Towards a Theory of Early Visual Processing

Joseph J. Atick
School of Natural Sciences, Institute for Advanced Study,
Princeton, NJ 08540, USA

A. Norman Redlich
Department of Physics and Center for Neural Science,
New York University, New York, NY 10003, USA

We propose a theory of the early processing in the mammalian visual pathway. The theory is formulated in the language of information theory and hypothesizes that the goal of this processing is to recode in order to reduce a "generalized redundancy" subject to a constraint that specifies the amount of average information preserved. In the limit of no noise, this theory becomes equivalent to Barlow's redundancy reduction hypothesis, but it leads to very different computational strategies when noise is present. A tractable approach for finding the optimal encoding is to solve the problem in successive stages where at each stage the optimization is performed within a restricted class of transfer functions. We explicitly find the solution for the class of encodings to which the parvocellular retinal processing belongs, namely linear and nondivergent transformations. The solution shows agreement with the experimentally observed transfer functions at all levels of signal to noise.

In the mammalian visual pathway, data from the photoreceptors are processed sequentially through successive layers of neurons in the retina and in the visual cortex. The early stages of this processing (the retina and the first few layers of the visual cortex) exhibit a significant degree of universality; they are very similar in many species and do not change as a mature animal learns new visual perceptual skills. This suggests that the early stages of the visual pathway are solving a very general problem in data processing, which is independent of the details of each species' perceptual needs. In the first part of this paper, we formulate a theory of early visual processing that identifies this general problem.

The theory is formulated in the language of information theory (Shannon and Weaver 1949) and was inspired by Barlow's redundancy reduction hypothesis for perception (Barlow 1961, 1989). Barlow's hypothesis is, however, applicable only to noiseless channels that are unrealistic. The

theory that we develop here is formulated for noisy channels. It agrees with Barlow's hypothesis in the limit of no noise but it leads to different computational strategies when noise is present. Our theory hypothesizes that the goal of visual processing is to recode the sensory data in order to reduce a redundancy measure, defined below, subject to a constraint that fixes the amount of average information maintained. The present work is an outgrowth of an earlier publication in which we addressed some of these issues (Atick and Redlich 1989). However, in that work the role of noise was not rigorously formulated, and although all solutions exhibited there did well in reducing redundancy, they were not proven to be optimal. For a related attempt to understand neural processing from information theory see Linsker (1986,1989) (see also Uttley 1979).

The problem of finding the optimal redundancy reducing code among all possible codes is most likely impossible to solve. A more tractable strategy is to reduce redundancy in successive stages, where at each stage one finds the optimal code within a restricted class. This appears to be the mechanism used in the visual pathway. For example, in the "parvocellular" portion of the pathway, which is believed to be concerned with detailed form recognition, the first recoding (the output of the retinal ganglion cells) can be characterized as linear and nondivergent (code dimension is unchanged). At the next stage, the recoding of the simple cells is still substantially linear but is divergent (for a review see Orban 1984). In this paper, we solve the problem of redundancy reduction for the class of linear and nondivergent codes and we find that the optimal solution is remarkably similar to the experimentally observed ganglion cell recoding. We leave the solution for the next stage of linear divergent codes, where one expects a simple cell like solution, for a future publication.

1 Formulation of the Theory

For concreteness, we shall start by formulating our theory within the specific context of retinal processing. The theory in its more general context will become clear later when we state our redundancy reduction hypothesis. It is helpful to think of the retinal processing in terms of a pair of communication channels, as pictured in Figure 1. In this flow chart, the center box represents the retinal transfer function A, with the signal x representing the visual input including noise ν, and y the output of the ganglion cells. Here, we do not concern ourselves with the detailed implementation of this transfer function by the retina, which involves a fairly complicated interaction between the photoreceptors and the layers of cells leading to the ganglion cells.

Although the input x is the actual input to the visual system, we have introduced an earlier input communication channel in the flow diagram with s representing an *ideal* signal. This earlier communication channel
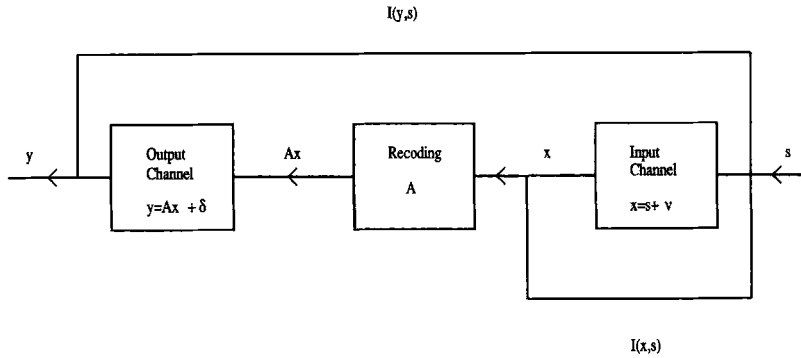
I(y,s)

I(x,s)

Figure 1: Schematic diagram showing the three stages of processing that the ideal signal s undergoes before it is converted to the output y. ν (δ) is the noise in the input (output) channel.

is taken to be the source of all forms of noise ν in the signal x, including quantum fluctuations, intrinsic noise introduced by the biological hardware, and semantic noise already present in the images. It must be kept in mind, however, that neither *noise* nor *ideal* signal is a universal concept, but depend on what is useful visual information to a particular organism in a particular environment. Here we assume that, minimally, the ideal signal does not include noise ν in the form of completely random fluctuations. It is reasonable to expect this minimal definition of noise to apply to the early visual processing of all organisms.

In this paper, we take the input x to be the discrete photoreceptor sampling of the two-dimensional luminosity distribution on the retina. For simplicity, we use a cartesian coordinate system in which the sampled signal at point $\mathbf{n} = (n_1, n_2)$ is $x[\mathbf{n}]$. However, all of the results below can be rederived, including the sampling, starting with spatially continuous signals, by taking into account the optical modulation transfer function of the eye (see the analysis in Atick and Redlich 1990).

The output y is the recoded signal Ax plus the noise δ in the output channel. This channel may be thought of as the optic nerve. Since the recoding of x into y is linear and nondivergent, its most general form can be written as $y[\mathbf{m}] = \sum_{\mathbf{n}} A[\mathbf{m}, \mathbf{n}] x[\mathbf{n}] + \delta[\mathbf{m}]$, where the transfer function $A[\mathbf{m}, \mathbf{n}]$ is a square matrix.

