METHODS

Recording. Multi-electrode extracellular recordings were obtained *in vitro* from a segment of isolated, peripheral macaque monkey retina, using preparation and recording methods described previously[15,26]. Analysis was restricted to two physiologically defined classes of cells; on the basis of light response properties and density, these were identified as ON and OFF parasol cells[27]. The cells shown were recorded in a square region of retina covered by 76 electrodes. A standard clustering-based spike-sorting procedure (see refs 15, 26) was used to estimate the number of units, and least-squares regression of the estimated spike times against multi-electrode voltage signal was used to estimate multi-electrode spike waveforms for each unit. Although this approach correctly and efficiently identifies isolated spikes, when two cells fire within a 1–2-ms window, the clustering approach can fail to identify the presence of both spikes. We solved this problem by using estimates of the elementary waveforms to detect the superposition of spikes. We performed maximum *a posteriori* estimation under the model that the multi-electrode voltage signal was the linear superposition of Gaussian white noise and the spike trains convolved with their associated spike waveforms, with a sparse (exponential) prior distribution on the spike trains. This corresponds to a tractable quadratic optimization problem under linear inequality constraints, which can be solved efficiently using existing methods. The real-valued solution vector was then binarized by greedily inserting spikes whenever the reduction in mean-squared error between predicted and actual voltage exceeded a threshold[28]. This procedure correctly identified simultaneous spikes in simulated data sets and corrected obvious cross-correlation artefacts appearing in real data sorted with standard clustering techniques.

Stimuli. The retina was stimulated with a photopic, achromatic image of a cathode ray tube display, refreshing at 120 Hz. The stimulus was a spatio-temporal pseudo-random binary sequence, where the intensity of each pixel was drawn independently from one of two values on each frame. The stimulus pixel size was 120×120 μm on the retina, and contrast (standard deviation divided by mean) was 96%.

Fitting. Model parameters were fitted by maximizing likelihood[5] using 7 min of spiking data recorded during presentation of a non-repeating stimulus. The parameters for each cell consisted of a stimulus filter $\mathbf{k}$, a spike-history filter $\mathbf{h}$, a set of incoming coupling filters $\{\mathbf{l}_i\}$ and a constant (specifying the log of the baseline firing rate) μ. The filter $\mathbf{k}$ was a 750-dimensional vector (5×5 spatial pixels $\times$ 30 time bins), parametrized using a lower-dimensional representation as a rank-2 matrix: $k(x, y, \tau) = k_{s,1}(x, y)k_{t,1}(\tau) - k_{s,2}(x, y)k_{t,2}(\tau)$, with $k_{s,i}(x, y)$ denoting a spatial filter (25 parameters) and $k_{t,i}(\tau)$ a temporal filter (10 parameters), giving $2 \times 35 = 70$ parameters. A rank-3 representation did not improve performance. These filters closely resembled a time-varying difference-of-Gaussians[30]; spatial filters were well-approximated (in a least-squares sense) by Gaussians, which were used to plot spatial ellipses shown in Fig. 1 and to summarize receptive field properties (Supplementary Figs 2 and 3). Gaussians fit to receptive field centres and surrounds had average standard deviations of 0.25 pixels and 0.7 pixels (1.0 pixels for the uncoupled model), respectively. Temporal filters $\mathbf{h}$ and $\{\mathbf{l}_i\}$ and the temporal components of $\mathbf{k}$ were represented using a basis of raised cosine 'bumps' of the form $\mathbf{b}_j(t) = (1/2)\cos(a\log[t + c] - \phi_j) + (1/2)$ for t such that $a\log(t + c) \in [\phi_j - \pi, \phi_j + \pi]$ and 0 elsewhere, with constants a and c set by hand to watch the structure observed in auto- and cross-correlation functions, and $\pi/2$ spacing between the ϕ_j (see Supplementary Information). This basis allows for the representation of fine temporal structure near the time of a spike and coarser/smoother dependency at later times (see ref. 22). The $\mathbf{h}$ filter was represented with ten such basis vectors, and the $\mathbf{l}_i$ coupling filters were represented with four. The 'uncoupled model' was fitted independently without coupling filters $\{\mathbf{l}_i\}$, and the inhomogeneous Poisson model (Fig. 4) was fitted without $\{\mathbf{l}_i\}$ or $\mathbf{h}$.

Conditional intensity (spike rate) is given by $\lambda(t) = \exp(\mathbf{k}\cdot\mathbf{x} + \mathbf{h}\cdot\mathbf{y} + \left(\sum_i \mathbf{l}_i\cdot\mathbf{y}_i\right) + \mu)$, where $\mathbf{x}$ is the stimulus, $\mathbf{y}$ the cell's own spike-train history, μ is the cell's baseline log-firing rate, and $\{\mathbf{y}_i\}$ the spike-train histories of other cells at time t. The population log-likelihood is the sum over single-cell log-likelihoods, each given by $L = \sum \log \lambda\left(t_{sp}\right) - \lambda(t)\mathrm{d}t$, where t_{sp} denotes the set of spike times and the integral is taken over the length of the experiment[4,5]. We added a penalty of the form $-\alpha \int \left|\sum_i \mathbf{l}_i(t)^2\right|^{1/2} \mathrm{d}t$ to eliminate unnecessary coupling filters (using a constrained Newton–Raphson algorithm to maximize the penalized log-likelihood), which regularizes and prevents overfitting. The regularization parameter α was selected by means of cross-validation on a novel 5-min data set, but results were robust with respect to both α and the choice of basis. (This reduced the number of coupling filters from 702 to 243 and recovered a roughly pairwise-adjacent structure; see Supplementary Information.)

