Systems/Circuits

Direction Selectivity in *Drosophila* Emerges from Preferred-Direction Enhancement and Null-Direction Suppression

Jonathan Chit Sing Leong,[1]* Jennifer Judson Esch,[1]* Ben Poole,[2]* Surya Ganguli,[3] and Thomas Robert Clandinin[1]

[1]Department of Neurobiology, School of Medicine, [2]Department of Computer Science, and [3]Department of Applied Physics, Stanford University, Stanford, California 94305

Across animal phyla, motion vision relies on neurons that respond preferentially to stimuli moving in one, preferred direction over the opposite, null direction. In the elementary motion detector of *Drosophila*, direction selectivity emerges in two neuron types, T4 and T5, but the computational algorithm underlying this selectivity remains unknown. We find that the receptive fields of both T4 and T5 exhibit spatiotemporally offset light-preferring and dark-preferring subfields, each obliquely oriented in spacetime. In a linear-nonlinear modeling framework, the spatiotemporal organization of the T5 receptive field predicts the activity of T5 in response to motion stimuli. These findings demonstrate that direction selectivity emerges from the enhancement of responses to motion in the preferred direction, as well as the suppression of responses to motion in the null direction. Thus, remarkably, T5 incorporates the essential algorithmic strategies used by the Hassenstein–Reichardt correlator and the Barlow–Levick detector. Our model for T5 also provides an algorithmic explanation for the selectivity of T5 for moving dark edges: our model captures all two- and three-point spacetime correlations relevant to motion in this stimulus class. More broadly, our findings reveal the contribution of input pathway visual processing, specifically center-surround, temporally biphasic receptive fields, to the generation of direction selectivity in T5. As the spatiotemporal receptive field of T5 in *Drosophila* is common to the simple cell in vertebrate visual cortex, our stimulus-response model of T5 will inform efforts in an experimentally tractable context to identify more detailed, mechanistic models of a prevalent computation.

Key words: elementary motion detection; Hassenstein–Reichardt correlator; two-photon calcium imaging *in vivo*

Significance Statement

Feature selective neurons respond preferentially to astonishingly specific stimuli, providing the neurobiological basis for perception. Direction selectivity serves as a paradigmatic model of feature selectivity that has been examined in many species. While insect elementary motion detectors have served as premiere experimental models of direction selectivity for 60 years, the central question of their underlying algorithm remains unanswered. Using *in vivo* two-photon imaging of intracellular calcium signals, we measure the receptive fields of the first direction-selective cells in the *Drosophila* visual system, and define the algorithm used to compute the direction of motion. Computational modeling of these receptive fields predicts responses to motion and reveals how this circuit efficiently captures many useful correlations intrinsic to moving dark edges.

Introduction

Perception relies critically on neural circuits that give rise to stimulus selectivity. Across brain regions and animal phyla, astonishing stimulus selectivity emerges only a few synapses into the brain. Starting from highly complex patterns of neuronal activity at the sensory epithelium, what computational algorithms enable downstream neurons to respond to specific sensory events and not others? Motion vision provides a paradigmatic example of stimulus selectivity, requiring that neurons compare visual contrast inputs across space and time to respond more strongly

Received April 16, 2016; revised May 22, 2016; accepted May 25, 2016.

Author contributions: J.C.S.L., J.J.E., B.P., S.G., and T.R.C. designed research; J.C.S.L., J.J.E., B.P., and S.G. performed research; J.C.S.L., J.J.E., and B.P. contributed unpublished reagents/analytic tools; J.C.S.L., J.J.E., and B.P. analyzed data; J.C.S.L., J.J.E., B.P., S.G., and T.R.C. wrote the paper.

J.C.S.L. was supported by Stanford Neuroventures, Stanford Bio-X, and the Stanford Medical Scientist Training Program. J.J.E. was supported by the National Science Foundation Graduate Research Fellowship Program. B.P. was supported by Stanford Bio-X, Stanford Interdisciplinary Graduate Fellowship. S.G. was supported by Burroughs Wellcome, McKnight, James S. McDonnell, Simons, and Sloan Foundations. T.R.C. was supported by National Eye Institute R01 Grant EY022638 and National Science Foundation Grant 1353956. We thank R. Schneeveis, K. Schreck, and C. Thom for expert technical assistance and S. Baccus, Y. Fisher, S. Lahiri, N. Maheswaranathan, B. Naecker, B. Wandell, H. Yang, and F. Zenke for helpful discussion.

The authors declare no competing financial interests.

*J.C.S.L., J.J.E., and B.P. contributed equally to this study.

Correspondence should be addressed to Dr. Thomas Robert Clandinin, Stanford University, Fairchild D200, 299 West Campus Drive, Stanford, CA 94305. E-mail: trc@stanford.edu.

DOI:10.1523/JNEUROSCI.1272-16.2016

 0270-6474/16/368078-15$15.00/0

to motion in one preferred direction (PD) over motion in the opposite null direction (ND) (Clifford and Ibbotson, 2002). However, this computational problem does not have a unique solution; a direction-selective neural circuit may enhance responses to motion in the PD, or it may suppress responses to motion in the ND (Borst and Euler, 2011). Here we demonstrate that the most upstream direction-selective neurons in *Drosophila* actually implement both of these solutions.

Studies of direction-selective behavioral and neuronal responses in insects have suggested that the direction-selective visual circuits of these animals implement a Hassenstein–Reichardt correlator (HRC) (Hassenstein and Reichardt, 1956; McCann, 1973; Buchner, 1976; Reichardt and Poggio, 1976; for review, see Egelhaaf and Borst, 1993; Borst et al., 2010). In the most upstream direction-selective stage of the fly HRC (Egelhaaf and Borst, 1993; Eichner et al., 2011), rectified, low-pass filtered input from one photoreceptor is multiplied by rectified, but temporally unfiltered input from the spatially offset, neighboring photoreceptor. Low-pass filtering results in delayed and more sustained kinetics. The particular combination of differential kinetics and spatial offset of the input pathways aligns them along the spacetime trajectory of PD motion. Multiplication nonlinearly enhances the PD response. Indeed, calcium imaging of the most upstream direction-selective neurons in *Drosophila*, T4 and T5, has demonstrated precisely this functional signature: nonlinear enhancement of responses to PD motion (Fisher et al., 2015b).

Despite >60 years of behavioral and neurophysiological evidence in support of the HRC, its specific neural correlates in the fly brain have been revealed only more recently. In the fly motion-processing stream, third-order medulla neurons are contrast selective to either light or dark increments but are not direction selective (Behnia et al., 2014; Meier et al., 2014; Strother et al., 2014; Serbe et al., 2016; Yang et al., 2016). Directly postsynaptic to these medulla neurons, T4, which is selective for moving light edges, and T5, which is selective for moving dark edges, exhibit direction selectivity (Maisak et al., 2013; Fisher et al., 2015b). What neural circuit configuration transforms non–direction-selective inputs into direction-selective outputs? Recent structural studies have revealed systematic anatomical offsets between the cell types presynaptic to T4 and T5 (Takemura et al., 2013; Shinomiya et al., 2014). Given the visuotopic organization of the fly visual system, these anatomical offsets correspond to offsets in visual space, consistent with the input pathways to an HRC. A recent functional study characterized the spatial and temporal filtering properties of the four major cell types immediately presynaptic to T5, each of which was selective for moving dark edges and was necessary for wild-type direction selectivity in T5 (Serbe et al., 2016).

Absent from this emerging characterization of the functional circuit architecture underlying direction selectivity in the fly is a quantitative model for visually evoked responses of T4 and T5. Despite extensive anatomical and physiological characterization of the fly motion-processing stream, the algorithm transforming spatiotemporal visual input into the earliest direction-selective outputs of T4 and T5 remains unidentified. Here we examined the visually evoked activity of T4 and T5 with two-photon calcium imaging *in vivo*, mapping T4 and T5 spatiotemporal receptive fields based on responses to spatiotemporal noise, and measuring responses to motion stimuli. We found that the stimulus-evoked activity of T5 is well described by a linear-nonlinear (LN) stimulus-response model. This model implements both PD enhancement and ND suppression as direction selectivity algorithms and accounts quantitatively for the activity of T5 in response to moving sinusoidal gratings.

Materials and Methods

Fly husbandry and preparation. Drosophila melanogaster were raised on molasses-based food at 25°C. "Dense" expression of GCaMP6f in T4 and T5 was achieved in the genotype *+/+; UAS-GCaMP6f/UAS-GCaMP6f, UAS-mtdTomato; R42F06-Gal4/+*. Expression of GCaMP6f in sparse single-cell clones of T4 and/or T5 was achieved in the genotype *+/yw, hs-FLP; UAS-GCaMP6f/tub-FRT-Gal80-FRT; R42F06-Gal4* (Gordon and Scott, 2009; Gruntman and Turner, 2013). Stochastic Flp-out was induced by heat-shock at 37°C for 5–10 min during the late third-instar stage of development.

Female flies were used within 24 h of eclosion for all experiments. To prepare them for imaging, flies were immobilized by cooling on ice, and affixed to a custom-machined 304 steel shim (Trinity Brand Industries) with UV-cured optical epoxy (NOA 63, Norland Optical Adhesives), which was also applied to the legs to prevent obstruction of the field of view of the fly. The cuticle, fat bodies, and trachea of the left hemisphere were removed by manual dissection under ice-cold, low-calcium artificial hemolymph to expose the left optic lobe for imaging from above. The ventral aspect of the fly, including eyes and spiracles, remained dry. Finally, ice-cold, low-calcium artificial hemolymph was exchanged with an excess volume of standard artificial hemolymph (Wilson et al., 2004) at room temperature (20°C–22°C). Artificial hemolymph at room temperature superfused with carbogen was exchanged continuously at 150 mL/h.

Imaging. Fluorescence was monitored *in vivo* by two-photon microscopy. Typically <10 mW of laser excitation (Chameleon Ti-Sapphire, Coherent) at 925 nm was delivered to the specimen via a Leica SP5 MP microscope mated to an HCX APO L 20× 1.0 water-immersion objective (Leica). Red and green emission from tdTomato and GCaMP6f, respectively, were separated from one another with a dichroic mirror (560dcxr, Chroma Technologies) and bandpass filters (ET525/50m, HC585/40m, Chroma Technologies). PMT gain was set to 1000 V, and raw PMT signals were acquired in direct data transfer mode. Bidirectional scanning at 1.4 kHz was used to acquire rectangular images (128 pixels × 256 pixels, rows × columns) at ≈15 Hz. Pixels measured ≈290 nm × 290 nm, resulting in a field of view of ≈37 μm × 74 μm. The FWHM of the axial point-spread function was <2.5 μm.

Delivery of visual stimuli. Stimuli were programmed with the OpenGL 1.0 API in Visual C#, rendered on an Nvidia K2200 AGP graphics card, and displayed by rear-projection using a LightCrafter 4500 DLP (Texas Instruments) configured to use exclusively blue LED illumination. The stimulus was attenuated with 447/60 bandpass (Semrock) and ND1 filters (Thorlabs). The mean radiance was 0.04 W sr^{-1} m^{-2}. The stimulus screen (gray, Rose Brand) measured 9 cm × 9 cm and was positioned ≈180 mm from the DLP so that the full height of the DLP output spanned the screen. The stimulus screen was positioned 65 mm from the fly to subtend ≈60° × 60° (azimuth × elevation) of the left visual field. The DLP presented stimuli at native spatial resolution, 912 pixels × 1140 pixels (width × height), at 300 Hz with a period of ≈333 μs in between frames, and at 6 bit pixel depth. The stimulus frame was updated at 100 Hz. For each stimulus frame, the timestamp obtained from the system clock and all parameters needed to regenerate the stimulus frame were saved directly to disk. The stimulus and data acquisition computers were linked via NIDAQ 6211 (National Instruments), and timestamps obtained from the system clock were saved directly to disk for each data frame as well. Stimulus presentation and data acquisition, however, were asynchronous.

Experiment and stimulus design. Each experiment was composed of a series of up to five recordings. Per recording, a single stimulus type was presented. Per stimulus type, a series of stimulus conditions, each with a different set of stimulus parameters, was presented in a pseudorandom order, which was reshuffled for each of five runs through the stimulus conditions. Total experiment time did not exceed 2 h.

