Supplemental Information

Experimental procedures

Cell identification and selection

In recordings made in wildtype retina, ooDS RGCs were identified by their soma size and shape, low maintained firing rate, transient ON and OFF responses to a light step, and direction selective spike responses to moving bars stimuli (see Figure S1).

In recordings made in TRHR-GFP and BDxTSY retinas, where two-photon targeting also aided ooDS cell identification, laser exposure was kept to a minimum, typically <1 minute exposure, <2 mW post-objective laser power at 960 nm. Recording from ON alpha RGCs at the light levels for these experiments (~50 R*) before and after typical laser exposure suggested that contrast sensitivity is not substantially affected by this targeting. For loose-patch (cell attached) spike recordings, only retinal mounts that met certain health and sensitivity requirements were used. Tissue health was judged by visual inspection, and contrast sensitivity was measured using loose patch recordings from ON alpha RGCs, which, in sufficiently sensitive tissue, produced robust spike responses to 10% contrast spots centered over their receptive fields.

Alternating voltage-clamp recording

DS RGC pairs and single cell conductances were measured using a previously described technique [Cafaro and Rieke, 2010] that allows simultaneous measurement of both excitatory and inhibitory synaptic conductances (see Fig. S3). Cells were whole-cell patched and voltage-clamped using a MultiClamp 700B amplifier (Molecular Devices) compensated at 50%. Series resistances were typically between 10 and 20 MΩ. Pipettes for voltage-clamp recordings were filled with a Cs-based internal solution (105 mM $CsCH_3SO_3$, 10 mM TEA-Cl, 20 mM HEPES, 10 mM EGTA, 5mM Mg-ATP, 0.5mM Tris-GTP and 2mM QX-314, pH 7.3, 280mOsm). Excitatory and inhibitory reversal potentials were assessed empirically during a full-field light step. Holding potentials used were chosen to be close to the assessed reversal potentials while remaining conservatively between the two, in order avoid noise correlation measurement error, as previously described. Sets of moving bars were then presented in three blocks while synaptic currents were recorded. The cell was first voltage-clamped near the inhibitory reversal potential for one trial, and then near the excitatory reversal potential, and then the holding potential was alternated every 5 ms between the excitatory and inhibitory reversal potentials. These blocks of three were repeated for 10-45 trials.

Excitatory and inhibitory synaptic currents were extracted from the current recorded during the alternating voltage trials by interpolating between the last sample point of current recorded before the next voltage change. This process effectively measures excitatory and inhibitory synaptic currents simultaneously at a sample rate of 100 Hz and is able to record the majority of both signal and noise.

Conductances were calculated from both non-simultaneous currents, recorded during the single hold voltage trials, and simultaneous currents, extracted from the alternating voltage trials. Conductances were calculated by dividing the measured excitatory and inhibitory currents by an assumed driving force, -62 mV and 62 mV respectively. The average excitatory and inhibitory synaptic conductances measured during the single hold trials are very similar to those measured during the alternating voltage trials, supporting previous results.

Dynamic clamp experiments

Synaptic input was mimicked and manipulated during dynamic clamp experiments [Sharp, et. al., 1993]. In brief, cells were whole-cell patched and current-clamped. Pipettes for dynamic-clamp experiments were filled with a K-based internal solution (110mM K aspartate, 1mM MgCl, 10mM HEPES, 5mM NMDG, 0.5mM CaCl2, 10mM phosphocreatine, 4mM Mg-ATP and 0.5mM Tris-GTP, pH ,7.2, ,280mosM). Current injected into a cell (I) during dynamic clamp experiments was calculated using equation (1) below.

$$I(t) = G_{exc}(t) \times (V(t - \Delta t) - E_{exc}) + G_{inh}(t) \times (V(t - \Delta t) - E_{inh}) \quad (1)$$

Where G_{exc} and G_{inh} are a set of conductances recorded during light stimulation, V is the cell's membrane potential, and E_{exc} and E_{inh} are reversal potentials set at 0 mV and -60mV respectively. The exact inhibitory reversal potential did not substantially impact the highlighted results. The absolute amplitude of the conductances were adjusted for each cell to achieve spike numbers near those recorded from the DSGCs from which the conductances were recorded and a constant current was often injected to help the cell maintain the appropriate resting potential.

Data analysis

Custom analysis scripts were written in MATLAB. In cell-attached recordings, we measured spike counts while the bar was swept over the receptive field. The response window was 2-3 seconds long. Because ooDS cells have a very low maintained (spontaneous) spike rate, our spike counts typically reflect the entirety of the response to each stimulus presentation.

Direction selectivity was quantified as follows. For each stimulus angle θ, we defined the two-dimensional direction vector $d(\theta) = (\cos(\theta),\sin(\theta))$. We then multiplied these direction vectors by the mean firing rate induced in the by that stimulus, to get the vector $v(\theta) = r(\theta)d(\theta)$. The length of this vector is the cell's mean firing rate (for this particular stimulus), and the direction indicates the direction of stimulus motion. We averaged these vectors over stimuli and computed the norm of the result. If the cell fired equally to stimuli in all directions, the horizontal and vertical components of the $v(\theta)$ vectors would cancel, leading to the average vector having small magnitude. Finally, we scaled the mean vector by the cell's overall average firing rate (averaged over all stimuli). This yields a number between 0 (cell responds equally to stimuli in all directions) and 1 (cell responds only to motion in 1 direction), that we called the Direction Selectivity Index (DSI).

$$DSI = \left| \frac{\langle \vec{v}(\theta) \rangle_\theta}{\langle r(\theta) \rangle_\theta} \right|$$

Over the course of paired recordings, both spike and conductance recordings are subject to slow, long-term drift that, if left uncorrected, can introduce artifactual noise correlation measurements. To correct for long term drift in cell attached recordings (Fig. 1), we subtracted the window (radius 2 trials) averaged spike count from each trial. By inspection of spike counts across trials and the trial autocorrelation function [Bair et al., 2001], this window-correction method effectively removed long-term drift from reported data. Qualitative results agree using uncorrected spike counts, window-corrected data (reported in main paper), and Bair *et al*.'s trial cross-covariance correction method. Similarly, for synaptic conductance (co)variance measurements, we used window-averaged conductance (radius 1 trial) for calculation of the residuals.

The simultaneously-measured synaptic currents from our alternating voltage experiments were used to estimate the synaptic gains and (co)variances in our network model (Fig. 3 & 4). For each input and each stimulus, we measured the mean conductance during the window of time that the mean conductance exceeded a value of 25% of its maximum for that stimulus. This ensured that we isolated the light-driven component of the responses. The mean conductance was used as the estimate of the gain for that input (in response to that stimulus), in our network model. We computed each (co)variance as a function of time delay by taking the cross-correlation, using MATLAB's built-in xcorr function, between the appropriate residuals at at each trial, and averaging across trials. The estimated (co)variance measure for our network model was defined as the peak of this (co)variance function.

The strength of the linear relationship between the appropriate gain product and (co)variance measurement (see Figs. 3&4), depended on the modulation of the (co)variance by the stimulus (Fig. S5). Fit quality was also affected by limited numbers of trials in these technically challenging paired alternating-voltage experiments: when only subsets of the total number of trials were used to fit the linear model, the fit qualities for all pairwise covariances declined (data not shown).

Structure of the Computational Model

The model follows the structure shown in Fig. 4A of the main paper, and supplemental Fig. S7.

Stimulus-dependent gain factors

In the literature, there is debate about whether or not excitatory input to DS RGCs is direction tuned. Recent optical approaches to this question suggest that excitatory input is untuned (Yonehara et al., 2013; Park et al, 2014). Apparent tuning of excitatory inputs is influenced by choosing an inappropriate holding potential as well as space clamp errors (Poleg-Polsky & Diamond, 2011). In our recordings, we see some modulation of excitation with stimulus direction, but that modulation is much smaller than the modulation of the inhibitory inputs (i.e., see the variation in peak heights of excitation and of inhibition as the stimulus changes in Fig. 3C). Given these concerns, and our motivation to simplify our computational model — which helps in fitting the free parameters to our neural data — we chose to construct the model with untuned excitation. The magnitude of the excitatory input, g_e, varies slightly from cell to cell. The population's mean g_e, is a model parameter (to be fit so as to match the experimental data) and individual cells' g_e values are drawn from a distribution with the given mean, and standard deviation 10% of the mean.

