# Filter-based models of suppression in retinal ganglion cells: comparison and generalization across species and stimuli

Neda Shahidi[1-4], Fernando Rozenblit[1,2], Mohammad H. Khani[1,2,§], Helene M. Schreyer[1,2,§], Matthias Mietsch[5,6], Dario A. Protti[7], Tim Gollisch[1,2,8,9,*]

*1. Department of Ophthalmology, University Medical Center Göttingen, Göttingen, Germany*

*2. Bernstein Center for Computational Neuroscience Göttingen, Göttingen, Germany*

*3. Georg-Elias-Müller-Institute for Psychology, Georg-August-Universität Göttingen, Germany*

*4. Cognitive Neuroscience Lab, German Primate Center, Göttingen, Germany*

*5. Laboratory Animal Science Unit, German Primate Center, Göttingen, Germany*

*6. German Center for Cardiovascular Research, Partner Site Göttingen, Göttingen, Germany*

*7. School of Medical Sciences (Neuroscience), The University of Sydney, Sydney, NSW, Australia*

*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*

** tim.gollisch@med.uni-goettingen.de*

§ *current address: Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland*

## Abstract

The dichotomy of excitation and suppression is one of the canonical mechanisms explaining the complexity of neural activity. Computational models of the interplay of excitation and suppression in single neurons aim at investigating how this interaction affects a neuron's spiking responses and shapes the encoding of sensory stimuli. Here, we compare the performance of three filter-based stimulus-encoding models for predicting retinal ganglion cell responses recorded from axolotl, mouse, and marmoset retina to different types of temporally varying visual stimuli. Suppression in these models is implemented via subtractive or divisive interactions of stimulus filters or by a response-driven feedback module. For the majority of ganglion cells, the subtractive and divisive models perform similarly and outperform the feedback model as well as a linear-nonlinear (LN) model with no suppression. Comparison between the subtractive and the divisive model depended on cell type, species, and stimulus components, with the divisive model generalizing best across temporal stimulus frequencies and visual contrast and the subtractive model capturing in particular responses for slow temporal stimulus dynamics and for slow axolotl cells. Overall, we conclude that the divisive and subtractive models are well suited for capturing interactions of excitation and suppression in ganglion cells and perform best for different temporal regimes of these interactions.

# Introduction

Sensory systems dynamically encode stimuli with a wide range of temporal characteristics, reflecting the complexity of changing scenes and self-motion in the natural world. A crucial mechanism for shaping and supporting the encoding of dynamic stimuli is thought to lie in the interplay of excitation and suppression, with the latter encompassing inhibitory synaptic signals as well as cell-intrinsic negative feedback, such as mediated by voltage-dependent conductances. A simple example of this interplay of excitation and suppression is the shortening of response duration, leading to markedly transient responses of cells in the retina[1], the thalamic lateral-geniculate nucleus[2], as well as in auditory[3] and somatosensory[4,5] systems in response to an abrupt change in the strength of a stimulus. In many cases, this response transiency has been explained by suppressive inputs coming with a short delay relative to the excitatory input, resulting in a brief window of opportunity for elevated activity before suppression dominates.

Similarly, the context dependence of stimulus encoding in the visual system, such as found by varying background illumination, contrast, stimulus size, or stimulus motion, has been captured by suppressive interactions that can act as a divisive gain modification[6–8]. Suppressive mechanisms may also allow for creating dynamic variations of stimulus integration time, as demonstrated by a conductance-based encoding model of suppression[9,10]. This provides a long temporal window for integrating sensory inputs when the signal-to-noise ratio is low, allowing a neuron to average out noise and generate a robust response. A short integration window at higher signal-to-noise ratio, on the other hand, enables the neuron to lock its response to the fast fluctuation of the sensory input.

To understand the role of the excitation–suppression interplay in generating various response dynamics, several filter-based encoding models have been proposed. The canonical idea behind these groups of models is to capture properties of excitatory and suppressive signals in independent temporal filters, whose signals are then combined to yield the activity of the modeled neuron. Here we focus on three classes of these models that have been proposed to capture the temporal response properties of retinal ganglion cell responses, using subtractive inhibition[11], divisive suppression[1], and activity-dependent negative feedback[10,12,13], respectively. For the subtractive and divisive models, the suppression is derived in a feedforward way by applying a filter to the same input signal that feeds the excitation. For the feedback model, on the other hand, suppression is triggered by the spikes that are stochastically generated by the model cell and that provide the input to the suppressive filter.

To maximize the similarity between the training algorithms and optimization constraints across the different models, we trained and tested the three suppressive models as well as a feedforward linear-nonlinear (LN) model within a unified computational structure, where each of the models could be implemented by fixing certain model parameters to specific values. We applied this framework to ganglion cell responses recorded from isolated retinas of axolotl, mouse, and common marmoset under stimulation with full-field white-noise flicker of light intensity. The unified implementation and datasets allowed us to compare the models directly and to identify the cell and stimulus types for which each model outperforms the others.

One of the main challenges for developing encoding models is their generalization to different classes of stimuli. While we train the models on data from white-noise stimuli, as is customarily done for these and comparable models, we here also test the performance of the suppressive models on stimuli that contain sinusoidal frequency and contrast sweeps. The ability of a model to generalize to such non-stationary stimuli helps provide an understanding of how the model captures retinal stimulus processing in a complex setting with varying stimulus statistics.

# Results

### Comparison of three suppression models in a unified framework

We implemented three filter-based neuron models of suppression, based on subtractive, divisive, and feedback interactions, respectively, as well as an LN model in a unified structure (Fig. S1; see Methods). The common structure aimed at maximizing similarities in the models' motifs as well as in training and testing procedures. In this framework, each model yields an integrated signal that encompasses the effects of excitation and suppression and that—after nonlinear rectification—is used for the stochastic generation of time-varying spike counts in small discrete time windows via a Poisson process. What differs between the models is how the integrated signal is obtained (Fig. 1A). The subtractive and divisive models both had two upstream branches, each with its own stimulus filter and subsequent nonlinear transformation. The two branches were then combined by a subtractive or divisive operation, respectively. The upstream filters were unconstrained, whereas the shapes of the nonlinear transformations were restricted to support the excitatory or suppressive action: The nonlinearities for the excitatory and suppressive branches of the subtractive model as well as for the excitatory branch of the divisive model were required to be monotonically increasing, whereas the suppressive branch of the divisive model had a nonlinearity that was "bump-shaped", i.e., monotonically increasing on the side of negative inputs and monotonically decreasing on the positive side. The feedback

model and the LN model each had a single (excitatory) upstream branch for stimulus filtering and subsequent nonlinear transformation of the filtered signal. For the feedback model, an additional filter was applied to the model-generated spikes, and the upstream and feedback filter signals were then summed.

The final output nonlinearity just prior to the stochastic spike generation is a common element in all models. For the LN model, this means two successive nonlinear stages that together constitute the typical single output nonlinearity of this model. That this is here subdivided into two stages is a concession to the unified structure of all models. We checked that the two-stage nonlinearity did not impair model performance by comparing this LN model to a more conventional LN model where the filter was obtained either as the spike-triggered average or as the first stimulus feature from a spike-triggered-covariance analysis (selecting from these two options the one that performed better). The nonlinearity was here obtained as a histogram of the measured spike probability depending on the filtered stimulus signal. We found that the LN model that was fitted via the unified structure of Fig. S1, including the two-stage nonlinearity, slightly outperformed the classical LN model implementation (Fig. S2), presumably because the regularization included in the model fit supports model performance for held-out data. For all further comparison, we there used the fitted LN model, according to the structure of Fig. S1.

To limit the number of free parameters, the models, the filters, and the upstream nonlinearities were parametrized using sets of basis-functions, and the output rectifier was a parametrized softplus function (see Methods). All models were trained using a maximum-likelihood approach with a block-coordinate ascent algorithm, in which we iterated the sequential update of each block of the model (a filter or a nonlinearity) to maximize the likelihood function while keeping other blocks constant.

The models were trained and tested on ganglion cell spiking activity in retinas of axolotls (n = 15 recordings), mice (n = 8), and marmosets (n = 10). Spikes were recorded with multi-electrode arrays while the retinas were stimulated with full-field illumination. For model fitting, we applied light intensity that varied according to Gaussian white noise. For axolotl and mouse, the stimulus consisted of an alternating sequence of non-repeating white noise, which was used for training the models, and “frozen noise”, i.e., a fixed white-noise segment that was presented repeatedly, which we held out for testing (see Methods). In the marmoset data, no frozen-noise stimulus was used. Therefore, we segmented the data into pseudo-trials of 33.3 s duration and used one fifth of each trial as held-out test data and trained the models on the rest. The model

performance was assessed by calculating the average information per spike provided by the model about the cell's spiking response[2,14], assessed on the held-out test data segments. Note that this measure of model performance does not require repeated trials.

Our analysis included 607 axolotl, 335 mouse, and 370 marmoset cells that responded reliably to the white-noise stimulation. ON-OFF cells, i.e., cells that responded to both step-wise increases and decreases in light intensity, were removed (see Methods) because their responses cannot be captured by the monotonically increasing excitatory nonlinearities, as applied in all four models.

Comparing the performance of the three suppressive models to that of the LN model (Fig. 1B) showed that each of the suppressive models predicted responses on held-out data better than the LN model for the majority of cells (subtractive: 97% of the cells, divisive: 96%, feedback: 85%). To quantify for each cell and each suppressive model how much predictive power was gained by the addition of the considered suppressive component, we computed the excess performance as the difference (in bits/spike) between the model performance of the considered suppressive model and that of the LN model (Fig. 1C). We found that, for 95% of the cells with the highest excess performance within each species, the excess performance of the best-performing suppressive model was in the range of [0.03, 0.94] bits/spike for axolotl, [0.01, 0.29] for mouse, and [0.001, 0.23] for marmoset. The best-performing model was the subtractive model for 46%, the divisive model for 42%, and the feedback model for 12% of the cells.

For recordings from axolotl and mouse retina, visual stimulation included repeatedly inserted sections of the same (randomly generated) stimulus sequence ("frozen noise"). These sections, which were not used for parameter fitting, allow measuring a firing-rate profile. Comparison with model-predicted firing rates then provided a measure of explained variance (Fig. S3). This yielded qualitatively similar findings as the evaluation via per-spike information (cf. Fig. 1B,C). In particular, the explained variance was generally higher for all suppressive models than for the LN model ($p \ll 0.001$). Moreover, the excess performance, as obtained here by subtracting the explained variance of the LN model from that of the considered suppressive model, was generally larger for the subtractive and the divisive models as compared to the feedback model, corroborating that the subtractive and divisive models on average outperformed the feedback model.