Unless otherwise noted, stimulus conditions were interleaved with a "blank" stimulus condition, comprised of uniform intermediate gray contrast for at least 2 s.

The following stimulus types were presented:

- "Full-field" contrast excursions of either maximal or minimal contrast for 3 s followed by full-field intermediate gray contrast for 3 s. Stimulus conditions were not interleaved with a blank.
- Sinusoidal gratings moving in one of eight evenly spaced directions, with full contrast, spatial frequency 1/30 cycles deg^{-1}, and temporal frequency 1 Hz.
- Edges of either maximal (light) or minimal (dark) contrast moving in one of the cardinal directions at 30°/s.
- Ternary XT noise. Each frame was composed of a 1×12 pixel array of rectangular pixels, each pixel measuring $60° \times 5°$ so that the stimulus spanned $60° \times 60°$. Each pixel was minimal, intermediate, or maximal contrast with equal probability independent of all other pixels. All pixel values were updated every 50 ms. Per experiment, ternary XT noise explored either azimuth or elevation.
- Sinusoidal gratings moving in a cardinal direction or its opposite at 10° s^{-1}, 25° s^{-1}, 50° s^{-1}, or 100° s^{-1}, with full contrast, and temporal frequency 1 Hz (1F). Per experiment, sinusoidal gratings explored either azimuth or elevation, matching the axis explored by ternary XT noise.

Data processing. For each recording, drift of the specimen within the imaging plane was modeled as a time series of Euclidean transformations estimated frame-wise using a parallelized adaptation of Robust Alignment by Sparse and Low-rank Decomposition applied to the red channel (Peng et al., 2012). Drift was corrected by applying the time series of inverse transformations to the data. Across recordings, drift of the specimen within the imaging plane was estimated per recording by applying Robust Alignment by Sparse and Low-rank Decomposition to the mean data frames of the within-frame corrected recordings. Drift across recordings was then corrected by applying each inverse transformation to all frames of its respective within-frame corrected recording. Experiments that included any axial drift appreciable by visual inspection either within a recording or across recordings were discarded.

Fluorescence time series $F(t)$ were high-pass filtered with a cutoff period of 150 data frames, or ≈10 s, averaged across pixels in the case of regions of interest (ROIs), and then normalized relative to the baseline fluorescence, F_0, to obtain the following:

$$\frac{\Delta F}{F_0} = \frac{F(t) - F_0}{F_0},$$

which is proportional to changes in calcium concentration. F_0 was calculated for each recording as the average fluorescence during the last few frames of all presentations of the blank stimulus in the recording. If no blank stimulus was presented, all stimulus conditions were used to calculate a "cocktail" blank. To accommodate asynchronous stimulus presentation and data acquisition, $\Delta F/F_0$ responses for all trials were converted to a stimulus-aligned time base before trial averaging.

For each stimulus condition s_i of a given stimulus type, the time series response $R_{s_i}(t)$ was typically summarized by a single scalar value R_{s_i}, for example, the maximum of the response, $R_{s_i} = \max R_{s_i}(t)$. Tuning curves refer to nonparametric maps between stimulus parameters and these response metrics. For stimuli s_θ uniformly spaced about the circle, the direction selectivity index (DSI) was computed as follows:

$$\mathrm{DSI} = \left| \frac{\sum_\theta R_{s_\theta} \exp(i\theta)}{\sum_\theta R_{s_\theta}} \right|,$$

where R_{s_θ} denotes the response to the stimulus s_θ moving in the direction θ. This DSI is a modification of the classical definition of circular variance, which was halved and then subtracted from 1 in order for a DSI of 0 to indicate no direction preference and a DSI of 1 to indicate responses exclusively in the PD (Batschelet, 1981). The contrast selectivity index (CSI) was computed as follows:

$$\mathrm{CSI} = \frac{R_{\mathrm{MAX}} - R_{\mathrm{NULL}}}{R_{\mathrm{MAX}} + R_{\mathrm{NULL}}},$$

where R_{MAX} and R_{NULL} denote, respectively, the maximal response evoked by any stimulus condition, and the response evoked by the same stimulus condition but with inverted contrast. Responsivity was assessed by an ANOVA performed across single-trial tuning curves, including the blank stimulus condition. Selectivity was assessed by an ANOVA performed across single-trial tuning curves, excluding the blank stimulus condition.

Identification and selection of ROIs. Based on raw fluorescence responses to sinusoidal gratings moving in one of eight evenly spaced directions, a peakiness measure was calculated per pixel, as follows:

$$\phi(x, y) = \left\langle \left| \frac{F - F_0}{F_0 + \langle F \rangle_{x,y,t}} \right| \right\rangle_t^{1.5},$$

where $\langle \cdot \rangle_{\cdot}$ denotes the average of the enclosed expression over the subscripted variables. This measure is related to $\Delta F/F_0$, and is analogous to previous peakiness measures (Ahrens et al., 2013) but for the weaker nonlinearity used to emphasize excursions in fluorescence, and the absolute value operation, which makes ϕ sensitive to both positive and negative excursions in fluorescence. The resulting peakiness pixel map was thresholded by Otsu's method (Otsu, 1979) to identify pixels eligible for further analysis. In practice, this procedure had the effect of removing background pixels where there was no tissue in the imaging plane.

Per pixel, direction tuning-curves of the 1F amplitude modulation and phase offset of the response were derived from trial-averaged $\Delta F/F_0$ responses to sinusoidal gratings moving in one of eight evenly spaced directions. These two eight-point tuning curves were then transformed into a 16-dimensional functional signature as follows:

$$(R_{0°}^{\mathrm{amplitude}} \cos R_{0°}^{\mathrm{phase}}, R_{0°}^{\mathrm{amplitude}} \sin R_{0°}^{\mathrm{phase}}, \ldots, R_{270°}^{\mathrm{amplitude}} \cos R_{270°}^{\mathrm{phase}}, R_{270°}^{\mathrm{amplitude}} \sin R_{270°}^{\mathrm{phase}})$$

where $R_\theta^{\mathrm{amplitude}}$ and $R_\theta^{\mathrm{phase}}$ denote the 1F amplitude modulation and phase offset, respectively, of the response to a sinusoidal grating moving in the direction θ. Based on this functional signature, the Euclidean distance between each pair of pixels was computed. Average-linkage agglomerative hierarchical clustering (Rokach and Maimon, 2005) of pixels based on these distances with an empirically chosen distance threshold, 0.5, yielded an initial set of ROIs. Occasionally, a ROI was largely if not completely surrounded by another ROI. In such events, the surrounding and surrounded ROIs were merged, although ROIs that had poor (correlation <0.3) or widely varying (variance >0.015) within-ROI pixel-pixel correlation were subsequently discarded. ROIs were also discarded based on their total size (<5 or >120 pixels), the size of their disconnected components (<3 pixels), and the number of their disconnected components (>8 regions). In practice, these criteria excluded morphologically unrealistic ROIs.

To isolate T5 single units, ROIs were further excluded from further analysis if their response properties did not match those of known T5 single units. Because T5 responds poorly to full-field contrast excursions (Maisak et al., 2013; Fisher et al., 2015a), ROIs were excluded from further analysis if they were responsive ($p < 0.01$) to this stimulus. Clustering pixels in a space of response properties guarantees that each cluster is defined by a response profile that is shared by all constituent pixels, but it does not guarantee that this response profile corresponds to a single unit. For example, a cluster could be defined by the mixture of response properties of multiple T4 and multiple T5 single units. T4 and T5, however, are strongly contrast selective for opposite contrasts. ROIs were excluded from further analysis if they were not comparable with T5 single-cell clones in terms of their contrast and direction selectivity index (CSI <0.6, DSI <0.6), which were computed based on maximum responses and 1F response amplitudes, respectively.

To avoid the analysis of multiple "duplicate" ROIs that correspond to the same T5 single unit, ROIs were clustered using single-linkage agglomerative hierarchical clustering (Rokach and Maimon, 2005) on the correlation between their spatiotemporal receptive fields (STRFs). Clusters were formed by cutting the dendrograms where the maximum intracluster correlations was >0.9, a conservative, empirically chosen

threshold. From each cluster, only the ROI with the highest SNR was selected for further analysis.

Receptive-field mapping. Responses to ternary noise were analyzed in the first instance by calculation of the fluorescence-weighted average stimulus, which we refer to as the STRF. Let $r(t)$ be the $\Delta F/F_0$-normalized response, $s(x, t)$ be the value of the stimulus at position x at time t. We computed the STRF as follows:

$$a(x, \tau) = \frac{1}{T - \tau} \sum_{t=\tau}^{T} r(t)\, s(x, t - \tau).$$

To compactly describe and denoise the STRFs, we fit difference of Gaussians (DoG) models with one positive and one negative component to each STRF. We initialized DoG fits with Gaussian components at the location of the minimum and maximum of the STRF. The initial widths were chosen to be 10° in space and 250 ms in time, with an angular orientation of 0 radians. To improve robustness to initialization, we fit models over 10 random perturbations of the initial parameters and chose the best-fitting model based on the Euclidean distance to the STRF.

Statistics of the STRF were computed from the linear filter corresponding to the best DoG approximation. Peak amplitudes were taken to be the maximum and minimum of the DoG filter. The DoG filter was split into an ON and OFF subfield by including the top 80% of positive and negative values, respectively. Temporal onset and spatial and temporal extents were computed for each subfield separately. The centroid of each subfield was used as the peak, and the slope of the axis between them used as the ON-OFF tilt, in units of $^\circ\ s^{-1}$. The tilts of the subfields were computed similarly as the slopes of their major axes.

Modeling. We modeled the mapping from stimulus to $\Delta F/F_0$ as an LN calcium response model followed by an LN indicator dynamics model. The first neuronal stage, the linear filter of the neuron, identifies a vector in XT stimulus space along which the neuron is sensitive. The second neuronal stage, a nonlinearity, enables otherwise linearly reinforcing or antagonizing inputs to enhance or suppress one another. The third stage, a linear filter over time, accounts for the delay of the binding kinetics of the calcium indicator. The final stage, a nonlinearity, reflects the cooperativity of calcium binding to the indicator.

The cell's linear filter, $w(x, \tau)$, depends on both the spatial location and temporal offset of the stimulus. The output of the linear filter is passed through a time-independent nonlinearity f to yield the time-dependent prediction of intracellular calcium concentration,

$$c(t) = f\left(\sum_{\tau=1}^{T_{\text{stim}}} \sum_{x=1}^{n_x} w(x, \tau)\, s(x, t - \tau)\right).$$

The predicted intracellular calcium concentration is then mapped to predicted $\Delta F/F_0$ through a temporal convolution with the filter, v, and an additional nonlinearity g, as follows:

$$r(t) = g\left(\sum_{\tau=1}^{T_{\text{calcium}}} v(\tau)\, c(t - \tau)\right).$$

We discretized visual space into 12 bins each with an extent of 5° to match the spatial resolution of the ternary noise stimulus. Time was discretized into 10 ms bins, matching the stimulus frame rate. The linear filter, $w(x, \tau)$, covers all 12 spatial bins, and extends 1 s (100 bins) in time containing a total of 1200 parameters.

The neuron nonlinearity f was parameterized as a smoothed half-wave rectifying nonlinearity raised to a power as follows:

$$f(x) = c \log(1 + \exp(ax + b))^k + d.$$

We constrained $a, c, d \geq 0$, so that $f(x) \geq 0$, allowing us to interpret the output of f as the predicted calcium response. We also explored a more flexible parameterization for f as a linear combination of Gaussian basis functions but found that it yielded worse performance.

The temporal convolution used to model the indicator dynamics was a difference of exponentials normalized to sum to 1 as follows:

$$v(t) = \frac{\exp\left(-\frac{t}{\tau_{\text{on}}}\right) - \exp\left(-\frac{t}{\tau_{\text{off}}}\right)}{\sum_{t'=1}^{T_{\text{calcium}}} \left(\exp\left(-\frac{t'}{\tau_{\text{on}}}\right) - \exp\left(-\frac{t'}{\tau_{\text{off}}}\right)\right)}.$$

The calcium nonlinearity was modeled using a polynomial with a learned exponent as follows:

$$g(y) = y^h + f.$$

Because of the nonconvex nature of the optimization problem, fits were highly sensitive to the initial parameters. To compensate for this sensitivity, we carefully initialized the parameters of our model. The linear filter of our model was initialized to the STRF, and additional parameters were initialized to the following: $\tau_{\text{on}} = 5$ ms, $\tau_{\text{off}} = 300$ ms, $a = 1, b = 0, c = 1, d = 0, k = 2, h = 2.3, e = 1, f = 0$.