Correlations. Spike responses of full and uncoupled models were simulated with the same 20-min stimulus (144,000 samples) presented experimentally. Pairwise cross-correlations were computed in 1-ms bins, according to $C(\tau) = [\langle y_1(t)y_2(t + \tau)\rangle - \langle y_1(t)\rangle\langle y_2(t)\rangle]/(\langle y_2(t)\rangle \mathrm{d}t)$, where $y_1(t)$ denotes the spike response of the first neuron in bins of width $\mathrm{d}t$, and $\langle\cdot\rangle$ denotes averaging over t. Triplet correlations were computed in 5-ms bins according to $C(\tau_1, \tau_2) = [\langle y_1(t)y_2(t + \tau_1)\, y_3(t + \tau_2)\rangle - \langle y_1(t)\rangle\langle y_2(t)\rangle\langle y_3(t)\rangle]/(\langle y_2(t)\rangle\langle y_3(t)\rangle \mathrm{d}t)$.

Encoding. Spike-train prediction was validated using the log-likelihood of novel spike trains under both models, computed on 5 min of data not used for fitting or setting α. The difference of log-likelihood under the model and log-likelihood under a homogeneous Poisson process, $\sum \log \bar{\lambda}\left(t_{sp}\right) - \int \bar{\lambda}(t)\mathrm{d}t$ (where $\bar{\lambda} = n_{sp}/T$ is the mean spike rate), divided by n_{sp}, gives prediction accuracy in bits per spike for each cell[25]. Repeat rasters were obtained using 200 presentations of a novel 10-s stimulus, and the time-varying average response (PSTH) was computed in 1-ms bins, smoothed with a Gaussian kernel of width $\sigma = 2$ ms. Conditional rasters were obtained from the coupled model by holding the responses of all but one neuron fixed, and sampling from the model-induced probability distribution on the remaining neuron's response. Samples were obtained by the Metropolis–Hastings algorithm, with spike 'proposals' drawn from a point process model as described in ref. 29. We kept only every 100th output sample of the algorithm to ensure independent samples.

Decoding. We decoded the population response using the Bayes' least-squares estimator, computed under each model (fully coupled, uncoupled with spike-history terms, and inhomogeneous Poisson) using 6,000 different 18-sample single-pixel stimulus segments (validation data that were not used for fitting). Each stimulus $\mathbf{x}_i$ (an 18-dimensional binary vector, given by the time series of light intensities for a centrally located stimulus pixel) was decoded by first extracting $\mathbf{y}_i$, the multi-neuronal spike response portion that was causally influenced by this stimulus. For each model, and for every one of the 2^{18} possible binary $\mathbf{x}_j$, we then computed $p_j = p(\mathbf{y}_i|\mathbf{x}_j)$, the likelihood of the observed population response given that it was generated by stimulus $\mathbf{x}_j$. By Bayes' rule, the posterior is $p(\mathbf{x}_j|\mathbf{y}_i) \propto p(\mathbf{y}_i|\mathbf{x}_j)p(\mathbf{x}_j)$, and the prior $p(\mathbf{x}_j)$ here is constant across binary stimuli. Thus, the posterior is proportional to p_j, and the Bayes' least-squares estimate is given by $\hat{\mathbf{x}}_i = \left(\sum p_j\mathbf{x}_j\right)/\left(\sum p_j\right)$. We also performed decoding on longer (30-sample) stimulus segments, where exhaustive evaluation of these sums is no longer tractable: in this case we used Gibbs sampling from $p(\mathbf{x}_j|\mathbf{y}_i)$ to approximately evaluate the sum. The results obtained using both methods were similar.

Linear decoding was performed using the optimal linear estimator[6], with the same training data as for model fitting. Decoding performance was quantified using the log SNR of each technique: $\log\left(\left|\left\langle \mathbf{x}_i\mathbf{x}_i^T\right\rangle\right| \Big/ \left|\left\langle \mathbf{r}_i\mathbf{r}_j^T\right\rangle\right|\right)$, where $\mathbf{r}_i = \hat{\mathbf{x}}_i - \mathbf{x}_i$ denotes the residual error for decoding stimulus vector $\mathbf{x}_i$, and $|\langle\cdot\rangle|$ denotes averaging over i followed by matrix determinant. Breakdown by temporal frequency was obtained by computing the Fourier power spectra of the stimuli $\tilde{\boldsymbol{x}}_i(\omega)^2$ and residuals $\tilde{\boldsymbol{r}}_i(\omega)^2$, and computing log SNR according to $\log\left(\left\langle \tilde{\boldsymbol{x}}_i(\omega)^2\right\rangle \Big/ \left\langle \tilde{\boldsymbol{r}}(\omega)^2\right\rangle\right)$. Integrating this log SNR across frequency, $(1/2)\int \log \mathrm{SNR}(\omega)\mathrm{d}\omega$, gives a commonly used estimate of the mutual information between the stimulus and the spike-train response[6], which is equivalent to the quantity shown in Fig. 4b.

30. Meister, M. & Berry, M. J. The neural code of the retina. *Neuron* **22**, 435–450 (1999).