The inhibition to a given cell follows a raised sinusoid function,

$$g_i = g_{i,base} + g_{i,max}(0.5 + 0.5 \times sin(\theta + \phi))^{\alpha} \quad (2)$$

where α is a parameter that sets the sharpness of the tuning curve for the inhibitory input, $g_{i,max}$ determines the amplitude of the tuned part of the inhibition, and $g_{i,base}$ determines the magnitude of the untuned part of the inhibition. The stimulus angle is θ, and the phase angle ϕ determines the stimulus angle at which inhibition peaks (the null direction of the cell). Each cell's phase angle ϕ is initially set to either 0˚, 90˚, 180˚, or 270˚. These are then all randomly jittered by up to +/- 10˚.

Similar to the excitatory tuning, the population mean $g_{i,base}$ and $g_{i,max}$ values are model parameters, chosen to make the model responses match the experimental ones. Individual cells' $g_{i,base}$ and $g_{i,max}$ values are drawn from a distribution with the given means and standard deviations 10% of the means.

The population mean of the tuning curve (TC) sharpness α value is another model parameter. Individual cells' α values are drawn from a distribution with the mean α_{mean} -- which is a free parameter of the model -- and standard deviation 0.5. For a given stimulus, the gain factors for all cells are computed based on the above functions. These then multiply noisy inputs.

Shared and independent noise

In accordance with Fig. 4A of the main paper, the incident noise to the cells is correlated, based on the overlap of their effective "dendritic fields". This is done by making a matrix of the correlations in the input noise to all cells. This matrix has a maximum of 0.95 for any cell pair, and the minimum value β_{min} is another free parameter in the model. The off-diagonal entries of the input correlation matrix are then drawn uniformly between β_{min} and 0.95. Because these matrices need to be positive semi-definite (PSD; as must any correlation matrix), the model keeps drawing random matrices until a PSD one is found. We then draw noise from N(0,1) Gaussian distributions (mean 0, variance 1) with the correlation matrix given by the matrix of β values.

The incident noise in the inhibitory (I) channel for each cell is then σ_c times the (correlated) input noise generated by the above procedure. σ_c is the standard deviation of the (correlated between cells) incident noise, and is a free parameter of the model. This parameter is the same for all cells. The incident inhibitory noise is then multiplied by the inhibitory gain, computed from the raised-sinusoid g_i tuning function (above), whereupon independent noise of standard deviation $\sigma_{ind,I}$ is added. This parameter -- describing the variance of the independent noise in the I channels is another free parameter of the model, and is the same for all cells.

The incident noise in the excitatory (E) channel is $\gamma\sigma_c$ times the (correlated) input noise generated by the above procedure. γ is a parameter that amplifies the common E noise relative to the common I noise, and is a free model parameter. γ is the same for all cells. Similar to inhibition, the incident noise in the E channel is multiplied by the gain factor g_e, and independent noise of standard deviation $\sigma_{ind,E}$ is added thereafter. $\sigma_{ind,E}$ is another free parameter of the model, and is the same for all cells.

Incident signal

In addition to multiplying the incident noise, the stimulus-dependent gain factors also multiply constant "incident signals". The amplitudes of these are the same for all cells, and are free model parameters: E_{in} and I_{in}.

Net E and I inputs to the cells -- signal plus noise

The E and I inputs to cell j on a given trial are thus

$$E_j = g_{e,j} \times (E_{in} + \gamma\sigma_c\eta_{input,j}) + \sigma_{ind,E}\eta_{ind,E,j} \quad (3)$$

and

$$I_j = g_{i,j} \times (I_{in} + \sigma_c\eta_{input,j}) + \sigma_{ind,I}\eta_{ind,I,j} \quad (4)$$

where the values of $\eta_{input,j}$ for each cell are drawn trial-by-trial from a joint Gaussian distribution with mean zero and variance 1, with the correlation matrix $\varrho_{jk} = \beta_{jk}$, and β_{jk} is the matrix of β values that describes the overlap in the pools of upstream noise to the two cells. The other noise values $\eta_{ind,E,j}$ and $\eta_{ind,I,j}$ are, for each cell, drawn trial-by-trial from independent and identically distributed (i.i.d.) Gaussian distributions with mean 0 and variance 1.

Integration to get (noisy) inputs

To summarize, for a given model population and stimulus, the correlated incident noise is generated, and enters each cell, in addition to the constant "incident signal" for each channel. These are then multiplied by the stimulus-dependent gain factors, and independent noise is added thereafter into each channel, to get the (noisy) E and I inputs to the cells (Eqs. 3,4).

These are then combined to yield the "net" input to the cell, which is x = 3E - I. The factor of 3, which is fixed in the model, is based on the fact that the reversal potential for E channels is relatively far from the cell's reversal potential, while the reversal potential of the I channels is relatively close to the cell's reversal potential. As a result, excitatory conductances generate (proportionately) more current than do inhibitory ones.

Generating the (output) responses

The net inputs are passed through a threshold-linear function. The population mean threshold (τ) is a model parameter, and each cell's individual threshold is drawn from a distribution with this mean, and standard deviation equal to 10% of the mean. The slope s of the threshold-linear function is the final free parameter, and is the same for all cells.

After the thresholding operation, independent noise of small amplitude (variance 1) is added to all the responses, and these (noisy) responses are positive-rectified. This last stage avoids the possibility of there being stimuli for which a cell's responses are 0 on all trials. In that case the covariance and correlations between that cell's responses and those of other cells are undefined. We also observe small variations in spike counts when we repeatedly inject the same current traces into ooDS cells (not shown), justifying this modeling choice.

Fitting the free parameters of the computational model

Our model (described above) has 13 free parameters: g_e, $g_{i,base}$, $g_{i,max}$, α_{mean}, β_{min}, σ_c, γ, $\sigma_{ind,E}$, $\sigma_{ind,I}$, E_{in}, I_{in}, τ, and s. We wished to choose the free parameters so that the model cells' responses closely match those observed in our experiments. To do this, we simulated the model population responses, and compute 15 statistics that describe them (discussed below and Fig. 6AB of the main paper). We then searched for sets of model parameters for which the response statistics were similar for both the models and the experimental data.

This search was challenging for a few reasons. Mainly, for a given set of parameters, the models are constructed probabilistically. This means that, even with the same parameters, not all models will yield identical responses. As a result, gradient-based optimization methods are ill-suited to our parameter-fitting problem.

To solve this problem, we used a simple evolutionary algorithm. In it, random sets of model parameters were chosen. We then computed the response statistics associated with these parameters -- averaged over 25 populations generated with the same underlying parameter values -- and assigned each parameter set a "score" based on how closely the statistics generated match the experimental data. Parameter sets were chosen to minimize their scores -- like in a golf game, high scores are associated with poor performance.

The score function that we chose was the mean Z-score (absolute value of model mean minus experimental mean divided by experimental S.E.M.) of the different statistics plus the maximum Z-score over all the different statistics. As a result, the models were forced to fit all of the statistics well, and could not achieve a low score by fitting most statistics well, at the expense of a few poorly fit ones.

In practice, it is easier to fit the single-cell statistics than the pairwise ones, so we multiplied all of the Z-scores associated with pairwise statistics (correlations, tuning curve overlaps, etc.) by 1.5 before computing the parameter set's score. This score function thus judges parameter sets based on the statistics that they fit the worst, and judges them more strongly on their ability to fit pair-wise statistics than single-cell statistics.

For the first generation of our evolutionary algorithm we chose 40 sets of random model parameters, and evaluated their scores as outlined above. We then chose the top two parameter sets, and copied them each 20 times, with random perturbations to each parameter, to create the second generation of parameter sets. This procedure (selection and copying with perturbations) was repeated 50 times, with the magnitude of the random perturbations decreasing slightly from generation-to-generation (similar to simulated annealing).

Not all of these 50-generation runs yielded good-fitting parameter sets. To resolve this, we stored the best-fitting parameter set found to-date, and after each 50-generation run (described above), the best-scoring parameter set was compared with the saved best parameter set to-date. If the current run yielded a fit better than the previous best, it superseded the previous best fit. At the end of the 50-generation run, the parameter values were re-initialized with the best-fitting parameters found to-date, and the process was repeated.

In this way, we could exploit a large number of random initial conditions, and the selection procedure, to find good-fitting parameter sets. We are not guaranteed that this procedure will find the minimum-score parameter set. However, in practice, this procedure did find parameter sets that reproduced our experimental data with high fidelity (Fig. 6).

