nature neuroscience

Decorrelation and efficient coding by retinal ganglion cells

Xaq Pitkow[1] & Markus Meister[2]

An influential theory of visual processing asserts that retinal center-surround receptive fields remove spatial correlations in the visual world, producing ganglion cell spike trains that are less redundant than the corresponding image pixels. For bright, high-contrast images, this decorrelation would enhance coding efficiency in optic nerve fibers of limited capacity. We tested the central prediction of the theory and found that the spike trains of retinal ganglion cells were indeed decorrelated compared with the visual input. However, most of the decorrelation was accomplished not by the receptive fields, but by nonlinear processing in the retina. We found that a steep response threshold enhanced efficient coding by noisy spike trains and that the effect of this nonlinearity was near optimal in both salamander and macaque retina. These results offer an explanation for the sparseness of retinal spike trains and highlight the importance of treating the full nonlinear character of neural codes.

The optic nerve limits how much visual information the eye can transmit to the brain. Early researchers postulated that the retina is designed to use that limited information capacity efficiently, reducing the redundancy in natural scenes by discarding information that the brain has already received from another source in space or time[1,2]. Subsequently, this idea was formalized mathematically[3–8]; images from the natural world have strong, uninformative correlations between the signals carried by different pixels[9]. An efficient encoder could suppress these by spatially filtering the image and thus optimize information transmission. Based on a model of the retina with several simplifying assumptions, one can compute the optimal spatial filter, which resembles the familiar center-surround receptive fields of retinal ganglion cells (RGCs)[5,10]. By computing the difference between the intensity at a point and the average intensity at nearby points, this filter indeed removes spatial correlations in the retinal image, up to some limit determined by photoreceptor noise. This idealized retina model correctly predicts the spatial sensitivity of human vision[6] and several other psychophysical laws[8].

Despite the decorrelation theory's successful predictions, there has been no experimental test of whether neural activity is in fact decorrelated at the putative bottleneck of the optic nerve. One study confirmed that neural firing in the cat's lateral geniculate nucleus is decorrelated in time[11], but there was no test of correlations across space. Another reported both spatial and temporal decorrelation by second-order fly visual neurons[7]. However, the stimuli in this study were still images scanned over the retina, confounding the spatial and temporal contributions to visual processing. A third study found that RGCs oversample visual space, resulting in substantial redundancy[12], but this oversampling may exist either with or without decorrelation relative to the stimulus. Thus, one is still left with these basic questions: does retinal processing indeed decorrelate signals at different spatial locations? If so, does this decorrelation improve coding efficiency?

We inspected spatial and temporal decorrelation in the retina by recording from a population of RGCs while presenting a stimulus with the spatio-temporal correlation structure of natural scenes[9]. We then compared the correlations among RGC spike trains to the correlations between corresponding image locations. To understand how the decorrelation occurs, we separately analyzed the contributions from center-surround receptive fields, noise and sparsifying nonlinearities in the retinal network. We conclude that the dominant effect comes not from the receptive field, but from the nonlinear stimulus-response relationship. These nonlinearities exhibited high response thresholds that led to sparse firing rates. We found that these attributes permitted neurons to transmit information with nearly optimal efficiency.

RESULTS

Our goal was to test whether retinal circuits remove the spatio-temporal correlations present in natural scenes and, if so, to explain whether this helps encode the stimulus efficiently. We measured correlation as a function of distance and time lag in both the visual input and the RGC output. We recorded spike trains from many ganglion cells in the isolated salamander retina under two visual stimuli: naturalistic, which consisted of pseudo-random Gaussian flicker with long-range spatio-temporal correlations such as those of natural scenes (**Fig. 1a** and **Supplementary Fig. 1**), and white noise, which consisted of a flicker stimulus without correlations (**Fig. 1b**). The stimuli were bright in the photopic regime, where the efficient coding theory predicts that decorrelation is the optimal strategy[4,7].

RGCs decorrelate the visual input

The typical ganglion cell responded to such displays with precisely timed bursts of spikes separated by complete silence (**Fig. 1c**).

[1]Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York, USA. [2]Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, Massachusetts, USA. Correspondence should be addressed to M.M. (meister@fas.harvard.edu).

Received 20 September 2011; accepted 13 February 2012; published online 11 March 2012; doi:10.1038/nn.3064

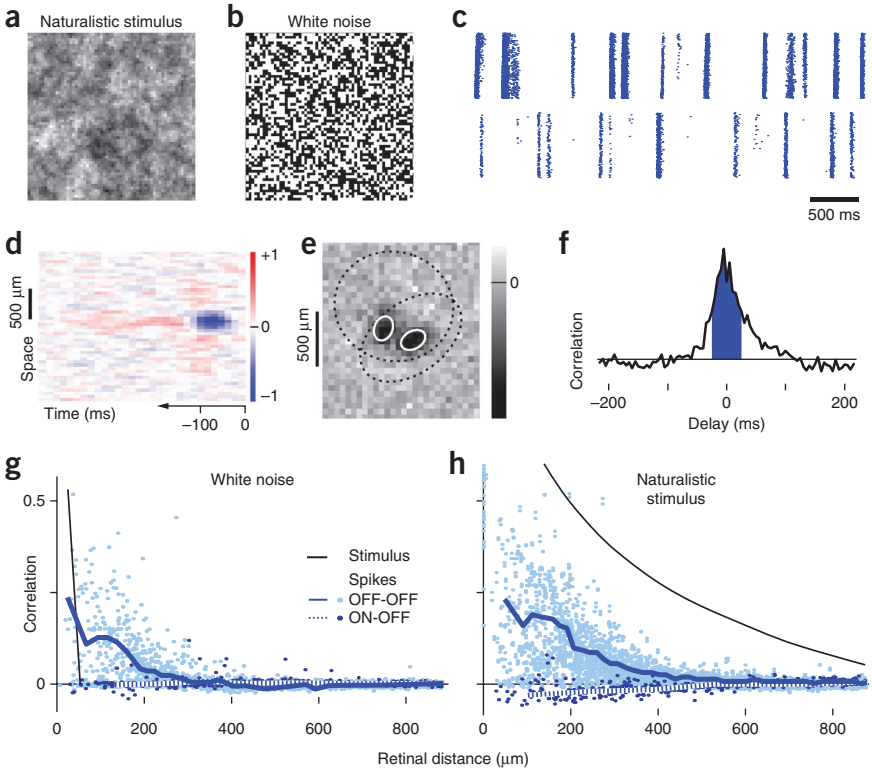

Figure 1 Decorrelation of naturalistic stimuli. (**a**,**b**) Sample frames of naturalistic and white noise stimuli, projected onto a 3.4-mm square on the retina. (**c**) Responses of two RGCs to a short segment of the naturalistic stimulus, displayed as rasters of spikes on 250 identical repeats. (**d**) A sample spatio-temporal receptive field for an OFF ganglion cell, measured as the spike-triggered average stimulus and integrated over one spatial dimension for ease of display. Note the spatial center-surround antagonism (red regions above and below blue) and the biphasic time course (red region left of blue). (**e**) Spatial receptive fields of two OFF cells, including 1-s.d. outlines of the receptive field centers (solid) and surrounds (dotted). (**f**) Cross-correlation function between two ganglion cell spike trains, indicating the frequency of spike pairs as a function of their delay. The shaded area encompasses most of the central peak and indicates the range of delays used to compute the quoted correlation coefficients. (**g**,**h**) Correlation coefficient between the responses of two ganglion cells as a function of their distance under a white noise (**g**) or naturalistic (**h**) stimulus. Each pair of cells contributes a point; lines represent median correlation for pairs at similar distance. Comparisons are restricted within a cell type (solid lines) or across cell types (dashed lines). For reference, the correlation between stimulus pixels is shown (thin lines).

We measured each neuron's spatio-temporal receptive field (**Fig. 1d**,**e**) using the standard reverse-correlation method[13] and then computed the correlation function between the spike trains of any two neurons (**Fig. 1f**). These correlation functions generally showed a central peak ~50 ms wide; this was also the characteristic timescale for variations in the firing rate (**Fig. 1c**). We therefore focused our analysis on the correlations of spike counts in 50-ms time windows (see Online Methods).

We plotted the firing correlation for every pair of ganglion cells against the retinal distance between their receptive field centers (**Fig. 1g**,**h**). This can be compared with the spatial correlations in the stimulus. During white-noise stimulation, the correlation in the firing of ganglion cells greatly exceeded the stimulus correlation out to ~300 µm (**Fig. 1g**). This is because the receptive field centers of nearby RGCs overlap (**Fig. 1e**), and they therefore receive correlated input from their shared photoreceptors. In contrast, under the naturalistic stimulus, neural responses were markedly less correlated than the stimulus pixels (**Fig. 1h**). The ganglion cells exhibited correlations only to ~400 µm distance, whereas the stimulus correlations extended at least twice as far. These observations held for distinct cell types[14,15] that were analyzed separately (data not shown). Thus, the retina decorrelates stimuli with natural image statistics while introducing excess correlation under the unnatural white-noise ensemble. This much is consistent with the classical efficient coding theory.

