

# A Synthesis of the Pribram Holonomic Theory of Vision With Quantum Associative Nets After Pre-Processing Using I.C.A. and Other Computational Models

Mitja Peruš

Institute BION, Stegne 21, SI-1000 Ljubljana, Slovenia

Fax/Phone +386-1-513-1147 - [mitja.perus@uni-lj.si](mailto:mitja.perus@uni-lj.si) - <http://www.bion.si/mitja.htm>

## Abstract

Statistically-Independent Component Analysis (ICA) and sparseness-maximization net are models which maximally preserve information ("infomax"). Research of relevance of these algorithms for modeling image-processing in V1 is reported in comparison with the Holonomic Brain Theory by Pribram which advocates dendritic processing and its connection to quantum processing. "Infomax" models are presented and discussed as a possible early-processing gateway to higher visual processing involving quantum associative nets (Peruš, 2000) and attractor dynamics.

**Keywords:** Independent Component Analysis (ICA), quantum associative networks, early vision, infomax, sparse coding

## 1 Introduction to the Holonomic Theory and Early Image Processing

V1. The striate cortex (V1), receiving inputs from retina over lateral geniculate nucleus (LGN), serves as the central brain area for image processing, before its outputs are projected to higher visual areas (V2, V3, V4, MT, ITC, PPC and others) (e.g., De Yoe & Van Essen, 1988; Tootell *et al.*, 1998). The form of V1 receptive fields are *Gabor wavelets* which minimize the uncertainty in phase space of space-time and spatial and temporal frequencies, and thus maximize information. Gabor wavelets serve as "patterns" which interfere in order to realize a Hebbian-like learning process incorporating phase-differences. The *holonomic brain theory* (HBT) by Pribram (1991, 1999) proposes that holography-like image processing and related Gabor transforms are implementable in synapto-dendritic and/or quantum substrates. We will try to complement HBT with new relevant computational models, based mainly on neural-net-like quantum or field computation, or ICA.

**Holonomic brain theory.** Pribram's HBT suggests biologically-plausible computational mechanisms for a triple convolution process along the visual pathway *retina – LGN – striate cortex*. At each of these three stages, inputs from the previous stage are convoluted with the receptive fields of neurons receiving those input at the present stage (Pribram & Carlton, 1986).

At the microscopic level, HBT considers synapto-dendritic networks (where synapses and dendrites belong to neurons at the above-mentioned three stages of the visual pathway, especially V1) as the most relevant information-processing executive

level. So, image processing in V1 using Gabor wavelets may be implemented by interacting dendritic polarization-fields arising from spine-produced oriented electrical dipoles. In retina and LGN a similar image *pre*-processing is done by convoluting inputs to a neuron with its “Mexican-hat”-shaped “filter”. Later, V1 cells use Gabor wavelets as the receptive-field-based “filters”.

HBT postulates quantum-rooted processes realizing *conscious* perception, not merely information dynamics. HBT considers processes at the synapto-dendritic-net level more relevant than processes at the usual neural-net level. This is supported by physiological experimental evidence (Pribram, 1991, 1995). Dendritic processes are also more flexible, faster, realize a higher rate of connectivity, or interaction, and greater parallelism than conventional nerve-signal processing (summation in soma and axonal propagation of impulses). Dendritic field processes are, over subcellular structures like cytoskeleton, “more directly connected” with quantum holism and unity which seems necessary for binding of perceptual features into an experiential whole.

**HBT-compatible models.** Neural-net models usually incorporate “amplitude”-information (quantified using real-valued numbers), but not phase-information (encoded in complex-valued variables), because they do not contain wave dynamics. However, for successful image processing, it is useful to incorporate *phase-information* into amplitude-correlation (second-order statistics) models like the Hopfield model or Principal Component Analysis (PCA). Latter associative memory models use (Hebbian) correlation matrices. But there are some modern neural-net-based models which successfully process phase-information in addition to amplitude-information and share many features with HBT. Such HBT-compatible models are:

- *Independent Component Analysis* (ICA) by Bell & Sejnowski (1995, 1996, 1997), and a similar sparse-coding model by Olshausen & Field (1996a,b; 1997);
- *field computation* model by MacLennan (1999; in Pribram, 1993; in Wang *et al.*, 1998) which also uses Gabor wavelets and many physics-inspired techniques;
- *Holographic Neural Technology* (HNeT) by Sutherland (1990) and AND Corp.: an application-effective neurobiology- and holography-inspired artificial model (in other words, simulated holography-based neurocomputation) which was recently renamed to *Holographic / Quantum Neural Technology*, because analogies between mathematical models of holography, associative neural networks and quantum fields were recently systematically listed and discussed (Gould, Peruš in Pykkänen & Pykkö, 1995; Peruš, 1996, 1998; Schempp in Pribram, 1993; Marcer & Schempp, 1998);
- *quantum neurodynamics* (QND) by Jibu & Yasue (in Pribram, 1991, 1993), and in another (artificial) model by Dawes (in Pribram, 1993);
- *quantum associative network* model (QAN) by Peruš (2000; in Wang, 1998): a quantum information-processing “algorithm”, inspired by the associative neural-net models of Hopfield and Haken, rewritten to quantum-mathematical formalism.

The QAN model was presented in Peruš (2000) (originally: Peruš in Wang *et al.*, 1998). It is *fundamental* in the sense of informational physics. To construct QAN from Hopfield-like simulated neural nets, the real-valued “neural” variables have been translated to analogous quantum complex-valued (phase-information-carrying) variables (Peruš, 1996, 1997). The translation to quantum formalism was deliberately and purposely done in the most “natural” way, i.e. so that the neural-net-like associative

information-processing can be implemented in a usual quantum-physical system, and hypothetically inside some brain tissues, not necessarily in some artificial quantum-based device. This characteristic distinguishes the QAN model from all similar quantum-computer models (Bonnell & Papini, 1997; Zak & Williams, 1998; Behrman *et al.*, Ventura in Wang *et al.*, 1998; Dawes in Pribram, 1993; Chrisley in Pylkkänen & Pylkkö, 1995) which are, at least in part, hypothetically-implementable only using some artificial devices or procedures.