The best-fitting parameter set, used for the model studies in the paper, has a score of 1.26. This reflects the fact that, for all of the statistics to which the model was fit, the model is within one S.E.M. of the experimental mean (Fig. 6). To make Figs. 6AB, which showcase the high quality of this fit, we generated 25 new models from the best-fitting parameter set, and show the statistics of these models in Figs. 6AB. This is important because otherwise, the comparison would be subject to selection bias.

Single-cell statistics to which the model was fit

The single-cell statistics to which the model was fit are as follows. First, we wanted the widths of the model tuning curves to match the experimental ones. This was achieved by matching the "normalized TC integral" parameter shown in Fig. 6A of the main paper. This parameter is the integral (over all stimuli) of the

TC divided by its maximum: wider tuning curves have larger normalized integrals. In particular, a TC that is constant for all stimuli has an integral of 2π, which is the range over which the integration takes place, whereas an infinitesimally narrow (δ function) TC has a normalized integral of 0. Finally, a tuning curve of the form $A(1 + \sin(\theta))$, for some amplitude A, has a width of π.

We also wanted the model to match the amplitudes of the neural responses, and so the model was selected to match both the mean responses (averaged over all stimuli, trials, and cells) and the population mean of the range of responses (TC peak minus minimum value of the TC).

Finally, the model parameters were chosen to match the observed level of trial-to-trial variability: the mean Fano factor (which is the variance of the responses to a given stimulus divided by the mean response to that stimulus); the mean (over all stimuli and cells) response variance; and the population mean variance range were all used in fitting the model parameters.

Pairwise statistics to which the model was fit

Our first goal was to match the observed noise correlations. Recall that there are 3 distinct types of cell pairs, identified by the offset in their tuning curve peaks, which can be either 0^o, 90^o, or 180^o. For each pair type, we computed the average noise correlation coefficient (averaged over all stimuli and pairs). We further computed, for each cell pair, the range of correlation coefficients (maximum minus minimum correlation coefficient for each pair, as the stimulus was varied), and averaged these over all pairs of the same type. These yield the first 6 of the pair-wise statistics shown in Fig. 6B.

Next, to ensure that the tuning curves are appropriately positioned with respect to each other, we computed the vector dot products between the tuning curves of all observed cell pairs, and divided those by the product of the cells' TC peaks. This gives a measure of how much the tuning curves overlap, and is influenced by, among other things, the sharpness of the tuning curves, and the offsets between the TC peaks. We then averaged these values over all cell pairs of the same type, yielding our final 3 response statistics.

Impressively, the model can fit all 15 of the response statistics to within 1 S.E.M. (Figs. 6AB). Admittedly, with 13 free parameters, the model has significant flexibility, but it is nevertheless not assured that a model with arbitrary structure could get all of these statistics right. As a result, we consider this good fit to confirm the appropriateness of our model's structure.

Further tests of the fitted model

As an independent test, we generated 250 model cell pairs, with the distribution of pair types matching the experimental data (36% 0^o pairs, 36% 90^o pairs, and 29% 180^o pairs; these numbers add up to 101% because they were rounded to the nearest integer). For each pair, we then generated 25 responses to each of 8 different stimuli.

For each stimulus, we computed the geometric mean of the two cells' mean responses, and the noise correlation coefficient. The results are shown, along with the experimental data, in Fig. 6D. To quantitatively compare the model to the data, we first performed a 2-dimensional KS test, using code written by Dylan Muir, based on work by Peacock (1983). This test computes the maximum deviation between the two cumulative distributions (in this case, the 2-D cumulative distributions), which can then be used to estimate the p-value. This test showed no significant difference between the distributions of the model and the experimental data points. We further showed the mean correlation coefficient in different bins of geometric mean response. The bins widths were chosen so as to have roughly 10 data points per bin.

Computing the stimulus coding performance: Fisher Information

For the model coding studies, we generated responses by our model populations (with either 4 or 8 cells) to 500 different stimuli, uniformly spaced between 0 and 2π. For the data in the main paper, we then estimated the mean responses and the covariance of the population in response to each stimulus. Because only these (one- and two-point) statistics were fitted to experimental data, we wished to remain agnostic to any higher-order structure in the responses, which was not directly fitted to experimental data. Fortunately, our linear Fisher

information measure (described in the Methods section) makes use of only the first- and second-order statistics, and thus is naturally agnostic to any higher-order structure.

Trial-shuffling and other perturbations to the response statistics

To compute the information content that the neural population activities *would* yield, if the tuning curves and trial-to-trial variances were the same, but the noise correlations were removed, we set all off-diagonal elements of the covariance matrix to zero, and then re-computed the linear Fisher information.

To generate responses with correlations that were stimulus-independent and matched to the mean over all stimuli, we first averaged the correlation coefficient for each cell pair over all stimuli. We then computed, for each stimulus, the covariance matrix with the appropriate variances on the diagonals, with the off-diagonal elements being $C_{ij}(\theta)= \varrho_{ij,mean}\, \sigma_i(\theta)\, \sigma_j(\theta)$. Finally, we used this covariance matrix in our Fisher information calculations.

Computing the stimulus coding performance: MLE and OLE Error (Fig. S8)

For the data in the main paper, we then estimated the mean responses and the covariance of the population in response to each stimulus. Because only these (one- and two-point) statistics were fitted to experimental data, we wished to remain agnostic to any higher-order structure in the responses, which was not directly fitted to experimental data. To do this, we drew responses from the least structured distribution that is consistent with the computed means and covariances -- a multivariate Gaussian. We also repeated our coding studies without this "Gaussianizing" step and obtained very similar results (not shown) -- our results are not very sensitive to this choice.

Optimal Linear Estimator

Following Salinas and Abbott (1994), we computed the optimal set of weights A that minimized the deviation between the estimated stimulus vector $\mathbf{s}_{est} = A\mathbf{r}$ and the true stimulus vector $\mathbf{s}$. Here, $\mathbf{r}$ is the vector of neural responses. The estimation of the weights is a standard regression problem and can be solved with the use of the Moore-Penrose pseudoinverse.

We used 1000 responses drawn from Gaussian distributions with the appropriate means and covariances to estimate the weights A. We then generated an additional 1000 responses, and estimated the corresponding stimulus vectors. It is important that we tested the model on these new responses, which were not used in training the estimator. Otherwise, we might overestimate the coding performance, due to overfitting of the matrix A.

Finally, we computed the angle between the estimated stimulus vector and the true stimulus vector. The results reported in Fig. S8 are the precision of the estimator, which is the inverse of the mean squared angle between the estimated stimulus vector and the true stimulus vector. Higher precision corresponds to better coding performance.

The coding performance was computed for 10 different models, each randomly generated from the same parameter set, and the reported results are averaged over these 10 models.

Maximum Likelihood Estimator

As in the OLE case (above), we generated 1000 responses to 500 different stimuli, from the Gaussian distributions. We then estimated the likelihood of each response, assuming the (Gaussian -- with the appropriate mean and covariance) conditional distribution from each of the 500 stimuli. The estimated stimulus for each response was the one that maximized the response likelihood.

We then computed the angle between the true stimulus vector and the estimated one, and computed the error (mean squared angular deviation), and the precision (inverse error) for each model. Finally, as with the OLE, we averaged these results over 10 randomly chosen models with the same underlying parameter values.