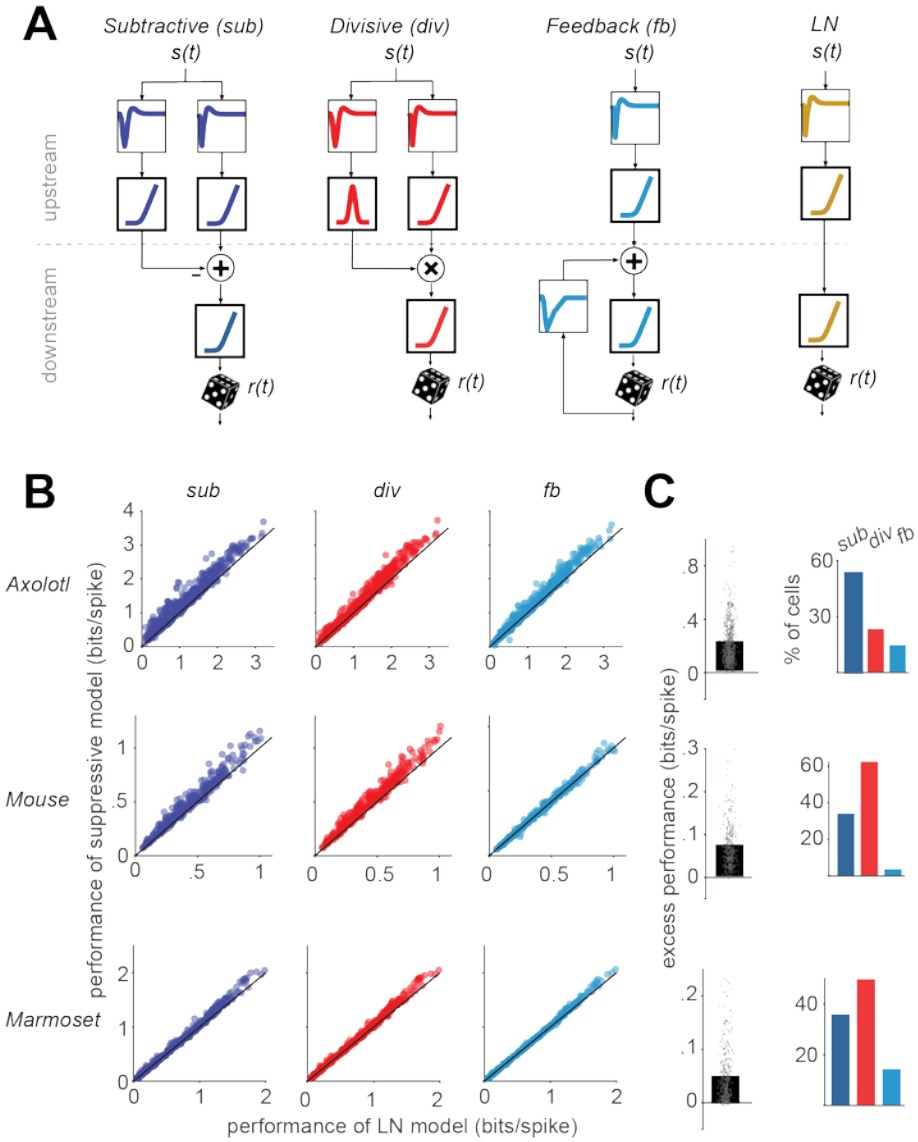


**Figure 1.** Schematics and overall performance of the models. **A)** Structure of the three suppressive models used in our study with subtractive, divisive, and feedback suppression and of the LN model. The filter components are shown in thin frames and the nonlinearities in thick frames. **B)** Comparison of the performance of each model, measured as the information provided by each spike, to the performance of the LN model. Each data point represents one cell. **C)** *Left*: The excess performance (performance of suppressive model minus LN-model performance) of the best-performing suppressive model, determined for each cell. *Right*: the percentage of cells for which each of the suppressive models outperformed all the other models.

### Cell-type specific comparison of subtractive and divisive model

As the subtractive and divisive models generally had the top performance among the four assessed models, we compared these two models more closely and aimed at determining for which cells one or the other of these two models was superior. Comparing the performances on a cell-by-cell basis (Fig. 2A) indicated that, in the axolotl retina, there is a group of cells for which the subtractive model outperformed the divisive model. For the rest of the axolotl cells as well as for mouse and marmoset cells, the performances of the two models were comparable.

For mouse and marmoset, the divisive model performed overall slightly better than the subtractive cells ($p < 0.003$ in both cases; here and in later comparisons between models: Wilcoxon signed rank test with false discovery rate for multiple comparison correction[15]; Fig. 2A).

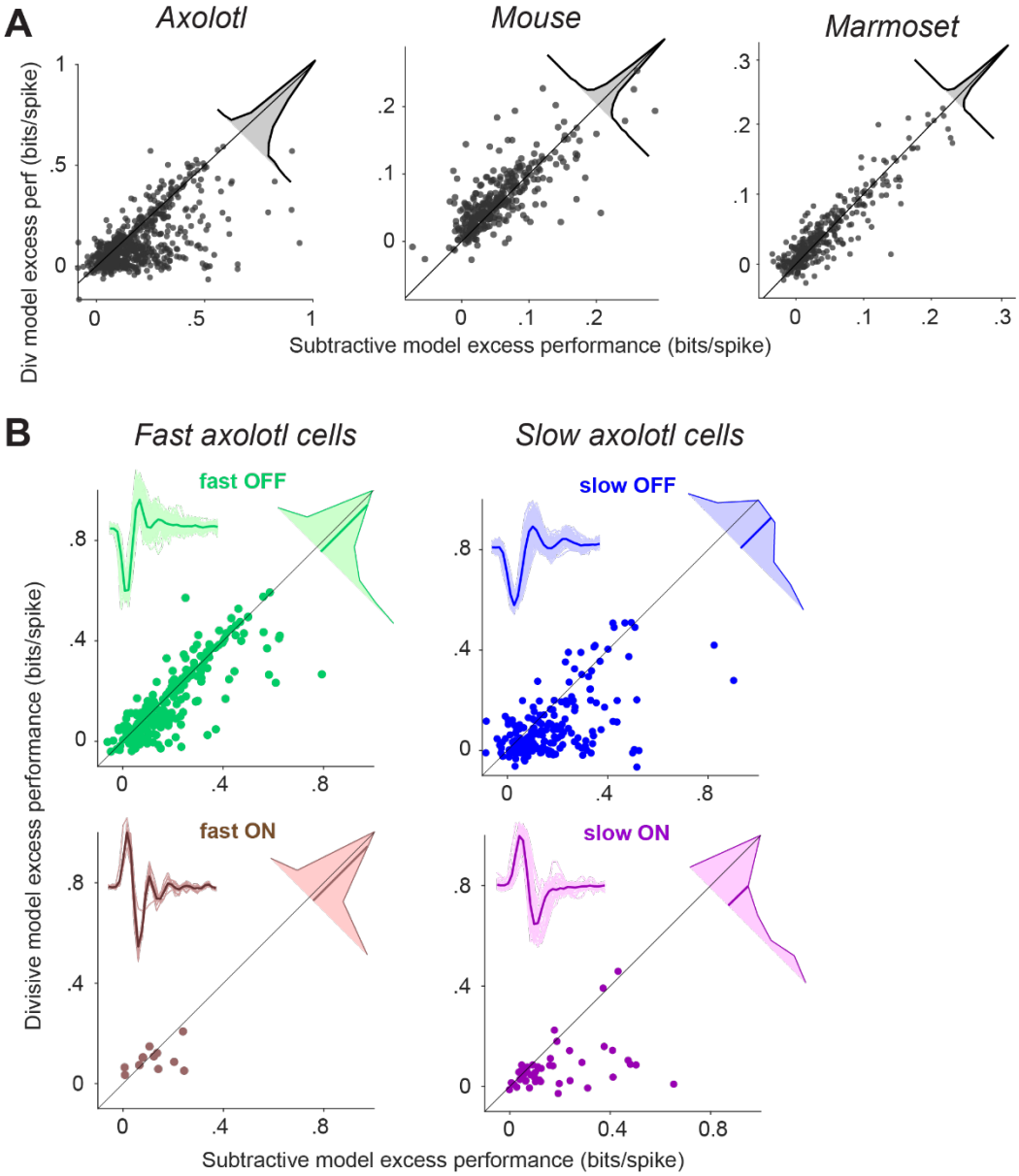


**Figure 2.** Comparison of the excess performance of the subtractive and the divisive model. **A)** Scatter plot of the excess performance of the subtractive model vs. the divisive model for each species. Each data point is a cell. Inset: The distribution of the performance difference between the subtractive and the divisive model. **B)** Same as A, but for slow/fast and ON/OFF subgroups of the cells recorded from axolotl retina. Insets: The filters of the LN model of all cells within the subgroup (thin lines) and their average (thick line).

For the axolotl retina, we aimed at understanding the diversity in model performance by classifying the cells into ON- and OFF-type classes as well as into slow and fast. To do so, we computed the temporal filter for each cell as the spike-triggered average under the temporal

white-noise stimulation and clustered cells in the space of the first two principal components of all filters into four clusters (Fig. S4), which were characterized according to the sign of the primary peak of the filter as ON- or OFF-type and according to the peak latency as fast or slow. Based on this classification, we found that the relative performance of the subtractive and the divisive model depended on the kinetics of the considered cell (Fig. 2B). For fast cells, the difference between the performance of the subtractive and the divisive model was small, with the subtractive model just slightly better for fast OFF cells (mean difference = 0.026 bits/spike, $p = 0.005$) and no significant difference for the small number of identified fast ON cells (mean difference = 0.02 bits/spike, $p = 0.7$). For slow ON and slow OFF cells, on the other hand, the subtractive model performed substantially better than the divisive model (slow OFF: mean difference = 0.08 bits/spike, $p \ll 0.0001$, slow ON: mean difference = 0.1 bits/spike, $p \ll 0.0001$).