Decorrelation is primarily achieved by retinal nonlinearities

However, the theory also specifies a decorrelation mechanism, namely RGC receptive fields with antagonistic center and surround regions[2–4,7]. Owing to this antagonism, a RGC fires less to stimuli with low spatial frequency, which drive center and surround equally, and more to those with high spatial frequency[16]. But the low-frequency patterns are precisely those that synchronize nearby neurons. Consequently, a center-surround receptive field should reduce spatial correlations in the retinal output.

To test this, we measured how much decorrelation could be attributed to receptive field filtering. We convolved each spatio-temporal receptive field (**Fig. 1d**) with the naturalistic visual stimulus and analyzed the remaining correlations (**Fig. 2**). Filtering by the receptive field center alone extended the range of correlations, but the addition of the antagonistic surround reduced them below the correlations in the stimulus (**Fig. 2a**), as predicted by the theory, especially at distances beyond one center diameter, ~300 µm (**Fig. 2a**). However, unlike the theoretical prediction, this decorrelation was far from complete. Under the high-contrast stimuli that we used, the optimal linear filters should reduce the correlations to nearly zero[4] for distances greater than the center diameter. Instead, the experimentally measured receptive fields left substantial correlations out to distances twice as great (**Fig. 2a**,**b**), falling far short of the theoretical prediction.

In comparison, the actual decorrelation achieved by the retina was very efficient. The measured correlations between ganglion cell spike trains were suppressed by a factor of ~3 even inside the receptive field center, and by more than tenfold outside (**Fig. 2a**,**b**). Clearly, something other than receptive field filtering is responsible.

Each ganglion cell fired in short stimulus-locked episodes, with some trial-to-trial variation (**Fig. 3**). The correlation between two such spike trains depends on the similarity of their firing events, and therefore on timing, sparseness, and trial-to-trial fluctuations or noise in each firing event (**Fig. 3d**).

The timing of a ganglion cell's firing events is largely determined by its spatio-temporal receptive field, as confirmed by comparing the peaks in the filtered stimulus to those in the firing rate (**Fig. 3b**,**c**). However, measured firing events were narrower than the positive excursions of the linear model (**Fig. 3b**,**c**), presumably resulting from the many documented nonlinearities in the retina's response, including synaptic rectification, depression, gain control, spiking threshold and refractoriness[17–20]. This sparsification means that firing events overlap in time much less than expected from receptive field processing.

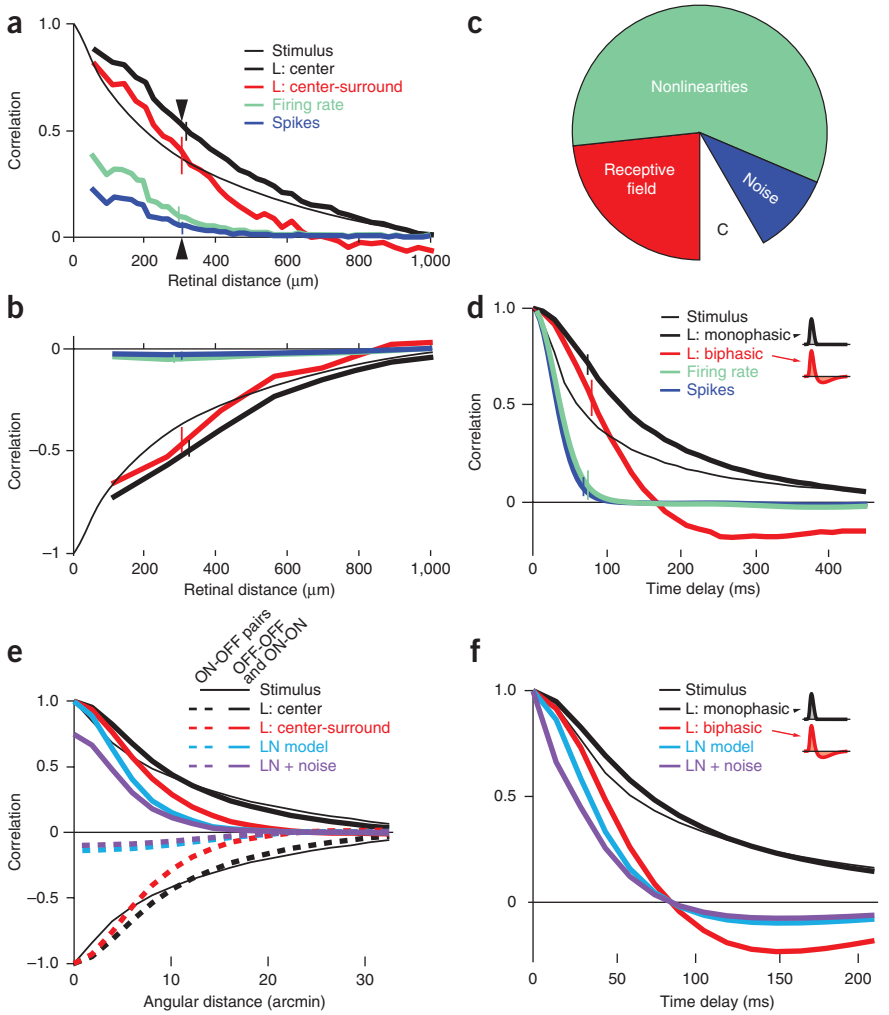

Figure 2 Nonlinearity accounts for much of decorrelation. (**a**,**b**) Spatial correlation functions for neurons and models under naturalistic stimulation. Cells with the same polarity preference (OFF-OFF or ON-ON pairs) have positive correlations (**a**) and those with opposite polarity preferences (OFF-ON pairs) have negative correlations (**b**). Curves are presented as in **Figure 1h** for the stimulus, trial-averaged firing rates, spike trains and linear models. The stimulus correlations are shown with opposite sign for ease of comparison in **b**. Results from many cell pairs are summarized by the median correlation for pairs at similar retinal distance; error bars indicate the central quartiles. L: center and L: center-surround designate linear models using receptive fields including the center component only or both center and surround. (**c**) The origins of decorrelation in different response components. The full circle represents the median correlation present in the stimulus after filtering by the receptive field centers at a retinal distance of 300 μm (arrowheads in **a**). The empty wedge (C) is the much smaller remaining correlation between the ganglion cell spike trains. The red wedge represents the decorrelation caused by lateral inhibition from receptive field surrounds. The difference between the linear response and the observed firing rate is a result of nonlinear processing and is responsible for over half the decorrelation implemented by the retina (green wedge). The trial-to-trial variation contributes an additional small amount of decorrelation (blue wedge). (**d**) Decorrelation in the time domain. Autocorrelation functions of salamander ganglion cell responses and linear models are plotted as a function of delay during naturalistic stimulation. The linear filter's first lobe, ~100 ms wide (inset, black), introduced excess correlation beyond that in the stimulus. The antagonistic second lobe (inset, red) counteracted those, but overcompensated, introducing anticorrelations at long delays. The observed correlations in the firing rate were much smaller still. (**e**,**f**) Spatial (**e**) and temporal (**f**) correlations in macaque RGCs, displayed as in **a**, **b** and **d**. Macaque RGC responses were approximated by an LN model[13,23], using published spatio-temporal receptive field parameters[36] (equations (4–6)) and sigmoidal nonlinearities[23] (equation (10)). The output noise was modeled as sub-Poisson variation (equation (11)) with parameters derived from published spike trains[21] (see Online Methods). The stimulus was scaled in space and time to compensate for the different scales of primate and salamander receptive fields. L, receptive field filter only; LN, including the nonlinearity; LN + noise, including the noise.

Indeed, correlations of the trial-averaged firing rates (**Fig. 3b**) lay far below those of the stimulus and those predicted from receptive field filtering alone (**Fig. 2a**). The effect is especially notable for neurons of opposite response polarity, where the retinal nonlinearities effectively abolish the pairwise correlations (**Fig. 2b**). Finally, we determined that the trial-to-trial fluctuations in different neurons were largely independent under the present stimulus conditions (**Supplementary Fig. 2**). This noise further decorrelates the ganglion cell output (**Fig. 2a**,**b**). Such noise-induced effects are detrimental to efficient coding, but downstream circuits in the brain cannot distinguish decorrelation by noise from that achieved by other means.

One can now compare how much the different aspects of retinal processing contribute to decorrelating the retinal output. For instance, at a distance of 300 μm, the natural stimulus contained strong correlations, but retinal processing subsequently reduced them by a total of 92% (**Fig. 2a**). Of this, the receptive field surround contributed ~25%, the sparsifying nonlinearities contributed ~60% and noise was responsible for ~15% (**Fig. 2c**). Thus, nonlinear processing in retinal circuits is by far the largest contributor to decorrelation at the retinal output, whereas the much-touted center-surround receptive field makes only a minor contribution.

