-negative parity asymmetry in the psychophysical data. In this numerical experiment, we tuned the coefficients of the weighted 4-quadrant model to optimize a fit to the psychophysical data. **(a)** The tuned model could reproduce the 2-point glider data well. **(b)** Although the tuned weighted 4-quadrant model could reproduce the signs of the 3-point glider data, it could not reproduce the differential amplitudes of the positive and negative parity responses. This demonstrates that the architecture of the weighted 4-quadrant model is too limited to reproduce the experimental response pattern.

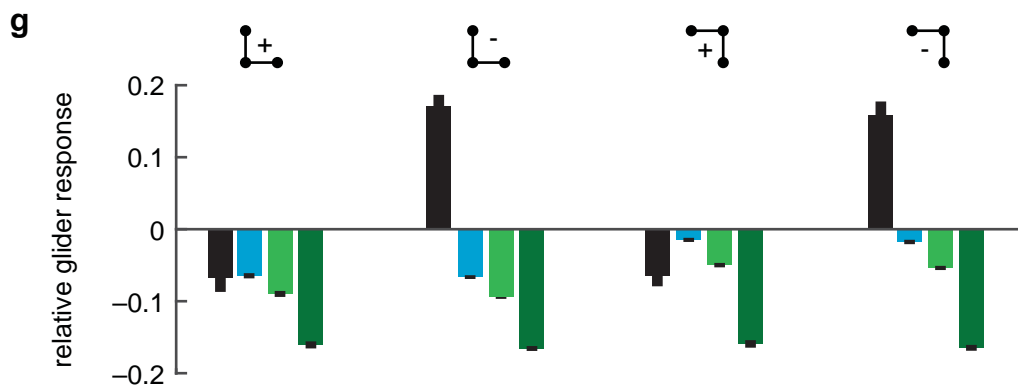
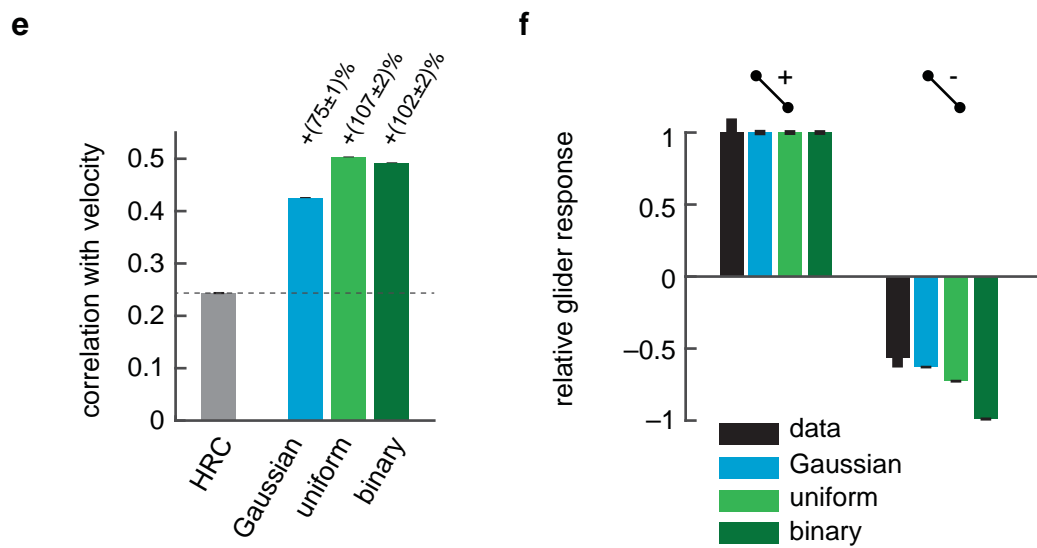