In order to formulate our hypothesis, we need to define some quantities from information theory (Shannon and Weaver 1949) that measure

how well the visual system is communicating information about the visual scenes. We first define the *mutual information* $I(x, s)$ between the ideal signal s and the actual signal x:

$$I(x, s) = \sum_{s,x} P(x, s) \log \left[\frac{P(x, s)}{P(s) P(x)} \right] \tag{1.1}$$

with a similar formula for $I(y, s)$. In equation 1.1, $P(s)$ [or $P(x)$] is the probability of the occurrence of a particular visual signal $s[\mathbf{n}]$ (or $x[\mathbf{n}]$), and $P(x, s)$ is the probability of the joint occurrence of $s[\mathbf{n}]$ together with $x[\mathbf{n}]$. $I(x, s)$ measures the actual amount of useful information available at the level of the photoreceptors x, given that the desired signal is s. Likewise, $I(y, s)$ measures the useful information available in the signal y. Also, for continuous signals, the mutual information is a well-defined, coordinate invariant quantity.

To calculate the mutual information $I(x, s)$ explicitly, it is necessary to know something about the probabilities $P(s)$, $P(x)$, and $P(x, s)$. These probability functions, together with the relationship $y = A x + \delta$, are also sufficient to calculate the mutual information $I(y, s)$. Although $P(s)$, $P(x)$, and $P(x, s)$ cannot be known completely, we do assume knowledge of the second-order correlators $\langle s[\mathbf{n}]s[\mathbf{m}] \rangle$, $\langle x[\mathbf{n}]x[\mathbf{m}] \rangle$, and $\langle x[\mathbf{n}]s[\mathbf{m}] \rangle$, where $\langle \ \rangle$ denotes the average over the ensemble of all visual scenes. We assume that these correlators are of the form

$$\begin{aligned}
\langle s[\mathbf{n}]s[\mathbf{m}] \rangle &= R_0[\mathbf{n}, \mathbf{m}] \\
\langle x[\mathbf{n}]x[\mathbf{m}] \rangle &= R_0[\mathbf{n}, \mathbf{m}] + N^2 \delta_{\mathbf{n},\mathbf{m}} \equiv R[\mathbf{n}, \mathbf{m}] \\
\langle x[\mathbf{n}]s[\mathbf{m}] \rangle &= \langle s[\mathbf{n}]s[\mathbf{m}] \rangle
\end{aligned} \tag{1.2}$$

where $R_0[\mathbf{n}, \mathbf{m}]$ is some yet unspecified correlation matrix, and we have defined $\langle \nu[\mathbf{n}]\nu[\mathbf{m}] \rangle \equiv N^2 \delta_{\mathbf{n},\mathbf{m}}$. Using $x = s + \nu$, equations 1.2 imply that there are no correlations between the noise ν and s. Given these correlators, we assume that the probability distributions are those with maximal entropy:

$$\begin{aligned}
P(u) &= \left[(2\pi)^d \det(R_{uu}) \right]^{-1/2} \\
&\quad \exp\left[-\frac{1}{2} \sum_{\mathbf{n},\mathbf{m}} (u[\mathbf{n}] - \bar{u}) R_{uu}^{-1}[\mathbf{n}, \mathbf{m}] (u[\mathbf{m}] - \bar{u}) \right]
\end{aligned} \tag{1.3}$$

for $u = s, x, y$ and $R_{uu}[\mathbf{n}, \mathbf{m}] \equiv \langle u[\mathbf{n}] u[\mathbf{m}] \rangle$ (d is the dimension of $\mathbf{n}$). We have included here the mean $\bar{u} \equiv \langle u \rangle$, although in all of our results it drops out. Equation 1.3 can also be used to determine the joint proba-

bilities $P(x, s)$ and $P(y, s)$, since these are equal to $P(z_{xs})$ and $P(z_{ys})$ for the larger sets of stochastic variables $z_{xs} = (x, s)$ and $z_{ys} = (y, s)$ whose correlators R_{zz} are calculated from R_{xx}, R_{xs}, R_{ss}, R_{yy}, and R_{ys}.

It is not difficult to show, using the explicit expressions for the various probability distributions, that

$$I(x, s) = \frac{1}{2} \log \left[\frac{\det \left(R_0 + N^2 \right)}{\det N^2} \right] \tag{1.4}$$

$$I(y, s) = \frac{1}{2} \log \left\{ \frac{\det \left[A(R_0 + N^2)A^T + N_\delta^2 \right]}{\det \left(AN^2 A^T + N_\delta^2 \right)} \right\} \tag{1.5}$$

(In 1.5, we used $\langle \delta[n] \, \delta[m] \rangle \equiv N_\delta^2 \delta_{n,m}$.) The mutual informations depend on both the amount of *noise* and on the amount of *correlations* in the signals. Noise has the effect of reducing $I(x, s)$ [or $I(y, s)$] because it causes uncertainty in what is known about s at x (or y). In fact, infinite noise reduces $I(x, s)$ and $I(y, s)$ to zero. This becomes clear in equations 1.4 and 1.5 as N^2 goes to infinity, since then the ratio of determinants goes to one causing I to vanish.

Increasing spatial correlations in equations 1.4 and 1.5 also has the effect of reducing I because correlations reduce the information in the signals. Correlations indicate that some scenes are far more common than others, and an ensemble with this property has lower average information than one in which all messages are equally probable. The effect of increasing correlations is most easily seen, for example, in $I(x, s)$ in the limit of N^2 very small, in which case $I(x, s) \sim \log[\det(R_0)]$. If the average signal strengths $\langle s^2[n] \rangle$ are held constant, then $\log[\det(R_0)]$ is maximum when R_0 is diagonal (no correlations) and vanishes when the signal is completely correlated, that is, $R_0[n, m] = $ constant, $\forall$ n, m. In fact, by Wegner's theorem (Bodewig 1956, p. 71), for positive definite matrices (correlation matrices are always positive definite) $\det(R_0) \leq \Pi_i(R_0)_{ii}$, with equality only when R_0 is completely diagonal.

Having introduced a measure $I(y, s)$ of the actual average information available at y, we now define the channel *capacity* $C_{\text{out}}(y)$ which measures the *maximal* amount of information that could flow through the output channel. Here, we define the capacity $C_{\text{out}}(y)$ as the maximum of $I(y, w)$ varying freely over the probabilities $P(w)$ of the inputs to the *output* channel, holding the average signal strengths $\langle y^2[n] \rangle$ fixed:

$$
\begin{aligned}
C_{\text{out}}(y) &= \max_{P(w)} \; I(y, w) \Big|_{\langle y^2 \rangle = \text{const.}} \\
&= \max_{P(w)} \; \log \det \left(\frac{R_{ww} + N_\delta^2}{N_\delta^2} \right) \Big|_{(R_{ww})_{ii} = \text{const.}} \tag{1.6}
\end{aligned}
$$

where $y = w + \delta$ and $(R_{ww})_{ii}$ are the diagonal elements of the autocorrelator of w (w is a dummy variable, which in Fig. 1 corresponds to Ax). A constraint of this sort is necessary to obtain a finite capacity for continuous signals and is equivalent to holding constant the average "power" expenditure or the variance in the number of electrochemical spikes sent along each fiber of the optic nerve.[1] Using Wegner's theorem, (1.6) the maximum occurs for the probability distribution $P(w)$ for which R_{ww}, or equivalently R_{yy}, is diagonal:

$$C_{\text{out}}(y) = \frac{1}{2} \log \prod_i \left[\frac{R_{yy}}{N_\delta^2} \right]_{ii} \tag{1.7}$$

which for $y = Ax + \delta$ is explicitly

$$C_{\text{out}}(y) = \frac{1}{2} \log \prod_i \left[\frac{A(R_0 + N^2)A^T + N_\delta^2}{N_\delta^2} \right]_{ii} \tag{1.8}$$

