Modeling spatial contrast sensitivity in responses of primate retinal ganglion cells to natural movies

Shashwat Sridhar [1,2] ✉, Michaela Vystrčilová [3], Mohammad H. Khani* [1,2], Dimokratis Karamanlis** [1,2], Helene M. Schreyer* [1,2], Varsha Ramakrishna [1,2,4], Steffen Krüppel [1,2], Sören J. Zapp [1,2], Matthias Mietsch [5,6], Alexander S. Ecker [2,3,7,8], and Tim Gollisch [1,2,8,9] ✉

[1]University Medical Center Göttingen, Department of Ophthalmology, Göttingen, Germany
[2]Bernstein Center for Computational Neuroscience Göttingen, Göttingen, Germany
[3]University of Göttingen, Institute of Computer Science and Campus Institute Data Science, Göttingen, Germany
[4]International Max Planck Research School for Neurosciences, Göttingen, Germany
[5]German Primate Center, Laboratory Animal Science Unit, Göttingen, Germany
[6]German Center for Cardiovascular Research, Partner Site Göttingen, Göttingen, Germany
[7]Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany
[8]Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, Göttingen, Germany
[9]Else Kröner Fresenius Center for Optogenetic Therapies, University Medical Center Göttingen, Göttingen, Germany

✉ For correspondence
shashwat.sridhar@med.uni-goettingen.de; tim.gollisch@med.uni-goettingen.de
* Present address: Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland
** Present address: University of Geneva, Department of Basic Neurosciences, Geneva, Switzerland

Abstract

Retinal ganglion cells, the output neurons of the vertebrate retina, often display nonlinear summation of visual signals over their receptive fields. This creates sensitivity to spatial contrast, letting the cells respond to spatially structured visual stimuli even when no net change in overall illumination of the receptive field occurs. Yet, computational models of ganglion cell responses are often based on linear receptive fields, and typical nonlinear extensions, which separate receptive fields into nonlinearly combined subunits, are often cumbersome to fit to experimental data. Previous work has suggested to model spatial-contrast sensitivity in responses to flashed images by combining signals from the mean and variance of light intensity inside the receptive field. Here, we extend and adjust this spatial contrast model for application to spatiotemporal stimulation and explore its performance on spiking responses that we recorded from ganglion cells of marmosets under artificial and naturalistic movies. We show how the model can be fitted to experimental data and that it outperforms common models with linear spatial integration to different degrees for different types of ganglion cells. Finally, we use the model framework to infer the cells' spatial scale of nonlinear spatial integration. Our work shows that the spatial contrast model can capture aspects of nonlinear spatial integration in the primate retina with only few free parameters. The model can be used to assess the cells' functional properties under natural stimulation and provides a simple-to-obtain benchmark for comparison with more detailed nonlinear encoding models.

Author Summary

Our visual experience depends on the retina's remarkable ability to detect light patterns and contrast in the world around us. Retinal ganglion cells, the output neurons of the retina, modulate their activity based on signals within small, specific regions of the visual scene, called their receptive fields. But many cells do not only encode overall brightness, summed linearly across the receptive field, but are also sensitive to local spatial contrast, that is, variations in brightness within the receptive field. Computational models that account for this nonlinear spatial integration exist, but require large amounts of data and are challenging to fit. We therefore developed the spatial contrast model, which takes a simple measure of light-intensity variations as an input, and tested it on measured responses of primate retinal ganglion cells to both artificial and naturalistic movies. The model substantially outperformed standard models with linear receptive fields, despite having only one additional tunable parameter. Furthermore, we used the model to investigate the spatial scale at which the cells integrate spatial contrast and found striking consistency across cell types. The spatial contrast model thus offers a practical tool for capturing retinal stimulus encoding and a simple-to-obtain benchmark for modeling nonlinear spatial integration.

1 Introduction

The ganglion cells of the vertebrate retina encode visual stimuli that occur inside their receptive fields. When they are activated by an appropriate stimulus, the cells report this by generating spikes for transmission along their axons, the fibers of the optic nerve. The most basic principle of this encoding is that ON ganglion cells are activated by an increase of light intensity in their receptive field centers, OFF ganglion cells by a decrease of light intensity, and ON-OFF cells by light-intensity changes of either sign (Hartline, 1938). In computational models of retinal ganglion cell activity, this is often captured by treating the receptive field as a (linear) filter of the incoming light-intensity pattern. Applying this filter amounts to computing a weighted spatial summation of the change in illumination, with weights given by the profile of the receptive field over space. The obtained filter activation can then be transformed into an actual response, such as a firing rate, in a way that takes the preferred contrast polarity of the ganglion cell — ON, OFF, or ON-OFF — into account. In its most basic form, this procedure corresponds to the commonly used linear–nonlinear (LN) model (Chichilnisky, 2001), where the receptive-field filtering represents the model's linear stage and the subsequent transformation into the cell's response the model's nonlinear stage.

Yet, despite the success of the LN model for describing retinal ganglion cells, as well as of its many variants that keep the linear receptive field for spatial signal integration, it has long been known that many ganglion cells display nonlinear receptive fields (Enroth-Cugell and Robson, 1966). Typically, this is demonstrated by responses to stimulation with contrast-reversing spatial gratings, showing that gratings can evoke sizeable responses even when the contrast reversals do not activate the linear receptive field because light-intensity increments and decrements symmetrically balance each other. The nonlinearities of the receptive field then lead to a characteristic frequency doubling in the response, as contrast reversals in both directions can activate the cell about equally strongly.

The receptive-field nonlinearities are, for some ganglion-cell types, considered a crucial component to be included in computational models for accurately capturing responses to natural stimuli (Gollisch, 2013; Heitman et al., 2016; Karamanlis et al., 2022; Schwartz & Rieke, 2011; Turner et al., 2019; Zapp et al., 2022). A typical approach is to separate the receptive field into multiple smaller spatial filters, so-called subunits, and to then nonlinearly combine the resulting filter signals (Victor & Shapley, 1979).

Several methods for inferring subunits from experimental data have been suggested (e.g., Freeman et al., 2015; Karamanlis et al., 2025; Krüppel et al., 2024; Maheswaranathan et al., 2018; Shah et al., 2020), which, in addition to serving as the basis for modeling, may also provide insights into retinal circuitry, as the subunits are thought to correspond to the receptive fields of presynaptic bipolar cells (Borghuis et al., 2001; Demb et al., 2001; Liu et al., 2017). Yet, subunit inference remains a formidable challenge, requiring long recordings and computationally demanding analyses. Moreover, it is not clear how much the actual precise subunit locations matter for response predictions (Freedland & Rieke, 2022).

As a simpler alternative to subunit models, one may therefore consider capturing nonlinear spatial integration by including information about higher-order statistics of light intensities inside the receptive field. Indeed, prior work on modeling responses from salamander and mouse retinal ganglion cells to flashed natural images has shown that a substantial improvement over the LN model can be obtained by simply providing a measure of spatial contrast inside the receptive field as an additional input channel (Liu et al., 2022). Here, we extend and modify this spatial contrast model to also include temporal integration, thus allowing application to spatiotemporal stimulation. Moreover, we apply the model to data recorded from the primate (marmoset) retina in response to artificial and natural movies and demonstrate that the spatial contrast model can outperform the LN model in a cell-type-dependent and stimulus-dependent manner. In addition, we use the model to probe the optimal scale of spatial contrast integration of different primate retinal ganglion cell types by introducing a smoothing step after temporal filtering and find that, across cell types, this optimal scale falls within a narrow range, which corresponds to known sizes of bipolar cell dendritic trees in the marmoset.

2 Results

2.1 Spatial-contrast-driven nonlinear integration in retinal ganglion cells

To investigate how spatial contrast inside the receptive fields of primate retinal ganglion cells affects stimulus encoding under natural movies, we recorded the spiking activity from marmoset ganglion cells in response to various stimuli with microelectrode arrays. To be able to assess the stimulus inside a cell's receptive field, we used reverse-correlation analysis under stimulation with spatiotemporal white noise to determine, for each recorded cell, its spatial receptive field and temporal filter (Fig. 1B).

In order to study cell-type dependent differences, we assigned the cells to clusters based on their receptive field sizes, their temporal filters, and the autocorrelograms of their spike trains (see *Methods*). We focused on four classes of cells: OFF midget cells, OFF parasol cells, ON parasol cells, and Large OFF cells (Fig. 1A). OFF midget, OFF parasol, and ON parasol cells were obtained as well-separated clusters, each with consistent autocorrelograms and temporal filters and with receptive fields of similar sizes, which approximately tiled the retinal surface (despite gaps in the mosaics from cells which were not recorded). The Large OFF cells in our analysis were a collection of cells from several clusters with somewhat more heterogeneous properties and likely belonged to several cell types, which were nonetheless pooled together here, as the boundaries between the contributing clusters were less clear than for midget and parasol cells.

These different classes of ganglion cells have different sensitivity to spatial structure inside their receptive fields, owing to different degrees of nonlinear spatial integration they perform. For example, compared to the relatively linear midget cells, the more nonlinear parasol cells typically displayed more pronounced and frequency-doubled responses under contrast-reversing gratings, as we also found in our recordings (Fig. 1C). We captured the nonlinearity of each cell's response to such gratings by a reversing-grating index, which was computed from the trial-averaged responses as the ratio of the spectral power at twice the stimulus frequency to that at the stimulus frequency. This index had similar values across

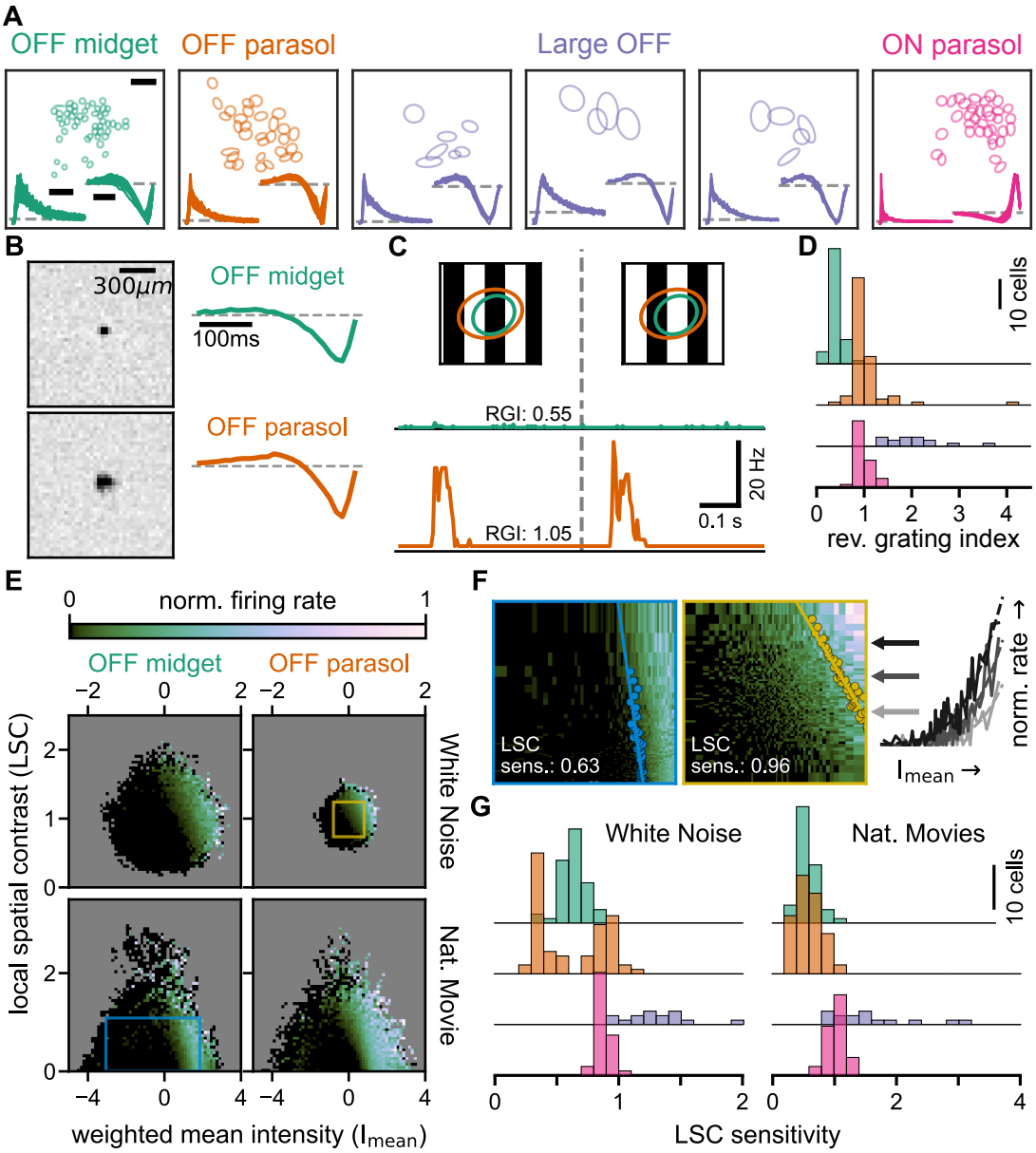

Figure 1: **Spatial receptive-field nonlinearities of marmoset retinal ganglion cells.** **A**: Layouts of receptive fields (shown as 1.5-σ contours), spike-train autocorrelograms (lower left), and temporal filters (lower right) of four analyzed cell classes, each showing data from one experiment. The Large OFF cells pool cells from three clusters, which are shown separately. Scale bars: spatial: 300 µm, temporal filters: 100 ms, autocorrelograms: 25 ms. **B**: Spatial receptive field (left) and temporal filter (right) of an OFF midget cell (top row) and an OFF parasol cell (bottom row). **C**: Firing rate profiles of the two sample cells from B (green trace: OFF midget cell; orange trace: OFF parasol cell) in response to reversing gratings (60 µm spatial period, shown on top; ellipses denoting the 1.5-σ receptive-field contours). RGI specifies the reversing grating index of the corresponding cell. **D**: Distributions of the reversing grating index for all four cell classes. **E**: Firing rate histograms with equally spaced bins from the training segments for white-noise (WN, top) and naturalistic-movie (NM, bottom) stimulation, based on mean light intensity (I_{mean}) and local spatial contrast (LSC) for the OFF midget (left) and OFF parasol (right) cells from B. Firing rates are normalized to the histogram maximum, with empty bins shown in gray. **F**: Firing rate histograms with quantile bins computed for regions corresponding to blue and yellow boxes in E. Data points used to compute the LSC sensitivity are overlaid, along with the linear fit. Right: Firing rate data (solid lines) of three selected histogram rows (marked by arrows of corresponding color) and fitted softplus functions (dashed lines). **G**: Distribution of LSC sensitivity for all four cell classes under white noise (left) and naturalistic movies (right).

cells within each class (Fig. 1D) and highlighted differences between midget, parasol, and Large OFF cell clusters.