These observations applied to temporal correlations as well. Filtering the stimulus through the receptive field produced a mild reduction in the autocorrelation at short time delays, but also introduced strong anticorrelations at long delays (**Fig. 2d**). This was a result of the biphasic time course of the receptive field (**Fig. 1d**), analogous to the spatial antagonism between center and surround. In comparison, both the trial-averaged firing rate and noisy spike trains showed almost complete decorrelation, down to delays <100 ms (**Fig. 2d**). Again, one concludes that the filtering by receptive fields reduces stimulus correlations only marginally, whereas the sparsifying nonlinearities account for the bulk of temporal decorrelation in the retina.

To assess the generality of these results, we asked whether they also extend to primate retinas and, thus, to our own visual processing. Published spike trains show that macaque RGCs similarly produce sparse bursts separated by silence[21,22], an indication that substantial nonlinear processing occurs in the macaque retina. By analyzing the

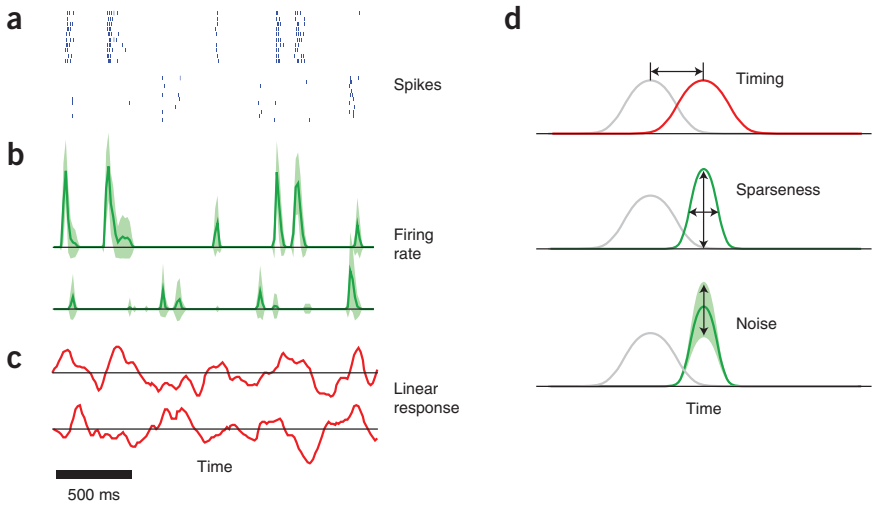

Figure 3 Sparseness in retinal responses. (**a**) Spike rasters for two salamander ganglion cells over ten repetitions of a naturalistic stimulus. Firing events are brief, separated by long silences, and have some trial-to-trial variability. (**b**) Mean firing rates for the same neurons, with shading that indicates the s.d. about the mean in time bins of 50 ms. (**c**) The linear response generated from convolving the stimulus with the spatiotemporal receptive fields of those two cells. This linear model generally captures the times of firing events, but differs markedly in sparseness. (**d**) Depiction of three factors contributing to decorrelation between two caricatured neural responses: event timing, sparseness and noise.

shapes of ganglion cell receptive fields[23,24], we confirmed that the center-surround filter explains only part of the decorrelation at the retinal output (**Fig. 2e**,**f**). On the other hand, the sparsifying nonlinearities in the response[23] make a substantial contribution. Again, they strongly decorrelate responses of opposite polarity (**Fig. 2e**) and they suppress negative temporal correlations at long delays (**Fig. 2f**).

Decorrelation, sparseness and efficient coding in LNP models

To build intuition for these effects and to prepare for further analysis, we considered a simple, tractable model of neural signaling that has enjoyed some popularity in the study of retina, visual cortex and other sensory modalities[25]. In the so-called LNP model, the visual stimulus is first convolved with a linear receptive field (L), producing a time-varying input signal. That signal is passed through an instantaneous nonlinearity (N), typically of sigmoid shape, producing a time-dependent firing rate from which the spike train is generated by a Poisson process (P). The LNP model offers perhaps the simplest instance in which one can analyze the contributions of receptive field, nonlinearity and noise to visual coding.

Consider two such neurons that process a Gaussian-distributed stimulus with different receptive field filters (**Fig. 4**). The outputs of the two filters will be jointly Gaussian variables with a statistical dependency that is fully characterized by the correlation coefficient. Passage through the subsequent nonlinearity always reduces the correlation of the two signals (**Fig. 4b–d**), regardless of the shape of the nonlinearity[26]. For a monotonic sigmoid nonlinearity, a higher threshold produces greater decorrelation (**Fig. 4c**,**d**). An increase in threshold also lowers the mean firing rate, accounting for earlier observations that correlations decrease when firing rates are low[27]. Note that a nonlinearity with a high threshold has qualitatively different effects from one with low threshold: although it suppresses positive correlation coefficients to a certain extent, it almost completely eliminates negative correlations (**Fig. 4c**,**d**). This is because two signals of opposite sign cannot cross threshold at the same time. These effects are very robust under different shapes of the nonlinearity (**Fig. 4c**), and likely explain why the observed anti-correlations between ON and OFF cells are so strongly suppressed by retinal nonlinearities (**Fig. 2b**). Finally, the effect of output noise is simply to reduce the correlation coefficient by a further factor (**Fig. 4e**). In sum, the basic relationships that we found for actual retinal spike trains can be understood in the context of a simple model of nonlinear stochastic processing.

The classical theory of retinal decorrelation attributed that phenomenon to filtering by center-surround receptive fields and explained its purpose as serving the efficient transmission of visual information through the optic nerve. Given that most of the observed decorrelation is instead furnished by the nonlinear response function of the retina, one wonders whether this version of decorrelation is equally beneficial for efficient coding. We explored this in the context of the LNP model and compared the resulting predictions with the measured spike trains.

In the LNP model, the nonlinearity decorrelates if it has a high threshold (**Fig. 4d**), ensuring that each neuron spends much of the time silent except for sharp and sparse firing events. This sparseness is prominent in the ganglion cell responses (**Fig. 1c**) and has been observed across species[11,21,22,28,29]. This seems to be counterproductive for efficient information transmission. Why don't ganglion cells modulate their firing rate continuously to encode different stimulus values? Suppose a neuron must transmit an input signal that changes every time interval Δt by producing spikes during each interval according to a Poisson process with some firing rate. What mapping from input to firing rate maximizes the information rate in the spike train?

To explore this, we compared different monotonic sigmoid nonlinearities, as are often observed in fitting the LNP model to visual neurons[23,30]. These can be described by three parameters: threshold, gain and peak rate (**Fig. 4b**). We took the filtered stimulus to have a normal distribution; this is guaranteed by construction for our Gaussian naturalistic stimulus and by the central limit theorem for the white noise stimulus because the receptive fields extend over many stimulus values in space and time. Next, we compute the mutual information between stimulus and spike train for any shape of the nonlinearity. The information can be increased arbitrarily by simply raising all of the firing rates, so we fixed the mean firing rate to a realistic value for RGCs. That constraint leaves only two free shape parameters for the nonlinearity, for example, the threshold and the gain.

At very high thresholds, the information transmission is poor (**Fig. 4f**). In this regime, the neuron reports only the rare threshold crossings, firing a burst of spikes each time to match the mean firing rate. Notably, transmission also drops at low thresholds. In this condition, the neuron fires in many of the time bins, and the spike counts must therefore be low to satisfy the average rate constraint. In a Poisson process, however, low spike counts are associated with high relative variability. Thus, the choice of threshold involves a trade-off between rarely using reliable symbols, such as high spike counts, or frequently using unreliable symbols, such as low spike counts. The optimum is found at an intermediate threshold value.