The ICA model resembles the receptive fields of so-called simple cells in V1 (Bell & Sejnowski, 1997). The “independent components”, selected by ICA-filters, are localized and oriented, as was also experimentally found for those simple cells which were denoted as “edge filters”. The unsupervised ICA learning algorithm, based on information maximization (“infomax”, ch. 2) which is agreement with HBT, provides maximal statistical independence of resulting neural states (outputs). This is achieved when their probability distribution gets factorized. It is equivalent to getting mutual information between the output-states zero. Here is a connection between ICA and HBT, since Gabor wavelets realize such optimizing characteristics. Another connection is the fact that ICA encodes and exploits the phase-information in addition to the amplitude-information. Actually, as ICA simulations show (chs. 3, 4, 5), the amplitude-information is less necessary for good image recall than the phase-information. This is similar to HNeT processing and to findings by MacLennan (in Wang *et al.*, 1998) for some Gabor-based field-computation models. In HNeT phases are considered as the main output; amplitudes just measure the rate of importance, confidence (or reliability, accuracy, respectively) or urgency of the phase-output.

**Table 1** compares characteristics of chosen models regarding vision research:

<b>Table 1</b>	<b>H.Ne.T</b>	<b>Q.A.N.</b>	<b>I.C.A.</b>	<b>Field computing</b>
<i>effectiveness</i>	very effective	effective, potentially very effective	very effective	a general model with potentially very effective “sub-branches”
<i>biological plausibility</i>	biol. plausible at the fundamental level only	biol. plausible at the fundamental level only	biol. implausible mechanism, but plausible output	biol. plausible at the fundamental level only
<i>possible quantum substrate</i>	indirect, but formally similar to Q.A.N.	direct quantum implementation	quantum implementation not yet known	quantum implem. indirect, partially direct
<i>main strength</i>	very applicative and developed	fundamentally quantum, “natural”	fits experiments on receptive field profiles	a general model with a concrete vision sub-model
<i>main weakness</i>	a mixture of natural and artificial features	limited to assoc. memory and pattern recognit.	algorithm’s biol. implementation unknown	(consciousness still missing)

Phase-information is necessary for detection of *edges* of perceived objects, because edges are situations where many sine-waves of different frequencies “get in phase”, i.e. they are all aligned in phase and sum together precisely where the edge lies.

Such a coherent constructive interference is not realizable without wavelets. Therefore ICA successfully detects edges, but the usual Hebbian amplitude-correlation methods (without phase-differences incorporated) do not. ICA gives biologically-plausible *results* (outputs), but its algorithm has not yet been found to be *implemented* in a biologically-plausible way. Our paper presents these issues in detail from a HBT view.

**From ICA to QAN.** After ICA-like *image pre-processing* by early visual perception (ch. 6), *object perception* seems to be realized by attractor-nets, and *conscious* associative processes seem to be grounded by quantum associative processes in a QAN-like system (ch. 7) which are strongly shaped by processes in neural circuits.

## 2 Maximal Preservation of Information (“Infomax”)

**From second- to higher-order statistics.** ICA is supposed to be important for cortical image processing because of taking into account the statistics of input data which is also of higher order than the second order. *Correlation* and *convolution*, for example, are “learning rules” of *second order* (i.e., based on multiplication of *two* state-variables). Such mathematical expressions have been named *Hebbian* (directly or in a generalized sense, e.g. the *covariance* learning rule and *PCA*).

**Phases needed for edges.** Phase-information is needed for successful approximation of V1-filters which were found to be *localized*, *oriented* and *band-pass* (i.e., selective to structure at different spatial scales). Local angle of orientation is described by local phase. Such filters are needed to trace segments of edges which are themselves oriented. Experiments show that individual *simple* cells of V1 with their specific receptive fields act as such filters, i.e. as selectors or edge-segments by having maximal response to specifically oriented stimuli.

An edge, a notable image-element, manifests specific relationships among many pixels being encoded in neurons or receptors, not only two (neighboring) ones. Second-order statistics (as in PCA) is sensitive *only to pair-wise relationships*, like correlations encoded in the Hebb rule. Higher-order statistics (as in ICA) is sensitive to multi-neuronal relationships, reflecting *multi-pixel gestalt-structures*, and thus goes beyond the two-pixel (or two-neuron, respectively) relations of PCA. Bell & Sejnowski (1996) show that the Hebbian models reflect only the amplitude spectrum of the signal and ignore the phase spectrum where most of the relevant *local* coincidences in natural signals take place. An edge in an image is a coincidence in the phase spectrum, since if one Fourier-analyses it, one would see many sine-waves of different frequencies, all aligned in phase where the edge occurred. If one Fourier-transforms (FT) the image, and then performs inverse FT, with (1) amplitude or (2) phase removed, the starting image will be reconstructed in case (1) badly, and in case (2) well enough. *Phases are thus necessary for image processing; “amplitudes”, on the other hand, are useful but not necessary.*

**Redundancy reduction.** Visual processing also has to be maximally efficient in the sense of *maximally preserving information by using as little structure as possible*. So, information has to be optimally condensed, not by reducing quality and relevance, but only by reducing quantity (where it is unnecessary), or by eliminating redundancy (i.e., double or multiple coding), respectively. This would enable faster processing and

decrease the number of errors because of getting trapped into spurious (irrelevant) attractors.

It turns out that an "infomax"-net is realizable if neurons have receptive fields shaped like Gabor functions. Computer simulations were done where model-neurons (as filters) resembled the Gabor-like receptive-field form which enables "infomax" and orientation- or edge-selectivity at the same time. Gabor wavelets (Daugman, 1985) are functions which have oriented stripes as seen from above. This means that the response of a simple cell with such a receptive field will be the biggest if stimulated by an edge with such an orientation.

**Two "infomax" methods.** To realize these characteristics of the primary visual system, *ICA* and *sparseness-maximization network* turned out to be the most successful methods. *ICA* is a new PDP method for searching unknown sources in complex data-structures, i.e. so-called blind separation into independent components. Starting from earlier *ICA*-models like Comon (1994), *ICA* was developed for vision modeling mainly by Bell & Sejnowski (1995, 1996, 1997), therefore it will be abbreviated by *BS*. *BS*-network results emphasized statistical independence more than sparseness of coding. The second method, that also resembled Gabor-like "infomax" orientation-filters, was developed by Olshausen & Field (1996a,b), therefore abbreviated by *OF*. *OF*-network achieved such results by emphasizing sparse coding over information maximization. *Sparse coding means that information is encoded by a small minority of neurons in a net.* These coding neurons are active ("non-zero") only, while all others are passive ("zero") in the end, although they also participate during the process of learning.