To understand how this difference manifests itself in responses to richer stimuli with more spatiotemporal complexity, we stimulated the retina with spatiotemporal white noise as well as with a naturalistic movie. The white-noise stimulus consisted of a flickering checkerboard pattern, while the naturalistic movie comprised segments from a publicly available movie which were modified by shifting the movie frames according to a model of gaze shifts, including saccades and fixational eye movements.

To assess how the spatial structure of the stimulus shapes the response of a ganglion cell to such stimuli beyond the simple, linear activation of its receptive field, we quantified the effective mean light intensity as well as the spatial structure inside the receptive field at each time point of the stimulus. To do so, we first filtered the stimulus by convolving it with the temporal filter of the cell, as obtained from the reverse-correlation analysis, to take care of the cell's temporal stimulus integration.

We then measured the mean light intensity signal, $\mathrm{I_{mean}}$, as the weighted average of pixel intensity values (given by their contrast relative to the mean stimulus intensity over time) in the receptive field of the cell. The weights were given by the cell's spatial filter, here defined as a Gaussian fit to the spatial component of the spike-triggered average under white noise. $\mathrm{I_{mean}}$ is a measure of the activation of the linear part of the receptive field and is equivalent to the input that is commonly used for computational models that do not account for receptive-field nonlinearities.

We furthermore measured the local spatial contrast, LSC, in each frame of the temporally filtered stimulus by computing the weighted standard deviation of the pixel intensity values within the receptive field of the cell, where the weights were again given by the spatial filter. The LSC provides a measure of the effective spatial structure in the stimulus that the cell was exposed to at a given point in time.

We analyzed how the cell's activity depended on $\mathrm{I_{mean}}$ and LSC by compiling two-dimensional histograms of average firing rates conditioned on different combinations of $\mathrm{I_{mean}}$ and LSC, as exemplified in Fig. 1E for one OFF midget (left) and one OFF parasol cell (right) and both stimulus conditions. Note that the regions of available combinations of $\mathrm{I_{mean}}$ and LSC differ between the two stimulus conditions. Histograms computed from spatiotemporal white-noise are distributed around $\mathrm{I_{mean}} = 0$ and $\mathrm{LSC} = 1$ with approximate circular symmetry, but those computed from naturalistic movies cover a broader range of $\mathrm{I_{mean}}$ values at low LSC and a narrower range of $\mathrm{I_{mean}}$ values at high LSC, indicating that the stimulus is dominated by low-spatial-contrast scenes. Furthermore, the range of covered $\mathrm{I_{mean}}$ and LSC values under white-noise stimulation can be different for different cells because larger receptive fields, as for the sample parasol cell versus the sample midget cell, lead to more averaging over stimulus pixels and thus smaller ranges of $\mathrm{I_{mean}}$ and LSC values.

The histograms reveal that the firing rates of the two cells not only depended on $\mathrm{I_{mean}}$, but are also affected by LSC. For example, if LSC had no effect on the firing rate, that is, if stimulus integration in the receptive field were completely linear, regions of increased firing rate, corresponding to brighter regions in the plot, should not depend on the value along the y-axis and rather be fully vertically oriented. Instead, these regions had a slanted orientation, indicating higher firing rate towards the upper right corner of the plot, where both $\mathrm{I_{mean}}$ and LSC are large, in particular for the sample parasol cell. For the sample midget cell, this tendency was less pronounced, suggesting that the LSC had somewhat less influence over the cell's firing.

To quantify the dependence of the firing rate on LSC for each cell and each stimulus condition, we extracted a threshold $\mathrm{I_{mean}}$ value for each row in the histogram, that is, for each level of LSC. We did so by fitting a family of parametrized curves to the firing rates of the histogram rows, which had identical shapes for all rows, but allowed for a horizontal shift, whose value defined the threshold value for the corresponding level of LSC (Fig. 1F, see *Methods*). We observed that these threshold values generally lay approximately on a straight line. We used the slope of this line, extracted from a linear fit (Fig. 1F),

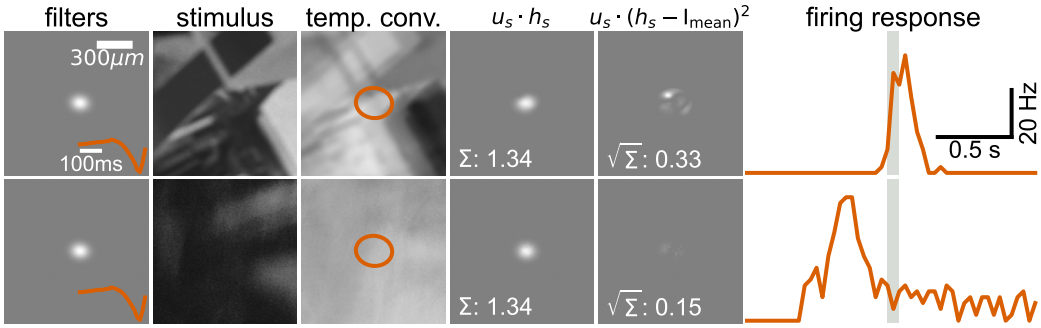

Figure 2: **Examples of the effect of local spatial contrast on responses to natural scenes.** Illustration of the stimulus processing to obtain $\mathrm{I}_{\mathrm{mean}}$ and LSC by applying the filters of a sample OFF parasol cell (1st column, same cell as in Fig. 1B–F) to the stimulus (2nd column, exemplified here by a single frame) to obtain a temporally convolved stimulus (3rd column, overlaid with the 3-σ contour of the spatial filter), from which the weighted light intensity signal (4th column) and the weighted spatial contrast (5th column) are extracted. The two rows correspond to two separate stimulus segments. The plots of the light-intensity and spatial-contrast signals display the results of the pixel-wise computation as indicated by the formulas at the top, with the final values given in the lower left of the plots. The corresponding measured firing rates of the sample cell are shown on the right, with the light gray region highlighting the time point of presentation of the analyzed stimulus frame for which the $\mathrm{I}_{\mathrm{mean}}$ and LSC values were calculated here.

to define the LSC sensitivity. The LSC sensitivity is a measure of the relative importance of the LSC to that of the $\mathrm{I}_{\mathrm{mean}}$ in determining the cell's spiking response, taking on a values of zero when the LSC has no effect and positive values when, for the same $\mathrm{I}_{\mathrm{mean}}$ values, larger LSC leads to higher firing rate. The LSC sensitivity can thus be seen as a measure of receptive field nonlinearity, equivalent to the reversing gratings index, but assessed under conditions of naturalistic stimulation rather than reversing gratings.

We found that the distribution of the LSC sensitivity is largely consistent across cells within a cell type (Fig. 1G). Moreover, cell-type specific distributions of the LSC sensitivity have relative positions similar to those of the reversing grating index (Fig. 1D), for example, Large OFF cells had the largest values in all conditions, whereas OFF midget cells generally had comparatively small values for both the reversing grating index and the LSC sensitivity. However, some discrepancies are also apparent. In particular, OFF parasol cells displayed a bimodal distribution of LSC sensitivity values under white-noise stimulation (Fig. 1G, left panel), with the two peaks corresponding to OFF parasol cells from different experimental datasets. The reason for this bimodality is unclear. While this might reflect experimental variability, differences in retinal locations or eccentricity of the extracted tissue might have contributed here, in particular considering the known influence of eccentricity on factors such as receptive field size (Masri et al., 2019), amacrine cell density (Weltzien et al., 2015) and cone photoreceptor density (Goodchild et al., 1996). Furthermore, we observed that ON parasol cells had relatively high values of LSC sensitivity under naturalistic movies, clearly larger than those of OFF parasol cells and nearly on the level of Large OFF cells, suggesting that the ON parasol cells displayed particularly strong receptive-field nonlinearities under naturalistic stimulation.

To illustrate the relation between LSC and firing rate, Fig. 2 shows examples of the $\mathrm{I}_{\mathrm{mean}}$ and LSC computation together with firing-rate responses from a sample OFF parasol cell for two different segments of the naturalistic movie. Each row displays the computation of $\mathrm{I}_{\mathrm{mean}}$ and LSC at a given point in time from the corresponding temporally convolved movie frame, which here resulted in nearly identical $\mathrm{I}_{\mathrm{mean}}$ signals (4th column), but substantially different LSC signals (5th column). Comparing the corresponding firing rates (Fig.2, right), shows that the larger LSC value (top row) yielded a much larger firing rate at the time point of interest (light gray region), compared to the scenario with smaller LSC value (bottom row). Note that here the earlier firing-rate peak visible in the bottom row corresponds to an earlier stimulus

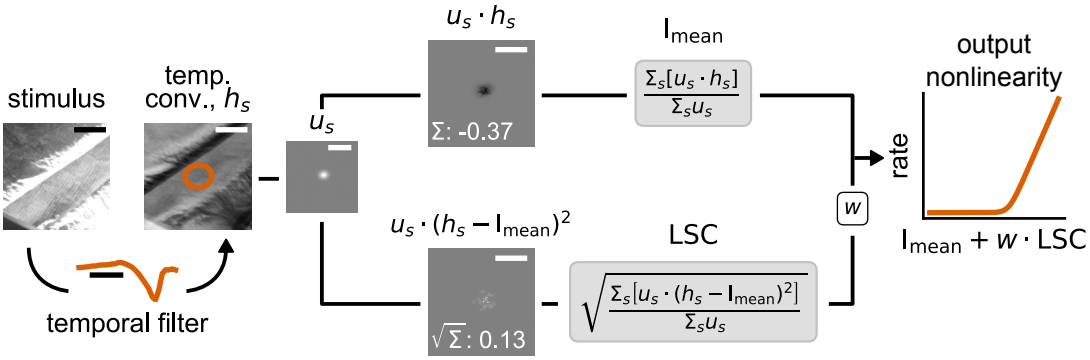

Figure 3: **Schematic of the spatial contrast (SC) model.** For a given cell, the incoming stimulus is first convolved with the cell's temporal filter to yield the effective stimulus frame h_s (overlaid here with the 3-σ contour of the spatial filter). From this, the mean light intensity signal $\mathrm{I}_{\mathrm{mean}}$ and the spatial contrast signal LSC are computed by using the the cell's spatial filter u_s as weights. The $\mathrm{I}_{\mathrm{mean}}$ and LSC are linearly combined with a weight parameter w used to scale the LSC contribution, and the combined signal is transformed by an output nonlinearity to yield the model's firing rate response. Scale bars: spatial: 300 µm, temporal: 100 ms.

segment and thus to different $\mathrm{I}_{\mathrm{mean}}$ and LSC values. These examples illustrate how spatial contrast can shape the firing response of ganglion cells, corroborating that the LSC can be used to capture these effects in computational models.

2.2 Predictive model with spatial-contrast sensitivity

In order to incorporate the LSC into a predictive model for responses to spatiotemporal stimuli, we extended the spatial contrast (SC) model that had previously been used to predict responses to flashed natural images in the salamander retina (Liu et al., 2022). The SC model itself is an extension of the commonly used linear–nonlinear (LN) model (Chichilnisky, 2001). In the linear part of the LN model, the stimulus is filtered with the receptive field of the cell to yield the mean light intensity in the stimulus, weighted by the receptive field, equivalent to computing the cell's $\mathrm{I}_{\mathrm{mean}}$ signal. The nonlinear part of the LN model then transforms the weighted mean light intensity into the model's response (e.g., firing rate or spike probability).

In the SC model, in addition to the $\mathrm{I}_{\mathrm{mean}}$ signal, the LSC is measured for the stimulus, and the two are linearly combined, using a weight parameter w, which determines the importance of the local spatial contrast relative to the mean light intensity signal. This enables the SC model to scale the dependence on LSC independently for each cell. The linearly combined signal is then transformed into the model's response by the output nonlinearity, much like in the LN model. For both models, we here applied a softplus rectifying nonlinearity that is parameterized with three free parameters.

In our spatiotemporal version of the SC model, the calculation of $\mathrm{I}_{\mathrm{mean}}$ and LSC is preceded by a temporal convolution of the stimulus with the cell's temporal stimulus filter. The full model is summarized in the schematic of Fig. 3. Together, implementing the SC model for a given cell requires estimating the cell's spatial receptive field and temporal filter, which can both be obtained by reverse correlation under spatiotemporal white-noise stimulation, as well as optimizing four free parameters, three for the softplus nonlinearity and one for the weight w of the LSC. We here optimized the free parameters simultaneously, using negative log-likelihood minimization under the assumption of a Poisson spiking process.