To better understand how the suppressive components of the subtractive and the divisive model help improve over the LN model, we examined the model components for individual sample cells more closely. For the sample axolotl slow OFF cell in Fig. 3A, the subtractive and the divisive model yielded qualitatively similar stimulus filters for both the excitatory and the suppressive model branches. However, the upstream suppressive functions of the subtractive and divisive models were substantially different for this cell. In the subtractive model, the monotonic shape of the nonlinearity in the suppression branch allowed for a gradual decline of suppression as the activation of the suppressive filter becomes more negative (Fig. 3Ai). This let the suppression contribute a gradual and nuanced component to the model activation, which supported the accurate prediction of spikes (Fig. 3Aii). In the divisive model, on the other hand, the shape of the nonlinearity in the suppressive branch led to a unilateral, rectified suppression with essentially no sensitivity to filter activations near to or smaller than zero (Fig. 3Ai). This follows from the constraint of the nonlinearity to be bump-like so that suppression cannot decrease as the filter activation deviates from zero. Therefore, the divisive model was not as sensitive as the subtractive model to fine fluctuations in the stimulus and often generated stretches of near-flat response predictions, which did not capture the response modulation for this cell (Fig. 3Aii).

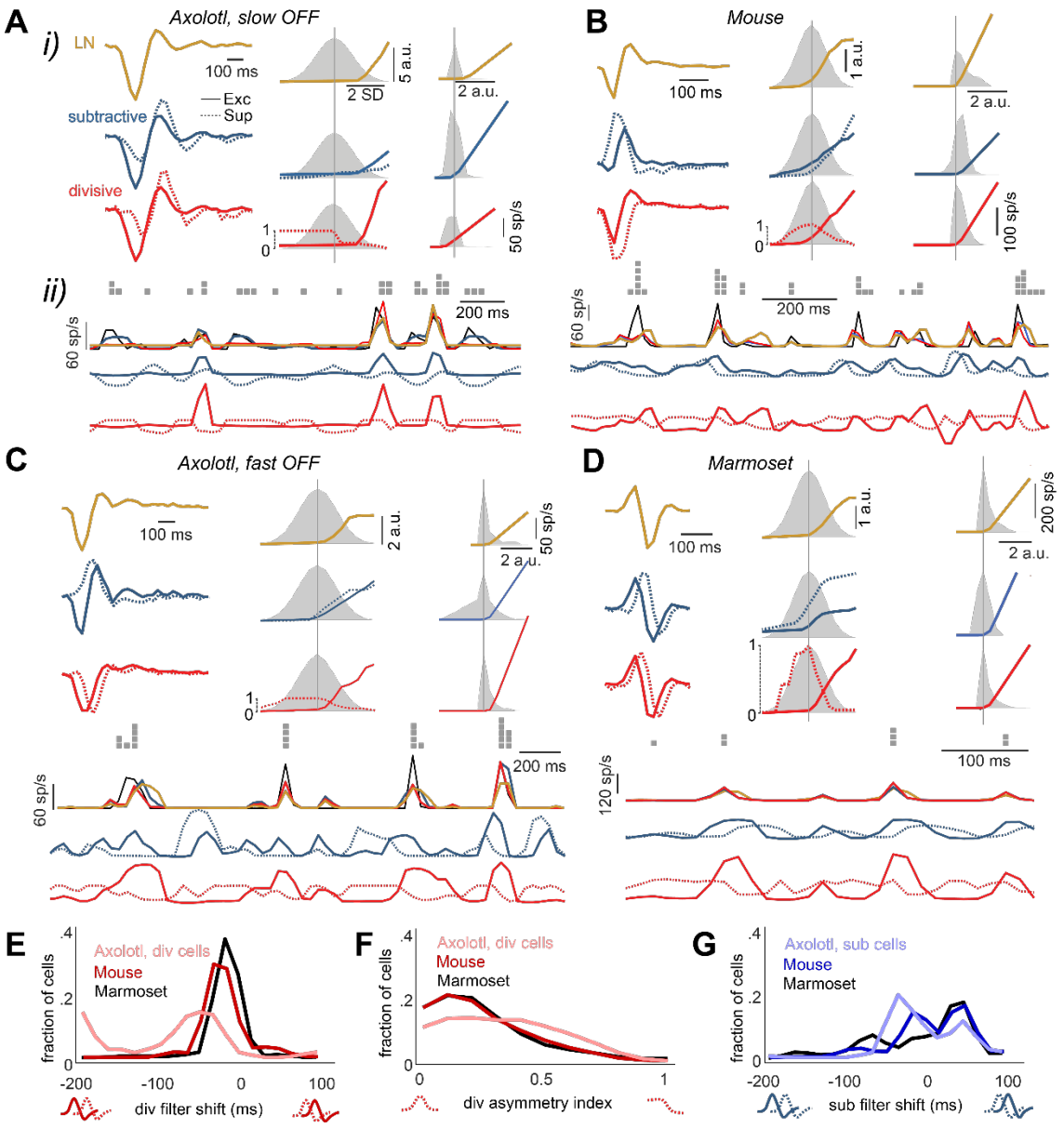


**Figure 3.** Model fits and response predictions for sample cells from each investigated species. **Ai)** Excitatory and suppressive filters *(left)*, upstream nonlinear functions *(middle)*, together with the distribution of the corresponding input (gray), and output rectifier *(right)*, together with the distribution of the corresponding input (gray). The feedback model was omitted, due to its lower performance, but the LN model was included for comparison. **Aii)** For a subsection of the test-trial period: *1st row:* spike counts per stimulus time bin for a sample test trial. *2nd row:* the cell's firing rate (black) and the predictions of the three models. *Below:* the output of the excitatory (solid line) and suppressive (dashed line) upstream branches for the subtractive (*3rd row*) and the divisive model (*4th row*). **B, C, D)** Same as A for other sample cells. Note that D has the single-trial spike train above the prediction curves as data, but no measured firing rate because the marmoset recordings did not contain frozen-noise sections. **E)** The shift in the suppressive filter of the divisive model, relative to the excitatory filter, across all cells. The shift was defined as the time lag for which the cross-correlogram between the excitatory and suppressive filters obtained its maximum. **F)** The asymmetry index for the suppressive nonlinear function of the divisive model. The index was calculated as the absolute difference between the sum of the function values over negative and positive inputs, divided by the total sum of the function values. **G)** The shift in the suppressive filter of the subtractive model, relative to the excitatory filter across all cells, calculated in a similar fashion as E.

By contrast, for the mouse cell shown in Fig. 3B, the axolotl fast OFF cell in Fig. 3C, and the marmoset cell in Fig. 3D, the bump-like shape of the divisive suppression nonlinearity seems more appropriate. Here, the suppressive filter for the divisive model resembled the excitatory filter, except for a small shift to the right. This resulted in a suppressive response that came with a slight delay compared to the excitatory response, leading to a timelier prediction of the falling edge of a firing-rate response peak compared to the LN model. Often, the suppressive filters lagged the associated excitatory filters by 1-2 stimulus frames (~30-60 ms for axolotl; ~16-33 ms for mouse and marmoset). This was the case for 57% of the axolotl cells among those for which the divisive model outperformed the others, 79% of mouse cells, and 70% of marmoset cells, as assessed by the peak location in the cross-correlogram between the two filters (Fig. 3E). Also, for 57% of the axolotl cells with a superior divisive model, 90% of the mouse cells, and 88% of the marmoset cells, the divisive nonlinearity was fairly symmetric (Fig. 3F; asymmetry index < 0.5), as measured by the absolute difference in average values of the nonlinearity on the left versus right side of zero (normalized by the corresponding sum). Thus, for these cells, both positive and negative activations of the suppressive filter led to suppression in the model, in contrast to the example of the slow axolotl cell in Fig. 3A.

The subtractive model for the three sample cells shown in Fig. 3B-C, on the other hand, yielded a more diverse set of suppressive filters, relative to their excitatory counterparts. The suppressive filter could be temporally trailing the excitatory filter (e.g., Fig. 3D), preceding it (Fig. 3B), or have an altogether different shape or relative strength in the positive and negative components (Fig. 3C). A suppressive filter lagging the excitatory one by 1-2 stimulus frames was found for 40% of the axolotl cells for which the subtractive model outperformed the others, for 12% of mouse cells, and for 5% of marmoset cells. By contrast, 32% of those axolotl cells, 23% of the mouse cells, and 28% of the marmoset cells had suppressive filters leading the excitatory filter by 1-3 stimulus frames (Fig. 3G). Yet, model performance was generally similar to the divisive model, suggesting that different computational motives might converge to inherently different, but equally good solutions.

**Differences in model performance between sustained and transient cells**

Delayed suppression that comes with a similar stimulus dependence as excitation is a proposed mechanism for generating sharp transient responses[1,2]. The delayed suppression may result, for example, from feedforward inhibition with additional synaptic transmission delays as compared to excitation, which generates a brief window for excitation to elicit spikes before the suppressive component terminates the response.

We thus hypothesized that cells with transient response characteristics have particularly strong suppression and that adding a suppressive component in the model is particularly important for these cells. For a subset of marmoset cells, which had also been stimulated using full-field steps in light intensity (100% contrast), we therefore determined the transiency of their responses to light-intensity steps and classified them as either transient or sustained by comparing the strength of the initial response peak to the sustained activity beyond 300 ms after stimulus onset (Fig. 4A). We then checked for each class of cells how important the addition of a suppressive component in the model was by collecting the excess performance values. As the model performances displayed considerable variability across experiments, most likely at least partly due to different sampling of ganglion cell types and retinal eccentricity, we made this comparison on an experiment-by-experiment basis. Furthermore, we focused on ON-type cells, as some of our recordings did not yield a sufficient number of OFF-type cells for separation into transient and sustained classes.

For the sample recording shown in Fig. 4B, the excess performance of suppression (from the best-performing suppressive model for each cell) was significantly larger for the transient-ON cells compared to the sustained-ON cells, suggesting weaker suppression in the sustained-ON cells (mean difference: 0.032 bits/spike; $p = 0.0083$). Four out of six recordings showed a similar difference between the transient and sustained-ON cells, whereas there was no substantial difference for two recordings (Fig. 4C). This indicates that, indeed, explicit suppressive model components can be particularly helpful for transient cells in explaining response characteristics.

**Generalization to stimuli with varying contrast and temporal frequency**

The presented findings so far were mostly based on the performance of the suppressive models on white-noise stimuli. Thus, an important question is how the derived models generalize to stimuli with different, non-stationary statistics, such as variations in contrast or temporal frequency. For a subset of recordings from marmoset cells, we applied a full-field chirp stimulus[16], which contained two sections of sinusoidal light-intensity modulations, one at 4 Hz with monotonically increasing contrast from 0% to 100% over 8 s (“contrast sweep”) and the other at 100% contrast with increasing frequency from 0 Hz to 15 Hz over 8 s (“frequency sweep”). We used measured firing rates in response to repeated presentations of these two sinusoidal light-intensity modulations to test the performance of the models after training on the white-noise stimulus.