*BS* and *OF* demonstrate how their phase-based methods transcend the limitations of *PCA* and Hebbian learning where neuronal configuration converges toward a basis of pattern-vectors (attractors) that are mutually orthogonal. Pairs of *PCA* output-patterns are therefore decorrelated ("have nothing in common"), but are not statistically independent, because the structure of input-data is usually not mutually orthogonal, especially not in natural scenes (Harpur & Prager, 1996).

**Statistically-independent coding (ICA).** Statistical independence is more than just decorrelation which ensures that stored images do not disturb each other when recalled, and that they can be selectively recognized in novel, similar circumstances. Decorrelation, achieved by *PCA*, is not sufficient in the case of natural-scene images which are rarely orthogonal. As numerous computer simulations of ours and of other modelers have shown, Hebbian nets and *PCA* still have some success in quasi-orthogonal cases, but in non-orthogonal cases they generally fail much earlier than *ICA*. Such cases also usually demand non-linear processing, like with edge problem.

*ICA* solves more difficult problems of vision by decomposing the natural image into its statistically independent component-images. The natural scene usually contains many intermixed features and parts of objects which cannot be stiffly categorized into orthogonal eigen-images, but share the data-structure. *ICA* achieves this more subtle decomposition on the level of finer data-statistics. Statistical independence means mathematically that the probability distribution of input-data can be factorized, i.e. written as a multiple product of probability distributions of some specific sub-sets of data. These sub-sets are called statistically independent components. Decomposition of the probability distribution (PD) of a natural image into independent factors is equivalent to decreasing their mutual information toward zero. *BS* and *OF* recognize

that their algorithms in practice do *not* achieve complete statistical independency, but merely try to get close to it as much as possible. (Such an practical maximum will from now on be meant when speaking about “independent components”).

**Table 2** summarizes the comparison of ICA and sparseness-maximization-net (SMN) versus PCA and Hebbian neural nets:

P.C.A., Hebbian neural nets	I.C.A. (by B.&S.), S.M.N. (by O.&F.)
second-order statistics: bi-linear coupling, e.g. correlation, convolution, covariance	higher-order statistics: multi-linear coupling and/or non-linear dependence
decorrelation into orthogonal eigen-images	statistical separation of non-orthogonal mixtures
optimal for gaussian input-data	optimal for super-gaussian data
“static activities of units (neurons)”	“oscillatory activities of units”
amplitude-information (activity rates) only	plus phase-information
edges and V1 receptive fields – problematic	edges detected; recept. fields reproduced
filters distributed, not Gabor-like	filters sparsely-distributed, Gabor-like
algorithm up-to-now satisfactorily efficient	algorithm more efficient than P.C.A.
implementation biologically more plausible	biological implementation unknown
outputs biologically not plausible (in vision)	outputs biologically plausible (in vision)

**Blind separation of images.** BS- and OF-nets search for the *amplitude-coefficients* or *codes* (“causes” or “sources”),  $s$ , of the *independent components*,  $Y$ , of an image  $X$ .  $X$  is in general representable by a linear combination of independent image-components  $Y$  which form the *columns* of a fixed matrix  $A$ . If we represent this relationship with equation  $X = A S$ , then each vector-component of vector  $S$  acts as a *weight* of the corresponding basis-function  $Y$ . Each pixel (denoted by index  $i$ ) of the image  $X$  can so be described as follows:

$$x_i = \sum_{k=1}^p a_{ik} s_k = a_{i1} s_1 + a_{i2} s_2 + a_{i3} s_3 + \dots + a_{ip} s_p \quad (1).$$

Each index  $i$  or  $j$  corresponds to a Cartesian coordinate-pair, like  $(x, y)$ , of the 2D image-space. For all pixels  $(x, y)$ ,  $X = A S$  is (in our notation using elements of vectors  $X, S$  (its elements are  $s$ ) and  $Y$ , hidden as column-vector in matrix  $A$ ) translated to

$$X(x, y) = \sum_{k=1}^p s_k Y_k(x, y) \quad (2).$$

If we “look in the opposite direction”, matrix  $W$  decomposes an image  $X$  into independent components  $Y$ , represented by their amplitudes or codes  $s$  (which are mathematically represented by individual vector-components, denoted by index  $k$ , of the vector  $S$ ). Ideally, we would get  $S = W X$ , but in practice we might get  $S' = W X$  (e.g., comparing to  $s$ , vector-components of  $S'$  might be in different order and rescaled). Matrix  $W$  should separate the image in such a way that codes  $s$  can encode as much information about image  $X$  as possible, but with as little redundancy as possible, so that we could reconstruct the initial image  $X$  as much as possible with available limited neural structure. Ideally,  $W$  is *inverse matrix* of  $A$ , so that  $WA=I$ , where  $I$  is the identical matrix which preserves or recovers the initial state (here  $S$ ), although making a (composite) transformation. In such a case, it is thus  $S' = S$ , and the reconstruction of the image is perfect.

The *filters*,  $\phi$ , which extract the independent codes  $S$ , are hidden in *rows* of matrix  $W$ . Filters  $\phi$  differ somewhat from independent components  $Y$ , although filters recover the amplitudes of independent components. This is because independent components  $Y$  of the image *overlap* strongly (they are not orthogonal), but filters concentrate their action to areas in spatial-(temporal)-frequency-space where the overlap of  $Y$  is minimal. With such filters, the amplitude of  $Y$  can be got without interference with other components  $Y'$ . Independent components  $Y$  can be considered as building blocks of the image  $X$ ; their filters  $\phi$ , on the other hand, are used for analyzing the image by determining how strongly each building block is present in it. Thus, it is more appropriate to say that the independent-component-*filters* ( $\phi$ , not  $Y$ ) correspond to V1-cells acting as Gabor-like “edge filters” (van Hateren & Ruderman, 1998).

In this “infomax” approximation, the role of primary visual system (from retina up to and including V1) is to find the proper *filter-matrix*  $W$  with the above-described properties.  $W$  is an extremely simplified and summarized mathematical description of results of the *cascade convolution processing weighted with receptive fields* (these are the filters!).