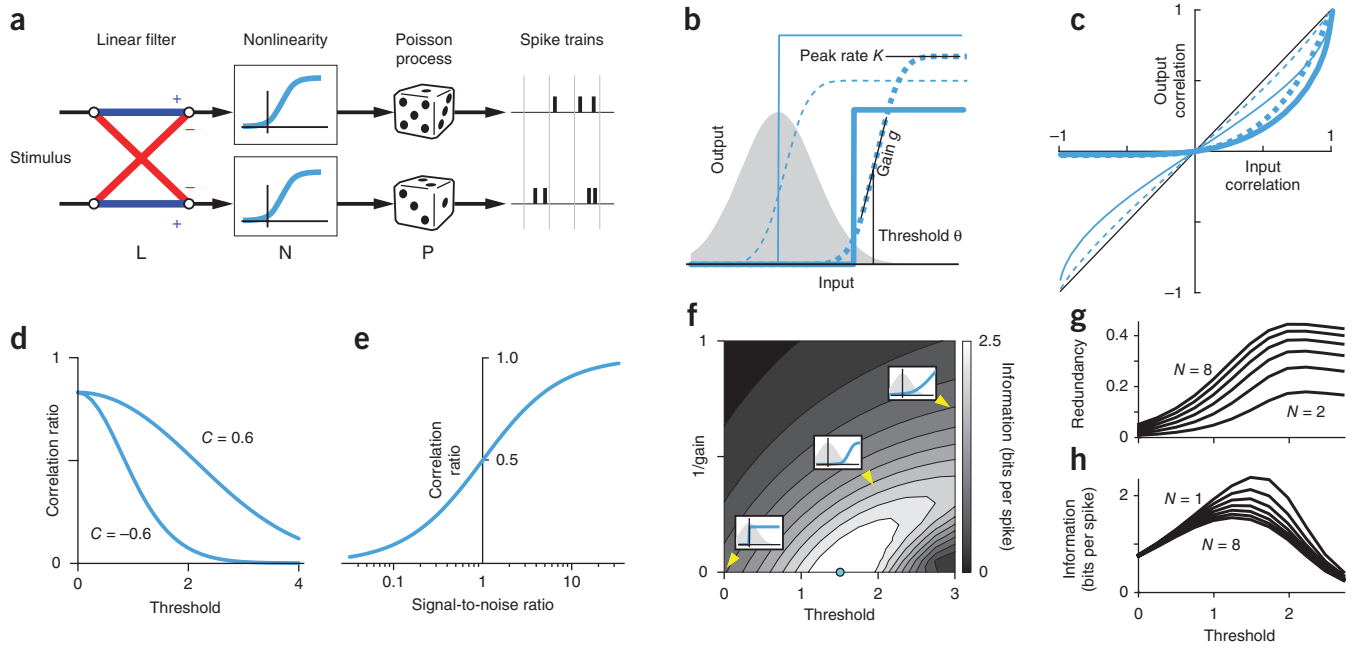

Figure 4 Decorrelation and efficient coding in the LNP model. (**a**) Schematic of two visual neurons that each respond according to the LNP model. For each cell (top and bottom), the stimulus is processed by a linear filter that includes lateral inhibition in space. This signal is passed through a sigmoid nonlinearity and the result modulates the rate of a Poisson process that generates spikes; the spike counts in discrete time windows are the response variable. (**b**) Four sample nonlinearities with sigmoid shape and high or low gain (solid or dashed lines), high or low threshold (thick or thin lines), and various peak rates. The shaded curve indicates the probability distribution of the filtered stimulus signal at the input to the nonlinearity. (**c**) The effects of such a nonlinear transform on the correlations between two jointly Gaussian variables. Note that the output correlation is always less than that of the input. A low threshold (thin lines) affects the correlation only weakly, but at high threshold (thick lines) the output correlation is greatly reduced, especially for negative values. The precise shape of the nonlinearity (dashed versus solid) is less important and the peak rate has no effect. (**d**) The ratio of output correlation to input correlation decreases with increasing threshold, shown here for the sigmoid nonlinearity applied to two variables with input correlation $C = \pm 0.6$. (**e**) When the two outputs are affected by independent additive noise, this reduces the output correlation by a factor determined by the signal-to-noise ratio (equation (14)). (**f**) Influence of the nonlinearity on information transmission. In the framework of the LNP model, the threshold and gain of the sigmoid nonlinearity determine how much information about the stimulus is transmitted by the spikes (grayscale and contour lines). The average firing rate was fixed at 1.1 Hz (the median over the salamander ganglion cells). Threshold and 1/gain are measured in s.d. of the input signal distribution. Insets illustrate nonlinearities (solid lines) at different thresholds and gains relative to the input distribution (shaded area). (**g**,**h**) When multiple neurons receive correlated inputs, raising the threshold makes their outputs more redundant (**g**) even as the total information increases (**h**) and correlation decreases (**d**). All neurons had pairwise correlation coefficients of 0.9, equal thresholds, optimal (infinite) gain and a fixed mean firing rate of 1.1 Hz. The optimal threshold varies only weakly with population size ($N = 1,\ldots,8$).

The optimal gain of the model neuron was infinite (**Fig. 4f**), with complete silence for stimuli below threshold and maximal firing rate for those above. This result runs counter to the conventional view of neural coding as a graded modulation of the firing rate, although related predictions have been in the theory literature for some time[31,32]. For the parameters that characterize a typical salamander ganglion cell from our experiments (coding window Δt=50 ms, average firing rate = 1.1 Hz), the optimal neuron should remain silent 94.5% of the time and fire at 20 Hz the remaining 5.5% of the time. Thus, efficient coding theory predicts that, under the present constraints on firing rate and dynamics, a neuron should indeed fire sparsely, with brief firing events being separated by periods of silence.

Sparse firing enhances coding efficiency of ganglion cells

How close do empirically observed firing rates come to optimal performance? We made the approximation that the dominant source of noise in ganglion cell responses arises at the output, after all of the retina's nonlinear processing has occurred, for example, during spike generation. In that case, the information transmission rate about the stimulus only depends on the probability distribution of the ganglion cell's firing rate, and not on how it is generated (equation (18)). Inspecting that distribution (**Fig. 5a**) reveals that, in most time bins, the measured rate was exactly zero, followed by a long tail in the distribution out to high values. These distributions are fit well by a three-parameter expression (equation (19)). How efficient are these distributions of the firing rate for information transfer?

For comparison, we identified the firing rate distribution with the same mean that used spikes most efficiently. Because real ganglion cell spike trains do not conform exactly to Poisson statistics[17,21,33], we used an empirically fit noise model (equation (11), **Supplementary Fig. 2**). The optimum firing rate distribution, as for the LNP model considered above, was a binary distribution that uses just two firing rates (**Fig. 5b**). But there was a corridor of high efficiency leading to that point, and almost all of the measured rate distributions lay in that domain. Indeed, when we computed the information transmission directly from the spike trains (Online Methods), the median RGC had a coding efficiency of 73% compared with the theoretical optimum (**Fig. 5c**).

Again, we found that these results extend to responses from primate RGCs. Although their mean firing rates were higher, the correlation time of the response, and thus the effective bin width, for spike train signaling was shorter, on the order of Δt = 10 ms[21]. We analyzed

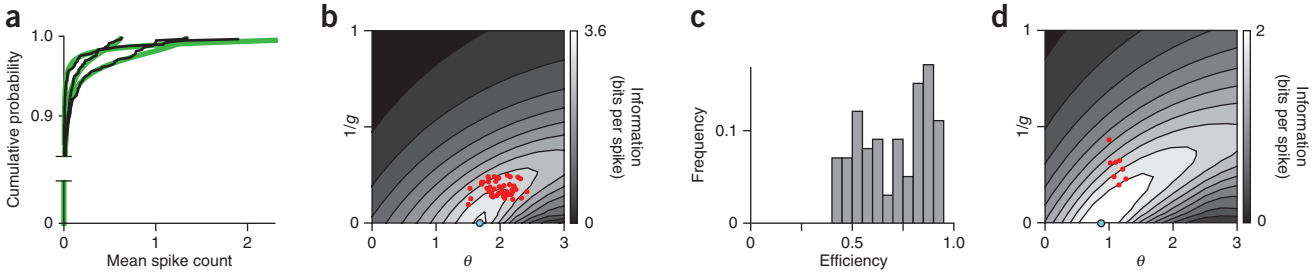

Figure 5 Efficiency of stimulus coding by RGCs. (**a**) Cumulative distribution of the spike count in 50-ms time bins, averaged over multiple repeats of the stimulus. Data (thin lines) for three sample ganglion cells and their fit with a model (thick lines) parametrized by θ, g and K (equation (20)). (**b**) The information transmitted by model firing rate distributions with a fixed mean firing rate of 1.1 Hz, whose shape is parametrized by θ and g. Noise was assumed to be sub-Poisson as observed empirically (equation (11), **Supplementary Fig. 3**). The blue dot indicates the globally maximal rate of information transmission at this mean rate. Red dots indicate the parameters of the rate distribution measured from salamander ganglion cells. These cells have widely varying mean firing rates. The contour plot of information transmission varies slightly with mean rate, but is shown here for illustration purposes only at one typical mean rate. (**c**) Histogram of information efficiencies over the population of salamander RGCs. For each cell, the information rate is calculated directly from the empirical spike counts. To calculate efficiency, we compared this information rate to the maximal information rate possible for the measured mean firing rate (Online Methods). (**d**) Information transmission estimated for macaque RGCs, displayed as in **b**. Red dots are parameters describing the firing rate distribution obtained from published spike rasters in response to white noise stimulation[21]. The contour plot shows the information transmission for different firing rate distributions while fixing the mean rate and time window to typical values, namely 30 Hz and 10 ms, respectively.

published distributions of the firing rate and the trial-to-trial noise[21] and computed information transmission rate as described above. The sparse responses of macaque neurons allowed a transmission rate close to the optimum, with a median efficiency of 81% (**Fig. 5d**). In summary, we found that a treatment of efficient coding theory that incorporates nonlinear transforms and noisy spike trains can explain the paradoxical nature of high-threshold nonlinearities and sparse responses in retinal processing.