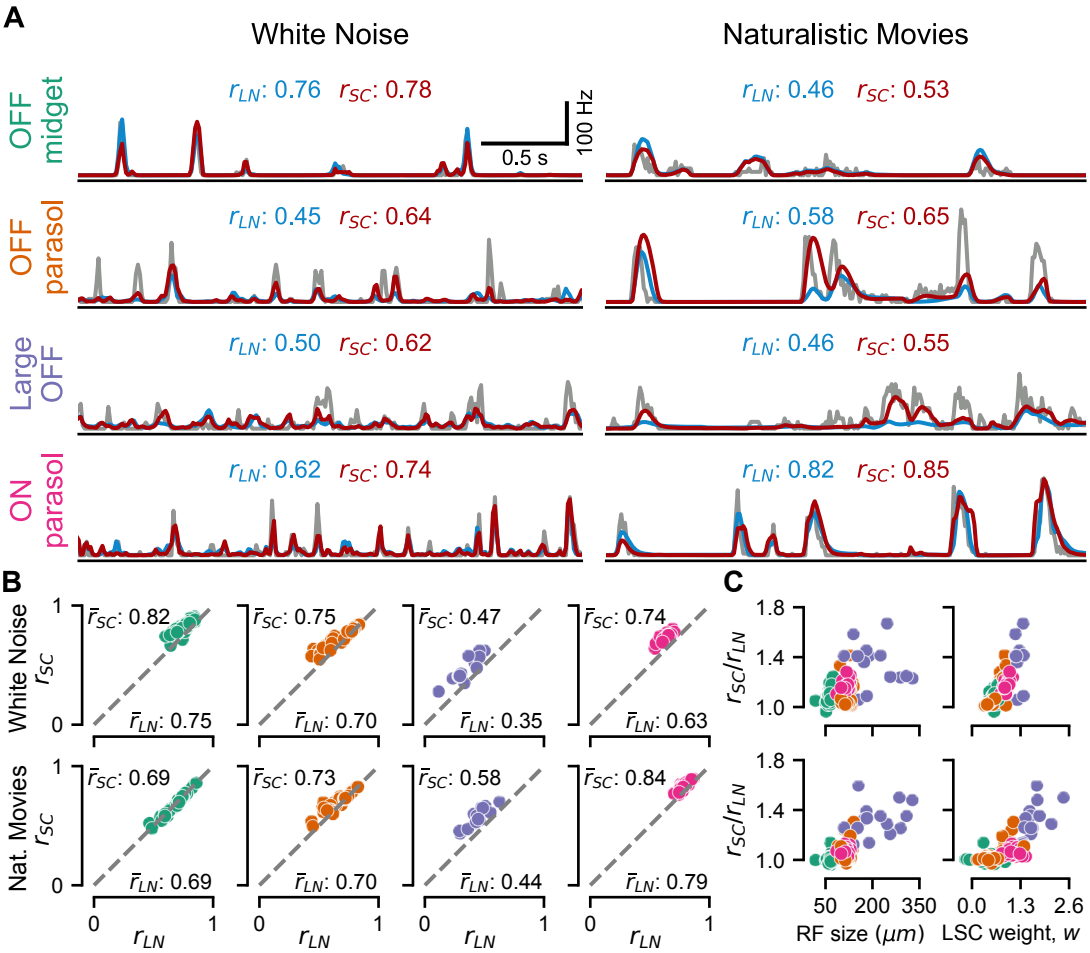

Figure 4: **Evaluation of model performance by cell type.** **A**: Responses of sample cells from each of the four cell classes in Fig. 1A to 3 s segments of test stimuli (gray) for white noise (left column) and naturalistic movie (right column), together with the predictions from the LN model (blue) and the SC model (red). The numbers above each plot show the Pearson's correlation coefficients between model prediction and cell response. **B**: Comparison of correlation coefficients for SC model (r_{SC}) and LN model (r_{LN}) for each cell in the four cell classes (colors corresponding to the labels in A) under spatiotemporal white noise (top row) and naturalistic movie (bottom row). The average correlation coefficients $\overline{r}_{SC}$ and $\overline{r}_{LN}$ are shown next to the corresponding axes. **C**: Relative improvement in model performance (r_{SC}/r_{LN}) under white noise (top row) and naturalistic movie (bottom row) for each cell versus receptive-field size (left) and the obtained LSC weight w (right). Colors indicate cell class as in B.

2.3 Cell-type-specific comparison of LN and SC model

We next tested whether the SC model improves model predictions as compared to the LN model for white noise as well as for naturalistic movies. Both stimuli were presented to the retina in several trials consisting of non-repeating segments used for model training, alternating with a fixed stimulus segment, displayed repeatedly and used for model testing. For both models, we used the same spatial filter, which was obtained as a two-dimensional Gaussian fit to the spatial receptive field. Model performances were evaluated for each cell on the time-dependent firing rate averaged across trials of the test segments by computing the Pearson correlation with the model prediction (Fig. 4A).

For spatiotemporal white noise, we found that the SC model yielded better model predictions across all cell classes (Fig. 4B, top row). The improvement was particularly notable for the ON parasol cells (average ratio of correlation values r_{SC}/r_{LN} ± standard deviation: 1.17 ± 0.03, $p < 0.0001$, one-sided Wilcoxon signed-rank test) and Large OFF cells (r_{SC}/r_{LN}: 1.41 ± 0.32, $p < 0.0002$), whereas OFF midget cells displayed a more modest improvement (r_{SC}/r_{LN}: 1.09 ± 0.05, $p < 0.0001$). OFF parasol cells exhibited greater variability in model improvement and a smaller average improvement compared to OFF midget cells (r_{SC}/r_{LN}: 1.08 ± 0.08, $p < 0.0001$), owing to the discrepancy of this cell type across the two analyzed datasets. On splitting the population of OFF parasol cells by dataset, we found that cells in one dataset were significantly more nonlinear than in the other, yet both populations displaying significantly better predictions by the SC model compared to the LN model (r_{SC}/r_{LN}: 1.16 ± 0.08, $p < 0.0001$, versus 1.02 ± 0.01, $p < 0.0001$). This aligns with our observation regarding the distribution of LSC sensitivity values for OFF parasol cells under white-noise stimulation (Fig. 1G).

For naturalistic movies (Fig. 4E), the performance gain by including LSC information in the model was generally lower than for white noise, yet the SC model still outperformed the LN model for ON parasol cells (r_{SC}/r_{LN}: 1.06 ± 0.03, $p < 0.0001$), OFF parasol cells (r_{SC}/r_{LN}: 1.05 ± 0.07, $p < 0.0001$) and Large OFF cells (r_{SC}/r_{LN}: 1.32 ± 0.13, $p < 0.0001$). However, no change in the prediction performance was observed for OFF midget cells (r_{SC}/r_{LN}: 1.00 ± 0.02, $p = 0.83$).

The finding that models of OFF midget cells profit least from including spatial contrast information is consistent with their weak responses to reversing gratings (Fig. 1D) and their smaller dependence on spatial contrast in firing rate histograms (Fig. 1E) as well as with the notion that midget cells are classically considered to have rather linear receptive field (Derrington et al., 1984). Note, though, that more recent reports have shown that midget cells also display substantial receptive-field nonlinearities, for example, as demonstrated in the peripheral macaque retina (Freedland and Rieke, 2022; Freeman et al., 2015).

The results suggest that cells with larger receptive fields have a greater improvement in model performance by including spatial contrast information. This is confirmed by plotting the ratio of model performances r_{SC}/r_{LN} against receptive field size (Fig. 4C, left). Cell types with larger receptive fields display greater relative improvement in model performance. For midget cells under white-noise stimulation, this trend is also apparent for cells of the same type (Pearson's r: 0.41, $p = 0.003$, two-tailed t-test). That larger cells are more sensitive to spatial contrast makes intuitive sense, as their larger receptive fields are more likely to actually experience substantial spatial structure. For OFF midget cells under natural stimuli, for example, there may simply be too little spatial structure available at the high spatial frequencies that would fit into their small receptive fields. Thus, unlike for white-noise stimuli where signal power extends to high frequencies, midget cells might simply not experience sufficient spatial contrast to shape their responses beyond the activation of the linear receptive field.

We also checked how the improvement in model performance is related to the weight parameter w, which sets the relative contribution of the spatial contrast information in determining the activation of the SC model. As expected, we found that larger relative improvements of the SC over the LN model

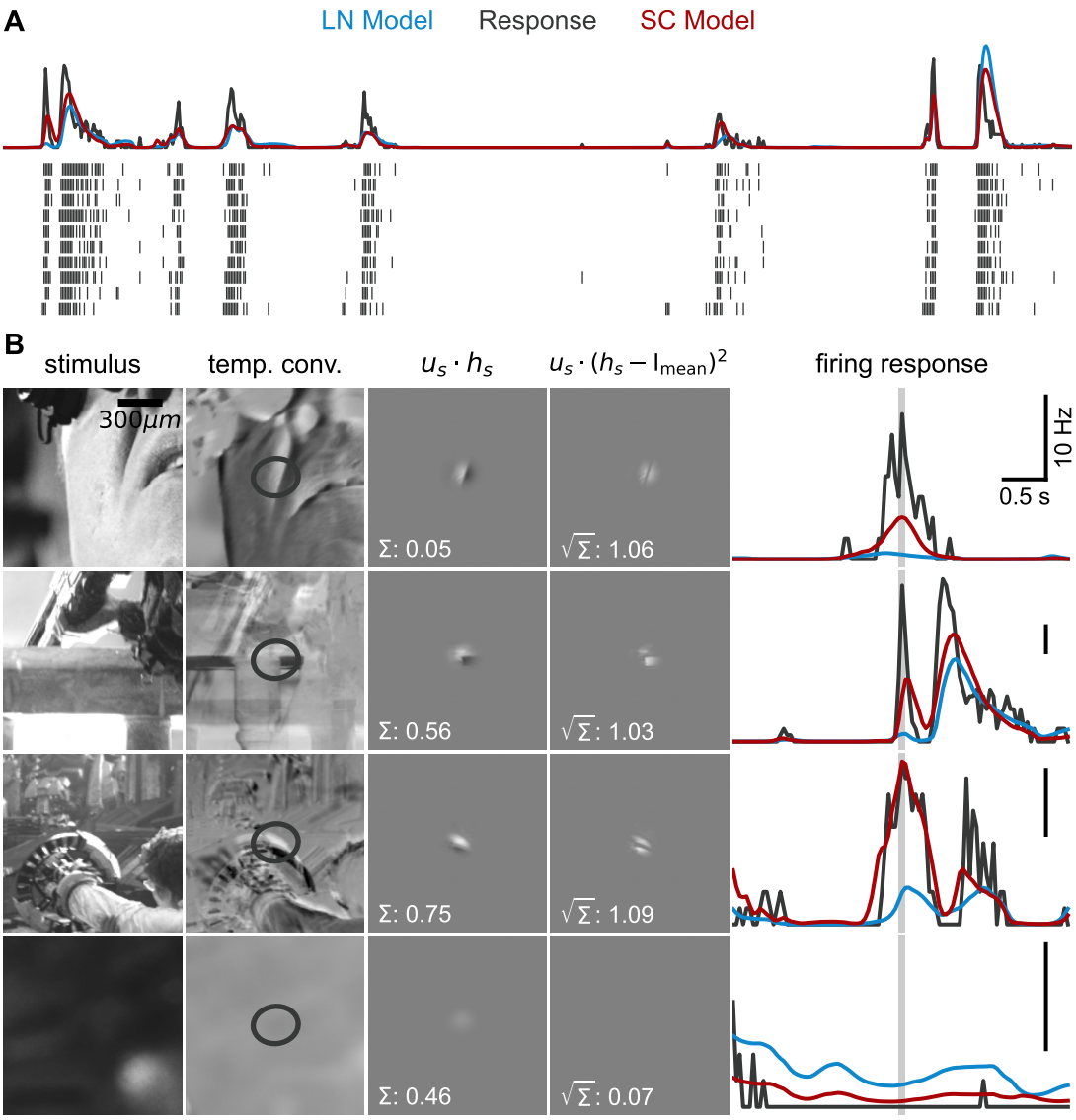

Figure 5: **Examples of improved predictions by SC model versus LN model.** **A**: Firing rate (top, black) and spike raster (bottom) for the sample OFF parasol from Fig. 1B to an 8 s segment of the naturalistic movie test stimulus. Overlaid are response predictions for the LN model (blue) and the SC model (red). **B**: Examples of episodes from the natural movie test stimulus with substantially better response prediction by the SC model versus the LN model for the same sample cell as in A. The columns are shown in the same way as in Fig. 2. Right-most panels are overlaid with response predictions with colors as in A.

were tightly associated with larger weights (Fig. 4C, right). A weight of zero reduces the SC model to an LN model, allowing no improvement, and larger positive weights reflect a stronger relevance of LSC on the cell's activity and thus a more nonlinear receptive field.

To illustrate how the SC model can lead to improvements in model prediction over the LN model, we examined the responses and model firing rates of a sample OFF parasol cell for different episodes taken from the test data of the naturalistic movie (Fig. 5A). In three of the four representative cases (Fig. 5B, first three rows), the LN model predicted a firing response lower than the ground truth, owing to the low average light intensity in the receptive field. Here, the temporally convolved stimulus, weighted by the cell's receptive field (Fig. 5B, 3rd column) displays bright as well as dark regions of similar extent, not unlike a grating stimulus. The substantial spatial structure inside the receptive field led to large values of the local spatial contrast, which in turn allowed the SC model to predict higher firing rates in response to the displayed stimulus episodes, which better matched the response of the cell.