### 3 I.C.A. by Bell and Sejnowski (BS)

**ICA-“infomax” by BS.** Bell & Sejnowski (1995, sec. 4) show precisely under what conditions an “infomax” computational algorithm minimizes the mutual information between resulting outputs called independent components. Namely, under specific conditions, mutual information  $I(a,b)$  between output-components  $a$  and  $b$  is *minimized* while overall information is *maximized*<sup>1</sup> or equivalently joint-entropy  $H(a,b)$  is maximized. This is evident from a well-known equation of information theory:  $H(a,b) = H(a) + H(b) - I(a,b)$ . Joint entropy  $H(a,b)$  is maximal if individual entropies  $H(a)$  and  $H(b)$  are maximized and mutual information  $I(a,b)$  is minimized. It was also proven in Bell & Sejnowski (1995, ch. 2, p. 1131) that the mutual information  $I(X,Y)$  between inputs  $X = (a', b', \dots)$  and outputs  $Y = (a, b, \dots)$  can be maximized by maximizing the entropy  $H(Y)$  of the outputs  $(a, b, \dots)$  alone.

As long as  $a$  and  $b$  can interfere, absolute minimum of  $I(a,b)$  can not be achieved for sure, therefore the probability distribution  $PD$  cannot be decomposed into independent factors, i.e. can not be written as a product of individual probability distributions  $PD(a)$  and  $PD(b)$ . Unwanted interference is eliminated, thus giving absolute minimum of  $I(a,b)$  and consequently realizing statistical independence where  $PD = PD(a) \times PD(b)$ , if  $PD$  of inputs are *super-gaussian*. Super-gaussian, as many natural signals are, means that  $PD$  has a sharper peak and longer tail than gaussian (“bell-shaped”). PCA gives optimal and complete results in the case of gaussian  $PD$ , but super-gaussian  $PD$  (where the bell-shaped function of statistics is more thin, except at the bottom where it is broader than the “normal bell”) is optimal for ICA. In statistical

<sup>1</sup> The “infomax” details are very subtle. When speaking about maximal overall information, we mean maximal information-*preservation* while transforming input-data into memory-codes or outputs, but at the same time being as “economical” as possible. So, input-output information should be preserved as much as possible (“infomax”), but at the same time all the redundant (multiple, unnecessary, common) data should be eliminated (i.e., mutual information in outputs should be minimized).

language, the kurtosis (the fourth-order standardized cumulant measuring deviation from normal gaussian PD) must be positive (but see Baddeley, 1996). In the opposite case which is undesired: negative kurtosis, i.e. sub-gaussian PD, means interference between image-components and so they will not get independent by applying an "infomax" algorithm for image-processing. Such rare negative-kurtosis images would be problematic while attempting redundancy-reducing image-recognition using ICA.

In simulations of Bell & Sejnowski (1995, 1996, 1997), receptive fields develop into Gabor-like wavelets which, seen from above, have stripes that are oriented like potential edges in the input-image are. So, Gabor stripes are parallel to edges. Not just orientation, but other subtle details of the edge-structure are also encoded in such a "Gabor filter" belonging to a simulated V1 simple cell and its tree-like connections. It will now be presented how BS-net incorporates one of many possible "hardwares and softwares" for computational modeling of evolution of edge-detecting V1 simple cells together with their receptive fields. Another option, OF-net, not based primarily on ICA, will be presented later.

**ICA-algorithm by BS.** BS-net maximizes information by maximizing the final joint entropy  $H(Y) = H[g(WX+B)]$ ;  $Y = g(U)$ ;  $U = WX+B$ . Input image is denoted by  $X$ , ICA-output is  $Y$ ,  $W$  is the matrix of ICA-filters,  $B$  is the so-called bias. The activation function  $g$  is the sigmoid function which enables optimal flow of information. Joint entropy  $H(Y)$  is maximized by performing stochastic gradient ascent, i.e. the elements of the filter-matrix  $W$  are changed in such a way that  $H(Y)$  increases as much as possible. Bell & Sejnowski (1995, ch. 2) derive  $W$  for the gradient ascent with the sigmoid  $g$ . For the whole BS-net the "learning rule" is

$$\Delta W \propto (W^T)^{-1} + (1 - 2\bar{Y})\bar{X}^T \quad (3),$$

$$\Delta \bar{B} \propto 1 - 2\bar{Y} \quad (4).$$

Here, output-vector  $Y$  is obtained from input-vector  $X$ , which is transformed by matrix  $W$ , and then filtered by  $g$  after bias-vector  $B$  has been added:  $Y = g(WX+B)$ .

This input-output ( $X$ -to- $Y$ ) transformation, taken from summation-process in soma, has no direct relation to dendritic processing, although the results of dendritic processing shape the inputs  $X$  giving signals  $WX$  being transmitted to soma for  $g$ -weighted summation. *BS-algorithm parallels HBT with "infomax"-result, but its implementation is, it seems, quite artificial, or it incorporates conventional neuronal computing rather than HBT-advocated dendritic computing.* Primarily, the learning rule (3)+(4) is *non-local*. This means that neurons must know information about synaptic weights (or, in other cases, activities) of many distant neurons without being directly connected to them.

Bell & Sejnowski (1997, p. 3330) improved the rule (3)+(4), making it less non-local, using an idea of Amari *et al.* (see *ibid.*) for a more local adaptation of filter-matrix  $W$ :

$$\Delta W \propto \frac{\partial H(\bar{Y})}{\partial W} W^T W = (I + \hat{Y}\bar{U}^T)W \quad (5).$$

$I$  is the identity-matrix (having 1s on diagonal and zeros elsewhere), and

$$\hat{y}_i = \frac{\partial}{\partial y_i} \frac{\partial y_i}{\partial u_i} = \frac{\partial}{\partial u_i} \ln \frac{\partial y_i}{\partial u_i} \quad (6).$$



For this new BS-algorithm, the above comment on implementation is still valid. After the entropy gradient-ascent factor, the “auxiliary”  $W^T W$ -factor of eq. (5) gives a positive-definite matrix which speeds up the gradient ascent and enables avoiding the problematic matrix-inversion in eq. (3).

**Biological implementation of BS?** BS, of course, recognize that the visual processing from analog retinal photoreceptors to spike-coding pyramidal cells in V1 is approximated very roughly by the filter-matrix  $W$  and its generalized or pre-processed ( $W_Z$  “whitened”) versions  $W_I = W W_Z$  (for mathematical details see: Bell & Sejnowski, 1997, pp. 3330-3332). BS claim that their pre-processing of data has counterparts in retinal ganglion cells and LGN’s relay-cells.

If the pre-processed filter-matrix  $W$  gives *relatively biologically-plausible outputs* (or *filters*, or *receptive fields*, respectively), this does *not* necessarily mean that a  $W$ -like retino-geniculate-cortical transformation is *implemented in the living tissue*. The learning rules of BS, the new one in eqs. (5)+(6) almost as much as the previous one in eqs. (3)+(4), are both *non-local*. (Such) non-local processing is not necessarily biologically implausible, but has more danger of being so than local learning rules which get physiological support easier.