DISCUSSION

Our findings extend the application of efficient coding theory in the retina to considerably more realistic conditions. The classic approach treated the early visual system as a linear filter, with graded output signals, Gaussian noise and an average power constraint[4,7], none of which describes the real retina. We allowed for nonlinear processing, a spiking output with stochastic noise and a constraint on the overall firing rate, as might be dictated by metabolic cost[34]. These extensions deliver new insights into the nature of retinal processing.

Two forms of redundancy reduction

We viewed the prominent decorrelation of signals in the retinal output as deriving primarily from two very different mechanisms (**Figs. 2–4**). The first is a linear spatio-temporal filter that implements lateral inhibition in space and biphasic responses in time. This conforms to the classic notion that the retina seeks to reduce redundancy between parallel channels in space and in a channel across time[2], although this reduction is incomplete (**Fig. 2**).

The second, more substantial contribution derives from nonlinear processing in each individual channel (**Figs. 2** and **3**), which efficiently matches visual signals to the available coding symbols (**Figs. 4** and **5**). This second stage reduces the coding redundancy in each output channel resulting from inefficient symbol use. These observations apply for ganglion cells of multiple types in different species, such as salamander and macaque (**Figs. 2** and **5**), suggesting that our extension of the efficient coding framework has some general utility.

Validity of the assumptions

Although the model of retinal processing that we used is considerably more realistic than that described in the classical linear decorrelation theory, it is worth inspecting the remaining approximations. For our first claim, that the receptive field filters contribute only a fraction of the decorrelation, we used a standard method to measure receptive fields, namely reverse correlation of the response to white noise stimuli[13]. Although receptive fields can adapt to the pattern of stimulation[35], we found that surrounds estimated under naturalistic stimulation narrowed only slightly and did not decorrelate any more than those obtained with white noise (data not shown).

Our second claim, that high-threshold nonlinearities enhance efficient coding, assumes that photoreceptor noise is negligible. This is the regime in which the classical theory predicts decorrelation of the retinal output. We also used this assumption to estimate the information rates in spike trains. The high light levels that we used in the experiments were designed to favor low photoreceptor noise. Any remaining input noise would be shared by ganglion cells with overlapping receptive fields, but we found noise correlations to be very small (**Supplementary Fig. 2**). This suggests that most of the noise in the RGC responses arises close to the output, rather than in shared presynaptic sources.

Our analysis of information transmission follows a classic approach[31] and requires choice of a coding window Δt, the timescale on which RGCs can completely change their firing rates to different values. We adopted $\Delta t = 50$ ms for salamander ganglion cells on the basis of the observed width of firing events (**Fig. 3a**,**b**) and their autocorrelation function (**Fig. 2f**). We varied Δt in the analysis and found that the general conclusions were insensitive to small changes in this parameter; sparse firing provides the most efficient code as long as the mean spike count remains considerably less than one spike per time bin.

Our models of signal and noise in the coding window do not specify a particular mechanism of spike generation. However, it is worth noting that the experimentally observed distributions of the firing rate (**Fig. 5a**) and the noise (**Supplementary Fig. 3**) are readily reproduced by mechanistic models such as a leaky integrate-and-fire neuron with Gaussian subthreshold noise (data not shown).

Incomplete decorrelation by receptive fields

We found that the spatial receptive fields of ganglion cells failed to decorrelate retinal signals as completely as would be expected from

the classical theory (**Fig. 2a**). Basically, the antagonistic surround of the receptive field is weaker than predicted. With the high luminance and high contrast that we used, the theory predicts that the integrated strength of the surround should precisely cancel the center (equation (2.4) of ref. 10). The receptive fields that we observed have much weaker surrounds, and thus decorrelate less. This is also evident in preceding work. In macaque RGCs, the surround amounts to only ~50% of the center (see Fig. 10 of ref. 36). Note that, although the original studies always wrote about "retinal filters" for spatial decorrelation, their tests of the theory used comparisons to human psychophysics, and therefore included post-retina stages of decorrelation (see Figs. 1 and 4 of ref. 4).

A plausible explanation of why the linear receptive fields fail to decorrelate much is that they don't entirely reflect what these RGCs compute. Many of these neurons are selective for quite specific visual features, such as motion in a particular direction[37], differential motion[38] or local edges[39]. This selectivity arises from diverse nonlinearities[40] and is poorly represented in the spatio-temporal receptive field. Thus, even neurons with strongly overlapping receptive fields may nonetheless never fire together. This recalls another feature of retinal organization that (so far) cannot be explained by efficient coding principles: the profusion of different ganglion cell types that each appear to compute a different visual message[41].

Nonlinearity and sparseness

Regardless of what a RGC computes, it must communicate the result downstream via noisy spike trains. To optimize information transmission using such a spiking process with a low mean activity, we found that RGCs should be silent most of the time and fire at a high rate only rarely. This expectation holds over all of the experimental conditions that we analyzed, for all of the salamander ganglion cells and for all but one of the macaque neurons. The actual measured nonlinearities were not quite infinitely sharp, but matched the expected threshold closely (**Fig. 5**).

The theory behind this was developed already some time ago. It was discovered by numerical methods that a Poisson process transmits maximal information using a discrete set of firing rates—only two if the maximal rate is strongly limited[31]. The result was later proved analytically in studies of fiber-optic communication[32]. Nevertheless, these facts are poorly appreciated among neuroscientists, even though Poisson models are used ubiquitously. Most of us (ourselves included) assumed that neurons should modulate their firing rate continuously to benefit from all possible rates. This intuition was formalized in an influential study[42] that derived a smooth sigmoid as the optimal shape of the response function. But that treatment was for a continuous output signal, such as membrane potential, and a constant additive noise level. The fact that the spike train is a point process with output-dependent noise ultimately leads to the counter-intuitive step-shaped nonlinearity. This behavior has been derived under a constraint on the maximal firing rate[32]. We found that discrete firing rate distributions also arise when the constraint applies instead to the mean rate (**Fig. 4f**).

The sparse responses of RGCs under naturalistic stimulation can be seen as maximizing coding efficiency in single spike trains in the optic nerve bottleneck. In the cortex, sparse coding has been interpreted differently, as a useful strategy for learning and processing spike patterns[43] or to extract large signals from background noise[44]. These arguments are plausible for highly overcomplete representations, where, unlike in the retina, the number of neurons greatly exceeds the stimulus dimensionality. Still, one might imagine that, even in the cortex, the driving force for sparseness is really communicating efficiently with Poisson spike trains[45].

Decorrelation and efficient coding

Correlation is often considered to be a proxy for information-theoretic redundancy, with the implication that decorrelation somehow improves efficiency. Certainly high correlation does imply strong statistical dependence, but weak correlation need not imply weak dependence; correlation is a second-order measure and fully reflects the redundancy between two signals only if they are normally distributed. For highly non-Gaussian signals, such as neural spike trains and natural images, correlation may be only weakly related to redundancy. For example, the nonlinearity of the LNP model markedly decreases the correlation between neural responses (**Fig. 4d**) while actually increasing their statistical dependency (**Fig. 4g**,**h**). Correlation and efficiency also have a complex relationship. For instance, if two signals are affected by independent noise, this decorrelates them without improving coding efficiency. Nonlinearities invariably decorrelate two Gaussian signals, but may not improve coding efficiency. Nonetheless, many studies of neural signaling simply measure correlation and leave the impression that decorrelation alone is evidence of improved efficiency[46–50].

These examples illustrate that all decorrelations are not created equal. Although many neural circuits perform some decorrelation of their inputs, one must distinguish the various forms of this phenomenon, as they are implemented by very different mechanisms and have different roles for the neural code.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank the members of the Meister laboratory, M. Berry, T. Toyozumi and J.-P. Nadal for helpful advice. This work was funded by grants from the US National Institutes of Health to M.M.