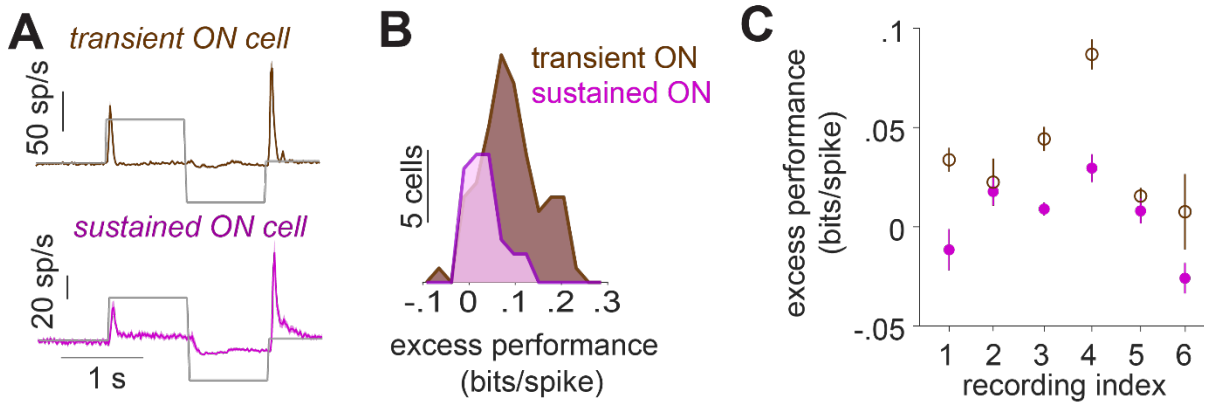


**Figure 4. A)** Responses of marmoset cells to a step stimulus for transient-ON and sustained-ON cells, averaged across all cells in a sample recording (recording index = 3). **B)** Distributions of the excess performance of the best-performing model, separately for transient and sustained ON cells for the same sample recording. **C)** Mean and standard error of the excess performance for transient- and sustained-ON cells in each recording.

As illustrated by the model responses for two sample cells shown in Fig. 5A-D, the degree to which the prediction of each model matched a cell's response varied with the frequency and contrast of the stimulus. For the sample cell shown for the contrast sweep (Fig. 5A), all models captured the response profile at low contrast quite well, displaying rectified response peaks with increasing amplitude for increasing contrast, which matched the recorded activity. For higher contrast, however, most models gave overshooting predictions, with response peaks that were too large and too long in time (Fig. 5A, right). Only the divisive model managed to yield good predictions here, reflecting how the cell's response events at high contrast were cut off and suppressed after the initial rising phase.

We quantified the dependence of model performance on contrast for the contrast sweep stimulus by calculating the explained variance ($R^2$) for each model within a sliding window of 150 ms duration, shifted along the time course of the stimulus in steps of 16.7 ms (1 stimulus frame). For the sample cell, the divisive model substantially outperformed the others for any contrast higher than about 20% (Fig. 5B). To quantify the performance across the population of cells, we calculated the performance range for each cell as the range of contrasts for which $R^2 > 0$. Over the population, the performance range was significantly larger for the divisive model, compared to other models ($p << 0.0001$ in each case; Fig. 5E), indicating that the divisive model generalized over a wider range of contrasts. This finding resonates well with a previous finding in mouse ON-alpha ganglion cells[1], demonstrating that the divisive model can capture adapting responses under contrast changes in a white-noise stimulus.

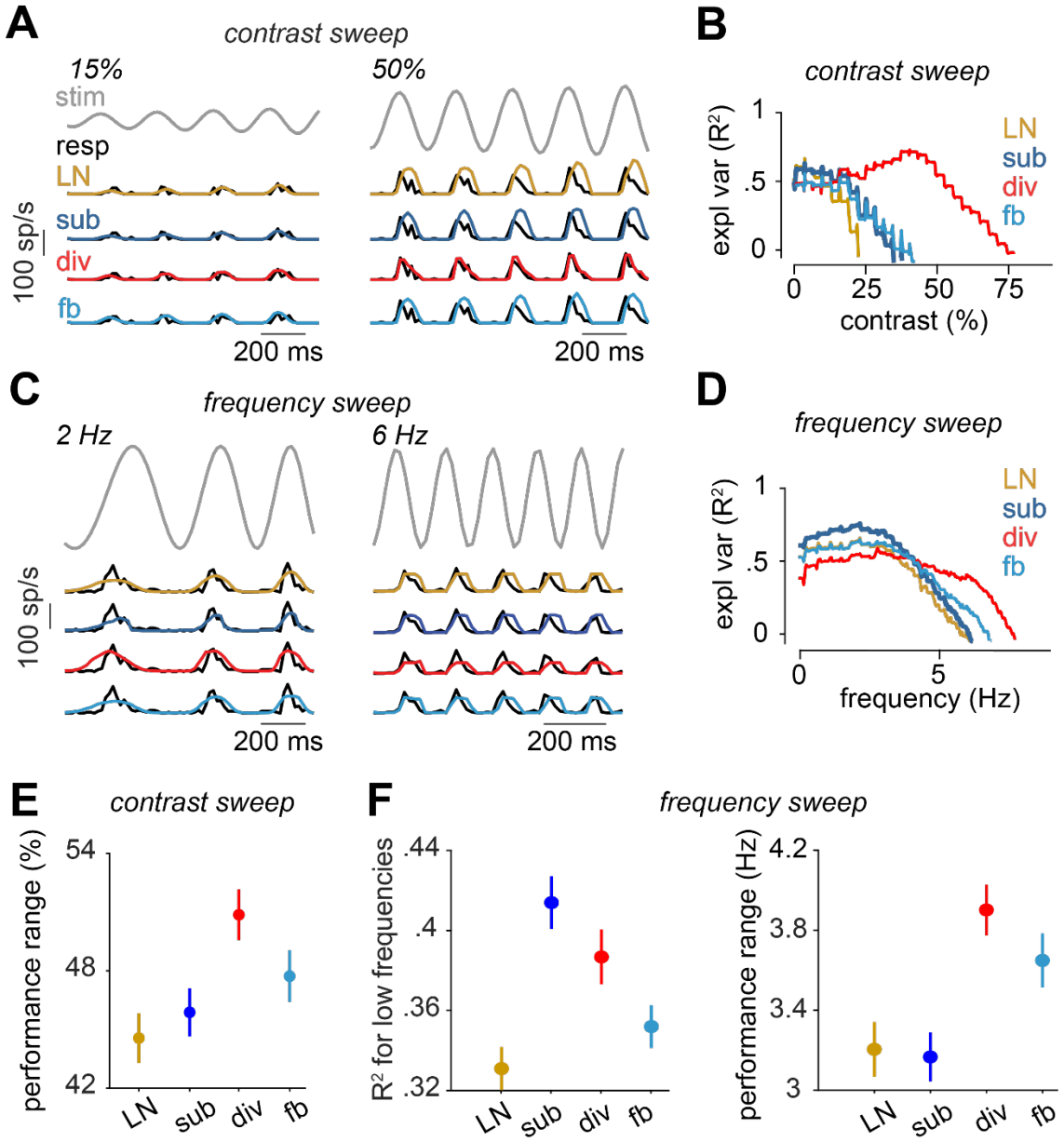


**Figure 5** Evaluation of the models on responses to contrast and frequency sweeps for marmoset cells. **A)** Part of the contrast-sweep stimulus together with the response of a sample cell and the model predictions. Two sections of the contrast sweep, with indicated starting contrasts and a duration of about 4-5 cycles, are shown. The complete range of the stimulus, the cell response, and the model predictions are provided in Fig. S5. **B)** Explained variance for the models for the cell in A, computed within a 150 ms sliding window. The x-axis shows the starting contrast of the sliding window. For each model, only the range with $R^2 > 0$ is shown. **C)** and **D)** Same as A and B, but for the frequency sweep and for a different sample cell. **E)** The contrast range for which each model had $R^2 > 0$ (mean and standard error over cells). **F)** *Left*: The $R^2$ of each model averaged over the first 10 data points of the frequency sweep analysis (mean and standard error over cells). *Right*: The frequency range for which each model had $R^2>0$ (mean and standard error over cells).

Similar to the contrast dependence of model performance, we observed that the relative performance of the models depended on temporal frequency. For the sample cell shown for the contrast sweep (Fig. 5C), model predictions under low frequency were often too wide and shallow. Here, the subtractive model managed best to generate fairly sharp responses,

approximately matching the time course of the data. For larger temporal stimulus frequencies, on the other hand, all models tended to overestimate the duration of response events (Fig. 5C, right). Yet, similar to the high-contrast scenario in Fig. 5A, the divisive model managed best to cut off some of the response prediction after the initial rising phase and thus yielded slightly more accurate predictions at higher frequencies.

We again quantified the frequency dependence of model predictions by computing $R^2$ within a sliding window, capturing different frequency ranges. As shown in Fig. 5D, the subtractive model outperformed the other models for the sample cell when the frequency of the stimulus was below about 4 Hz, whereas the divisive model had the highest performance among the tested models for higher frequencies. To assess the generality of these findings on the population level, we calculated for each cell the low-frequency performance, i.e., the average explained variance for the first 317 ms of the frequency sweep (corresponding to 10 shifts by one stimulus frame of the 150-ms analysis window) as well as the performance range, i.e., the range of frequencies for which $R^2 > 0$. For the subtractive model, the low-frequency performance was indeed significantly higher than for the other models ($p < 0.03$ for each comparison; Fig. 5F, left), indicating that this model performed best in capturing the cells' responses to low frequencies in the chirp stimulus. On the other hand, it was again the divisive model—like for the contrast sweep—that had the largest performance range ($p < 0.004$ for each comparison; Fig. 5F, right), indicating that this model generalized better over the frequency range of the chirp stimulus.

# Discussion

The interplay of excitation and suppression is a central theme for the generation and formation of neuronal responses to sensory stimulation. How to reflect the suppression in computational models of single cells is thus a question of broad interest. By testing models on spike-train data from retinal ganglion cells of three different species, we here found that three previously introduced approaches to incorporate suppression into filter-based models can all be used to improve response predictions over the commonly applied single-filter model, the linear-nonlinear model. Moreover, the subtractive and the divisive model generally outperform the feedback model for most cells in axolotl, mouse, and marmoset.