Moreover, the additional spatial contrast information allowed the SC model to also predict a lower firing response than the LN model. This is due to the fact that, at the same mean light intensity, the SC model can predict different responses to different local spatial contrast values in the scene, whereas the LN model predicts the same, average response across the different values of local spatial contrast. The final example case (Fig. 5B, last row) illustrates this for one episode where the lack of spatial structure within the receptive field suppressed the prediction of the SC model, which was closer to the ground truth.

2.4 Identifying spatial scales of stimulus integration with the SC model

Visual input can contain information at a wide range of spatial frequencies, but not all of these are relevant for downstream sensory processing. Spatial contrast at very high spatial frequencies, corresponding to scales below the size of bipolar cell or even photoreceptor receptive fields, for example, is unlikely to matter even for a nonlinear ganglion cell. One may thus ask what the relevant spatial scales of integrating spatial contrast are for the analyzed marmoset retinal ganglion cells. This question can be targeted with the SC model as a tool by analyzing what the optimal spatial scale is for assessing local spatial contrast. To do so, we used a modified version of the SC model, in which the LSC is computed from the temporally convolved stimulus after additional spatial smoothing, here performed with a two-dimensional circular Gaussian filter (Fig. 6A). The smoothing eliminates spatial-contrast information on scales smaller than the spatial scale of the Gaussian filter but retains spatial-contrast information on larger scales. The computation of I_{mean}, on the other hand, is kept the same as in the original model.

For each cell, we thus trained the modified SC model for various spatial scales of smoothing (defined as three times the standard deviation of the Gaussian filter, in line with our definition of the size of a cell's receptive field). The resulting performance in predicting responses to the test stimulus was then compared with that of the original SC model without stimulus smoothing (Fig. 6B). For spatiotemporal white noise, a clear pattern emerged. Across all cell types, increasing the scale of spatial smoothing led, on average, to an initial increase in performance followed by a steady decrease (Fig. 6C, top row). The relative performance peaked at a spatial scale of around 40–50 µm (distribution of peaks across cells shown in insets of Fig. 6C, top row; mean ± standard deviation in units of µm: all cells: 48 ± 19; OFF midget cells: 45 ± 33, OFF parasol cells: 51 ± 4, Large OFF cells: 44 ± 12, ON parasol cells: 46 ± 2).

The magnitude of the model improvement obtained by assessing the local spatial contrast at its optimal scale can be read off from the peak size of the performance-improvement curves. This shows that the improvement of SC model performance by adjusting the spatial scale can be considerable, up to around 10% for individual ON and OFF parasol cells and even 20% for individual Large OFF cells. Over the population of all cells, the improvement was consistently substantial for OFF parasol cells (relative

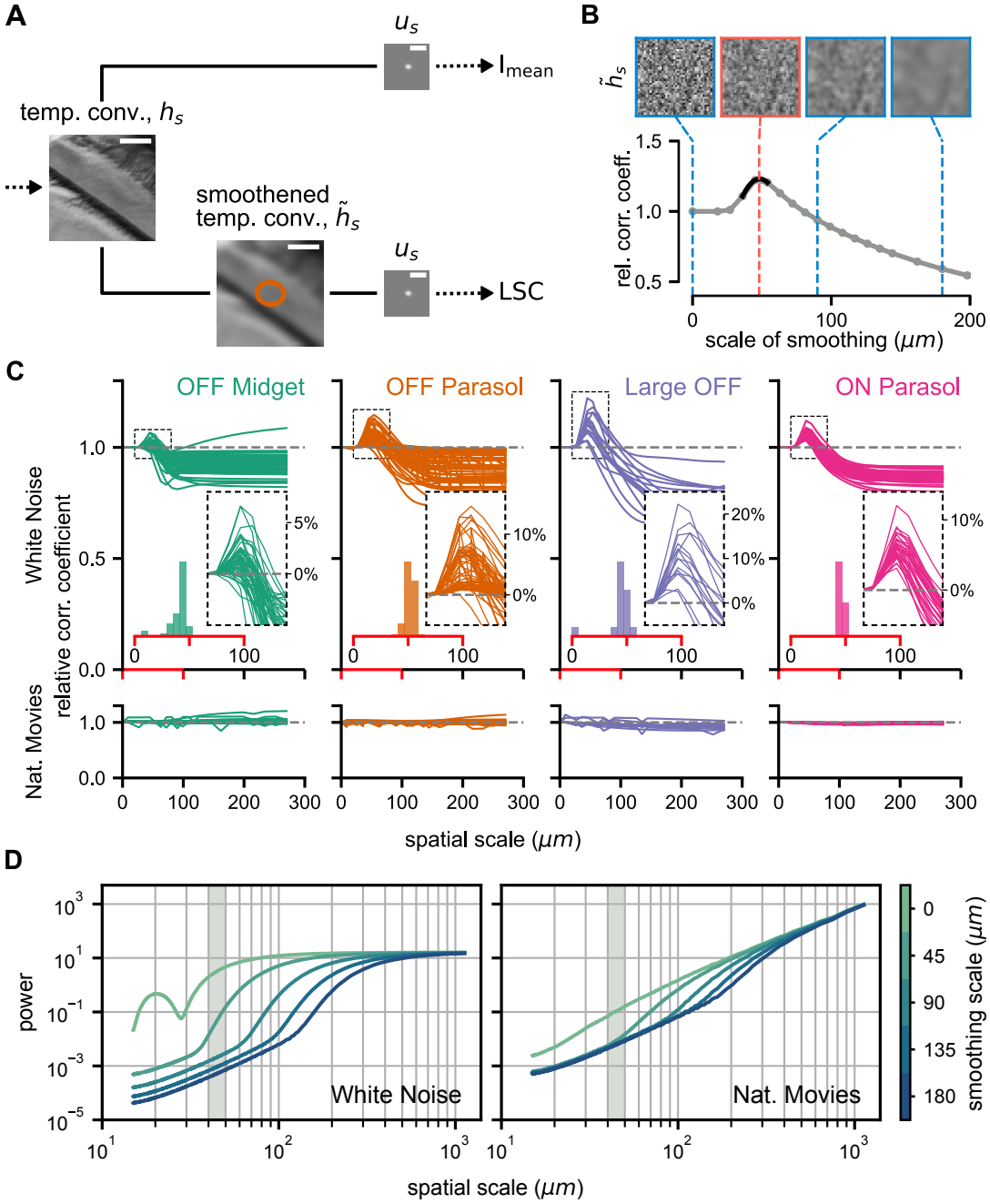

Figure 6: **Analysis of spatial scale of contrast sensitivity.** **A**: Schematic of the SC model from Fig. 2, modified to include spatial stimulus smoothing before the computation of LSC. **B**: Spatial smoothing analysis for a sample cell in response to white noise. Gray curve shows the SC model performance for different smoothing scales, normalized to performance without smoothing. The optimal scale is obtained by interpolation (black segment). Sample white-noise frames at different smoothing scales are shown on top, with the red outline marking the approximate optimal scale. **C**: Spatial smoothing analysis for all cells under white-noise (top) and naturalistic-movie (bottom) stimulation, with the region around the peak shown enlarged in the inset with dashed outline and relative model performance shown as percentage improvements over a model without smoothing. Distributions of optimal scales of spatial integration for each cell class under white noise are shown in the inset histograms. The x-axis here corresponds to the range of the red segment of main plot. **D**: Radially-averaged power spectral density (RAPSD) curves averaged across frames of white noise (left) and of the naturalistic movie (right) for different scales of spatial smoothing. Spatial scale (x-axis) is here defined as the inverse of spatial frequency. Shaded gray areas mark the range of 40–50 µm, matching approximately the range of optimal integration scales from C. Note that the RAPSD of the non-smoothed white-noise stimulus falls off for small spatial scales because RAPSD curves were here computed based on monitor-pixel resolution and the stimulus was close to white only for scales beyond the length of two stimulus squares.

Pearson's r: 1.05 ± 0.04, mean ± standard deviation, $p < 0.0001$, one-sided Wilcoxon signed-rank test), ON parasol cells (relative Pearson's r: 1.06 ± 0.02, $p < 0.0001$), and Large OFF cells (relative Pearson's r: 1.10 ± 0.06, $p < 0.0005$) cells, and less so (though still significant) for OFF midget cells (relative Pearson's r: 1.02 ± 0.02, $p < 0.0001$). This likely reflects the more nonlinear receptive fields of the parasol and Large OFF cells, whose models can therefore benefit more from properly assessing spatial contrast information.

For the naturalistic movie stimulus, the same analysis of different smoothing scales did not reveal optimal scales for any of the four cell classes (Fig. 6C, bottom row). Increasing the scale of spatial smoothing overall had little effect on model performance and did not systematically improve model performance for any of the cell types. This stark difference between the two stimuli can be explained by the fact that natural stimuli are dominated by low spatial frequencies. To illustrate this, we computed the radially-averaged power spectral density (RAPSD) for each stimulus and compared the spectra at different scales of spatial smoothing (Fig. 6D). For the white-noise stimulus, the spectrum of the non-smoothed stimulus is flat, as expected, for spatial scales that are larger than twice the size of the stimulus squares (30 µm), and smoothing has a strong effect on the power in the range of several tens of micrometers, in particular for spatial scales near the optimal spatial scales of around 40–50 µm extracted from Fig. 6C, thus substantially affecting the computation of LSC. For the naturalistic movie, on the other hand, the RAPSD curve falls off strongly towards smaller spatial scales even without smoothing, owing to the dominance of low spatial frequencies in natural scenes (Ruderman and Bialek, 1994; Simoncelli and Olshausen, 2001). Thus, the LSC is largely determined by spatial structure at larger scales and smoothing has little effect on the LSC or may even deteriorate the LSC signal, as it diminishes relevant spatial structure at scales beyond the optimal spatial scales of the cells.

3 Discussion

Despite the ubiquity of spatial nonlinearities within the receptive fields of retinal ganglion cells (Zapp et al., 2022), current models of retinal stimulus encoding still often start with a linear receptive field, implemented by a spatial filter that is applied to the incoming stimulus. The widespread use (and success) of these linear-receptive-field models is undoubtedly related to the relative ease with which their parameters, in particular the filters themselves, can be obtained from experimental data. Under Gaussian white noise stimulation, for example, the spike-triggered average can provide an accurate and easy-to-obtain estimate of the filter (Chichilnisky, 2001; Paninski, 2003). In addition, natural stimuli allow obtaining the filter via maximum-likelihood procedures (Paninski, 2004; Pillow et al., 2008; Truccolo et al., 2005), backed by a good mathematical understanding of the parameter optimization challenge, or via information-theoretic tools (Sharpee et al., 2004). This has supported the success of simple single-filter-based models, such as the LN model and its various extensions, which add model components that capture, for example, cell-intrinsic feedback (Keat et al., 2001; Pillow et al., 2005), dependencies on other neurons (Pillow et al., 2008), biophysical adaptation models (Ozuysal et al., 2018), and spike-timing dynamics (Aldworth et al., 2005; Gollisch, 2006).

Extending the models to include receptive-field nonlinearities, on the other hand, has presented a substantial challenge. A typical approach is to split up the receptive field into spatial subunits, which act as multiple, parallel filters and whose signals are nonlinearly combined, often by applying a threshold-linear rectification to the individual subunit signals before additive pooling (Karamanlis et al., 2025; Liu et al., 2017; Maheswaranathan et al., 2018; Real et al., 2017; Schwartz et al., 2012; Shah et al., 2020; Shi et al., 2019; Zapp et al., 2022). Yet, although such subunit models of retinal ganglion cells have been introduced almost half a century ago (Victor and Shapley, 1979), it has remained difficult to optimize subunit models with respect to data from actual, experimentally measured cells. Among the problems

here are that the optimal number of subunits is not known beforehand, which makes parameterizing subunit models difficult, and that the small spatial scale of subunits requires finely structured visual stimuli to resolve their boundaries, which often results in inefficient stimulation and the need for long recording durations.

We therefore turned to a simpler, phenomenological approach for including effects of nonlinear spatial integration into computational models, aiming to keep the advantage of conceptual model simplicity and relatively easy parameter optimization as in linear-receptive-field models. To do so, we consider the local spatial contrast (LSC) inside a cell's receptive field, which can be viewed as a first-order extension of the linear receptive field, as it depends on the square-deviation of pixel intensities from the mean intensity. Including LSC information can thereby capture spatial nonlinearities of the receptive field. A square-wave spatial grating that symmetrically covers the receptive field with its bright and dark regions, for example, yields zero activation of the linear receptive field but a high LSC signal.

We here first demonstrated how the LSC can be used to measure the level of spatial nonlinearity and sensitivity to spatial contrast in ganglion cells of the primate retina (Fig. 1 and Fig. 2). We then extended the spatial contrast (SC) model previously suggested for flashed images (Liu et al., 2022) to spatiotemporal stimuli by accounting for temporal integration before assessing the LSC (Fig. 3) and applying a modified measure of LSC that is more appropriate for the variations in receptive field sizes as encountered, for example, for the cell types of the primate retina. Evaluating on data that we recorded from the marmoset retina, we found that, compared to an LN model with linear receptive field, the SC model yielded improved response predictions under artificial as well as naturalistic stimulation in a cell-type-specific fashion (Fig. 4), owing to its ability to capture nonlinear spatial integration in the receptive field of the cell (Fig. 5). Finally, we demonstrated that the SC model can be employed to identify the spatial scale of nonlinear stimulus integration in the primate retina by comparing model performance with different levels of spatial smoothing prior to the computation of the local spatial contrast (Fig. 6).