AUTHOR CONTRIBUTIONS

X.P. and M.M. designed the study. X.P. performed all of the experiments, analysis and modeling. X.P. and M.M. wrote the article.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
2. Barlow, H.B. Possible principles underlying the transformation of sensory messages. in *Sensory Communication* (ed. Rosenblith, W.A.) 217–234 (MIT Press, Cambridge, MA, 1961).
3. Srinivasan, M.V., Laughlin, S.B. & Dubs, A. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* **216**, 427–459 (1982).
4. Atick, J.J. & Redlich, A.N. What does the retina know about natural scenes? *Neural Comput.* **4**, 196–210 (1992).
5. Atick, J.J. & Redlich, A.N. Convergent algorithm for sensory receptive field development. *Neural Comput.* **5**, 45–60 (1993).
6. Atick, J.J. & Redlich, A.N. Could information theory provide an ecological theory of sensory processing? *Network* **3**, 213–251 (1992).
7. van Hateren, J.H. Real and optimal neural images in early vision. *Nature* **360**, 68–70 (1992).
8. van Hateren, J.H. Spatiotemporal contrast sensitivity of early vision. *Vision Res.* **33**, 257–267 (1993).
9. Field, D.J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
10. Atick, J.J. & Redlich, A.N. Toward a theory of early visual processing. *Neural Comput.* **2**, 308–320 (1990).
11. Dan, Y., Atick, J.J. & Reid, R.C. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.* **16**, 3351–3362 (1996).

12. Puchalla, J.L., Schneidman, E., Harris, R.A. & Berry, M.J. Redundancy in the population code of the retina. *Neuron* **46**, 493–504 (2005).
13. Chichilnisky, E.J. A simple white noise analysis of neuronal light responses. *Network* **12**, 199–213 (2001).
14. Warland, D.K., Reinagel, P. & Meister, M. Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.* **78**, 2336–2350 (1997).
15. Segev, R., Puchalla, J. & Berry, M.J. Functional organization of ganglion cells in the salamander retina. *J. Neurophysiol.* **95**, 2277–2292 (2006).
16. Enroth–Cugell, C. & Robson, J.G. Functional characteristics and diversity of cat retinal ganglion cells. Basic characteristics and quantitative description. *Invest. Ophthalmol. Vis. Sci.* **25**, 250–267 (1984).
17. Berry, M.J. & Meister, M. Refractoriness and neural precision. *J. Neurosci.* **18**, 2200–2211 (1998).
18. Burrone, J. & Lagnado, L. Synaptic depression and the kinetics of exocytosis in retinal bipolar cells. *J. Neurosci.* **20**, 568–578 (2000).
19. Demb, J.B., Zaghloul, K., Haarsma, L. & Sterling, P. Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (Y) ganglion cell in mammalian retina. *J. Neurosci.* **21**, 7447–7454 (2001).
20. Field, G.D. & Rieke, F. Nonlinear signal transfer from mouse rods to bipolar cells and implications for visual sensitivity. *Neuron* **34**, 773–785 (2002).
21. Uzzell, V.J. & Chichilnisky, E.J. Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol.* **92**, 780–789 (2004).
22. Pillow, J.W. *et al*. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature* **454**, 995–999 (2008).
23. Chichilnisky, E.J. & Kalmar, R.S. Functional asymmetries in ON and OFF ganglion cells of primate retina. *J. Neurosci.* **22**, 2737–2747 (2002).
24. Croner, L.J., Purpura, K. & Kaplan, E. Response variability in retinal ganglion cells of primates. *Proc. Natl. Acad. Sci. USA* **90**, 8128–8130 (1993).
25. Schwartz, O., Pillow, J.W., Rust, N.C. & Simoncelli, E.P. Spike-triggered neural characterization. *J. Vis.* **6**, 484–507 (2006).
26. Lancaster, H.O. Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* **44**, 289–292 (1957).
27. de la Rocha, J., Doiron, B., Shea–Brown, E., Josic, K. & Reyes, A. Correlation between neural spike trains increases with firing rate. *Nature* **448**, 802–806 (2007).
28. Berry, M.J., Warland, D.K. & Meister, M. The structure and precision of retinal spike trains. *Proc. Natl. Acad. Sci. USA* **94**, 5411–5416 (1997).
29. Reinagel, P. How do visual neurons respond in the real world? *Curr. Opin. Neurobiol.* **11**, 437–442 (2001).
30. Baccus, S.A. & Meister, M. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36**, 909–919 (2002).
31. Stein, R.B. The information capacity of nerve cells using a frequency code. *Biophys. J.* **7**, 797–826 (1967).
32. Shamai, S. Capacity of a pulse amplitude modulated direct detection photon channel. *IEE Proc. Commun. Speech Vis.* **137**, 424–430 (1990).
33. 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).
34. Balasubramanian, V. & Berry, M.J. A test of metabolically efficient coding in the retina. *Network* **13**, 531–552 (2002).
35. Hosoya, T., Baccus, S.A. & Meister, M. Dynamic predictive coding by the retina. *Nature* **436**, 71–77 (2005).
36. Croner, L.J. & Kaplan, E. Receptive fields of P and M ganglion cells across the primate retina. *Vision Res.* **35**, 7–24 (1995).
37. Barlow, H.B. & Levick, W.R. The mechanism of directionally selective units in rabbit's retina. *J. Physiol. (Lond.)* **178**, 477–504 (1965).
38. Ölveczky, B.P., Baccus, S.A. & Meister, M. Segregation of object and background motion in the retina. *Nature* **423**, 401–408 (2003).
39. Levick, W.R. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbits retina. *J. Physiol. (Lond.)* **188**, 285–307 (1967).
40. Gollisch, T. & Meister, M. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* **65**, 150–164 (2010).
41. Dacey, D.M. Origins of perception: retinal ganglion cell diversity and the creation of parallel visual pathways. in *The Cognitive Neurosciences* (ed. Gazzaniga, M.S.) 281–301 (MIT Press, Cambridge, Massachusetts, 2004).
42. Laughlin, S.B. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C* **36c**, 910–912 (1981).
43. Olshausen, B.A. & Field, D.J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14**, 481–487 (2004).
44. Ringach, D.L. & Malone, B.J. The operating point of the cortex: neurons as large deviation detectors. *J. Neurosci.* **27**, 7673–7683 (2007).
45. van Vreeswijk, C.A. Whence sparseness? *Adv. Neural Inf. Process. Syst.* **13**, 189–195 (2001).
46. Vinje, W.E. & Gallant, J.L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
47. Wang, X.J., Liu, Y., Sanchez–Vives, M.V. & McCormick, D.A. Adaptation and temporal decorrelation by single neurons in the primary visual cortex. *J. Neurophysiol.* **89**, 3279–3293 (2003).
48. Rucci, M. & Casile, A. Fixational instability and natural image statistics: implications for early visual representations. *Network* **16**, 121–138 (2005).
49. Cleland, T.A. Early transformations in odor representation. *Trends Neurosci.* **33**, 130–139 (2010).
50. Wiechert, M.T., Judkewitz, B., Riecke, H. & Friedrich, R.W. Mechanisms of pattern decorrelation by recurrent neuronal circuits. *Nat. Neurosci.* **13**, 1003–1010 (2010).

ONLINE METHODS

Recording. Experiments were performed on the isolated retina of the larval tiger salamander, superfused with oxygenated Ringer's solution, following protocols approved by the Institutional Animal Care and Use Committee at Harvard University. Action potentials from RGCs were recorded extracellularly with a multi-electrode array[51]. Neurons were selected for analysis if they maintained steady firing rates throughout the 2-h experiments and their spike waveforms could be sorted unambiguously. 103 cells from 9 retinas satisfied this criterion. Classification into different cell types was achieved by agglomerative maximum-linkage clustering according to the Euclidean distances between temporal receptive fields[52]. Of the recorded cells, 6 were classified as ON cells, 18 as slow OFF cells, and 79 as fast OFF cells. Altogether, this yielded 5,356 response pairs, including comparisons across experiments.

Stimulation. Light was projected from a computer monitor onto the photoreceptor layer. The stimulus was a square grid with fields of size 54 μm^2 covering a total area of 3.4 mm^2. The monitor refresh interval was 15 ms. The mean light level at the retina (7×10^{-3} W m^{-2}) was in the regime of photopic vision[51].

The decorrelation theories assumed that light intensities in visual stimuli are drawn from a correlated multivariate normal distribution exhibiting the spatial power spectrum measured for natural scenes, which varies with spatial frequency $\mathbf{k}$ as $1/|\mathbf{k}|^2$. These assumptions neglect objects, edges and textures, but capture pairwise intensity correlations in the visual world. To address these theories directly, we designed spatiotemporal stimuli that approximated the pairwise correlations in natural stimuli and neglected all higher order structure.