At this point, we are ready to state our generalized redundancy reduction hypothesis. We propose that the purpose of the recoding A of the visual signal in the early visual system is to *minimize the "redundancy"*

$$\mathcal{R} = 1 - I(y, s)/C_{\text{out}}(y) \tag{1.9}$$

subject to the constraint that $I(y, s)$ be equal to the minimum average information I^* that must be retained to meet an organism's needs. $I(y, s)$ is therefore constrained to be a fixed quantity and redundancy is reduced by choosing an A that minimizes $C_{\text{out}}(y)$. To avoid confusion, we should emphasize that C_{out} is fixed only at fixed "power," but can be lowered by choosing A to lower the output "power."

Although, in practice we do not know precisely what the minimal I^* is, we assume here that it is the information available to the retina, $I(x, s)$, lowered slightly by the presence of the additional noise δ in the output channel. We therefore choose the constraint

$$I(y, s) = I^* = I(x + \delta, s) \tag{1.10}$$

but our results below do not depend qualitatively on this precise form for the constraint. Since I^* does not depend on A, it can be determined from physiological data, and then used to predict independent experiments (see Atick and Redlich 1990).

The reader should be cautioned that equation 1.9 is *not* the conventional definition of redundancy for the total channel from s to y. The standard redundancy would be $\mathcal{R} = 1 - I(y, s)/C_{\text{tot}}(y)$ where C_{tot} is the maximum of $I(y, s)$ varying freely over the input probabilities $P(s)$,

---

[1]Since a ganglion cell has a nonvanishing mean output, "power" here is actually the cell's dynamic range.

keeping $\langle y^2 \rangle$ fixed. In contrast to C_{tot}, C_{out} is directly related to the "power" in the optic fiber, so reducing equation 1.9 in the manner just described always leads to lower "power" expenditure. Also, since $C_{\text{out}} >$ C_{tot}, lowering C_{out} necessarily lowers the "power" expenditure at all stages up to y, which is why we feel equation 1.9 could be biologically more significant.

Our hypothesis is similar to Barlow's redundancy reduction hypothesis (Barlow 1961), with the two becoming identical when the system is free of noise ν. In this limit, redundancy is reduced by diagonalizing the correlation matrix R_0 by choosing the transfer matrix A such that $R_{yy} = AR_0A^T$ is diagonal. With R_{yy} diagonal, the relationship $\det(R_{yy}) \leq \Pi_i(R_{yy})_{ii}$ becomes an equality giving $C(y) = I(y,s)$ so the redundancy (1.9) is eliminated. [In reality, the redundancy (1.9) is a lower bound reflecting the fact that we chose probability distributions which take into account only second-order correlators. More complete knowledge of $P(s)$ would lower $I(x,s)$ and $I(y,s)$ and therefore increase $\mathcal{R}$.]

Where reducing $\mathcal{R}$ in equation 1.9 differs considerably from Barlow's hypothesis is in the manner of redundancy reduction when noise is significant. Under those circumstances, $\mathcal{R}$ in equation 1.9 is sizable, not because of correlations in the signal, but because much of the channel capacity is wasted carrying noise. Reducing equation 1.9 when the noise is large has the effect of increasing the signal-to-noise ratio. To do this the system actually *increases* correlations (more precisely increases the amplitude of the correlated signal relative to the noise amplitude), since correlations are what distinguish signal from noise. For large enough noise, more is gained by lowering the noise in this way than is lost by increasing correlations. For an intermediate regime, where signal and noise are comparable, our principle leads to a compromise solution, which locally accentuates correlations, but on a larger scale reduces them. All these facts can be seen by examining the properties of the explicit solution given below.

Before we proceed, it should also be noted that Linsker (1986) has hypothesized that the purpose of the encoding A should be to maximize the mutual information $I(y,s)$, subject to some constraints. This differs from the principle in this paper which focuses on lowering the output channel capacity while maintaining the *minimum* information needed by the organism. While both principles may be useful to gain insight into the purposes of neural processing in various portions of the brain, in the early visual processing, we believe that the primary evolutionary pressure has been to reduce output channel capacity. For example, due to much lower resolution in peripheral vision, the amount of information arriving at the retina is far greater than the information kept. It is difficult to believe that this design is a consequence of inherent local biological hardware constraints, since higher resolution hardware is clearly feasible, as seen in the fovea.

2 Explicit Solution

To actually minimize $\mathcal{R}$ we use a lagrange multiplier λ to implement the constraint (equation 1.10) and minimize

$$E\{A\} = C(y) - \lambda[I(y, s) - I(x + \delta, s)] \tag{2.1}$$

with respect to the transfer function A, where $C(y)$, $I(x, s)$, and $I(y, s)$ are given in equations 1.8, 1.4, and 1.5, respectively. One important property of $R[n, m]$ (in equation 1.2) that we shall assume is translation invariance, $R[n, m] = R[n - m]$, which is a consequence of the homogeneity of the ensemble of all visual scenes. We can take advantage of this symmetry to simplify our formulas by assuming $A[n, m] = A[n - m]$. With this assumption, the diagonal elements $(R_{yy})_{ii}$ in equation 1.7 are all equal and hence minimizing $C(y)$ is equivalent to minimizing the simpler expression $\text{Tr}(A R A^T)$.

Using the identity $\log(\det B) = \text{Tr}(\log B)$ for any positive definite matrix B, and replacing $C(y)$ by $\text{Tr}(A R A^T)$, equation 2.1 becomes

$$\begin{aligned}
E\{A\} = \quad & \frac{1}{N_\delta^2} \int_{-\pi}^{\pi} d\mathbf{w}\, A(\mathbf{w})\, R(\mathbf{w})\, A(-\mathbf{w}) \\
& - \frac{\lambda}{2} \left\{ \int_{-\pi}^{\pi} d\mathbf{w} \log \left[\frac{A(\mathbf{w})R(\mathbf{w})A(-\mathbf{w}) + N_\delta^2}{A(\mathbf{w})N^2 A(-\mathbf{w}) + N_\delta^2} \right] \right. \\
& \left. - \int_{-\pi}^{\pi} d\mathbf{w} \log \left[\frac{R(\mathbf{w}) + N_\delta^2}{N^2 + N_\delta^2} \right] \right\}
\end{aligned} \tag{2.2}$$

where all variables are defined in momentum space through the standard discrete two-dimensional fourier transform, for example,

$$A(\mathbf{w}) \equiv A(w_1, w_2) = \sum_{\mathbf{m}} e^{-i\mathbf{m} \cdot \mathbf{w}} A[\mathbf{m}]$$