3.1 Advantages and limitations of the SC model

An important feature of the SC model, in particular in comparison to subunit models, is that matching the model to a given recorded cell mostly requires an estimate of the receptive field and temporal filter, which can be obtained from standard stimuli and analyses, such as reverse correlation under spatiotemporal white-noise stimulation. Beyond this, the model has few free parameters – the weight factor of the local spatial contrast along with the parameters that describe the nonlinearity. Thus, as receptive field measurements are among the typical standard characterizations that form part of any recording in sensory systems, obtaining the SC model essentially amounts to fitting a handful of parameters, which is feasible even with little training data.

Obtaining the receptive field and the temporal filter from white-noise stimulation, however, disregards that adaptation to different stimulus statistics, such as those of natural stimuli, may change a cell's filtering characteristics (Lesica et al., 2007; Sharpee et al., 2006). An interesting future extension might thus be to directly fit the full model, including its spatial and temporal filter, under the stimulus statistics of interest and compare with the filtering characteristics under white noise. It may even be feasible to allow for different spatial filters for computing the mean light intensity signal and the local spatial contrast, though from a biological perspective it seems plausible that evaluation of light intensity and spatial contrast occurs with the same spatial weighting.

Another current simplification is the assumption that spatial and temporal filtering are separable. This does not only reduce the number of parameters required for describing the filters (and therefore makes the parameter estimation more robust to noise) but also conceptually simplifies the assessment of the local spatial contrast. Without assuming space-time separability, for example, it is not immediately clear how

the spatiotemporal weighting in the computation of the standard deviation should occur, as temporal variations in illumination most likely contribute differently (or not at all) to nonlinear signal processing as compared to spatial structure. Similarly, the current approach neglects the receptive-field surround, which may – besides standard (linear) surround suppression – also contribute nonlinear components (Takeshita and Gollisch, 2014) or influence the nonlinearities of the receptive field center (Turner et al., 2018). An interesting extension could thus be to apply the SC model in the context of a difference-of-Gaussians receptive field (Rodieck, 1965), with local spatial contrast independently evaluated in the center and in the surround Gaussian. This may, however, require a different estimation of the receptive field, as spatiotemporal white noise often does not reveal proper receptive-field surrounds (Wienbar and Schwartz, 2018), consistent with the present data (cf. Fig. 1B).

3.2 Alternative approaches

Despite the difficulties associated with subunit models discussed above, a number of approaches have successfully tackled the challenge of identifying subunits from recordings of retinal ganglion cells. These include direct fits of cascaded LN models (Karamanlis et al., 2025; Maheswaranathan et al., 2018; Real et al., 2017), statistical inference techniques (Liu et al., 2017; Shah et al., 2020), tomographic reconstruction (Krüppel et al., 2024), and making use of anatomical data (Schwartz et al., 2012) or prior determination of photoreceptor locations (Freeman et al., 2015). Given that subunits are thought to correspond to presynaptic bipolar cells (Borghuis et al., 2001; Demb et al., 2001; Liu et al., 2017), these efforts of subunit inference also aim at providing insight into the layout of bipolar cell receptive fields and bipolar-to-ganglion cell connectivity. Yet, how much the precise knowledge of subunit number and position matters for response predictions of computational models of ganglion cells has been debated (Freedland and Rieke, 2022). Moreover, obtaining sufficient recording data for the challenging task of subunit identification may often be impractical.

Subunit models are essentially a form of multi-filter models, which apply several parallel filters before combining their signals nonlinearly. Thus, in principle, other approaches to obtaining multi-filter models could tackle the challenge of nonlinear spatial integration. Standard techniques like spike-triggered covariance analysis (Bialek and de Ruyter van Steveninck, 2005; Samengo and Gollisch, 2013; Schwartz et al., 2006) or most-informative dimensions (Sharpee et al., 2004), however, do not easily provide functional, interpretable models (Gollisch, 2013; Kaardal et al., 2013). An alternative is a likelihood-based model fit, such as in the Nonlinear Input Model (McFarland et al., 2013), which has had success in capturing excitatory and suppressive components as well as ON–OFF interactions of retinal ganglion cells (Shi et al., 2019), but which has not yet been applied to nonlinear spatial integration within the receptive field center.

Conceptually, the closest alternative approach would be to fit a generalized quadratic model (GQM; Park et al., 2013; Park and Pillow, 2011; Rajan et al., 2013) to the data. In the GQM, a neuron's activation is described by a quadratic form (general second-order polynomial function) of the stimulus components, such as here the pixel-wise intensity values over space and time. By a nonlinear transformation (typically an exponential function) and a stochastic process (typically Poisson), the activation is then turned into an output spike train, and feedback from generated spikes can be integrated as well via a filtering process that feeds into the model activation. It is the use of a quadratic form of the stimulus components, rather than a linear function as is the case in the generalized linear model (GLM; Paninski, 2004; Pillow et al., 2008; Truccolo et al., 2005) or the LN model, which renders the model sensitive to spatial structure beyond the linear activation of the receptive field and thus incorporates effects of receptive-field nonlinearities. This is equivalent to our use of the local spatial contrast as a way to capture nonlinear spatial integration. In fact, the GQM can be viewed as a more general form of the SC model. This generality, on the other

hand, makes fitting the GQM much harder as compared to the more constrained SC model, owing to the many free parameters in the quadratic form, in particular in the spatiotemporal scenario considered here. Thus, this would likely require large amounts of experimental data obtained in long recordings of spiking activity under visual stimulation as well as the exploration of appropriately strong parameter regularization.

3.3 Spatial scale of nonlinear integration

Using the SC model with a spatial stimulus smoothing prior to computing the local spatial contrast allowed us, at least for the case of white-noise stimulation, to determine the optimal spatial scale for assessing receptive-field nonlinearities in the model. Interestingly, all four analyzed cell classes yielded a similar optimal scale of around 40–50 µm. This matches typical scales of receptive fields and dendritic trees of presynaptic bipolar cells in the peripheral monkey retina. In macaque, for example, receptive field sizes of midget bipolar cells seem to lie around 30–50 µm (Dacey et al., 2000), and diffuse bipolar cells, the putative main inputs of parasol ganglion cells, have been reported with dendritic tree sizes, which may be expected to match their receptive fields, around 30–50 µm (Boycott and Wässle, 1991) and receptive fields of 70–110 µm (Dacey et al., 2000). Note, though, that these measures might differ between macaque and marmoset. For the marmoset retina, data are sparser, but anatomical measurements in the retinal periphery suggest bipolar cell dendritic tree sizes in the range of 20–80 µm (Chan et al., 2001). Despite partly comparing across different species (macaque versus marmoset) and across likely different retinal eccentricities as well as different measures (anatomical versus electrophysiological), these numbers indicate that the 40–50 µm spatial scale of nonlinear stimulus integration lies in or near the expected range, given that the spatial contrast sensitivity is thought to result from nonlinear signal transmission between bipolar and ganglion cells.

3.4 Nonlinearity of ON and OFF parasol cells

We found that ON parasol cells in our marmoset data displayed stronger spatial nonlinearities than OFF parasol cells (Fig. 1D,G and Fig. 4B), which matches our previous observations in a different study of the marmoset retina (Karamanlis et al., 2025). This contrasts findings in the macaque retina, for which rather linear spatial integration by ON parasol cells has been reported when stimulated with natural or complex scenes (Manookin et al., 2018; Turner and Rieke, 2016; Yu et al., 2022). Yet, the degree of spatial nonlinearity may not be a fixed cell-type feature, but rather be scene-dependent (Freedland and Rieke, 2022), letting linear spatial integration also fail for macaque ON parasol cells in many examples (Heitman et al., 2016), and depend on stimulus context, such as ambient light intensity (Grimes et al., 2014), surround activation (Turner et al., 2018), or adaptation state (Yu et al., 2022).

It also seems plausible that the relative strength of spatial nonlinearities between ON and OFF parasol cells is species-dependent and differs between macaque and marmoset, given their evolutionary divergence approximately 43 million years ago (Steiper and Young, 2006). Among the reported disparities are not only the well-known dichromacy in all male and some female marmosets (Jacobs, 2008) but also, for example, the density of peripheral cone photoreceptors and their convergence onto parasol cells (Goodchild et al., 1996) as well as the expression of certain transcription factors in non-parasol and non-midget ganglion cells (Nasir-Ahmad et al., 2021; Nasir-Ahmad et al., 2022). Furthermore, the ON and OFF pathways distinctly differ for macaque versus marmoset in stainings for choline acetyltransferase (ChAT), with marmoset retinas displaying considerably more cells marked in the ON ChAT band than the OFF ChAT band and vice versa in marmosets (Grünert and Martin, 2020).

3.5 Artificial versus naturalistic stimuli

It may seem surprising that the analysis under natural stimuli did not yield an optimal scale of nonlinear spatial integration. Rather, stimulus smoothing before computing the local spatial contrast generally had no or a detrimental effect. The likely reason is that natural scenes, such as those in the applied naturalistic movie, are dominated by low spatial frequencies, owing to the approximate $1/f^2$-scaling of the power spectrum of light intensities in natural images with spatial frequency f (Ruderman and Bialek, 1994; Simoncelli and Olshausen, 2001). Thus, smoothing at small spatial scales does little to the stimulus and leads to no appreciable change in the model's prediction accuracy. These results suggest that, despite the many advantages of using naturalistic stimuli to analyze neuronal coding, specific aspects of stimulus processing might still be easier to investigate with artificial stimuli that emphasize the stimulus feature of interest, such as spatial contrast at high spatial frequency.

The relative lack of high-spatial-frequency signals in the naturalistic movie as compared to the white-noise stimulus is also apparent in the reduced importance of local spatial contrast in the SC model under the naturalistic movie. Performance gains relative to the LN model and values of the parameter for weighting the local spatial contrast contribution were generally smaller than under white noise. This may simply reflect that responses are, on average, determined more by the light intensity signals inside the receptive field than by spatial contrast, so that the model naturally profits from putting more weight on the former when determining its activation.

On the other hand, it also seems feasible that the retinal neurons adapt to the different stimulus statistics under naturalistic versus white-noise stimulation (Lesica et al., 2007; Sharpee et al., 2006) and that changes in the models between those two scenarios reflect actual changes in nonlinear spatial integration, as can be observed for other alterations of stimulus context (Grimes et al., 2014; Turner et al., 2018; Yu et al., 2022). Yet, our analysis of how firing rates depend jointly on $\mathrm{I_{mean}}$ and LSC (Fig. 1) suggests that nonlinearities of spatial integration under natural stimuli are comparable to those under white-noise stimulation. Future investigations may aim at investigating this in more detail by adding appropriate probe stimuli to the artificial and naturalistic stimulation or by frequently switching between the two, as done to analyze other complex adaptation phenomena (Hosoya et al., 2005). Finally, applying different types of naturalistic stimuli, for example, containing different levels and types of motion signals, and comparing model performance and the importance of spatial-contrast information under these scenarios may help understand how general or specific the observed effects are.

4 Methods & Materials

4.1 Ethics statement

All experimental procedures were performed in strict accordance with national and institutional guidelines. Electrophysiological recordings from marmoset retinas were performed on tissue obtained from animals used by other researchers, as approved by the institutional animal care committee of the German Primate Center and by the responsible regional government office (Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, permit number 33.19-42502-04-20/3458).

4.2 Electrophysiological recordings

We used two retinas from male adult common marmosets (*Callithrix jacchus*). The eyes were enucleated and the cornea, lens, and vitreous humor were carefully removed immediately after the sacrifice of the animal. Eyecups were transported to the electrophysiology lab in a light-sealed container filled with freshly oxygenated (95% O_2 and 5% CO_2) Ames' medium (Sigma-Aldrich), supplemented with 6 mM

D-glucose and buffered with 20 mM $NaHCO_3$ to maintain a pH of 7.4. Retinas were then dark-adapted for 1–2 hours in a light-sealed container while submerged in continually replenished oxygenated Ames' medium. Care was taken to avoid oxygen depletion and to reduce the amount of light incident onto the retinal tissue to a minimum.

Each retina was first hemisected along the superior-inferior axis. For recording, a smaller piece of few millimeters in width was cut off, carefully separated from the sclera, choroid, and pigment epithelium, and gently flattened onto a transparent dialysis membrane held taut across a hollow, plastic support, with the photoreceptor side of the tissue towards the membrane. The support, along with the membrane and retinal piece, was slid into a plastic bath chamber, glued onto a planar microelectrode array (MEA; MultiChannel Systems) and designed to snugly fit the support. Thereby, the inverted support positioned the retinal tissue on top of the recording electrodes with ganglion cell-side down. MEAs had 252 electrodes with 10 µm electrode diameter and 60 µm electrode spacing and were coated prior to experiments with a thin layer of poly-D-lysine, which helped keep the retinal tissue in place. The support with the membrane was carefully removed, leaving the retina piece on the MEA, and replaced with a support without a membrane, but with a small horizontal glass plate a few millimeters above the retinal tissue. This minimized refractive effects of the stimulus light when the bath chamber was filled with perfusion solution up to the lower surface of the glass plate. The MEA was then transferred to the recording setup and connected with an amplifier and analog-to-digital converter (MultiChannel Systems). Once mounted, the retina was perfused with the oxygenated Ames' medium at a rate of around 400–450 ml/h while being held at a temperature of 32–34°C with the help of an inline heater and a heating element under the array. The remaining retina pieces were stored for later experiments in a light-sealed container at the same temperature range under constant perfusion with the Ames' medium.