“Infomax” algorithms do not (yet) have biologically plausible implementation. However, van Hateren (1992), Wainwright (1999) and Porrill *et al.* (1999), among others, studied optimal coding of natural images, comparing model with physiological or psychophysical measurements, and found that “infomax” ideas are the most suitable for image processing.

#### 4 Sparseness Maximization By Olshausen and Field (OF)

**Sparse coding is related to statistical independence.** Bell & Sejnowski (1995, 1997) emphasize that *maximal preservation of information while reducing it to statistically-independent components* (primary focus of BS) is *very similar to maximizing sparseness of information-encoding* (primary focus of OF). Sparse coding means that a small percentage of neurons in the network is firing (is having non-zero activities), while the great majority of neurons is quiescent (has zero activity). The bulk of coding is thus realized by specialized minority of neurons, although the whole network participates in sorting the activity into such a restricted result or output (Olshausen & Field, 1996a,b). OF and BS achieve similar “infomax”-results using different methods and algorithms. Sparseness is much related to super-gaussian PD of input-data with positive kurtosis. Very sparse PD are more peaky and have longer tails than gaussian PD. The OF-versus-BS difference in trying to *maximize information* is that BS force the net to *minimize mutual information of outputs*, but OF force the net to *produce sparse output-codes*.

**Sparse-coding “infomax” by OF.** Like BF, OF-method also decomposes the image  $X$  into a linear superposition of basis functions  $Y$  weighted by amplitude-coefficients  $s$ , just like in eq. (2). Codes  $s$ , which in OF’s case do not necessarily represent *statistically-independent image-components*, but some basis-images of *very similar* type, are dynamic variables which change from one image to another (Olshausen & Field, 1996a,b).

OF's search for maximal sparseness of the codes  $s$  is done "globally and by force". They use minimization of the "cost function"  $E$  for regulation of learning:

$$E = -(\text{preserve information}) - \lambda (\text{code-sparseness}),$$

where  $\lambda$  is a positive constant determining the importance-rate of the sparseness-term relative to the first term. The first term  $PI$ , measuring preservation of information by optimal coding, incorporates the error or difference between the actual image  $X$  and the reconstructed image (sum of  $s$ -weighted components  $Y$ ):

$$PI = -\sum_{x,y} [X(x,y) - \sum_{k=1}^p s_k Y_k(x,y)]^2 \quad (7).$$

The difference should be as small as possible. If so,  $E$  is kept small, as desired.

The second term  $CS$ , sparseness of coding, "determines the cost" depending on how information is distributed among the codes  $s$ . If distribution is limited to few coefficients  $s$ , then  $CS$  will be big, and its contribution to total "cost"  $E$  will be small, as desired.

$$CS = -\sum_k S\left(\frac{s_k}{\sigma}\right)^2 \quad (8)$$

This sparseness-term to be maximized was constructed by OF to concentrate encoding to as few components as possible, "globally and by force", to increase sparseness.  $S$  is a chosen nonlinear function,  $\log(1+x^2)$  or something similar, and  $\sigma$  is a scaling constant. Optimal  $S$  should promote such a set of components  $Y$  with equal variance, which would have maximal number of zero coefficients  $s$  (i.e., codes which could be eliminated to economize processing).

**Implementation of OF.** The OF learning procedure goes as follows. After each image presentation, code-coefficients  $s$  are changed in such a way that  $E$  is minimized. After many image presentations,  $Y$  evolve towards a convenient final form which ensures maximally-sparse coding directly and statistically-independent coding indirectly in trade-off with requirement of minimal reconstruction-error.

Let us see details of iterative dynamics of codes  $s$ , which are similar to Gabor coefficients, and of components  $Y$ , which are similar to Gabor wavelets. Olshausen & Field (1996a,b) derived, from the requirement of gradient-descent of  $E$ , the equation of iterative adaptation of coefficients  $s_k$ :

$$\frac{\partial s_k(t)}{\partial t} = \eta \left[ c_k - \sum_j C_{kj} s_j(t) - \frac{\lambda}{\sigma} S' \left( \frac{s_k(t)}{\sigma} \right) \right] \quad (9)$$

where  $\eta$  is the learning-constant (bigger  $\eta$  makes learning faster); the auxiliary scalar-product coefficients  $c_k$  and  $C_{kj}$  are:

$$c_k = \sum_x \sum_y Y_k(x,y) X(x,y) \quad (10),$$

$$C_{kj} = \sum_x \sum_y Y_k(x,y) Y_j(x,y) \quad (11).$$

After several input-images  $X$  have been computed using eq. (9),  $Y_k$  are updated by making a stepwise decrease of  $E$  (gradient descent of  $E$ ):

$$\Delta Y_k(x_m, y_n) = \eta_w \left\langle [X(x_m, y_n) - X'(x_m, y_n)] s_k \right\rangle \quad (12).$$

$\eta_w$  is the learning-rate, and  $X'$  is the reconstructed image:  $X'(x_m, y_n) = \sum_k s_k Y_k(x_m, y_n)$ .

**Comparison OF - MacLennan.** There are similarities of OF-algorithm and the *dendritic field computation model* by MacLennan (in Pribram, 1993, ch. 6, p. 179). MacLennan started from the model of Daugman (1988) and considered a possibility to implement it in a dendritic net. Daugman at an early stage used just the first left-side term in eq. (9), i.e.  $\partial s_k(t)/\partial a = \eta c_k$ . Following Daugman, MacLennan then made improvement by using the first and the second (with  $C_{kj}$ ) left-side term of eq. (11). This linear model, possibly implementable in dendritic nets (as shown in: MacLennan in Pribram, 1993, ch. 6), is *equivalent to OF-model with exception of the sparseness-term*. OF invented CS-term (8) which makes their algorithm “more infomax” than the Daugman and MacLennan algorithms, focused on PI-term (7), are. We should pay attention to a possibility that the linear model with OF’s extension, which uses the sparseness-maximization (CS) term, could be *dendritically* implemented as MacLennan has shown for the information-preservation (PI) term.