It is straightforward to see from equation 2.2 that the optimal transfer function $A(\mathbf{w})$ satisfies the following quadratic equation:

$$[F(\mathbf{w})R(\mathbf{w}) + 1][F(\mathbf{w})N^2 + 1] = \frac{\lambda}{2} \frac{R_0(\mathbf{w})}{R(\mathbf{w})} \tag{2.3}$$

where we have defined $F(\mathbf{w}) = A(\mathbf{w}) \cdot A(-\mathbf{w})/N_\delta^2$. The fact that A appears only through F, is a manifestation of the original invariances of I and C under orthogonal transformations on the transfer function $A[\mathbf{m}]$, that is, under $A \to U A$ with $U^T U = 1$. Equation 2.3 has only one positive solution for F, which is given explicitly by

$$N^2 F = \frac{1}{2} \frac{R_0}{R} \left(1 + \sqrt{1 + \frac{2\lambda N^2}{R_0}} \right) - 1 \tag{2.4}$$

where λ is determined by solving $I(y, s) = I(x + \delta, s)$. After eliminating F the latter equation becomes

$$\int d\mathbf{w} \log \left(\sqrt{\frac{R_0}{2\lambda N^2}} + \sqrt{1 + \frac{R_0}{2\lambda N^2}} \right) = \frac{1}{2} \int d\mathbf{w} \log \left(\frac{R}{N^2} \cdot \frac{N^2 + N_\delta^2}{R + N_\delta^2} \right) \quad (2.5)$$

In general, equation 2.5 must be solved for λ numerically.

The fact that the transfer function A appears only through F leads to a multitude of degenerate solutions for A, related to each other by orthogonal transformations. What chooses among them has to be some principle of minimum effort in implementing such a transfer function. For example, some of the solutions are nonlocal (by local we mean a neighborhood of a point n on the input grid is mapped to the neighborhood of the corresponding point n on the output grid), so they require more elaborate hardware to implement; hence we examine local solutions. Among these is a unique solution satisfying $A(\mathbf{w}) = A(-\mathbf{w})$, which implies that it is rotationally invariant in coordinate space. We compare it to the observed retinal transfer function (ganglion kernel), known to be rotationally symmetric.

Since rotation symmetry is known to be broken at the simple cell level, it is significant that this formalism is also capable of producing solutions that are not rotationally invariant even when the correlation function is. It may be that the new features of the class of transfer functions at that level (for example, divergence factor) will lift the degeneracy in favor of the nonsymmetric solutions. (In fact, in one dimension we find solutions that break parity and look like one-dimensional simple cells kernels.)

The rotationally invariant solution is obtained by taking the square root of F in equation 2.4 (we take the positive square root, corresponding to on-center cells). In what follows, we examine some of its most important properties. To be specific, we parameterize the correlation function by a decaying exponential

$$R[\mathbf{n}, \mathbf{m}] = N^2 \delta_{\mathbf{n-m},0} + S^2 e^{-\|\mathbf{n-m}\|/D} \quad (2.6)$$

with D the correlation length measured in acuity units and S the signal amplitude. We have done numerical integration of equations 2.4 and 2.5 and determined $A[\mathbf{m}]$ for several values of the parameters. In Figure 2, we display one typical solution, which was obtained with $S/N = 2.0$, $D = 50$, and $N_\delta = 0.025$. In that figure, empty disks represent positive (excitatory), while solid disks represent negative (inhibitory) components of $A[\mathbf{m}]$. Also, the logarithm of the area of a disk is directly related to the amplitude of the component of $A[\mathbf{m}]$ at that location. As one can see, the solution has a strong and rather broad excitatory center with a weaker and more diffuse surround. A very significant feature of the theoretical profiles is their insensitivity to D (and to N_δ), which is necessary to account for the fact that the observed profiles measured in acuity units are similar in different species and at different eccentricities.
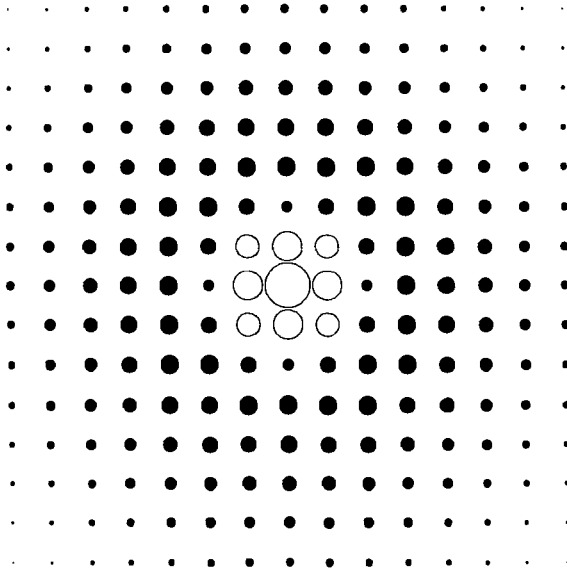
Figure 2: Optimal transfer function, $A[\mathbf{m}]$, for nondivergent linear codes, with $D = 50$, $S/N = 2$, and $N_\delta = 0.025$. Open disks denote positive (excitatory) components of $A[\mathbf{m}]$ while solid disks denote negative (inhibitory) components. The area of a disk is directly related to the logarithm of $A[\mathbf{m}]$ at that location.

To get more insight into this solution, let us qualitatively examine its behavior as we change S/N (for a detailed quantitative comparison with physiological data see Atick and Redlich 1990). For that, we find it more convenient to integrate out one of the dimensions (note this is not the same as solving the problem in one dimension). The resulting profile, corresponding to Figure 2, is shown in Figure 3b. In Figure 3, we have also plotted the result for two other values of S/N, namely for low and high noise regimes (Fig. 3a and c, respectively). These show that an interpolation is happening as S/N changes between the two extremes. Analytically, we can also see this from equation 2.4 for any R_0 by taking the limit $N/S \to 0$, where $A(\mathbf{w})$ becomes equal to