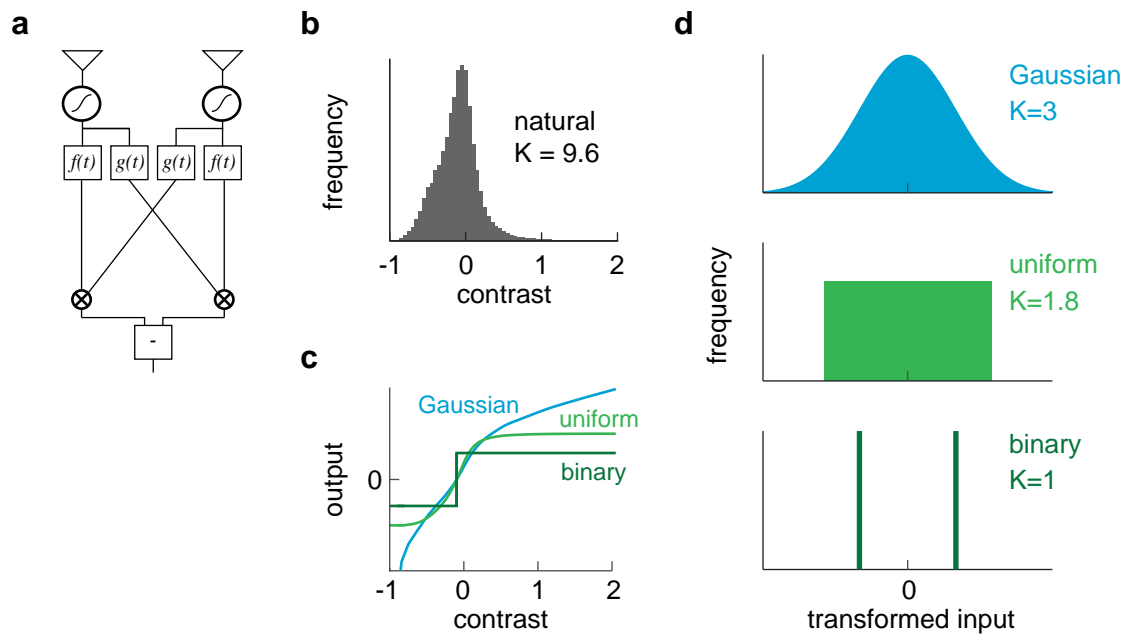
Figure 4-figure supplement 1. The non-multiplicative nonlinearity model can be tuned to account for the positive-negative parity asymmetry in the psychophysical data. In this numerical experiment, we tuned the model nonlinearity to optimize a fit to the psychophysical data. **(a)** In this case, the tuned model could reproduce the 2-point glider data well. **(b)** This tuned model could also reproduce the differential amplitudes of the positive and negative parity responses. Thus, the non-multiplicative nonlinearity model repairs an architectural defect of the weighted 4-quadrant model (**Figure 3-figure supplement 2**).