### Comparisons of model performances

The fact that the subtractive and the divisive model overall performed quite similarly for most cells might seem a bit disappointing at first glance, as the lack of a clear "winner" makes

inference about the type of excitation–suppression interaction difficult. On the other hand, this null-result means that, from a phenomenological perspective, the details of how suppression is included in the model matter much less than having a suppressive component at all, indicating that different interactions can achieve similar functions by different means. The functional equivalence of different model structures is reminiscent of studies of the stomatogastric ganglion in lobster where the potential variable implementations of the same circuit function has been forcefully demonstrated[17].

More detailed comparisons between these two models, however, revealed systematic differences in how performance depended on species and stimulus features. The subtractive model performed better for capturing responses of slow axolotl cells to white-noise stimuli and responses of marmoset cells to low-temporal-frequency parts of the chirp stimulus. The divisive model, on the other hand, displayed greater potential for generalizing over wider ranges of contrast and frequency. This suggests that different neural suppression mechanisms may dominate for different temporal scales of stimulation and response and that the two models differ in how well they capture the effects of these specific mechanisms. In particular, the subtractive model may specifically account for biological mechanisms that prevail in slow cells or that take effect when the stimulus is predominated by slow frequencies.

A technical aspect that may contribute to limiting the performance of the feedback model is the fact that suppression in this model depends on the occurrence of spikes, which are sparse and stochastic in the model. This limits the effectiveness of the suppression mechanism (e.g., it cannot suppress a spiking event that is not preceded by another spiking event) and induces variability in the predicted responses. Alternative implementations might let the feedback depend on the continuous internal signal before spike generation[18] or might apply a deterministic spike generation process[13,19].

**Model complexity and free parameters**

In the present study, the implementation of the compared models was guided by the goal of a common framework that made the model structures and the training most comparable. Yet, the number of free parameters was not identical between all models. Given the complexity of the models, this raises the question of whether the comparison might be influenced by over-fitting, which could reduce performance on the test data for models with a higher number of degrees of freedom. However, all evaluations occurred on held-out data not used for model fitting, and the subtractive and the divisive model actually had the same number of free parameters so the

observed differences are unlikely to arise from differences in fitting accuracy. Therefore, we believe that over-fitting is not likely to limit the interpretability of our findings.

A limitation of the present study is that we restricted our analysis to the use of full-field stimuli without any spatial structure. We made this choice in order to limit the number of free parameters in the model as well as to allow for comparisons between cell types and models independent of how well the spatial filtering and potential nonlinearities in spatial stimulus integration[20–23] are captured. On the downside, the use of full-field stimuli makes our analysis insensitive to the spatiotemporal structure of suppression, in particular by not providing for a dedicated suppressive signal pathway from the receptive-field surround.

**Comparisons across different species: general features and species-specific variations**

A particular aspect of our study was the comparison of the models across ganglion cells from different species. This showed an overall high level of consistency of the findings; the similar performance of the subtractive and the divisive model and the generally lower performance of the feedback model was observed in axolotl as well as mouse and marmoset. One somewhat exceptional group of cells was the set of slow axolotl cells that displayed a clear preference for the subtractive model over the divisive model. For these cells, the slowness is particularly pronounced, as filters and response kinetics in the axolotl are overall slower than in the other species, perhaps partly owing to the lower temperature at which the axolotl experiments were performed. This may make the subtractive model, which appeared to generally capture slow temporal dynamics better, particularly suited for the slow cells in this slow retina.

For marmoset ganglion cells, we found that including suppression in the response model, either through the subtractive or the divisive model, was more relevant for transient cells than for sustained cells, in particular when considering responses to steps in light intensity. In the primate retina, the ganglion cells types of midget and parasol cells numerically make up the majority of the ganglion cell population[24,25], and among these, parasol cells tend to have transient responses and midget cells sustained responses[25–27]. It thus seems likely that it is the parasol cells for which including suppression in the model is particularly relevant, though we have no independent classification of parasol and midget cells for the present datasets.

**Generalization beyond white-noise stimuli**

Like many other investigations of neuronal response models to sensory stimulation, we based our model fits on recordings under stimulation with white-noise sequences of light intensity. There is increasing awareness, however, that an important follow-up step is to assess how well

the obtained model descriptions and deduced insights hold in the context of different, typically more complex stimulus characteristics, such as natural stimuli. In the present study, rather than using natural stimuli, we investigated the generalization to non-white-noise, non-stationary stimuli, with variations in specific stimulus components, such as contrast and temporal frequency. Some studies of models without suppressive components have been used for different contrast levels by recalculating the stimulus components for each contrast[20]. Other investigations have suggested suppressive elements by feedback[18,28,29], by divisive gain control[30], or by complex biologically motivated combinations of nonlinear filtering and feedback[31] as mechanisms that capture contrast adaptation and thereby promote generalization across contrast levels. These findings are in line with our observation that suppressive model elements help generalization across contrast levels and frequency ranges and that the divisive model is particularly suited for this purpose.

## Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 724822) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project numbers 432680300 (SFB 1456, project B05) and 515774656.

## Contributions

FR, MHK, HMS, and TG designed the experiments. FR established the CMOS recording setup in the lab and collected and pre-processed the marmoset data, with help from NS. MHK, MM, and DAP performed preparations of marmoset retinas and helped with data collection. MHK performed preparations of mouse retinas and collected and pre-processed the mouse data. HMS collected and pre-processed the axolotl data. NS and TG designed the computational analysis. NS performed the analysis and the modeling. NS and TG wrote the manuscript.

## References


1. Cui, Y., Wang, Y. V, Park, S. J. H., Demb, J. B. & Butts, D. A. Divisive suppression explains high-precision firing and contrast adaptation in retinal ganglion cells. *Elife* **5**, e19460 (2016).
2. Butts, D. A., Weng, C., Jin, J., Alonso, J.-M. & Paninski, L. Temporal Precision in the Visual Pathway through the Interplay of Excitation and Stimulus-Driven Suppression. *J. Neurosci.* **31**, 11313–11327 (2011).
3. Wehr, M. & Zador, A. M. Balanced inhibition underlies. *Nature* **426**, 860–863 (2003).
4. Gabernet, L., Jadhav, S. P., Feldman, D. E., Carandini, M. & Scanziani, M. Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* **48**, 315–327 (2005).
5. Wilent, W. B. & Contreras, D. Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nat. Neurosci.* **8**, 1364–1370 (2005).
6. Heeger, D. J. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).
7. Albrecht, D. G. & Geisler, W. S. Motion Selectivity And The Contrast-Response Function Of Simple Cells In The Visual Cortex. *Vis. Neurosci.* **7**, 531–546 (1991).
8. Roska, B. & Werblin, F. Rapid global shifts in natural scenes block spiking in specific ganglion cell types. *Nat. Neurosci.* **6**, 600–608 (2003).
9. Roska, B., Molnar, A. & Werblin, F. S. Parallel processing in retinal ganglion cells: How integration of space-time patterns of excitation and inhibition form the spiking output. *J. Neurophysiol.* **95**, 3810–3822 (2006).
10. Latimer, K. W., Rieke, F. & Pillow, J. W. Inferring synaptic inputs from spikes with a conductance-based neural encoding model. *Elife* **8**, e47012 (2019).
11. McFarland, J. M., Cui, Y. & Butts, D. A. Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. *PLoS Comput. Biol.* **9**, 1003143 (2013).
12. Pillow, J. W. *et al.* Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995–999 (2008).
13. Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P. & Chichilnisky, E. J. Prediction and Decoding of Retinal Ganglion Cell Responses with a Probabilistic Spiking Model. *J. Neurosci.* **25**, 11003–11013 (2005).
14. Brenner, N., Strong, S. P., Koberle, R. & Bialek, W. Synergy in a neural code. *Neural Comput.* **12**, 1531–1552 (2000).

15. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
16. Baden, T. *et al.* The functional diversity of retinal ganglion cells in the mouse. *Nature* **529**, 345–350 (2016).
17. Prinz, A. A., Bucher, D. & Marder, E. Similar network activity from disparate circuit parameters. *Nat. Neurosci.* **7**, 1345–1352 (2004).
18. Berry, M. J., Brivanlou, I. H., Jordan, T. A. & Meister, M. Anticipation of moving stimuli by the retina. *Nature* **398**, 334–338 (1999).
19. Keat, J., Reinagel, P., Reid, R. C. & Meister, M. Predicting Every Spike: A Model for the Responses of Visual Neurons. *Neuron* **30**, 803–817 (2001).
20. Carandini, M. *et al.* Do We Know What the Early Visual System Does? *J. Neurosci.* **25**, 10577–10597 (2005).
21. Wienbar, S. & Schwartz, G. W. The dynamic receptive fields of retinal ganglion cells. *Prog. Retin. Eye Res.* **67**, 102–117 (2018).
22. Zapp, S. J., Nitsche, S. & Gollisch, T. Retinal receptive-field substructure: scaffolding for coding and computation. *Trends Neurosci.* **45**, 430–445 (2022).
23. Karamanlis, D., Schreyer, H. M. & Gollisch, T. Retinal Encoding of Natural Scenes. *Annu. Rev. Vis. Sci.* **8**, 171–193 (2022).
24. Dacey, D. Origins of Perception: Retinal Ganglion Cell diversity and the creation of parallel visual pathways. in *the Cognitive Neurosciences* 281–301 (2011).
25. Field, G. D. & Chichilnisky, E. J. Information processing in the primate retina: circuitry and coding. *Annu. Rev. Neurosci.* **30**, 1–30 (2007).
26. Gouras P. Antidromic responses of orthodromically identified gangion cells in monkey retina. *J Physiol* **204**, 407–419 (1969).
27. de Monasterio, F. M. & Schein, S. J. Protan-like spectral sensitivity of foveal Y ganglion cells of the retina of macaque monkeys. *J. Physiol.* **299**, 385–396 (1980).
28. Liu, J. K. & Gollisch, T. Spike-Triggered Covariance Analysis Reveals Phenomenological Diversity of Contrast Adaptation in the Retina. *PLoS Comput. Biol.* **11**, e1004425 (2015).
29. Gaudry, K. S. & Reinagel, P. Contrast adaptation in a nonadapting LGN model. *J. Neurophysiol.* **98**, 1287–1296 (2007).
30. Shapley, R., Enroth-Cugell, C., Bonds, A. B. & Kirby, A. Gain control in the retina and retinal dynamics. *Nature* **236**, 352–353 (1972).
31. Van Hateren, J. H., Rüttiger, L., Sun, H. & Lee, B. B. Processing of natural temporal

stimuli by macaque retinal ganglion cells. *J. Neurosci.* **22**, 9945–9960 (2002).