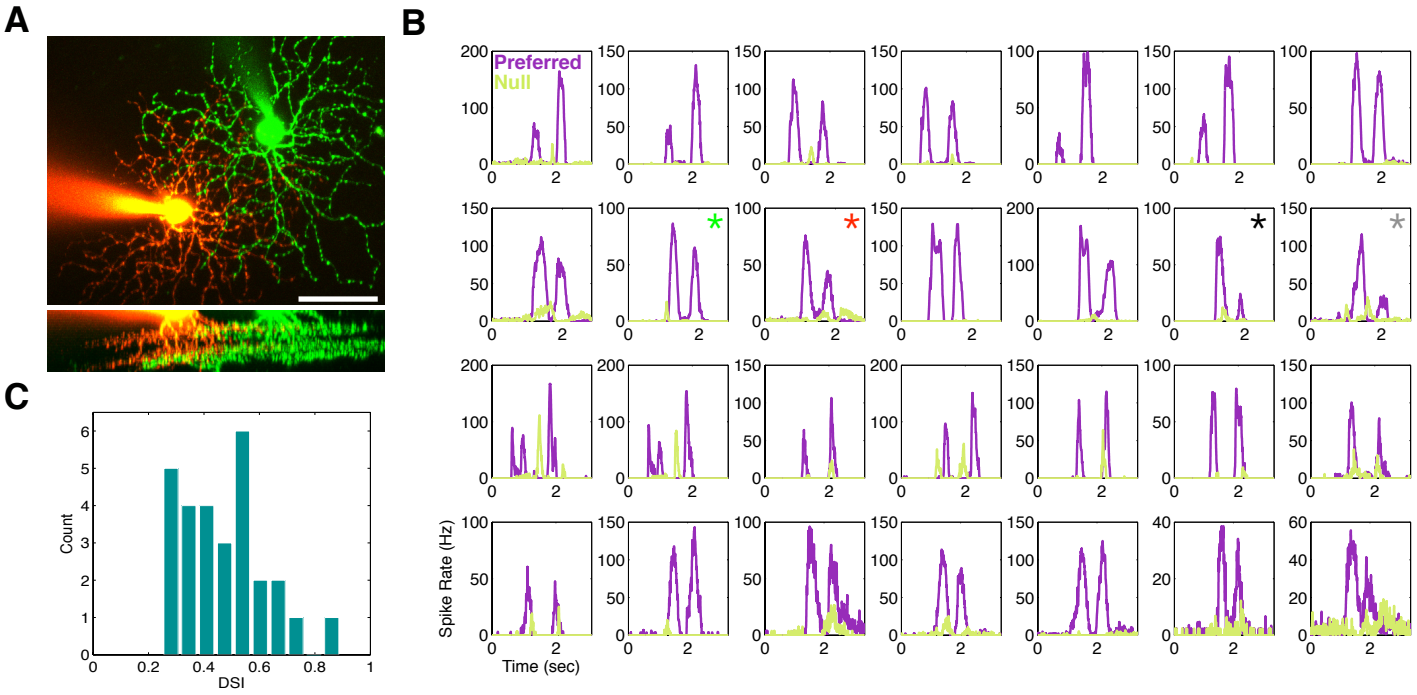

Figure S1. Identification of ON-OFF dsRGCs - related to Figure 1. (**A**) Confocal image of a pair of ON-OFF Direction Selective RGCs that were filled with dye using a patch pipette and imaged after a paired loose-patch spike recording. Top: maximum intensity projection image showing dendritic overlap. Bottom: cross-section maximum intensity projection showing characteristic bistratified dendritic arbors. (**B**) For each cell in the cell-attached dataset, average PSTH traces show ON-OFF spike responses. Responses are shown to bars moving in the preferred (purple) and null (yellow) directions. Panels marked with an (*) indicate example cells: red and green correspond to the dye-filled pair in (A), which is the 0° pair in Figure 1d; black and grey correspond to the example cells from the 180° pair in Figure 1B-E. (C) Histogram of vector-sum DSI values (see methods) for all 28 cells in the cell-attached dataset.

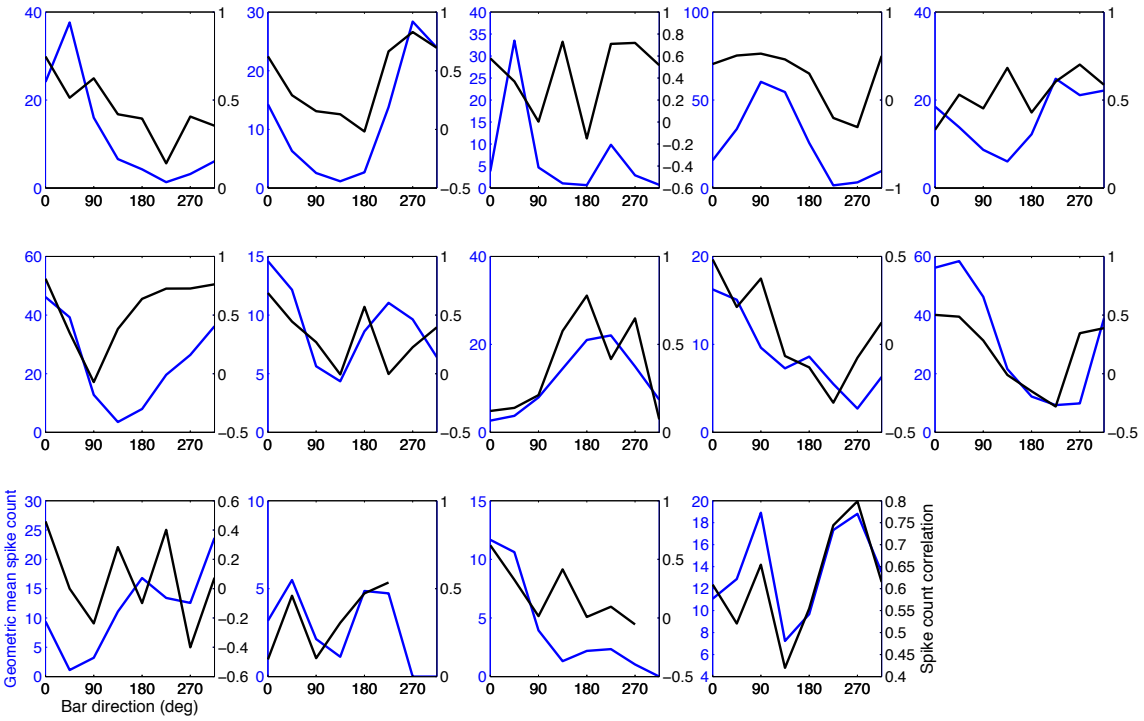

Figure S2. Direction dependence of responses and correlations for all recorded cell pairs - related to Figure 1. Each panel corresponds to one cell-attached recorded ooDS cell pair. Blue traces are the geometric mean of the mean spike counts and black traces are the spike count noise correlation coefficients, each as a function of bar angle (as in Fig. 1E). Panels are organized by relative preferred direction. Top row shows pairs (n = 5 pairs total) with roughly the same direction tuning ("0° pairs"); middle row shows 90° pairs (n = 5), and bottom row shows 180° pairs (n = 4).

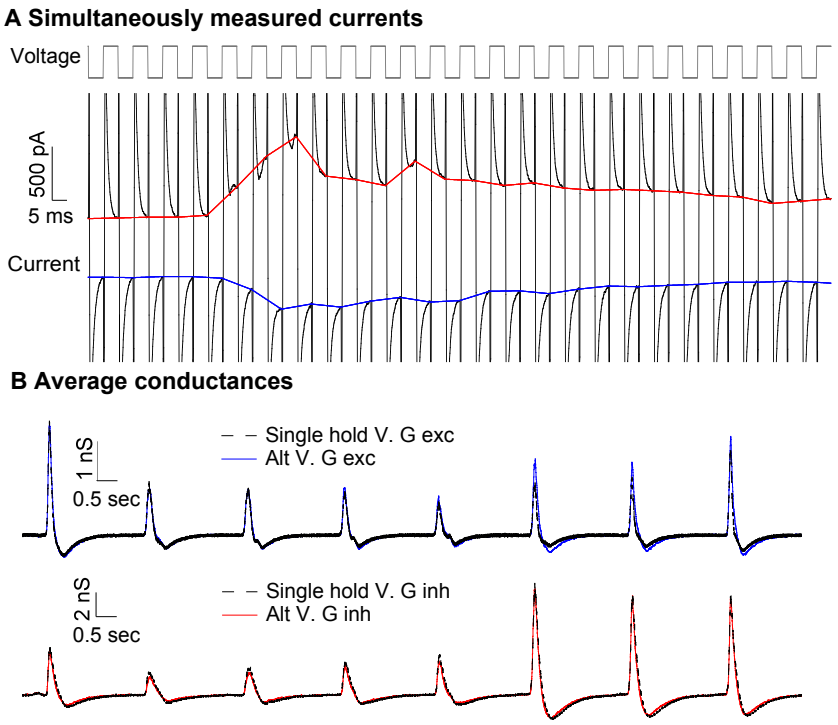

Figure S3. Alternating voltage-clamp conductance measure - related to Figures 3 & 4. (A) The voltage command (top) was alternated between the excitatory and inhibitory reversal potentials while the current (gray line) was measured. Interpolating between the last current values measured on each cycle (colored lines) provides an estimate of the neurons excitatory and inhibitory currents. (B) The average conductances calculated using this procedure were very similar to the average conductances measured using a single holding potential at both the excitatory and inhibitory reversal potentials.

<u>A simple non-linear model explains the relationship between (co)variance and mean conductance shown in Figs. 3-4</u> *(See Figures 3 & 4)*

Consider a (direction-dependent) signal s, together with (direction-independent) noise n1 arriving at a nonlinearity N, as shown in the left panel below. The nonlinearity N will reshape the incoming mean and variance, as illustrated in the middle panel. Then, a (direction-independent) noise is added, so that the final output is N(s+n1) + n2.