Models were trained to minimize the mean squared error between the predicted and true responses. Because of the large number of parameters and small amount of data, we used additional regularization to prevent models from overfitting the training dataset. To encourage connectedness and sparsity in the cell's linear filter, we used a penalty combining total variation (TV) and the ℓ_1 norm, respectively (Chambolle et al., 2010), as follows:

$$\lambda_{\text{time}} \sum_{x,\tau} |w(x, \tau) - w(x, \tau + 1)| + \lambda_{\text{space}} \sum_{x,\tau} |w(x, \tau) - w(x + 1, \tau)| + \lambda_1 \sum_{x,\tau} |w(x, \tau)|.$$

Because of the anisotropic resolution in spacetime, we optimized over different weightings for the spatial and temporal components of the TV penalty, λ_{space} and λ_{time}. These weightings and additional hyperparameters were optimized over an additional held-out set of noise for one ROI, and then used across all ROIs.

Models were fit using only the responses to the ternary noise stimulus, with the first minute of the ternary noise stimulus held out for validation. We validated the models by comparing the measured and predicted responses to both held-out ternary noise and sinusoidal grating motion stimuli. The models used to predict the sinusoidal grating motion stimuli were trained exclusively on the ternary noise stimulus without ever being exposed to a structured motion stimulus.

Model-fitting code was implemented in Python using Theano and the Lasagne deep-learning framework (Bergstra et al., 2010; Bastien et al., 2012). Models were fit using minibatch optimization with the Adam optimization algorithm (Kingma and Ba, 2014) with a learning rate optimized through cross-validation, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. We performed 500 full passes through the dataset. Training datasets were split contiguously in time to form 10 equally sized minibatches, and parameters were fit using all 10 minibatches.

Results

The spatiotemporal receptive field of T4 and T5 single-cell clones

T4 and T5 extend light- and dark-edge selective pathways from the medulla and lobula neuropils into the lobula plate (Fig. 1*A*) (Fischbach and Dittrich, 1989; Maisak et al., 2013). To study the algorithm for direction selectivity in T4 and T5, we recorded the visually evoked calcium dynamics of T4 and T5 using *in vivo* two-photon microscopy of the genetically encoded calcium indicator GCaMP6f, which we targeted to these cells using the R42F06 driver line (Maisak et al., 2013; Fisher et al., 2015b). R42F06 drives strong expression in most T4 and T5 neurons, revealing strongly contrast- and direction-selective responses in T4 and T5 axon terminals in the lobula plate (Maisak et al., 2013). However, the neuronal processes of T4 and T5 are fine and interdigitating, such that the axon terminals of individual cells are not clearly segregated from one another in the dense R42F06 expression pattern (Fig. 1*B*). To achieve single-neuron resolution of T4

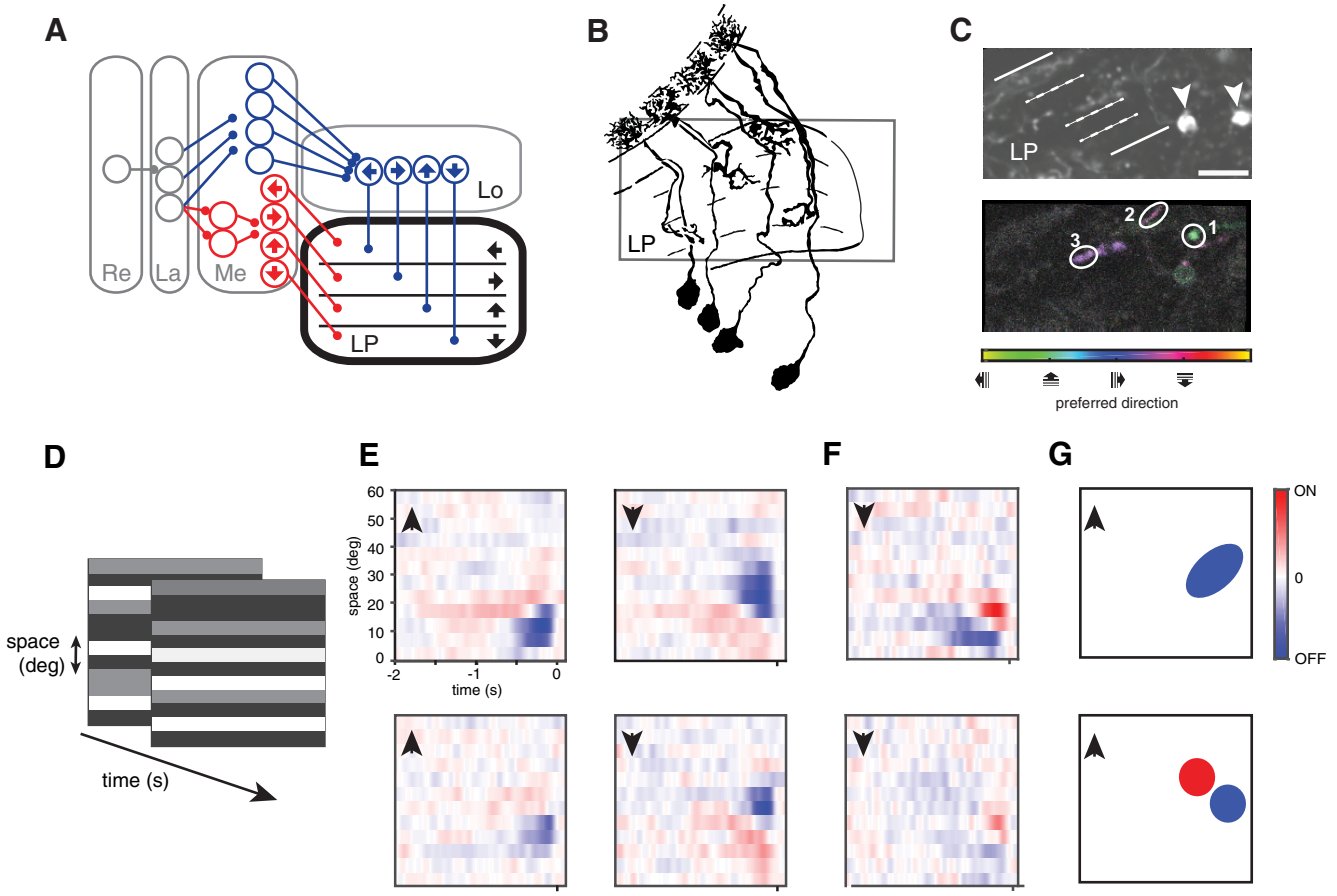

Figure 1. Tilted organization of ON and OFF subfields in the spatiotemporal receptive fields of genetically isolated T4 and T5 single-cell clones. ***A***, Schematic of the functional anatomy of motion-processing pathways in the fly. Photoreceptors in the retina (Re) synapse (terminated line segments) onto second-order interneurons in the lamina (La), which synapse onto third-order interneurons in the medulla (Me), which synapse onto T4 dendrites in the medulla and T5 dendrites in the lobula (Lo). T4 and T5 send their axons into the lobula plate neuropil (LP), where their axons are organized retinotopically across one axis and into four layers according to their cardinal direction preferences. Light and dark edge pathways shown in red and blue, respectively. Arrows indicate PD. ***B***, Illustration of the morphology of T4 and T5 single cells, indicating the approximate field of view (FOV) used for imaging in the lobula plate neuropil. ***C***, FOV containing single-cell clones isolated from the R42F06 expression pattern by stochastic flp-out. Top, Mean fluorescence image. Scale bar, 10 μm. Bottom, Hue-saturation value (HSV) map of the same FOV calculated from responses evoked by sinusoidal gratings moving in eight evenly spaced directions. For each pixel, hue indicates the PD (legend), saturation the DSI, and value the response amplitude (see Materials and Methods). Axons terminals from three different neurons, labeled, are seen in this FOV. The LP lies between the two solid white lines. White dashed lines indicate LP layers. Two T4 or T5 cell bodies (white arrowheads) are also visible in the FOV. ***D***, Two frames of an XT noise stimulus. Each frame is composed of twelve 5° × 60° bars. The contrast of each bar is updated every 50 ms to be black, gray, or white with equal probability. ***E***, STRFs of T5 single-cell clones for which the spatial axis of the stimulus was aligned with the axis of the PD. ***F***, STRFs of T4 single-cell clones for which the spatial axis of the stimulus was aligned with the axis of the PD. ***E***, ***F***, All STRFs exhibit two oppositely signed subfields. Both subfields are tilted to have slopes corresponding to object motion in the PD. The spatiotemporal offset between the subfields corresponds to object motion in the ND. ***G***, Top, Schematic of an STRF exhibiting a single subfield oriented along the axis corresponding to PD motion. This STRF enhances responses to PD motion. Bottom, Schematic of an STRF exhibiting two subfields of opposite sign oriented along the axis corresponding to ND motion. This STRF suppresses responses to ND motion. Both STRFs have the same PD. ***E–G***, Positive values (red) indicate points in spacetime where contrast increments evoke a positive response (ON), and negative values (blue) indicate points in spacetime where contrast decrements evoke a positive response (OFF). Color map is the same across all STRFs. Inset, Arrows indicate the PD.

and T5, we first isolated single-cell clones using a stochastic Flp-out method, which limited expression of GCaMP6f to a sparse subset of the R42F06 expression pattern (Fig. 1*C*) (Gordon and Scott, 2009; Chen et al., 2013; Gruntman and Turner, 2013; Fisher et al., 2015b). From the sparsened R42F06 expression pattern, we identified visually responsive single-cell clones by their periodic responses to moving sinusoidal gratings. Many single-cell clones were strongly contrast and direction selective, consistent with previously reported tuning properties of T4 and T5 (Fig. 2*B–E*) (Maisak et al., 2013). Of 49 single-cell clones from eight flies in which direction selectivity was assessed, 25 had a DSI of >0.6, which we defined to be strongly direction selective. Of 35 single-cell clones from six flies in which edge-contrast selectivity was assessed, 13 were essentially completely light-edge selective (CSI > 0.6; see Materials and Methods), consistent with their being T4 single-cell clones, whereas 13 were essentially completely dark-edge selective (CSI > 0.6; see Materials and Methods), consistent with their being T5 single-cell clones. Nine single-cell clones did not exceed our contrast selectivity threshold (CSI ≤ 0.6).