We generated the spatial structure of the stimulus $S(\mathbf{x},t)$ by drawing spatial frequency coefficients $\tilde{S}_0(\mathbf{k},t)$ independently every 15 ms from a Gaussian distribution with variance proportional to $1/|\mathbf{k}|^2$. Temporal correlations were introduced by low-pass filtering the spatial frequency coefficients with an exponential of time constant $\tau = 1/|\mathbf{k}|v$, where v is a constant with units of velocity that determines the scaling between space and time. This constant was set to $v = 10°\ s^{-1}$, corresponding to a typical velocity that elicits neural and behavioral responses in salamanders in visual tasks[53]. The spatial frequency coefficients were given by

$$\tilde{S}(\mathbf{k},t) = Ae^{-t/\tau} \circ \tilde{S}_0(\mathbf{k},t) \tag{1}$$

where $\circ$ represents a temporal convolution and the constant A fixed the overall contrast (the ratio of s.d. of luminance to mean luminance) at 35%. An inverse spatial Fourier transform generated each image frame for display (**Fig. 1a**). The overall spatiotemporal power spectrum at spatial frequency $\mathbf{k}$ and temporal frequency ω is

$$\tilde{P}(\mathbf{k},\omega) \propto \frac{1}{|\mathbf{k}|^2} \cdot \frac{1-e^{-2|\mathbf{k}|v\Delta t}}{1+e^{-2|\mathbf{k}|v\Delta t}-2e^{-|\mathbf{k}|v\Delta t}\cos\omega\Delta t} \tag{2}$$

(**Supplementary Fig. 1**). This spectrum closely approximates that of natural movies[54]; in the limit of low spatial frequency it becomes $|\mathbf{k}|^{-3} f(\omega/|\mathbf{k}|)$ for a function f peaked at zero in the ratio $\omega/|\mathbf{k}|$. In this limit the power varies with temporal frequency approximately as ω^{-2}, as observed[54]. The stimulus has qualitative similarities with natural scenes: Large features are more prominent, and persist longer than small details. To compare the results from macaque and salamander, we used the same stimuli, except that the stimulus checker size was scaled in proportion to the mean ganglion cell receptive field radius.

Correlation. The correlation between two signals, x and y, was quantified by the second-order correlation function

$$C_{xy}(\tau) = \frac{\langle \Delta x(t) \cdot \Delta y(t+\tau) \rangle}{\sqrt{\langle \Delta x^2(t) \rangle \langle \Delta y^2(t) \rangle}} \tag{3}$$

where Δx and Δy represent deviations of x and y from their respective means and $\langle \cdot \rangle$ symbolizes an average over time. To reduce high-frequency noise, we first binned each signal into windows of width $\Delta t = 50$ ms for salamander neurons and 10 ms for primate neurons. This sets the time resolution on which the neural responses are analyzed. These values were chosen because they reflect the timescale on which ganglion cell firing varies: the typical duration of a stimulus-evoked burst of spikes (**Fig. 1c**) and the width of the peak in the temporal receptive field (**Fig. 1d**).

As the shared noise sources were small (**Supplementary Fig. 2**), we focused on stimulus-driven correlations, by presenting the same stimulus twice and computing correlations between the spike trains across the two repeats. The correlation measure (equation (3)) was computed the same way for pairs of stimulus values, trial-averaged firing rates, spike trains or the outputs of various functional models. In graphs of correlation versus spatial distance, we plotted the correlation at zero delay, $C_{xy}(0)$. For visualization, we binned the cell pairs by distance into groups of 100 and plotted the median for each group (**Figs. 1g,h** and **2**). Distances were quantified as the separation between the midpoint of the receptive fields.

Receptive fields. To map the receptive fields, we applied a random checkerboard stimulus[51] with a temporal sampling rate of 22 Hz and (54 μm)^2 black or white checkers. To reduce noise in the receptive field estimate, we fitted each neuron's spatiotemporal receptive field with a direct product of a spatial receptive field and a temporal kernel

$$F(\mathbf{x},t) \approx X(\mathbf{x})T(t) \tag{4}$$

using singular value decomposition. Each neuron's position was assigned as the midpoint of a two-dimensional Gaussian fit to its spatial receptive field $X(\mathbf{x})$ (**Fig. 1e**).

For modeling primate receptive fields, we parametrized $X(\mathbf{x})$ and $T(t)$ as

$$X(\mathbf{x}) = \frac{1}{2\pi\sigma_c^2} e^{-|\mathbf{x}-\mu_c|^2/2\sigma_c^2} - a\frac{1}{2\pi\sigma_s^2} e^{-|\mathbf{x}-\mu_s|^2/2\sigma_s^2} \tag{5}$$

$$T(t) = (t/\tau_1)^{n_1} e^{-n_1(t/\tau_1-1)} - b(t/\tau_2)^{n_2} e^{-n_2(t/\tau_2-1)} \tag{6}$$

with spatial parameters drawn from a previous study[55] for the parafovea (5°–10° eccentricity) and temporal parameters drawn from ref. 56.

Given a receptive field $F(\mathbf{x},t)$, we computed the linear prediction $r(t)$ for the neural response by convolution with the stimulus

$$r(t) = \iiint d^2\mathbf{x}\, d\tau\, F(\mathbf{x},\tau) S(\mathbf{x},t-\tau) \tag{7}$$

Nonlinearities. In the LNP model, the linear prediction $r(t)$ is transformed into a firing rate $\rho(t)$ by an instantaneous nonlinearity $N(\cdot)$,

$$\rho(t) = N(r(t)) \tag{8}$$

and then into a spike count n by drawing from a Poisson distribution with that rate

$$P(n|\rho) = \frac{e^{-\rho\Delta t}(\rho\Delta t)^n}{n!} \tag{9}$$

where Δt is the time bin. We parametrized the nonlinearity as a sigmoid using the logistic function

$$N(r) = K/\left(1+e^{-g(r-\theta)}\right) \tag{10}$$

with peak firing rate K, gain g and threshold θ.

If the linear input $r(t)$ follows a normal distribution, one can constrain the mean firing rate of the model neuron to a value μ by setting the peak rate to

$$K = \mu\sqrt{2\pi} \Big/ \int dr\, e^{-\frac{1}{2}r^2} \left(1+e^{-g(r-\theta)}\right)^{-1}$$

Noise. Large bursts of spikes from ganglion cells are more regular than expected from Poisson statistics[57,58], so the Poisson model generally overestimates the noise. For some computations (**Fig. 5b–d**) we used a noise distribution that was measured empirically. For a given mean spike count ρ at a given time during the trial, the measured spike count distributions $P(n|\rho)$ had a width that stayed constant with ρ after an initial Poisson-like growth (**Supplementary Fig. 3**).

These distributions were well-described as a Gaussian distribution on non-negative integer spike counts

$$P(n|\rho) \propto \exp\left[-\frac{(n-n_0(\rho))^2}{2\sigma^2}\right] \quad n=0,1,2,\ldots \tag{11}$$

where

$$n_0(\rho) = a\log\left(1+e^{\rho\Delta t/a}\right)$$

is the center of the Gaussian and σ is the width of the noise distribution. For each σ, the parameter a was set so the conditional mean of the model noise distribution closely approximated the desired mean ρ. The noise width σ was fit by numerically maximizing the log-likelihood,

$$\sum_{t,i}\log P\left(n(t,i)|\rho(t);\sigma\right)$$

where $n(t,i)$ is the measured spike count in bin t during stimulus repetition i.

Our models assume that noise affects the spiking of each neuron independently, whereas nearby ganglion cells share certain noise sources, especially at low light levels[59]. We found that noise correlations at photopic intensities were very small, <0.01 for 90% of pairwise comparisons (**Supplementary Fig. 2**). This justified the independent noise approximation for the great majority of cells, which simplifies the treatment of optimal coding. Another study reported that response models that account for noise correlations in ganglion cell spike trains can extract additional (~20%) visual information[60].

Decorrelation by nonlinearities and noise. The correlation between two LNP model neurons depends on both the nonlinearities and the noise (**Fig. 4**). Suppose that the inputs x and y to two neurons are both normally distributed with zero mean and unit variance and correlation coefficient c. After transformation by the nonlinear function $N(\bullet)$, the correlation coefficient becomes

$$C_{N(x)N(y)} = \frac{\langle N(x)N(y)\rangle - \mu^2}{\sigma^2} \tag{12}$$

where the nonlinear output has mean $\mu = \langle N(x)\rangle$ and variance $\sigma^2 = \langle N^2(x)\rangle - \mu^2$, and where $\langle\ \rangle$ is an expectation over the input distribution

$$P(x,y) = \frac{1}{2\pi\sqrt{1-c^2}}\exp\left(-\frac{1}{2}\frac{x^2-2xyc+y^2}{1-c^2}\right) \tag{13}$$

We computed these expectation values by numerical integration (**Fig. 4b–d**).