If the nonlinearity is an exponential $N(x) = c \exp(k x)$, then this model can produce a linear relationship between conductance variance and (squared) mean. This is illustrated in the right panel below, where parameters are chosen to roughly reproduce the excitatory variance data in Fig. 3 of the main paper.

To see why this works, we first assume that the variance v1 of n1 is sufficiently small so that N(s+n1) can be approximated by its linearization around the mean input s [N.B. this appears to be the case for our measured excitatory data at least -- fits to that were used in figure below]. This implies that the mean of N(s +n1) will be N(s). Moreover, the variance will be scaled by the slope squared, so that it is $(N'(s)^2)v1 = (k\ N(s))^2$ v1. This shows that there is a linear relationship between the between conductance variance and (squared) mean -- as in the data (Fig. 3). The slope of this relationship is k^2v1. The post-nonlinearity noise always adds a constant variance (v2) to this -- this is the intercept of the relationship between conductance variance and (squared) mean.

Moreover, the choice $N(x) = c \exp(k x)$, is the only nonlinearity that will produce a linear relationship between conductance variance and (squared) mean, within the constraints of our model.

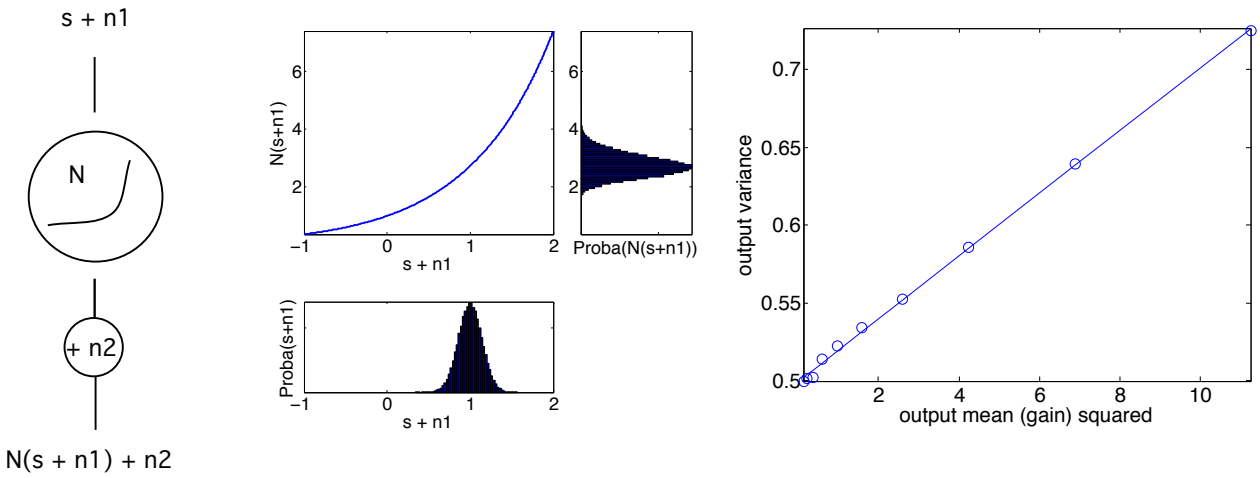

Figure S4. Simple non-linear model for the observed scaling of variance with mean conductance - related to Figures 3 & 4.

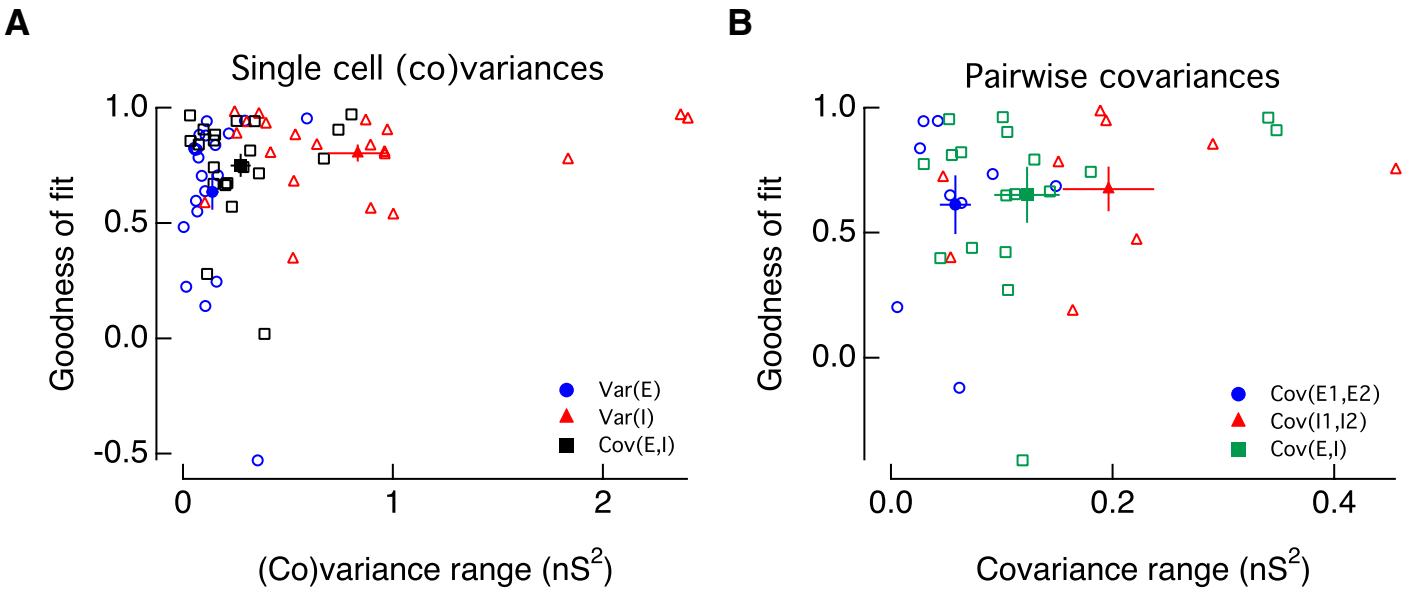

Figure S5. Gain model goodness of fit as a function of (co)variance range - related to Figures 3 & 4. The linear correlation between measured conductance gain and (co)variance is shown as a function of the (co)variance range for each single cell (A) and pair (B) in the alternating voltage clamp experiments. The linear correlation is stronger for conductances that are more strongly modulated by the stimulus. Error bars represent the S.E.M.

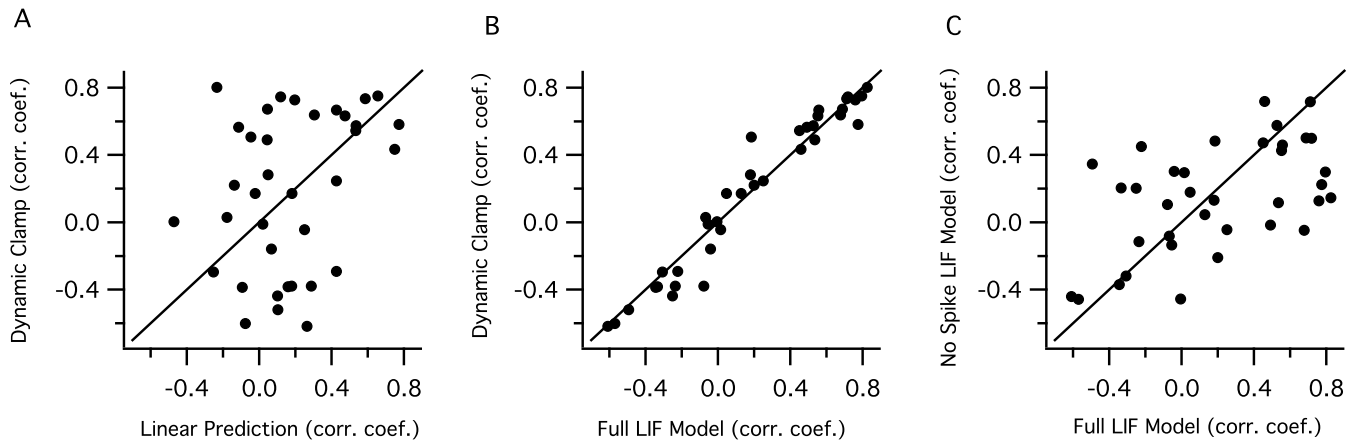

Figure S6. Impact of spike generation on spike count correlations - related to Figure 5. (A) The linear prediction (Eq. 6) failed to predict the dynamic clamp output spike count correlations. (B) A simple LIF model was highly predictive of the dynamic clamp output spike count correlations. (C) Preventing spike generation in the LIF model caused both increases and decreases in the spike count correlation compared with the spiking model.