32. Liu, J. K. *et al.* Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization. *Nat. Commun.* **8**, 149 (2017).
33. Schreyer, H. M. & Gollisch, T. Nonlinear spatial integration in retinal bipolar cells shapes the encoding of artificial and natural stimuli. *Neuron* **109**, 1692-1706.e8 (2021).
34. Khani, M. H. & Gollisch, T. Diversity in spatial scope of contrast adaptation among mouse retinal ganglion cells. *J. Neurophysiol.* **118**, 3024–3043 (2017).
35. Khani, M. H. & Gollisch, T. Linear and nonlinear chromatic integration in the mouse retina. *Nat. Commun.* **12**, 1900 (2021).
36. Pouzat, C., Mazor, O. & Laurent, G. Using noise signature to optimize spike-sorting and to assess neuronal classification quality. *J. Neurosci. Methods* **122**, 43–57 (2002).
37. Krüppel, S. *et al.* Diversity of Ganglion Cell Responses to Saccade-Like Image Shifts in the Primate Retina. *J. Neurosci.* **43**, 5319–5339 (2023).
38. Muthmann, J. O. *et al.* Spike detection for large neural populations using high density multielectrode arrays. *Front. Neuroinform.* **9**, 28 (2015).
39. Hilgen, G. *et al.* Unsupervised Spike Sorting for Large-Scale, High-Density Multielectrode Arrays. *Cell Rep.* **18**, 2521–2532 (2017).
40. Samengo, I. & Gollisch, T. Spike-triggered covariance: Geometric proof, symmetry properties, and extension beyond Gaussian stimuli. *J. Comput. Neurosci.* **34**, 137–161 (2013).
41. Schwartz, O., Pillow, J. W., Rust, N. C. & Simoncelli, E. P. Spike-triggered neural characterization. *J. Vis.* **6**, 484–507 (2006).

# Methods

*Electrophysiology*

All the experiments and procedures conformed to national and institutional guidelines. Recordings from axolotl and mouse retina were approved by the institutional animal care committee of the University Medical Center Göttingen (protocol number T11/35). Marmoset retinas were 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 33.19-42502-04-17/2496).

The axolotl recordings were performed using 60-electrode perforated multielectrode arrays (MultiChannel Systems, Reutlingen, Germany; 30 µm electrode diameter and 100 µm minimum electrode spacing) as described previously[32,33]. The eyes were enucleated after 1 hour of dark-adaptation, and the retina was mounted over a 1.5-2 mm hole on a nitrocellulose filter membrane, which was then placed on the array with the ganglion cells facing the electrodes. During the experiment, the retina was perfused constantly with oxygenated (95% $O_2$ and 5% $CO_2$) Ringer's medium (110 mM NaCl, 2.5 mM KCl, 1 mM $CaCl_2$, 1.6 mM $MgCl_2$, 22 mM $NaHCO_3$, 10 mM D-Glucose monohydrate) at room temperature (~22°C).

The mouse recordings were done with 252-electrode arrays (MultiChannel Systems, Reutlingen, Germany; 30 µm electrode diameter, and 100 or 200 µm minimum electrode spacing) as described previously[34,35]. The mice (aged 8-13 weeks, of either sex) were dark-adapted for at least 1 hour and subsequently euthanized by cervical dislocation. The eyes were removed quickly and transferred to a chamber with oxygenated (95% $O_2$ and 5% $CO_2$) Ames' medium (Sigma-Aldrich, Munich, Germany), buffered with 22 mM $NaHCO_3$ (to maintain a pH of 7.4) and supplemented with 6 mM D-glucose. The eyes were dissected, and the cornea, lens, and vitreous humor were carefully removed. The retina was isolated from the pigmented epithelium and transferred to a multi-electrode array. The temperature of the recording chamber was kept constant at around 32–34°C, using an inline heater (PH01, MultiChannel Systems, Reutlingen, Germany) and a heating element below the array (controlled by TCX-Control 1.3.4, MultiChannel Systems), while the retina was perfused continuously with oxygenated Ames' medium (4–5 ml/min).

For both axolotl and mouse recordings, the activity of the ganglion cells was amplified, band-pass filtered (300 Hz to 5 kHz), and recorded digitally at 25 kHz (60 MEAs) or 10 kHz (252

MEAs) using the MC-Rack software (Version 4.6.2, MultiChannel Systems). The spikes were extracted from the recorded voltage traces with a custom-made spike-sorting program based on a Gaussian mixture model and an expectation-maximization algorithm[36].

The recordings from marmoset retina were performed as described previously[37]. Briefly, eyes were quickly enucleated and dissected to allow direct supply to the retina with oxygenated (95% O2 and 5% CO2) Ames' medium (Sigma- Aldrich, Munich, Germany), supplemented with 6 mM D-glucose, and buffered with 22 mM NaHCO3 to maintain a pH of 7.4. Retinas were then dark-adapted for 1-2 hours. For each recording, a small peripheral retina piece, with pigment epithelium removed, was placed onto a 4096-channel CMOS-based microelectrode array (21 µm electrode size and 42 µm pitch; 3Brain AG). The retina piece was mounted with the ganglion cell-side down, covered with a piece of translucent polyester filter membrane (Pieper Filter #PE0104700; 0.1 µm pores), and weighted down with a metal ring. During the recording, the retina was perfused with the oxygenated Ames' medium (4-5 ml/min), and the temperature of the recording chamber was kept constant at around 33°C in the same way as for the mouse recordings. Spike extraction and sorting were performed by HerdingSpikes2 (github.com/mhhennig/HS2). In short, spiking events were detected independently at each electrode using a fixed threshold, then simultaneous events in neighboring electrodes were triangulated to identify the location and shape of the spike. Finally, events were clustered into units, using a mean-shift algorithm, based on the location and first two principal components of the spike waveform[38,39]. Repeated units were detected and removed, based on the similarity between their responses to white-noise stimuli (see next section for stimulus details). To that end, the firing rate of each cell was binned according to the stimulus frames, then the correlation coefficient between the binned firing rates (with no time lag) was calculated for all pairs of cells within a simultaneously recorded population. Next, a graph of the simultaneously recorded cells was made with one node for each cell and one edge between two nodes if the response correlation between the two cells was > 0.3. Repeated cells were detected by finding cliques within this graph. Within each group of repeated cells, all except one were removed.

All retina preparations were performed under infrared illumination with a stereomicroscope equipped with night-vision goggles.

*Visual stimulation*

Visual stimuli for all recordings were generated and controlled by custom-made software written in C++ and using the OpenGL library and presented at low photopic light levels. For

both axolotl and marmoset recordings, the stimuli were presented on a monochromatic gamma-corrected white OLED monitor (eMagin, 800 × 600 pixels, 60 Hz refresh rate, 75 Hz in a few axolotl experiments). For the mouse recordings, the stimuli were displayed via modified DLP Lightcrafter projectors (evaluation module, 864 × 480 pixels, 60 Hz refresh rate, Texas Instruments Company, Dallas, USA), with the blue LED replaced by a UV LED (emission peak at 365 nm), as described previously[35]. Sizes of the projections of monitor pixels on the retina were 2.5 µm × 2.5 µm for axolotl, 7.5 µm × 7.5 µm for marmoset, and 8 µm × 8 µm for mouse. All stimuli were presented with a mean light intensity in the mesopic to low-photopic range (2.5 mW/m$^2$ for axolotl, 0.75 to 2.8 mW/m$^2$ for marmoset, and 5.06 to 5.2 mW/m$^2$ for mouse).

The white-noise stimulus was obtained by drawing new screen intensities randomly from a normal distribution at either 30 Hz (axolotl; except for 25 Hz and 37.5 Hz in one recording each) or 60 Hz (mouse and marmoset). The normal distribution of screen intensities was centered at half of the maximum intensity of the display and had a standard deviation of 30% of the mean light intensity. In the recordings from axolotl and mouse retinas, a fixed white-noise sequence of 150 or 300 intensity values ("frozen noise") was repeatedly inserted after every 750, 900, or 1500 intensity values (with variations coming from different recording sessions) of non-repeating white noise. For model evaluation, the data was subdivided into training and held-out test data. When frozen noise data was available (axolotl and mouse data), it was held out as the test data. Otherwise (marmoset data), 6.7 s of data were taken from every 33.3 s as held-out pseudo-trials for testing. The total amount of data varied between experiments, but was typically around 7 min for training and 2 min for testing (average over all analyzed cells: 427 ± 187 s (mean ± SD) for training, 113 ± 46 s for testing).

The contrast-step stimulus, the contrast sweep, and the frequency sweep were presented consecutively, in a combo called the chirp stimulus[16]. In each trial of the chirp stimulus, following 120 frames (i.e., 2 s) of darkness (0% contrast), ON and OFF stimuli (100% and 0% contrast, respectively) were presented for 60 frames each. After another 60 frames of darkness, the sinusoidal frequency sweep was presented, starting at 0 Hz and linearly ramping up to 15 Hz over a duration of 480 frames, fluctuating around the mean stimulus intensity with peak light levels at 0 and 100% contrast. After another 60 black frames, the sinusoidal contrast sweep was presented, starting at 0% peak contrast and linearly ramping to 100% at a frequency of 4 Hz over a duration of 480 frames. At the end, there were another 60 black frames. The chirp stimulus was repeated for 15 trials. The step part of the chirp stimulus was used for classifying

cells into transient and sustained cells. The frequency and contrast sweeps were used for evaluating the generalization of the models to a range of stimulus frequencies and contrasts.

*Model construction and training*

All model analyses were performed with time bins discretized at the update rate of the stimulus and responses given by the number of spikes in each time bin. Each model contains one branch (for the feedback and the LN model) or two convergent branches (for the subtractive and the divisive model) that represent upstream computation. After combining the two branches additively for the subtractive model and multiplicatively for the divisive model, the upstream computation in each model is followed by an output rectifier and a spike generator. Each upstream branch consists of a stimulus filter and a subsequent nonlinear function. The feedback model's feedback loop consists of a filter that is applied to the recent history of spiking activity with no nonlinear function (Fig. 1A). To maximize the consistency in implementation and training, we combined all four models in one master structure (Fig. S1). The input to this master structure was the contrast of the stimulus as a one-dimensional time series, with zero corresponding to mean light intensity. Each filter was parametrized by sampling at the stimulus update rate, using 30 samples for axolotl, 25 for mouse, and 15 for marmoset. Each upstream nonlinear function was parametrized with a set of basis functions as done previously[11]. The output rectifier was described as a "softplus" function, $f(x) = m \ln(1 + e^{ax+b}) + c$, with $a$, $b$, $c$, and $m$ representing free parameters.