To obtain the STRFs of T4 and T5 single-cell clones, we recorded their responses to one-dimensional spatiotemporal (XT) ternary noise (Fig. 1*D*; see Materials and Methods). This stimulus comprised a series of bars, each 5° wide and with one of three contrasts (−1, 0, 1) chosen randomly every 50 ms. We then computed the average spacetime history of the stimulus preceding the neural response weighted by the amplitude of the response, an analysis known as reverse correlation that yields the STRF (Chichilnisky, 2001). When the spatial axis of the stimulus was aligned with the axis of the PD, across four flies 9 of 10 visually responsive single-cell clones had STRFs that exhibited two oppositely signed subfields, with both subfields tilted to have slopes cor-

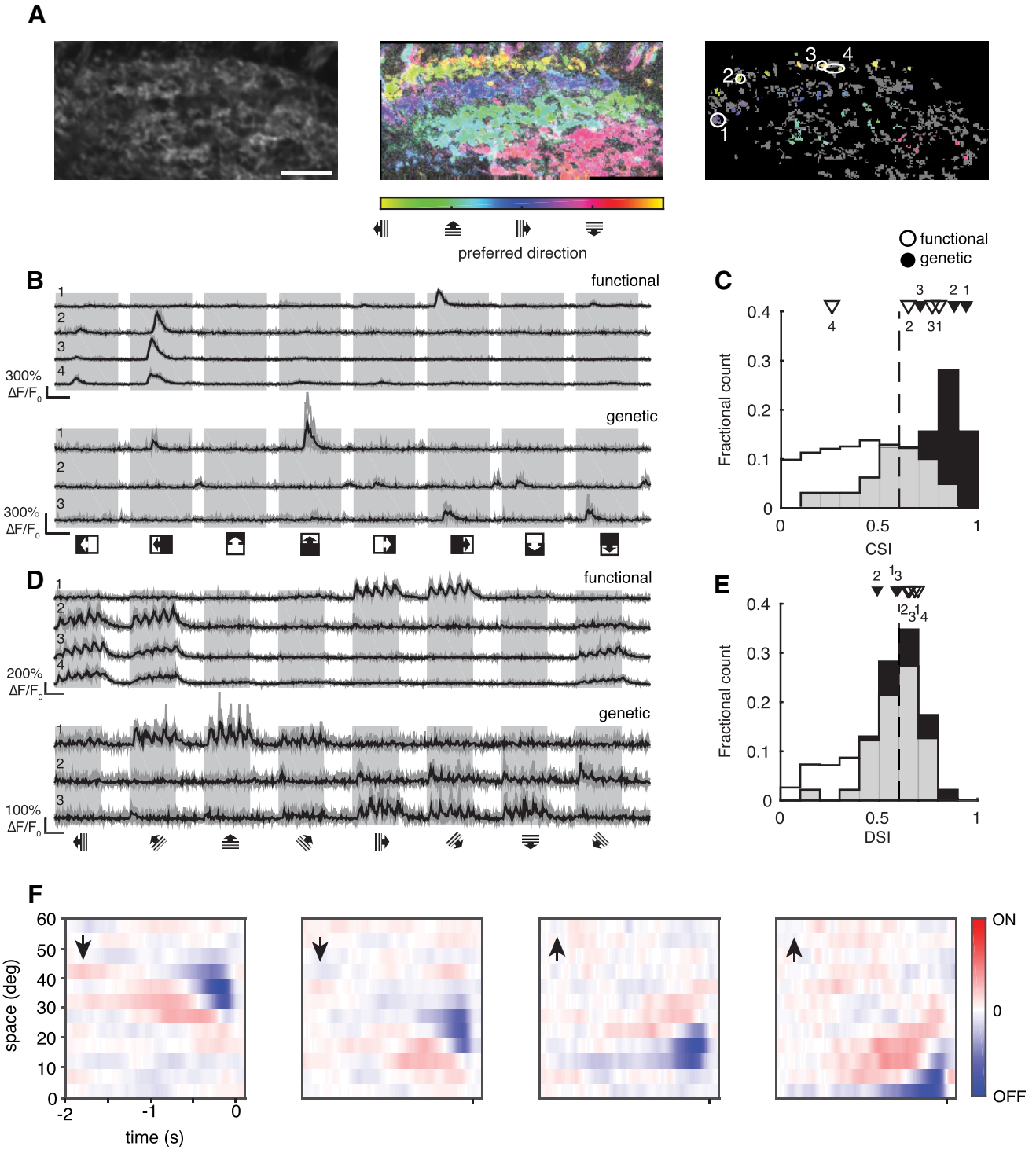

Figure 2. Isolation of T5 single units based on their functional properties. ***A***, FOV of the R42F06 expression pattern. Left, Mean fluorescence image. Scale bar, 10 μm. Middle, Hue-saturation value (HSV) map of the same FOV calculated from response evoked by sinusoidal gratings moving in eight evenly spaced directions. Color map as in Figure 1*C* (bottom). Right, Functionally isolated ROIs derived from the same FOV (see Materials and Methods). Only ROIs that were morphologically realistic and functionally indistinguishable from single-cell clones in terms of their contrast and direction selectivity were considered further, here colored as in Figure 1*C* (bottom). All other ROIs were discarded, here colored gray. ***B***, Response time courses of functionally isolated single units and genetically isolated single-cell clones to light and dark edges moving in the cardinal directions. Gray represents single-trial response time courses. Black represents trial-averaged response time courses. Each shaded epoch corresponds to one stimulus condition as indicated schematically below each shaded region. Top, Functionally isolated single units, indicated in ***A*** (right). Bottom, Single-cell clones from a different fly. Scale bars indicate 2.5 s. ***C***, Distribution of CSI for functionally isolated single units (white) and genetically isolated single-cell clones (black). CSIs of single units and single-cell clones from ***B*** indicated with white and black arrowheads, respectively. Sample size for functionally isolated single units: $N = 37$ single units from 9 flies. Sample size for genetically isolated single-cell clones: $N = 35$ single-cell clones from 6 flies. ***D***, For the same functionally isolated single units and genetically isolated single-cell clones as in ***B***, response time courses to sinusoidal gratings moving in eight evenly spaced directions. Conventions as in ***B***. ***E***, Distribution of DSI for functionally isolated single units (white) and genetically isolated single-cell clones (black). DSIs of single units and single-cell clones from ***D*** indicated with white and black arrowheads, respectively. Sample size for functionally isolated single units as in ***C***. Sample size for genetically isolated single-cell clones: $N = 49$ single-cell clones from 8 flies. ***F***, STRFs of T5 single units for which the spatial axis of the stimulus was aligned with the axis of the PD (arrow). Color map is as in Figure 1*E–G* and is the same across all four STRFs. All STRFs exhibit two oppositely signed subfields. Both subfields are tilted to have slopes corresponding to object motion in the PD. The spatiotemporal offset between the subfields corresponds to objection motion in the ND.

responding to object motion in the PD. Moreover, the two subfields were spatiotemporally offset along an axis corresponding to object motion in the ND (Fig. 1*E*,*F*). In the STRFs of four single-cell clones, the more temporally offset, or delayed, subfield was OFF and the nondelayed subfield was ON. These were presumably T4 light-edge preferring single-cell clones. In the STRFs of five single-cell clones, the delayed subfield was ON and the nondelayed subfield was OFF. These were presumably T5 dark-edge preferring single-cell clones. The organization of these STRFs was specific to the spatial axis of the PD. For the 13 single-cell clones that were visually responsive but whose STRFs were mapped orthogonally to the axis of PD motion, the STRFs were weak and disorganized (data not shown), consistent with the known orientation selectivity of T4 and T5 (Fisher et al., 2015b).

One prominent feature of the STRFs of T4 and T5 single-cell clones is that they are bilobed, with two subfields of opposite sign and with the axis connecting the centers of the two oppositely signed subfields aligned with the ND. Visual features moving in the PD contact only one subfield at a time, whereas features moving in the ND simultaneously contact both oppositely signed subfields, resulting in mutual cancellation of the subfield activations. Thus, the organization of the STRF contributes to direction selectivity by suppressing responses to ND motion (Fig. 1*G*, bottom). In general, there is ambiguity in the interpretation of subfield sign relative to the contrast polarity of the stimulus. For example, an OFF subfield may be due to depolarizing responses to contrast decrements (OFF "facilitation"), or it may be due to hyperpolarizing responses to contrast increments (ON "suppression"), or both. However, dark edges moving in the PD evoke depolarizing transients in T5 (Maisak et al., 2013), demonstrating that the stimulus-response relationship underlying the OFF subfield includes OFF facilitation. In order for T5 to be direction-selective to moving dark edges, and for T5 not to respond to full-field contrast steps or wide, dark stimuli (Fisher et al., 2015a), the ON subfield must necessarily reflect OFF suppression. We have also observed that T5 depolarizes in response to moving dark stripes as well as moving light stripes, suggesting that the ON subfield may also reflect ON facilitation under some conditions (data not shown).

A second prominent feature of the STRFs of T4 and T5 single-cell clones is that their subfields are tilted in spacetime along the PD. Visual features moving in the PD have spacetime trajectories that align more perfectly with individual subfields than features moving in the ND. Thus, this spacetime tilt contributes to direction selectivity by enhancing responses to PD motion (Fig. 1*G*, top). Further, because the inputs to T4 and T5 are not themselves direction selective, there cannot be a one-to-one correspondence between tilted subfields of the STRF and non–direction-selective input pathways. Each direction-selective subfield in the STRF must receive contributions from multiple non–direction-selective input pathways.

T4 and T5 single units can be isolated from mixed recordings based on functional properties

While functional imaging of single-cell clones enabled the isolation of single units, this approach was hampered by low throughput and low signal-to-noise ratio. In the majority of flies, the GCaMP6f expression pattern was either too dense to reveal single-cell clones, or so sparse or weak that no single-cell clone could be found that responded to the visual stimulus. We developed a method to isolate many single T4 and T5 neurons from within the dense expression pattern of the R42F06 driver, thereby enabling us to describe the STRFs of many T4 and T5 single units quantitatively. Although the morphological segmentation of the R42F06 expression pattern is challenging, the exquisite functional organization of the fly optic lobe suggested the possibility of resolving single units by functional segmentation of anatomically dense, even non-cell type specific, expression patterns. Both T4 and T5 axon terminals are organized into retinotopic columns and the four direction-selective layers of the lobula plate neuropil (Figs. 1*A*, 2*B*) (Maisak et al., 2013). Moreover, T4 is strongly light-edge preferring, whereas T5 is strongly dark-edge preferring (Maisak et al., 2013; Fisher et al., 2015b). Thus, even though the morphology of T4 and T5 essentially prohibits the isolation of single units based on the R42F06 expression pattern alone (Fig. 2*A*, first and second panels), each neuron exhibits a uniquely identifying combination of direction preference, contrast preference, and spatial receptive field.

We isolated T4 and T5 single units by probing these functional properties using moving sinusoidal gratings presented over eight evenly spaced directions (Fig. 2*A*, third panel). These gratings evoke periodic responses in T4 and T5, with the amplitude of the response reflecting direction preference, and the phase of the response reflecting contrast preference and spatial receptive field (Fig. 2*D*). Each T4 and T5 single unit has a uniquely identifying tuning profile comprised of these eight amplitudes and phases. Clustering of pixels based on their tuning profiles yielded ROIs. Each ROI is uniquely characterized by its direction preference, contrast preference, and spatial receptive field, all shared across its constituent pixels (Fig. 2*A*, third panel; see Materials and Methods). In the R42F06 expression pattern, the vast majority of pixels include a mixture of multiple T4 and T5 single units. ROIs derived from these pixels, despite having a uniquely identifying tuning profile shared across pixels, do not correspond to T4 or T5 single units. Such multiunit ROIs were poorly contrast selective, inconsistent with the strong contrast selectivity of T4 and T5 (Fig. 2*B*,*C*). Based on the known morphology and functional properties of T4 and T5, we applied a series of filtering steps (see Materials and Methods) to recover ROIs that were morphologically realistic and functionally indistinguishable from single-cell clones in terms of their contrast and direction selectivity (Fig. 2*B*–*E*). We refer to such functionally isolated and validated ROIs as single units. Although we recovered both light-edge preferring T4 and dark-edge preferring T5 single units, T5 single units were far more abundant (13 T4 single units and 63 T5 single units from 9 flies), likely because of the T5-bias of the R42F06 expression pattern (Fisher et al., 2015b). We focused subsequent analyses on T5 given the relatively large number of T5 single units that we identified.

The spatiotemporal receptive field of functionally isolated T5 single units

With the spatial axis of the stimulus aligned with the axis of the PD, 37 T5 single units that we functionally isolated replicated the STRF organization of T5 single-cell clones (Figs. 1*E*, 2*F*). We note that T5 single units were functionally isolated based on their responses to non-noise stimuli. In contrast, the STRFs of these units depend only on their responses to noise stimuli. Therefore, the similarity of the STRFs of functionally isolated T5 single units to those of the genetically isolated T5 single-cell clones is a nontrivial check of our functional isolation procedure.