Predicting output correlations using a linear integration of correlated input *(See Figure 5)*

We calculated a linear prediction of the spike number correlation coefficient given the variance and covariance of the synaptic conductances in a pair of ooDS cells. The linear prediction assumes that the spike number generated by a cell is proportional to the average net inward synaptic current the cell receives. Assuming that changes in driving force do not substantially change synaptic currents compared to changes in the conductances then the spike output in cell 1, s_1, can be written as

$$s_1 = k_E E_1 - k_I I_1 \tag{5}$$

where E_1 and I_1 are the measured conductances averaged over time and k_E and k_I are the driving forces on those conductances.

Under these assumptions, the spike number correlation coefficient can be written as

$$\begin{aligned}\rho &= \frac{\sigma^2_{s_1,s_2}}{\sqrt{\sigma^2_{s_1} \times \sigma^2_{s_2}}} \\ &= \frac{\sigma^2_{E_1,E_2} \; + \; \alpha^2\sigma^2_{I_1,I_2} \; - \; \alpha\sigma^2_{E_1,I_2} \; - \; \alpha\sigma^2_{E_2,I_1}}{\sqrt{(\sigma^2_{E_1} + \alpha^2\sigma^2_{I_1} - \alpha\sigma^2_{E_1,I_1}) \times (\sigma^2_{E_2} + \alpha^2\sigma^2_{I_2} - \alpha\sigma^2_{E_2,I_2})}}, \\ \alpha &= k_I/k_E\end{aligned} \tag{6}$$

where σ^2_{xy} is the covariance of x and y and σ^2_x is the variance of x.

From equation (6) above, we can see that the covariance between the excitatory and inhibitory synaptic input received by two cells reduces their predicted spike covariance while the excitatory-excitatory and inhibitory-inhibitory covariances increase the predicted spike covariance.

We tested the ability of equation (6) to predict spike correlation coefficients from dynamic clamp data. Although the linear prediction accounts for the broad trends (i.e. decorrelating effect of EI and correlating effect of EE, II) it did not accurately predict the spike correlation coefficients (Fig. S6A). However, the spike correlation coefficients generated in the dynamic clamp experiments could be reproduced using the same conductances as inputs in a Linear Integrate and Fire (LIF) model (Fig. S6B), suggesting a relatively simple current-to-spike transformation in the neurons and providing a model that we could manipulate to better understand the failure of the linear prediction. Correlation coefficients in a version of the LIF model in which spike generation was disabled were both higher and lower than those of the full LIF model (Fig. S6C). These results indicate that spike generation can both correlate and decorrelate the recorded conductances and suggest that the failure of the linear prediction is caused by the nonlinear effects of spike generation.

Attempts to directly assemble 8-cell populations from the recorded data

Recall that the challenge is to take pairwise measurements of ooDS cell responses, and assemble 8-cell populations for which the covariance matrices (for all stimuli) are logically consistent -- being positive semi-definite. This is made difficult by the large correlation coefficients for some pairs, in response to some stimuli.

To attempt this, we took sets of 8 randomly-chosen tuning curves from our cell-attached recordings, and shifted their centers so that two of the cells' TC peaks were at each of the four cardinal directions. For each cell, we kept the measured variances associated with each stimulus (shifted in the same way as the tuning curves, so the mean-variance relationship was preserved), allowing us to fill in the diagonal elements of the stimulus-dependent covariance matrices. To fill in the off-diagonal elements of the covariance matrices, we needed to choose correlation coefficients that would match our measured rate-correlation relationship (Fig. 1F), and lead to logically consistent, positive semi-definite (PSD) response covariance matrices for all stimuli.

For each stimulus, we computed the geometric mean response of each cell pair (horizontal axis in Fig. 1F), and then randomly selected correlation coefficients from the measured responses that were within 0.1 of the given geometric mean response. Doing this for all pairs, we filled in the off-diagonal entries of the covariance matrices. We then checked that the covariance matrices so-generated were PSD for all stimuli. If they were not, the putative model was rejected, 8 more randomly-chosen tuning curves were selected, and the process was repeated. After repeating this procedure 10^8 times, no logically consistent models were obtained.

This is a testament to how tightly constrained pairwise responses can be when they need to fit into larger neural populations, and highlights the importance of the mechanistic model that we used to construct logically consistent 8 cell populations.

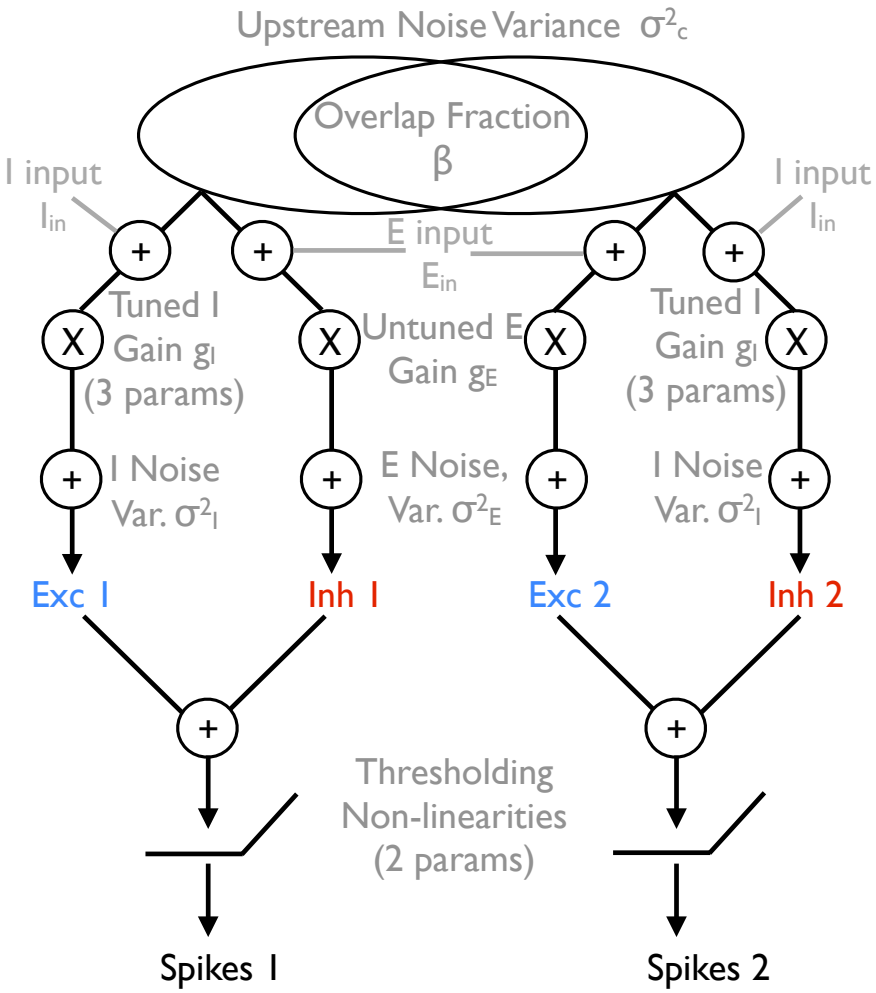

Fig. S7: Diagrammatic description of the computational model - related to Figure 6. The model follows the architecture (Fig. 4A of the main paper) revealed by our alternating voltage clamp experiments, and is shown here schematically for 2 cells, although the "full" model can contain any number of cells. Each cell pair received correlated input noise, with the correlation dictated by parameter β (which varies from pair-to-pair). This noise, and (constant for all stimuli) excitatory and inhibitory signals are then added, and multiplied by the stimulus-dependent gain factors. As such, the stim.-dependent gain factors are proportional to the mean inputs of each type for each stimulus. The E gain is untuned, and the I gain is parameterized by 3 parameters that describe the amplitude of the tuning, the (constant) offset in the gain, and the sharpness of the tuning (see Supplemental Eqs. 3,4). After the multiplicative gain stage, independent noise is added to all channels. Finally, excitation and inhibition are combined, to yield the net inputs, which are 3E - I. The factor of 3 comes in because the E or I inputs describe conductances, and excitation typically has a larger driving force, as its reversal potential is further from the resting potential than is the inhibitory reversal potential. The inputs are then passed through thresholding nonlinearities, parameterized by the threshold value and the slope of the linear portion of the curve. The parameters can vary from cell-to-cell in the model.