$$\sqrt{(\lambda/2 - 1)/R_0}$$

One recognizes that this is the square root of the solution one gets by carrying out *prediction* on the inputs, a signal processing technique which we advocated for this regime of noise (see also Srinivasan et al. 1982)
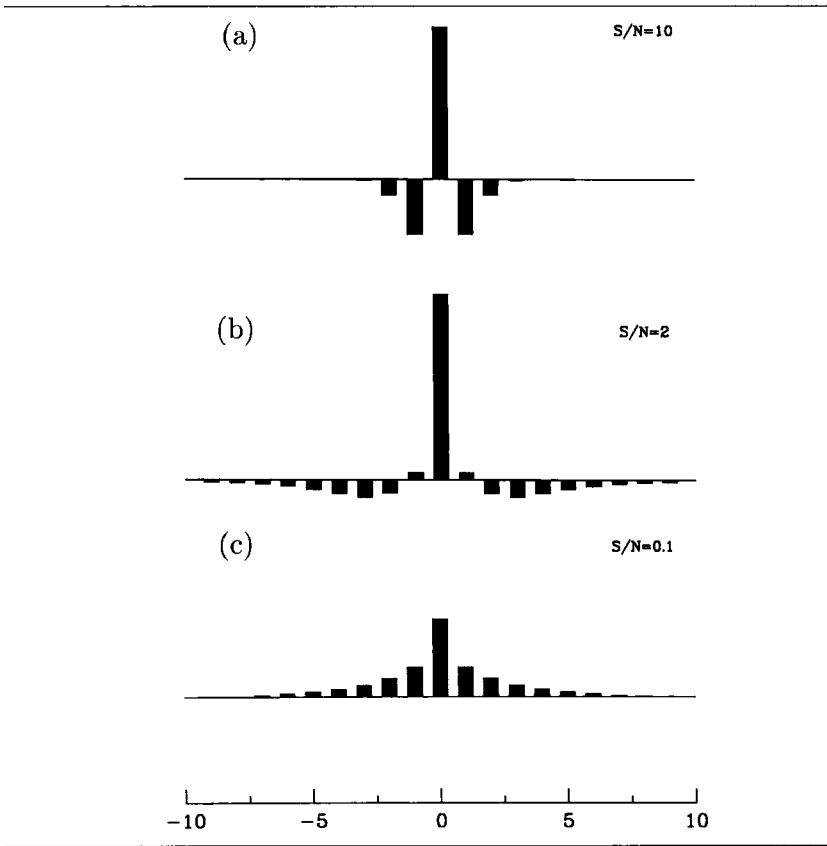
Figure 3: (a–c) Optimal solution at three different values of S/N. These profiles have been produced from the two-dimensional solution by summing over one direction and normalizing the resulting profile such that the height central point is equal to the center height in the two-dimensional solution.

as a redundancy reduction technique in our previous paper (Atick and Redlich 1989). The spatial profiles for the square root solution are very similar to the prediction profiles, albeit a bit more spread out in the surround region. This type of profile reduces redundancy by reducing the amount of correlations present in the signal.

In the other regime, where noise is very large compared to the signal, the solution for $A(\mathbf{w}) \sim (R_0/N^2)^{1/4}$ and has the same qualitative features as the *smoothing* solution (Atick and Redlich 1989) which in that limit is $A_{\text{smoothing}} = R_0/N^2$. Smoothing increases the signal to noise of the output and, in our earlier work, we argued that it is a good redundancy reducing technique in that noise regime. Moreover, in that work, we

argued that to maintain redundancy reduction at all signal-to-noise levels a process that interpolates between prediction and smoothing has to take place. We proposed a convolution of the prediction and the smoothing profiles as a possible interpolation (SPI-coding), which was shown to be better than either prediction or smoothing. In the present analysis, the optimal redundancy reducing transfer function is derived, and, although it is not identical to SPI-coding, it does have many of the same qualitative properties, such as the interpolation just mentioned and the overall center-surround organization.

The profiles in Figures 2 and 3 are very similar to the kernels of ganglions measured in experiments on cats and monkeys. We have been able to fit these to the phenomenological difference of gaussian kernel for ganglions (Enroth-Cugell and Robson 1966). The fits are very good with parameters that fall within the range that has been recorded. Another significant way in which the theory agrees with experiment is in the behavior of the kernels as S/N is decreased. In the theoretical profiles, one finds that the size of the center increases, the surround spreads out until it disappears, and finally the overall scale of the profile diminishes as the noise becomes very large. In experiment, these changes have been noted as the luminosity of the incoming light (and hence the signal to noise) is decreased and the retina adapts to the lower intensity (see, for example, Enroth-Cugell and Robson 1966). This active process, in the language of the current theory, is an adjustment of the optimal redundancy reducing processing to the S/N level.

In closing, we should mention that many of the techniques used to derive optimal encoding for the spatial properties of visual signals can be directly applied to temporal properties. In that case, for low noise the theory would lead to a reduction of temporal correlations, which would have the effect of taking the time derivative, while in the high noise case, the theory would lead to integration. Both types of processing play a significant role in visual perception, and it will be interesting to see how well they can be accounted for by the theory. Another issue that should be addressed is the question of how biological organisms evolved over time to have optimal redundancy reducing neural systems. In our previous paper, we discovered an anti-Hebbian unsupervised learning routine which converges to the *prediction* configuration and a Hebbian routine which converges to the *smoothing* profiles. We expect that there exist reasonably local learning algorithms that converge to the optimal solutions described here.

Acknowledgments

References

Atick, J. J., and Redlich, A. N. 1989. Predicting the ganglion and simple cell receptive field organizations from information theory. Preprint no. IASSNS-HEP-89/55 and NYU-NN-89/1.

Atick, J. J., and Redlich, A. N. 1990. Quantitative tests of a theory of early visual processing: I. Spatial contrast sensitivity profiles. Preprint no. IASSNS-HEP-90/51.

Barlow, H. B. 1961. Possible principles underlying the transformation of sensory messages. In *Sensory Communication*, W. A. Rosenblith, ed. M.I.T. Press, Cambridge, MA.

Barlow, H. B. 1989. Unsupervised learning. *Neural Comp.* **1**, 295–311.

Bodewig, E. 1956. *Matrix Calculus*. North-Holland, Amsterdam.

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

Linsker, R. 1986. Self-organization in a perceptual network. *Computer* (March), 105–117.

Linsker, R. 1989. An application of the principle of maximum information preservation to linear systems. In *Advances in Neural Information Processing Systems*, D. S. Touretzky, ed., Vol. 1, pp. 186–194. Morgan Kaufmann, San Mateo.

Orban, G. A. 1984. *Neuronal Operations in the Visual Cortex*. Springer-Verlag, Berlin.

Shannon, C. E., and Weaver, W. 1949. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana.

Srinivisan, M. V., Laughlin, S. B., and Dubs, A. 1982. Predictive coding: A fresh view of inhibition in the retina. *Proc. R. Soc. London Ser. B* **216**, 427–459.

Uttley, A. M. 1979. *Information Transmission in the Nervous System*. Academic Press, London.