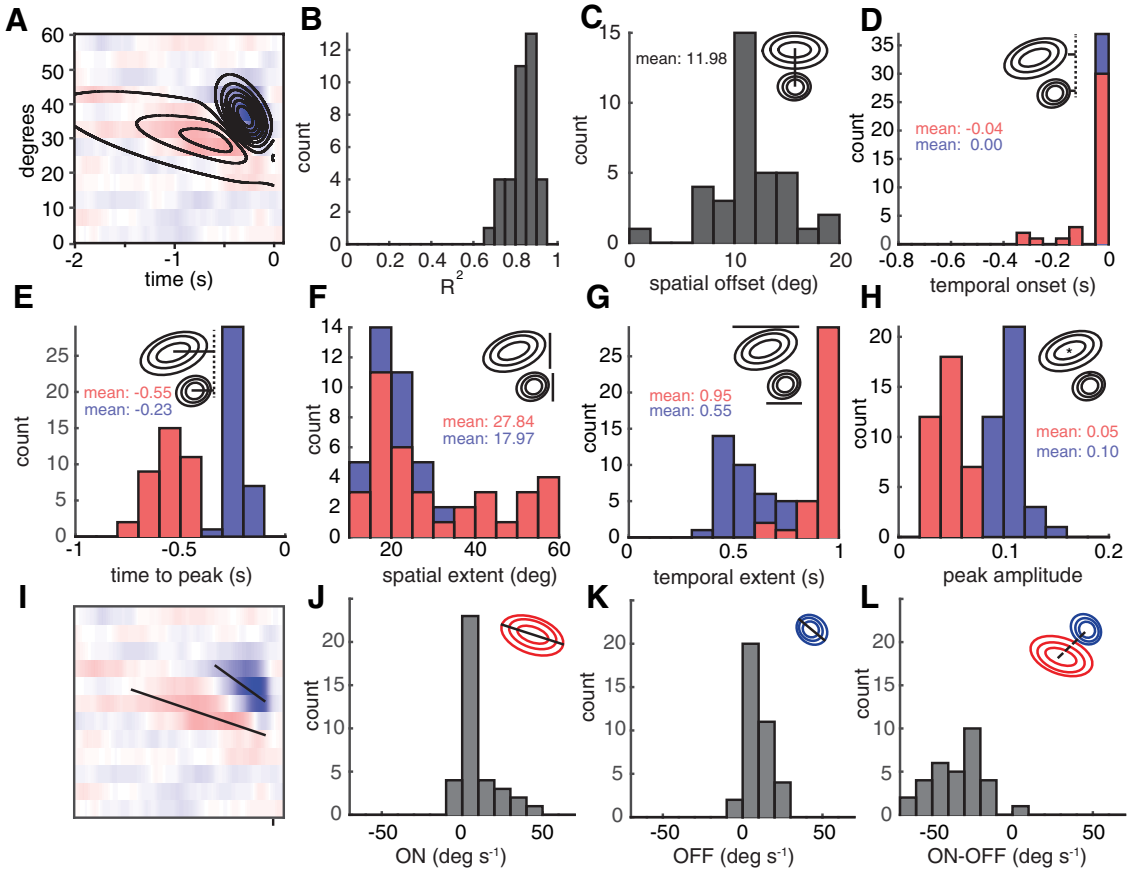

Figure 3. The spatiotemporally offset, tilted organization of ON and OFF subfields is consistent across the population of T5 STRFs. ***A***, Example difference of Gaussians (DoG, contour plot) fit to a T5 STRF (color map). ***B***, R^2 of fits of DoGs to T5 STRFs. $N = 37$ single units from 9 flies. ***C***, Distribution of the spatial offsets between the centers of ON and OFF subfields. ***D–H***, Distributions of additional DoG parameters corresponding to the ON (red) and OFF (blue) subfields of T5 STRFs: ***D***, temporal onset; ***E***, time to peak; ***F***, spatial extent; ***G***, temporal extent; ***H***, peak amplitude. ***I***, Spatiotemporal tilts of the ON (dashed) and OFF (solid) subfields for the same T5 STRF as in ***A***. ***J***, Distribution of the velocity of the ON subfield, corresponding to the slope of its tilt. ***K***, Distribution of the velocity of the OFF subfield, corresponding to the slope of its tilt. ***L***, Distribution of the slope of the spatiotemporal axis between the ON and OFF subfields. ***J–L***, Positive slopes correspond to object motion in the PD, whereas negative slopes correspond to object motion in the ND.

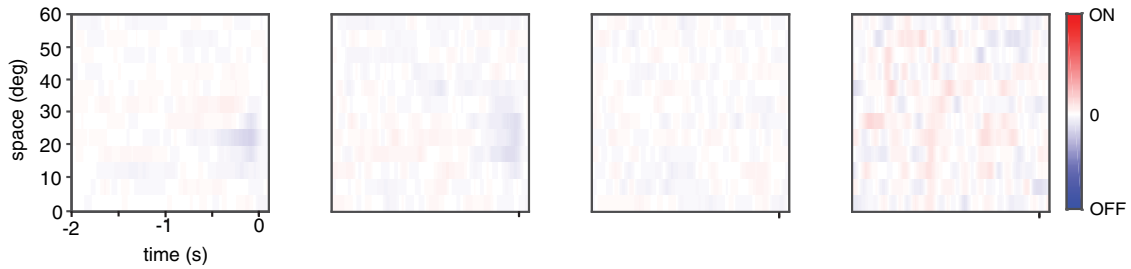

Figure 4. T5 STRFs measured along the axis orthogonal to the axis of the PD are weak and disorganized relative to those measured along the axis of the PD. Four representative STRFs measured orthogonal to the axis of the PD are shown. Color map is as in Figure 1*E–G* and is the same across all four STRFs.

We further characterized the spatiotemporal organization of T5 STRFs by fitting them with differences of Gaussians (Fig. 3*A*), which provided an accurate parametric description (Fig. 3*B*). The parameters of the fit did not vary substantially across cells, confirming that all T5 single units share a common STRF organization. In this organization, the centers of the two subfields were offset in space by ≈12° (Fig. 3*C*). The OFF subfield was spatiotemporally more compact with a faster and higher peak, whereas the ON subfield extended further back in time and had a more delayed and broader peak (Fig. 3*D–H*). Each individual subfield was tilted along a spatiotemporal axis consistent with the trajectory of a feature moving in the PD of the unit (Figs. 1*E*,*F*, 2*F*, 3*I–L*). Furthermore, the centers of mass of the two subfields were separated along a spatiotemporal axis consistent with the trajectory of a feature moving in the ND of the unit (Figs. 1*E*,*F*, 2*F*, 3*I–L*). Again, consistent with expectations based on the orientation selectivity of T5 (Fisher et al., 2015b), when the spatial axis of the stimulus was orthogonal to the axis of the PD, 26 T5 single units had STRFs that were weak and disorganized (Fig. 4).

An elaborated LN model of the T5 stimulus-response function

The T5 STRF, as a linear approximation of the stimulus-response function of T5, explained little stimulus-evoked variance in the

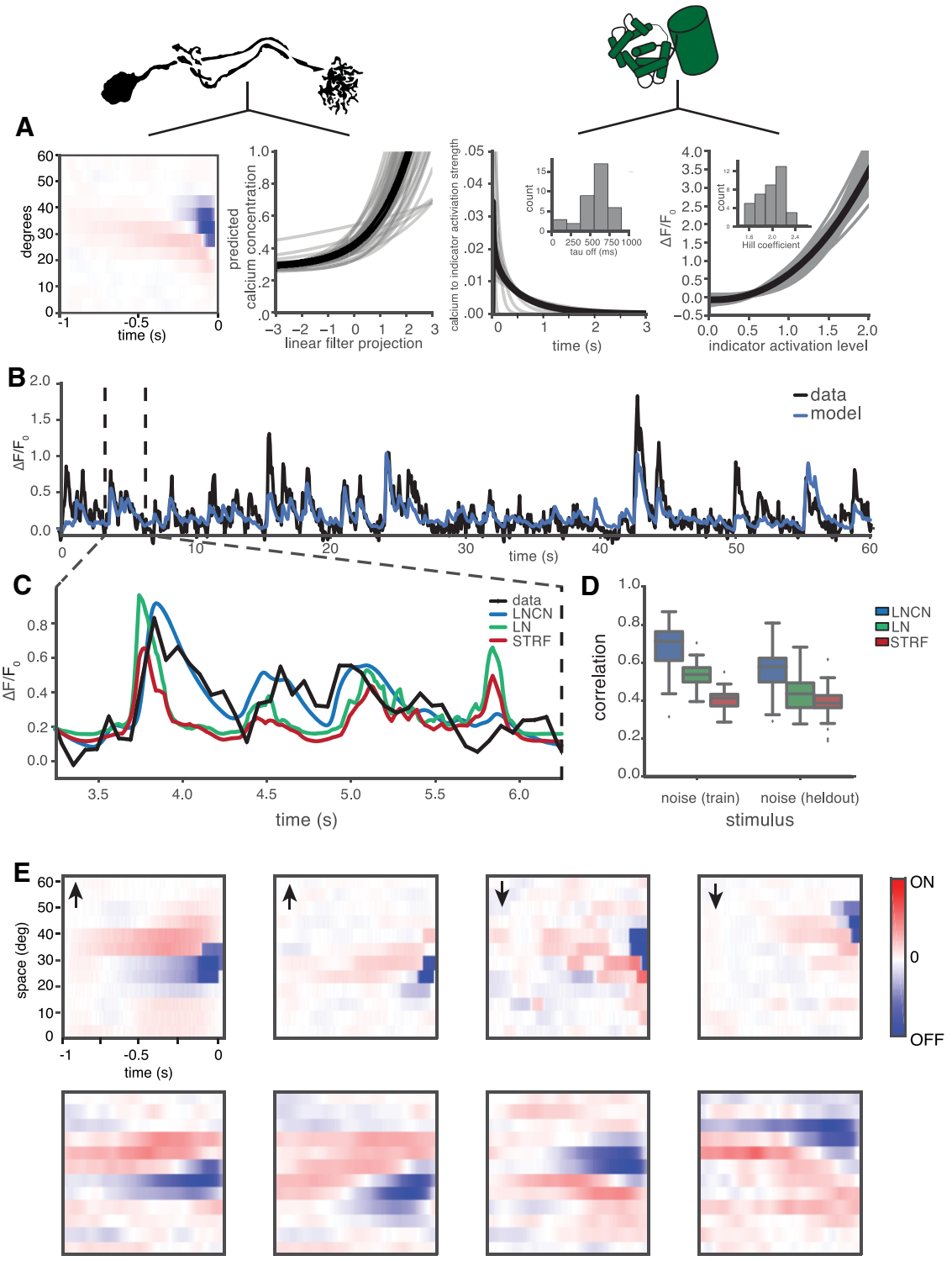

Figure 5. The LNCN model captures T5 responses. ***A***, The LNCN model architecture consists of a sequence of four components, shown from left to right: a linear filter that integrates contrast inputs over spacetime, a static nonlinearity, a linear convolution of the calcium signal over time, and a nonlinearity reflecting the cooperativity of indicator-calcium binding. The linear stimulus filter shown is the average of the linear stimulus filters for 37 T5 single units aligned in space (see Materials and Methods). For all other components of the model, the average is in black, and values for individual models are in gray. Insets, Distributions of the decay time (tau_{off}) for the convolutional filter, and the Hill coefficient corresponding to indicator-calcium binding cooperativity. ***B***, The experimentally observed fluorescence response of a T5 single unit (black) and the fluorescence response predicted by the respective model (blue) when probed with a held-out noise stimulus. The linear stimulus filter for this example is the leftmost unit in Figure 2*F*. ***C***, Comparison of the experimentally observed fluorescence response (black) with responses predicted by the LNCN model (blue), the linear prediction based on the measured STRF alone (red), and a LN model without an account of calcium indicator dynamics (green). ***D***, Comparison of model performance as measured by the correlation coefficient with data on training noise versus held-out noise. ***E***, Top, Linear filters from LNCN models of four representative T5 ROIs (from Fig. 2*F*). Color map is as in Figure 1*E–G* and is the same across all four linear filters. Arrows indicate the PD. Bottom, STRFs computed from the responses of the LNCN models to noise. Color map is as in Figure 1*E–G* and is the same across all four STRFs.