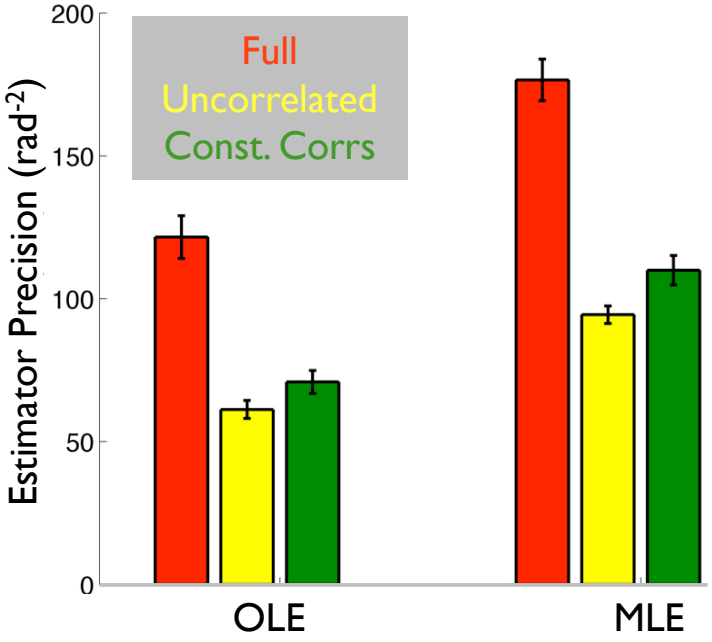

Fig. S8: Coding performance of the 8-cell model using measures other than Fisher information — related to Fig. 6E. Similar to Fig. 6E ,we simulated responses of our 8-ooDS-cell model to 500 different directions of stimulus motion. We then attempted to decode those responses to estimate the stimulus, and quantified the precision (inverse of mean-squared error) of the estimator. We used both the optimal linear estimator (OLE; Salinas and Abbott, 1994), and the maximum likelihood estimator (MLE). We repeated this calculation for the responses with stimulus-dependent correlations generated by our model ("Full"; red), for uncorrelated (trial-shuffled) responses (yellow), and for responses with stimulus-independent correlations that matched, pair-by-pair, the average of the stimulus-dependent correlations ("const corrs; green). In all cases, the estimators were optimized for the given data type (for example, for the MLE, we used the joint distribution of population responses from the 8-cell model), and the performance was computed on held-out neural responses, in a cross-validation procedure. Error bars are the S.E.M. over 10 randomly-drawn model populations.

Stimulus-dependent correlations increase variability in the radial direction in the neural response space

We assume that the cells have mean responses to each stimulus given by their tuning curves (**f**(θ) is the vector of such mean responses over the population, for a given stimulus s), with Poisson-like variability (trial-to-trial variance equal to mean response), and that that (stimulus-dependent) correlation coefficient between cell pairs is given by

$$\rho_{ij}(\theta) = a_{ij}\sqrt{f_i(\theta)f_j(\theta)},$$
$$a_{ij} = \rho_{max}\left[\sqrt{max_\theta(f_i(\theta))max_\theta(f_j(\theta))}\right]^{-1} \quad (7)$$

where i and j are the cells' indices. In this case, the covariance for the cell pairs is given by (correlation times geometric mean of variances)

$$\Sigma_{i\sim j} = a_{ij}\, f_i(\theta)\, f_j(\theta)$$
$$\Sigma_{ii} = f_i(\theta) \quad (8)$$

The second line follows from the fact that the diagonal elements of the covariance matrix are the cells' variances, and that the cells have Poisson-like variability (variance equal to mean).

Assuming the same a_{ij} for all pairs, we write this as the matrix

$$\Sigma(\theta) = a\vec{f}(\theta)\vec{f}^{T}(\theta) + \mathrm{diag}\left[\vec{f}(\theta) - a\vec{f}^{2}(\theta)\right] \tag{9}$$

where diag[.] denotes the diagonal part of the argument, and the vector $\mathbf{f}^2(\theta)$ is the element-wise square of the vector $\mathbf{f}(\theta)$. We note that the parameter "a" in the covariance matrix controls the strength of of the stimulus-dependent correlations; larger "a" corresponds to stronger stimulus-dependent correlations.

From this covariance matrix, we compute its variance along the "radial" direction (along the vector $\mathbf{f}(\theta)$ given by the mean responses) in the neural response space:

$$\mathrm{RV} = [\mathbf{f}(\theta)\, \Sigma(\theta)\, \mathbf{f}^{T}(\theta)]/\|\mathbf{f}(\theta)\|^2 , \tag{10}$$

and the total variance in the neural population (trace of the covariance matrix) is given by

$$\mathrm{TV} = \Sigma_i f_i(\theta). \tag{11}$$

Thus the fraction of variance along the radial direction is (RV/TV) given by

$$\begin{aligned} \mathrm{RV/TV} &= [\mathbf{f}(\theta)\, \Sigma(\theta)\, \mathbf{f}^{T}(\theta)]/\, [\, \|\mathbf{f}(\theta)\|^2 \Sigma_i f_i(\theta)] \\ &= \{\Sigma_{ij} f_i(\theta) f_j(\theta)\, [a\, f_i(\theta) f_j(\theta) + \delta_{ij}(f_i(\theta) - a\, f^2{}_i(\theta))]\} \,/\, \{\Sigma_i f_i(\theta)\, \Sigma_j f^2{}_j(\theta)\} \end{aligned} \tag{12}$$

using the $\Sigma(\theta)$ given above (Eq. 9). Taking derivatives of (RV/TV) with respect to the parameter a, we find that

$$\begin{aligned} \mathrm{d[RV/TV]/da} &= \{\Sigma_{ij} f_i(\theta) f_j(\theta)\, [f_i(\theta) f_j(\theta) - \delta_{ij}\, f^2{}_i(\theta))]\} \,/\, \{\Sigma_i f_i(\theta)\, \Sigma_j f^2{}_j(\theta)\} \\ &= \{\Sigma_{ij}\, f^2{}_i(\theta) f^2{}_j(\theta) - \Sigma_i f^4{}_i(\theta)\} \,/\, \{\Sigma_i f_i(\theta)\, \Sigma_j f^2{}_j(\theta)\} \\ &> 0, \end{aligned} \tag{13}$$

where the last line follows because the second sum ($\Sigma_i f^4{}_i(\theta)$) is a subset of the first, and all elements of the sums are non-negative (we further assume that at least one pair of neural responses is non-zero (having $f_i(\theta) f_j(\theta) > 0$) so that the inequality is a strict inequality). This analysis shows that, as the strength "a" of the stimulus-dependent correlations increases, so too does the fraction of the population-wide trial-to-trial variability that is along the "radial" direction in the space of neural responses. Consequently, whenever the "signal" (stimulus-induced change in population activity) is orthogonal to this radial direction, the stimulus-dependent correlations should improve the neural population code (relative to the case of a=0, in which there are no correlations). This is shown empirically in Fig. 8 of the main paper.

<u>The stimulus-dependent gain mechanism naturally orients noise in the radial direction</u>

The stimulus-dependent gain model that accounts for the correlation in the ooDS cells' synaptic inputs (Fig. 4A) naturally leads to a significant amount of trial-to-trial variability along the radial direction in the neural response space. To see this, let input to each cell be s_j. Trial-by-trial, this input will be composed of both a mean (m_j), and some noise n_j,

$$s_j = m_j + n_j \tag{14}$$

Following the circuit architecture in Fig. 4A, we have the following equation for the mean input,

$$m_j = k_E\, g_{E,j} - k_I\, g_{I,j}, \tag{15}$$

where k_E and k_I are the driving forces pushing current through the E and I channels, respectively. Note (in Fig. 4A) that the noise n_j is a sum of independent and common parts (n_c),

$$n_j = n_{ind,j} + n_{c,j} \tag{16}$$

where the common part is common to all 8 cells, and is multiplied by the relevant stimulus-dependent gain factors, i.e.

$$n_j = n_{ind,j} + n_c\,(k_E\, g_{E,j} - k_I\, g_{I,j}) \tag{17}$$

Comparing equations (15) & (17), we see that trial by trial, $n_{c,j}$ is proportional to m_j, for each cell. Thus, on every trial, the common component of the noise points, in the space of neural responses, in exactly the same direction as the mean. In other words, the common noise is oriented along the radial axis, by virtue of the circuit architecture. The same is not true of other circuit architectures in which, unlike the one we uncovered from our experiments (Fig. 4A), the common noise is not effectively multiplied by the mean response.