In the unified structure, only one model component, i.e., a filter, an upstream nonlinearity, or the rectifier function, was trained at a time. To this end, we write the output of this structure as the expansion in the basis functions of the branch that is under training, while describing the rest of the model by a set of fixed contributions, e.g.:

$$r(t) = f\left(\left[\pm \sum_i \left(w_i . g_i \left(\sum_\tau \left(k_\tau X(t-\tau)\right)\right)\right) \pm P(t)\right] \times M(t) + A(t)\right)$$

in which $k_\tau$ are the elements of the filter under training and $w_i$ the weights for the basis functions of the corresponding nonlinear function, $f$ is the output rectifier, and $g_i$ are the pre-specified basis functions of the nonlinearities. For training each branch of each model, variables $X$, $P$, $M$, and $A$ were defined from Table 1. This table shows, for example, that when the excitatory branch of the subtractive model was being updated, it was represented as the weighted sum in the formula above, while the suppressive branch was represented by variable $P$ and vice versa. Similarly, for the divisive model, when the excitatory branch was being

updated, the suppressive branch was represented by M and vice versa. Note that M is multiplicative in the formula, but acts as suppression, because the output of the suppressive branch is constrained to be in the range of zero to unity, owing to the constraint on its nonlinear function.

***Table 1*** *Assignment of the variables to derive each model from the unified structure.*

| **Model** | **Components** | ***X*** | ***P*** | ***M*** | ***A*** |
|---|---|---|---|---|---|
| **Subtractive** | exc (Excitatory) | stim | sup output | 1 | 0 |
| | sup (Suppressive) | stim | exc output | 1 | 0 |
| **Divisive** | exc | stim | 0 | sup output | 0 |
| | sup | stim | 0 | exc output | 0 |
| **Feedback** | exc | stim | 0 | 1 | feedback output |
| | sup | spikes | 0 | 1 | exc output |
| **LN** | exc | stim | 0 | 1 | 0 |

The branch being trained was updated using a block-coordinate descent algorithm, in which $k_\tau$ were updated to minimize the negative log-likelihood function, while $w_i$ were constant and vice versa. The components of the model were trained in this order: upstream excitatory, upstream suppressive (for subtractive and divisive models), the feedback loop (for the feedback model), and the output rectifier.

The objective function was the negative of the log-likelihood, $-LL$, and was calculated as the natural logarithm of the total probability that the number of spikes in each time bin $t$ would be drawn from a Poisson distribution with an expectation value of $r(t)$ for each bin $t$. In each training iteration, we used the function fmincon in Matlab to take 10 steps per model component toward minimizing the objective function. We trained each model for either 100 iterations or until the change in the objective function was smaller than 0.01%. In fmincon, we used the 'trust region reflective' algorithm. The constraints on the $w_i$ were applied using the sequential quadratic programming algorithm. The derivative of the objective function was provided to fmincon by calculating the derivatives of $-LL$ with respect to the $w_i$ as well as to the $k_\tau$ and to the free parameters of $f(\cdot)$. The constraints for each variable are specified in Table 2.

***Table 2*** *The optimization constraint for the set of variables of each model component*

| **Model** | **Exc filter** | **Exc function** | **Sup filter** | **Sup function** |
|---|---|---|---|---|
| **Subtractive** | Norm = 1<br>tail* = 0 | Monotonically increasing<br>Lower bound = $10^{-16}$ | Norm = 1<br>tail = 0 | Monotonically increasing<br>Lower bound = $10^{-16}$ |
| **Divisive** | Norm = 1<br>tail = 0 | Monotonically increasing<br>Lower bound = $10^{-16}$ | Norm = 1<br>tail = 0 | Monotonically increasing for input < 0 &<br>decreasing for input > 0<br>Scaled to [$10^{-16}$, 1] |
| **Feedback** | Norm = 1<br>tail = 0 | Monotonically increasing<br>Lower bound = $10^{-16}$ | tail = 0 | - |
| **LN** | Norm = 1<br>tail = 0 | Monotonically increasing<br>Lower bound = $10^{-16}$ | - | - |

* The tail of the filter was the average of the values of the last 5 time bins.

To reduce the risk of finding non-optimal solutions corresponding to local minima of the objective function, the training was repeated 5 times with different initializations that were guided by simple spike-triggered analyses. We then selected the solution for which the value of the objective function (negative log-likelihood) for the training data was the lowest. The initial values were selected from the spike-triggered average, the most relevant stimulus features extracted from a spike-triggered-covariance analysis[40,41], pre-defined deterministic functions and Gaussian noise, as specified in Table 3 (when different initial values were used for each model, the entries in the table follow the format of sub / div / fb / LN) where STA was the spike-triggered average, $STC_1$ and $STC_N$ were the first and the last principal components of the spike-triggered covariance (with the STA projected-out from the stimulus segments before the covariance analysis), N(0,1) was Gaussian white noise with mean of zero and standard deviation of unity, $W_{mono}$ was the softplus function, as defined above, with a = 10, b,c = 0, and m = 0.1, $W_{bi}$ was a bell-shaped function (Gaussian spanning approximately 2 standard deviations over the range [-3, 3], normalized to a magnitude range of [0, 1]). To add variability in the initialization, Gaussian noise with a mean of zero and standard deviation of 0.1 was added to the original output values of $W_{mono}$ and $W_{bi}$ at each time.

***Table 3*** *Initial values for 5 training iterations*

| **Run number** | **Exc filter** | **Exc function** | **Sup filter** | **Sup function** |
|---|---|---|---|---|
| **1** | STA | $W_{mono}$ | $STC_1$/ $STC_1$ / 0 / - | $W_{mono}$ / $W_{bi}$ / - / - |
| **2** | STA | $W_{mono}$ / $W_{bi}$ / $W_{mono}$ / $W_{bi}$ | $STC_N$/ $STC_N$ / 0 / - | $W_{mono}$ / $W_{bi}$ / - / - |
| **3** | $STC_1$ | $W_{mono}$ | $STC_N$ / $STC_N$ / 0 / - | $W_{mono}$ / $W_{bi}$ / - / - |
| **4** | $STC_N$ | $W_{mono}$ / $W_{bi}$ / $W_{mono}$ / $W_{bi}$ | $STC_1$/ $STC_1$ / 0 / - | $W_{mono}$ / $W_{bi}$ / - / - |
| **5** | N(0,1) | $W_{mono}$ | N(0,1) | $W_{mono}$ / $W_{bi}$ / - / - |

For all models and iterations, the initial value of the output rectifier was $10*\ln(1+e^{0.1x})$.

*Evaluation of the fitted models*

The best solution of each model for each cell was finally evaluated on and the held-out test data. To do so, the model performance was assessed by calculating the average information per spike provided by the model about the cell's spiking response[2,14]. The information per spike was calculated as

$$\frac{LL\,(r,\hat{r}) - LL\,(r,\bar{r})}{\bar{r}\ln(2)}$$

Here, $r$ denotes the spike train data (spikes per bin), $\bar{r}$ the corresponding average, and $\hat{r}$ the prediction by the evaluated model. The denominator ensures normalization to obtain the information in bits/spike. Note that this measure of model performance does not require repeated trials. For the feedback model, the performance was averaged across 100 evaluations, each with spike counts obtained stochastically according to a Poisson distribution with expectation value given by the model's output.

Poisson explained variance was used as a secondary measure of the model performance when firing rates over repeated trials of frozen noise were available as held-out test data. This was computed as

$$\text{Poisson explained var} = 1 - \text{dev}_{\text{Poisson}}(r,\hat{r})/\text{dev}_{\text{Poisson}}(r,\bar{r})$$

in which $\text{dev}_{\text{Poisson}}$ was the Poisson deviance, calculated as

$$\text{dev}_{\text{Poisson}}(r,\hat{r}) = 2\sum\left(\ln\left(\left(\frac{r}{\hat{r}}\right)^{r}\right) - (r-\hat{r})\right)$$

where the sum runs over time bins. Here, $r$ is the trial-averaged spike count per time bin, $\hat{r}$ the corresponding model prediction, and $\bar{r}$ the average firing rate (per time bin) over the frozen-noise section.

The performance of the models for the frequency and contrast sweep stimuli were computed using the explained variance ($R^2$) according to

$$R^2 = 1 - \frac{\sum(r-\hat{r})^2}{\sum(r-\bar{r})^2}$$

Here, again, the sums run over all time bins of the analyzed time window as specified in the main text, $r$ is the trial-averaged firing rate per time bin, $\hat{r}$ the corresponding model prediction, and $\bar{r}$ the average firing rate over the analyzed time window.

*Selection of units*

We selected units for further analysis based on the following criteria: 1. The average firing rate of the unit under the white-noise stimulus was required to be larger than 2 Hz for axolotl and larger than 5 Hz for mouse and marmoset. 2. Responses to the frozen-noise presentation for units from axolotl or mouse were required to be sufficiently reliable across trials. To this end, response reliability was assessed by dividing the responses of a cell into two subsets of equal sizes, computing a PSTH separately for each subset, and calculating the variance of the first PSTH explained by the second. The reliability was obtained as the average explained variance across 20 random subset divisions and was required to be larger than 0.5. 3. The difference between the average firing rate in the early (first 30% of the stimulus duration) and late (last 30% of the stimulus duration) part of the recording under the white-noise stimulus was required to be less than 50% of the overall average firing rate. 4. The unit was required to not correspond to an ON-OFF cell, i.e., it should not respond to both light increments and light decrements. ON-OFF cells were detected by a U-shape nonlinear function of the LN model: The function was divided into the left (input $< 0$) and the right (input $> 0$) part, then a line was fitted to the left (slope = sL) and another line was fitted to the right part (slope = sR). A cell was classified as ON-OFF, if $sL / (|sL| + |sR|) < -0.2$. 5. Performance on the test data was required to be at least 60% of the performance on the training data for all four models to reduce the effects of overfitting.

*Statistical analysis*

Two-sided Wilcoxon signed-rank test was used, except where another test was indicated. When multiple groups of data were compared, false discovery rate (FDR) correction for multiple comparisons[15] was used to correct the *p*-values.

*Data and code availability*

The data and source code for model fitting and generating figures of this manuscript have been made publicly available at the GRO.data repository, DOI: https://doi.org/10.25625/AP5BRQ.