The recorded extracellular signals were amplified, bandpass-filtered between 300 Hz and 5 kHz, and stored digitally at a sampling rate of 25 kHz. Spikes were extracted using a modified version of the Kilosort spike-sorting algorithm (Pachitariu et al., 2024; Pachitariu et al., 2016, modified code available at https://github.com/dimokaramanlis/KiloSortMEA). The resulting clusters were manually curated using the software package Phy2 (https://github.com/cortex-lab/phy). Only well-isolated clusters with minimal refractory period violations were chosen for further analysis, and each such cluster was deemed to correspond to a single cell. For the purpose of modeling, spike times of each cell were binned at the resolution of the projector refresh rate (85 Hz).

4.3 Visual stimulation and basic cell characterizations

Visual stimuli were rendered by a custom-made software, written in C++ and OpenGL. The stimuli were displayed on a gamma-corrected monochrome grayscale OLED monitor (eMagin) at a refresh rate of 85 Hz and a resolution of 800x600 pixels. A telecentric lens (Edmund Optics) was used to project the stimuli onto the retina, resulting in a pixel size of 7.5 µm x 7.5 µm on the retinal surface. The mean light level on the surface of the retina was either 5.5 mW/m^2 or 4.6 mW/m^2, depending on the specific recording setups. Based on photoreceptor spectral sensitivity profiles (Lamb, 1995), peak sensitivities of marmoset rods and cones (Tovée et al., 1992; Travis et al., 1988), and collecting areas of 0.37 µm^2 for cones and 1.0 µm^2 for rods, here applying values from macaques (Schnapf et al., 1990; Schneeweis and Schnapf, 1995), we estimated the isomerization rates per photoreceptor and second for the applied range of mean light intensities as 380–450 for S-cones, 2,400–2,900 for M-cones, and 7,200–8,600 for rods.

4.3.1 Spatiotemporal white-noise stimulus

We stimulated each retina with a spatiotemporal white-noise stimulus, which consisted of a sequence of frames of black and white squares (100% Michelson contrast) arranged in a uniform checkerboard pattern

of 150x200 stimulus squares, with each square spanning 30 µm x 30 µm on the surface of the retina. The contrast values for each square were independently and randomly obtained from a binary distribution and updated at the monitor refresh rate of 85 Hz. For analysis, the stimulus was denoted by the Weber contrast values $+1$ or -1 of each pixel. The stimulus was divided into trials, each consisting of a 150 s- or 300 s-long (value depending on the particular recording) distinct, non-repeating part, which was used for model training, and a 30 s- or 60 s-long fixed part ('frozen noise'), which was repeated in each trial and used for model testing. Total recording duration with this stimulus in each recording was either about 30 min or 60 min.

4.3.2 Estimation of receptive fields and stimulus filters

We estimated the receptive field of each cell using the responses to the training set of the spatiotemporal white-noise stimulus. We computed the spatiotemporal spike-triggered average (STA) from the binned responses (85 Hz) over the past 30 stimulus frames (corresponding to 353 ms). Spikes during the first 29 frames in each trial were excluded in this computation.

From the spatiotemporal STA, we extracted a spatial and a temporal component in the following way, as previously described (Karamanlis et al., 2025). For each STA, we selected pixels whose absolute peak intensity over time exceeded six times the robust standard deviation of the absolute peak intensities of all pixels around their median. (The robust standard deviation is defined as 1.4826 times the median absolute deviation and is considered a consistent estimator of the standard deviation of a normal distribution.) The temporal component was taken as the average of the time course of the selected pixels in the STA. The spatial component was computed by projecting the three-dimensional STA onto the temporal component. The receptive field center was estimated by determining the pixel in the STA with the largest variance along the temporal dimension. Lastly, the spatial component of the STA was cropped to a region of 40x40 stimulus squares (1200 µm x 1200 µm) around the receptive field center. The spatial and the temporal components were normalized by their respective Euclidean norms.

For modeling responses to the naturalistic movie stimulus, we followed the same procedure as above to obtain the STA using spatiotemporal white noise and then upsampled the spatial dimensions of the STA from 200x150 squares of the checkerboard layout to 800x600 pixels, the resolution of the movie stimulus, before extracting the spatial and temporal components. The upsampling was performed by simply using the same STA value for all pixels in the region of a given white-noise stimulus square. The resulting spatial component was again cropped around the receptive field center to a region of 160x160 pixels (1200 µm x 1200 µm). We then fitted a two-dimensional elliptical Gaussian function to each spatial component and used the fit as a spatial filter. The fitting was done using the `curve_fit` function of Python's `scipy` library. Pixels beyond a 3-σ boundary around the center of the Gaussian fit were set to zero. From the Gaussian fit, we also estimated the receptive-field size as the diameter of a circle that has the same area as the 1.5-σ ellipse of the fit.

4.3.3 Autocorrelograms

To aid cell-type classification, we computed the autocorrelogram of the spike times for each cell, based on its responses to the training segments of spatiotemporal white noise. We used a bin size of 0.5 ms, denoting the occurrence of a spike by a value of unity and no-spike by zero, and retained 150 bins, corresponding to a range of temporal lags of 0–75 ms. Each autocorrelogram was normalized to unit sum.

4.3.4 Naturalistic movie stimulus

To create a naturalistic movie stimulus, we utilized a short, publicly accessible science-fiction film released by the Blender Foundation under the CCA 3.0 license (Tears of Steel, accessible at https://mango.blender.org/). The movie was chosen for its diverse array of textures and light intensities. For experiments, the movie frames, provided in the Portable Network Graphics (PNG) format, were sequentially projected onto the retina. Excluding frames from the movie's introductory and credit sequences, as well as extended periods of homogeneous black, left us with 14,544 movie frames, equivalent to approximately 10 minutes of the movie, given the original frame rate of 24 Hz. Of these, 1,434 continuous frames were reserved as a test stimulus, with the remaining 13,110 movie frames constituting the training set. The test segment was chosen through visual inspection, focusing on scenes featuring a balanced mix of fast and slow moving objects and diverse textures. To reduce memory requirements, we cropped each movie frame from a resolution of 1920x800 pixels to 1000x800 pixels, by removing 460 pixels from each end along the horizontal axis.

The movie's frame rate of 24 Hz differed from the projector's refresh rate of 85 Hz. For presentation to the retina, we therefore upsampled the cropped frames by a factor of 3.54 by repeating individual frames three to four times, randomly selecting for each frame the presentation duration of three or four projector frames with probabilities of 0.46 and 0.54, respectively. Each cropped frame was converted from the RGB color space to grayscale using a weighted sum of R, G, and B values to compute linear luminance as $0.2126R + 0.7152G + 0.0722B$.

To generate a unique, long stimulus sequence from a short movie and to increase the diversity of stimulus patterns within the receptive fields of individual ganglion cells, we manipulated the stimulus in the following ways. First, we repeated the grayscale frames of the training set as many times as needed to yield a training set size of 255,000 stimulus frames, separated into 10 trials of 300 s each. Next, we added gaze-shift-like global motion signals to the stimulus by shifting the center of each frame independently using simulated eye movements (see below). Occasionally, due to the overlayed movement, the shifted frame exceeded the bounds of the cropped movie frame, leading to the replacement of missing pixels with gray pixels at the edges of the presentation region, but on the retina this did not overlap with the region from which ganglion cells were recorded. Lastly, to add further variability, 40% of the training segments were flipped vertically. The test stimulus sequence was treated in the same way as the training stimulus and repeated 10 times (interleaved with the training stimulus sequences), always with the same trajectory of simulated eye movements and sequence of frame repeats and not flipped.

The final stimulus was projected onto the retina at a resolution of 800x600 pixels, with each pixel covering 7.5 µm x 7.5 µm on the surface of the retina, and a mean light level of 76% of the white-noise stimulus in order to accommodate for the range of pixel intensities and the limited maximal intensity of the projector. The root–mean–square contrast of the movie over space and time was 45% of the mean intensity.

For analysis, we converted the light intensity values of each pixel in the stimulus to Weber contrast values by subtracting its mean intensity over time and subsequently normalizing by this mean intensity.

4.3.5 Simulated eye movements

We introduced global motion to the movie stimulus by spatially shifting the center position of each frame, mimicking saccades, fixations, and fixational eye movements, consistent with the statistics of active vision in marmosets (Mitchell et al., 2014). Saccades were characterized by a combination of amplitude, direction, and duration. Saccade amplitudes were randomly sampled from an exponential distribution with a scale parameter of 200 µm. The directions of saccades were randomly drawn from a uniform distribution between 0 and 360 degrees. Given that the displayed movie's frame rate was constrained by

the monitor's refresh rate, saccade duration was sampled from three possible values (23.5 ms, 35.2 ms, 47.1 ms), corresponding to 2, 3, and 4 frame updates, with probabilities of 0.35, 0.4, and 0.25, respectively. During these periods the frame's center position was linearly shifted from the original position to the target of the saccade.

Fixation durations were randomly sampled from an exponential distribution with a refractory period. This type of distribution was originally proposed as a model for fixation times in humans (Harris et al., 1988) and has since been used by others to describe the statistics of fixation duration in other animal models (Rucci et al., 2000, Angueyra et al., 2022). We here applied a scale parameter of 200 ms and a refractory period of 100 ms as used previously for primate-retina recordings (Angueyra et al., 2022). Additionally, to simulate fixational eye movements, the center of each cropped frame was jittered in the x and y directions for every frame update during periods of fixation at 85 Hz. The amplitude of the jitter was sampled independently for each direction from a normal distribution with a mean of zero pixels and a standard deviation of two pixels.

Simulated eye movement traces were generated in chunks of 10 s for both the training and the test segments. The fixation center was reset to the center of the monitor at the beginning of every chunk, preventing the fixation center from drifting out of the cropped frame. Additionally, chunks where the fixation center drifted more than 200 monitor pixels in any direction were rejected and regenerated.

4.3.6 Reversing grating stimulus and index

We also recorded responses of RGCs to a standard reversing grating, a stimulus widely used to probe spatial nonlinearities of the cells. Here, the stimulus consisted of alternating vertical black and white stripes of a fixed width (100% Michelson contrast; mean intensity same as for white-noise stimulus) that change polarity every 500 ms. We used gratings with stripe widths of 1, 2, 4, 8, 16, 32, 64, and 800 pixels and a range of spatial phase shifts, which depended on the applied stripe width. Each grating was shown for 12 temporal periods of the reversal, and the whole sequence was repeated twice. Responses were binned in 5 ms windows for further analysis.

To compute the reversing grating index for each cell, we averaged its binned responses across all 48 repeats for each combination of stripe width and phase shift to yield 1 s-long peri-stimulus time histograms (PSTHs). We computed the amplitude of the harmonics of each PSTH using a Fourier transform and extracted the amplitude F_1 at the stimulus frequency (1 Hz) and the amplitude F_2 at twice the stimulus frequency (2 Hz). We computed the largest F_2 and the largest F_1 across all widths and phases for each cell, aiming at capturing the maximum mean-light-intensity-induced and maximum spatial-contrast-induced activity modulations, and took the ratio F_2/F_1 to be the reversing grating index for that cell (Karamanlis and Gollisch, 2021).

4.4 Cell selection

Since model fits will be compromised if responses are unreliable across trials, we selected cells based on the reliability of their responses to the test stimuli. We evaluated reliability separately for the white-noise stimulus and the naturalistic movie and required cells to pass reliability criteria for both stimuli.

We used two measures of reliability. The first was the fraction of explainable variance FEV $= (\sigma^2_{total} - \sigma^2_{noise})/\sigma^2_{total}$ (Cadena et al., 2019), used as a measure of the variability in the data that cannot be attributed to noise and is thus considered a reliable signal. Here, the total variance σ^2_{total} is the variance of spike counts per time bin across all time bins and trials, and the noise variance σ^2_{noise} is the spike-count variance computed across trials for a given time bin, averaged across time bins.

The second measure was the coefficient of determination (R^2) between the binned firing-rate responses to odd versus even trials of the test set. Here, we used a symmetrized version of the coefficient of

determination (Karamanlis and Gollisch, 2021), where the average responses to either the odd or the even trials were taken as the prediction of the other and the average of both versions was used.

Based on visual inspection, we selected conservative thresholds and discarded cells with FEV less than 0.15 or with negative R^2. In total, 180 recorded cells passed our reliability criterion and were included for further analyses.

4.5 Cell-type classification

We used a semi-automated approach to identify cell types in our data set. To start, we constructed a feature vector for each reliable cell, which consisted of the 30 elements of the temporal filter, the receptive-field size, and the scores corresponding to the first ten components from a principal component analysis (PCA) of the autocorrelograms (with each entry Z-scored across cells) of all cells. After Z-scoring each entry in the feature vector independently across cells, we reduced the dimensionality of the feature vectors using PCA, retaining as many components as necessary to explain 90% of the variance (26 components). Finally, we used KMeans++ on the reduced-dimensionality feature vectors to determine 6–8 initial clusters of cells. These clusters were manually curated to ensure each cluster had similarly-sized receptive fields that tiled the retinal surface and showed consistency in the shapes of the temporal filters and autocorrelograms. We used functions from the Python library `scikit-learn` to run the PCA and KMeans++ algorithms.

We identified three clusters as corresponding to OFF midget cells (n = 51), OFF parasol cells (n = 69) and ON parasol cells (n = 35), according to receptive field size and temporal filters. For analysis, we selected three additional clusters with fairly consistent features and containing multiple cells and grouped these into a single class of Large OFF cells (n = 15). This multi-cluster class is not considered to correspond to a single cell type. Cells not included in one of these groups were discarded as they either displayed unreliable receptive field measurements or came from clusters with only few samples.