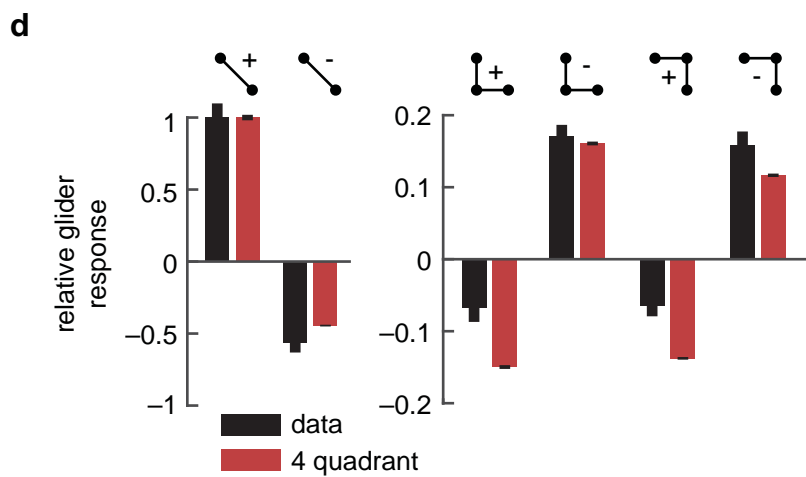
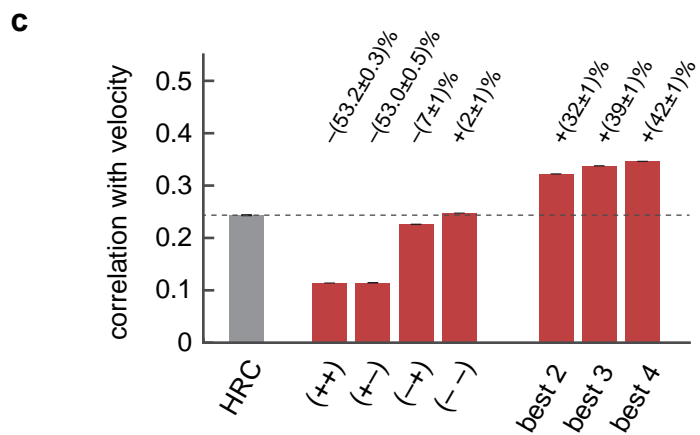
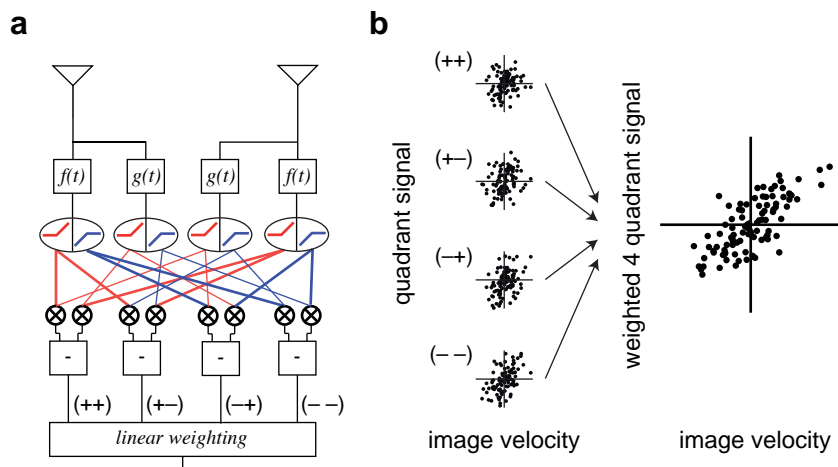
Figure 4-figure supplement 2. The performance of the non-multiplicative nonlinearity model is plotted against the order of the fitted polynomial. With only zeroth or first order terms, the model cannot predict motion. With second order terms, it can perform slightly better than the HRC (**Appendix X**). The biggest performance increase occurred when third-order terms were included, and the fourth-order terms also improved performance.