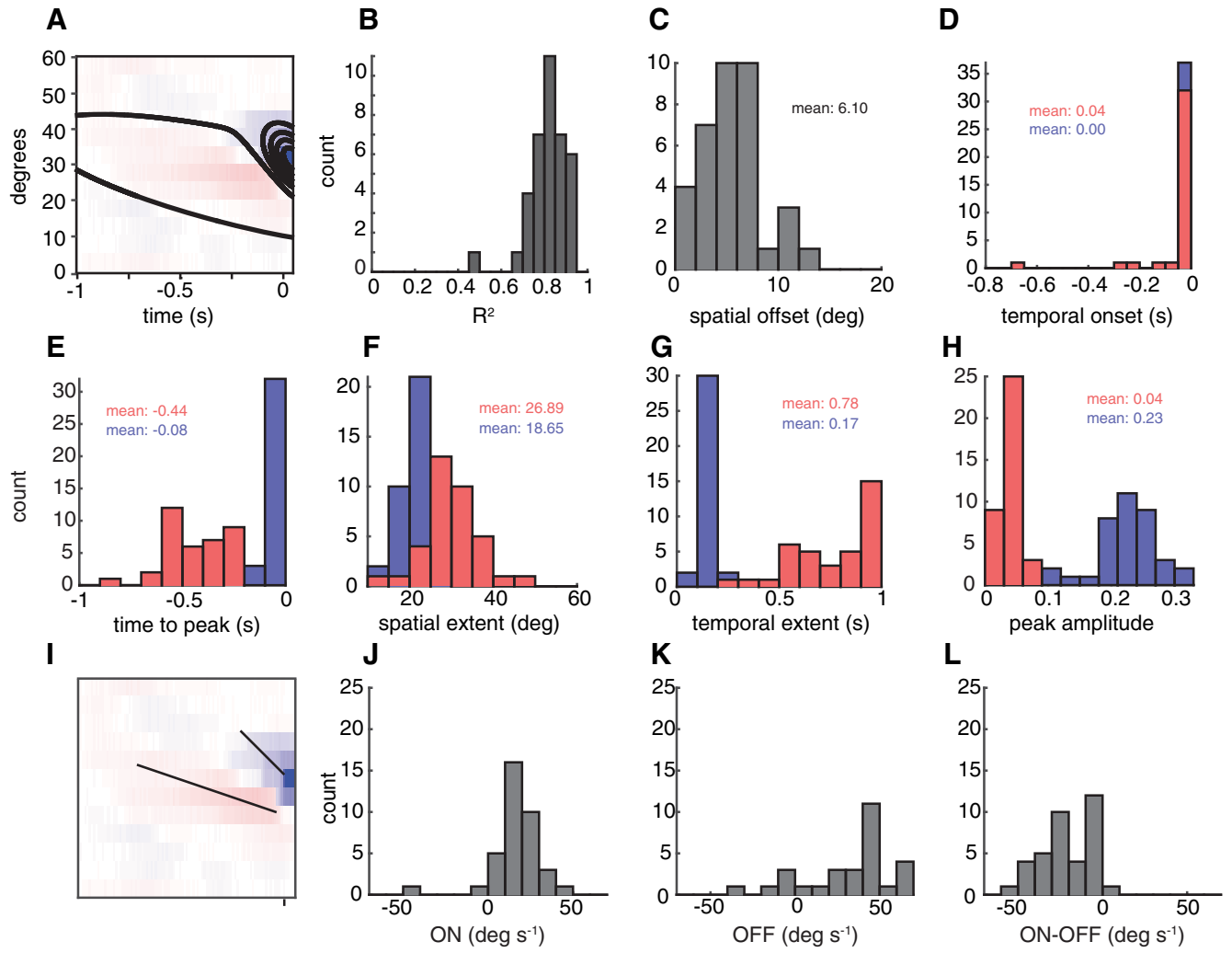

Figure 6. Linear stimulus filters of LNCN models of T5 have similar organization to measured STRFs, although more compact in time. Quantitative measurements of differences of Gaussians fit to T5 model linear stimulus filters. ***A–L***, Same as in Figure 3*A–L*.

T5 response, with a correlation coefficient on held-out noise of only 0.38 ± 0.01 (mean ± SEM). To capture nonlinearities in the neuronal calcium response and to further isolate the sensory encoding transformation reflected in the T5 response from indicator-induced transformation of the calcium signal, we modeled the stimulus-evoked T5 calcium response as an LN model, a linear stimulus filter (L) followed by a static nonlinearity (N), feeding into a temporal convolution (C) and a static nonlinearity (N) that reflect the kinetics of cooperative GCaMP6f-calcium binding (Fig. 5*A*) (Kato et al., 2014). We trained LNCN models to predict the response of T5 single units to XT noise. These models explained a considerable fraction of stimulus-evoked variance in the T5 response to XT noise held out from the training set (correlation of 0.56 ± 0.02; Fig. 5*B–D*). In these models, parameters corresponding to the calcium-binding kinetics and cooperativity of GCaMP6f were consistent with corresponding *in vitro* measurements (Chen et al., 2013), providing ground-truth validation of the stimulus-response model (Fig. 5*A*). LNCN models performed better than the best LN models (correlation of 0.43 ± 0.01), in large part by better matching the time course of fluorescence decay (Fig. 5*C*).

Echoing the T5 STRF, the linear stimulus filter of the LNCN model of T5 exhibited two spatiotemporally offset, oppositely signed subfields, with both subfields tilted to have slopes corresponding to object motion in the PD (Figs. 1*E*,*F*, 2*F*, 3*I–L*, 5*E*). As before, the OFF subfield was spatiotemporally more compact, whereas the ON subfield extended further across space and back in time (Figs. 5*E*, 6). The neuronal nonlinearity was half-wave rectifying and expansive, and was well approximated by a cubic polynomial (Fig. 7). Alternative choices for the parameterization of the nonlinearity diminished the predictive performance of the models but retained the characteristic structure of the linear stimulus filter (data not shown). Thus, a nonlinear elaboration of the spatiotemporal organization of the T5 STRF, augmented by a model of signal transduction by GCaMP6f, serves as a highly predictive stimulus-response model of T5.

The LNCN model quantitatively accounts for the responses of T5 to motion stimuli

If the LNCN model captures fundamental properties of the algorithm by which the fly EMD generates direction-selective responses, then the model should reproduce T5 responses to directional motion stimuli. In addition to spatiotemporal noise, we probed a subset of the T5 single units with classical motion stimuli. For 30 T5 single units from 7 flies, up to an overall scaling factor the LNCN model explained a considerable fraction of stimulus-evoked variance in the T5 response to moving sinusoi-

dal gratings (Fig. 8*A*), quantitatively reproducing the responses of T5 to a diversity of 10 stimulus conditions (average correlation coefficient, 0.87 ± 0.01). The LNCN model of T5 predicted the spatial frequency tuning of T5, as well as its strong direction selectivity across the range of spatial periods that we probed (Fig. 8*B*). Critically, the LNCN model of T5 was a highly accurate predictor of the PD across the range of spatial periods that we probed (Fig. 8*C*). To assess the ability of the LNCN model of T5 to predict not only the PD, but the degree of direction selectivity, we quantified direction tuning with a DSI. A DSI of 0 indicates no direction preference, whereas a DSI of 1 indicates responses exclusively in the PD. The LNCN model quantitatively predicted DSI in the majority of conditions that we studied, although it underpredicted DSI by an increasing margin at the two largest spatial periods that we studied, 100° and 250° (Fig. 8*D*). This underprediction of DSI resulted from a slight overprediction of the responses to fast moving, large spatial period ND stimuli. Because our stimulus screen subtended a relatively limited visual angle, these large spatial period, low spatial frequency stimuli behaved similarly to "full-field" contrast excursions, which are known to drive T4 and T5 poorly. Importantly, we note that LNCN models were trained only to predict the response of T5 to noise stimuli. Thus, their ability to also predict the response to a diversity of gratings with different directions and speeds constitutes a remarkable level of generalization across stimulus types. Overall, we conclude that the LNCN model does indeed capture fundamental properties of the direction selectivity algorithm of the fly EMD, as evidenced by the model's ability to quantitatively predict T5 responses to moving sinusoidal gratings, a classical motion stimulus.

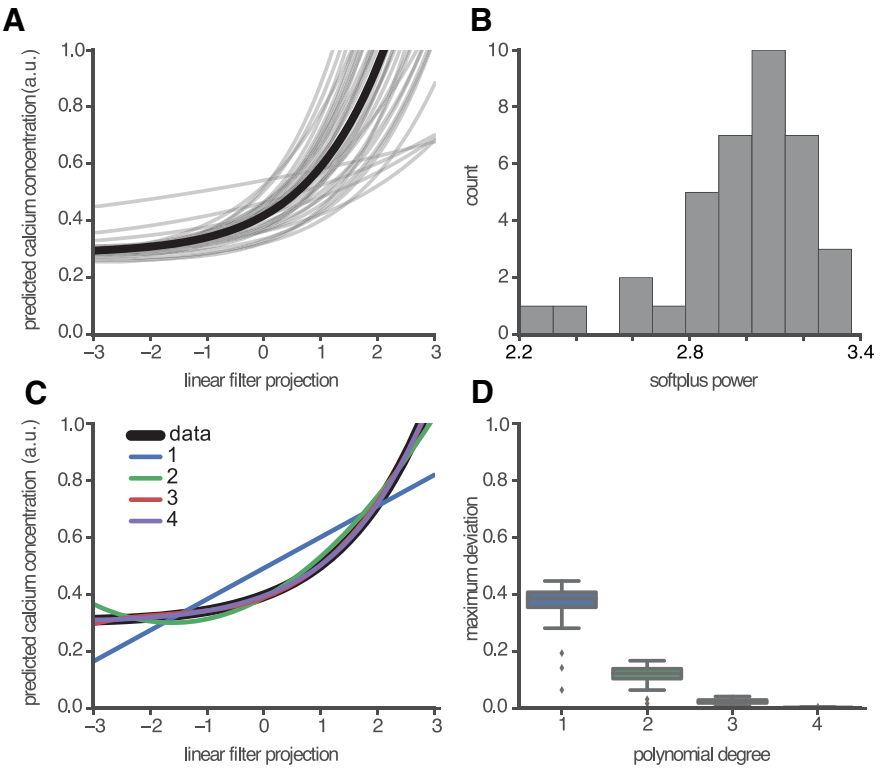

Figure 7. The rectified neuronal nonlinearity is well fit by polynomials of order ≥ 3. ***A***, Average (black) and individual (gray) neuronal nonlinearities from fits to 37 T5 ROIs from 9 flies. ***B***, Distribution of softplus exponents for neuronal nonlinearities (see Materials and Methods). ***C***, Best-fitting polynomials of order 1, 2, 3, and 4 to the average neuronal nonlinearity of LNCN models. ***D***, Maximum deviation of nonlinearity from the best-fitting polynomial of orders 1–4.

T5 performs multiple spatiotemporal correlations that support direction and contrast selectivity

Moving objects induce not only second-order correlations at two neighboring points in spacetime, but also higher-order spatiotemporal correlations. Third-order correlations contain information about the motion of natural scenes, and stimuli with only third-order spacetime correlations can induce motion percepts in both flies and humans (Hu et al., 2010; Fitzgerald et al., 2011; Clark et al., 2014; Nitzany and Victor, 2014; Fitzgerald and Clark, 2015). To understand how the LNCN model exploits second- and third-order spatiotemporal correlations to achieve direction and contrast selectivity, consider an idealized linear stimulus filter that combines stimulus contrast at four points in spacetime (Fig. 9*A*), two points from within the ON subfield of the original filter having positive weight, and two points from within the OFF subfield having negative weight. If the model response is generated by a weighted sum of stimulus contrast at these four points, passed through a polynomial nonlinearity with linear, quadratic, and cubic terms, then the resulting computation can be organized into a sum of signed multiplications, or correlations, of stimulus contrasts at one, two, and three points in spacetime (Fig. 9*B–D*).

The linear term in the polynomial corresponds to a weighted sum of stimulus contrasts at the four points in spacetime (Fig. 9*B*). Computationally, this is a linear direction-selective cell, whose filter is the idealized linear stimulus filter. The quadratic term leads to two classes of direction-selective second-order correlators and a class of non–direction-selective but contrast-selective correlators. ON-ON and OFF-OFF correlators at two points in spacetime originating from a single subfield of the original filter (Fig. 9*C*, top) compute the positive product of stimulus contrast at two points in spacetime along the PD of the cell. ON-OFF correlators at two points in spacetime originating from distinct subfields of the original filter, aligned along the ND of the cell (Fig. 9*C*, middle), compute the negative product of stimulus contrast at two points in spacetime along the ND. ON-OFF correlators at two points in spacetime aligned along either the temporal or spatial axis (Fig. 9*C*, bottom) are not direction selective but are contrast selective. The cubic term results in four third-order correlators, involving signed products of stimulus contrast at three spacetime points arranged in spatially or temporally, diverging or converging patterns (Fig. 9*D*). The divergent correlators, by having one point in the OFF subfield, compute the negative product, whereas the convergent correlators, by having two points in the OFF subfield, compute the positive product. Whereas the linear and quadratic terms of the nonlinearity are relatively large, the cubic term is smaller, indicating a smaller contribution of third-order correlations to the response of the LNCN model of T5.