4.6 Two-dimensional firing rate histograms and LSC sensitivity

We investigated the combined influence of the mean light intensity and the high-frequency spatial contrast in a cell's receptive field on the cell's firing response using two-dimensional firing rate histograms. For each cell and each stimulus condition (white noise and naturalistic movie), we first computed two corresponding measures, the weighted mean intensity ($\mathrm{I_{mean}}$) and the local spatial contrast (LSC), for each frame of the stimulus (described below). Then, we binned both signals independently and constructed two-dimensional firing rate histograms by averaging the cell's firing response over all time points that corresponded to each pair of $\mathrm{I_{mean}}$ and LSC bins. For visualization, we used 40 equally-spaced bins for each signal and normalized the entries to the largest entry in each histogram.

To quantify how the LSC affected the firing rate in these histograms, we aimed at measuring by how much the firing rate-versus-$\mathrm{I_{mean}}$ curves in the rows of the histogram were shifted relative to each other, depending on the LSC value. For this purpose and to have a better representation of the histogram values near the edges, we recomputed the histograms by binning both $\mathrm{I_{mean}}$ and LSC independently using 60 bins, such that for each of the two signals, all the corresponding bins had about the same number of values. We excluded the first and last bins for each signal in further computation and used the remaining bins to construct two-dimensional firing rate histograms as described above. In effect, the rows of the histogram describe changes in the firing response of the cell with increasing mean light intensity at different levels of spatial contrast in the receptive field. We used the softplus function parametrized as

$$f_i(x) = a \cdot \log_e \left[1 + \exp\left(b \cdot (x - c_i)\right)\right], i = 1, ..., N \tag{1}$$

to fit a family of curves to the N rows of the histogram, with parameters a and b common across the rows and the parameters c_i depending on the rows i. These can be viewed as LSC-dependent threshold values of mean luminosity beyond which the cell's firing response rapidly increased.

We observed that the obtained c_i values depended approximately linearly on the LSC level, l_i, and used the linear relationship between the values of l_i and of c_i as a measure of the cell's sensitivity to LSC. To ensure a robust estimate of the linear relationship, we removed outlier (c_i, l_i) pairs using the Local Outlier Factor score with 5 nearest neighbours (as implemented in the Python library `scikit-learn`). Then, we used linear regression to determine the gradient m of the straight line fit to all remaining points, taking the c_i values as independent variables and the l_i values as dependent variables. Finally, we defined our metric as LSC sensitivity $= -1/m$ in order to obtain a measure where larger, positive index values correspond to more nonlinear cells.

4.7 Modeling

We developed a spatiotemporal version of the spatial contrast (SC) model (Liu et al., 2022) to predict responses of individual ganglion cells to spatiotemporal white noise and natural stimuli. The model takes spatial contrast inside a cell's receptive field into account by providing a measure of pixel-wise intensity range as an input component and thus aims at capturing nonlinear spatial integration. For comparison, we employed a classical linear-nonlinear (LN) model, which is based on linear spatiotemporal filtering and thus not sensitive to high-frequency spatial contrast inside the receptive field.

4.7.1 Linear-nonlinear model

The LN model involved a linear filtering operation of the stimulus by the cell's temporal and spatial filters and a subsequent nonlinear activation function, describing the relationship between the linear-filter output and the cell's firing response. We separated the spatiotemporal filter of a cell into a temporal and a spatial filter and obtained each, as described above, from the STA of the cell under spatiotemporal white noise. For filtering, we first convolved the stimulus z_{st} (denoting the Weber contrast of a pixel with spatial index s at time bin t) pixel-wise with the cell's temporal filter v_t of length M to obtain the convolved stimulus h_{st}:

$$h_{st} = \sum_{\tau=0}^{M-1} z_{s(t-\tau)} \cdot v_\tau. \tag{2}$$

We then applied the spatial filter u_s, taken as the Gaussian fit to the spatial component of the receptive field, by computing the scalar product of the pixel-wise elements of the stimulus and the filter to obtain the linear activation x_t of the LN model at time bin t:

$$x_t = \sum_s u_s \cdot h_{st}. \tag{3}$$

Finally, the filtered signal is transformed by the nonlinear function $f(\cdot)$ to obtain the model's expected spike-count response λ_t^{LN}:

$$\lambda_t^{LN} = f\left(x_t\right). \tag{4}$$

We here take the output nonlinearity f to be a softplus function, parameterized as

$$f(x) = a \cdot \log_e \left[1 + \exp(b \cdot x + c)\right] \tag{5}$$

The parameters a, b, and c are learned during model fitting. The details of the fitting procedure are described below.

4.7.2 Spatial contrast (SC) model

In our version of the SC model, the temporal component of the stimulus is accounted for by convolving it with the cell's temporal filter as in Eq. 2. From the convolved stimulus, we computed for every time point t the mean light intensity signal $\mathrm{I}_{\mathrm{mean}\,t}$ and the local spatial contrast LSC_t within the receptive field of the cell and combined them linearly. $\mathrm{I}_{\mathrm{mean}\,t}$ is the linearly filtered stimulus, similar to Eq. 3, defined as a weighted average of the (temporally convolved) stimulus h_{st}, with weights coming from the spatial filter u:

$$\mathrm{I}_{\mathrm{mean}\,t} = \frac{\sum_s u_s \cdot h_{st}}{\sum_s u_s}. \tag{6}$$

where r also indexes the spatial dimension.

LSC_t is analogously obtained as a weighted standard deviation of the pixel intensities from the (temporally convolved) stimulus h_{st}:

$$\mathrm{LSC}_t = \sqrt{\frac{\sum_s u_s \cdot (h_{st} - \mathrm{I}_{\mathrm{mean}\,t})^2}{\sum_s u_s}}. \tag{7}$$

Note that this definition of LSC differs from the original one suggested in Liu et al., 2022, where LSC was simply defined as a standard deviation of the spatially filtered stimulus. For large receptive fields relative to pixel size, as was the case in Liu et al., 2022, the two definitions are nearly equivalent. However, for smaller receptive field, as we encounter for the cell types of the primate retina analyzed here, Eq. 7 is superior, as it is not sensitive to variance of pixel intensities induced by the spatial filtering.

To obtain the model response from $\mathrm{I}_{\mathrm{mean}}$ and LSC, a free parameter w was used to determine the relative importance of spatial contrast information for each cell. Finally, an activation function, again taken as a parameterized softplus function, modeled the relationship between the linearly combined signal and the expected spike-count response of the cell, similar to the LN model:

$$\lambda_t^{SC} = f\left(\mathrm{I}_{\mathrm{mean}\,t} + w \cdot \mathrm{LSC}_t\right). \tag{8}$$

Together, the SC model has four free parameters – a, b, and c for the shape of the output nonlinearity (Eq. 5) and the weight factor w. Optimization of the parameters is described below.

4.7.3 Model fitting and evaluation

For both models described above, the responses of cells were binned by the update times of the stimulus (85 Hz). Model parameters were optimized by a maximum-likelihood fit under the assumption that spike counts per bin are determined according to a Poisson process with the expectation value given by the model response for that bin, λ_t^{LN} (Eq. 4) and λ_t^{SC} (Eq. 8) for the LN model and the SC model, respectively. The probability under each model of observing the complete binned spike train with spike counts n_t is given by the product of the probabilities for individual bins:

$$P(\{n_1, n_2, \ldots\}) = \prod_t \frac{\lambda_t^{n_t} e^{-\lambda_t}}{n_t!} \tag{9}$$

where λ_t is either λ_t^{LN} or λ_t^{SC}. This yields the negative log-likelihood function (up to a constant)

$$\mathcal{L}(\Theta) = -\sum_t n_t \ln \lambda_t + \sum_t \lambda_t, \tag{10}$$

where Θ represents the parameters of the model to be optimized, that is the a, b, and c from the softplus nonlinearity, and, in the case of the SC model, the weight w. The minimization of Eq. 10 was carried out for both models by employing the `minimize` function provided in Python's `scipy` library, using the `L-BFGS-B` algorithm – a quasi-Newtonian constrained optimization algorithm that supports upper and lower bounds on variables, which we found to provide reliable solutions for our problem.

Before fitting each model, we Z-scored the activation signal obtained from the spatiotemporal filtering in the LN model as well as the $\mathrm{I}_{\mathrm{mean}}$ and the LSC signals for the SC model by their respective means and standard deviations. This allowed us to apply a simple, largely common set of initial parameters for optimization of each model across all cells. We used the following initial values for the model parameters: $a = max(n_t)$, $b = 1.0$, $c = -2.0$, as well as $w = 0.0$ in the case of the SC model. Since the firing response of a cell cannot be negative, a lower bound of zero was set for the parameter a in both models by passing it as an argument to the `minimize` function.

For each cell, model predictions were evaluated against the average response across trials of the test set using the Pearson correlation coefficient r.

4.7.4 Spatial smoothing and optimal scale of spatial integration

We used a modified version of the SC model to determine the scale at which spatial nonlinearities in the ganglion cell receptive fields occur. In this version of the model, the temporally convolved stimulus (h_{st}, Eq. 2) was smoothed before extracting the LSC_t from it. The computation of $\mathrm{I}_{\mathrm{mean}\,t}$ remained unaffected. The smoothing was performed by convolving h_{st} with a circular Gaussian filter. To probe different scales of spatial smoothing, we applied different standard deviations of the Gaussian filter and defined the scale of a given spatial smoothing as three standard deviations of the filter. This definition of scale was chosen so as to be in line with our definition of receptive-field size, as the applied 1.5-σ contour of the receptive field would yield a size of three standard deviations for a circular receptive field. To perform the smoothing, we used the function `gaussian_filter` provided by the Python library `scipy`, with the diameter of the applied region of the filter taken to be six standard deviations of the filter.

For both white noise and naturalistic movie, we tested 20 different values of standard deviation of the smoothing filter, ranging from 0.8 to 12.0 pixels (in steps of 0.4 pixels from 0.8 to 6.4 and then 7.2, 8.0, 8.8, 10.4, and 12.0 pixels), corresponding to spatial scales of 18 to 270 μm. In each case, model parameters were optimized independently. The change in model performance compared to the case without stimulus smoothing was quantified for each cell as the ratio of the two corresponding Pearson's r correlation values between prediction and response. The optimal smoothing scale for a cell was determined as the spatial scale where this ratio was maximized. To obtain the optimal scale and the corresponding maximum value at finer resolution, we interpolated between the three datapoints around the largest measured ratio of correlation values using cubic spline interpolation at a resolution of 0.1 μm before identifying the maximum.

4.7.5 Power spectrum analysis

To illustrate the effect of stimulus smoothing, we examined the radially-averaged power spectral density (RAPSD) of each of the two stimuli after temporal filtering and for various smoothing scales. We analyzed a square region around the center for each stimulus, resulting in a resolution of 600x600 pixels for both the naturalistic movie and the white-noise stimulus. The white noise frames were upscaled from the 150x150 stimulus squares to 600x600 pixels without interpolation. Using a sample OFF parasol cell's temporal filter, we computed the temporally convolved stimulus to account for any effects of temporal filtering on the power spectrum. Stimulus smoothing was here applied as described earlier for the analysis of optimal spatial scales. For each resulting NxN stimulus frame, h_{mn} (where m and n index the stimulus pixels

in the two spatial dimensions and the temporal dimension is left out for convenience), we then obtained the discrete two-dimensional Fourier transform $\hat{h}_{kl}$ as:

$$\hat{h}_{kl} = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} h_{mn} \exp\{-2\pi i \frac{mk + nl}{N}\}, k = -N/2, ..., N/2; l = -N/2, ..., N/2 \tag{11}$$

where N is the number of pixels in each dimension (here 600 pixels). We then computed the power $|\hat{h}_{kl}|^2$ in each $\hat{h}_{kl}$-component. Next, we binned the two-dimensional power spectrum using radial bins of unit width, centered on the zero-frequency term, and computed the mean power in each bin, yielding the RAPSD for each frame. Finally, we averaged the RAPSDs across all frames for each individual stimulus.

Acknowledgment

We thank Fred Rieke for advice on experiments with primate retina and Fernando Rozenblit for help with the stimulus generation.

Author contributions

Conceptualization: S.S., A.E., T.G.; Methodology: S.S., M.K., D.K., M.M., T.G.; Software: S.S., M.V.; Formal analysis: S.S.; Investigation: S.S., M.V., M.K., D.K., H.M.S., V.R., S.K., S.J.Z.; Resources: A.E., T.G.; Writing - original draft: S.S., T.G.; Writing - review & editing: S.S., M.V., M.K., D.K., H.M.S., V.R., S.K., S.J.Z., M.M., A.E., T.G.; Visualization: S.S.; Supervision: T.G.; Funding acquisition: A.E., T.G.

Data availability statement

The data of this is study have been made publicly available on G-Node (https://doi.org/10.12751/g-node.3dfiti) together with information for reconstructing the visual stimuli. The code for fitting and evaluating the LN and SC models has been made available on GitHub (https://github.com/gollischlab/SpatiotemporalSCModel).

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project numbers 432680300 (SFB 1456, project B05) and 515774656 – and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 101041669). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

Aldworth, Z. N., Miller, J. P., Gedeon, T., Cummins, G. I., & Dimitrov, A. G. (2005). Dejittered spike-conditioned stimulus waveforms yield improved estimates of neuronal feature selectivity and spike-timing precision of sensory interneurons. *J Neurosci*, *25*, 5323–5332.