Response noise increases the variance without altering the covariance, lowering the correlation. For two conditionally independent signals x and y with (time dependent) trial averages of $\bar{x}$ and $\bar{y}$, the noise is $\delta x = x - \bar{x}$ and $\delta y = y - \bar{y}$. The correlation between the noisy signals x and y is then

$$\begin{aligned} C_{xy} &= \frac{\langle\bar{x}\bar{y}\rangle - \langle\bar{x}\rangle\langle\bar{y}\rangle}{\sqrt{\langle\bar{x}^2\rangle - \langle\bar{x}\rangle^2 + \langle\delta x^2\rangle}\sqrt{\langle\bar{y}^2\rangle - \langle\bar{y}\rangle^2 + \langle\delta y^2\rangle}} \\ &= C_{\bar{x}\bar{y}}\frac{1}{\sqrt{\left(1+1/SNR_x\right)\left(1+1/SNR_y\right)}} \end{aligned} \tag{14}$$

where $C_{\bar{x}\bar{y}}$ is the correlation of the trial-averaged responses and $SNR_x = \left(\langle\bar{x}^2\rangle - \langle\bar{x}\rangle^2\right)\Big/\langle\delta x^2\rangle$ is the ratio of signal variance to noise variance (**Fig. 4e**).

Information and efficiency for a single neuron. To analyze the role of nonlinearity in efficient coding, we computed the mutual information between the stimulus and the ganglion cell spike count in single windows of width Δt. This approximation neglects correlations between spike counts in different bins and spike timing within a bin. The mutual information between stimulus s and the spike count n is

$$I(n;s) = H(n) - H(n|s) \tag{15}$$

where the unconditional entropy $H(n)$ is

$$H(n) = -\sum_{n=0}^{\infty} p(n)\log p(n) \tag{16}$$

and the conditional entropy $H(n|s)$ is

$$H(n|s) = -\int ds\, p(s)\sum_{n=0}^{\infty} p(n|s)\log p(n|s) \tag{17}$$

We calculated the mutual information in two ways: directly from neural responses, and using a response model. For the former, the integrals over all possible stimuli were replaced by temporal averages over the stimulus presentation. For the latter, the integrals over the high-dimensional stimulus ensemble are intractable. However, the model responses depend on the stimulus only through the time-varying firing rate $\rho(t)$. Assuming again that input noise is negligible, this firing rate is a deterministic function of the stimulus. Thus, the conditional entropy given the stimulus equals the conditional entropy given the firing rate, $H(n|s) = H(n|\rho)$, and the mutual information is fully specified by the distribution of firing rates $p(\rho)$, regardless of how those rates arise

$$I(n;s) = H(n) - H(n|s) = H(n) - H(n|\rho) = I(n;\rho) \tag{18}$$

Thus we compute entropies (equations (16–17)) using $p(n) = \int d\rho\, p(n|\rho)\, p(\rho)$ and $p(n|\rho)$ instead of $p(n) = \int ds\, p(n|s)\, p(s)$ and $p(n|s)$.

For the LNP model, the firing rate distribution is produced by the sigmoid nonlinearity $N(r)$ (equation (10)) acting on the Gaussian distributed linear input r. These distributions are parametrized like the logistic function, by the peak rate K, gain g and threshold θ

$$p(\rho) = \frac{K\exp\left[-\frac{1}{2}\left(\frac{1}{g}\log\left(K/\rho-1\right)-\theta\right)^2\right]}{\sqrt{2\pi}g\rho^2\left(K/\rho-1\right)} \tag{19}$$

This family of distributions encompasses a wide range of unimodal and bimodal shapes, including binary rate distributions when $g = \infty$.

To fit each ganglion cell response distribution (**Fig. 5b,d**), we minimized the mean squared difference between the cumulative distribution of the parametric model (equation (19))

$$D(\rho) = \frac{1}{2}\mathrm{erfc}\left(\frac{\log\left(K/\rho-1\right)-g\theta}{\sqrt{2}g}\right) \tag{20}$$

and the cumulative distribution of the measured firing rates. The median parameters over the recorded salamander cells were $K = 48$ Hz, $g = 5.8$ and $\theta = 2.0$. For primate neurons, fits were derived from published spike rasters[58], with median parameters $K = 72$ Hz, $g = 2.8$ and $\theta = 0.95$.

We numerically calculated the mutual information for the response model (**Figs. 4f** and **5b,d**) by substituting the rate distribution (equation (19)) and noise model (equation (11)) into equations (15–17).

Given a neuron's mean firing rate μ, we determined the firing rate distribution that optimizes information transmission by numerically maximizing mutual information (equation (18)) over the parameters g and θ in equation (19), setting

$$K = \mu\sqrt{2\pi}\Big/\int dr\, e^{-\frac{1}{2}r^2}\left(1+e^{-g(r-\theta)}\right)^{-1} \tag{21}$$

to preserve the mean firing rate. Finally, we computed coding efficiency (**Fig. 5c**) by dividing the mutual information for the measured neural responses by this maximal information.

doi:10.1038/nn.3064

Information and redundancy for multiple correlated neurons. To compute the mutual information for a population of N LNP model neurons (**Fig. 4g**,**h**), we allowed the spike counts and firing rates in equations (15–17) to be N-dimensional vectors. We made several simplifications for tractability. First, all models had identical thresholds θ, gains g and peak firing rates K. Next, we assumed the input to the nonlinearities was a multivariate Gaussian with uniform correlation matrix $\mathbf{\Sigma} = c + (1-c)\mathbf{I}$. Finally, we restricted the nonlinearity to have optimal (infinite) gain; each neuron was either silent or fired at a maximal rate K in each time bin.

Given these simplifying assumptions, we can calculate the mutual information for the population. The unconditional probabilities of the vector of binary firing rates are

$$p(\boldsymbol{\rho}) = \int_{O(\boldsymbol{\rho})} d^N\mathbf{r}\, \frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{r}-\theta)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{r}-\theta)\right) \tag{22}$$

with integration over orthants

$$O(\boldsymbol{\rho}) = \bigcap\nolimits_i \left\{\mathrm{sgn}\left(\rho_i - \tfrac{K}{2}\right) r_i > 0\right\}$$

We computed these integrals numerically, exploiting the model's permutation symmetry to reduce the number of integrals.

The model neurons are silent and have zero noise entropy if $\rho_i = 0$, and emit spikes with probability

$$p(n_i|\rho_i = K) = \begin{cases} q & n_i = 0 \\ 1-q & n_i > 0 \end{cases} \tag{23}$$

and noise entropy

$$h(q) = -q\log q - (1-q)\log(1-q) \tag{24}$$

otherwise, with $q = \exp(-K\Delta t)$ for Poisson noise. The conditional entropy (equation (17)) is the average noise entropy across all firing rate patterns

$$H(\mathbf{n}|\boldsymbol{\rho}) = \sum_{\boldsymbol{\rho}} p(\boldsymbol{\rho}) h(q) \sum_i \frac{\rho_i}{K} \tag{25}$$

The unconditional entropy (equation (16)) is computed from the marginal spike count probability over spikes and silences,

$$p(\mathbf{n}) = \sum\nolimits_{\boldsymbol{\rho}} p(\boldsymbol{\rho}) \prod\nolimits_i p(n_i|\rho_i)$$

Redundancy (**Fig. 4g**) measures the difference between the total information conveyed by each neuron considered independently and the information all neurons convey together, compared to the information that could be conveyed if all neurons were independent,

$$R = \left(\sum\nolimits_i I(\rho_i;s) - I(\boldsymbol{\rho};s)\right) \Big/ \sum\nolimits_i I(\rho_i;s)$$

51. Meister, M., Pine, J. & Baylor, D.A. Multi–neuronal signals from the retina: acquisition and analysis. *J. Neurosci. Methods* **51**, 95–106 (1994).
52. Segev, R., Puchalla, J. & Berry, M.J. Functional organization of ganglion cells in the salamander retina. *J. Neurophysiol.* **95**, 2277–2292 (2006).
53. Himstedt, W. Prey selection in salamanders. in *Analysis of Visual Behavior* (eds. Ingale, D.J., Goodale, M.A. & Mansfield, R.J.W.) 47–66 (MIT Press, Cambridge, Massachusetts, 1982).
54. Dong, D.W. & Atick, J.J. Statistics of natural time-varying images. *Network* **6**, 345–358 (1995).
55. Croner, L.J. & Kaplan, E. Receptive fields of P and M ganglion cells across the primate retina. *Vision Res.* **35**, 7–24 (1995).
56. Chichilnisky, E.J. & Kalmar, R.S. Functional asymmetries in ON and OFF ganglion cells of primate retina. *J. Neurosci.* **22**, 2737–2747 (2002).
57. Berry, M.J. & Meister, M. Refractoriness and neural precision. *J. Neurosci.* **18**, 2200–2211 (1998).
58. Uzzell, V.J. & Chichilnisky, E.J. Precision of spike trains in primate retinal ganglion cells. *J. Neurophysiol.* **92**, 780–789 (2004).
59. Schneidman, E., Bialek, W. & Berry, M.J. Synergy, redundancy and independence in population codes. *J. Neurosci.* **23**, 11539–11553 (2003).
60. Pillow, J.W. *et al.* Spatio–temporal correlations and visual signaling in a complete neuronal population. *Nature* **454**, 995–999 (2008).