**Biological implementation of OF?** OF claim that their learning algorithm can be implemented in a network with local connections. This is true, but the adaptation dynamics are regulated by a global requirement to minimize an abstract, artificial “cost”-parameter  $E$ . In the case of Hopfield’s neural nets,  $E$  is similar to physical energy. In physical systems, energy is minimized naturally. In complex and biophysical systems, free energy is minimized naturally. Brain, however, seems such a special biosystem, that tries to maximize efficiency by maximization of information (“infomax”), or equivalently, by reducing redundancy and information-loss during net-processing which transforms inputs to outputs. As OF show, this may be done by sparseness maximization. This seems plausible and natural in principle, but how this is concretely implemented in the brain tissue is not clear. Minimization of OF’s  $E$  is not so natural as in Hopfield’s case. It will remain an open question, whether brain uses such a global regulation, until anatomical and physiological research reveals more details on visual-brain’s structures and cooperation between them. The fact that the needed connectivity and processing is not nonlocal raise the possibility of being biologically implementable. A special sort of lateral inhibition could even be sufficient, but this is not necessarily the case at all. Other options include top-down constraining of rather specific type, since the form of OF’s  $E$  is quite specific, and this raises the question which higher visual centers could regulate it and how. Nevertheless, OF’s algorithm seems to have more chance to be some day found to be biologically plausible than those by BS. Namely, BS-algorithm is nonlocal, more complicated and more artificial than OF’s. BS derived their algorithm by purely mathematical means as abstract as factorization of probability distribution. OF, on the other hand, derived theirs by more natural and simple, it seems, means like sparse representation.

**Overcomplete coding.** OF-algorithm allows overcompleteness, i.e. the number of basis-vectors (memory-patterns) is greater than the dimensionality of input-images (Olshausen & Field, 1997; Olshausen & Millman, 2000). This is achievable by OF even with high degree of sparseness. OF’s basis-vectors need not be linearly independent. BS, in contrast, do not realize overcomplete representation (Bell & Sejnowski, 1997). Pribram (in Wang *et al.*, 1998) emphasizes physiological evidence on a tight-frame representation through oversampling.

## 5 Spatio-Temporal and Extended "Infomax"

**Natural image sequences.** Van Hateren & Ruderman (1998) simulated ICA in spatio-temporal domain. They successfully reproduced (van Hateren & van der Schaaf, 1998) experimental findings that V1 simple cells have receptive fields localized in space and time (Artun, Shouval & Cooper, 1998), bandpass in spatial and temporal frequencies, tuned in orientation, and selective for *direction of movement*. A Cray-supercomputer was used for running an BS-based ICA-algorithm by Hyvärinen & Oja (1997)<sup>2</sup> on sequences of natural images taken from television programs. Instead of detection of static edges, the simulation of van Hateren & Ruderman (1998) with input data-sets having spatial plus temporal dimensions reproduced spatio-temporal Gabor-like receptive fields responsive to *moving* edges. Corresponding edge-like, edge-detecting wavelets moved with fixed velocity perpendicularly to their own orientation-axis.

**Moving Gabors.** Beside adding temporal dimension to data-vectors and successfully reproducing V1-like "moving Gabor-wavelets" (i.e., moving sinusoids in a steady gaussian envelope), this simulation did not go essentially beyond algorithms by BS and by Hyvärinen & Oja (1997) - in the sense of satisfying the need for more biologically-plausible implementation, not merely output.<sup>3</sup> Filters focused on low spatial frequencies tend to be tuned to faster movements in natural-scene movies, in contrast to filters centered at high spatial frequencies. Van Hateren & Ruderman (1998) somewhat reinterpreted BS by emphasizing that the independent-component filters  $\phi$  (in rows of  $W$ ), and not the independent components  $Y$  (in columns of  $A$ ) which differ somewhat from filters, are comparable to V1 simple cells with their receptive fields.

**Extended "infomax".** T.-W. Lee (1998) made an extension of BS-algorithm and roughly reproduced the results of Bell & Sejnowski (1995, 1997). In the case of super-gaussian data, his ICA learning rule is:  $\Delta W \propto (I - th u u^T - u u^T) W$  (Lee, 1998, eq. 2.56). In the case of sub-gaussian data, the learning rule is the same, but the minus before  $th$  is replaced with a plus. Recall that sub-gaussian data were problematic for BS-net, but for Lee's extended ICA this is not so. Nadal & Parga (1997, sec. 4.5) show how "infomax" can be related to the BCM theory of synaptic plasticity<sup>4</sup>. It was also found that if data-bases are overcomplete to a higher degree, then input-data are processed more efficiently.

<sup>2</sup> The fast algorithm of Hyvärinen & Oja (1997), which has been used in spatio-temporal image simulations, realizes the "infomax" ideas by special statistical procedures, different from BS, which can be considered as a fixed-point algorithm for maximum likelihood estimation of the ICA data model. The new versions of this extended ICA algorithm are in: Hyvärinen (1999) and, applied to modeling complex cells of V1, in Hyvärinen & Hoyer (2000).

<sup>3</sup> New versions, like Hyvärinen (1999) and Hyvärinen & Hoyer (2000), bring improvement of computational efficiency and speed only. In latter paper, however, interesting new relations to V1-modeling were proposed.

<sup>4</sup> BCM-theory (original in: E.L. Bienenstock, L.N. Cooper & P.W. Munro (1982): *J. Neurosci.* 2, 32-48) bases on a nonlinear generalization of the Hebbian (covariance) rule with significant modifications. While the Hebbian rule prescribes strengthening of synaptic connection if presynaptic and postsynaptic activity are coactivated, BCM-model only does so if the coactivity exceeds a threshold, else the synapse is weakened. The threshold is variable and depends on the mean-activity level of the neuron.

## 6 Discussion on the Relevance of "Infomax" Models of Vision

**Outputs plausible.** Many different models with *different algorithms* (usually ICA-based), using different mathematical or statistical means and different computational "implementations", have succeeded to reproduce satisfactorily (for present criteria) the receptive fields of V1 cells in their static, spatio-temporal and complex variants.<sup>5</sup>

**Implementation presently implausible.** None of the models can be considered as satisfactorily biological in the sense of implementation. This is recognized by the authors of models themselves as well as their reviewers like Lee (1998, ch. 8). ICA models are currently less supported by physiological evidence than Hebbian models. This might change if more experimental attention will be paid to "infomax" models in the future, since they have more support of psychophysical experiments and information-theoretic studies (e.g., Wainwright, 1999; van Hateren, 1992; Porrill *et al.*, 1999) than Hebbian models. Maybe "infomax"-processing is realized on a "bio-software" level more than on a "bio-hardware" level as Hebbian models seem to be (at least more than "infomax"-models). This could be the reason why there is the unresolved problem of biological implementation of "infomax" models like ICA.