Angueyra, J. M., Baudin, J., Schwartz, G. W., & Rieke, F. (2022). Predicting and Manipulating Cone Responses to Naturalistic Inputs. *J Neurosci*, *42*, 1254–1274.

Bialek, W., & de Ruyter van Steveninck, R. R. (2005). Features and dimensions: Motion estimation in fly vision. *arXiv*, q–bio/0505003.

Borghuis, B. G., Marvin, J. S., Looger, L. L., & Demb, J. B. (2001). Two-photon imaging of nonlinear glutamate release dynamics at bipolar cell synapses in the mouse retina. *J Neurosci*, *33*, 10972–10985.

Boycott, B. B., & Wässle, H. (1991). Morphological classification of bipolar cells of the primate retina. *Eur J Neurosci*, *3*, 1069–1088.

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., & Ecker, A. S. (2019). Deep convolutional models improve predictions of macaque V1 responses to natural images. *PLoS Comput Biol*, *15*, e1006897.

Chan, T. L., Martin, P. R., Clunas, N., & Grünert, U. (2001). Bipolar cell diversity in the primate retina: Morphologic and immunocytochemical analysis of a new world monkey, the marmoset *Callithrix jacchus*. *J Comp Neurol*, *437*, 219–239.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Netw Comput Neural Syst*, *12*, 199–213.

Dacey, D. M., Packer, O. S., Diller, L., Brainard, D., Peterson, B., & Lee, B. (2000). Center surround receptive field structure of cone bipolar cells in primate retina. *Vis Res*, *40*, 1801–1811.

Demb, J. B., Zaghloul, K., Haarsma, L., & Sterling, P. (2001). Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (Y) ganglion cell in mammalian retina. *J Neurosci*, *21*, 7447–7454.

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *J Physiol*, *357*, 241–265.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *J Physiol*, *187*, 517–552.

Freedland, J., & Rieke, F. (2022). Systematic reduction of the dimensionality of natural scenes allows accurate predictions of retinal ganglion cell spike outputs. *Proc Natl Acad Sci USA*, *119*, e2121744119.

Freeman, J., Field, G. D., Li, P. H., Greschner, M., Gunning, D. E., Mathieson, K., Sher, A., Litke, A. M., Paninski, L., Simoncelli, E. P., & Chichilnisky, E. J. (2015). Mapping nonlinear receptive field structure in primate retina at single cone resolution. *eLife*, *4*, e05241.

Gollisch, T. (2006). Estimating receptive fields in the presence of spike-time jitter. *Netw Comput Neural Syst*, *17*, 103–129.

Gollisch, T. (2013). Features and functions of nonlinear spatial integration by retinal ganglion cells. *J Physiol Paris*, *107*, 338–348.

Goodchild, A. K., Ghosh, K. K., & Martin, P. R. (1996). Comparison of photoreceptor spatial density and ganglion cell morphology in the retina of human, macaque monkey, cat, and the marmoset callithrix jacchus. *J Comp Neurol*, *366*, 55–75.

Grimes, W. N., Schwartz, G. W., & Rieke, F. (2014). The synaptic and circuit mechanisms underlying a change in spatial encoding in the retina. *Neuron*, *82*, 460–473.

Grünert, U., & Martin, P. R. (2020). Cell types and cell circuits in human and non-human primate retina. *Prog Retin Eye Res*, *78*, 100844.

Harris, C. M., Hainline, L., Abramov, I., Lemerise, E., & Camenzuli, C. (1988). The distribution of fixation durations in infants and naive adults. *Vis Res*, *28*, 419–432.

Hartline, H. K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am J Physiol*, *121*, 400–415.

Heitman, A., Brackbill, N., Greschner, M., Sher, A., Litke, A. M., & Chichilnisky, E. J. (2016). Testing pseudo-linear models of responses to natural scenes in primate retina. *bioRxiv*, 045336v2.

Hosoya, T., Baccus, S. A., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, *436*, 71–77.

Jacobs, G. H. (2008). Primate color vision: A comparative perspective. *Vis Neurosci*, *25*, 619–633.

Kaardal, J., Fitzgerald, J. D., Berry II, M. J., & Sharpee, T. O. (2013). Identifying functional bases for multidimensional neural computations. *Neural Comput*, *25*, 1870–1890.

Karamanlis, D., & Gollisch, T. (2021). Nonlinear spatial integration underlies the diversity of retinal ganglion cell responses to natural images. *J Neurosci*, *41*, 3479–3498.

Karamanlis, D., Khani, M. H., Schreyer, H. M., Zapp, S. J., Mietsch, M., & Gollisch, T. (2025). Nonlinear receptive fields evoke redundant retinal coding of natural scenes. *Nature*, *637*, 394–401.

Karamanlis, D., Schreyer, H. M., & Gollisch, T. (2022). Retinal encoding of natural scenes. *Annu Rev Vis Sci*, *8*, 171–193.

Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron*, *30*, 803–817.

Krüppel, S., Khani, M. H., Schreyer, H. M., Sridhar, S., Ramakrishna, V., Zapp, S. J., Mietsch, M., Karamanlis, D., & Gollisch, T. (2024). Applying super-resolution and tomography concepts to identify receptive field subunits in the retina. *PLoS Comput Biol*, *20*, e1012370.

Lamb, T. D. (1995). Photoreceptor spectral sensitivities: Common shape in the long-wavelength region. *Vis Res*, *35*, 3083–3091.

Lesica, N. A., Jin, J., Weng, C., Yeh, C.-I., Butts, D. A., Stanley, G. B., & Alonso, J.-M. (2007). Adaptation to stimulus contrast and correlations during natural visual stimulation. *Neuron*, *55*, 479–491.

Liu, J. K., Karamanlis, D., & Gollisch, T. (2022). Simple model for encoding natural images by retinal ganglion cells with nonlinear spatial integration. *PLoS Comput Biol*, *18*, e1009925.

Liu, J. K., Schreyer, H. M., Onken, A., Rozenblit, F., Khani, M. H., Krishnamoorthy, V., Panzeri, S., & Gollisch, T. (2017). Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nat Commun*, *8*, 149.

Maheswaranathan, N., Kastner, D. B., Baccus, S. A., & Ganguli, S. (2018). Inferring hidden structure in multilayered neural circuits. *PLoS Comput Biol*, *14*, e1006291.

Manookin, M. B., Patterson, S. S., & Linehan, C. M. (2018). Neural mechanisms mediating motion sensitivity in parasol ganglion cells of the primate retina. *Neuron*, *97*, 1327–1340.

Masri, R. A., Percival, K. A., Koizumi, A., Martin, P. R., & Grünert, U. (2019). Survey of retinal ganglion cell morphology in marmoset. *J Comp Neurol*, *527*, 236–258.

McFarland, J. M., Cui, Y., & Butts, D. A. (2013). Inferring nonlinear neuronal computation based on physiologically plausible inputs. *PLoS Comput Biol*, *9*, e1003143.

Mitchell, J. F., Reynolds, J. H., & Miller, C. T. (2014). Active vision in marmosets: A model system for visual neuroscience. *J Neurosci*, *34*, 1183–1194.

Nasir-Ahmad, S., Lee, S. C. S., Martin, P. R., & Grünert, U. (2021). Identification of retinal ganglion cell types expressing the transcription factor Satb2 in three primate species. *J Comp Neurol*, *529*, 2727–2749.

Nasir-Ahmad, S., Vanstone, K. A., Novelli, M., Lee, S. C. S., Do, M. T. H., Martin, P. R., & Grünert, U. (2022). Satb1 expression in retinal ganglion cells of marmosets, macaques, and humans. *J Comp Neurol*, *530*, 923–940.

Ozuysal, Y., Kastner, D. B., & Baccus, S. A. (2018). Adaptive feature detection from differential processing in parallel retinal pathways. *PLoS Comput Biol*, *14*, e1006560.

Pachitariu, M., Sridhar, S., Pennington, J., & Stringer, C. (2024). Spike sorting with Kilosort4. *Nat Methods*, *21*, 914–921.

Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M., & Harris, K. D. (2016). Fast and accurate spike sorting of high-channel count probes with KiloSort. *Adv Neural Inf Process Syst*, *29*.

Paninski, L. (2003). Convergence properties of three spike-triggered analysis techniques. *Netw Comput Neural Syst*, *14*, 437–464.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Netw Comput Neural Syst*, *15*, 243–262.

Park, I. M., Archer, E., Priebe, N., & Pillow, J. W. (2013). Spectral methods for neural characterization using generalized quadratic models. *Adv Neural Inf Process Syst*, *26*, 2454–2462.

Park, I. M., & Pillow, J. W. (2011). Bayesian spike-triggered covariance. *Adv Neural Inf Process Syst*, *24*, 1692–1700.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci*, *25*, 11003–11013.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*, 995–999.

Rajan, K., Marre, O., & Tkacik, G. (2013). Learning quadratic receptive fields from neural responses to natural stimuli. *Neural Comput*, *25*, 1661–1692.

Real, E., Asari, H., Gollisch, T., & Meister, M. (2017). Neural circuit inference from function to structure. *Curr Biol*, *27*, 189–198.

Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vis Res*, *5*, 583–601.

Rucci, M., Edelman, G. M., & Wray, J. (2000). Modeling LGN responses during free-viewing: A possible role of microscopic eye movements in the refinement of cortical orientation selectivity. *J Neurosci*, *20*, 4708–4720.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Phys Rev Lett*, *73*, 814–817.

Samengo, I., & Gollisch, T. (2013). Spike-triggered covariance: Geometric proof, symmetry properties, and extension beyond gaussian stimuli. *J Comput Neurosci*, *34*, 137–161.

Schnapf, J. L., Nunn, B. J., Meister, M., & Baylor, D. A. (1990). Visual transduction in cones of the monkey macaca fascicularis. *J Physiol*, *427*, 681–713.

Schneeweis, D. M., & Schnapf, J. L. (1995). Photovoltage of rods and cones in the macaque retina. *Science*, *268*, 1053–1056.

Schwartz, G. W., & Rieke, F. (2011). Nonlinear spatial encoding by retinal ganglion cells: When $1+1 \neq 2$. *J Gen Physiol*, *138*, 283–290.

Schwartz, G. W., Okawa, H., Dunn, F. A., Morgan, J. L., Kerschensteiner, D., Wong, R. O., & Rieke, F. (2012). The spatial structure of a nonlinear receptive field. *Nat Neurosci*, *15*, 1572–1580.

Schwartz, O., Pillow, J. W., Rust, N. C., & Simoncelli, E. P. (2006). Spike-triggered neural characterization. *J Vis*, *6*, 484–507.

Shah, N. P., Brackbill, N., Rhoades, C., Kling, A., Goetz, G., Litke, A. M., Sher, A., Simoncelli, E., & Chichilnisky, E. J. (2020). Inference of nonlinear receptive field subunits with spike-triggered clustering. *eLife*, *9*, e45743.

Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, *439*, 936–942.

Sharpee, T. O., Rust, N. C., & Bialek, W. (2004). Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Comput*, *16*, 223–250.

Shi, Q., Gupta, P., Boukhvalova, A. K., Singer, J. H., & Butts, D. A. (2019). Functional characterization of retinal ganglion cells using tailored nonlinear modeling. *Sci Rep*, *9*, 8713.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural Image Statistics and Neural Representation. *Annu Rev Neurosci*, *24*, 1193–1216.

Steiper, M. E., & Young, N. M. (2006). Primate molecular divergence dates. *Mol Phylogenetics Evol*, *41*, 384–394.

Takeshita, D., & Gollisch, T. (2014). Nonlinear spatial integration in the receptive field surround of retinal ganglion cells. *J Neurosci*, *34*, 7548–7561.

Tovée, M. J., Bowmaker, J. K., & Mollon, J. D. (1992). The relationship between cone pigments and behavioural sensitivity in a new world monkey (callithrix jacchus jacchus). *Vis Res*, *32*, 867–878.

Travis, D. S., Bowmaker, J. K., & Mollon, J. D. (1988). Polymorphism of visual pigments in a callitrichid monkey. *Vis Res*, *28*, 481–490.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol*, *93*, 1074–1089.

Turner, M. H., Sanchez Giraldo, L. G., Schwartz, O., & Rieke, F. (2019). Stimulus- and goal-oriented frameworks for understanding natural vision. *Nat Neurosci*, *22*, 15–24.

Turner, M. H., & Rieke, F. (2016). Synaptic rectification controls nonlinear spatial integration of natural visual inputs. *Neuron*, *90*, 1257–1271.

Turner, M. H., Schwartz, G. W., & Rieke, F. (2018). Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. *eLife*, *7*, e38841.

Victor, J. D., & Shapley, R. M. (1979). The nonlinear pathway of Y ganglion cells in the cat retina. *J Gen Physiol*, *74*, 671–689.

Weltzien, F., Percival, K. A., Martin, P. R., & Grünert, U. (2015). Analysis of bipolar and amacrine populations in marmoset retina. *J Comp Neurol*, *523*, 313–334.

Wienbar, S., & Schwartz, G. W. (2018). The dynamic receptive fields of retinal ganglion cells. *Prog Retin Eye Res*, *57*, 102–117.

Yu, Z., Turner, M. H., Baudin, J., & Rieke, F. (2022). Adaptation in cone photoreceptors contributes to an unexpected insensitivity of primate on parasol retinal ganglion cells to spatial structure in natural images. *eLife*, *11*, e70611.

Zapp, S. J., Nitsche, S., & Gollisch, T. (2022). Retinal receptive-field substructure: Scaffolding for coding and computation. *Trends Neurosci*, *45*, 430–445.