Discussion

To characterize the computational algorithm underlying the emergence of direction selectivity in the fly EMD, we studied the

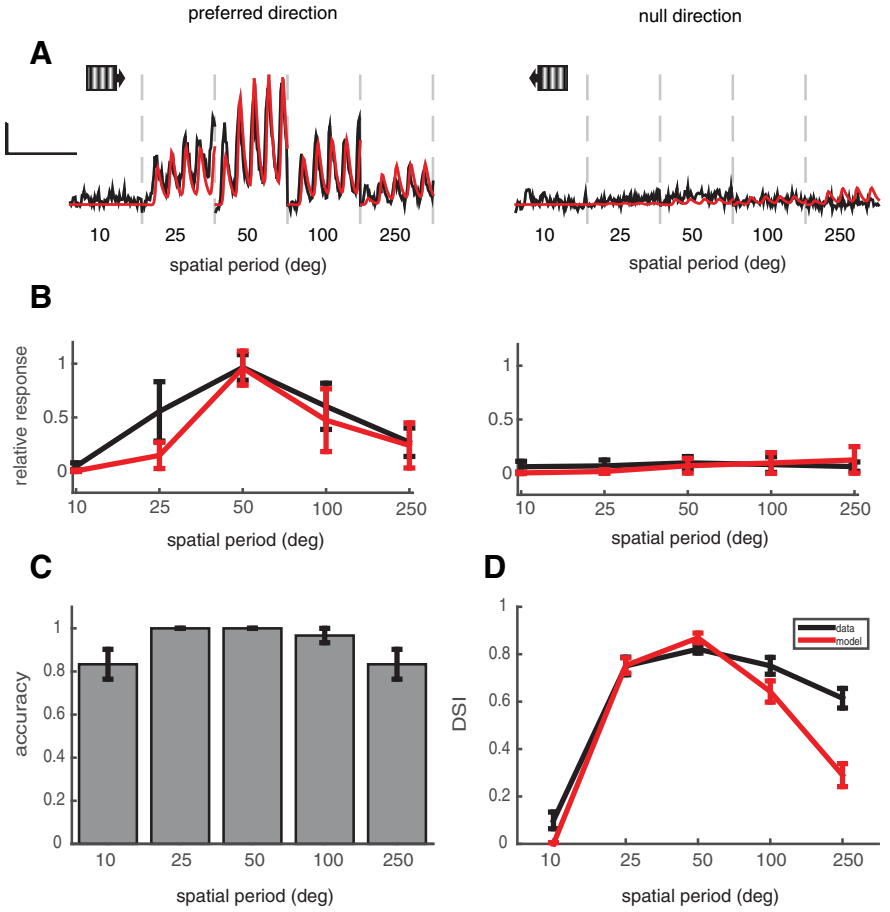

Figure 8. LNCN models predict T5 responses to motion stimuli. ***A***, Representative time course for a single T5 unit (black) and its model (red) responding to moving sinusoidal gratings with a temporal frequency of 1 Hz and increasing spatial period. Scale bars indicate 100% $\Delta F/F_0$ and 5 s. In all panels, PD stimuli are on the left and ND stimuli are on the right. ***B***, Population tuning curves for model (red) and data (black) responses to moving sinusoidal gratings with temporal frequency of 1 Hz and varying spatial period, as measured by the mean of the response to each grating ($N = 7$ (30)). Responses for each model or unit are scaled to the fraction of the maximum mean response observed across spatial periods. ***C***, Average model prediction accuracy of the preferred direction for each spatial period, using the direction with the largest mean response as the preferred direction. ***D***, DSI as a function of spatial period for data (red) and model (black).

stimulus-evoked activity of key computational intermediates, T4 and T5, the most upstream direction-selective neurons in *Drosophila*. Our results from this study add to the understanding of motion detection in the fly in three ways. First, we measured the STRFs of T4 and T5 and found them to be organized into two subfields of opposite sign, spatiotemporally offset and obliquely oriented in spacetime. Because the two subfields are oppositely signed and spatiotemporally offset along the spacetime direction corresponding to the ND, responses to motion in the ND are suppressed through mutually cancelling interactions. This suppression corresponds exactly to the algorithmic strategy for direction selectivity used by the Barlow–Levick detector in the vertebrate retina (Barlow and Levick, 1965). By contrast, motion in the PD aligns with each subfield, creating mutually enhancing interactions that can be further amplified by downstream nonlinearities. This enhancement corresponds to the algorithmic strategy used by the HRC. Thus, T5 incorporates the essential algorithmic strategies used by both the HRC and the Barlow–Levick detector. Second, we found, using an LN stimulus-response model, that the combination of the T5 STRF and a cubic nonlinearity predicts the responses of T5 to behaviorally relevant motion stimuli. The cubic nonlinearity enables our model of T5 to detect specific first-, second-, and third-order spatiotemporal correlations in the stimulus, discussed below. Third, our results highlight the crucial influence of input pathway filtering properties in generating T5 direction selectivity. As we will discuss, the combination of inputs that have center-surround, temporally biphasic receptive fields yields a T5 receptive field that contains two subfields of spatiotemporal contrast sensitivity. These subfields do not correspond in a one-to-one manner with T5 inputs; multiple inputs contribute to each subfield, with the temporal dynamics and spatial structure of the inputs allowing single inputs to contribute differently signed signals at different spacetime locations.

Complementary genetic and functional approaches to isolating single units for calcium imaging

Essential to our characterization of the T5 STRF was the isolation of T5 single-unit responses. To meet this requirement, we used complementary genetic and functional methods. Stochastic Flp-out methods are a well-known approach to labeling and recording from single cells (Gordon and Scott, 2009; Gruntman and Turner, 2013; Fisher et al., 2015a). However, such methods typically modulate both the extent and the level of gene expression, limiting their utility, because the genetic effectors used are often affected by protein perdurance. Consequently, functional imaging of sparse neuronal subsets derived by stochastic Flp-out is often hampered by both low throughput and low signal-to-noise ratio. We developed a functional method to isolate stronger single-unit responses. Each T4 and T5 neuron is identifiable by a unique functional profile: direction preference, contrast preference, and spatial receptive field. By mapping these functional properties across a dense expression pattern, we could locate and extract high SNR activity due to single neurons. In previous studies using calcium imaging *in vivo*, the segmentation of ROIs was either performed manually (Euler et al., 2002; Wang et al., 2003) or based on morphological priors about ROIs (Ohki et al., 2005; Mank et al., 2008). More recent attempts to extract signals from dense functional imaging datasets have leveraged general functional priors (Mukamel et al., 2009; Pnevmatikakis et al., 2016). Here we demonstrate that genetically isolated single-cell recordings can provide highly specific functional priors that can guide the segmentation of high SNR ROIs from dense functional imaging datasets. We expect that this methodological advance will be widely applicable in genetic model organisms.

Understanding motion-processing pathways in flies

In flies as in humans, moving light edges and moving dark edges are detected by separate circuits (Joesch et al., 2010; Clark et al., 2011, 2014; Maisak et al., 2013; Fisher et al., 2015b). Before the present study, behavioral measurements and physiological re-

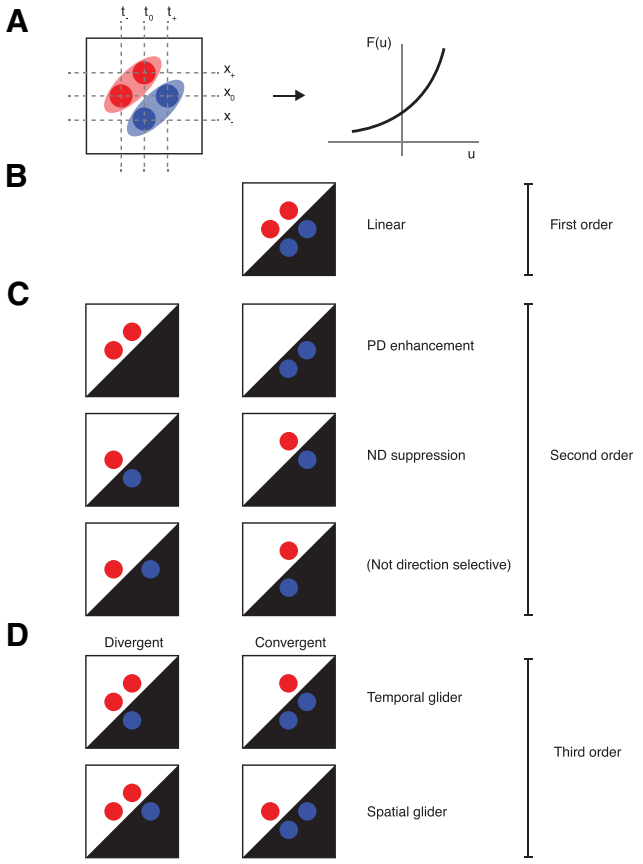

Figure 9. T5 uses both second- and third-order correlations to respond direction-selectively to moving dark edges. ***A***, XT representation of an idealized version of the linear spacetime filter (left, light ellipses) that combines stimulus contrast at four spacetime points (dark circles). These points are arranged across three points in time ($t-$, t_0, $t+$) and three points in space ($x-$, x_0, $x+$). The output of this linear filter is passed through an idealized positive cubic nonlinearity resembling that we observe in T5 (right). The response of the idealized model can be factored into three classes of correlators: first-order (***B***), second-order (***C***), and third-order (***D***). The first-order term linearly combines contrasts at the four spacetime points. The second-order terms combine contrasts at two spacetime points. Two of these combinations correspond to PD enhancement, two correspond to ND suppression, and two are not direction-selective. The third-order terms combine contrasts at three spacetime points and can be arranged into four types: divergent and convergent temporal gliders, and divergent and convergent spatial gliders. Each second- and third-order correlator computes the signed product of contrasts at the associated spacetime points where the sign is positive if the number of OFF points is even, and negative if the number of OFF points is odd.

cordings demonstrated that the first direction-selective neurons in the fly, T4 and T5, are selective for moving edges of specific polarity, but the full repertoire of contrast comparisons made by these neurons was unknown (Maisak et al., 2013; Fisher et al., 2015b). In T5, we identify a single neuronal substrate that computes the signed, weighted sum of up to 11 correlators of orders 1, 2, and 3, indeed all possible individual dark-edge specific contrast comparisons that the fly can detect, as demonstrated in previous studies of either downstream neurons or behavior (Clark et al., 2011, 2014; Eichner et al., 2011; Haikala et al., 2013; Joesch et al., 2013). When a dark edge moving along the PD crosses the STRF of the LNCN model (Fig. 5), the output of every one of these correlators is positive, yielding a very large summed response. On the other hand, for an identically moving light edge, the second-order correlators produce the same positive output, but the odd order correlators produce negative outputs, suppressing the response. For a dark edge moving in the ND, the linear term becomes zero and the outputs of the second-order and temporal third-order correlators become negative, again suppressing the response. Thus the core features of the LNCN model of T5 provide a remarkably simple and robust way to compute the sum of multiple correlators with the appropriate signs to be both direction and contrast selective for moving dark edges. Although we have not directly modeled T4 responses, our measurements of the STRF of T4 suggest that the moving light-edge pathway utilizes the same correlational structures as the moving dark-edge pathway, but with oppositely signed inputs.

The contribution of input pathway filtering properties, and likely parallels with vertebrate vision

Each of the four major input cell types immediately presynaptic to T5 has been shown to exhibit a center-surround, temporally biphasic receptive field, but the functional consequence of this visual processing has remained unclear for T5, let alone neurons further downstream (Fisher et al., 2015a; Serbe et al., 2016). In light of the spatiotemporal organization of the T5 STRF, the visual processing of T5 inputs gains new significance as a mechanistic cornerstone of direction selectivity. An input with a temporally biphasic receptive field will contribute to each of the oppositely signed subfields of the T5 STRF, albeit at different latencies. Similarly, an input with a center-surround spatial receptive field will contribute to each of the oppositely signed subfields of the T5 STRF, albeit at different spatial locations. While oppositely signed subfields can thus be inherited directly from single T5 inputs, the oblique orientations of the subfields and the spatiotemporal offset between them result from spatial offsets between multiple inputs with distinct kinetics. The resulting tilted geometry of the T5 STRF yields direction selectivity. Each individual subfield will enhance the T5 response to motion along its corresponding spacetime tilt, defining the PD. The spatiotemporal offset between the two oppositely signed subfields will result in suppression of the T5 response to motion along the axis of the offset, defining the ND.