This article has been cited by:

1. Shaul Druckmann, Dmitri B. Chklovskii. 2012. Neuronal Circuits Underlying Persistent Representations Despite Time Varying Activity. *Current Biology* **22**:22, 2095-2103. [CrossRef]

2. Wallace B. Thoreson, Stuart C. Mangel. 2012. Lateral interactions in the outer retina. *Progress in Retinal and Eye Research* **31**:5, 407-441. [CrossRef]

3. Mitsuhiko Hanada. 2012. Investigation of center–surround interaction in motion with reaction time for direction discrimination. *Vision Research* **59**, 34-44. [CrossRef]

4. Xaq Pitkow, Markus Meister. 2012. Decorrelation and efficient coding by retinal ganglion cells. *Nature Neuroscience* . [CrossRef]

5. Alfonso Renart, Mark C. W. van Rossum. 2012. Transmission of Population-Coded Information. *Neural Computation* **24**:2, 391-407. [Abstract] [Full Text] [PDF] [PDF Plus]

6. Eizaburo Doi, Michael S. Lewicki. 2011. Characterization of Minimum Error Linear Coding with Sensory and Neural Noise. *Neural Computation* **23**:10, 2498-2510. [Abstract] [Full Text] [PDF] [PDF Plus]

7. Gašper Tkačik, Aleksandra M Walczak. 2011. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter* **23**:15, 153102. [CrossRef]

8. Alexander G. Dimitrov, Aurel A. Lazar, Jonathan D. Victor. 2011. Information theory in neuroscience. *Journal of Computational Neuroscience* . [CrossRef]

9. G. Tkacik, J. S. Prentice, J. D. Victor, V. Balasubramanian. 2010. Local statistics in natural scenes predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences* **107**:42, 18149-18154. [CrossRef]

10. C. P. Ratliff, B. G. Borghuis, Y.-H. Kao, P. Sterling, V. Balasubramanian. 2010. Retina is structured to process an excess of darkness in natural scenes. *Proceedings of the National Academy of Sciences* **107**:40, 17368-17373. [CrossRef]

11. G. Tkacik, J. S. Prentice, V. Balasubramanian, E. Schneidman. 2010. Optimal population coding by noisy spiking neurons. *Proceedings of the National Academy of Sciences* **107**:32, 14419-14424. [CrossRef]

12. Mikhail Y. Lipin, Robert G. Smith, W. Rowland Taylor. 2010. Maximizing contrast resolution in the outer retina of mammals. *Biological Cybernetics* **103**:1, 57-77. [CrossRef]

13. A. Benoit, A. Caplier. 2010. Fusing bio-inspired vision data for simplified high level scene interpretation: Application to face motion analysis. *Computer Vision and Image Understanding* **114**:7, 774-789. [CrossRef]

14. A. Benoit, A. Caplier, B. Durette, J. Herault. 2010. Using Human Visual System modeling for bio-inspired low level image processing. *Computer Vision and Image Understanding* **114**:7, 758-773. [CrossRef]

15. S. X. Luo, R. Axel, L. F. Abbott. 2010. Generating sparse and selective third-order responses in the olfactory system of the fly. *Proceedings of the National Academy of Sciences* **107**:23, 10713-10718. [CrossRef]

16. Simo Vanni, Tom Rosenström. 2010. Local non-linear interactions in the visual cortex may reflect global decorrelation. *Journal of Computational Neuroscience* . [CrossRef]

17. Aleksandra M. Walczak, Gašper Tkačik, William Bialek. 2010. Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Physical Review E* **81**:4. . [CrossRef]

18. Y. S. Liu, C. F. Stevens, T. O. Sharpee. 2009. Predictable irregularities in retinal receptive fields. *Proceedings of the National Academy of Sciences* **106**:38, 16499-16504. [CrossRef]

19. Gašper Tkačik, Aleksandra Walczak, William Bialek. 2009. Optimizing information flow in small genetic networks. *Physical Review E* **80**:3. . [CrossRef]

20. Robert G. Smith, Narender K. Dhingra. 2009. Ideal observer analysis of signal quality in retinal circuits. *Progress in Retinal and Eye Research* **28**:4, 263-288. [CrossRef]

21. Anca Raˇdulescu, Kingsley Cox, Paul Adams. 2009. Hebbian errors in learning: An analysis using the Oja model. *Journal of Theoretical Biology* **258**:4, 489-501. [CrossRef]

22. Siwei Lyu, Eero P. Simoncelli. 2009. Nonlinear Extraction of Independent Components of Natural Images Using Radial Gaussianization. *Neural Computation* **21**:6, 1485-1519. [Abstract] [Full Text] [PDF] [PDF Plus] [Supplementary Content]

23. JAN WILTSCHUT, FRED H. HAMKER. 2009. Efficient coding correlates with spatial frequency tuning in a model of V1 receptive field organization. *Visual Neuroscience* **26**:01, 21. [CrossRef]

24. G. Tkacik, C. G. Callan, W. Bialek. 2008. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences* **105**:34, 12265-12270. [CrossRef]

25. Gašper Tkačik, Curtis Callan, William Bialek. 2008. Information capacity of genetic regulatory elements. *Physical Review E* **78**:1. . [CrossRef]

26. Michele Rucci. 2008. Fixational eye movements, natural image statistics, and fine spatial vision. *Network: Computation in Neural Systems* **19**:4, 253-285. [CrossRef]

27. Karl J. Friston, Klaas E. Stephan. 2007. Free-energy and the brain. *Synthese* **159**:3, 417-458. [CrossRef]

28. SUSMITA CHATTERJEE, DAVID K. MERWINE, FRANKLIN R. AMTHOR, NORBERTO M. GRZYWACZ. 2007. Properties of stimulus-dependent synchrony in retinal ganglion cells. *Visual Neuroscience* **24**:06. . [CrossRef]

29. C CLIFFORD, M WEBSTER, G STANLEY, A STOCKER, A KOHN, T SHARPEE, O SCHWARTZ. 2007. Visual adaptation: Neural, psychological and computational aspects. *Vision Research* . [CrossRef]

30. Nada Yousif, Mike Denham. 2007. The role of cortical feedback in the generation of the temporal receptive field responses of lateral geniculate nucleus neurons: a computational modelling study. *Biological Cybernetics* **97**:4, 269-277. [CrossRef]

31. S NIRENBERG, J VICTOR. 2007. Analyzing the activity of large populations of neurons: how tractable is the problem?. *Current Opinion in Neurobiology* **17**:4, 397-400. [CrossRef]

32. Odelia Schwartz, Anne Hsu, Peter Dayan. 2007. Space and time in visual context. *Nature Reviews Neuroscience* **8**:7, 522-535. [CrossRef]

33. JUYANG WENG, TIANYU LUWANG, HONG LU, XIANGYANG XUE. 2007. A MULTILAYER IN-PLACE LEARNING NETWORK FOR DEVELOPMENT OF GENERAL INVARIANCES. *International Journal of Humanoid Robotics* **04**:02, 281-320. [CrossRef]

34. J. D. Victor, E. M. Blessing, J. D. Forte, P. Buzás, P. R. Martin. 2007. Response variability of marmoset parvocellular neurons. *The Journal of Physiology* **579**:1, 29-51. [CrossRef]

35. Fred H. Hamker, Jan Wiltschut. 2007. Hebbian learning in a model with dynamic rate-coded neurons: An alternative to the generative model approach for learning receptive fields from natural scenes. *Network: Computation in Neural Systems* **18**:3, 249-266. [CrossRef]