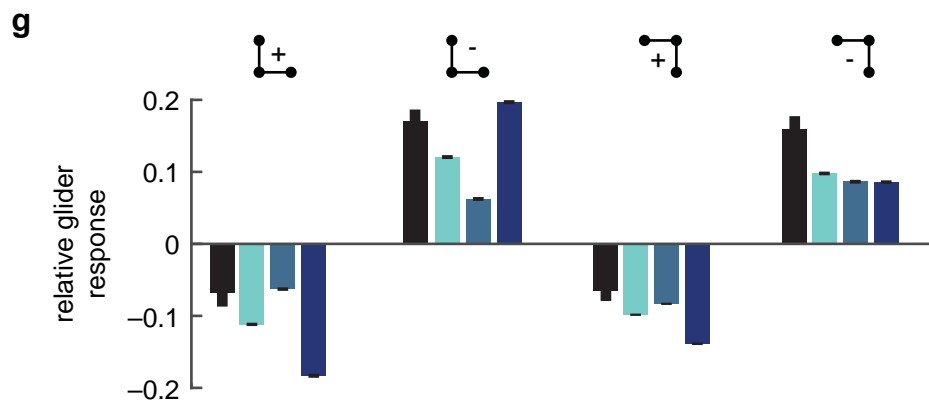
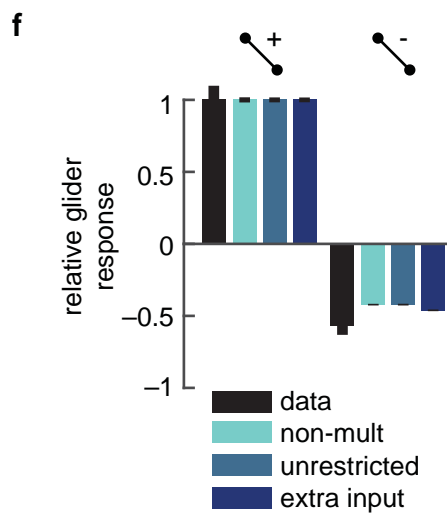
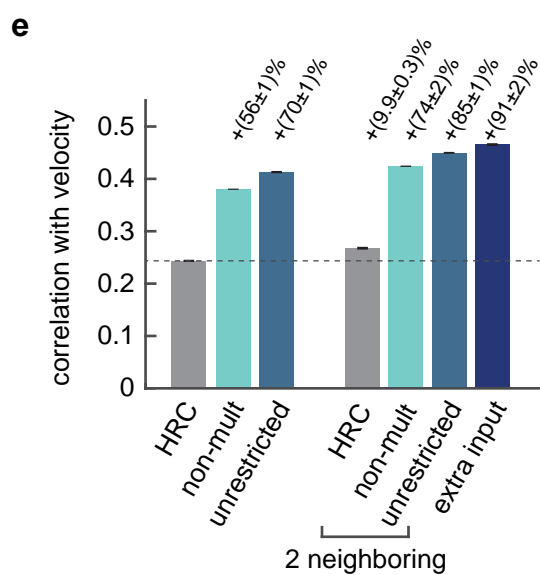
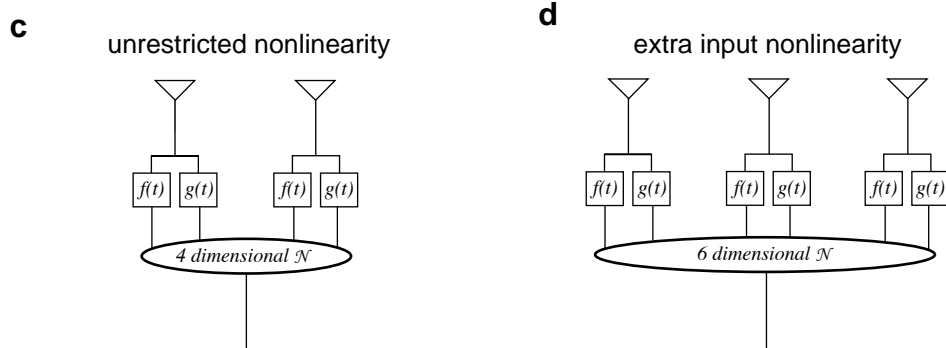
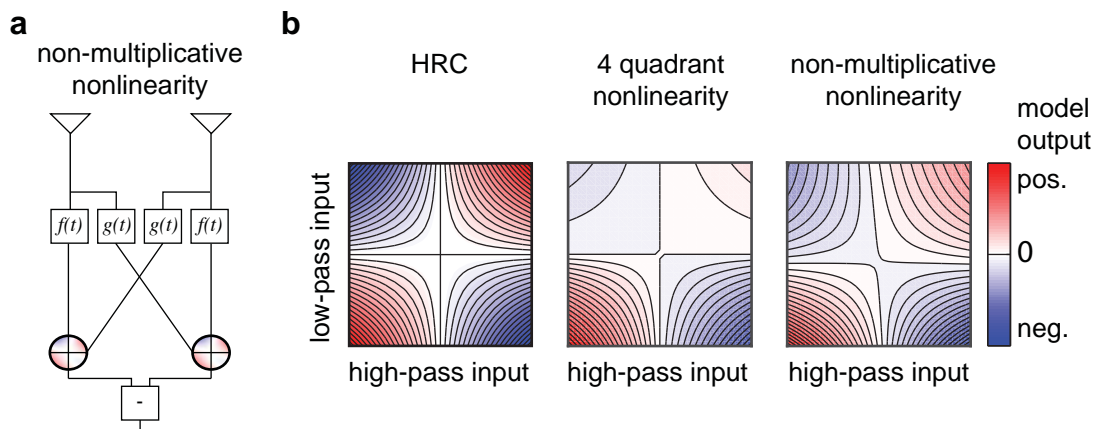
Figure 5-figure supplement 1. Structure of non-multiplicative nonlinearities in the extra input model of **Figure 5b**. **(a)** The nearest-neighbor non-multiplicative nonlinearity was made up of a standard HRC and a 3-point correlator in which the low-pass filtered input was squared before being multiplied by the adjacent receptor's high-pass filtered signal. **(b)** The next-nearest-neighbor non-multiplicative nonlinearity combined the analogous long-range terms. **(c)** Structure of the 2-dimensional nonlinearities, shown according to the same conventions as **Figure 4b**.