It is tempting to compare the receptive fields of T4 and T5 to receptive fields of direction-selective simple cells in vertebrate primary visual cortex, a model system in which spatiotemporal tilt is widely accepted as the mechanistic basis for direction selectivity (DeAngelis et al., 1993; Carandini et al., 2005). However, whereas the simple cell generates direction selectivity using a fundamentally linear mechanism (Jagadeesh et al., 1993) and subsequently enhances direction selectivity via a nonlinearity, we have not provided direct evidence for such a linear mechanism underlying T5 responses. While the T5 STRF clarifies the hybrid algorithm underlying direction selectivity in T5, and the LN model of T5 clarifies its sensitivity to higher order correlations, we consider these models to be purely algorithmic descriptions of the transformation of spatiotemporal visual input into direction-selective T5 output. We do not consider these models to correspond directly to underlying mechanisms for the generation of T5 responses, although any future mechanistic model of the fly EMD must be consistent with the algorithm we describe.

From the algorithmic perspective, T4, T5, and the simple cell all combine PD enhancement and ND suppression at the level of their net computation to become direction selective. Independent of the potentially different internal workings of T4, T5, and the simple cell, what common fundamental mechanisms underlie this hybrid algorithmic approach to direction selectivity? Despite all that is known about the functional circuit architecture underlying direction selectivity in V1, certain mechanistic details remain elusive. In particular, the sources of differentially delayed input to direction-selective V1 simple cells have been challenging

to identify (De Valois and Cottaris, 1998; De Valois et al., 2000; Livingstone and Conway, 2007; Priebe et al., 2010), and, as a consequence, the relative spatial configuration of these inputs remains unknown. We believe the stereotyped and well-described input architecture to T4 and T5, combined with our measurements and modeling of their STRFs, will pave the way for new mechanistic models to reveal the specific contributions of individual cell types to the generation of the highly structured, spatiotemporally tilted T4 and T5 STRFs.

References

Ahrens MB, Orger MB, Robson DN, Li JM, Keller PJ (2013) Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat Methods 10:413–420. CrossRef Medline

Barlow HB, Levick WR (1965) The mechanism of directionally selective units in rabbit's retina. J Physiol 178:477–504. CrossRef Medline

Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow IJ, Bergeron A, Bouchard N, Bengio Y (2012) Theano: new features and speed improvements. arXiv:1211.5590 [cs.SC].

Batschelet E (1981) Circular statistics in biology. San Diego: Academic.

Behnia R, Clark DA, Carter AG, Clandinin TR, Desplan C (2014) Processing properties of ON and OFF pathways for *Drosophila* motion detection. Nature 512:427–430. CrossRef Medline

Bergstra J, Breuleux O, Bastien F, Lamblin P, Pascanu R, Desjardins G, Turian J, Warde-Farley D, Bengio Y (2010) Theano: a CPU and GPU math expression compiler. In Proceedings of the Python for Scientific Computing Conference, Austin, Texas.

Borst A, Euler T (2011) Seeing things in motion: models, circuits, and mechanisms. Neuron 71:974–994. CrossRef Medline

Borst A, Haag J, Reiff DF (2010) Fly motion vision. Annu Rev Neurosci 33:49–70. CrossRef Medline

Buchner E (1976) Elementary movement detectors in an insect visual system. Biol Cybern 24:85–101. CrossRef

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do we know what the early visual system does? J Neurosci 25:10577–10597. CrossRef Medline

Chambolle A, Caselles V, Cremers D, Novaga M, Pock T (2010) An introduction to total variation for image analysis. In: Theoretical foundations and numerical methods for sparse recovery, pp 263–340. Berlin: Walter de Gruyter.

Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499:295–300. CrossRef Medline

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network 12:199–213. CrossRef Medline

Clark DA, Bursztyn L, Horowitz MA, Schnitzer MJ, Clandinin TR (2011) Defining the computational structure of the motion detector in *Drosophila*. Neuron 70:1165–1177. CrossRef Medline

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 17:296–303. CrossRef Medline

Clifford CW, Ibbotson MR (2002) Fundamental mechanisms of visual motion detection: models, cells and functions. Prog Neurobiol 68:409–437. CrossRef Medline

DeAngelis GC, Ohzawa I, Freeman RD (1993) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. J Neurophysiol 69:1118–1135. Medline

De Valois RL, Cottaris NP (1998) Inputs to directionally selective simple cells in macaque striate cortex. Proc Natl Acad Sci U S A 95:14488–14493. CrossRef Medline

De Valois RL, Cottaris NP, Mahon LE, Elfar SD, Wilson JA (2000) Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. Vision Res 40:3685–3702. CrossRef Medline

Egelhaaf M, Borst A (1993) A look into the cockpit of the fly: visual orientation, algorithms, and identified neurons. J Neurosci 13:4563–4574. Medline

Eichner H, Joesch M, Schnell B, Reiff DF, Borst A (2011) Internal structure of the fly elementary motion detector. Neuron 70:1155–1164. CrossRef Medline

Euler T, Detwiler PB, Denk W (2002) Directionally selective calcium signals in dendrites of starburst amacrine cells. Nature 418:845–852. CrossRef Medline

Fischbach KF, Dittrich AP (1989) The optic lobe of *Drosophila melanogaster*. Cell Tissue Res 258:441–475.

Fisher YE, Leong JC, Sporar K, Ketkar MD, Gohl DM, Clandinin TR, Silies M (2015a) A class of visual neurons with wide-field properties is required for local motion detection. Curr Biol 25:3178–3189. CrossRef Medline

Fisher YE, Silies M, Clandinin TR (2015b) Orientation selectivity sharpens motion detection in *Drosophila*. Neuron 88:390–402. CrossRef Medline

Fitzgerald JE, Clark DA (2015) Nonlinear circuits for naturalistic visual motion estimation. eLife Sci e09123.

Fitzgerald JE, Katsov AY, Clandinin TR, Schnitzer MJ (2011) Symmetries in stimulus statistics shape the form of visual motion estimators. Proc Natl Acad Sci U S A 108:12909–12914. CrossRef Medline

Gordon MD, Scott K (2009) Motor control in a *Drosophila* taste circuit. Neuron 61:373–384. CrossRef Medline

Gruntman E, Turner GC (2013) Integration of the olfactory code across dendritic claws of single mushroom body neurons. Nat Neurosci 16:1821–1829. CrossRef Medline

Haikala V, Joesch M, Borst A, Mauss AS (2013) Optogenetic control of fly optomotor responses. J Neurosci 33:13927–13934. CrossRef Medline

Hassenstein B, Reichardt W (1956) Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. Z Naturforschung B 11:513–524.

Hu Q, Victor JD (2010) A set of high-order spatiotemporal stimuli that elicit motion and reverse-phi percepts. J Vis 9:1–16. CrossRef Medline

Jagadeesh B, Wheat HS, Ferster D (1993) Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. Science 262:1901–1904. CrossRef Medline

Joesch M, Schnell B, Raghu SV, Reiff DF, Borst A (2010) ON and OFF pathways in *Drosophila* motion vision. Nature 468:300–304. CrossRef Medline

Joesch M, Weber F, Eichner H, Borst A (2013) Functional specialization of parallel motion detection circuits in the fly. J Neurosci 33:902–905. CrossRef Medline

Kato S, Xu Y, Cho CE, Abbott LF, Bargmann CI (2014) Temporal responses of *C. elegans* chemosensory neurons are preserved in behavioral dynamics. Neuron 81:616–628. CrossRef Medline

Kingma D, Ba J (2014) Adam: a method for stochastic optimization. 1412:6980. arXiv:1412.6980 [cs.LG].

Livingstone MS, Conway BR (2007) Contrast affects speed tuning, space-time slant, and receptive-field organization of simple cells in macaque V1. J Neurophysiol 97:849–857. CrossRef Medline

Maisak MS, Haag J, Ammer G, Serbe E, Meier M, Leonhardt A, Schilling T, Bahl A, Rubin GM, Nern A, Dickson BJ, Reiff DF, Hopp E, Borst A (2013) A directional tuning map of *Drosophila* elementary motion detectors. Nature 500:212–216. CrossRef Medline

Mank M, Santos AF, Direnberger S, Mrsic-Flogel TD, Hofer SB, Stein V, Hendel T, Reiff DF, Levelt C, Borst A, Bonhoeffer T, Hübener M, Griesbeck O (2008) A genetically encoded calcium indicator for chronic in vivo two-photon imaging. Nat Methods 5:805–811. CrossRef Medline

McCann GD (1973) The fundamental mechanism of motion detection in the insect visual system. Kybernetik 12:64–73. CrossRef Medline

Meier M, Serbe E, Maisak MS, Haag J, Dickson BJ, Borst A (2014) Neural circuit components of the *Drosophila* OFF motion vision pathway. Curr Biol 24:385–392. CrossRef Medline

Mukamel EA, Nimmerjahn A, Schnitzer MJ (2009) Automated analysis of cellular signals from large-scale calcium imaging data. Neuron 63:747–760. CrossRef Medline

Nitzany EI, Victor JD (2014) The statistics of local motion signals in naturalistic movies. J Vis 14:pii10. CrossRef Medline

Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC (2005) Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. Nature 433:597–603. CrossRef Medline

Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Systems Man Cybern 9:62–66. CrossRef

Peng Y, Ganesh A, Wright J, Xu W, Ma Y (2012) RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Trans Pattern Anal Mach Intell 34:2233–2246. CrossRef Medline

Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D,

Reardon T, Mu Y, Lacefield C, Yang W, Ahrens M, Bruno R, Jessell TM, Peterka DS, Yuste R, Paninski L (2016) Simultaneous denoising, deconvolution, and demixing of calcium imaging data. Neuron 89: 285–299. CrossRef Medline

Priebe NJ, Lampl I, Ferster D (2010) Mechanisms of direction selectivity in cat primary visual cortex as revealed by visual adaptation. J Neurophysiol 104:2615–2623. CrossRef Medline

Reichardt W, Poggio T (1976) Visual control of orientation behaviour in the fly: I. A quantitative analysis. Q Rev Biophys 375:428 – 438. Medline

Rokach L, Maimon O (2005) Clustering methods. In: Data mining and knowledge discovery handbook, pp 321–352. New York: Springer.

Serbe E, Meier M, Leonhardt A, Borst A (2016) Comprehensive characterization of the major presynaptic elements to the *Drosophila* OFF motion detector. Neuron 89:829 – 841. CrossRef Medline

Shinomiya K, Karuppudurai T, Lin TY, Lu Z, Lee CH, Meinertzhagen IA (2014) Candidate neural substrates for off-edge motion detection in *Drosophila*. Curr Biol 24:1062–1070. CrossRef Medline

Strother JA, Nern A, Reiser MB (2014) Direct observation of ON and OFF pathways in the *Drosophila* visual system. Curr Biol 24:976 – 983. CrossRef Medline

Takemura SY, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P, Zhao T, Horne JA, Fetter RD, Takemura S, Blazek K, Chang LA, Ogundeyi O, Saunders MA, Shapiro V, Sigmund C, et al. (2013) A visual motion detection circuit suggested by *Drosophila* connectomics. Nature 500:175–181. CrossRef Medline

Wang JW, Wong AM, Flores J, Vosshall LB, Axel R (2003) Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. Cell 112:271–282. CrossRef Medline

Wilson RI, Turner GC, Laurent G (2004) Transformation of olfactory representations in the *Drosophila* antennal lobe. Science 303:366 – 370. CrossRef Medline

Yang HH, St-Pierre F, Sun X, Ding X, Lin MZ, Clandinin TR (2016) Subcellular imaging of voltage and calcium signals reveals neural processing in vivo. Cell 166:1–13. CrossRef Medline