36. Bijoy K. Ghosh, Ashoka D. Polpitiya, Wenxue Wang. 2007. Bio-Inspired Networks of Visual Sensors, Neurons, and Oscillators. *Proceedings of the IEEE* **95**:1, 188-214. [CrossRef]

37. Ryotaro Kamimura, Osamu Uchida, Seiki Hashimoto. 2007. Greedy network-growing algorithm with Minkowski distances. *International Journal of General Systems* **36**:2, 157-177. [CrossRef]

38. Ha Youn Lee, Mehran Kardar. 2006. Patterns and Symmetries in the Visual Cortex and in Natural Images. *Journal of Statistical Physics* **125**:5-6, 1243-1266. [CrossRef]

39. Valentin Dragoi, Mriganka Sur. 2006. Image Structure at the Center of Gaze during Free Viewing. *Journal of Cognitive Neuroscience* **18**:5, 737-748. [Abstract] [PDF] [PDF Plus]

40. Bruno B. Averbeck, Peter E. Latham, Alexandre Pouget. 2006. Neural correlations, population coding and computation. *Nature Reviews Neuroscience* **7**:5, 358-366. [CrossRef]

41. Li Zhaoping. 2006. Theoretical understanding of the early visual processes by data compression and data selection. *Network: Computation in Neural Systems* **17**:4, 301-334. [CrossRef]

42. Ryotaro Kamimura. 2005. Improving information-theoretic competitive learning by accentuated information maximization. *International Journal of General Systems* **34**:3, 219-233. [CrossRef]

43. K. Friston. 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**:1456, 815-836. [CrossRef]

44. M Hanada. 2004. Effects of the noise level on induced motion. *Vision Research* . [CrossRef]

45. References 483-514. [CrossRef]

46. K Friston. 2003. Learning and inference in the brain. *Neural Networks* **16**:9, 1325-1352. [CrossRef]

47. R Balboa. 2003. Power spectra and distribution of contrasts of natural images from different habitats. *Vision Research* **43**:24, 2527-2537. [CrossRef]

48. Ryotaro Kamimura, Fumihiko Yoshida. 2003. Teacher-directed learning: information-theoretic competitive learning in supervised multi-layered networks. *Connection Science* **15**:2, 117-140. [CrossRef]

49. B Vincent. 2003. Synaptic energy efficiency in retinal processing. *Vision Research* **43**:11, 1285-1292. [CrossRef]

50. C.A. Perez, C.A. Salinas, P.A. Estevez, P.M. Valenzuela. 2003. Genetic design of biologically inspired receptive fields for neural pattern recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)* **33**:2, 258-270. [CrossRef]

51. E Simoncelli. 2003. Vision and the statistics of the visual environment. *Current Opinion in Neurobiology* **13**:2, 144-149. [CrossRef]

52. A Turiel. 2003. Orientational minimal redundancy wavelets: from edge detection to perception. *Vision Research* **43**:9, 1061-1079. [CrossRef]

53. Yury Petrov, L. Zhaoping. 2003. Local correlations, information redundancy, and sufficient pixel depth in natural images. *Journal of the Optical Society of America A* **20**:1, 56. [CrossRef]

54. K Friston. 2002. Functional integration and inference in the brain. *Progress in Neurobiology* **68**:2, 113-143. [CrossRef]

55. Ryotaro Kamimura, Taeko Kamimura, Haruhiko Takeuchi. 2002. Greedy information acquisition algorithm: A new information theoretic approach to dynamic information acquisition in neural networks. *Connection Science* **14**:2, 137-162. [CrossRef]

56. J TROY, T SHOU. 2002. The receptive fields of cat retinal ganglion cells in physiological and pathological states: where we are after half a century of research. *Progress in Retinal and Eye Research* **21**:3, 263-302. [CrossRef]

57. Karl Friston. 2002. BEYOND PHRENOLOGY: What Can Neuroimaging Tell Us About Distributed Circuitry?. *Annual Review of Neuroscience* **25**:1, 221-250. [CrossRef]

58. I Fine. 2002. Visual function before and after the removal of bilateral congenital cataracts in adulthood. *Vision Research* **42**:2, 191-210. [CrossRef]

59. Ryotaro Kamimura, Taeko Kamimura, Osamu Uchida. 2001. Flexible feature discovery and structural information control. *Connection Science* **13**:4, 323-347. [CrossRef]

60. R Lestienne. 2001. Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Progress in Neurobiology* **65**:6, 545-591. [CrossRef]

61. J van Hateren. 2001. Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Research* **41**:14, 1851-1865. [CrossRef]

62. R Balboa. 2001. Occlusions contribute to scaling in natural images. *Vision Research* **41**:7, 955-964. [CrossRef]

63. S. Raudys, A. Saudargiene. 2001. First-order tree-type dependence between variables and classification performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**:2, 233-239. [CrossRef]

64. G Brown. 2001. Independent component analysis at the neural cocktail party. *Trends in Neurosciences* **24**:1, 54-63. [CrossRef]

65. Ryotaro Kamimura, Taeko Kamimura, Thomas R. Shultz. 2001. Information Theoretic Competitive Learning and Linguistic Rule Acquisition. *Transactions of the Japanese Society for Artificial Intelligence* **16**, 287-298. [CrossRef]

66. Christoph Zetzsche, Gerhard Krieger. 2001. Nonlinear mechanisms and higher-order statistics in biological vision and electronic image processing: review and perspectives. *Journal of Electronic Imaging* **10**:1, 56. [CrossRef]

67. Christoph Zetzsche, Gerhard KriegerIntrinsic Dimensionality 403-448. [CrossRef]

68. Emilio Salinas, L. F. Abbott. 2000. Do Simple Cells in Primary Visual Cortex Form a Tight Frame?. *Neural Computation* **12**:2, 313-335. [Abstract] [PDF] [PDF Plus]

69. M Wainwright. 1999. Visual adaptation as optimal information transmission. *Vision Research* **39**:23, 3960-3974. [CrossRef]

70. Peter Adorján,, Christian Piepenbrock,, Klaus Obermayer,. 1999. Contrast Adaptation and Infomax in Visual Cortical Neurons. *Reviews in the Neurosciences* **10**:3-4, 181-200. [CrossRef]

71. Nick Chater. 1999. The Search for Simplicity: A Fundamental Cognitive Principle?. *The Quarterly Journal of Experimental Psychology Section A* **52**:2, 273-302. [CrossRef]

72. Friedrich O. Huck, Carl L. Fales, Rachel Alter-Gartenberg, Stephen K. Park, Zia-ur Rahman. 1999. Information-theoretic assessment of sampled imaging systems. *Optical Engineering* **38**:5, 742. [CrossRef]

73. D. Obradovic, G. Deco. 1998. Information Maximization and Independent Component Analysis: Is There a Difference?. *Neural Computation* **10**:8, 2085-2101. [Abstract] [PDF] [PDF Plus]