Supplementary File 1. Nonlinear Circuits for Naturalistic Visual Motion Estimation -
Supplementary Appendices

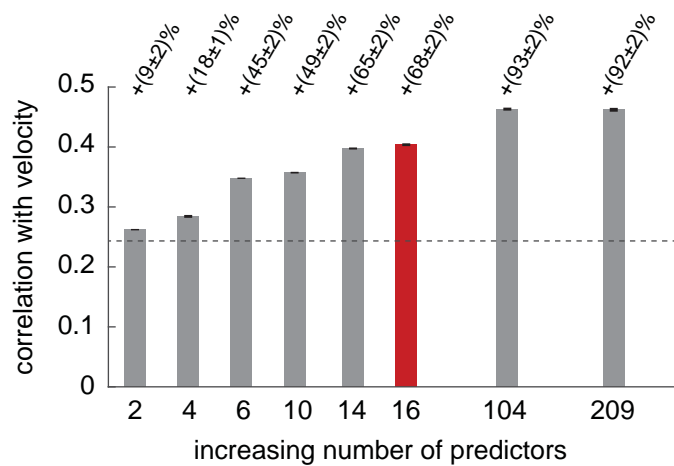




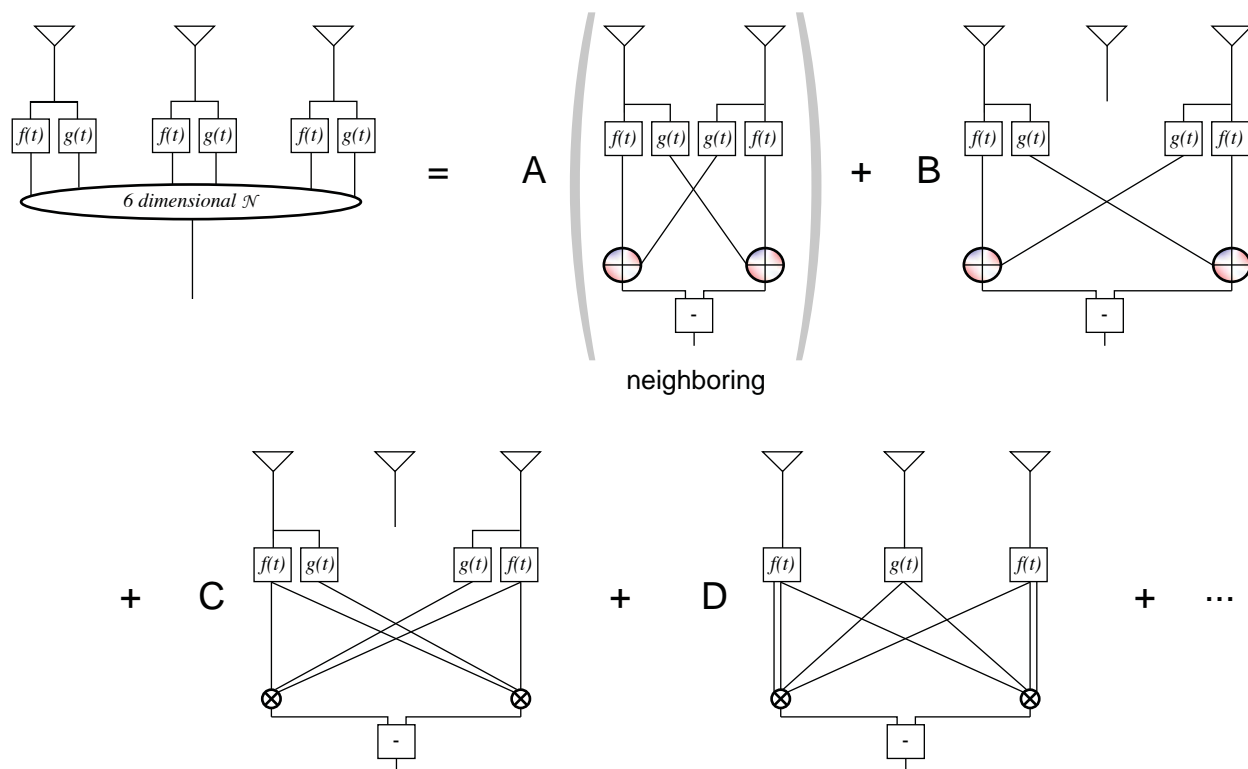




a



b



c