# Supplemental figures

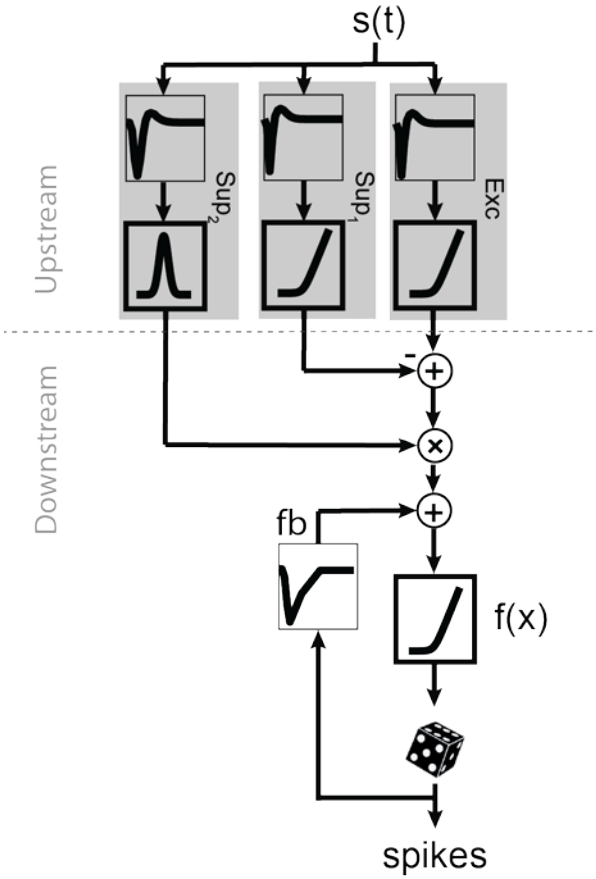


**Figure S1** The unified model structure that incorporates all components of the models in Fig. 1A. The filter components are shown in thin frames and the nonlinearities in thick frames. Exc upstream branch represents the excitation, $Sup_1$ represents the subtractive suppression, and $Sup_2$ represents the divisive suppression. The feedback (fb) was added to the downstream part of the model.

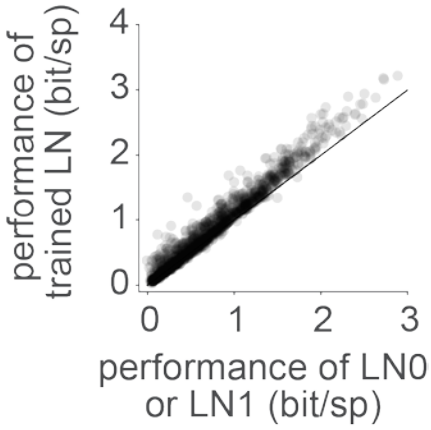


**Figure S2** Comparing the performance of our trained LN model to the performance of an LN model that uses the STA as the filter (LN0) or the first principal component of the STC (LN1). For both LN0 and LN1, the nonlinear function is computed as the histogram of filter outputs versus spike counts. For the comparison here, either LN0 or LN1 was chosen, depending on which model had better performance.

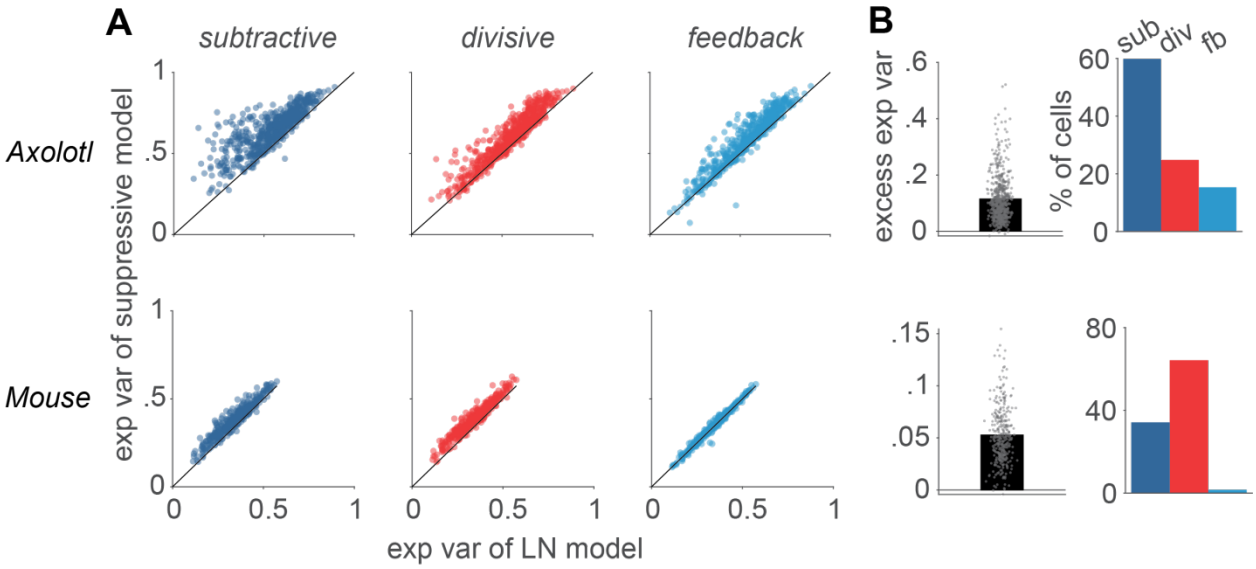


**Figure S3** Same as Fig. 1B-C, but for Poisson explained variance (exp var), as measured on the frozen-noise section of the data. **A)** Comparison of the Poisson explained variance of each model to the performance of the LN model. Each data point represents one cell. **B)** *Left*: The excess explained variance (computed as explained variance of suppressive model minus explained variance of the LN model) of the best performing suppressive model, determined for each cell. *Right*: The percentage of cells for which each of the explained variance of the suppressive models was higher than all the other models.

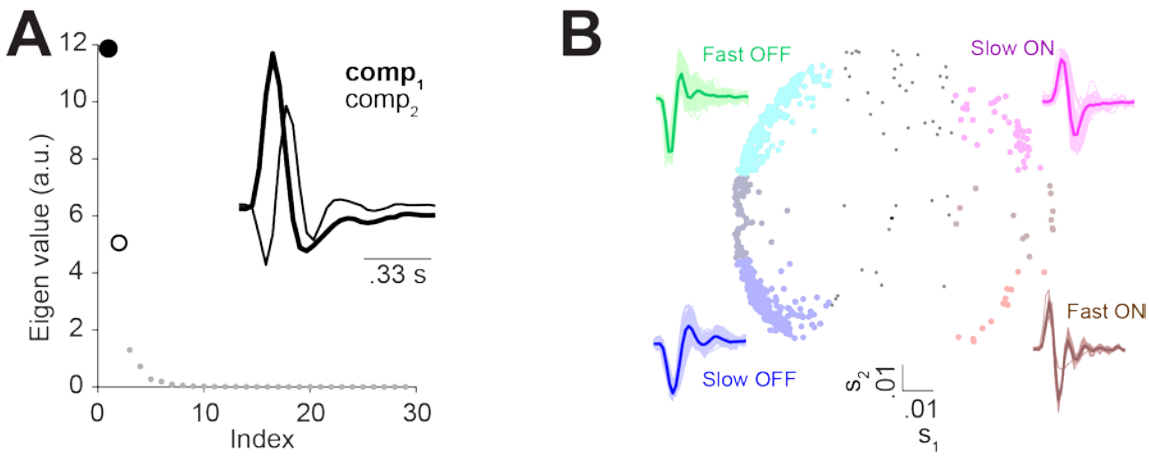


**Figure S4** Principal component analysis of the excitatory filter shapes to separate cells recorded from axolotl retina into functional classes. **A)** The eigenvalues associated with 30 principal components of the filter of the LN model. *Inset:* The first and the second principal components, which were used for classification. **B)** *Center:* All analyzed axolotl cells in the space of $s_1$, the score for the first principal component, and $s_2$, the score for the second principal component. To determine the classes, the scores for each cell were compared to threshold of ±0.02: for slow ON cells (n = 48) $s_1 > 0.02$ and $s_2 > 0.02$, for fast ON cells (n = 13) $s_1 > 0.02$ and $s_2 < -0.02$, for slow OFF cells (n = 195) $s_1 < -0.02$ and $s_2 < -0.02$ and for the fast OFF cells (n = 208) $s_1 < -0.02$ and $s_2 > 0.02$. 143 cells remained unclassified, as they did not pass any of the threshold combinations. *Insets surrounding the scatter plot:* LN-model filters (thin lines) for each of the identified groups, together with their average (thick line).

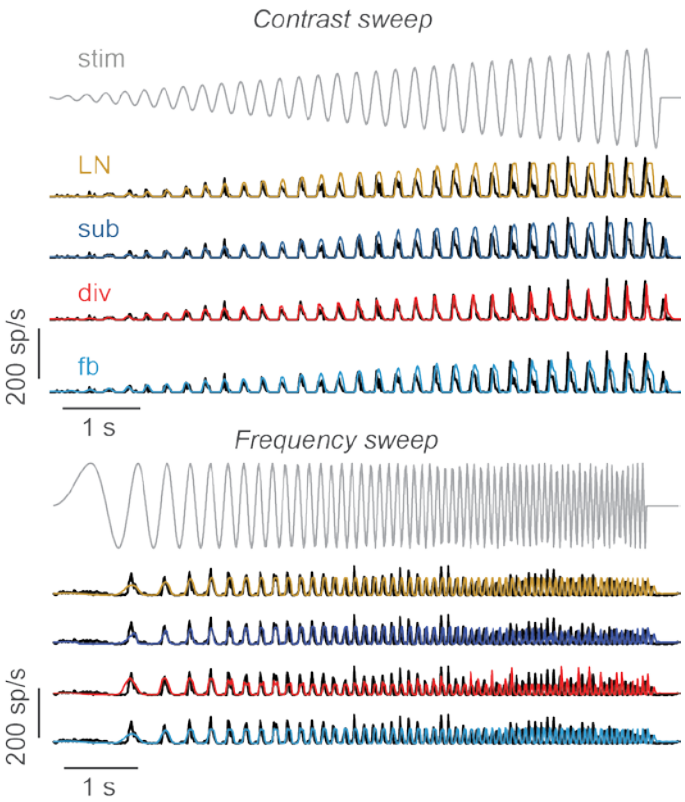


**Figure S5** The responses of the sample cells from Fig. 5A-D (black lines) and the model predictions (colored lines) for the entire ranges of the contrast sweep and the frequency sweep. The stimulus traces (gray lines on top) are displayed with temporal sampling corresponding to the monitor update. **A)** The cell from Fig. 5A-B for the contrast sweep. **B)** The cell from Fig. 5C-D for the frequency sweep.