74. Alexander Dimitrov, Jack D. Cowan. 1998. Spatial Decorrelation in Orientation-Selective Cortical Cells. *Neural Computation* **10**:7, 1779-1795. [Abstract] [PDF] [PDF Plus]

75. H. Attias, C. E. Schreiner. 1998. Blind Source Separation and Deconvolution: The Dynamic Component Analysis Algorithm. *Neural Computation* **10**:6, 1373-1424. [Abstract] [PDF] [PDF Plus]

76. Daniel L. Ruderman, Thomas W. Cronin, Chuan-Chin Chiao. 1998. Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A* **15**:8, 2036. [CrossRef]

77. Daniel L. Ruderman. 1997. Origins of scaling in natural images. *Vision Research* **37**:23, 3385-3398. [CrossRef]

78. D. Osorio, N. Justin Marshall, Thomas W. Cronin. 1997. Stomatopod photoreceptor spectral tuning as an adaptation for colour constancy in water. *Vision Research* **37**:23, 3299-3309. [CrossRef]

79. Bruno A. Olshausen, David J. Field. 1997. Sparse coding with an overcomplete basis set: A strategy employed by V1?. *Vision Research* **37**:23, 3311-3325. [CrossRef]

80. Anthony J. Bell, Terrence J. Sejnowski. 1997. The "independent components" of natural scenes are edge filters. *Vision Research* **37**:23, 3327-3338. [CrossRef]

81. J.H. van Hateren. 1997. Processing of natural time series of intensities by the visual system of the blowfly. *Vision Research* **37**:23, 3407-3416. [CrossRef]

82. M Andrade. 1997. Receptive Field Map Development by Anti-Hebbian Learning. *Neural Networks* **10**:6, 1037-1052. [CrossRef]

83. G. Deco, L. Parra. 1997. Non-linear Feature Extraction by Redundancy Reduction in an Unsupervised Stochastic Neural Network. *Neural Networks* **10**:4, 683-691. [CrossRef]

84. Christoph Zetzsche, Erhardt Barth, Gerhard Krieger, Bernhard Wegmann. 1997. Neural network models and the visual cortex: the missing link between orientation selectivity and the natural environment. *Neuroscience Letters* **228**:3, 155-158. [CrossRef]

85. A VANDERSCHAAF, J VANHATEREN. 1996. Modelling the power spectra of natural images: Statistics and information. *Vision Research* **36**:17, 2759-2770. [CrossRef]

86. Suzanna Becker, Mark Plumbley. 1996. Unsupervised neural network learning procedures for feature extraction and classification. *Applied Intelligence* **6**:3, 185-203. [CrossRef]

87. Zhaoping Li. 1996. A Theory of the Visual Motion Coding in the Primary Visual Cortex. *Neural Computation* **8**:4, 705-730. [Abstract] [PDF] [PDF Plus]

88. Suzanna Becker. 1996. *Network: Computation in Neural Systems* **7**:1, 7-31. [CrossRef]

89. Joseph J Atick, Paul A Griffin, A Norman Redlich. 1996. The vocabulary of shape: principal shapes for probing perception and neural response. *Network: Computation in Neural Systems* **7**:1, 1-5. [CrossRef]

90. Stefano Panzeri, Alessandro Treves. 1996. *Network: Computation in Neural Systems* **7**:1, 87-107. [CrossRef]

91. Michael Haft, Martin Schlang, Gustavo Deco. 1995. Information theory and local learning rules in a self-organizing network of Ising spins. *Physical Review E* **52**:3, 2860-2871. [CrossRef]

92. J. Gerard Wolff. 1995. Computing as compression: An overview of the SP theory and system. *New Generation Computing* **13**:2, 187-214. [CrossRef]

93. Marco Idiart, Barry Berk, L. F. Abbott. 1995. Reduced Representation by Neural Networks with Restricted Receptive Fields. *Neural Computation* **7**:3, 507-517. [Abstract] [PDF] [PDF Plus]

94. Gustavo Deco, Bernd Schürmann. 1995. Learning time series evolution by unsupervised extraction of correlations. *Physical Review E* **51**:3, 1780-1790. [CrossRef]

95. G Deco. 1995. Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures. *Neural Networks* **8**:4, 525-535. [CrossRef]

96. G DECO. 1995. Linear redundancy reduction learning. *Neural Networks* **8**:5, 751-755. [CrossRef]

97. Bartlett W. Mel. 1994. Information Processing in Dendritic Trees. *Neural Computation* **6**:6, 1031-1085. [Abstract] [PDF] [PDF Plus]

98. Zhaoping Li, Joseph J. Atick. 1994. Toward a Theory of the Striate Cortex. *Neural Computation* **6**:1, 127-146. [Abstract] [PDF] [PDF Plus]

99. S LAUGHLIN. 1994. Matching coding, circuits, cells, and molecules to signals: General principles of retinal design in the fly's eye. *Progress in Retinal and Eye Research* **13**:1, 165-196. [CrossRef]

100. Dawn M. Adelsberger-Mangan, William B. Levy. 1993. Adaptive synaptogenesis constructs networks that maintain information and reduce statistical dependence. *Biological Cybernetics* **70**:1, 81-87. [CrossRef]

101. A. Norman Redlich. 1993. Supervised Factorial Learning. *Neural Computation* **5**:5, 750-766. [Abstract] [PDF] [PDF Plus]

102. A. Norman Redlich. 1993. Redundancy Reduction as a Strategy for Unsupervised Learning. *Neural Computation* **5**:2, 289-304. [Abstract] [PDF] [PDF Plus]

103. Joseph J. Atick, A. Norman Redlich. 1993. Convergent Algorithm for Sensory Receptive Field Development. *Neural Computation* **5**:1, 45-60. [Abstract] [PDF] [PDF Plus]

104. John G. Taylor, Mark D. PlumbleyInformation Theory and Neural Networks **51**, 307-340. [CrossRef]

105. M PLUMBLEY. 1993. Efficient information transfer and anti-Hebbian neural networks. *Neural Networks* **6**:6, 823-833. [CrossRef]

106. J ATICK, Z LI, A REDLICH. 1993. What does post-adaptation color appearance reveal about cortical color representation?. *Vision Research* **33**:1, 123-129. [CrossRef]

107. Dawn M. Adelsberger-Mangan, William B. Levy. 1992. Information maintenance and statistical dependence reduction in simple neural networks. *Biological Cybernetics* **67**:5, 469-477. [CrossRef]

108. R Monasson. 1992. *Journal of Physics A: Mathematical and General* **25**:13, 3701-3720. [CrossRef]

109. H. P. Snippe, J. J. Koenderink. 1992. Information in channel-coded systems: correlated receivers. *Biological Cybernetics* **67**:2, 183-190. [CrossRef]

110. W SOFTKY, D KAMMEN. 1991. Correlations in high dimensional or asymmetric data sets: Hebbian neuronal processing#. *Neural Networks* **4**:3, 337-347. [CrossRef]