**Chances for biological support?** Phase-Hebbian models, including HNeT (Sutherland, 1990) and QAN (Peruš, 2000, in Wang *et al.*, 1998), combine the Hebb correlation-rule  $\sum_k A_i^k A_j^k$  with phase-differences  $S_i^k - S_j^k$  (because the complex-valued activities  $V = A \exp[\sqrt{-1} S]$  are correlated or convoluted). The phase-Hebbian learning rule  $\sum_k A_i^k A_j^k \exp[\sqrt{-1} (S_i^k - S_j^k)]$  does not contradict current physiological and psychophysical experiments and can support "infomax" by processing higher-order correlations. For early-vision problems (Poggio *et al.*, 1985) like edge-detection, a "phase-Hebbian infomax" compromise-model has more chance in getting physiological support than for higher-order problems like recognizing occlusions of objects.

**Neuropsychological consideration.** ICA seems to be a good model for image processing, but not necessarily for object perception (which is well distinguished from image processing by HBT and other models of vision, and this has psycho-physiological reasons). Object recognition, based on search for perceptual invariances (Hoffman, 1966, 1968, 1979; Caelli, 1976), might need a combination of ICA and associative processing (Haken, 1991), probably in attractor-nets which manifest gestalt-like structures. Phase-Hebbian processing might be the searched-for biologically-implementable associative "infomax algorithm" for higher image-processing *after natural images have been successfully preprocessed and abstracted (including extraction of edges) by an ICA-like "infomax algorithm"*. So, first stage would be ICA-like preprocessing of images, and then another (phase-Hebbian based?) associative / synergetic process would take over to recognize objects from images. Simulations by Bartlett & Sejnowski (1997) support this hypothesis.

Quantum nonlocality and coherence might help to diminish the problem of nonlocal ICA-processing, but not necessarily at all, because very specific information-access is needed in algorithms.

<sup>5</sup> Their information-theoretic meta-analyses begin to emerge (e.g., Lewicki & Olshausen, 1999).

## 7 Dendritic Computing and Synthesis With QAN and Attractor-Nets

**MacLennan's linear field model.** MacLennan (1999; in Pribram, 1993) has demonstrated how could parallel-distributed processing be realized in more-or-less continuous physical media so that information would be processed in analog rather than digital manner. Possible implementations of *field computation* in dendritic nets were systematically studied in MacLennan (in Pribram, 1993). Taking into account HBT, it is argued that function of axons is communication, but function of dendrites is computation. In spite of increasing awareness of importance of nonlinear processing (mainly at axon hillock), MacLennan (in Pribram, 1993, espec. pp. 170-172) shows precisely where linear approximation is not to be abandoned, especially not in dendritic nets. Since quantum processing is (basically) linear, MacLennan's model might have connections with quantum associative dynamics (Peruš, 1996, 1997, 1998, 2000).

**Possible relations with QAN.** Approximate *global* linearity of dendritic processing, in spite of many *local* (i.e., subcellular) nonlinear ingredients, indicates that the well-known model of Hopfield, and those by Haken (1991), are better implementable in dendritic nets than on neuronal level (meaning emphasis on processes in soma and axon). In addition to proposals by Hameroff, Penrose and others (like Marcer & Schempp, 1997), Peruš (2000) presented a purely-quantum implementation of the linear Hopfield model with phase-Hebbian learning rule (QAN).

After presenting a general set of dynamic equations for dendritic-field processing, MacLennan (in Pribram, 1993, sec. 6.1) presents a concrete linear algorithm based on Gabor wavelets as eigen-vectors (for our purposes: eigen-images). They are not orthogonal in general, but MacLennan (in Pribram, 1993, sec. 6.2) analyses an often-occurring case of optimal processing of "*nearly* orthogonal" data. Gabor wavelets are remarkably similar to certain wave-packets used in quantum theory. "Almost orthogonal" input-data, which are usually optimal for Hebbian-based associative processing and memory, can be used also in QAN-model, as proposed in Peruš (2000). It is shown there that such a "fuzzification" of quantum-mechanics- or Hebb-based information processing on a complete orthonormal set of eigenvectors is possible and useful in the case of an open system.

**Image processing plus object perception.** Bartlett & Sejnowski (1997) have quite successfully (and better than with PCA - see also Gray *et al.*, 1997) attempted to combine *ICA-based image processing* (V1-like) and *object-perception with attractor-networks* (ITC-like?) into an unified computational model. Specifically, results of ICA-preprocessing of input-images (in this case, faces from various views when a subject moves his head or we move around him) were put into attractor-neural-net, after passing a low-pass temporal filter. As a result of inter-spatial and inter-temporal Hebbian associations, faces neighboring in temporal sequence landed into the same attractor. Thus, view-independent categories of individual objects (faces) were formed as corresponding attractors - an attractor for each (head-moving) person.

In the attractor-net, Hebbian associations between Gabor-filtered proximal time-sequences made the neurons' tuning to specific facial poses more broad. Addition of lateral connections, however, achieved invariance to pose of a face. This pose-invariance was manifested by all single-person's poses "falling into" the same attractor.

One could speculate that *phase*-Hebbian learning would improve the results of such an encouraging integrative model which attempts to bridge the gap between image-perception, which is view-dependent, and object-perception which is view-invariant. QAN-implementation of such attractor-dynamics is also realizable.

## 8 Conclusions

“Infomax” preprocessing of images for further quantum-holography-like visual processing has been presented and discussed in the context of Pribram’s holonomic brain theory. Since holography is an universal process, the microstructure of junctional slow-wave potentials attributed to polarization fields in dendritic nets (Pribram, 1971, 1991) may have various implementations, but the basic one would be quantum. The QAN model could realize it *if Gabor wavelets would be used as QAN eigenstates  $\psi$*  to be processed as detailed in Peruš (2000; in Wang *et al.*, 1998). If this is the case, then *the neural brain serves as an encoder and decoder into and out of the Gabor representation*, and such ICA-produced Gabor wave-packets are processed in quantum interference dynamics (Ahn *et al.*, 2000). Detailed biological structures and processes (HNeT, ICA-like?) which maximize (*de*)coding performance (as ICA does) remain to be researched. In our model, the *core* of *conscious* visual information processing is of quantum nature, but modulated by classical neural processes (Peruš, 1997). Further computer-simulation-based research and detailed reports are forthcoming.

**Acknowledgements** to Dr. P.J. Marcer and to Professors K.H. Pribram, D.M. Dubois, R. Dawes, D. Raković, and to BCS-CMSG, asbl CHAOS, MZT and NIC for support.