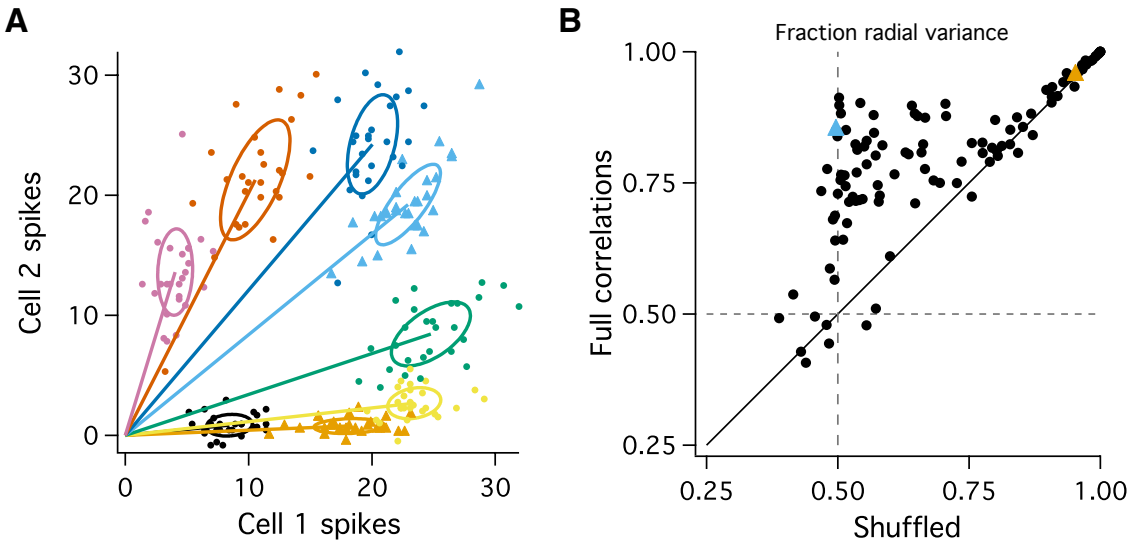

Fig. S9: Variability in ooDS spike responses lies along the radial direction in response space - related to Fig. 8. (A) Spike data from an example ooDS cell pair showing the conditional response distributions to 8 different bar directions, each indicated by a different color. The ellipses are the 1 standard deviation contours of these distributions, and the solid lines denote the mean response vector corresponding to each stimulus. The pairwise correlations tend to orient response variability in the direction of the mean response vector. (B) We computed the variability for each conditional distribution along the radial direction (aligned with the mean response vector) and divided it by the total variance in that distribution for the original data and for trial-shuffled data. The radial variance, v_r, was computed as $v_r = u^T \Sigma u$, where u is the mean response vector, normalized to be unit length (u^T denotes its transpose) and Σ is the covariance matrix of the conditional response distribution. Geometrically, we are computing the variance of the distribution projected along the mean response vector. The full pairwise correlated data has a greater fraction of its variance oriented radially than shuffled data ($p = 3x10^{-16}$; for 14 recorded pairs, 8 stimuli each). Colored triangles denote the corresponding conditional response distributions in (A).

Assembling model-neuronal populations with a range of average-cosine values (Fig. 8)

Model-neuronal populations for which the tuning curve positions are randomly (and uniformly) generated tend to have low average-cosine values. This is because the overall magnitude of the population response is relatively constant as the stimulus changes (for all stimuli, *some* of the cells are firing strongly), and thus the signal lies near a spherical shell.

To generate Fig. 8, however, we wanted model populations covering a range of average-cosine values, to explore the relationship between average-cosine value and the impact of stimulus-dependent correlations on the population code. To generate the populations with low average-cosine values, we chose the tuning curves to be clustered together within some (small) range of angles. In other words, for the "fully" random populations, the tuning curve peaks were uniformly drawn over the range $[0, 2\pi]$, whereas for these "clustered" populations, the tuning curve peaks were uniformly drawn over the range $[\pi - \Delta/2, \pi + \Delta/2]$, where the parameter Δ determined the degree of clustering. The clustered (vs "fully" random) populations tend to have larger average-cosine values because the signal correlations are typically larger in the clustered populations. To generate the data in Fig. 8, we generated models with a range of Δ values, and thus with a range of different average-cosine values.

Impact of correlations on stimulus coding for 2-cell populations

We used a Gaussian Maximum Likelihood Estimator to guess which of the 8 stimuli was responsible for each response by our ooDS cell pairs. We then computed the difference between the true stimulus and this MLE estimate. The precision (inverse variance; higher precision reflects better coding performance) is shown below (Fig. S10) for both the full (correlated) data and the trial-shuffled (uncorrelated by otherwise identical to the full data) surrogate. Overall, correlations slightly improve direction coding ($p = 0.01$, paired t-test). The 90° pairs have the most stimulus coding power, and are the most affected by noise correlations. Note that we could not use linear Fisher information for this investigation, like we did with the 8-cell and larger models in the paper (Figs. 6,7). This is because the Fisher information calculation requires derivatives of the tuning curves with respect to the stimulus. Our experimentally-measured tuning curves were sampled at only 8 stimulus locations, and thus we did not have enough resolution to compute their derivatives with respect to the stimulus.

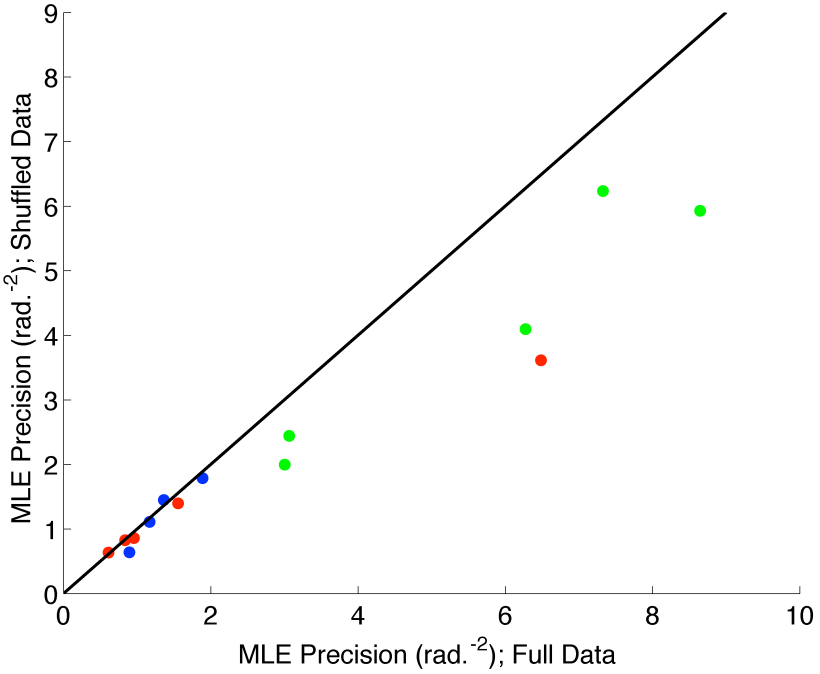

Fig. S10: Impact of correlations on direction coding by ooDS cell pairs - related to Discussion. . Precision (inverse variance) of the MLE estimates of stimulus direction from the measures ooDS cell pair responses. Different colors indicate pairs with different spacings between their tuning curve peaks: either 0° (red), 90° (green), or 180° (blue). Correlations slightly improve direction coding (points tend to fall below unity line; $p = 0.01$, paired t-test).

Supplemental references

Park, S.J.H., Kim, I.-J., Looger, L.L., Demb, J.B., and Borghuis, B.G. (2014). Excitatory Synaptic Inputs to Mouse On-Off Direction-Selective Retinal Ganglion Cells Lack Direction Tuning. *J. Neurosci.* 34, 3976–3981.

Poleg-Polsky, A., and Diamond, J.S. (2011). Imperfect space clamp permits electrotonic interactions between inhibitory and excitatory synaptic conductances, distorting voltage clamp recordings. *PLoS One* 6, e19463.

Yonehara, K., Farrow, K., Ghanem, A., Hillier, D., Balint, K., Teixeira, M., Jüttner, J., Noda, M., Neve, R., Conzelmann, K.K., and Roska, B. (2013). The first stage of cardinal direction selectivity is localized to the dendrites of retinal ganglion cells. *Neuron* 79, 1078–1085.