## 9 References

- Ahn J., T.C. Weinacht, P.H. Bucksbaum (2000): *Science* **287**, 463-465.  
Artun Ö.B., H.Z. Shouval, L.N. Cooper (1998): *Proceed. Natl. Acad. Sci. USA* **95**, 11999-12003.  
Baddeley R. (1996): *Network: Computat. Neural Sys.* **7**, 409-421.  
Bartlett M.S., T.J. Sejnowski (1997): *Advances Neural Info. Proces. Sys.* **9**, 817-823.  
Bell A.J., T.J. Sejnowski (1995): *Neural Computat.* **7**, 1129-1159.  
Bell A.J., T.J. Sejnowski (1996): *Network: Computat. Neural Sys.* **7**, 261-266.  
Bell A.J., T.J. Sejnowski (1997): *Vision Res.* **37**, 3327-3338.  
Bonnell G., G. Papini (1997): *Int. J. Theor. Phys.* **36**, 2855-2875.  
Caelli T.M. (1976): *Math. Biosci.* **30**, 191-  
Comon P. (1994): *Signal Proces.* **36**, 287-314.  
Daugman J.G. (1985): *J. Opt. Soc. Amer. A* **2**, 1160-  
Daugman J.G. (1988): *IEEE Transac. Acoust., Speech & Signal Proces.* **36**, 1169-  
De Yoe E.A., D.C. Van Essen (1988): *Trends Neurosci.* **11**, 219-226.  
Gray M.S., J.R. Movellan, T.J. Sejnowski (1997): *Advances Neural Info. Proc. Sys.* **9**, 751-757.  
Haken H. (1991): *Synergetic Computers and Cognition*. Springer, Berlin.  
Harpur G.F., R.W. Prager (1996): *Network: Computat. Neural Sys.* **7**, 277-284.  
Hoffman W.C. (1966): *J. Math. Psychol.* **3**, 65-98.

- Hoffman W.C. (1968): *Quart. J. Appl. Math.* **25**, 423-.
- Hoffman W.C. (1979): *Math. Biosci.* **6**, 437-.
- Hyvärinen A. (1999): *IEEE Transac. Neural Net.* **10**, 626-634.
- Hyvärinen A., P. Hoyer (2000): *Advances Neural Info. Proces. Sys.* **12**, 827-833.
- Hyvärinen A., E. Oja (1997): *Neural Computat.* **9**, 1483-1492.
- Lee T.-W. (1998): *Independent Component Analysis*. Kluwer, Boston.
- Lewicki M.S., B.A. Olshausen (1999): *J. Opt. Soc. Amer. A* **16**, 1587-1601.
- MacLennan B.J. (1999): *Inform. Sci.* **119**, 73-89.
- Marcer P., W. Schempp (1997): *Informatica* **21**, 517-532.
- Marcer P., W. Schempp (1998): *Int. J. General Sys.* **27**, 231-248.
- Nadal J.-P., N. Parga (1997): *Neural Computat.* **9**, 1421-1456.
- Olshausen B.A., D.J. Field (1996a): *Network: Computat. Neural Sys.* **7**, 333-339.
- Olshausen B.A., D.J. Field (1996b): *Nature Lett.* **381**, 607-609.
- Olshausen B.A., D.J. Field (1997): *Vision Res.* **37**, 3311-3325.
- Olshausen B.A., K.J. Millman (2000): *Advances Neural Info. Proces. Sys.* **12**, in press.
- Peruš M. (1996): *Informatica* **20**, 173-183.
- Peruš M. (1997) in: M. Gams *et al.* (Eds.): *Mind Versus Computer*. IOS Press & Ohmsha, Amsterdam; pp. 156-170.
- Peruš M. (1998): *Zeitschr. angewan. Math. & Mech.* **78**, S1, 23-26.
- Peruš M. (2000): *AIP Proceedings of CASYS'99 – 3<sup>rd</sup> Int. Conf.*, Liege (B), vol. **517**, ed. by D. Dubois; American Institute of Physics, Woodbury (NY); pp. 289-295.
- Poggio T., V. Torre, C. Koch (1985): *Nature* **317**, 314-319.
- Porrill J., J.P. Frisby, W.J. Adams, D. Buckley (1999): *Nature Lett.* **397**, 63-66.
- Pribram K.H. (1971): *Languages of the Brain*. Orig.: Prentice-Hall, Englewood Cliffs (NJ); 5<sup>th</sup> publ.: Brandon, New York.
- Pribram K.H. (1991): *Brain and Perception*. Lawrence Erlbaum Assoc., Hillsdale (NJ).
- Pribram K.H. (Ed.) (1993): *Rethinking neural networks: Quantum fields and biological data*. Lawrence Erlbaum Assoc., Hillsdale (NJ).
- Pribram K.H. (1995): *Proceed. 39<sup>th</sup> Congress German Soc. Psychol.*, pp. 53-69.
- Pribram K.H. (1999): *Inform. Sci.* **115**, 97-102.
- Pribram K.H., E.H. Carlton (1986): *Acta Psychologica* **63**, 175-210.
- Pyllkkänen P., Pyllkö P. (Eds.) (1995): *New Directions in Cognitive Science*. Finnish AI Soc., Helsinki.
- Sutherland J.G. (1990): *Int. J. Neural Sys.* **1**, 259-267.
- Tootell R.B.H., N.K. Hadjikhani, J.D. Mendola, S. Marrett & A.M. Dale (1998): *Trends Cognit. Sci.* **2**, 174-183
- van Hateren J.H. (1992): *Nature* **360**, 68-70.
- van Hateren J.H., D.L. Ruderman (1998): *Proceed. Roy. Soc. Lond. B* **265**, no. 1412, 2315-2320.
- van Hateren J.H., A. van der Schaaf (1998): *Proceed. Roy. Soc. Lond. B* **265**, 359-366.
- Wainwright M. J. (1999): *Vision Res.* **39**, 3960-3974.
- Wang P.P. *et al.* (Eds.) (1998): *Proceed. 4<sup>th</sup> Joint Conf. on Information Sci. '98*. AIM, Research Triangle Park (NC). Volume II: *Proceed. 3<sup>rd</sup> Int. Conf. on Comput. Intellig. & Neurosci.* (Ed. G. Georgiou); sections on neuro-quantum information processing; pp. 167-224.
- Zak M., C.P. Williams (1998): *Int. J. Theor. Phys.* **37**, 651-684.